Title page

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 Ebisawa[1], Kaori Tanaka[1], Tetsutaro Hayashi[1], Akira Kurisaki[2*], and Itoshi Nikaido[1,3*]

[†]These authors contributed equally to this work.

*Corresponding author

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

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

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

Yohei Sasagawa: youhei.sasagawa@riken.jp

Hiroki Danno: hiroki.danno@riken.jp, redgrapefruit@mac.com

Hitomi Takada: htakada@bs.naist.jp

Masashi Ebisawa: support-bit@riken.jp, masashi.ebisawa@riken.jp

Kaori Tanaka: kaori.tanaka@riken.jp

Tetsutaro Hayashi: tetsutaro.hayashi@riken.jp

Akira Kurisaki: akikuri@bs.naist.jp

Itoshi Nikaido: itoshi.nikaido@riken.jp

Abstract

High-throughput single-cell RNA-seq methods assign limited unique molecular identifier (UMI) counts as gene expression values to single cells from shallow sequence reads and detect limited gene counts. We thus developed a high-throughput single-cell RNA-seq method, Quartz-Seq2, to overcome these issues. Our improvements in the reaction steps make it possible to effectively convert initial reads to UMI counts (at a rate of 30%–50%) and detect more genes. To demonstrate the power of Quartz-Seq2, we analyzed approximately 10,000 transcriptomes in total from *in vitro* embryonic stem cells and an *in vivo* stromal vascular fraction with a limited number of reads.

Keywords

High-throughput single-cell RNA-seq

Flow cytometry

Cell sorter

Quartz-Seq

Poly-A tagging

Stromal vascular fraction

Mesenchymal stem cell

Background

Single-cell transcriptome analysis is a powerful tool to identify nongenetic cellular heterogeneity, which includes differences in cell type due to differentiation and differences in cell state within a cell population. In previous studies, various methods for single-cell RNA-seq were developed [1-19]. Some of these that generate read coverage across all transcripts have been exploited to detect alternative transcription splicing isoforms [14,17], and others using a unique molecular identifier (UMI) have been applied to quantify the number of transcripts expressed in a cell [1-3,6,7,9,11-13,20]. To extract substantial information on a cell population, such as the composition of different cell types or the distribution of cell states, it is necessary to analyze hundreds or thousands of cells. Cell barcoding is a key technology for this, which enables us to deal with samples from numerous cells in a single tube. Cell barcoding technology, which tags nucleotides unique to each cell to target RNA molecules from that cell, is a key technology for increasing the throughput of single-cell RNA-seq [16,18]. Mixing cDNA tagged with cell barcodes before whole-transcript amplification decreases the cost of reaction reagents and the laboriousness of experimental steps. There are two types of cell barcoding technology according to the method of cell sampling used. One method involves single cells being selectively sorted to multi-well plates using flow cytometry, which allows us to remove dead or aggregated cells. Besides, transcriptome data can be linked to cellular information obtained by flow cytometry. The other method involves single cells and barcoded beads being captured in water-in-oil droplets using droplet-generation microfluidic devices [6,7]. In this latter method, thousands of cells can probabilistically be captured in half an hour. However, the total number of sequence reads generated by a deep sequencer is still limited. To increase the number of analyzed cells, each cell is assigned a limited number of initial sequence reads. For example, approximately 400 million initial fastq reads were sequenced for 3,000–4,000 cells in several previous studies [6,7]. In this case, the input data size for a single cell involves shallow initial reads (100,000 fastq reads) (Figure 1b). UMI counts were converted from shallow initial sequence reads for each cell, and the conversion ratio of that was limited to approximately one-tenth [21]. Ideally, greater UMI counts should be generated from limited sequence reads because the increase in UMI count

assigned to each cell leads to the detection of low-copy genes and the identification of cell-type-specific genes using statistical tests.

In this study, we developed a novel high-throughput single-cell RNA-seq method, Quartz-Seq2. As Quartz-Seq is a sensitive and reproducible single-cell RNA-seq method, Quartz-Seq2 was developed based on it [15]. Quartz-Seq is based on a poly-A tagging strategy. By the combination of molecular biological improvements including major improvement of poly-A tagging, Quartz-Seq2 resulted in an increase in the effectiveness with which the initial sequence reads were converted to the expression UMI counts (UMI conversion efficiency: 30%–50%). To demonstrate the highly effective use of initial reads in Quartz-Seq2, we analyzed a population of approximately 9,000 mouse embryonic stem (ES) cells as *in vitro* cells and approximately 1,000 cells from the stromal vascular fraction (SVF) as *in vivo* cells.

Results

Outline of Quartz-Seq2 experiment

To increase the UMI conversion efficiency (Figure 1), we improved several steps in the preparation of the single-cell RNA-seq library (Additional file 1: Figure S1), resulting in the development of Quartz-Seq2. Below, we explain the five steps of the Quartz-Seq2 procedure (Figure 1).

(1) The first step is single-cell collection using a cell sorter. We selectively sort living single cells into lysis buffer in a 384-well PCR plate without dead cells. In cell sorting, various types of channel information, such as the intensity of fluorescence, are obtained for each cell. This enables us to link the transcriptome to cellular information from a cell sorter for each cell.

(2) The second step is cell barcoding. Each well contains lysis buffer and reverse-transcription (RT) primer, which includes a cell barcode sequence (14- or 15-mer), a UMI sequence (8-mer), and an oligo-dT sequence (24-mer). Using these RT primers, respective RNA from single cells is converted to cDNA with unique cell barcodes. Note that a long RT primer resulted in a severe problem regarding the synthesis of byproducts at the downstream reaction in our system (Additional file 1: Figure S2 and Supplemental Note). Therefore, we use a relatively short RT primer (73- or 74-mer), which allows us to skip the step of removing byproducts using exonuclease I. We design two types of RT primer set (v3.1: 384 barcodes; v3.2: 1536 barcodes). Within each primer set, barcode sequences are designed such that the minimum Sequence–Levenshtein distance between two sequences should be greater than 5, which leads to the correction of mutations of two nucleotides, including substitution, insertion, or deletion in sequence reads [22]. We also optimize the buffer and temperature in the RT reaction, leading to an improvement of RT efficiency from that of the original Quartz-Seq (Additional file 1: Figure S3 and Supplemental Note). We apply a low enzyme concentration in reverse transcription to Quartz-Seq2. These conditions reduce the cost and technical variability of Quartz-Seq2. We describe the details of this in the subsection entitled “Reduction of enzyme concentration in reverse transcription decreased the experimental cost of Quartz-Seq2”.

(3) The third step involves the pooling of cell-barcoded cDNA. By cell barcoding, Quartz-Seq2 can pool cDNA of up to 1,536 individual cells into one mixture. We developed a

174 rapid and high-throughput method for collecting small volumes of cDNA in multiwell plates. This
175 method also achieved higher efficiency of collection than dispensing with pipettes and tips
176 (Additional file 1: Figure S4 and Supplemental Note). As the efficiency of cDNA purification after
177 pooling was 93.77%, we estimated that approximately 80% of cell-barcoded cDNA could be
178 used for subsequent whole-transcript amplification in our system (Additional file 1: Figure S3a
179 and Supplemental Note).
180 (4) The fourth step is whole-transcript amplification based on an improved poly-A tagging
181 strategy. Poly-A tagging is one of the methods of converting first-strand cDNA to amplifiable
182 cDNA. First-strand cDNA is extended with a poly-A tail by the terminal transferase. Subsequently,
183 second-strand cDNA is synthesized with tagging primers that contain a poly-dT sequence,
184 followed by PCR amplification. Here, we improve the efficiency of poly-A tagging by 3.6-fold.
185 This improvement is a crucial point in the development of Quartz-Seq2. We describe the details
186 of this in the next subsection.
187 (5) The fifth step is library preparation for deep sequencing. Amplified cDNA is fragmented,
188 ligated with the sequence adapter, and amplified by PCR. In sequencing using Illumina
189 sequencers, a sequence for a cell barcode and a UMI is read in Read1, while a sequence for a
190 region of a transcript (mRNA) is read in Read2.
191
192 **Improvement of poly-A tagging efficiency**
193 We previously reported Quartz-Seq based on the poly-A tagging strategy, which has significant
194 potential for detecting a large number of genes expressed in a cell [15]. However, the efficiency
195 of poly-A tagging itself for single-cell RNA-seq has not been improved. We hypothesized that the
196 improvement of poly-A tagging would lead to high UMI conversion efficiency. Therefore, we
197 attempted to improve the efficiency of this tagging step. Poly-A tagging is composed of two
198 processes: (1) The first-strand cDNA is modified with a poly-A tail by terminal deoxynucleotidyl
199 transferase. (2) Next, the poly-A-tailed cDNA is annealed with a tagging primer, which has a
200 poly-dT sequence. Then, the second-strand cDNA is extended. The resulting second-strand
201 cDNA is amplifiable cDNA, which has a PCR primer sequence at both ends of it.

202 It is known that the DNA yield of amplified cDNA generally reflects the quantitative
203 performance of single-cell RNA-seq methods [3,13,14]. Thus, we determined the effects of
204 various buffers for the poly-A tailing step on the amplified cDNA yield. We performed the poly-A
205 tailing reaction with various buffers using purified first-strand cDNA from 1 ng of total RNA
206 (Figure 2a, Additional file 1: Figure S5). Finally, we obtained amplified cDNA. We found that the
207 use of T55 buffer in the poly-A tailing reaction efficiently improved the cDNA yield (Figure 2a).
208 The amount of amplified cDNA increased 2.88-fold using T55 buffer compared with the level
209 using Quartz-Seq buffer (Figure 2a). In these buffer conditions, we did not observe any obvious
210 byproducts derived from the RT primer (Figure 2a, Additional file 1: Figures S2 and S5, and
211 Supplemental Note).
212 Next, we added an “Increment” temperature condition for the tagging and second-strand
213 synthesis steps (see Methods). In this condition, the reaction temperature of these steps was
214 gradually increased. As a result, the amount of cDNA tended to increase, by approximately
215 1.2-fold (Figure 2a). Moreover, upon combining T55 buffer and the Increment condition, the
216 amount of cDNA increased approximately 3.6-fold. We also confirmed the reproducibility of this
217 phenomenon of cDNA increment in additional experiments (Additional file 1: Figure S5).
218 Moreover, we confirmed the amplified cDNA yield of various genes by qPCR analysis as another
219 assay. Specifically, we determined the qPCR scores of eight genes from amplified cDNA and
220 nonamplified cDNA (Additional file 1: Figure S5c). Spearman’s rank correlation coefficients
221 (SCCs) between amplification and nonamplification were approximately 0.79 in the condition
222 (T55+Increment). The SCC was approximately 0.66 in Quartz-Seq-like conditions. We also
223 observed clear increments of qPCR scores for almost all genes. These results show that the
224 combination of T55 buffer and this temperature condition improved the efficiency of the poly-A
225 tagging step. We also found that other conditions (NBF40+Increment) improved the cDNA yield.
226 However, under these conditions, byproducts were clearly synthesized (Additional file 1: Figures
227 S2c and S5b). Moreover, the amount of cDNA with T55 buffer was slightly greater than that with
228 RH55 (Figure 2 and Additional file 1: Figure S5a). Therefore, we used the combination of T55
229 buffer and the “Increment” temperature condition for the poly-A tagging strategy for
230 Quartz-Seq2.

231

232 **Reduction of enzyme concentration in reverse transcription decreased the experimental**
233 **cost of Quartz-Seq2**

234 The cost of experiments for the single-cell RNA-seq method is one of the most important
235 benchmarks regarding high-throughput performance. The cost of experimental preparation per
236 cell was approximately ¥2,600 ($23) for our previously reported Quartz-Seq, which does not use
237 cell barcoding (Additional file 1: Figure S6a). To improve on this value, we first applied the
238 "RT100" enzyme condition in reverse transcription to Quartz-Seq2. In this condition, we used
239 approximately 20U reverse transcriptase in 2 μL of solution for reverse transcription. This
240 enzyme concentration in the RT reaction is broadly used for various molecular biological
241 applications including single-cell RNA-seq methods [3,14,15]. By using the cell barcoding
242 strategy, the cost of experimental preparation for Quartz-Seq2 under the "RT100" condition was
243 reduced (¥122 or $1.08 per cell) (Additional file 1: Figure S6a).

244 We found that 65% of the cost of experimental steps is derived from reverse transcription
245 in Quartz-Seq2 under the "RT100" condition (Additional file 1: Figure S7a). To further reduce the
246 cost of experimental preparation of sequence libraries on a large scale, we investigated the
247 effect of a low enzyme concentration in reverse transcription in Quartz-Seq2. In the assessment
248 assay, we noticed that a low enzyme concentration did not markedly affect the efficiency of
249 reverse transcription in T100 buffer (Additional file 1: Figure S7b). We performed a similar
250 experiment with a broader range of concentrations of enzymes in reverse transcription under
251 conditions with T100 buffer (Figure 2b and Additional file 1: Figure S7c). We found that the RT25
252 condition maintained the efficiency of reverse transcription at the 99% level on average, which
253 was comparable to that in the RT100 condition. Therefore, we prepared three technical
254 replicates of a 384-well PCR plate with 10 pg of total RNA with the "RT100" condition or the
255 "RT25" low-enzyme condition. In the "RT25" condition, the cDNA yield showed a tendency for a
256 slight increase of approximately 1.17-fold, which although not being a major improvement did at
257 least not involve a decrease (Additional file 1: Figure 7d). The cost of experimental preparation
258 per cell became approximately ¥46–¥63 ($0.40–$0.56) in the "RT25" condition (Figure S6a). We
259 thus mainly used the "RT25" condition for Quartz-Seq2.

260

261 **Evaluation of the quantitative performance of Quartz-Seq2 using 10 pg of purified total**
262 **RNA**

263 Finally, we adopted three molecular biological improvements [poly-A tailing buffers (T55), the
264 "Increment" temperature condition, and low-enzyme concentration (RT25) in reverse
265 transcription] for Quartz-Seq2 (Additional file 1: Figure S1). To determine the technical variability
266 and specificity of Quartz-Seq2, we performed whole-transcript amplification using 10 pg of
267 diluted mouse total RNA as a single-cell-like averaged sample with the v3.1 384 cell-barcode RT
268 primer in a 384-well plate (Figure 2). In this experiment, we used the "RT25" enzyme
269 concentration for Quartz-Seq2. The effect of RT enzyme concentration on the quantitative
270 performance is specifically described in the last paragraph in this subsection of the paper. We
271 analyzed 10 pg of total RNA in all wells at approximately 0.19 M fastq reads on average per well.
272 In the case of the Quartz-Seq-like conditions, we detected 18,407±4,040 UMI counts and
273 5,728±604 gene counts (n=768 wells from two 384-well plates). In the case of Quartz-Seq2, we
274 achieved a UMI count of 44,100±7,521 and a gene count of 7,442±484 (n=1,152 wells from three
275 384-well plates)(Figure 2c). We observed similar results at the level of individual 384-well plates
276 (Additional file 1: Figure S8a). We also calculated the UMI conversion efficiency for the
277 respective protocol at approximately 0.19 million initial fastq reads on average per well. We
278 defined the formula for calculating the "UMI conversion efficiency," which indicates how
279 effectively initial fastq reads can be converted to UMI counts (Figure 1c and Additional file 1:
280 Figure S9). The UMI conversion efficiency levels of Quartz-Seq2 and Quartz-Seq-like were
281 about 22.88% and 9.55%, respectively. The UMI conversion efficiency was reproducible among
282 individual 384-well plates (Additional file 1: Figure S8a). In addition, the UMI conversion
283 efficiency depended on the Read2 length and the number of initial fastq reads (Additional file 1:
284 Figure S8b). These results indicated that the combination of the molecular biological
285 improvements for Quartz-Seq2 clearly improved the conversion ratio from target RNA to
286 sequence library DNA.

287 We also validated the technical reproducibility of gene expression. In the mean-CV
288 (coefficient of variation) plot, technical gene expression variability of Quartz-Seq2 became fairly

289 close to the theoretical variability of a Poisson distribution (Figure 2d). The mean of gene
290 expression variability for Quartz-Seq2 was lower than that for the Quartz-Seq-like method
291 (Figure 2f). To quantify the technical reproducibility of internal gene expression and external
292 gene (ERCC spike-in RNA) expression, we used a pairwise comparison of technical replicates
293 for each protocol. Regarding the reproducibility of internal gene expression, the SCCs of
294 Quartz-Seq2 and Quartz-Seq-like were 0.71±0.01 and 0.59±0.02, respectively. Regarding the
295 reproducibility of external gene expression, the SCCs of Quartz-Seq2 and Quartz-Seq-like were
296 0.90±0.02 and 0.83±0.04, respectively. Subsequently, we validated the accuracy of
297 internal/external gene expression. We observed that the average internal gene expression of
298 Quartz-Seq2 highly correlated with the internal gene expression of conventional RNA-seq
299 (Figure 2e). We calculated all combinations of pairwise correlation between the internal gene
300 expression of Quartz-Seq2 with 10 pg of total RNA and that of conventional RNA-seq with 1 μg
301 of total RNA. Moreover, we calculated the pairwise correlation between the external gene
302 expression of Quartz-Seq2 and the input molecule count of external genes (ERCC spike-in RNA).
303 Regarding the accuracy of internal gene expression, the SCCs of Quartz-Seq2 and
304 Quartz-Seq-like were 0.70±0.01 and 0.60±0.02, respectively. Regarding the accuracy of external
305 gene expression, the SCCs of Quartz-Seq2 and Quartz-Seq-like were 0.90±0.01 and 0.74±0.03,
306 respectively. These results indicate that the combination of the molecular biological
307 improvements for Quartz-Seq2 clearly improved the technical reproducibility and accuracy of
308 gene expression.

309 We noted that a low enzyme concentration improved the quantitative performance of
310 Quartz-Seq2. We validated the quantitative performance with 10 pg of total RNA in the “RT100”
311 and “RT25” conditions at various input data sizes. Unexpectedly, we found that the “RT25”
312 condition improved the quantitative performance (Additional file 1: Figure S8a). We compared
313 RT25 with RT100 at approximately 0.096 M fastq reads on average per well. In the case of
314 Quartz-Seq2 in the “RT25” condition, we achieved a UMI count of 30,117±789 and a gene count
315 of 6,320±35 (three 384-well plates). In the case of Quartz-Seq2 in the “RT100” condition, we
316 achieved a UMI count of 25,671±1020 and a gene count of 5,889±35 (three 384-well plates). We
317 also observed that the well-to-well technical variability for the UMI count and gene count clearly

318 decreased in the “RT25” condition (Additional file 1: Figure S8a). These results showed that the
319 “RT25” low-enzyme condition clearly reduced the experimental cost and improved the
320 quantitative performance. We thus applied the “RT25” condition in subsequent experiments
321 using real single cells. Note that, in the RT1.5625-RT6.25 condition, reverse-transcription
322 efficiency was maintained at an average level of over 90% (Figure 2b). The average
323 reverse-transcription efficiency rapidly decreased below the RT1.56 low enzyme concentration.
324 In addition, in actual experiments (not serially diluted experiments) at conditions below RT3.12, it
325 was not guaranteed that enzymes could be collected given the viscosity and low volume of the
326 mixture. Therefore, we validated the RT6.25 condition for subsequent analysis. It seems that
327 UMI counts, gene counts, and ERCC capture efficiency (also called as “ERCC spike-in RNA
328 detection efficiency”) slightly increased in the RT6.25 condition, which although not being a
329 major improvement did at least not involve a decrease (Additional file 1: Figure S10). The cost of
330 experimental preparation per cell became approximately ¥31–¥48 ($0.27–$0.43) in the “RT6.25”
331 condition (Additional file 4: Table S3, Additional file 1: Figure S6a). The “RT6.25” low-enzyme
332 condition thus further reduced the experimental cost and slightly improved the quantitative
333 performance.
334
335 **Quartz-Seq2 shows higher efficiency of UMI conversion and detects more biological**
336 **pathways than Drop-seq**
337 In the high-throughput single-cell RNA-seq methods, the total number of sequence reads
338 generated by a deep sequencer is limited. Ideally, greater UMI counts should be generated from
339 limited sequence reads because the increase in UMI count assigned to each cell leads to the
340 detection of low-copy genes and the identification of cell-type-specific genes using statistical
341 tests. We compared Quartz-Seq2 to the high-throughput single-cell RNA-seq method using two
342 distinct cell types. To prepare these two distinct cell types, we cultured G6GR ES cells and
343 Dex-treated G6GR cells because it has been reported that almost all G6GR ES cells
344 differentiated into primitive endoderm-like cells upon dexamethasone treatment [15,23].
345 Drop-seq is one of the high-throughput single-cell RNA-seq methods, which can capture
346 thousands of cells and barcoded beads in half an hour [6]. We performed Drop-seq experiments,

347 which were validated by species-mixing analysis (Figure S11a, b). We performed Quartz-Seq2
348 and Drop-seq on a mixture of mouse ES cells and Dex-treated mouse ES cells [primitive
349 endoderm (PrE) cells] and calculated the UMI conversion efficiency (Figure 3a and Additional file
350 1: Figure S11c). The effectiveness of Quartz-Seq2 ranged from 25% to 35% depending on the
351 initial fastq read depth (Figure 3), which was higher than that of Drop-seq.
352 As Figure 3a shows, Quartz-Seq2 detected more genes with high UMI conversion
353 efficiency. To examine the power for identifying differentially expressed genes, we performed
354 principal component analysis (PCA) and clustering. We randomly selected single cells from ES
355 cluster cells and PrE cluster cells. The number of genes differentially expressed between two
356 distinct cell types was calculated (ES and PrE cells) (Additional file 1: Figure S11h). We also
357 identified differentially expressed genes (DEGs) between bulk ES cells and that PrE cells by
358 using conventional bulk RNA-seq. We then counted the number of overlapping DEGs between
359 single-cell RNA-seq methods and bulk RNA-seq. We observed that this number linearly
360 correlated with the cell number (Figure 3b). Quartz-Seq required fewer cells for the detection of
361 overlapping DEGs. These results showed that more genes that were differentially expressed
362 between the two cell types were identified with Quartz-Seq2 (Figure 3b, c). In addition, more
363 biological pathways particularly associated with the differentially expressed genes were detected
364 in Quartz-Seq2 (Figure 3c). Furthermore, we calculated the highly variable expression of genes
365 in Quartz-Seq2 and Drop-seq, which potentially include not only genes that are differentially
366 expressed between cell types but also genes of which the expression changes depending on the
367 cell state in a cell type. Terms related to the cell cycle state were only enriched for the genes
368 calculated with Quartz-Seq2 (Figure 3d). Note that simulation-based power analysis also
369 showed that Quartz-Seq2 detected more DEGs than Drop-seq (Additional file 1: Figure S12a,
370 see Methods). These results suggest that high UMI conversion efficiency with limited initial reads
371 leads to more biological information being revealed, such as functional terms and biological
372 pathways.
373
374 **Superiority of Quartz-Seq2 regarding quantitative performance in the same experimental**
375 **design compared with other methods**

376 To obtain additional evidence for the superiority of Quartz-Seq2 in terms of the UMI conversion
377 efficiency and gene count, we compared it to other methods. Using mouse J1 ES cells,
378 Ziegenhain et al. systematically compared the quantitative performance of several single-cell
379 RNA sequencing (RNA-seq) methods [CEL-seq2(C1), SCRB-seq, MARS-seq, and Drop-seq]
380 that use the UMI technique [21]. Therefore, we cultured J1 ES cells under 2i/LIF conditions in
381 accordance with the procedure described in this previous paper [21]. We sorted J1 ES cells into
382 five 384-well plates. Subsequently, we prepared sequence library DNA of Quartz-Seq2 (RT25)
383 with 1,152 wells (three 384-well plates) or 768 wells (two 384-well plates) on different days. For
384 comparison between our data and those obtained in the previous study mentioned above, we
385 used the same analytical conditions as previously applied, such as a HiSeq sequencer platform,
386 Read2 length, the same genome file, and the same transcript annotation file.
387 First, we compared the quantitative performance between Quartz-Seq2 and other
388 methods at 0.1 million initial fastq reads on average per cell (Figure 4). The results indicate that
389 the UMI conversion efficiency levels of Quartz-Seq2 were approximately 32.55% (Day 1) and
390 32.25% (Day 2); the UMI conversion efficiency of the other methods ranged from 7.11% to
391 22.45%. The average gene counts using Quartz-Seq2 were approximately 6,636 (Day 1) and
392 6,584 (Day 2), while the average gene counts for the other methods ranged from 2,738 to 5,164.
393 We also validated the quantitative performance for external control RNA. The levels of ERCC
394 capture efficiency for Quartz-Seq2 were approximately 6.12% (Day 1) and 6.38% (Day 2), while
395 the ERCC capture efficiencies for the other methods ranged from 0.76% to 3.22% (Figure 4a
396 and Additional file 1: Figure S13d). We also calculated the copy number of the detection limit of
397 ERCC spike-in RNA. The detection limits for Quartz-Seq2 were approximately 6.82 (Day 1) and
398 6.57 (Day 2), while those for the other methods ranged from 13.26 to 710.24 (Additional file 1:
399 Figure S13d).
400 We also performed simulation-based power analysis to estimate the power for identifying
401 differentially expressed genes (Additional file 1: Figure S12b, see Methods). The true positive
402 ratio (TPR) at the data point of 256 cells ranged from 0.93 and 0.95, and the false discovery rate
403 (FDR) at the data point of 256 cells ranged from 0.07 to 0.09. Therefore, TPR and FDR for each

404 method were comparable to each other. We found that Quartz-Seq2 detected more simulated
405 DEGs than the different methods.
406 In particular, the UMI conversion efficiency and gene count of the Quartz-Seq2 method
407 were significantly better than those of the other methods at approximately 0.1 million initial reads.
408 We also estimated the UMI and gene counts and the UMI conversion efficiency at various
409 numbers of initial fastq reads (Figure 4a). We found that Quartz-Seq2 is greatly advantageous
410 for detecting the UMI and gene counts from limited initial amounts of data (under 0.2 million fastq
411 reads). These results showed that the quantitative performance of Quartz-Seq2 was almost
412 always better than that of other methods under conditions with a limited number of initial fastq
413 reads.
414
415 **Quartz-Seq2 achieves high UMI conversion efficiency at relatively low cost**
416 The experimental cost for sequence library preparation is an important benchmark for single-cell
417 RNA-seq methods because it is highly correlated with the throughput of single-cell RNA-seq
418 methods within a limited budget. In the case of using 384 indexes, the library preparation cost of
419 Quartz-Seq2 (RT25) is ¥63 ($0.56) (Additional file 1: Figure S6a). For Quartz-Seq2 with 384
420 indexes, we used three purification columns per 384-well plate at the step of cDNA purification.
421 To reduce the amount of enzyme solution in the downstream reaction after cDNA purification, we
422 used one purification column per 384-well plate (Additional file 4: Table 3). In the previous
423 subsection, we also described that the “RT6.25” low-enzyme condition improved the
424 experimental cost of Quartz-Seq2. The combination of the “RT6.25” low-enzyme condition and
425 reduction of the number of purification columns thus further improved the library preparation cost
426 of Quartz-Seq2 (RT6.25+column1) to ¥37 ($0.32) (Additional file 1: Figure S6a). For comparison,
427 the library preparation costs of other methods that use a cell sorter range from ¥146 ($1.30) to
428 ¥372 ($3.29) (Figure S6a). Regarding the cost, Quartz-Seq2 is thus highly competitive with other
429 single-cell RNA-seq methods that use a cell sorter.
430 To show the potential of Quartz-Seq2 for high-throughput performance, we additionally
431 carried out Quartz-Seq2 (RT6.25+column1) and Quartz-Seq2 (RT6.25) using J1 ES cells (Figure
432 4b). We estimated that the number of processable single cells per 1,000 US dollars spent on

433 Quartz-Seq2 ranges from 1,785 to 3,058 (Figure 4b). We also calculated the total cost, including
434 the sequencing and library preparation costs, for each method. The total cost per cell is
435 extremely low (approximately ¥71–¥131, or $0.63–$1.16) for Drop-seq. The total cost per cell is
436 ¥97–¥183 ($0.85–$1.62) for Quartz-Seq2. Thus, the total cost of Quartz-Seq2, a
437 cell-sorter-based method, approaches that typical of droplet-based methods (Additional file 1:
438 Figure S6b). In summary, the Quartz-Seq2 method achieves the highest UMI conversion
439 efficiency and high-sensitivity detection of genes under conditions with limited numbers of initial
440 reads, providing single-cell RNA-seq data in a high-throughput manner.
441
442 **Demonstration of high-throughput Quartz-Seq2 analysis on 4,484 mouse embryonic stem**
443 **cells and differentiated conditions**
444 To demonstrate the capability of Quartz-Seq2 for quantifying the transcriptome of a large
445 number of cells and identifying rare cell populations, we analyzed approximately 4,484 cells from
446 a mixture of mouse ES cells and differentiation-induced cells. Cells were prestained with
447 Hoechst 33342 and/or Calcein-AM as an indicator of DNA content and culture condition (ES or
448 PrE cells), respectively (Figure 5a). Calcein-AM-positive and -negative cells were sorted to
449 twelve 384-well plates (4,608 well in total) in a checkered pattern (Figure 5a). These conditions
450 maximize the likelihood of detecting the cell doublets caused by mis-sorting. In this analysis, as
451 an average for cells, 0.1 M initial fastq reads were effectively converted to 35,915±1,188 UMI
452 counts in a final gene expression matrix. The UMI conversion efficiency was 35.91±1.18%.
453 Dimensionality reduction of the expression UMI matrix resulting from Quartz-Seq2 showed
454 clear separation of six clusters, including the main populations of ES cells and PrE cells, as well
455 as four small populations. We did not observe a clear batch effect among the twelve 384-well
456 plates (Additional file 1: Figure S15). One cluster contained a relatively high proportion of
457 mitochondrial RNA (Additional file 1: Figure S16), which was judged to reflect low-quality cells
458 [24]. Another cluster showed high values of detected UMI counts and gene counts, and
459 expressed both ES and PrE marker genes, which were judged to represent doublets of an ES
460 cell and a PrE cell (Additional file 1: Figure S16). To characterize the identified populations, we
461 determined the genes that were specific or enriched for each cluster using binomial tests (Figure

462 5c). Although cells in cluster 3 shared gene expression with the main population of ES cells
463 (cluster 1), some genes including *Stmn2* and *Rhox5* were additionally expressed, suggesting
464 that cluster 1 and cluster 3 shared the same cell type but had different cell states. Cluster 5 was
465 a small population (0.17%, Figure 6b), but was characterized by many specific marker genes
466 (Figure 5c), including *Zscan4c* and *Zscan4d*, which are known to be expressed in a
467 subpopulation of mouse ES cells [25]. However, among our culture conditions, these genes were
468 more expressed in the Dex-treatment condition than in the ES-maintenance condition. Flow
469 cytometry information on cells in cluster 5 also showed no fluorescence of Calcein-AM,
470 indicating that these cells were from the suspension of Dex-treated cells (Figure 5d). These
471 observations were consistent with a previous study demonstrating that *Zscan4* was more
472 expressed in a differentiation condition induced by the withdrawal of leukemia inhibitory factor
473 [7].
474 In the sampling of cells using a cell sorter, flow cytometry information including the
475 intensity of Hoechst 33342 staining for each cell was collected. Using this information, we
476 reconstructed the distribution of Hoechst 33342 staining intensity for each cluster (Figure 5e).
477 The main population of ES cells (cluster 1) showed an embryonic pattern of distribution of DNA
478 content, which is characterized by a high ratio of cells in G_2 and M phases to cells in the G_1
479 phase. In contrast, the main population of PrE cells (cluster 4) showed a somatic pattern of
480 distribution of DNA content. This was consistent with PrE cells being differentiated from ES cells.
481 The cells in cluster 3, the gene expression of which was similar to that of the ES main population
482 (cluster 1), showed an embryonic pattern rather than a somatic one. As the number of cells in
483 cluster 5 that expressed *Zscan4c/d* was small, it was difficult to classify the observed pattern as
484 the embryonic or somatic type. These findings indicated the usefulness of single-cell RNA-seq
485 using flow cytometry for the reconstruction of population information after transcriptome-based
486 clustering. For good interpretation of the distribution, a large number of cells for each cluster are
487 required.
488
489 **Quartz-Seq2 classified genes with variable expression within a cell type**

490 When cells were plotted on t-SNE (t-distributed stochastic neighbor embedding) space using
491 transcriptome analysis and the intensity of Hoechst 33342 staining was depicted using a
492 color-based scale, the gradient pattern was easily observed for clusters 1, 3, and 4 (Figure 6a),
493 suggesting that Quartz-Seq2 was highly sensitive for detection of the cell cycle state. To
494 examine this, we plotted the gene expression of several cell cycle markers against the intensity
495 of Hoechst 33342 staining. We observed a strong relationship between the gene expression of
496 several cell cycle markers and DNA content (Figure 6b). Using the intensity of Hoechst 33342
497 staining as reference of the cell cycle phase, we calculated the relationship between the cell
498 cycle phase and all genes detected in Quartz-Seq2 experiments (see Methods). The results
499 enabled us to identify numerous genes for which the expression level changed in relation to the
500 DNA content (Figure 6c, d and Additional file 1: Figure S17). We called these genes "cell cycle
501 associated variable genes". These genes included cell cycle markers. Again, it was confirmed
502 that the ratio of cells in G_1 and S phases to those in G_2 and M phases differed between ES and
503 PrE cells (Figure 6c). As we showed that the efficiency of UMI conversion was high in the
504 Quartz-Seq2 method, we examined whether 10,000 initial reads are sufficient for cell cycle
505 analysis (Additional file 1: Figure S18). In this analysis, the final UMI count was 4,774±62 and the
506 UMI conversion efficiency was 47.74±0.66% (n=3, 1,536 wells). Plotting cells on t-SNE space
507 while using color to depict the intensity of Hoechst 33342 staining revealed a gradient pattern,
508 which was positively correlated with the expression of G_2/M-phase marker genes including
509 *Ccnb1* and *Top2a* and negatively correlated with the expression of G_1-phase marker genes such
510 as *Ccne1* (Additional file 1: Figure S18). These observations demonstrated that Quartz-Seq2
511 with few initial sequence reads could analyze the cell cycle due to the high UMI conversion
512 efficiency.
513 Within a single cell type, there are two types of cell state categories. One is the different
514 stages of the cell cycle as mentioned above, and the other refers to cellular heterogeneity, which
515 is only remotely related to the cell cycle phases. This means that genes with variable expression
516 in a cell type include "cell cycle associated variable genes" and "variable genes the expression of
517 which is less associated with the cell cycle phase" (Additional file 1: Figure S24). We determined
518 the number of the latter group of genes by subtracting the former group from the genes with

519 variable expression (Figure 6e). For the main population of ES cells (cluster 1), we identified tens
520 of variable genes the expression of which is less associated with the cell cycle phase. These
521 genes included *Spp1*, *Rhox5*, and *S100a6*, which were previously identified as genes with highly
522 variable expression by scRNA-seq (Figure 6) [7,13,15]. We also identified *Stmn2*, *Dnmt3l*,
523 *Tmsb4x*, and several other genes as novel variable genes the expression of which is less
524 associated with the cell cycle phase. Single-cell RT-qPCR also showed high variability for *Stmn2*
525 and *Rhox5* compared with that for *Nanog* (Figure 6f). In previous studies, *Sgk1* and *Actb* were
526 identified as genes with variable expression by scRNA-seq, which analyzed cells only in the G_1
527 phase or cells without measuring the intensity of Hoechst 33342 [7,15]. In this study, these were
528 classified as cell cycle associated variable genes (Additional file 1: Figure S17).
529 These results show that Quartz-Seq2 globally classified genes with variable expression
530 within a cell type into “cell cycle associated variable genes” and “variable genes the expression
531 of which is less associated with the cell cycle phase” by using Hoechst 33342 staining intensity
532 data obtained by flow cytometry.
533
534 **The global picture of cell-type composition in the stromal vascular fraction as revealed**
535 **by Quartz-Seq2**
536 Compared with cultured cells, fresh tissue samples tend to be composed of multiple cell types,
537 whose sizes are highly variable. The use of fresh samples is one promising strategy to assess
538 the robustness of high-throughput single-cell RNA-seq for various samples. Therefore,
539 Quartz-Seq2 was applied to cells from the mouse stromal vascular fraction (SVF), which was
540 taken from adipose tissue (Figure 7a). SVF is thought to be one of the most important sources of
541 mesenchymal stem cells (MSCs) due to its potential for use in cell therapies and the stimulation
542 of endogenous repair [26-29]. However, the global picture of cell-type composition in SVF has
543 not been clarified. To achieve this, we first observed the distribution of cell size in the SVF
544 population and found it to be broad (5–13 μm, average 6.43±1.35 μm; n=200) (Figure 7a). This
545 was confirmed by flow cytometry data that were obtained using a cell sorter. It is known that the
546 typical amount of total RNA in a mammalian cell is approximately 10 pg [15,17]. Thus, the
547 average amount of total RNA in the SVF population of approximately 1.7 pg is relatively small.

548 We collected cells in multiple 384-well PCR plates and analyzed their transcriptome using
549 Quartz-Seq2. In this analysis, 38,450±3513 initial fastq reads were converted to UMI counts in
550 the final digital expression matrix. The UMI conversion efficiency was 26.85±2.70% (n=3). We
551 observed broader distributions of UMI counts and gene counts in SVF than in ES and 10 pg of
552 total RNA. The gene counts correlated well with side scatter (SSC) values, which were
553 associated with cell size (Additional file 1: Figure S19).
554 Dimensionality reduction and clustering of single-cell transcriptome data showed the clear
555 separation of 11 clusters (Figure 7b). To annotate the cell type for each cluster, we identified
556 differentially expressed genes and functional terms that were either specific or enriched for each
557 cluster (Figure 7c, d and Additional file 1: Figure S20). As Figure 7c shows, clusters 1, 2, 5, and
558 8–11 expressed cluster-specific genes, and clusters 3, 4, 6, 7, and 9 shared expressed genes
559 (marker class C) as well as differentially expressed genes of classes D–G. Taking these markers
560 and functional terms as well as previous knowledge together, each cluster was identified as
561 follows: two types of *Cd34*-positive MSC (cluster 1: 7.8% of cells from SVF population, cluster 8:
562 6.4%), two types of *Cd4*-positive T cells (cluster 3: 45.0%, cluster 4: 1.0%), *Cxcr6*-positive T cells
563 (cluster 9: 1.71%), B cells (cluster 5: 15.9%), killer T cells (cluster 6: 11.0%), natural killer cells
564 (cluster 7: 3.0%), macrophages (cluster 2: 5.5%), dendritic cells (cluster 10: 0.86%), and
565 plasmacytoid dendritic cells (cluster 11: 0.57%). We noted that the sizes of the T cells and B
566 cells from clusters 3–5, which were measured using a cell sorter, were smaller than those of
567 MSCs and macrophages from clusters 1, 2, and 8 (Additional file 1: Figure S19). Using
568 immunofluorescence observation, we confirmed that the sizes of CD4-positive T cells and
569 CD79-positive B cells were smaller in SVF (Additional file 1: Figure S19). These results show
570 that Quartz-Seq2 with flow cytometry data is useful for defining respective cell clusters with cell
571 size information.
572 Subsequently, we investigated the gene expression of receptors and ligands in cells from
573 the SVF. The variety of receptors expressed in each cluster was similar (Additional file 1: Figure
574 S21). However, we found a wide variety of ligands expressed in clusters 1 and 8 (Additional file
575 1: Figure S21). It is known that MSCs secrete many paracrine mediators, which have a

576 therapeutic effect [30,31]. These results suggest that clusters 1 and 8 are MSCs, which secrete
577 many paracrine mediators.
578 We focused on these two potential MSC clusters (1 and 8) because the heterogeneity of
579 the MSC population has been discussed in many papers [32,33]. By using Quartz-Seq2, clusters
580 1 and 8 were shown to have the *Cd31−*/*Cd34+*/*Cd45−* phenotype (Additional file 1: Figure S22).
581 In a previous study, qPCR-based single-cell transcript analysis of the *Cd31−*/*Cd34+*/*Cd45−*
582 population of SVF was reported. By analyzing 140 genes, the authors identified the
583 *Cd55+*/*Dpp4+*/*Cd31−*/*Cd34+*/*Cd45−* cell population and showed that the administration of this
584 population was effective for normalizing diabetic wound healing in mice [32]. They identified 13
585 marker genes for the *Cd55+*/*Dpp4+*/*Cd31−*/*Cd34+*/*Cd45−* population in mouse SVF. We
586 compared the expression of those 13 genes between clusters 1 and 8 (Additional file 1: Figure
587 S23). The results showed that almost all of the genes were more highly expressed in cluster 1
588 than in cluster 8, suggesting that the cluster 1 population was similar to the
589 *Cd55+*/*Dpp4+*/*Cd31−*/*Cd34+*/*Cd45−* population. To further analyze the relationship between our
590 identified clusters and previously reported heterogeneity of the MSC population, we examined
591 the expression of typical MSC markers (*CD90*, *CD105, PDGFRa*, and *Sca-1*) [32,34] as well as
592 subpopulation-specific markers (*Pou5f1*, *Nanog*, *Sox2*, *Tnnt2*, and *Myog*) [35,36]. *PDGFRa*
593 (*Pdgfra*) and *Sca1* were expressed in both clusters 1 and 8, whereas *Cd90* was expressed at a
594 higher level in cluster 8 than in cluster 1, and *Cd105* was not strongly expressed in either cluster.
595 *Cd90* and *Cd105* were, however, not specific to these MSC clusters. Neither pluripotent markers
596 (*Pou5f1*, *Nanog*, and *Sox2*) nor skeletal and cardiomyogenic markers (*Tnnt2* and *Myog*) were
597 detected in either cluster (Additional file 1: Figure S22). Collectively, our transcriptome analysis
598 showed that the MSC population is divided into two clusters, suggesting that there is less
599 heterogeneity of MSCs in the SVF than expected.
600 Next, we investigated the difference between the two MSC clusters in SVF. We detected
601 182 differentially expressed genes between clusters 1 and 8 (Additional file 5: Table S4). Cluster
602 8 was characterized by the enriched expression of genes that encode extracellular matrix
603 proteins, including collagens, *Hspg2*, and *Bgn*. By contrast, cluster 1 was characterized by the
604 expression of *IGFBPs* (*Igfbp4*, *Igfbp5*, *Igfbp6*), prostaglandin endoperoxidase synthases (*Ptgs2*),

605 and secreted factors *Fgf2* and Bmp2. Notably, the expression of stem-cell-characteristic genes
606 *Aldh1a3*, *Fgf2*, *Bmp2* and *Tgfbr2* were identified as important modules in cluster 1, suggesting
607 that this cluster included more stem-like phenotypes that secrete medicinal factors.
608 To leverage the power of the entire transcriptome analysis, we performed an enrichment
609 analysis of biological pathways using the 182 genes differentially expressed between the two
610 clusters (Figure 7e, Additional file 6: Table S5). Cluster 1 was characterized by elastic fiber
611 pathways and IGF transport pathways. Elastic fibers are constituents of various connective
612 tissues and are related to vascular function [37]. Growth factors including FGF2, BMP2, IGFBPs,
613 and PGE2 regulate growth and differentiation [38]. A tightly linked network of collagen
614 metabolism pathways was found to be highly enriched in cluster 8. These results suggest that
615 two MSC clusters of SVF have different functions. Cluster 1 could feature medicinal cells
616 characterized by the expression of growth and development-related genes. In contrast, cluster 8
617 may be characterized by the production of extracellular matrix.
618 Taking the obtained findings together, our transcriptome analysis by Quartz-Seq2
619 successfully identified cell populations that consisted of various types of both large and small
620 cells, which were isolated from *in vivo* tissue, and demonstrated that fresh SVF contains two
621 closely related types of MSCs that have distinct characteristics.
622
623
624
625
626
627
628
629
630
631
632
633

634 **Discussion**

635 The rate of conversion from initial reads to UMI counts was low, especially in high-throughput
636 single-cell RNA-seq. Ideally, greater UMI counts should be generated from limited sequence
637 reads because the increase in UMI count assigned to each cell leads to the detection of low-copy
638 genes and the identification of cell-type-specific genes using statistical tests. To overcome these
639 issues, we developed a novel high-throughput single-cell RNA-seq method, Quartz-Seq2. We
640 calculated the UMI conversion efficiency, which indicated how effectively initial reads can be
641 converted to UMI counts. The UMI conversion efficiency of Quartz-Seq2 was 1.44–4.55-fold
642 higher than those of other single-cell RNA-seq methods (Figure 4). Quartz-Seq2 could detect
643 1.28–2.41-fold more genes from limited fastq reads and for a lower cost (Figure 4). Moreover, we
644 decreased the cDNA preparation cost per cell of Quartz-Seq2 by 97.5%–98.5% compared with
645 that of Quartz-Seq (Figure S6a). The total cost (cDNA preparation cost and sequence cost per
646 cell) of Quartz-Seq2 as a cell-sorter-based method approached that of a droplet-based method
647 (Additional file 1: Figure S6b). Furthermore, the use of evaporation-preventing oil might reduce
648 the cDNA preparation cost of Quartz-Seq2 by 75% [2]. By cell barcoding, Quartz-Seq2 can pool
649 cDNA of up to 1,536 individual cells into one mixture. Streamlined and simplified experimental
650 processes of Quartz-Seq2 allowed us to handle thousands of single cells within a few days.
651 Quartz-Seq2 also showed high accuracy and high reproducibility, leading to precise
652 measurement of the transcriptome from single cells (Figure 2d and e). We demonstrated
653 Quartz-Seq2 analyses on a total of approximately 9,000 mouse ES cells as an *in vitro* sample,
654 and approximately 1,000 cells from mouse SVF samples including MSCs as an *in vivo* sample.

655 We proposed that UMI conversion efficiency could be a useful variable for evaluating
656 performance for the further development of high-throughput single-cell RNA-seq methods with
657 shallow reads. The increase of initial sequence reads by additional sequencing does not cancel
658 out the low efficiency of UMI conversion at an equal rate. For example, approximately 24,000
659 fastq reads were converted to approximately 11,000 UMI counts in Quartz-Seq2, while Drop-seq
660 required approximately 169,000 fastq reads to generate the same UMI counts. Under these
661 conditions, the difference of UMI conversion efficiency was 2.5-fold but the difference of required
662 fastq reads to generate the same UMI count was 7-fold. This is because UMI counts do not

663 increase linearly with an increase of initial sequence reads due to UMI filtering. Unfortunately,
664 the rate of increase of sequence throughput was lower than that of the processing ability for
665 single cells. It will thus continue to be important to utilize limited initial fastq reads for
666 high-throughput single-cell RNA-seq methods.

667 Quartz-Seq2 is based on poly-A tagging, which is one of the strategies of converting
668 first-strand cDNA to amplified cDNA [19,39]. However, the poly-A tagging efficiency itself has not
669 been improved for single-cell RNA-seq. In this study, three molecular biological improvements of
670 Quartz-Seq2 contributed to the increase of amplified cDNA, which led to high UMI conversion
671 efficiency (Figure 2 and Additional file 1: Figure S1). In addition, the improvement of poly-A
672 tagging steps was the most efficient approach to increase amplified cDNA for Quartz-Seq2 (up
673 360%, Additional file 1: Figure S1). Several previous studies improved the efficiency of
674 conversion from an mRNA molecule to amplifiable cDNA, and showed that the increase of
675 amplified cDNA is a good guidepost to improve UMI counts or quantitative performance for
676 single-cell RNA-seq [3,14,15]. These results suggest that improved poly-A tagging was the most
677 important feature for the high UMI conversion efficiency of Quartz-Seq2.

678 Several improvements proposed in this study could contribute to the further
679 development of other single-cell RNA-seq methods, as follows. (1) We showed that a decrease
680 of enzyme concentration in RT solution led to decreases in technical error and cDNA preparation
681 cost. (2) We applied the Sequence–Levenshtein distance for the design of cell barcodes
682 containing RT primer sets. The use of primer sets designed in this study allows the user to
683 correct mutations of at most two nucleotides of a cell barcode, including substitution, insertion, or
684 deletion; the correction capability of these primer sets is higher than that of barcode sets used in
685 previous studies [1,3]. Such correction of the cell barcode increased the average UMI count by
686 3%–5%. (3) We hope that the spin-down collection system developed in this study can be
687 applied to other cell-sorter-based single-cell RNA-seq methods [2,3,11,12]. (4) In Quartz-Seq2,
688 we improved the efficiency of poly-A tagging itself (Figure 2). Several single-cell RNA-seq
689 methods including Quartz-Seq were based on poly-A tagging [4,5,10,15,19,40]. Such
690 improvement may directly contribute to these methods having increased quantitative
691 performance.

692 We achieved a marked improvement in UMI conversion efficiency at a low cost from
693 initial reads to UMI counts, allowing us to analyze 3- to 10-fold more single cells in limited
694 sequence experiments. However, it is difficult to sort more than 20,000 single cells in a day when
695 using one cell sorter. While there is greater scalability for cell sampling in droplet-based
696 single-cell RNA-seq methods than in cell-sorter-based methods, these latter methods can utilize
697 additional information obtained in flow cytometry [41,42], which cannot be determined from
698 transcriptome data alone. These two types of method provide complementary approaches for
699 investigating complex biological phenomena. C1 Single-Cell Auto Prep System (Fluidigm) is
700 another widely used platform for single-cell RNA-seq methods [43]. CEL-seq2 shows high UMI
701 conversion efficiency, so it is a very convenient method for users of the C1 platform as a
702 single-cell RNA-seq method based on UMI count [3]. Quartz-Seq2 cannot be performed with the
703 C1 platform, but it can assimilate 1,536 cell barcodes in a cell sorter in a high-throughput manner.
704 We think that both methods can be effective as long as the most appropriate one is selected for
705 each situation.
706 In this study, we showed that Quartz-Seq2 has advantages in gene detection and the
707 identification of biological pathways via high UMI conversion efficiency (Figure 3). Moreover, we
708 analyzed thousands of single cells from all cell cycle phases within a cell type (Figure 6). These
709 specifications with Quartz-Seq2 allow us to perform the global classification of genes with
710 variable expression in a cell type into “cell cycle associated variable genes” and “variable genes
711 the expression of which is less associated with the cell cycle phase”.
712 While Hoechst 33342 is used as an indicator of DNA content, FUCCI reporter systems
713 monitor the ratio of activities for two different cell-cycle-associated proteins. Therefore, the latter
714 has an advantage in cell cycle analysis via providing more detailed resolution, although the
715 introduction of reporter constructs into the cells cannot be applied for all purposes, especially for
716 the analysis of human samples. A previous study utilizing the FUCCI fluorescent reporter system
717 reported that competence to respond to specific differentiation signaling was limited to only an
718 early or a late window of the G_1 phase in human ES cells [44]. In the future, combining
719 Quartz-Seq2 and FUCCI/FUCCI2 or other fluorescent reporter systems should lead to an
720 understanding of the global picture of differentiation dynamics regarding competence, response,

transition, and commitment because Quartz-Seq2 can analyze changes of cell state for thousands of cells.

In summary, Quartz-Seq2 can be used to obtain continuous data on cell states because of the large number of cells with which it deals and the high efficiency of use of initial sequence reads. Quartz-Seq2 can facilitate investigation of the cell state within a cell type, such as gradated or stochastic changes of the cell population in organism development and disease progression.

750 **Conclusions**

751 In this study, we developed a high-throughput single-cell RNA-seq method, Quartz-Seq2, which
752 can analyze cells numbering up to 1,536 that are pooled together in a single sample.
753 Quartz-Seq2 allows us to effectively utilize initial sequence reads from a next-generation
754 sequencer. The UMI conversion efficiency in Quartz-Seq2 ranged from 32% to 35%, which is
755 much higher than for other single-cell RNA-seq methods (7%–22%). This was caused by the
756 improvements in several molecular biological steps including poly-A tagging. The technical gene
757 expression variability of Quartz-Seq2 was close to the theoretical variability of a Poisson
758 distribution. As we showed in the analysis of SVF and the ES/PrE mixture, cell types in the
759 population were identified with marker genes and functional terms that characterized each cell
760 type. We identified two types of *Cd34*-positive MSCs in SVF, namely, those that express
761 numerous transcription-factor-encoding and secreted-protein-encoding genes specific to each.
762 Quartz-Seq2 can also be used to provide continuous data on cell states because of the large
763 number of cells with which it deals and the high efficiency in the use of initial sequence reads.
764 Quartz-Seq2 should facilitate investigations of the cell state within a cell type, such as gradated
765 or stochastic changes of cell populations in organism development and disease progression.
766
767
768
769
770
771
772
773
774
775
776
777
778

779 **Methods**

780 **Cell culture**

781 Mouse embryonic stem cells were cultured as described previously [15]. Briefly, 1.0 x 10^5 cells
782 were seeded on a 60-mm dish coated with gelatin (0.1%). Cells were maintained in
783 GMEM-based medium containing 10% FBS and 1000 units/ml leukemia inhibitory factor
784 (Millipore ESGRO). The cell line used in this study was G6GRGFP, which was established in a
785 previous study [23]. It has been reported that almost all G6GRGFP differentiated into primitive
786 endoderm-like cells upon dexamethasone treatment [15,23]. To differentiate ES cells into
787 primitive endoderm-like cells, 1.0 x 10^5 cells were seeded on a 60-mm dish coated with gelatin
788 and cultured in GMEM-based medium containing 10% FBS, 1000 units/ml LIF, and 100 mM
789 dexamethasone for 72 h. We confirmed that almost all of the Dex-treated G6GR ES cells
790 differentiated into primitive endoderm-like cells. Moreover, we cultured J1 ES cells in accordance
791 with a procedure described in a previous report [21].

792

793 **Cell staining**

794 To identify dead or damaged cells, propidium iodide (PI) was added to the cell suspension (final
795 concentration 1–2 μg/ml). As an indicator of an undifferentiated state in a suspension containing
796 a mixture of cells, Calcein-AM was used. Suspensions for cells cultured under maintenance and
797 differentiation conditions were prepared independently, and cells in the maintenance condition
798 were treated with 1 μg/ml Calcein-AM for 10 min on ice. After washing with PBS, cell
799 suspensions of ES cells and PrE cells at the same concentration were mixed. Hoechst 33342
800 was used as an indicator of the DNA content in a cell. The procedures were performed as
801 described previously [15].

802

803 **Single-cell preparation for stromal vascular fraction**

804 These experiments were carried out in accordance with the protocol approved by the National
805 Institute of Advanced Industrial Science and Technology (AIST) Animal Care and Use
806 Committee and the RIKEN Animal Experiment Committee. Subcutaneous fat tissues from three-
807 to four-month-old ICR male mice (n=3 per sample, two biological replicate samples) were

808 minced into small pieces and incubated with 0.4 U/mL collagenase NB4G (Serva) at 37°C for 35
809 min in a shaking water bath. The digested solution was sequentially filtered through 100-μm and
810 40-μm cell strainers (Corning), followed by centrifugation at 250 × *g* for 5 min to remove mature
811 adipocytes. The pellet was treated with erythrocyte lysis buffer (BD Biosciences) and centrifuged
812 at 180 × *g* for 5 min. The nucleated cells were suspended in HBSS with 0.1% BSA, filtered
813 through a 20-μm cell strainer (pluriSelect), and then kept on ice (Cell solution A). The cell
814 aggregates that did not pass through the 20-μm strainer were further treated with Accutase
815 (Thermo Fisher Scientific) at 37°C for 15 min to dissociate them into single cells, centrifuged at
816 180 × *g* for 5 min, and suspended in HBSS with 0.1% BSA (Cell solution B). Cell solutions A and
817 B were mixed and again filtered through the 20-μm cell strainer, followed by centrifugation at 180
818 × *g* for 5 min. The pellet was resuspended with HBSS with 0.1% BSA, stained with PI, and used
819 for single-cell analysis.
820
821 **RNA preparation**
822 Total RNA was purified from cultured cells using Direct-zol RNA MiniPrep kit (Zymo Research)
823 with TRIzol RNA Isolation Reagents (Thermo). We measured the concentration of purified total
824 RNA using a NanoDrop 1000 Spectrophotometer (Thermo). We confirmed that the RNA integrity
825 number of total RNA was over 9.5 using an Agilent RNA 6000 Nano Kit (Agilent). Estimation of
826 the average amount of total RNA per single cell for respective samples was performed in
827 accordance with our previous study [15]. For high-throughput single-cell RNA-seq, we prepared
828 total RNA from a single cell with ERCC spike-in RNA. First, we diluted the ERCC spike-in RNA
829 tenfold. We then added 6 μL of 1:10 diluted ERCC spike-in RNA per 10 μg of total RNA. We
830 used diluted 10 pg of total RNA with ERCC spike-in RNA for the technical validation of
831 Quartz-Seq2. For single cells, we used the same concentration of ERCC spike-in RNA.
832
833 **Bulk RNA-seq methods for populations of cells**
834 We prepared sequence library DNA with 1 μg of total RNA using NEBNext Poly(A) mRNA
835 Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit. The total RNA
836 did not contain ERCC spin-in mix I. In addition, we used SuperScript III instead of ProtoScript in

837 the reverse-transcription step and KAPA HiFi DNA polymerase instead of NEBNext High-Fidelity
838 PCR DNA polymerase in the PCR step. The resulting sequence library DNA was analyzed by
839 HiSeq2500.
840
841 **Design of cell barcodes**
842 The selection of 384 or 1,536 sequences for v3.1 and v3.2 barcode primer sets was performed
843 as described below. First, 1,582 or 4,714 candidate sequences were created using the
844 DNABarcodes package of R Bioconductor for the v3.1 set with 14-mer and the v3.2 set with
845 15-mer, respectively (version: 1.0.0). To reduce the loss of reads converted into UMI counts, we
846 applied the Sequence–Levenshtein distance as an edit distance in order to maximize the ability
847 to correct errors that occur during the synthesis of oligonucleotides or sequencing [22]. The
848 minimum distance between any two sequences was controlled to 5, leading to the correction of a
849 maximum of two errors of substitution, insertion, or deletion. As the base composition of the
850 created sequences was not uniform, we selected 384 or 1,536 sequences from among 1,582 or
851 4,714 created sequences so that the variance of base composition decreased. These sequences
852 are listed in Additional file 3: Table S2.
853
854 **Single-cell collection using flow cytometry**
855 Stained cells were analyzed using flow cytometry SH800 (Sony) or MoFlo Astrios EQ (Beckman
856 Coulter). For SH800, we used 130-μm microfluidic sorting chips. For MoFlo Astrios EQ, we used
857 100-μm nozzle sizes. Each cell sorter was equipped with a custom-made splash-guard to
858 prevent contamination from unexpected droplets between the target well and neighboring wells.
859 We used two sets of RT primer for Quartz-Seq2. The set of v3.1 RT primers has 384 kinds of
860 unique cell barcodes, with a length of 14 nucleotides (OPC purification, FASMAC). The set of
861 v3.2 RT primers has 1,536 kinds of unique cell barcodes, with a length of 15 nucleotides (OPC
862 purification, Sigma). The set of v3.1 RT primers corresponds to one set of the 384-well PCR
863 plate with lysis buffer, the wells of which have unique barcodes. The set of v3.2 RT primers
864 corresponds to four sets of the 384-well PCR plate with lysis buffer. The RT primer position in the
865 384-well plate and the sequence were as described in Additional file 3: Table S2. Single cells

866 were isolated in the 384-well PCR plate with 1 μL of lysis buffer (0.1111 μM respective RT
867 primers, 0.12 mM dNTP mix, 0.3% NP-40, 1 unit/μL RNasin plus) containing ERCC spike-in RNA.
868 During single-cell sorting, the 384-well PCR plate was kept on a 384 aluminum stand at 4ºC. For
869 Moflo Astrios EQ, we used G5498B-060 Insert 384 Eppendorf twin.tec PCR as a 384 aluminum
870 stand (Agilent). For SH800, we used an SH800 384 aluminum stand (Sony). Immediately after
871 the cell collection, the plate was temporarily sealed with LightCycler 480 Sealing Foil (Roche)
872 and the sealed 384-well PCR plate was centrifuged at 10,000 g and 4ºC for 1 min using TOMY
873 MX307, equipped with a Rack-in-Rotor and centrifugation rack PCR96-02. These steps were
874 very important to collect the droplet with a single cell in lysis buffer in an efficient manner. By
875 altering the volume of lysis buffer from 0.4 μL (as described in the original Quartz-Seq paper) to
876 1 μL (Quartz-Seq2), bubbling of lysis buffer was not required before single-cell sorting and the
877 lysis buffer could easily be handled. We then peeled open the temporary seal and re-sealed it
878 with Agilent PlateLoc Thermal Microplate Sealer (Agilent). We agitated the plate at 2,600 rpm
879 and 4ºC for 1 min using ThermoMixer C (Eppendorf), after which we centrifuged the plate again.
880 The resulting 384-well plate was then immediately cryopreserved at −80ºC and maintained
881 under such conditions until subsequent reverse transcription for cell barcoding. After
882 cryopreservation, we performed subsequent whole-transcript amplification using the
883 cryopreserved 384-well PCR plate within a few months.
884
885 **Whole-transcript amplification of Quartz-Seq2**
886 Cryopreserved 384-well plates with single-cell lysate were centrifuged at 10,000 g and 4ºC for 1
887 min. Subsequently, we denatured total RNA in each 384 plate at 70ºC for 90 s and hybridized the
888 RT primer to poly-adenylated RNA at 35ºC for 15 s using the C1000/S1000 thermal cycler. The
889 resulting plates were again centrifuged at 10,000 g and 4ºC for 1 min. Next, the plates were
890 placed on the 384 aluminum plate at 0ºC. We peeled away the seal and added 1 μL of RT
891 premix (2x Thermopol buffer, 5 units/μL SuperScript III, 0.55 units/μL RNasin plus) to 1 μL of
892 lysis buffer for each well using a Mantis microfluidic dispensing system (Formulatrix) or a 384
893 Transfer Plate system (1859-384S, Watson). The above RT solution was used for the RT25
894 condition. For the RT100 condition, we used the following RT solution: 2x Thermopol buffer, 20

895 units/μL SuperScript III, and 2.2 units/μL RNasin plus. We sealed the plates again and agitated
896 them at 2,600 rpm and 4ºC for 1 min. The plates were then centrifuged at 10,000 g and 4ºC for 1
897 min. We then performed reverse transcription at 35ºC for 5 min and 50ºC for 50 min. The reverse
898 transcription was stopped at 70ºC for 15 min. Then, the plates were placed on a prechilled
899 aluminum block, after which we peeled off their seals. Subsequently, we turned the plates upside
900 down on the assembled collector type A or type B (supplemental figure). We mainly used type A.
901 We centrifuged the plates with an assemble collector at 3,010 g and 4ºC for 3 min with
902 swing-bucket rotors. Subsequently, we collected the cDNA solution into a disposable reservoir.
903 Typically, we obtained 650–700 μL of cDNA solution from one 384-well PCR plate. We purified
904 and concentrated the cDNA solution using the DNA Clean & Concentrator™-5 kit (Zymo
905 Research). We used three purification columns for one 384-well PCR plate in the case of the
906 v3.1 RT primer system (384-cell barcode). Purified cDNA was extracted into 20 μL of
907 nuclease-free water from one column purification and transferred into an eight-linked PCR tube
908 (TaKaRa). The PCR tubes were placed on an aluminum PCR stand at 0ºC. We added 25 μL of
909 TdT solution [1x Thermopol buffer, 2.4 mM dATP, 0.0384 units/μL RNase H (Invitrogen), 26.88
910 units/μL terminal transferase (Roche)] into 20 μL of extracted cDNA using a pipette at 0ºC. The
911 resulting 45 μL of TdT solution was mixed with a pipette at 0ºC or ThermoMixer at 2,000 g and
912 0ºC for 1 min. Immediately thereafter, the PCR tubes were centrifuged at 10,000 g and 0ºC for 1
913 min. We used a C1000/S1000 thermal cycler equipped with the 96-Deep Well Reaction Module
914 for the following steps. The PCR tubes were placed on the block of the thermal cycler, which had
915 been prechilled to 0ºC. We then performed a poly-A tailing reaction at 37ºC for 75 s. The solution
916 was inactivated at 65ºC for 10 min. The PCR tubes were placed on an aluminum PCR stand at
917 0ºC. We then dispensed approximately 11 μL of solution into four wells from 45 μL of TdT
918 solution. We added 46.16 μL of PCR I premix (1.08492x MightyAmp Buffer version 2, 0.06932
919 μM Tagging primer, 0.05415 units/μL MightyAmp DNA polymerase) to 11 μL of TdT solution for
920 the respective wells of the PCR tube. MightyAmp DNA polymerase, which used in Quartz-Seq
921 and Quartz-Seq2, are marketed as Terra PCR Direct polymerase [15]. We performed gentle
922 inversion mixing on the resulting solution in the PCR tube. The tubes were then centrifuged at
923 10,000 g and 4ºC for 1 min. Subsequently, the solution was mixed with ThermoMixer at 2,000

924 rpm and 4ºC for 2 min. Then, we spun down the tube again. Next, we denatured the solution at
925 98ºC for 130 s and hybridized Tagging primer to poly-A-tailed cDNA at 40ºC for 1 min. After that,
926 we performed the “Increment step” by heating to 68ºC at 0.2ºC every second and performed
927 second-strand synthesis at 68ºC for 5 min. The tubes were placed on an aluminum PCR stand at
928 0ºC. We added 50.232 μL of PCR II premix (0.99697x MightyAmp Buffer version.2, 1.8952 μM
929 gM primer) to 56.16 μL of PCR I solution. We performed gentle inversion mixing on the resulting
930 solution in the PCR tube. The tubes were then centrifuged at 10,000 g and 4ºC for 1 min.
931 Subsequently, the solution was mixed with ThermoMixer at 2,000 rpm and 4ºC for 2 min, after
932 which we spun down the tube again. We then placed it on the block of the thermal cycler at 68ºC.
933 Subsequently, we amplified the cDNA for 11 cycles under the following conditions: 98ºC for 10 s,
934 65ºC for 15 s, and 68ºC 5 min. We then incubated the tube at 68ºC for an additional 5 min.
935 Finally, we transferred all of the PCR solution, derived from one 384-well PCR plate, to a 50-mL
936 Polypropylene Centrifuge Tube (Watson). Typically, we obtained approximately 1.2 mL of PCR
937 solution per 384-well PCR plate. We added 32 μL of 3 M sodium acetate (pH 5.2) and 6420 μL
938 PB-Buffer (Qiagen) to the PCR solution. The mixture was then purified using a MinElute Spin
939 Column (Qiagen). Purified cDNA was extracted into 40 μL of nuclease-free water. We
940 additionally purified the cDNA with 32 μL of Ampure XP beads. Finally, we obtained 32 μL of
941 purified cDNA. We checked the length distribution of amplified cDNA with an Agilent High
942 Sensitivity DNA Kit (Agilent). The typical average size of the amplified cDNA in Quartz-Seq2 was
943 approximately 1,400 bp (Additional file 1: Figure S2c). Primer sequences are listed in Additional
944 file 2: Table S1.
945 In the case of the usage of the v3.2 RT primer in Quartz-Seq2, we modified the above
946 steps as follows. After reverse transcription, we collected cDNA solution into a disposable
947 reservoir from four sets of 384-well plates, which corresponded to 1,536 wells. We purified and
948 concentrated the cDNA solution using eight purification columns for four 384-well PCR plates in
949 the case of the v3.2 RT primer system. In the PCR step, we amplified cDNA for nine cycles with
950 the following conditions: 98ºC for 10 s, 65ºC for 15 s, and 68ºC for 5 min. Finally, we transferred
951 all of the PCR solution derived from four 384-well plates to a 50-mL Polypropylene Centrifuge
952 Tube (Watson). Typically, we obtained approximately 3.5 mL of PCR solution per four 384-well

953 PCR plates. We added 88 μL of 3 M sodium acetate (pH 5.2) and 17.6 mL of PB-Buffer (Qiagen)
954 to the PCR solution. The mixture was purified using the MinElute Spin Column (Qiagen).
955 Subsequently, cDNA was again purified using Ampure XP magnetic beads.
956 For the Quartz-Seq1-like reaction, we performed the following procedure in
957 accordance with our previous study. We added 1 μL of RT premix (2x PCR buffer, 5 units/μL
958 SuperScript III, 0.55 units/μL RNasin plus) to 1 μL of lysis buffer. We then performed reverse
959 transcription at 35ºC for 5 min and 45ºC for 20 min. This reverse transcription was stopped at
960 70ºC for 15 min. We added 5 μL of ExoIB solution (1.6x Exonuclease I buffer, 3.2x PCR buffer,
961 16 mM DTT) and 20 μL of TdT solution [1x PCR buffer, 3 mM dATP, 0.0384 units/μL RNase H
962 (Invitrogen), 33.6 units/μL terminal transferase (Roche)] into 20 μL of extracted cDNA using a
963 pipette at 0ºC. The PCR tubes were placed on the block of the thermal cycler that had been
964 prechilled to 0ºC. We performed a poly-A tailing reaction at 37ºC for 75 s. We then denatured the
965 solution at 98ºC for 130 s and hybridized Tagging primer to poly-A-tailed cDNA at 40ºC for 1 min.
966 After that, we performed second-strand synthesis at 68ºC for 5 min. Subsequently, we amplified
967 the cDNA via a PCR reaction for 12 cycles.
968
969 **Preparation of truncated sequence adaptor**
970 The truncated sequence adaptor was composed of an rYshapeP5 primer (HPLC-purified) and
971 rYshapeP7LT primers (HPLC-purified), which had a TruSeqLT-compatible pool barcode. We
972 prepared 100 μM respective primers with adaptor buffer (10 mM Tris-HCl pH7.8, 0.1 mM EDTA
973 pH8.0, 50 mM NaCl). We added 5 μL of 100 μM rYshapeP5 primer and rYshapeP7LTxx primer
974 into a single PCR tube. We denatured the solution at 90ºC for 90 s. After that, we achieved
975 annealing by cooling to 10ºC by 0.5ºC every 30 s and then maintaining the sample at 4ºC. We
976 then placed the tube on an aluminum PCR stand at 0ºC. Subsequently, we added adaptor buffer,
977 which was prechilled at 0ºC, to 10 μL of 50 μM truncated adaptor. Finally, we obtained about 50
978 μL of 10 μM truncated adaptor. We cryopreserved 1 μL of 10 μM truncated adaptor in a PCR
979 tube at −80ºC until usage in the adaptor ligation step. Primer sequences are listed in Additional
980 file 2: Table S1.
981

982 **Sequence library preparation of Quartz-Seq2**

983 We added 130 μL of nuclease-free water including 5–10 ng of amplified cDNA into a Crimp-Cap
984 microTUBE with AFA Fiber. We then performed cDNA fragmentation using an LE220
985 Focused-ultrasonicator (Covaris) under the following conditions: duty factor 15%, peak incident
986 power 450 W, cycles per burst 200, and treatment time 80 s. We purified and concentrated the
987 cDNA solution using the DNA Clean & Concentrator™-5 kit. Purified cDNA was extracted into 10
988 μL of nuclease-free water from one column purification and transferred into an eight-linked PCR
989 tube (TaKaRa).

990 We then added 2 μL of End-Repair premix [1.4 μL of End repair & A-tailing buffer and
991 0.6 μL of End repair & A-tailing Enzyme (KAPA Biosystems)] to 10 μL of fragmented cDNA
992 solution. Subsequently, we mixed the solution by pipetting on an aluminum PCR stand at 0ºC.
993 The PCR tubes were placed on the block of the thermal cycler, which had been prechilled to
994 20ºC. We then incubated the tubes at 37ºC for 60 min and 65ºC for 30 min. After this, we added
995 2 μL of adaptor buffer (1.5 μM truncated adaptor, 10 mM Tris-HCl pH7.8, 0.1 mM EDTA pH8.0,
996 50 mM NaCl) and 8 μL of ligation premix [6 μL of ligation buffer, 2 μL of DNA ligase (KAPA
997 Biosystems)] at 4ºC. The 22 μL of ligation solution was well mixed by pipetting at 4ºC. Then, we
998 performed adaptor ligation at 20ºC for 15 min. After ligation, we added 18 μL of Ampure XP
999 beads to 22 μL of adaptor ligation solution and mixed them well. Adaptor-ligated cDNA was
1000 extracted into 20 μL of nuclease-free water. We added 32 μL of PCR premix [25 μL of 2xKAPA
1001 HiFi ready mix, 1.75 μL of 10 μM TPC2 primer (HPLC-purified), 10 μM P5-gMac hybrid primer
1002 (HPLC-purified)] to 18 μL of adaptor-ligated cDNA. We denatured the solution at 98ºC for 45 s.
1003 Subsequently, we amplified cDNA for eight cycles under the following conditions: 98ºC for 15 s,
1004 60ºC for 30 s, and 72ºC for 1 min. Finally, we additionally incubated the tube at 72ºC for 5 min.

1005 We added 40 μL of Ampure XP beads to 50 μL of PCR solution. Purified
1006 sequence-library DNA was eluted into 20–30 μL of nuclease-free water. We checked the DNA
1007 concentration and DNA size of sequence library DNA using the Agilent High Sensitivity DNA Kit
1008 (Agilent) and QuantiFluor® dsDNA System (Promega). We also prepared sequence library DNA
1009 with Tn5 transposase, in accordance with a modified version of a procedure used in a previous
1010 study [6]. Specifically, the modifications were as follows. We used 0.75 ng of amplified cDNA for

1011 the Nextera XT library preparation kit. We amplified sequence library DNA with the P5-gMac
1012 hybrid primer and the Nextera XT primer, which has the P7 sequence.
1013
1014 **Deep sequencing for Quartz-Seq2**
1015 We analyzed the sequence library DNA from Quartz-Seq2 using NextSeq500 (Illumina) and
1016 HiSeq2500 (Illumina). For the Read1 sequence, we used a custom sequence primer named
1017 Read1DropQuartz primer (HPLC-purified). In the case of the v3.1 RT primer, sequence
1018 specification was as follows (Read1: 22 cycles, Index1: 6 cycles, Read2: 64–118 cycles). In the
1019 case of the v3.2 RT primer, sequence specification was as follows (Read1: 23 cycles, Index1: 6
1020 cycles, Read2: 63 cycles). When Read1 length was within 64 nt we mainly used the NextSeq
1021 500/550 High Output v2 Kit (75 cycles), which can be used for sequencing up to 92 nt. For long
1022 sequences (over 64 nt), we analyzed sequence library DNA using the HiSeq Rapid SBS Kit v2.
1023 Sequence analysis with HiSeq2500 was supported by the Genomic Network Analysis Support
1024 facility (GeNAS) in RIKEN and Phyloinformatics Unit of RIKEN Center for Life Science
1025 Technologies (formerly GRAS) in RIKEN.
1026
1027 **Amplification-free single-cell qPCR analysis**
1028 Amplification-free single-cell qPCR was performed in accordance with our previous study with
1029 the following modifications [15]. We collected a single cell into 1 μL of lysis buffer (2.5 μM
1030 random hexamer, 1 mM dNTP mix, 0.3% NP40, 2 units/μL RNasin plus) of a 384-well PCR plate
1031 (Eppendorf) using SH800. The resulting 384-well PCR plate was cryopreserved at −80ºC. We
1032 denatured total RNA in the 384-well PCR plate at 70ºC for 1.5 min using a C1000/S1000 thermal
1033 cycler and hybridized the random primer to RNA at 0ºC for 2 min. We added 1 μL of
1034 reverse-transcription solution (2x SSIV buffer, 10 mM DTT, 2 units/μL RNasin plus, 20 units/μL
1035 SuperScript IV) to the lysis buffer. Subsequently, we performed reverse transcription at 23ºC for
1036 10 min and 50ºC for 10 min. The reverse transcription was stopped at 80ºC for 10 min. The
1037 resulting solution was diluted with qPCR solution (10 mM Tris-HCl pH 8.0, 0.05% Tween-20).
1038 The obtained diluted solution was then used for qPCR detection with QuantiTect SYBR Green

1039 PCR Master Mix and the LightCycler480 system. For primer sets for each gene, see Additional
1040 file 3: Table S2.
1041
1042 **Drop-seq experiments and data analysis**
1043 Drop-seq was performed as reported previously [6] and in line with an online protocol
1044 (http://mccarrolllab.com/dropseq/), but with the following modifications: flow rates for oil and
1045 aqueous suspensions were 14,000 µl/h and 4,000 µl/h, respectively. The diameter of droplets
1046 was 95–100 µm. Microfluidic devices were fabricated by Fluidware Technologies (Japan). The
1047 lot number of barcoded beads was 051415 (ChemGenes). Data analysis for Drop-seq was
1048 performed as described online (http://mccarrolllab.com/dropseq/). The versions of the software
1049 and databases were as follows: STAR: v2.5.1b; mouse genome: GRCm38/mm10; genome
1050 annotation: gencode GRCm38.p4 vM9; and Drop-seq tools: v1.11.
1051
1052 **Quartz-Seq2 read alignment and generation of digital expression data**
1053 The structure of the sequence library and data processing was designed based on those of
1054 Drop-seq ([6] and online as referenced above). BCL files generated by Illumina NextSeq500
1055 were converted to fastq files by bcl2fastq2 (v2.17.1.14) with demultiplexing pool barcodes. The
1056 --mask-short-adapter-reads parameter was set to 20. If needed, fastq reads were randomly
1057 downsampled by seqtk software (version: sgdp). We mainly trimmed read2 length to 62 nt for
1058 Quartz-Seq2 by using FASTX-Toolkit (version: 0.0.14). Fastq files for Read1 and Read2 were
1059 converted to a bam file by FastqToSam of Picard tools (version: 1.134). Extracting cell barcodes
1060 and UMI (also called as molecular barcode) and filtering out of reads with low barcode quality
1061 were performed using Drop-seq tools (version: 1.11). The resulting bam files were re-converted
1062 to fastq files by SamToFastq of Picard tools, and mapped to the mouse genome
1063 (GRCm38/mm10) using STAR (version: 2.5.1b). After sorting the resulting bam files using
1064 SortSam of Picard tools, the unaligned bam and aligned bam were merged by
1065 MergeBamAlignment of Picard tools. Then, gene names were assigned to each record using
1066 TagReadWithGeneExon of Drop-seq tools and gtf file (version: gencode GRCm38.p4 vM9). For
1067 the correction of errors of cell barcodes considering the Sequence–Levenshtein distance, a

1068 custom-made Python program (correct_bacode.py) was used, which enabled the correction of
1069 up to two nucleotide errors of substitution, insertion, or deletion. This program used Python2
1070 (version: 2.7.11+), PypeR (version: 1.1.2), R (version: 3.2.3), Bioconductor (version: 3.2), and
1071 the Bioconductor package DNABarcodes (version: 1.0.0). Finally, the UMI for each gene for
1072 each cell was counted using DigitalExpression of Drop-seq tools, which generated the digital
1073 expression matrix. To generate a non-UMI-filtered matrix, a custom-made Python program that
1074 counts reads for each gene for each cell was used. To compare the quantitative performance
1075 between Quartz-Seq2 and reported data in Figure 4, we used Ensembl75 as a reference
1076 transcript. Moreover, we trimmed the read2 length to 45 nt for the comparison.
1077
1078 **Dimensionality reduction, clustering, and term analysis**
1079 Cells with low detected gene counts were removed for further analysis (Quartz-Seq2 on ES/PrE
1080 mixture: 4,000 genes; Quartz-Seq2 on SVF: 500 genes). Total counts of each cell were
1081 normalized to 10,000 or the mean of total UMI counts for cells. UMI counts had 1 added to them
1082 and were then log-transformed (natural logarithm), after which they were scaled to have mean
1083 and variance for each gene of 0 and 1, respectively. For PCA, UMI counts for all detected genes
1084 (for ES/PrE mixture) or genes with highly variable expression (for SVF) were used. For t-SNE,
1085 the top 10 to 40 principal components of matrices produced by PCA were used. For clustering,
1086 the DBSCAN algorithm was used. The values of the parameter epsilon were 0.69 for
1087 Quartz-Seq2 on the 4,500 mouse ES/PrE mixture, 5 for Quartz-Seq2 on the 384 mouse ES/PrE
1088 mixture, 3 for Drop-seq on the 500 mouse ES/PrE mixture, and 2.2 for SVF analysis.
1089 Significantly enriched GO terms with the top principal components were calculated using the
1090 GO-PCA package [45].
1091
1092 **Identification of differentially expressed genes for each cluster**
1093 Marker genes for each cluster were identified based on a generalized linear model. The
1094 difference in deviance between two models, in which the gene was or was not differentially
1095 expressed between two clusters, was calculated for the genes. To filter out noise, pseudocount 1
1096 was added to the averaged gene expression in clusters and only genes with fold change of 2 or

1097 more between two clusters were further analyzed. P-values were calculated as 1 − cumulative
1098 density function of a chi-squared continuous random variable for difference in deviance. After
1099 corrections for multiple testing, genes with FDR of less than 0.05 were identified as differentially
1100 expressed genes for the cluster. For comparison between differentially expressed genes
1101 identified by scRNA-seq methods and those identified by bulk RNA-seq, this pseudocount was
1102 not added.
1103
1104 **Identification of variable genes in same cell-type**
1105 To identify genes for which the expression changes depending on the phase of the cell cycle, we
1106 used the intensity of Hoechst 33342 staining measured using a cell sorter. First, cells were
1107 discretized into 40 equal-sized buckets based on the rank of Hoechst 33342 staining intensity.
1108 Then, the CV of averaged UMI counts for a gene in each bin was calculated. After z-scaling of
1109 this, genes with high z-scores were identified. To identify genes whose expression fluctuates in a
1110 manner not relating to the cell cycle phase, we calculated the CV of UMI counts for a gene in
1111 each cell and the CV of averaged UMI counts for a gene in each bin. After z-scaling of such data,
1112 the genes for which the difference between two scaled CVs was large were identified as
1113 "variable genes the expression of which is less associated with the cell cycle phase".
1114
1115 **Enrichment analysis on pathway and Gene Ontology (GO) terms**
1116 Pathways that were particularly enriched for the differentially expressed genes were calculated
1117 using the ReactomePA package of R Bioconductor (version: 1.14.4, [46]) and Metascape
1118 (http://metascape.org/ [47]). The cut-off parameter for the q value was 0.05. Terms that were
1119 enriched for genes with highly variable expression were calculated using DAVID, the Database
1120 for Annotation, Visualization and Integrated Discovery (version: 6.8, [48,49]). Ontology used for
1121 calculation was GOTERM_BP_FAT, GOTERM_MF_FAT, and GOTERM_CC_FAT. The terms
1122 with FDR < 0.05 were identified as enriched terms.
1123
1124 **Quantitative analysis for external RNA molecules**

1125 To calculate the ERCC capture efficiency, we determined the slope of regression between the
1126 input ERCC molecules and detected UMI counts for each ERCC. We tested our calculation
1127 using digital expression matrix (GSM1599501) of the inDrop single-cell sequence method as a
1128 control, and confirmed that our calculated ERCC capture efficiency (7.2%) approximately
1129 matched their reported ERCC capture efficiency (7.1%). To calculate the detection limit of ERCC
1130 molecules, the method of Svensson et al. was used [50]. We confirmed that our calculated
1131 results and those in the paper by Svensson et al. were approximately similar. For calculation of
1132 ERCC capture efficiency and detection limit for previous data, we used the concentration of
1133 ERCC spike-in RNA, which reported by original authors. For the paper by Ziegenhain et al., we
1134 confirmed the concentration by personal communication.
1135
1136 **Simulation-based power analysis**
1137 PowsimR was used to perform power analysis [51]. The parameters of the negative binomial
1138 distribution, namely, mean, dispersion, and dropout probability, were estimated by the function
1139 estimateParam with the following parameters: sigma = 1.96, distribution = "NB," RNAseq =
1140 "singlecell," and normalization = "scran." To simulate differential expression between two groups
1141 using these parameters, the number of genes is represented by the number of detected genes
1142 (count > 0) in each method. Differentially expressed genes set up simulations with number of
1143 detected genes, 23.6% genes being DE, log fold change sample from a narrow gamma
1144 distribution (shape: 1.02, rate: 0.78). That parameter was derived from parameter estimates
1145 based on bulk RNA-seq data from mouse ES and PE cells (n = 3). The function simulateDE was
1146 used to run power analysis with the following parameters: DEmethod = MAST and normalization
1147 = scran. The results of power simulation were plotted using R with ggplot2 packages.
1148
1149 **Data resources**
1150 Raw and processed data files for Quartz-Seq2 and Drop-seq experiments are available under
1151 GEO: GSE99866 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99866) for
1152 single-cell RNA-seq data and DRA002954 for bulk-RNA-seq data. A custom-made python

program (correct_bacode.py) has been deposited under GitHub (DOI: 10.5281/zenodo.1118151, https://github.com/rikenbit/correct_barcode/).

1182 **List of abbreviations**

1183 UMI: unique molecular identifier

1184 PCR: polymerase chain reaction

1185 PCC: Pearson's correlation coefficient

1186 SCC: Spearman's rank correlation coefficient

1187 CV: coefficient of variation

1188 ES cell: embryonic stem cell

1189 PrE cell: primitive endoderm cell

1190 SVF: stromal vascular fraction

1191 RT: reverse transcription

1192 Dex: dexamethasone

1193 PCA: principal component analysis

1194 MSC: mesenchymal stem cell

1195 TdT: terminal deoxynucleotidyl transferase

1196 GO: Gene Ontology

1197 FDR: false discovery rate

1198 t-SNE: t-distributed stochastic neighbor embedding

1199 nt: nucleotide

1200 bp: base pair

1201 SSC: side scatter

1202 TPR: true positive rate

1203 DEG: differentially expressed gene

1204

1205

1206

1207

1208

1209

1210

1211 **Declarations**

1212 **Ethics approval and consent to participate**

1213 Mouse experiments were carried out in accordance with the protocol approved by the National
1214 Institute of Advanced Industrial Science and Technology (AIST) Animal Care and Use
1215 Committee and the RIKEN Animal Experiment Committee.

1216

1217 **Consent for publication**

1218 Not applicable.

1219

1220 **Availability of data and material**

1221 The data have been deposited under GEO accession number GSE99866 and DRA accession
1222 number: DRA002954. A custom-made python program (correct_bacode.py) has been deposited
1223 under GitHub (DOI: 10.5281/zenodo.1118151, https://github.com/rikenbit/correct_barcode/).

1224

1225 **Competing interests**

1226 The authors declare that they have no competing interests.

1227

1228 **Funding**

1229 This work was supported by the Projects for Technological Development, Research Center
1230 Network for Realization of Regenerative Medicine by Japan, the Japan Agency for Medical
1231 Research and Development (AMED), and the Japan Science and Technology Agency (JST).
1232 The sequence analyses of bulk RNA-seq were supported by MEXT KAKENHI (No. 221S0002)
1233 and JSPS KAKENHI Grant Number JP24651218 and JSPS KAKENHI Grant number
1234 JP15K16329. The sequence operations of Quartz-Seq2 using HiSeq2500 were supported by the
1235 Platform Project for Supporting Drug Discovery and Life Science Research (Platform for Drug
1236 Discovery, Informatics, and Structural Life Science) from AMED. This work was partially
1237 supported by JST CREST Grant Number JPMJCR16G3, Japan.

1238

1239 **Authors' contributions**

1240 YS, HD, and IN designed the study. YS designed and developed the Quartz-Seq2 reaction. YS
1241 and KT performed experiments of single-cell RNA-seq. HD, YS, and IN designed and performed
1242 data analysis. HT and AK designed and performed experiments on the mouse stromal vascular
1243 fraction. ME and KT assisted with the experiments. TH assisted with the development of the
1244 spin-down collection system. YS, HD, HT, AK, and IN wrote the manuscript. All authors read and
1245 approved the final manuscript.
1246

1247 **Acknowledgements**
1248 We greatly appreciate the Drop-seq team's useful comments about our technical introduction to
1249 Drop-seq. We thank Mr. Akihiro Matsushima and Mr. Manabu Ishii for their assistance with the
1250 infrastructure for the data analysis. We are also grateful to Dr. Christoph Ziegenhain and Dr.
1251 Wolfgang Enard for giving us their quantitative data from single-cell RNA-seq methods. We also
1252 thank Dr. Rudolf Jaenisch and Dr. Yoichi Shinkai for giving us the J1 ES cell line. Advice and
1253 comments from Dr. Tempei Sato were also of great help in the culture of J1 ES cells. We thank
1254 Dr. Yutaka Suzuki for his assistance with the bulk RNA-seq. We are also grateful to Dr. Mika
1255 Yoshimura and Dr. Haruka Ozaki (RIKEN) for their assistance with the bulk RNA-seq analysis.
1256 Moreover, we thank the staff at the Phyloinformatics Unit of RIKEN Center for Life Science
1257 Technologies (formerly GRAS) for their assistance with the HiSeq2500 sequencer. Our thanks
1258 also go to Sony's cell sorter team for specially customizing the 384-well sorting mode of the
1259 SH800. Finally, we thank Edanz Group (www.edanzediting.com/ac) for editing a draft of this
1260 manuscript.
1261
1262
1263
1264
1265
1266
1267
1268

1269 **References**

1270 1. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel
1271 digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.
1272 2. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A Single-Cell
1273 Transcriptome Atlas of the Human Pancreas. Cell Syst. 2016;3:385–394.e3.
1274 3. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al.
1275 CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016;17:892.
1276 4. Fan X, Zhang X, Wu X, Guo H, Hu Y, Tang F, et al. Single-cell RNA-seq transcriptome
1277 analysis of linear and circular RNAs in mouse preimplantation embryos. Genome Biol.
1278 2015;16:236.
1279 5. Nakamura T, Yabuta Y, Okamoto I, Aramaki S, Yokobayashi S, Kurimoto K, et al. SC3-seq: a
1280 method for highly parallel and quantitative measurement of single-cell gene expression. Nucleic
1281 Acids Res. 2015;43:e60–0.
1282 6. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel
1283 Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell.
1284 2015;161:1202–14.
1285 7. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet Barcoding for
1286 Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell. 2015;161:1187–201.
1287 8. Fan HC, Fu GK, Fodor SPA. Combinatorial labeling of single cells for gene expression
1288 cytometry. Science. 2015;347:1258367–7.
1289 9. Bose S, Wan Z, Carr A, Rizvi A, Vieira G, Pe'er D, et al. Scalable microfluidics for single-cell
1290 RNA printing and sequencing. Genome Biol. 2015;16:120.
1291 10. Streets AM, Zhang X, Cao C, Pang Y, Wu X, Xiong L, et al. Microfluidic single-cell
1292 whole-transcriptome sequencing. Proc. Natl. Acad. Sci. 2014;111:7048–53.
1293 11. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively
1294 Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types.
1295 Science. 2014;343:776–9.
1296 12. Soumillon M, Cacchiarelli D, Semrau S, van Oudenaarden A, Mikkelsen TS. Characterization
1297 of directed differentiation by high-throughput single-cell RNA-Seq. 2014.

1298 13. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell
1299 RNA-seq with unique molecular identifiers. Nat. Methods. 2013;11:163–6.
1300 14. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for
1301 sensitive full-length transcriptome profiling in single cells. Nat. Methods. 2013;10:1096–8.
1302 15. Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T, et al. Quartz-Seq: a highly
1303 reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic
1304 gene-expression heterogeneity. Genome Biol. 2013;14:R31.
1305 16. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed
1306 Linear Amplification. Cell Rep. 2012;2:666–73.
1307 17. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq
1308 from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol.
1309 2012;30:777–82.
1310 18. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, et al. Characterization of the
1311 single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21:1160–
1312 7.
1313 19. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq
1314 whole-transcriptome analysis of a single cell. Nat. Methods. 2009;6:377–82.
1315 20. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell
1316 transcriptomics. Nat. Methods. 2014;11:637–40.
1317 21. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al.
1318 Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell. 2017;65:631–4.
1319 22. Buschmann T. DNABarcodes: an R package for the systematic construction of DNA sample
1320 tags. Bioinformatics. 2017;:btw759.
1321 23. Shimosato D, Shiki M, Niwa H. Extra-embryonic endoderm cells derived from ES cells
1322 induced by GATA Factors acquire the character of XEN cells. BMC Dev. Biol. 2007;7:80.
1323 24. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification
1324 of low quality cells from single-cell RNA-seq data. Genome Biol. 2016;17:57.
1325 25. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee S-L, et al. Zscan4 regulates
1326 telomere elongation and genomic stability in ES cells. Nature. 2010;464:858–63.

1327 26. Pak J, Lee JH, Park KS, Park M, Kang L-W, Lee SH. Current use of autologous adipose
1328 tissue-derived stromal vascular fraction cells for orthopedic applications. J. Biomed. Sci.
1329 2017;24:1496.
1330 27. Trounson A, McDonald C. Stem Cell Therapies in Clinical Trials: Progress and Challenges.
1331 Cell Stem Cell. 2015;17:11–22.
1332 28. Baer PC, Geiger H. Adipose-Derived Mesenchymal Stromal/Stem Cells: Tissue Localization,
1333 Characterization, and Heterogeneity. Stem Cells Int. 2012;2012:1–11.
1334 29. Lindroos B, Suuronen R, Miettinen S. The Potential of Adipose Stem Cells in Regenerative
1335 Medicine. Stem Cell Rev. 2010;7:269–91.
1336 30. Kwon HM, Hur S-M, Park K-Y, Kim C-K, Kim Y-M, Kim H-S, et al. Multiple paracrine factors
1337 secreted by mesenchymal stem cells contribute to angiogenesis. Vascular Pharmacology.
1338 2014;63:19–28.
1339 31. Madrigal M, Rao KS, Riordan NH. A review of therapeutic effects of mesenchymal stem cell
1340 secretions and induction of secretory modification by different culture methods. J Transl Med.
1341 2014;12:71.
1342 32. Rennert RC, Januszyk M, Sorkin M, Rodrigues M, Maan ZN, Duscher D, et al. Microfluidic
1343 single-cell transcriptional analysis rationally identifies novel surface marker profiles to enhance
1344 cell-based therapies. Nat Commun. 2016;7:11945.
1345 33. Baer PC. Adipose-derived mesenchymal stromal/stem cells: An update on their phenotype in
1346 vivo and in vitro. World J Stem Cells. 2014;6:256.
1347 34. Mabuchi Y, Houlihan DD, Akazawa C, Okano H, Matsuzaki Y. Prospective Isolation of
1348 Murine and Human Bone Marrow Mesenchymal Stem Cells Based on Surface Markers. Stem
1349 Cells Int. 2013;2013:1–7.
1350 35. Dromard CC, Barreau C, Andr M, Berger-M ller S, Casteilla L, Planat-Benard V. Mouse
1351 adipose tissue stromal cells give rise to skeletal and cardiomyogenic cell sub-populations. Front
1352 Cell Dev Biol. 2014;2:697.
1353 36. Heneidi S, Simerman AA, Keller E, Singh P, Li X, Dumesic DA, et al. Awakened by Cellular
1354 Stress: Isolation and Characterization of a Novel Population of Pluripotent Stem Cells Derived
1355 from Human Adipose Tissue. Ljubimov AV, editor. PLoS ONE. 2013;8:e64752.

1356 37. Wachi H. Role of Elastic Fibers on Cardiovascular Disease. J. Health Sci. 2011;57:449–57.
1357 38. Hughes-Fulford M, Li C-F. The role of FGF-2 and BMP-2 in regulation of gene induction, cell
1358 proliferation and mineralization. J Orthop Surg Res. 2011;6:8.
1359 39. Kurimoto K. An improved single-cell cDNA amplification method for efficient high-density
1360 oligonucleotide microarray analysis. Nucleic Acids Res. 2006;34:e42–2.
1361 40. Yang L, Ma Z, Cao C, Zhang Y, Wu X, Lee R, et al. MR-Seq: measuring a single cell's
1362 transcriptome repeatedly by RNA-seq. Science Bulletin. 2017.
1363 41. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human
1364 haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol.
1365 2017;Complex Sy:1695.
1366 42. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq
1367 reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem
1368 cells. Genome Res. 2015;25:1860–72.
1369 43. Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, et al. Quantitative
1370 assessment of single-cell RNA-sequencing methods. Nat. Methods. 2013;11:41–6.
1371 44. Pauklin S, Vallier L. The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity.
1372 Cell. 2013;155:135–47.
1373 45. Wagner F. GO-PCA: An Unsupervised Method to Explore Gene Expression Data Using Prior
1374 Knowledge. Li X, editor. PLoS ONE. 2015;10:e0143196.
1375 46. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis
1376 and visualization. Mol. BioSyst. 2016;12:477–9.
1377 47. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and
1378 Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding.
1379 Cell Host Microbe. 2015;18:723–35.
1380 48. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
1381 lists using DAVID bioinformatics resources. Nat Protoc. 2008;4:44–57.
1382 49. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the
1383 comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2008;37:1–13.

50. 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.

51. Vieth B, Ziegenhain C, Parekh S, Enard W, Hellmann I. powsimR: power analysis for bulk and single cell RNA-seq experiments. Bioinformatics. 2017;33:3486–8.

1413 **Figure legends**

1414 **Figure 1. Overview of Quartz-Seq2 experimental processes**

1415 a) Quartz-Seq2 consisted of five steps. (1) Each single cell in a droplet was sorted into lysis
1416 buffer in each well of a 384-well PCR plate using flow cytometry analysis data. (2)
1417 Poly-adenylated RNA in each well was reverse-transcribed into first-strand cDNA with reverse
1418 transcription primer, which had a unique cell barcode. We prepared 384 or 1,536 kinds of cell
1419 barcode (CB) with a unique sequence based on the Sequence–Levenshtein distance (SeqLv).
1420 The edit distance of SeqLv was 5. RT primer also had a UMI sequence for reduction of PCR bias
1421 (MB) and a poly-dT sequence for binding to poly-A RNA. (3) Cell-barcode-labeled cDNAs from
1422 all 384 wells were promptly collected by centrifugation using assembled collectors. (4) Collected
1423 first-strand cDNAs were purified and concentrated for subsequent whole-transcript amplification.
1424 In the poly-A tailing step, purified cDNA was extended with a poly-A tail by terminal
1425 deoxynucleotidyl transferase (TdT). Subsequently, second-strand cDNA was synthesized with a
1426 tagging primer, which had a poly-dT sequence. The resulting second-strand cDNA had a PCR
1427 primer sequence (M) at both ends of it. The cDNA was amplifiable in subsequent PCR
1428 amplification. (5) For conversion from amplified cDNA to sequence library DNA, we fragmented
1429 the amplified cDNA using the ultrasonicator Covaris. Such fragmented cDNA was ligated with a
1430 truncated Y-shaped sequence adaptor, which had an Illumina flow-cell binding sequence (P7)
1431 and a pool barcode sequence (PB). PB made it possible to mix different sets of
1432 cell-barcode-labeled cDNA. Ligated cDNA, which had CB and MB sequences, was enriched by
1433 PCR amplification. The resulting sequence library DNA contained P7 and P5 flow-cell binding
1434 sequences at respective ends of the DNA. We sequenced the cell-barcode site and the UMI site
1435 at Read1, the pool-barcode site at Index1, and the transcript sequence at Read2. b) The
1436 relationship between initial fastq reads and the number of single cells for sequence analysis in
1437 NextSeq500 runs. Typically, one sequence run with NextSeq 500/550 High Output v2 Kit reads
1438 out 400–450 M fastq reads. The x-axis represents the input cell number for one sequence run.
1439 The y-axis represents the initial data size (fastq reads) on average per cell. Red circle represents
1440 the typical range of shallow input read depth for a single cell. c) We define the formula for
1441 calculating the UMI conversion efficiency. Each parameter is defined as follows: UMI_{sc}: the

1442 number of UMI count, assigned to a single-cell sample, $fastq_{sc}$: the number of fastq reads
1443 derived from each single-cell sample, $fastq_{non\text{-}sc}$: the number of fastq reads derived from
1444 non-single-cell samples, which include experimental byproducts such as WTA adaptors, WTA
1445 byproducts, and non-STAMPs. Initial fastq reads are composed of $fastq_{sc}$ and $fastq_{non\text{-}sc}$.
1446
1447 **Figure 2. Sequence performance of Quartz-Seq2 with molecular biological improvements**
1448 a) Improvement of poly-A tagging efficiency. Bar plot represents the relative DNA yield in various
1449 poly-A tagging conditions using purified first-strand cDNA from 1 ng of total RNA. T55 buffer as
1450 the terminal deoxynucleotidyl transferase (TdT) buffer and the temperature condition “Increment”
1451 for the poly-A tagging step improved the cDNA yield of whole-transcript amplification. Buffer
1452 compositions are indicated in Additional file 3: Table S2. QuartzB represents use of a
1453 Quartz-Seq-like buffer as a positive control, in accordance with the approach described in the
1454 original Quartz-Seq paper. Finally, we quantified cDNA yield (300–9,000 bp) and byproduct DNA
1455 yield (50–300 bp) using Bioanalyzer (Agilent). The presented p-value was obtained using
1456 two-tailed Welch's t-test. b) Reverse transcription efficiency with serially diluted RT enzymes.
1457 The x-axis represents the average relative RT qPCR score from 10 genes. Detailed
1458 concentrations of RT enzymes are presented in Additional file 1: Figure S7. c,f–h) Comparison
1459 between Quartz-Seq2 in the RT25 condition and Quartz-Seq-like conditions regarding sequence
1460 performance. c) We analyzed 384 wells with 10 pg of total RNA and used approximately 0.19 M
1461 fastq reads on average per well. We show the UMI count and gene count in box plots. d) The
1462 panel shows a scatter plot between the mean of gene expression and the variability of gene
1463 expression with 10 pg of total RNA in 384 wells. Red lines represent the theoretical variability of
1464 gene expression in the form of a Poisson distribution. e) Gene expression reproducibility
1465 between bulk poly-A-RNA-seq (1 μg of total RNA) and Quartz-Seq2 (10 pg of total RNA,
1466 averaged over 384 wells). f) Dispersion of gene expression. The x-axis represents gene
1467 expression variability. g) Reproducibility of gene expression for internal gene and external
1468 control RNA. h) Accuracy of gene expression for internal gene and external control RNA.
1469

1470 **Figure 3. Quartz-Seq2 with high UMI conversion efficiency identified more differentially**
1471 **expressed genes and biological pathways**
1472 a) UMI counts, gene counts, and UMI conversion efficiency for Quartz-Seq2 and Drop-seq
1473 experiments. These values depended on the initial fastq reads on average per cell. b) We
1474 calculated overlapping differentially expressed genes between bulk RNA-seq data and
1475 single-cell RNA-seq data. We randomly picked up the indicated number of single cells and
1476 calculated differentially expressed genes 20 times. c,d) We randomly selected 100 ES cells and
1477 100 PrE cells for each method. c) Venn diagram of genes that were differentially expressed
1478 between the ES cluster and the Dex-treated ES (PrE) cluster, as identified by Quartz-Seq2,
1479 Drop-seq, and bulk RNA-seq (left). The number of genes that differed in expression level
1480 between ES and PrE cells by at least twofold was determined (FDR < 0.05). We also present a
1481 Venn diagram of the Reactome pathway (right). d) Venn diagram of GO terms for genes with
1482 highly variable expression, as identified by Quartz-Seq2 and Drop-seq.
1483
1484 **Figure 4. Quantitative comparison among Quartz-Seq2 and previously reported methods**
1485 **using embryonic stem cells**
1486 a) We determined the UMI and gene counts with Quartz-Seq2 in the “RT25” condition using J1
1487 ES cells. We performed Quartz-Seq2 (RT25) with three sets of 384-well plates and two sets of
1488 384-well plates on different days. We also estimated the UMI conversion efficiency of other
1489 single-cell RNA-seq methods [CEL-seq2(C1), SCRB-seq, MARS-seq, and Drop-seq] from a
1490 previous study that used mouse ES cells [21]. In our comparison, the Read2 length for transcript
1491 mapping was 45 nt for all of the methods, including Quartz-Seq2. We estimated the average UMI
1492 and gene counts and the UMI conversion efficiency with various numbers of initial fastq reads for
1493 each method. The findings indicate that, compared with the other methods, Quartz-Seq2 has a
1494 superior ability to detect UMI and gene counts from limited initial amounts of data (under 0.2
1495 million fastq reads). b) To investigate the throughput capacity for establishing sequence library
1496 DNA, we estimated the number of processable single cells per $1,000 spent on each method:
1497 Quartz-Seq2 (384 indexes, RT25) yielded 1,785 cells, Quartz-Seq2 (384 indexes, RT6.25)
1498 yielded 2,325 cells, Quartz-Seq2 (384 indexes, RT6.25+column1) yielded 3,058 cells,

1499 Quartz-Seq2 (1,536 indexes, RT25) yielded 2,500 cells, CEL-seq2(C1) yielded 111 cells,
1500 SCRB-seq yielded 500 cells, MARS-seq yielded 769 cells, and Drop-seq yielded 10,000 cells.
1501 The UMI conversion efficiency was approximately 32.55% (n = 1,152), 32.25% (n = 768) and
1502 32.12% (n = 192) for Quartz-Seq2 (384 indexes, RT25), 35.48% (n = 2,304) for Quartz-Seq2
1503 (1,536 indexes, RT25), 34.04% (n = 768) for Quartz-Seq2 (384 indexes, RT6.25), 35.51% (n =
1504 768) for Quartz-Seq2 (384 indexes, RT6.25+column1), 22.4% for CEL-seq2(C1), 13.3% for
1505 SCRB-seq, 10.6% for MARS-seq, and 7.1% for Drop-seq. The average gene count was
1506 approximately 6,636 (n = 1,152), 6,584 (n = 768) and 6,529 (n = 192) for Quartz-Seq2 (384
1507 indexes, RT25), 6,712 (n = 2,304) for Quartz-Seq2 (1,536 indexes, RT25), 6,753 (n = 768) for
1508 Quartz-Seq2 (384 indexes, RT6.25), 6,794 (n = 768) for Quartz-Seq2 (384 indexes,
1509 RT6.25+column1), 5,164 for CEL-seq2(C1), 4,044 for SCRB-seq, 3,252 for MARS-seq, and
1510 2,738 for Drop-seq.
1511
1512 **Figure 5. High-throughput Quartz-Seq2 analysis of 4,484 cells from mouse embryonic**
1513 **stem cells and differentiated cells**
1514 a) We successfully analyzed 97.3% of 4,608 wells. The procedures for cell suspension as used
1515 in this assay are shown. Cells cultured under ES-maintenance and Dex-treatment conditions
1516 were separately dissociated into single cells, stained with Hoechst 33342 and/or Calcein-AM,
1517 and mixed evenly. Calcein-AM-positive and -negative cells were sorted to 384-well plates in a
1518 checkered pattern. White scale bars represent 100 μm. b) Clustering of 4,484 single cells
1519 according to the transcriptome. Plotting of cells on t-SNE space with color labeling for each
1520 cluster. The percentage indicates the proportion of cells for each cluster relative to all cells
1521 analyzed. Numbers in parentheses indicate the numbers of cells making up the cluster. c)
1522 Marker genes for each cluster identified by Quartz-Seq2. Cluster-specific or cluster-enriched
1523 genes were calculated for each cluster, and their expression is displayed as a color in a heatmap.
1524 No more than 50 cells are shown for simplicity. d) Reconstructed distribution of Calcein-AM
1525 intensity for each cluster. The x-axis represents the intensity of Calcein-AM dye staining. e)
1526 Reconstructed distribution of Hoechst 33342 intensity for each cluster. The y-axis represents the
1527 density of cells. The x-axis represents the intensity of Hoechst 33342 dye staining.

1528

1529 **Figure 6 Quartz-Seq2 classified genes with variable expression within a cell type**

1530 a) Plotting cells on t-SNE space with the colors representing the intensity of Hoechst 33342
1531 staining. b) Dependence of expression of several genes on the intensity of Hoechst 33342
1532 staining [G_2/M phase: *Ccnb1*; G_1 phase: *Ccne1*; housekeeping gene: *Gnb2l1*; upper panels: ES
1533 main population (cluster 1); lower panels: PrE main population (cluster 4)]. c) Identification of
1534 cell-cycle-associated genes. Average gene expression of cells in a bin is depicted by color in the
1535 heatmap. Bins are arranged in order of Hoechst 33342 staining intensity. d) The number of
1536 cell-cycle-associated genes identified for each z-threshold. CV values from which the z-score
1537 was calculated were as follows. ES cells (z=4: 1.17, z=3: 0.91, z=2, 0.68, z=1: 0.4). PrE cells
1538 (z=4: 1.1, z=3: 0.88, z=2: 0.64, z=1: 0.39). e) Several examples for genes whose expression was
1539 variable, but not associated with the phase of the cell cycle. f) Single-cell qPCR detection of
1540 genes with variable expression.

1541

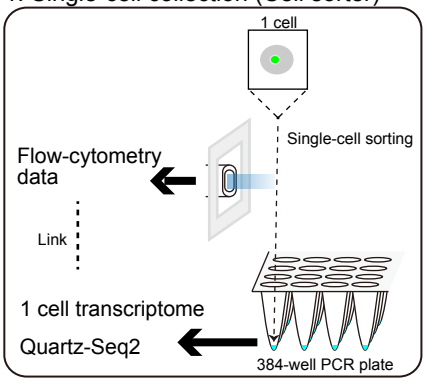
1542 **Figure 7 Quartz-Seq2 analysis of stromal vascular fraction (SVF) from mouse adipose**
1543 **tissue**

1544 a) Morphology of SVF cells. Adipose tissue from a cell suspension of SVF was prepared. Upper
1545 panels present a photograph of adipose tissues and dissociated SVF samples. Yellow scale bar
1546 represents 1 cm. White scale bar represents 10 μm. Lower panels represent the distribution of
1547 cell size information with different platforms (left: diameter of cell size using photography, right:
1548 flow cytometry information using a cell sorter). The diameter of cell size for SVF samples was
1549 6.43±1.35 μm (n=200). b) Clustering of cells included in SVF. The transcriptome of
1550 approximately 1,000 cells was quantified by Quartz-Seq2 and clustering on t-SNE space was
1551 performed. In accordance with the genes and functional terms enriched in each cluster, the cell
1552 type was annotated. The percentage indicates the proportion of cells for each cluster relative to
1553 all cells analyzed. Numbers in parentheses indicate the numbers of cells constituting the cluster.
1554 c) Marker genes for each cluster were identified by Quartz-Seq2. Cluster-specific or
1555 cluster-enriched genes were calculated for each cluster, with their expression being displayed as
1556 color in a heatmap. No more than 50 cells are shown for simplicity. d) The results of GO-PCA

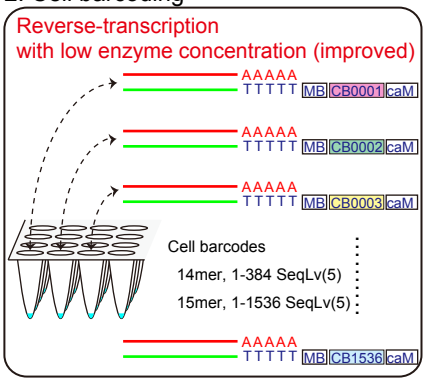
1557 analysis. Functional terms enriched in the genes with high factor loadings of PCA were
1558 calculated and the enrichment is displayed as color in the heatmap. No more than 50 cells are
1559 shown for simplicity. e) Reactome pathway with genes differentially expressed between cluster 1
1560 and cluster 8.

a)

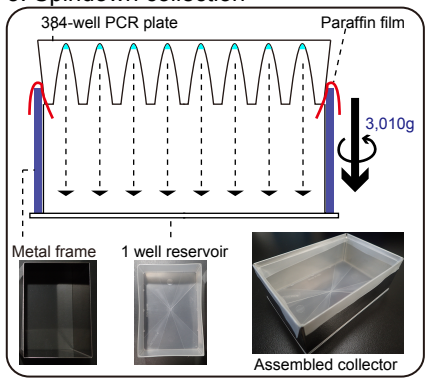
1. Single-cell collection (Cell sorter)



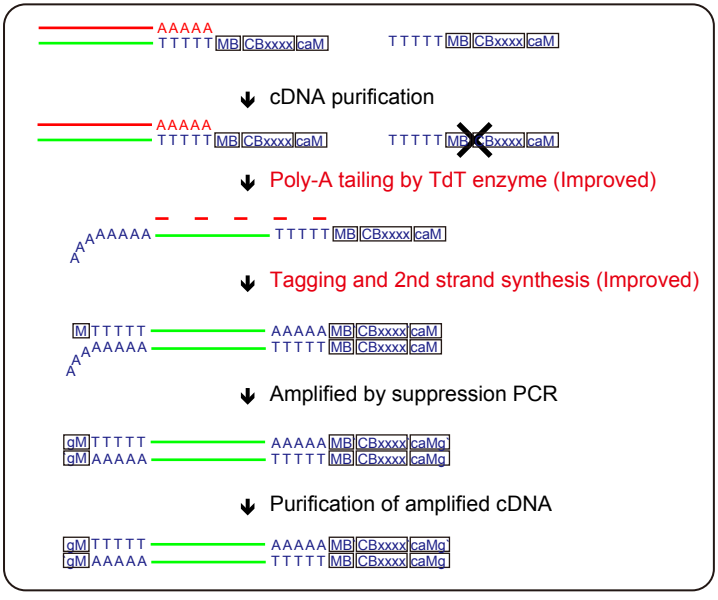
2. Cell barcoding



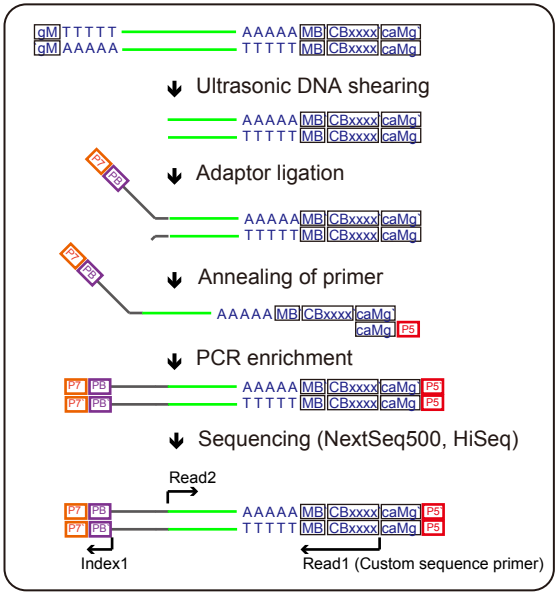
3. Spindown collection



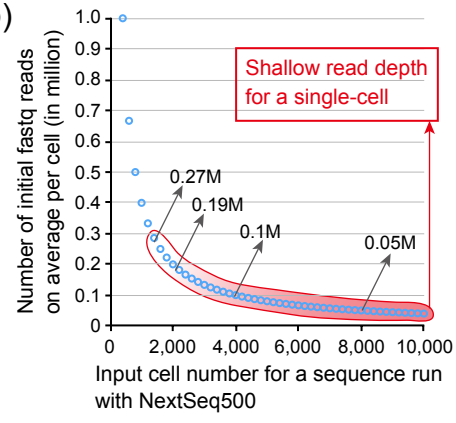
4. Whole transcript amplification based on poly-A tagging



5. Sequence library preparation

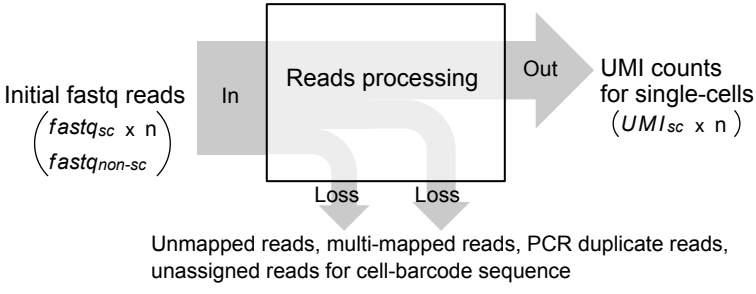


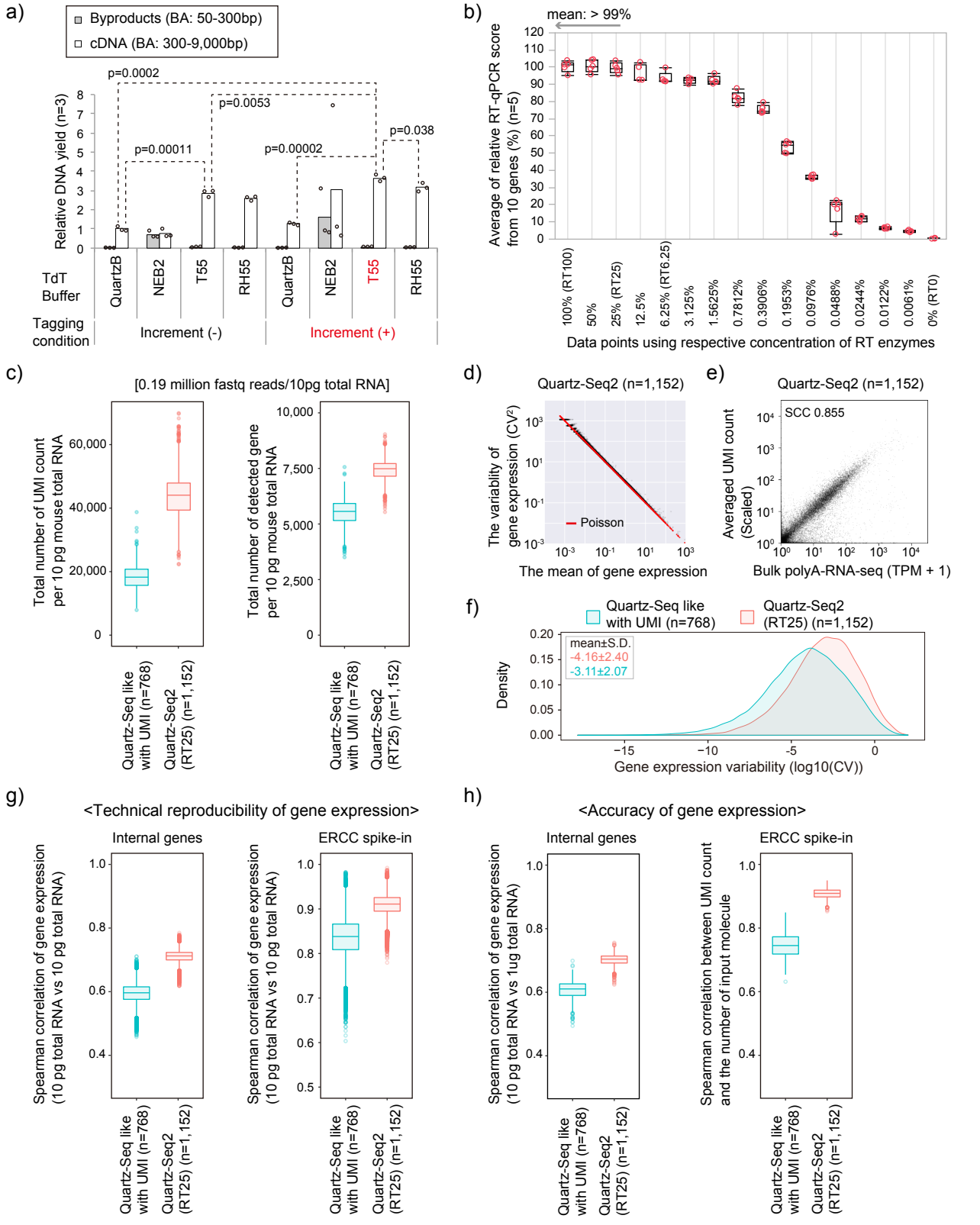
b)

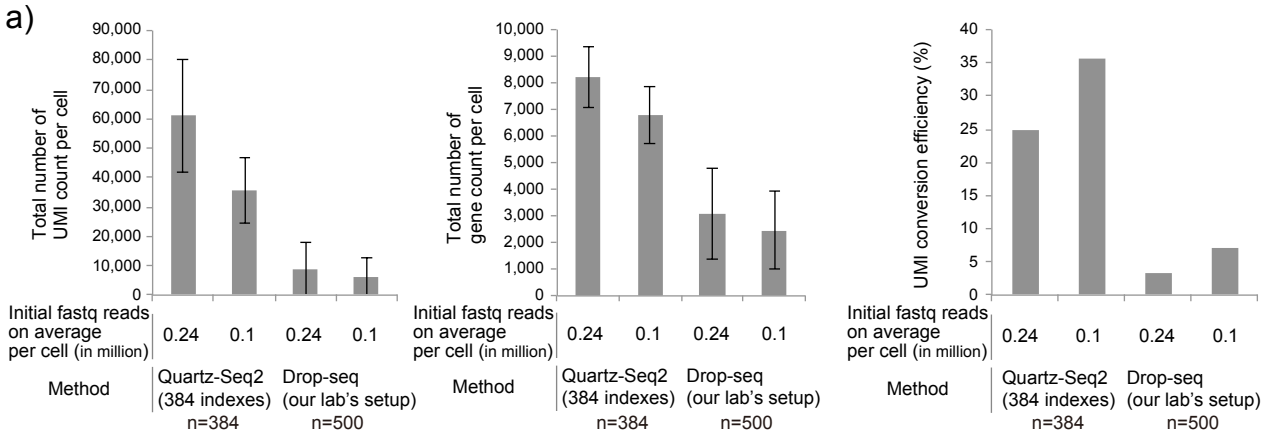


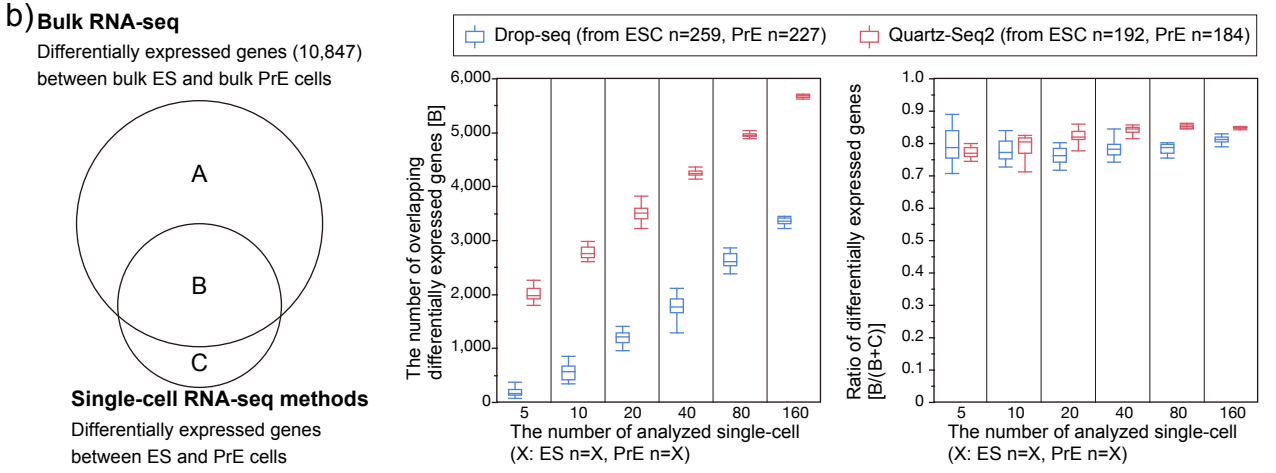
c)

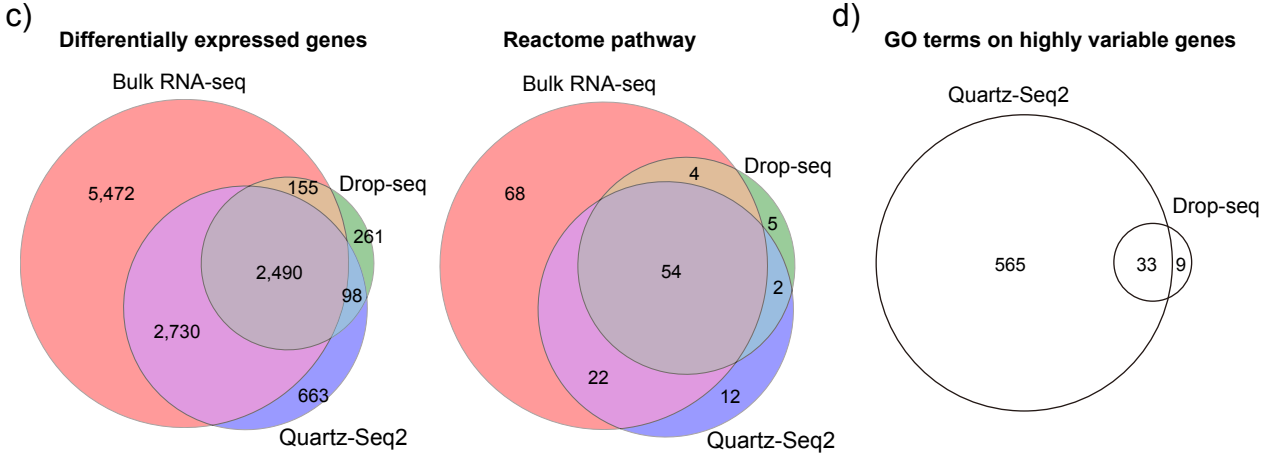
$$\text{UMI conversion efficiency} = \frac{\sum UMI_{sc}}{\sum fastq_{sc} + fastq_{non\text{-}sc}}$$



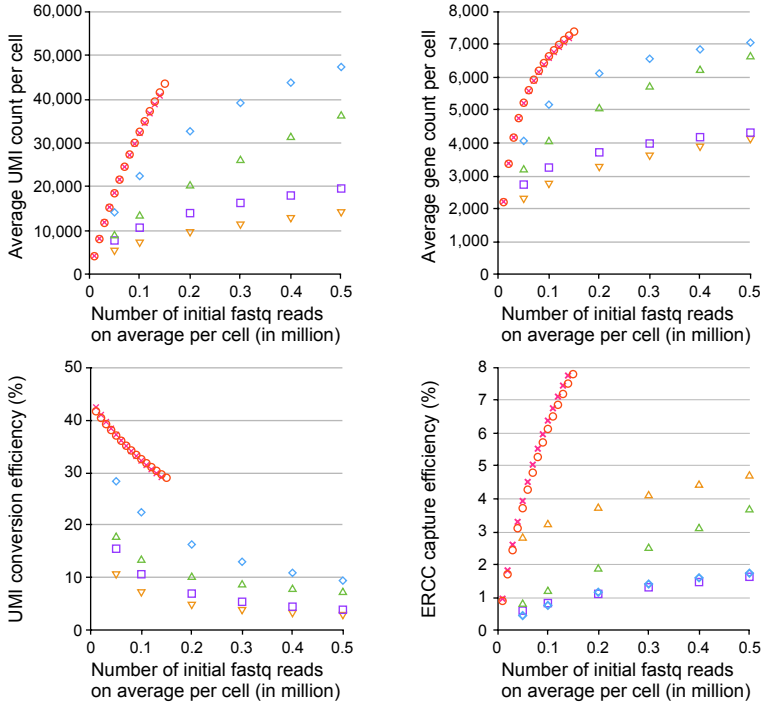

a)
Byproducts (BA: 50-300bp)
cDNA (BA: 300-9,000bp)
p=0.0002
p=0.0053
p=0.038
p=0.00011
p=0.00002
Relative DNA yield (n=3)
TdT Buffer
QuartzB
NEB2
T55
RH55
QuartzB
NEB2
T55
RH55
Tagging condition
Increment (-)
Increment (+)
b)
mean: > 99%
Average of relative RT-qPCR score from 10 genes (%) (n=5)
100% (RT100)
50%
25% (RT25)
12.5%
6.25% (RT6.25)
3.125%
1.5625%
0.7812%
0.3906%
0.1953%
0.0976%
0.0488%
0.0244%
0.0122%
0.0061%
0% (RT0)
Data points using respective concentration of RT enzymes
c)
[0.19 million fastq reads/10pg total RNA]
Total number of UMI count per 10 pg mouse total RNA
Total number of detected gene per 10 pg mouse total RNA
Quartz-Seq like with UMI (n=768)
Quartz-Seq2 (RT25) (n=1,152)
d)
Quartz-Seq2 (n=1,152)
The variability of gene expression (CV²)
Poisson
The mean of gene expression
e)
Quartz-Seq2 (n=1,152)
SCC 0.855
Averaged UMI count (Scaled)
Bulk polyA-RNA-seq (TPM + 1)
f)
Quartz-Seq like with UMI (n=768)
Quartz-Seq2 (RT25) (n=1,152)
mean±S.D.
-4.16±2.40
-3.11±2.07
Density
Gene expression variability (log10(CV))
g)
<Technical reproducibility of gene expression>
Internal genes
ERCC spike-in
Spearman correlation of gene expression (10 pg total RNA vs 10 pg total RNA)
Quartz-Seq like with UMI (n=768)
Quartz-Seq2 (RT25) (n=1,152)
h)
<Accuracy of gene expression>
Internal genes
ERCC spike-in
Spearman correlation of gene expression (10 pg total RNA vs 1ug total RNA)
Spearman correlation between UMI count and the number of input molecule
Quartz-Seq like with UMI (n=768)
Quartz-Seq2 (RT25) (n=1,152)


a)
Total number of UMI count per cell
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
Total number of gene count per cell
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
UMI conversion efficiency (%)
40
35
30
25
20
15
10
5
0
Initial fastq reads on average per cell (in million)
0.24
0.1
0.24
0.1
Method
Quartz-Seq2 (384 indexes)
Drop-seq (our lab's setup)
n=384
n=500


b) Bulk RNA-seq
Differentially expressed genes (10,847) between bulk ES and bulk PrE cells
A
B
C
Single-cell RNA-seq methods
Differentially expressed genes between ES and PrE cells
Drop-seq (from ESC n=259, PrE n=227)
Quartz-Seq2 (from ESC n=192, PrE n=184)
The number of overlapping differentially expressed genes [B]
6,000
5,000
4,000
3,000
2,000
1,000
0
Ratio of differentially expressed genes [B/(B+C)]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
20
40
80
160
The number of analyzed single-cell (X: ES n=X, PrE n=X)


c)
Differentially expressed genes
Bulk RNA-seq
Drop-seq
Quartz-Seq2
5,472
155
261
2,490
98
2,730
663
Reactome pathway
Bulk RNA-seq
Drop-seq
Quartz-Seq2
68
4
5
54
2
22
12
d)
GO terms on highly variable genes
Quartz-Seq2
Drop-seq
565
33
9

a)



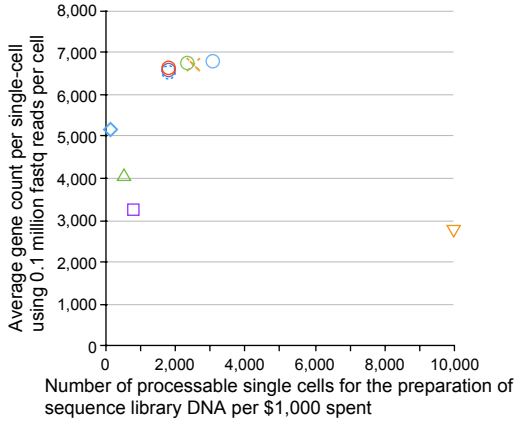
J1 ES cell 2i/LIF, HiSeq platform

Our data	○ Quartz-Seq2, Day1 (n=1,152)
	× Quartz-Seq2, Day2 (n=768)
Data from *Ziegenhain et al. 2017*	◇ CEL-seq2(C1) (n=71)
	△ SCRB-seq (n=84)
	□ MARS-seq (n=65)
	▽ Drop-seq (w/o ERCC) (n=76)

without cell (ERCC data), NextSeq platform

Data from *Macosko et al. 2015*	△ Drop-seq (w ERCC) (n=75)

b)



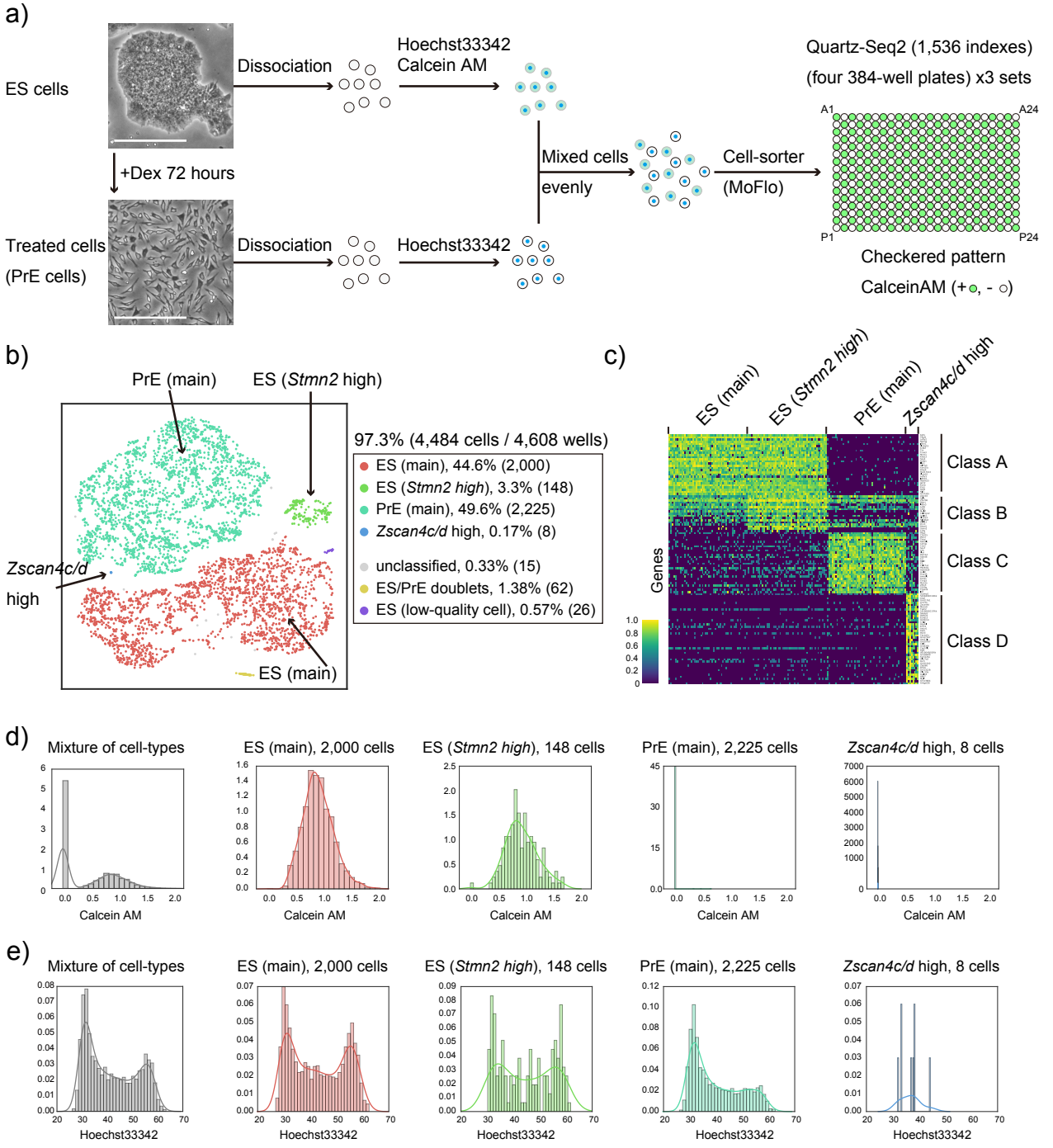
Our data

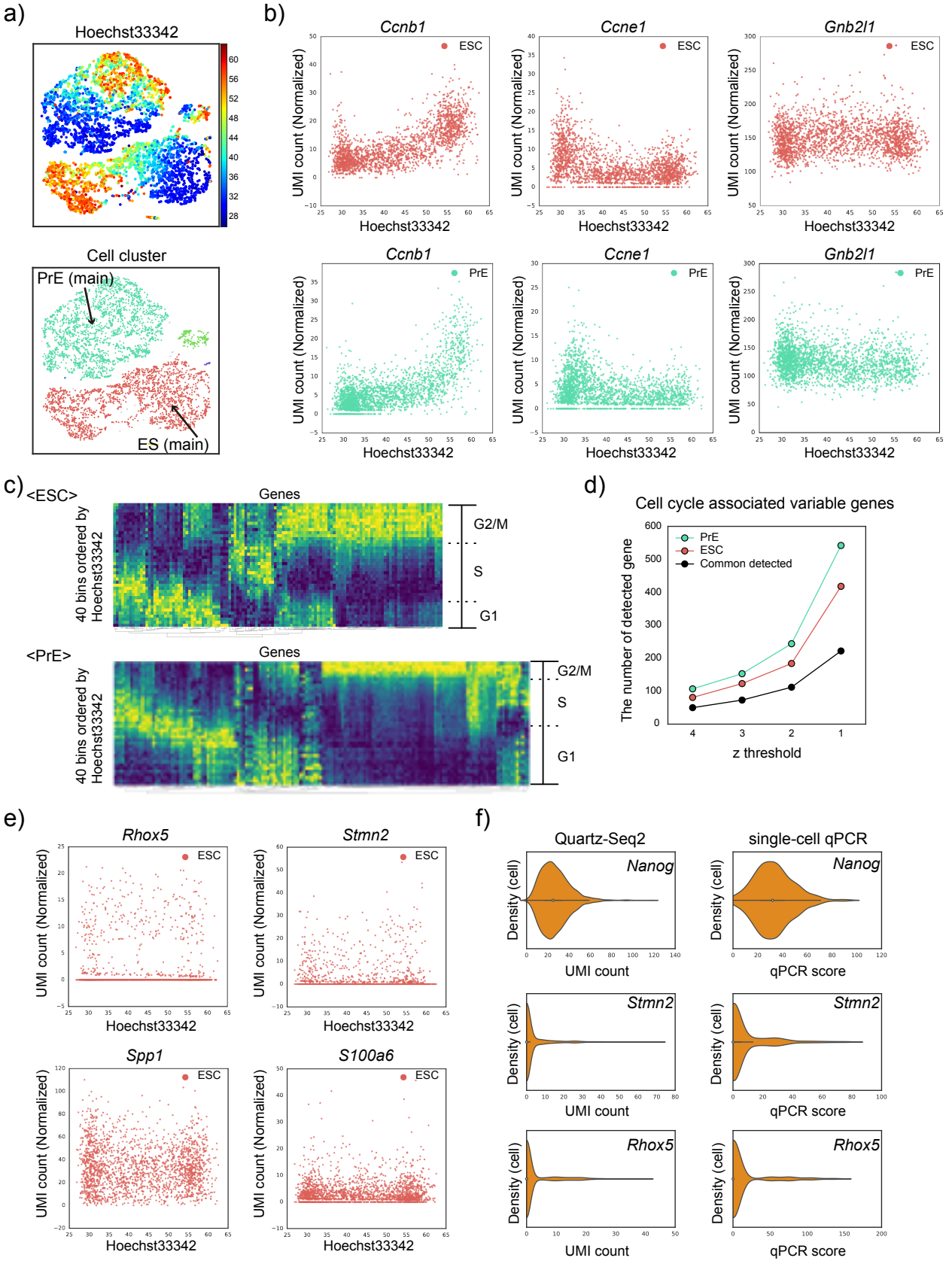
Cell line	Method name	Conditions	Index type
J1 ES cell ○	Quartz-Seq2 default (n=1,152)*	RT25	384 indexes
J1 ES cell ○	Quartz-Seq2 default 2 (n=768)*	RT25	384 indexes
J1 ES cell ○	Quartz-Seq2 RT6.25 (n=768)**	RT6.25	384 indexes
J1 ES cell ○	Quartz-Seq2 cheapest (n=768)**	RT6.25+Column1	384 indexes
G6GR ES cell ⊙	Quartz-Seq2 default (n=192)**	RT25	384 indexes
G6GR ES cell ×	Quartz-Seq2 default (n=2,304)**	RT25	1,536 indexes

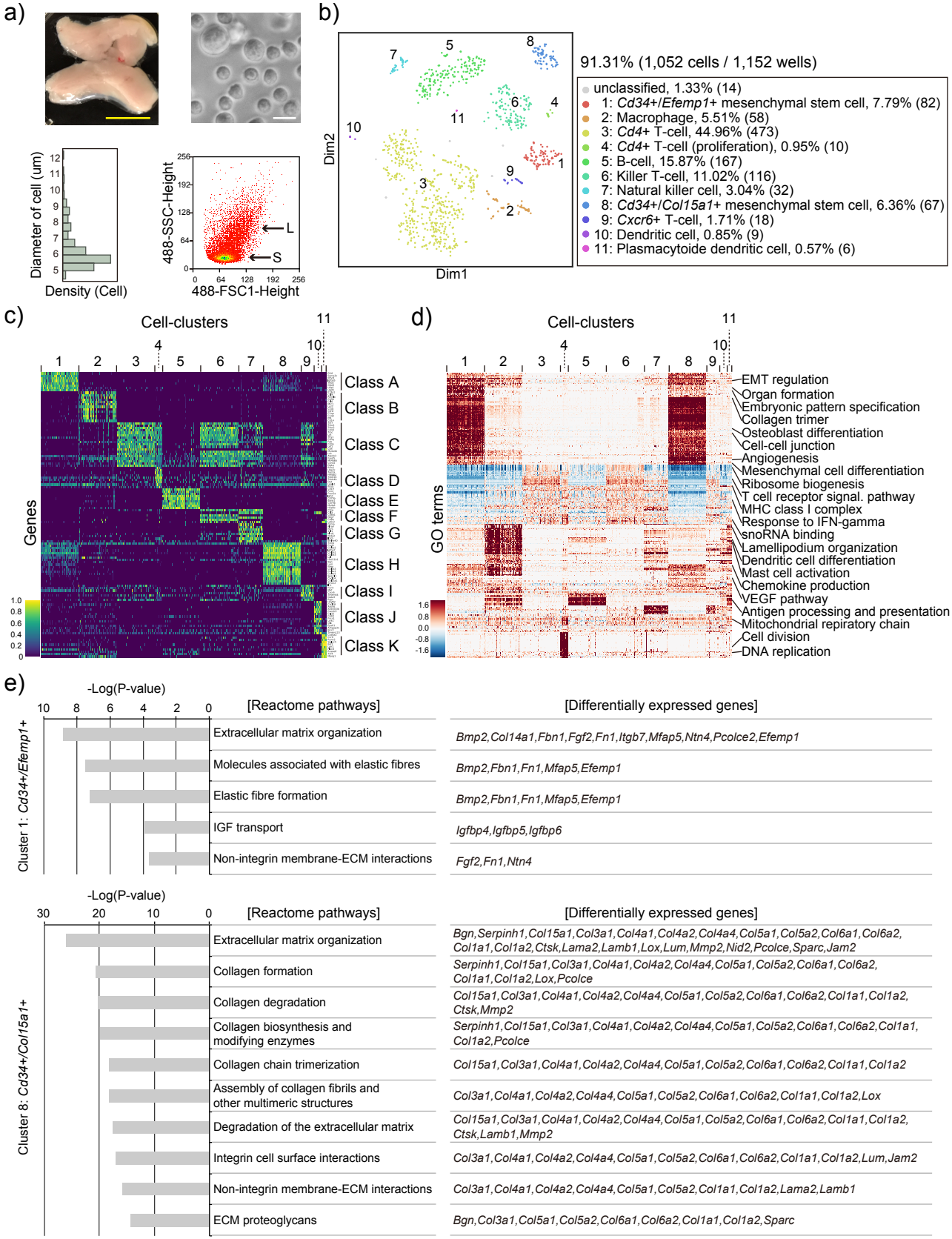
Data from *Ziegenhain et al. 2017*

Cell line	Method name
J1 ES cell ◇	CEL-seq2(C1) (n=71)*
J1 ES cell △	SCRB-seq (n=84)*
J1 ES cell □	MARS-seq (n=65)*
J1 ES cell ▽	Drop-seq (n=76)*

* HiSeq platform ** NextSeq platform


a)
ES cells
Dissociation
Hoechst33342
Calcein AM
+Dex 72 hours
Treated cells
(PrE cells)
Dissociation
Hoechst33342
Mixed cells
evenly
Cell-sorter
(MoFlo)
Quartz-Seq2 (1,536 indexes)
(four 384-well plates) x3 sets
A1
A24
P1
P24
Checkered pattern
CalceinAM (+, -)
b)
PrE (main)
ES (Stmn2 high)
Zscan4c/d
high
ES (main)
97.3% (4,484 cells / 4,608 wells)
ES (main), 44.6% (2,000)
ES (Stmn2 high), 3.3% (148)
PrE (main), 49.6% (2,225)
Zscan4c/d high, 0.17% (8)
unclassified, 0.33% (15)
ES/PrE doublets, 1.38% (62)
ES (low-quality cell), 0.57% (26)
c)
ES (main)
ES (Stmn2 high)
PrE (main)
Zscan4c/d high
Genes
Class A
Class B
Class C
Class D
d)
Mixture of cell-types
ES (main), 2,000 cells
ES (Stmn2 high), 148 cells
PrE (main), 2,225 cells
Zscan4c/d high, 8 cells
Calcein AM
e)
Mixture of cell-types
ES (main), 2,000 cells
ES (Stmn2 high), 148 cells
PrE (main), 2,225 cells
Zscan4c/d high, 8 cells
Hoechst33342


a)
Hoechst33342
Cell cluster
PrE (main)
ES (main)
b)
Ccnb1
Ccne1
Gnb2l1
ESC
PrE
UMI count (Normalized)
Hoechst33342
c)
<ESC>
<PrE>
Genes
40 bins ordered by Hoechst33342
G2/M
S
G1
d)
Cell cycle associated variable genes
PrE
ESC
Common detected
The number of detected gene
z threshold
e)
Rhox5
Stmn2
Spp1
S100a6
ESC
UMI count (Normalized)
Hoechst33342
f)
Quartz-Seq2
single-cell qPCR
Nanog
Stmn2
Rhox5
Density (cell)
UMI count
qPCR score


a)
Diameter of cell (um)
Density (Cell)
488-SSC-Height
488-FSC1-Height
L
S
b)
Dim1
Dim2
91.31% (1,052 cells / 1,152 wells)
unclassified, 1.33% (14)
1: Cd34+/Efemp1+ mesenchymal stem cell, 7.79% (82)
2: Macrophage, 5.51% (58)
3: Cd4+ T-cell, 44.96% (473)
4: Cd4+ T-cell (proliferation), 0.95% (10)
5: B-cell, 15.87% (167)
6: Killer T-cell, 11.02% (116)
7: Natural killer cell, 3.04% (32)
8: Cd34+/Col15a1+ mesenchymal stem cell, 6.36% (67)
9: Cxcr6+ T-cell, 1.71% (18)
10: Dendritic cell, 0.85% (9)
11: Plasmacytoide dendritic cell, 0.57% (6)
c)
Cell-clusters
Genes
Class A
Class B
Class C
Class D
Class E
Class F
Class G
Class H
Class I
Class J
Class K
d)
Cell-clusters
GO terms
EMT regulation
Organ formation
Embryonic pattern specification
Collagen trimer
Osteoblast differentiation
Cell-cell junction
Angiogenesis
Mesenchymal cell differentiation
Ribosome biogenesis
T cell receptor signal. pathway
MHC class I complex
Response to IFN-gamma
snoRNA binding
Lamellipodium organization
Dendritic cell differentiation
Mast cell activation
Chemokine production
VEGF pathway
Antigen processing and presentation
Mitochondrial repiratory chain
Cell division
DNA replication
e)
-Log(P-value)
[Reactome pathways]
[Differentially expressed genes]
Cluster 1: Cd34+/Efemp1+
Extracellular matrix organization
Bmp2,Col14a1,Fbn1,Fgf2,Fn1,Itgb7,Mfap5,Ntn4,Pcolce2,Efemp1
Molecules associated with elastic fibres
Bmp2,Fbn1,Fn1,Mfap5,Efemp1
Elastic fibre formation
Bmp2,Fbn1,Fn1,Mfap5,Efemp1
IGF transport
Igfbp4,Igfbp5,Igfbp6
Non-integrin membrane-ECM interactions
Fgf2,Fn1,Ntn4
-Log(P-value)
[Reactome pathways]
[Differentially expressed genes]
Cluster 8: Cd34+/Col15a1+
Extracellular matrix organization
Bgn,Serpinh1,Col15a1,Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Ctsk,Lama2,Lamb1,Lox,Lum,Mmp2,Nid2,Pcolce,Sparc,Jam2
Collagen formation
Serpinh1,Col15a1,Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Lox,Pcolce
Collagen degradation
Col15a1,Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Ctsk,Mmp2
Collagen biosynthesis and modifying enzymes
Serpinh1,Col15a1,Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Pcolce
Collagen chain trimerization
Col15a1,Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2
Assembly of collagen fibrils and other multimeric structures
Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Lox
Degradation of the extracellular matrix
Col15a1,Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Ctsk,Lamb1,Mmp2
Integrin cell surface interactions
Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Lum,Jam2
Non-integrin membrane-ECM interactions
Col3a1,Col4a1,Col4a2,Col4a4,Col5a1,Col5a2,Col1a1,Col1a2,Lama2,Lamb1
ECM proteoglycans
Bgn,Col3a1,Col5a1,Col5a2,Col6a1,Col6a2,Col1a1,Col1a2,Sparc