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], Tetsutaro Hayashi[1], Akira Kurisaki[2*], and Itoshi Nikaido[1*]

[†]Contributed equally

*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

[†]Contributed equally

*Corresponding author

Yohei Sasagawa: youhei.sasagawa@riken.jp

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

Hitomi Takada: htakada@bs.naist.jp

Masashi Ebisawa: masashi.ebisawa@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 several of the reaction steps of Quartz-Seq2 allow us to effectively convert initial reads to UMI counts (at a rate of 30%–50%). To demonstrate the power of Quartz-Seq2, we analyzed transcriptomes from a cell population of *in vitro* embryonic stem cells and an *in vivo* stromal vascular fraction with a limited number of sequence 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 allows us 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 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 measured in 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 reads were sequenced for 3,000–4,000 cells in several previous studies [6,7]. UMI counts were converted from 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 mouse embryonic stem (ES) cells as *in vitro* cells and the cell population from the stromal vascular fraction (SVF) as *in vivo* cells.

146 **Results**

147

148 **Outline of Quartz-Seq2 experiment**

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

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

156 (2) The second step is cell barcoding. Each well contains lysis buffer and
157 reverse-transcription (RT) primer, which includes a cell barcode sequence (14- or 15-mer), a
158 UMI sequence (8-mer), and an oligo-dT sequence (24-mer). Using these RT primers,
159 respective RNA from single cells is converted to cDNA with unique cell barcodes. Note that a
160 long RT primer resulted in a severe problem regarding the synthesis of byproducts at the
161 downstream reaction in our system (Additional file 1: Figure S2 and Supplemental Note).
162 Therefore, we use a relatively short RT primer (73- or 74-mer), which allows us to skip the step
163 of removing byproducts using exonuclease I. We design two types of RT primer set (v3.1: 384
164 barcodes; v3.2: 1536 barcodes). Within each primer set, barcode sequences are designed
165 such that the minimum Sequence–Levenshtein distance between two sequences should be
166 greater than 5, which leads to the correction of two nucleotides of mismatch, insertion and
167 deletion in sequence reads [22]. We also optimize the buffer and temperature in the RT
168 reaction, leading to an improvement of RT efficiency from that of the original Quartz-Seq
169 (Additional file 1: Figure S3 and Supplemental Note).

170 (3) The third step involves the pooling of cell-barcoded cDNA. By cell barcoding,
171 Quartz-Seq2 can pool cDNA of up to 1,536 individual cells into one mixture. We developed a
172 rapid and high-throughput method for collecting small volumes of cDNA in multiwell plates. This
173 method also achieved higher efficiency of collection than dispensing with pipettes and tips
174 (Additional file 1: Figure S4 and Supplemental Note). As the efficiency of cDNA purification

175 after pooling was 93.77%, we estimated that approximately 80% of cell-barcoded cDNA could
176 be used for subsequent whole-transcript amplification in our system (Additional file 1: Figure
177 S3a and Supplemental Note).
178 (4) The fourth step is whole-transcript amplification based on an improved poly-A tagging
179 strategy. Poly-A tagging is one of the methods of converting first-strand cDNA to amplifiable
180 cDNA. First-strand cDNA is extended with a poly-A tail by the terminal transferase.
181 Subsequently, second-strand cDNA is synthesized with tagging primers that contain a poly-dT
182 sequence, followed by PCR amplification. Here, we improve the efficiency of poly-A tagging by
183 3.6-fold. This improvement is a crucial point in the development of Quartz-Seq2. We describe
184 the details of this in the next subsection.
185 (5) The fifth step is library preparation for deep sequencing. Amplified cDNA is
186 fragmented, ligated with the sequence adapter, and amplified by PCR. In sequencing using
187 Illumina sequencers, a sequence for a cell barcode and a UMI is read in Read1, while a
188 sequence for a region of a transcript (mRNA) is read in Read2.
189
190 **Improvement of poly-A tagging efficiency**
191 We previously reported Quartz-Seq based on the poly-A tagging strategy, which has a
192 significant potential for detecting a large number of genes expressed in a cell [15]. However,
193 the efficiency of poly-A tagging itself for single-cell RNA-seq has not been improved. We
194 hypothesized that the improvement of poly-A tagging would lead to high UMI conversion
195 efficiency. Therefore, we attempted to improve the efficiency of this tagging step. Poly-A
196 tagging is composed of two processes: (1) The first-strand cDNA is modified with a poly-A tail
197 by terminal deoxynucleotidyl transferase. (2) Next, the poly-A-tailed cDNA is annealed with a
198 tagging primer, which has a poly-dT sequence. Then, the second-strand cDNA is extended.
199 The resulted second-strand cDNA is amplifiable cDNA, which has a PCR primer sequence at
200 both ends of it.
201 It is known that the DNA yield of amplified cDNA generally reflects the quantitative
202 performance of single-cell RNA-seq methods [3,13,14]. Thus, we determined the effects of
203 various buffers for the poly-A tailing step on the amplified cDNA yield. We performed the poly-A

204 tailing reaction with various buffers using purified first-strand cDNA from 1 ng of total RNA
205 (Figure 2a, Additional file 1: Figure S5). Finally, we obtained amplified cDNA. We found that the
206 use of T55 buffer and RH55 buffer in the poly-A tailing reaction efficiently improved the cDNA
207 yield (Figure 2a). The amount of amplified cDNA increased 2.88-fold and 2.55-fold using T55
208 buffer and RH55 buffer, respectively, compared with the level using Quartz-Seq buffer (Figure
209 2a). In these buffer conditions, we did not observe any obvious byproducts derived from the RT
210 primer (Figure 2a, Additional file 1: Figures S2 and S5, and Supplemental Note).
211 Next, we added an “Increment” temperature condition for the tagging and second-strand
212 synthesis steps (see Material and Methods). In this condition, the reaction temperature of these
213 steps was gradually increased. As a result, the amount of cDNA tended to increase, by
214 approximately 1.2-fold (Figure 2a). Moreover, upon combining T55 buffer and the Increment
215 condition, the amount of cDNA increased approximately 3.6-fold. We also confirmed the
216 reproducibility of this phenomenon of cDNA increment in the additional experiments (Additional
217 file 1: Figure S5). Moreover, we confirmed the amplified cDNA yield of various genes by qPCR
218 analysis as another assay. Specifically, we determined the qPCR scores of eight genes from
219 amplified cDNA and nonamplified cDNA (Figure 2b). Spearman's rank correlation coefficients
220 (SCCs) between amplification and nonamplification were approximately 0.79 in both conditions
221 (T55+Increment, RH55+Increment). The SCC was approximately 0.66 in Quartz-Seq-like
222 conditions. We also observed clear increments of qPCR scores for almost all genes. These
223 results show that the combination of T55 buffer and this temperature condition improved the
224 efficiency of the poly-A tagging step. We also found that other conditions (NBF40+Increment)
225 improved the cDNA yield. However, under these conditions, byproducts were clearly
226 synthesized (Additional file 1: Figures S4c and S5b). Moreover, the cDNA amount with T55
227 buffer was slightly greater than the cDNA amount with RH55 (Figure 2 and Additional file 1:
228 Figure S5a). Therefore, we used the combination of T55 buffer and the “Increment”
229 temperature condition for the poly-A tagging strategy.
230
231 **Evaluation of the quantitative performance of Quartz-Seq2 using single-cell total RNA**

232 Finally, we adopted three molecular-biological improvements [modified RT condition, poly-A
233 tailing buffers (T55), and the “Increment” temperature condition] for Quartz-Seq2 (Additional file
234 1: Figure S1). To determine the technical variability and specificity of Quartz-Seq2, we
235 performed whole-transcript amplification using 10 pg of diluted mouse total RNA as a
236 single-cell-like averaged sample with the v3.1 384 cell-barcode RT primer in a 384-well plate.
237 In this experiment, we used the “RT100” enzyme concentration for reverse transcription. The
238 effect of enzyme concentration is specifically described in the next subsection of this paper. We
239 analyzed 10 pg of total RNA in all wells at approximately 0.19 M fastq reads on average per
240 well. In the case of Quartz-Seq-like conditions, we detected 19,614±3516 UMI counts and
241 5726±494 gene counts (n=384 wells). In the case of Quartz-Seq2, we achieved a UMI count of
242 35,391±8719 and a gene count of 6621±711 (n=384 wells). We confirmed the reproducibility of
243 the quantitative performance by a further batch experiment with other 384-well plates, which
244 had 10 pg of total RNA in all wells (Additional file 1: Figure S6). These results indicated that the
245 combination of the three molecular-biological improvements for Quartz-Seq2 clearly improved
246 the conversion ratio from target RNA to sequence library DNA. We also confirmed the
247 correlation between single-cell RNA-seq data and bulk RNA-seq data. SCC between
248 Quartz-Seq2 and bulk RNA-seq was 0.85 (Figure 2d). This result indicates that Quartz-Seq2
249 data are comparable to conventional bulk RNA-seq data. We also determined the variability of
250 expression between technical replicates to investigate the UMI filtering effect. The mean-CV
251 (coefficient of variation) plot of Quartz-Seq2 clearly showed that UMI filtering reduced
252 expression variability derived from PCR bias (Figure 2e). Moreover, technical gene expression
253 variability of Quartz-Seq2 became fairly close to the theoretical variability of a Poisson
254 distribution (Figure 2e).
255 We also focused on the relationship between sequencing platform and Quartz-Seq2.
256 We did not observe clear differences between the NextSeq500 and HiSeq2500 sequencing
257 platforms for determination of the UMI count and the gene count for Quartz-Seq2 (Figure S7a).
258 On the other hand, the unique mapping ratio and UMI count clearly depended on the Read2
259 length for transcript mapping (Figure S7c, d). In this regard, the number of detected genes
260 slightly depended on Read2 length (Figure 7d). Long Read2 length led to better UMI counts,

261 but high sequencing cost (Figure 7b). Therefore, in this study, we mainly used the NextSeq500
262 sequence platform and a Read2 length of 62 due to the cost performance of this combination.
263 Moreover, we mainly used the adaptor ligation method for sequence library preparation of
264 Quartz-Seq2. We did not observe a clear difference between the two sequence library
265 preparation methods (ligation and Nextera/Tn5 transposase) regarding determination of the
266 UMI count and the gene count. However, we could easily identify specific familiar genes (such
267 as *Pou5f1* and *Gapdh*) using the ligation method (Figure S7e). Therefore, we mainly used the
268 ligation method for Quartz-Seq2.
269
270 **Reduction of enzyme concentration in reverse transcription decreased cost and**
271 **technical noise of Quartz-Seq2**
272 The cost of experimental preparation per cell was approximately ¥2600 ($23) for our previously
273 reported Quartz-Seq, which does not use cell barcoding (Figure S8c). CEL-seq2 is one of the
274 PCR plate-based single-cell RNA-seq methods using a cell sorter. This method has the
275 advantage of reducing costs via cell barcoding. Indeed, the experimental cost of CEL-seq2 is
276 lower (approximately ¥372 or $3.35) than the cost of Quartz-Seq. On the other hand, the
277 experimental cost per cell is extremely low (approximately ¥10 or $0.08) for droplet-based
278 high-throughput single-cell RNA-seq methods, such as Drop-seq ([6] and Figure S8c). Ideally,
279 the experimental cost for high-throughput single-cell RNA-seq methods using a cell sorter
280 should be reduced to approach the cost of droplet-based methods. As mentioned earlier, we
281 used the “RT100” enzyme condition in reverse transcription for Quartz-Seq2. This enzyme
282 concentration in RT reaction is broadly used for various molecular biological application
283 including single-cell RNA-seq methods. By using the cell barcoding strategy, the cost of
284 experimental preparation for Quartz-Seq2 under the “RT100” condition was reduced (¥121 or
285 $1.09 per cell). We found that 65% of the cost in experimental steps is derived from reverse
286 transcription in Quartz-Seq2 under the “RT100” condition (Figure 3a). Therefore, we
287 investigated the effect of a low enzyme concentration in reverse transcription in Quartz-Seq2.
288 In the assessment assay, we noticed that a low enzyme concentration did not markedly affect
289 the efficiency of reverse transcription and the amount of amplified cDNA yield (Figure S8a right

290 panel and S8b). Therefore, we prepared three technical replicates of a 384-well PCR plate with
291 10 pg of total RNA with the “RT100” condition or the “RT25” low-enzyme condition. In the
292 “RT25” condition, the amount of cDNA yield showed a tendency for a slight increase of
293 approximately 1.17-fold, which although not being a major improvement did at least not involve
294 a decrease. The cost of experimental preparation per cell became approximately ¥45–¥62
295 ($0.40–$0.55) in the “RT25” condition.
296 We validated the quantitative performance with 10 pg of total RNA in the “RT100” and
297 “RT25” conditions at various input data sizes. Unexpectedly, we found that the “RT25”
298 condition improved the quantitative performance (Figure 3c and Additional file 1: Figure S9). In
299 a typical pattern, approximately 4,000 single cells were analyzed by the NextSeq500
300 sequencer in the high-throughput single-cell RNA-seq method [6,7,21]. In this case, the input
301 data size for a single cell involves shallow initial reads (0.1 M fastq reads) (Figure 3d). We
302 compared RT25 with RT100 at approximately 0.096 M fastq reads on average per well. In the
303 case of Quartz-Seq2 in the “RT25” condition, we achieved a UMI count of 30,117±789 and a
304 gene count of 6,320±35 (n=3 plates). In the case of Quartz-Seq2 in the “RT100” condition, we
305 achieved a UMI count of 25,671±1020 and a gene count of 5,889±35 (n=3 plates). The capture
306 ratios of ERCC spike were 11.49±0.46% and 10.28±0.31% in the “RT25” and “RT100”
307 conditions, respectively (Figure 3d and e). In addition, the capture ratio of ERCC spike was
308 16.38±0.66% in the “RT25” condition at approximately 0.27 M fastq reads on average per well
309 (Figure 3c, e). We also investigated the technical variability of UMI count and gene count. The
310 well-to-well technical variability for UMI count and gene count clearly decreased in the “RT25”
311 condition (Figure 3). These results showed that the “RT25” low-enzyme condition clearly
312 reduced the experimental cost and improved the quantitative performance, including that for
313 the UMI count. We thus applied the “RT25” condition in subsequent experiments using real
314 single cells.
315
316 **Quartz-Seq2 shows high efficiency of UMI conversion**
317 The increase of UMI counts generated from limited initial sequence reads leads to the
318 detection of low-copy genes and the identification of more marker genes using statistical tests.

319 We calculated the UMI conversion efficiency, which indicates how effectively initial reads can
320 be converted to UMI counts (Figure 4a and Additional file 1: Figure S10). The UMI conversion
321 efficiency of Quartz-Seq2 with 10 pg of purified total RNA (from mouse ES cells) ranged from
322 31% to 39% at approximately 0.096 M initial reads depending on the Read2 length. These
323 results show that major improvements (such as poly-A tagging) in Quartz-Seq2 increase the
324 efficiency of UMI conversion.

325 We performed Quart-Seq2 and Drop-seq on a mixture of mouse ES cells and the
326 ES-derived primitive endoderm (PrE) cells, and calculated the UMI conversion efficiency
327 (Figure 4b and Additional file 1: Figure S11a). Drop-seq is one of the high-throughput
328 single-cell RNA-seq methods, which can capture thousands of cells and barcoded beads in half
329 an hour. However, there are many barcoded beads that are not co-captured with a cell in a
330 droplet, leading that 10%–40% of initial sequence reads could be lost [6]. We performed
331 Drop-seq experiments, which were validated by species-mixing analysis (Figure S11b, c). The
332 effectiveness of Quartz-Seq2 ranged from 25% to 35% depending on the initial read depth
333 (Figure 4), which was higher than that of Drop-seq. In previous studies, average UMI counts
334 and gene counts were compared among several single-cell RNA-seq methods [1,6,21]. Using
335 such data, we calculated the UMI conversion efficiency of each method and compared them
336 with that of Quartz-Seq2 (Additional file 1: Figure S12). UMI conversion efficiency depended on
337 the length of Read2 that corresponds to the transcript sequence because the rate of unique
338 mapping depends on it (Additional file 1: Figure S7). As Figure S12 shows, the UMI conversion
339 efficiency of Quartz-Seq2 was higher than those of the other single-cell RNA-seq methods.

340 As Figure 4b shows, Quartz-Seq2 detected more genes with high UMI conversion
341 efficiency. To examine the power for identifying marker genes, we performed principal
342 component analysis (PCA) and clustering. The number of genes differentially expressed
343 between two distinct cell types was calculated (ES and PrE cells) (Additional file 1: Figure
344 S11d). The results showed that more genes that were differentially expressed between the two
345 cell types were identified with Quartz-Seq2 (Figure 4b). In addition, more biological pathways
346 particularly associated with the differentially expressed genes were detected in Quartz-Seq2
347 (Figure 4c). Furthermore, we calculated genes with highly variable expression in Quartz-Seq2

348 and Drop-seq, which potentially include not only genes that are differentially expressed
349 between cell types but also genes of which the expression changes depending on the cell state
350 in a cell type. The results also showed that more Gene Ontology (GO) terms enriched among
351 the genes with highly variable expression were identified with Quartz-Seq2. Terms on cell cycle
352 state were only enriched in the genes calculated with Quartz-Seq2 (Figure 4d). These results
353 suggest that high UMI conversion efficiency with limited initial reads leads to more biological
354 information being revealed, such as functional terms and biological pathways.
355
356 **Quartz-Seq2 analysis on 4,484 mouse embryonic stem cells in maintenance and**
357 **differentiated conditions**
358 We did not observe any clear differences in UMI counts and gene counts between the v3.1 and
359 v3.2 RT primer sets (Additional file 2: Table S1). To demonstrate the capability of Quartz-Seq2
360 for quantifying the transcriptome of a large number of cells and identifying rare cell populations,
361 we analyzed approximately 4,484 cells from a mixture of mouse ES cells and
362 differentiation-induced cells. Cells were prestained with Hoechst 33342 and/or Calcein-AM as
363 an indicator of DNA content and culture condition (ES or PrE cells), respectively (Figure 5a).
364 Calcein-AM-positive and -negative cells were sorted to 12 of 384-well plates in a checkered
365 pattern (Figure 5a). In this condition, the cell doublets caused by mis-sorting can be detected at
366 the highest probability. In this analysis, as an average for cells, 0.1 M initial fastq reads were
367 effectively converted to 35,915±1,188 UMI counts in a final gene expression matrix. The UMI
368 conversion efficiency was 35.91±1.18%.
369 Dimensionality reduction of the expression UMI matrix resulting from Quartz-Seq2
370 showed clear separation of six clusters, including the main populations of ES cells and PrE
371 cells and four small populations. We did not observe a clear batch effect between 12 of the
372 384-well plates (Additional file 1: Figure S13). One cluster contained a relatively high ratio of
373 mitochondrial RNA (Additional file 1: Figure S14), which was judged to reflect low-quality cells
374 [23]. Another cluster showed high values of detected UMI counts and gene counts, and
375 expressed both ES and PrE marker genes, which were judged to represent doublets of an ES
376 cell and a PrE cell. To characterize the identified populations, we determined the genes that

377 were specific or enriched for each cluster using binomial tests (Figure 5c and Additional file 1:
378 Figure S15). Although cells in cluster 3 shared gene expression with the main population of ES
379 cells (cluster 1), some genes including *Stmn2* and *Rhox5* were additionally expressed,
380 suggesting that cluster 1 and cluster 3 shared the same cell type but had different cell states.
381 Cluster 5 was a small population (0.17%, Figure 5b), but was characterized by many specific
382 marker genes (Figure 5b), including *Zscan4c* and *Zscan4d*, which are known to be expressed
383 in a subpopulation of mouse ES cells [24]. However, among our culture conditions, these
384 genes were more expressed in the Dex-treatment condition than in the ES-maintenance
385 condition. Flow cytometry information on cells in cluster 5 also showed no fluorescence of
386 Calcein-AM, indicating that these cells were from the suspension of Dex-treated cells (Figure
387 5d). These observations were consistent with a previous study demonstrating that *Zscan4* was
388 more expressed in a differentiation condition induced by withdrawal of leukemia inhibitory
389 factor [7].
390 In sampling of cells using a cell sorter, flow cytometry information including the intensity
391 of Hoechst 33342 staining for each cell was collected. Using this information, we reconstructed
392 the distribution of Hoechst 33342 staining intensity for each cluster (Figure 5e). The main
393 population of ES cells (cluster 1) showed an embryonic pattern of distribution of DNA content,
394 which is characterized by a high ratio of cells in G2 and M phases to cells in the G1 phase. On
395 the other hand, the main population of PrE cells (cluster 4) showed a somatic pattern of
396 distribution of DNA content. This was consistent with PrE cells being differentiated from ES
397 cells. The cells in cluster 3, the gene expression of which was similar to that of the ES main
398 population (cluster 1), showed an embryonic pattern rather than a somatic one. As the number
399 of cells in cluster 5 that expressed *Zscan4c/d* was small, it was difficult to classify the observed
400 pattern as the embryonic or somatic type. These findings indicated the usefulness of single-cell
401 RNA-seq using flow cytometry for the reconstruction of population information after
402 transcriptome-based clustering. For good interpretation of the distribution, a large number of
403 cells for each cluster are required.
404

405 **Quartz-Seq2 revealed two types of cell state: cell cycle-dependent and -independent**
406 **heterogeneity**
407 When cells were plotted on t-SNE (t-distributed stochastic neighbor embedding) space using
408 transcriptome analysis and the intensity of Hoechst 33342 staining was colored, the gradient
409 pattern was easily observed for clusters 1, 3, and 4 (Figure 6a), suggesting that Quartz-Seq2
410 was highly sensitive for detection of the cell cycle state. To examine this, we plotted the gene
411 expression of several cell cycle markers against the intensity of Hoechst 33342 staining, which
412 indicated the strong dependence of such expression on DNA content (Figure 6b). Using the
413 intensity of Hoechst 33342 staining, cell cycle dependence was calculated for all genes
414 detected in Quartz-Seq2 experiments, resulting in the identification of numerous genes for
415 which the expression level changed depending on the DNA content (Figure 6c, d and
416 Additional file 1: Figure S16). We called these genes cell cycle-associated genes. Again, it was
417 confirmed that the ratio of cells in G1 and S phases to those in G2 and M phases was different
418 between ES and PrE cells (Figure 6c). As we showed that the efficiency of UMI conversion was
419 high in the Quartz-Seq2 method, we examined whether 10,000 initial reads are sufficient for
420 cell cycle analysis (Additional file 1: Figure S17). In this analysis, final UMI count was 4,774±62
421 and the UMI conversion efficiency was 47.74±0.66%. Plotting cells on t-SNE space with
422 coloring the intensity of Hoechst 33342 staining revealed a gradient pattern, which was
423 positively correlated with G2/M marker genes including *Ccnb1* and *Top2a* and was negatively
424 correlated with G1-phase marker genes such as *Ccne1* (Additional file 1: Figure S17). These
425 observations demonstrated Quartz-Seq2 with few initial sequence reads could analyze cell
426 cycle due to the high UMI conversion efficiency.
427 Within a single cell type, there are two types of cell state. One is the different stages of
428 the cell cycle as mentioned above, and the other refers to cellular heterogeneity that is only
429 remotely related to the cell cycle phases. This means that genes with variable expression in a
430 cell type include cell cycle-associated genes and gene with variable expression that is
431 independent of the cell cycle phases (cell cycle-phase independent variable genes). We
432 determined the number of this latter group of genes by subtracting the cell cycle-associated
433 genes from the genes with variable expression (Figure 6e). For the main population of ES cells

434 (cluster 1), we identified tens of cell cycle-phase independent variable genes. These genes
435 included *Spp1*, *Rhox5*, and *S100a6*, which were previously identified as genes with highly
436 variable expression by scRNA-seq (Figure 6) [7,13,15]. We also identified *Stmn2*, *Dnmt3l*,
437 *Tmsb4x*, and several genes as novel cell cycle phase-independent variable genes. Single-cell
438 RT-qPCR also showed high variability for *Stmn2* and *Rhox5* compared with that for *Nanog*
439 (Figure 6f). In previous studies, *Sgk1* and *Actb* were identified as genes with variable
440 expression by scRNA-seq that analyzed cells only in G1 phase or cells without measuring the
441 intensity of Hoechst 33342 [7,15]. In this study, these were classified as cell cycle-associated
442 genes (Additional file 1: Figure S16).
443 These results show that Quartz-Seq2 globally classified genes with variable expression
444 within a cell type into cell cycle-associated genes and cell cycle phase-independent genes with
445 variable expression by using Hoechst 33342 staining intensity data obtained in flow cytometry.
446
447 **The global picture of cell-type composition in stromal vascular fraction as revealed by**
448 **Quartz-Seq2**
449 Next, Quartz-Seq2 was applied to cells from the mouse stromal vascular fraction (SVF), which
450 was taken from adipose tissue (Figure 7a). SVF is thought to be one of the most important
451 sources of mesenchymal stem cells (MSC) due to its potential for cell therapies and the
452 stimulation of endogenous repair [25-28]. However, the global picture of cell-type composition
453 in SVF has not been clarified. To do so, we first observed the distribution of cell size in the SVF
454 population and found it to be broad (5–13 μm, average 6.43 ± 1.35 μm; n=200) (Figure 7a).
455 This was confirmed by flow cytometry data that were obtained using a cell sorter. It is known
456 that the typical amount of total RNA in a mammalian cell is approximately 10 pg [15,17]. Thus,
457 the average amount of total RNA in the SVF population of approximately 1.7 pg is relatively
458 small.
459 We collected cells in multiple 384-well PCR plates and analyzed their transcriptome
460 using Quartz-Seq2. In this analysis, 38,450±3513 initial fastq reads were converted to UMI
461 counts in the final digital expression matrix. The UMI conversion efficiency was 26.85±2.70%
462 (n=3). We observed broader distributions of UMI counts and gene counts in SVF than in ES

463 and 10 pg of total RNA. The gene counts correlated well with side scatter (SSC) values, which
464 were associated with cell size (Additional file 1: Figure S19).
465 Dimensionality reduction and clustering of single-cell transcriptome data showed the
466 clear separation of 11 clusters (Figure 7b). To annotate the cell type on each cluster, we
467 calculated marker genes and functional terms that were either specific or enriched for each
468 cluster (Figure 7c, d and Additional file 1: Figures S18 and S20). As Figure 7c shows, clusters 1,
469 2, 5, and 8-11 expressed cluster-specific genes, and clusters 3, 4, 6, 7, and 9 shared
470 expressed genes (marker class C) as well as differentially expressed genes of classes D-G.
471 Taking these markers and functional terms as well as previous knowledge together, each
472 cluster was identified as follows: two types of *Cd34*-positive MSC (cluster 1: 7.8%, cluster 8:
473 6.4%), two types of *Cd4*-positive T cells (cluster 3: 45.0%, cluster 4: 1.0%), *Cxcr6*-positive T
474 cells (cluster 9: 1.71%), B cells (cluster 5: 15.9%), killer T cells (cluster 6: 11.0%), natural killer
475 cells (cluster 7: 3.0%), macrophages (cluster 2: 5.5%), dendritic cells (cluster 10: 0.86%), and
476 plasmacytoid dendritic cells (cluster 11: 0.57%). We noted that the cell sizes of the T cells and
477 B cells from clusters 3, 4 and 5, which were measured using a cell sorter, were smaller than
478 those of MSCs and macrophages from clusters 1, 2, and 8 (Additional file 1: Figure S19). We
479 confirmed that the cell sizes of CD4-positive T cells and CD79-positive B cells were smaller in
480 SVF via immunofluorescence observation (Additional file 1: Figure S19). These results show
481 that Quartz-Seq2 with flow-cytometry data is useful for defining respective cell clusters with cell
482 size information.
483 Subsequently, we investigated the gene expression of receptors and ligands in stromal
484 vascular fractions. The variety of receptors expressed in each cluster was similar (Additional
485 file 1: Figure S21). However, we found a wide variety of ligands expressed in clusters 1 and 8
486 (Additional file 1: Figure S21). It is known that MSCs secrete many paracrine mediators, which
487 have a therapeutic effect [29,30]. These results suggest that clusters 1 and 8 are MSCs, which
488 secretes many paracrine mediators.
489 We focused on two potential MSC clusters (1 and 8) because the heterogeneity of the
490 MSC population has been discussed in many papers [31,32]. By using Quartz-Seq2, clusters 1
491 and 8 had the *Cd31-*/*Cd34+*/*Cd45-* phenotype (Figure 7, Additional file 1: Figure S22). In a

492 previous study, qPCR-based single-cell transcript analysis of the *Cd31-*/*Cd34+*/*Cd45-*
493 population of SVF was reported. By analyzing 140 genes, the authors identified the
494 *Cd55+*/*Dpp4+*/*Cd31-*/*Cd34+*/*Cd45-* cell population and showed that administration of this
495 population was effective for normalizing diabetic wound healing in mice [31]. They identified 13
496 marker genes for the *Cd55+*/*Dpp4+*/*Cd31-*/*Cd34+*/*Cd45* population in mouse SVF. We
497 compared the expression of those 13 genes between clusters 1 and 8 (Additional file 1: Figure
498 S23). The results showed that almost all the genes were more highly expressed in cluster 1
499 than in cluster 8, suggesting that the cluster 1 population was similar to the
500 *Cd55+*/*Dpp4+*/*Cd31-*/*Cd34+*/*Cd45-* population. To further analyze the relationship between our
501 identified clusters and previously reported heterogeneity of the MSC population, we examined
502 the expression of typical MSC markers (*CD90*, *CD105*, *PDGFRa*, and *Sca-1*) [31,33] as well as
503 subpopulation specific markers (*Pou5f1*, *Nanog*, *Sox2*, *Tnnt2*, and *Myog*) [34,35].
504 *PDGFRa* (*Pdgfra*) and *Sca1* were expressed in both clusters 1 and 8, whereas *Cd90* was
505 expressed at a relatively higher level in cluster 8 than in cluster 1, and *Cd105* was not strongly
506 expressed in either cluster. Neither pluripotent markers (*Pou5f1*, *Nanog*, and *Sox2*) nor skeletal
507 and cardiomyogenic markers (*Tnnt2* and *Myog*) were detected in either cluster (Additional file
508 1: Figure S22). Collectively, our transcriptome analysis showed that the MSC population is
509 divided into two clusters, suggesting that there is less heterogeneity of MSC in the SVF than
510 expected.
511 Next, we investigated the difference between the two MSC clusters in SVF. We detected
512 182 differentially expressed genes between clusters 1 and 8 (Additional file 1: Figure S24,
513 Additional file 5: Table S4). Cluster 8 was characterized by the enriched expression of genes
514 that encode extracellular matrix proteins, including collagens, *Hspg2*, and *Bgn* (Additional file
515 1: Figure S22). By contrast, cluster 1 was characterized by the expression of *IGFBPs* (*Igfbp4*,
516 *Igfbp5*, *Igfbp6*), prostaglandin endoperoxidase synthases (*Ptgs2*), and secreted factors *Fgf2*
517 and Bmp2. Of note, the expression of stem cell characteristic genes *Aldh1a3*, *Fgf2*, *Bmp2*, and
518 *Tgfbr2* were identified as important modules in cluster 1 (Additional file 1: Figure S22),
519 suggesting that cluster 1 had more stem-like phenotypes that were secreting with medicinal
520 factors.

521 To leverage the power of the entire transcriptome analysis, we performed an enrichment
522 analysis of biological pathways using the 182 differentially expressed genes between the two
523 clusters (Figure 7e, Additional file 6: Table S5). Cluster 1 was characterized by elastic fiber
524 pathways and IGF transport pathways. Elastic fibers contribute various connective tissues and
525 are related to vascular function [36]. In cluster 8, a tightly linked network of collagen
526 metabolism pathways was highly enriched. These results show that two MSC clusters of SVF
527 can have different functions in extracellular matrix formation.
528 Taken together, our transcriptome analysis by Quart-seq2 successfully identified cell
529 populations that consisted of various types of both large and small cells, which were isolated
530 from *in vivo* tissue, and demonstrated that fresh SVF contains two closely related types of
531 MSCs that have distinct characteristics.
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549

550 **Discussion**

551 The rate of conversion from initial reads to UMI counts was low, especially in high-throughput
552 single-cell RNA-seq. An increase in UMI counts assigned to each cell theoretically leads to the
553 detection of low-copy genes and the identification of cell-type-specific genes using statistical
554 tests. Therefore, in this study, we developed a novel high-throughput single-cell RNA-seq
555 method, Quartz-Seq2. We calculated the UMI conversion efficiency, which indicated how
556 effectively initial reads can be converted to UMI counts. The UMI conversion efficiency of
557 Quartz-Seq2 was higher than those of other single-cell RNA-seq methods (Additional file 1:
558 Figure S12). In other words, Quartz-Seq2 could detect more genes from limited fastq reads and
559 for a lower cost. By cell barcoding, Quartz-Seq2 can pool cDNA of up to 1,536 individual cells
560 into one mixture. Streamlined and simplified experimental processes of Quartz-Seq2 allowed us
561 to handle thousands of single cells within a few days. Quartz-Seq2 also showed high accuracy
562 and high reproducibility, leading to precise measurement of the transcriptome from single cells
563 (Figure 2d and e). We demonstrated Quartz-Seq2 analyses on mouse ES cells as an i*n vitro*
564 sample, and mouse SVF including MSC as an *in vivo* sample. Moreover, we decreased the
565 cDNA preparation cost per cell of Quartz-Seq2 by 97.6%-98.2% compared with that of
566 Quartz-Seq (Figure S8c). The total cost (cDNA preparation cost and sequence cost per cell) of
567 Quartz-Seq2 as a plate-based method approached that of a droplet-based method (Additional
568 file 1: Figure S8d). Furthermore, the use of evaporation-preventing oil might reduce the cDNA
569 preparation cost of Quartz-Seq2 by 75% [2].

570 We proposed that UMI conversion efficiency could be a useful variable for evaluating
571 performance for the further development of high-throughput single-cell RNA-seq methods with
572 shallow reads. The increase of initial sequence reads by additional sequencing does not cancel
573 out the low efficiency of UMI conversion at equal rate. For example, approximately 24,000 fastq
574 reads were converted to approximately 11,000 UMI counts in Quartz-Seq2, while Drop-seq
575 required approximately 169,000 fastq reads to generate the same UMI counts. At this condition,
576 the difference of UMI conversion efficiency was 2.5-fold but the difference of required fastq reads
577 to generate the same UMI counts was 7-fold. This is because UMI counts do not linearly
578 increase with an increase of initial sequence reads due to UMI filtering. Unfortunately, the rate of

579 increase of sequence throughput was lower than that of the processing ability for single cells. It
580 will thus continue to be important to utilize limited initial fastq reads for high-throughput
581 single-cell RNA-seq methods.

582 Quartz-Seq2 is based on the poly-A tagging, which is one of the strategies of
583 converting first-strand cDNA to amplified cDNA [19,37]. However, the poly-A tagging efficiency
584 itself has not been improved for single-cell RNA-seq. In this study, three molecular biological
585 improvements on Quartz-Seq2 contributed to the increase of amplified cDNA, which led to high
586 UMI conversion efficiency (Figure 2 and Additional file1: Figure S1). In addition, improvement of
587 poly-A tagging steps was most efficient to increase amplified cDNA for Quartz-Seq2 (up 360%,
588 Additional file 1: Figure S1). Several previous studies improved the efficiency of conversion from
589 an mRNA molecule to amplifiable cDNA, and showed that the increase of amplified cDNA is a
590 good guidepost to improve UMI counts or quantitative performance for single-cell RNA-seq
591 [3,14,15]. These results suggest that improved poly-A tagging was the most important point for
592 high UMI conversion efficiency of Quartz-Seq2.

593 Several improvements proposed in this study could contribute to the further
594 development of other single-cell RNA-seq methods, as follows. (1) We showed that a decrease
595 of enzyme concentration in RT solution led to decreases in technical error and cDNA preparation
596 cost. (2) We applied the Sequence–Levenshtein distance for the design of cell barcodes
597 containing RT primer sets. The use of primer sets designed in this study allows the user to
598 correct mutations of at most two nucleotides of cell barcode, including substitution, insertion, or
599 deletion; the correction capability of these primer sets is higher than that of barcode sets used in
600 previous studies [1,3]. Such correction of the cell barcode increased the average UMI count by
601 3%–5%. (3) We hope that the spin-down collection system developed in this study can be
602 applied to other plate-based single-cell RNA-seq methods [2,3,11,12]. (4) In Quartz-Seq2, we
603 improved the efficiency of poly-A tagging itself (Figure 2). Several single-cell RNA-seq methods
604 including Quartz-Seq were based on poly-A tagging [4,5,10,15,19,38]. Such improvement may
605 directly contribute to these methods having increased quantitative performance.

606 We achieved a marked improvement in UMI conversion efficiency from initial reads to
607 UMI counts, allowing us to analyze 3- to 10-fold more single cells in limited sequence

608 experiments. However, it is difficult to sort more than 20,000 single cells using one cell-sorter
609 machine in a day. While there is greater scalability for cell sampling in droplet-based single-cell
610 RNA-seq methods than in cell sorter-based methods, these latter methods can utilize additional
611 information obtained in flow cytometry [39,40], which cannot be determined from transcriptome
612 data alone. These two types of method provide complementary approaches for investigating
613 complex biological phenomena.
614 In this study, we showed that Quartz-Seq2 has advantages in gene detection and the
615 identification of biological pathways via high UMI conversion efficiency (Figures 4, 5, and 7).
616 Moreover, we analyzed thousands of single cells from all cell cycle phases within a cell type
617 (Figure 6). These specifications with Quarz-Seq2 allow us to perform the global classification of
618 genes with variable expression in a cell type into those whose expression varies in association
619 with the cell cycle or independently of it.
620 A previous study utilizing the FUCCI fluorescent reporter system reported that
621 competence to respond to specific differentiation signaling was limited to only an early or a late
622 window of the G1 phase in human ES cells [41]. In the future, combining Quartz-Seq2 and
623 FUCCI/FUCCI2 or other fluorescent reporter systems will lead to an understanding of the
624 global picture of differentiation dynamics on competence, response, transition, and
625 commitment because Quartz-Seq2 can analyze changes of cell state for thousands of cells.
626 In summary, Quartz-Seq2 can be used to detect continuous cell states because of the
627 large number of cells with which it deals and the high efficiency of use of initial sequence reads.
628 Quartz-Seq2 can facilitate investigation of the cell state within a cell type, such as gradient or
629 stochastic changes of the cell population in organism development and disease progression.
630
631
632
633
634
635
636

637 **Conclusions**

638 In this study, we developed a high-throughput single-cell RNA-seq method, Quartz-Seq2, which
639 can analyze cells numbering up to 1,536 that are pooled together in a single sample.
640 Quartz-Seq2 allows us to effectively utilize initial sequence reads from a next-generation
641 sequencer. The UMI conversion efficiency in Quartz-Seq2 ranged from 30% to 50%, which is
642 much higher than for other single-cell RNA-seq methods (5%–20%). This was caused by the
643 improvements in several molecular biological steps including poly-A tagging. The technical gene
644 expression variability of Quartz-Seq2 was sufficiently low, and that was close to the theoretical
645 variability of a Poisson distribution. As we showed in the analysis of SVF and the ES/PrE mixture,
646 cell types in the population were identified with marker genes and functional terms that
647 characterized each cell type. We identified two types of *Cd34*-positive MSC in SVF, namely,
648 those that express numerous transcription factor-coding and secreted protein-coding genes
649 specific to each. Quartz-Seq2 can also be used to provide continuous data on cell states
650 because of the large number of cells with which it deals and the high efficiency in the use of initial
651 sequence reads. Quartz-Seq2 should facilitate investigations of the cell state within a cell type,
652 such as gradient or stochastic changes of cell populations in organism development and disease
653 progression.
654
655
656
657
658
659
660
661
662
663
664
665

666 **Methods**

667

668 **Cell culture**

669 Mouse embryonic stem cells were cultured as described previously [15]. Briefly, 1.0 x 10^5 cells
670 were seeded on a 60-mm dish coated with gelatin (0.1%). Cells were maintained in
671 GMEM-based medium containing 10% FBS and 1000 units/ml leukemia inhibitory factor
672 (Millipore ESGRO). The cell line used in this study was G6GRGFP, which was established in a
673 previous study [42]. To differentiate ES cells into primitive endoderm-like cells, 1.0 x 10^5 cells
674 were seeded on a 60-mm dish coated with gelatin and cultured in GMEM-based medium
675 containing 10% FBS, 1000 units/ml LIF, and 100 mM dexamethasone for 72 h.

676

677 **Cell staining**

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

686

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

688 These experiments were carried out in accordance with the protocol approved by the National
689 Institute of Advanced Industrial Science and Technology (AIST) Animal Care and Use
690 Committee and the RIKEN Animal Experiment Committee. Subcutaneous fat tissues from three-
691 to four-month-old ICR male mice (n=3 per sample, two biological replicate samples) were
692 minced into small pieces and incubated with 0.4 U/mL collagenase NB4G (Serva) at 37°C for 35
693 min in a shaking water bath. The digested solution was sequentially filtered through 100-μm and
694 40-μm cell strainers (Corning), followed by centrifugation at 250 × *g* for 5 min to remove mature

695 adipocytes. The pellet was treated with erythrocyte lysis buffer (BD Biosciences) and centrifuged
696 at 180 × *g* for 5 min. The nucleated cells were suspended in HBSS with 0.1% BSA, filtered
697 through a 20-μm cell strainer (pluriSelect), and then kept on ice (Cell solution A). The cell
698 aggregates that did not pass through the 20-μm strainer were further treated with Accutase
699 (Thermo Fisher Scientific) at 37°C for 15 min to dissociate them into single cells, centrifuged at
700 180 × *g* for 5 min, and suspended in HBSS with 0.1% BSA (Cell solution B). Cell solutions A and
701 B were mixed and again filtered through the 20-μm cell strainer, followed by centrifugation at 180
702 × *g* for 5 min. The pellet was resuspended with HBSS with 0.1% BSA, stained with PI, and used
703 for single-cell analysis.
704
705 **RNA preparation**
706 Total RNA was purified from cultured cells using Direct-zol RNA MiniPrep kit (Zymo Research)
707 with TRIzol RNA Isolation Reagents (Thermo). We measured the concentration of purified total
708 RNA using a NanoDrop 1000 Spectrophotometer (Thermo). We confirmed that the RNA integrity
709 number of total RNA was over 9.5 using an Agilent RNA 6000 Nano Kit (Agilent). For
710 high-throughput single-cell RNA-seq, we prepared total RNA from a single cell with ERCC
711 spike-in mix I. First, we diluted the ERCC spike mix I by tenfold. We then added 6 μL of 1:10
712 diluted ERCC spike-in mix I per 10 μg of total RNA. We used diluted 10 pg of total RNA with
713 ERCC spike mix I for the technical validation of Quartz-Seq2.
714
715 **Bulk RNA-seq methods for populations of cells**
716 We prepared sequence library DNA with 1 μg of total RNA using NEBNext Poly(A) mRNA
717 Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit. The total RNA
718 did not contain ERCC spin-in mix I. In addition, we used SuperScript III instead of ProtoScript in
719 the reverse-transcription step and KAPA HiFi DNA polymerase instead of NEBNext High-Fidelity
720 PCR DNA polymerase in the PCR step. The resulting sequence library DNA was analyzed by
721 HiSeq2500.
722
723 **Design of cell barcodes**

724 The selection of 384 or 1,536 sequences for v3.1 and v3.2 barcode primer sets was performed
725 as described below. First, 1,582 or 4,714 candidate sequences were created using the
726 DNABarcodes package of R Bioconductor for the v3.1 set with 14-mer and the v3.2 set with
727 15-mer, respectively (version: 1.0.0). To reduce the loss of reads converted into UMI counts, we
728 applied the Sequence–Levenshtein distance as an edit distance in order to maximize the ability
729 to correct errors that occur during the synthesis of oligonucleotides or sequencing [22]. The
730 minimum distance between any two sequences was controlled to 5, leading to the correction of a
731 maximum of two errors of substitution, insertion, or deletion. As the base composition of the
732 created sequences was not uniform, we selected 384 or 1,536 sequences from among 1,582 or
733 4,714 created sequences so that the variance of base composition decreased. These sequences
734 are listed in Additional file 3: Table S2.
735
736 **Single-cell collection using flow cytometry**
737 Stained cells were analyzed using flow cytometry SH800 (Sony) or MoFlo Astrios EQ (Beckman
738 Coulter). For SH800, we used 130-μm microfluidic sorting chips. For MoFlo Astrios EQ, we used
739 100-μm nozzle sizes. Each cell sorter was equipped with a custom-made splash-guard to
740 prevent contamination from unexpected droplets between the target well and neighboring wells.
741 We used two sets of RT primer for Quartz-Seq2. The set of v3.1 RT primers has 384 kinds of
742 unique cell barcodes, with a length of 14 nucleotides (OPC purification, FASMAC). The set of
743 v3.2 RT primers has 1,536 kinds of unique cell barcodes, with a length of 15 nucleotides (OPC
744 purification, Sigma). The set of v3.1 RT primers corresponds to one set of the 384-well PCR
745 plate with lysis buffer, the wells of which have unique barcodes. The set of v3.2 RT primers
746 corresponds to four sets of the 384-well PCR plate with lysis buffer. The RT primer position in the
747 384-well plate and the sequence were as described in Additional file 3: Table S2. Single cells
748 were isolated in the 384-well PCR plate with 1 μL of lysis buffer (0.1111 μM respective RT
749 primers, 0.12 mM dNTP mix, 0.3% NP-40, 1 unit/μL RNasin plus) containing ERCC spike mix I.
750 During single-cell sorting, the 384-well PCR plate was kept on a 384 aluminum stand at 4ºC. For
751 Moflo Astrios EQ, we used G5498B-060 Insert 384 Eppendorf twin.tec PCR as a 384 aluminum
752 stand (Agilent). For SH800, we used an SH800 384 aluminum stand (Sony). Immediately after

753 the cell collection, the plate was temporarily sealed with LightCycler 480 Sealing Foil (Roche)
754 and the sealed 384-well PCR plate was centrifuged at 10,000 g and 4ºC for 1 min using TOMY
755 MX307, equipped with a Rack-in-Rotor and centrifugation rack PCR96-02. These steps were
756 very important to collect the droplet with a single cell in lysis buffer in an efficient manner. By
757 altering the volume of lysis buffer from 0.4 μL (as described in the original Quartz-Seq paper) to
758 1 μL (Quartz-Seq2), bubbling of lysis buffer was not required before single-cell sorting and the
759 lysis buffer could easily be handled. We agitated the plate at 2,600 rpm and 4ºC for 1 min using
760 ThermoMixer C (Eppendorf), after which we centrifuged the plate again. We then peeled open
761 the temporary seal and re-sealed it with Agilent PlateLoc Thermal Microplate Sealer (Agilent).
762 The resulting 384-well plate was then immediately cryopreserved at −80ºC and maintained
763 under such conditions until subsequent reverse transcription for cell barcoding. After
764 cryopreservation, we performed subsequent whole-transcript amplification using the
765 cryopreserved 384-well PCR plate within a few months.
766
767 **Whole-transcript amplification of Quartz-Seq2**
768 Cryopreserved 384-well plates with single-cell lysate were centrifuged at 10,000 g and 4ºC for 1
769 min. Subsequently, we denatured total RNA in each 384 plate at 70ºC for 90 s and hybridized the
770 RT primer to poly-adenylated RNA at 35ºC for 15 s using the C1000/S1000 thermal cycler. The
771 resulting plates were again centrifuged at 10,000 g and 4ºC for 1 min. Next, the plates were
772 placed on the 384 aluminum plate at 0ºC. We peeled away the seal and added 1 μL of RT
773 premix (2x Thermopol buffer, 5 units/μL SuperScript III, 0.55 units/μL RNasin plus) to 1 μL of
774 lysis buffer for each well using a Mantis microfluidic dispensing system (Formulatrix) or a 384
775 Transfer Plate system (1859-384S, Watson). The above RT solution was used for the RT25
776 condition. For the RT100 condition, we used the following RT solution: 2x Thermopol buffer, 20
777 units/μL SuperScript III, and 2.2 units/μL RNasin plus. We sealed the plates again and agitated
778 them at 2,600 rpm and 4ºC for 1 min. The plates were then centrifuged at 10,000 g and 4ºC for 1
779 min. We then performed reverse transcription at 35ºC for 5 min and 50ºC for 50 min. The reverse
780 transcription was stopped at 70ºC for 15 min. Then, the plates were placed on a prechilled
781 aluminum block, after which we peeled off their seals. Subsequently, we turned the plates upside

782 down on the assembled collector type A or type B (supplemental figure). We mainly used type A.
783 We centrifuged the plates with an assemble collector at 3,010 g and 4ºC for 3 min with
784 swing-bucket rotors. Subsequently, we collected the cDNA solution into a disposable reservoir.
785 Typically, we obtained 650–700 μL of cDNA solution from one 384-well PCR plate. We purified
786 and concentrated the cDNA solution using the DNA Clean & Concentrator™-5 kit (Zymo
787 Research). We used three purification columns for one 384-well PCR plate in the case of the
788 v3.1 RT primer system (384-cell barcode). Purified cDNA was extracted into 20 μL of
789 nuclease-free water from one column purification and transferred into an eight-linked PCR tube
790 (TaKaRa). The PCR tubes were placed on an aluminum PCR stand at 0ºC. We added 25 μL of
791 TdT solution [1x Thermopol buffer, 2.4 mM dATP, 0.0384 units/μL RNase H (Invitrogen), 26.88
792 units/μL terminal transferase (Roche)] into 20 μL of extracted cDNA using a pipette at 0ºC. The
793 resulting 45 μL of TdT solution was mixed with a pipette at 0ºC or ThermoMixer at 2,000 g and
794 0ºC for 1 min. Immediately thereafter, the PCR tubes were centrifuged at 10,000 g and 0ºC for 1
795 min. We used a C1000/S1000 thermal cycler equipped with the 96-Deep Well Reaction Module
796 for the following steps. The PCR tubes were placed on the block of the thermal cycler, which had
797 been prechilled to 0ºC. We then performed a poly-A tailing reaction at 37ºC for 75 s. The solution
798 was inactivated at 65ºC for 10 min. The PCR tubes were placed on an aluminum PCR stand at
799 0ºC. We then dispensed approximately 11 μL of solution into four wells from 45 μL of TdT
800 solution. We added 46.16 μL of PCR I premix (1.08492x MightyAmp Buffer version 2, 0.06932
801 μM Tagging primer, 0.05415 units/μL MightyAmp DNA polymerase) to 11 μL of TdT solution for
802 the respective wells of the PCR tube. We performed gentle inversion mixing on the resulting
803 solution in the PCR tube. The tubes were then centrifuged at 10,000 g and 4ºC for 1 min.
804 Subsequently, the solution was mixed with ThermoMixer at 2,000 rpm and 4ºC for 2 min. Then,
805 we spun down the tube again. Next, we denatured the solution at 98ºC for 130 s and hybridized
806 Tagging primer to poly-A-tailed cDNA at 40ºC for 1 min. After that, we performed the “Increment
807 step” by heating to 68ºC at 0.2ºC every second and performed second-strand synthesis at 68ºC
808 for 5 min. The tubes were placed on an aluminum PCR stand at 0ºC. We added 50.232 μL of
809 PCR II premix (0.99697x MightyAmp Buffer version.2, 1.8952 μM gM primer) to 56.16 μL of PCR
810 I solution. We performed gentle inversion mixing on the resulting solution in the PCR tube. The

811 tubes were then centrifuged at 10,000 g and 4ºC for 1 min. Subsequently, the solution was
812 mixed with ThermoMixer at 2,000 rpm and 4ºC for 2 min, after which we spun down the tube
813 again. We then placed it on the block of the thermal cycler at 68ºC. Subsequently, we amplified
814 the cDNA for 11 cycles under the following conditions: 98ºC for 10 s, 65ºC for 15 s, and 68ºC 5
815 min. We then incubated the tube at 68ºC for an additional 5 min. Finally, we transferred all of the
816 PCR solution, derived from one 384-well PCR plate, to a 50-mL Polypropylene Centrifuge Tube
817 (Watson). Typically, we obtained approximately 1.2 mL of PCR solution per 384-well PCR plate.
818 We added 32 μL of 3 M sodium acetate (pH 5.2) and 6420 μL PB-Buffer (Qiagen) to the PCR
819 solution. The mixture was then purified using a MinElute Spin Column (Qiagen). Purified cDNA
820 was extracted into 40 μL of nuclease-free water. We additionally purified the cDNA with 32 μL of
821 Ampure XP beads. Finally, we obtained 32 μL of purified cDNA. We checked the length
822 distribution of amplified cDNA with an Agilent High Sensitivity DNA Kit (Agilent). The typical
823 average size of the amplified cDNA in Quartz-Seq2 was approximately 1,400 bp (Additional file
824 1: Figure S2c). Primer sequences are listed in Additional file 3: Table S2.
825 In the case of the usage of the v3.2 RT primer in Quartz-Seq2, we modified the above
826 steps as follows. After reverse transcription, we collected cDNA solution into a disposable
827 reservoir from four sets of 384-well plates, which corresponded to 1,536 wells. We purified and
828 concentrated the cDNA solution using eight purification columns for four 384-well PCR plates in
829 the case of the v3.2 RT primer system. In the PCR step, we amplified cDNA for nine cycles with
830 the following conditions: 98ºC for 10 s, 65ºC for 15 s, and 68ºC for 5 min. Finally, we transferred
831 all of the PCR solution derived from four 384-well plates to a 50-mL Polypropylene Centrifuge
832 Tube (Watson). Typically, we obtained approximately 3.5 mL of PCR solution per four 384-well
833 PCR plates. We added 88 μL of 3 M sodium acetate (pH 5.2) and 17.6 mL of PB-Buffer (Qiagen)
834 to the PCR solution. The mixture was purified using the MinElute Spin Column (Qiagen).
835 Subsequently, cDNA was again purified using Ampure XP magnetic beads.
836 For the Quartz-Seq1-like reaction, we performed the following procedure in
837 accordance with our previous study. We added 1 μL of RT premix (2x PCR buffer, 5 units/μL
838 SuperScript III, 0.55 units/μL RNasin plus) to 1 μL of lysis buffer. We then performed reverse
839 transcription at 35ºC for 5 min and 45ºC for 20 min. This reverse transcription was stopped at

840 70ºC for 15 min. We added 5 μL of ExoIB solution (1.6x Exonuclease I buffer, 3.2x PCR buffer,
841 16 mM DTT) and 20 μL of TdT solution [1x PCR buffer, 3 mM dATP, 0.0384 units/μL RNase H
842 (Invitrogen), 33.6 units/μL terminal transferase (Roche)] into 20 μL of extracted cDNA using a
843 pipette at 0ºC. The PCR tubes were placed on the block of the thermal cycler that had been
844 prechilled to 0ºC. We performed a poly-A tailing reaction at 37ºC for 75 s. We then denatured the
845 solution at 98ºC for 130 s and hybridized Tagging primer to poly-A-tailed cDNA at 40ºC for 1 min.
846 After that, we performed second-strand synthesis at 68ºC for 5 min. Subsequently, we amplified
847 the cDNA via a PCR reaction for 12 cycles.
848
849 **Preparation of truncated sequence adaptor**
850 The truncated sequence adaptor was composed of an rYshapeP5 primer (HPLC-purified) and
851 rYshapeP7LT primers (HPLC-purified), which had a TruSeqLT-compatible pool barcode. We
852 prepared 100 μM respective primers with adaptor buffer (10 mM Tris-HCl pH7.8, 0.1 mM EDTA
853 pH8.0, 50 mM NaCl). We added 5 μL of 100 μM rYshapeP5 primer and rYshapeP7LTxx primer
854 into a single PCR tube. We denatured the solution at 90ºC for 90 s. After that, we achieved
855 annealing by cooling to 10ºC by 0.5ºC every 30 s and then maintaining the sample at 4ºC. We
856 then placed the tube on an aluminum PCR stand at 0ºC. Subsequently, we added adaptor buffer,
857 which was prechilled at 0ºC, to 10 μL of 50 μM truncated adaptor. Finally, we obtained about 50
858 μL of 10 μM truncated adaptor. We cryopreserved 1 μL of 10 μM truncated adaptor in a PCR
859 tube at −80ºC until usage in the adaptor ligation step. Primer sequences are listed in Additional
860 file 3: Table S2.
861
862 **Sequence library preparation of Quartz-Seq2**
863 We added 130 μL of nuclease-free water including 5–10 ng of amplified cDNA into a Crimp-Cap
864 microTUBE with AFA Fiber. We then performed cDNA fragmentation using an LE220
865 Focused-ultrasonicator (Covaris) under the following conditions: duty factor 15%, peak incident
866 power 450 W, cycles per burst 200, and treatment time 80 s. We purified and concentrated the
867 cDNA solution using the DNA Clean & Concentrator™-5 kit. Purified cDNA was extracted into 10

868 μL of nuclease-free water from one column purification and transferred into an eight-linked PCR
869 tube (TaKaRa).
870 We then added 2 μL of End-Repair premix [1.4 μL of End repair & A-tailing buffer and
871 0.6 μL of End repair & A-tailing Enzyme (KAPA Biosystems)] to 10 μL of fragmented cDNA
872 solution. Subsequently, we mixed the solution by pipetting on an aluminum PCR stand at 0ºC.
873 The PCR tubes were placed on the block of the thermal cycler, which had been prechilled to
874 20ºC. We then incubated the tubes at 37ºC for 60 min and 65ºC for 30 min. After this, we added
875 2 μL of adaptor buffer (1.5 μM truncated adaptor, 10 mM Tris-HCl pH7.8, 0.1 mM EDTA pH8.0,
876 50 mM NaCl) and 8 μL of ligation premix [6 μL of ligation buffer, 2 μL of DNA ligase (KAPA
877 Biosystems)] at 4ºC. The 22 μL of ligation solution was well mixed by pipetting at 4ºC. Then, we
878 performed adaptor ligation at 20ºC for 15 min. After ligation, we added 18 μL of Ampure XP
879 beads to 22 μL of adaptor ligation solution and mixed them well. Adaptor-ligated cDNA was
880 extracted into 20 μL of nuclease-free water. We added 32 μL of PCR premix [25 μL of 2xKAPA
881 HiFi ready mix, 1.75 μL of 10 μM TPC2 primer (HPLC-purified), 10 μM P5-gMac hybrid primer
882 (HPLC-purified)] to 18 μL of adaptor-ligated cDNA. We denatured the solution at 98ºC for 45 s.
883 Subsequently, we amplified cDNA for eight cycles under the following conditions: 98ºC for 15 s,
884 60ºC for 30 s, and 72ºC for 1 min. Finally, we additionally incubated the tube at 72ºC for 5 min.
885 We added 40 μL of Ampure XP beads to 50 μL of PCR solution. Purified
886 sequence-library DNA was eluted into 20–30 μL of nuclease-free water. We checked the DNA
887 concentration and DNA size of sequence library DNA using the Agilent High Sensitivity DNA Kit
888 (Agilent) and QuantiFluor® dsDNA System (Promega). We also prepared sequence library DNA
889 with Tn5 transposase, in accordance with a modified version of a procedure used in a previous
890 study [6]. Specifically, the modifications were as follows. We used 0.75 ng of amplified cDNA for
891 the Nextera XT library preparation kit. We amplified sequence library DNA with the P5-gMac
892 hybrid primer and the Nextera XT primer, which has the P7 sequence.
893
894 **Deep sequencing for Quartz-Seq2**
895 We analyzed the sequence library DNA from Quartz-Seq2 using NextSeq500 (Illumina) and
896 HiSeq2500 (Illumina). For the Read1 sequence, we used a custom sequence primer named

897 Read1DropQuartz primer (HPLC-purified). In the case of the v3.1 RT primer, sequence
898 specification was as follows (Read1: 22 cycles, Index1: 6 cycles, Read2: 64–118 cycles). In the
899 case of the v3.2 RT primer, sequence specification was as follows (Read1: 23 cycles, Index1: 6
900 cycles, Read2: 63 cycles). When Read1 length was within 64 nt we mainly used the NextSeq
901 500/550 High Output v2 Kit (75 cycles), which can be used for sequencing up to 92 nt. For long
902 sequences (over 64 nt), we analyzed sequence library DNA using the HiSeq Rapid SBS Kit v2.
903 Sequence analysis with HiSeq2500 was supported by the Genomic Network Analysis Support
904 facility (GeNAS) in RIKEN.
905
906 **Amplification-free single-cell qPCR analysis**
907 Amplification-free single-cell qPCR was performed in accordance with our previous study with
908 the following modifications [15]. We collected a single cell into 1 μL of lysis buffer (2.5 μM
909 random hexamer, 1 mM dNTP mix, 0.3% NP40, 2 units/μL RNasin plus) of a 384-well PCR plate
910 (Eppendorf) using SH800. The resulting 384-well PCR plate was cryopreserved at −80ºC. We
911 denatured total RNA in the 384-well PCR plate at 70ºC for 1.5 min using a C1000/S1000 thermal
912 cycler and hybridized the random primer to RNA at 0ºC for 2 min. We added 1 μL of
913 reverse-transcription solution (2x SSIV buffer, 10 mM DTT, 2 units/μL RNasin plus, 20 units/μL
914 SuperScript IV) to the lysis buffer. Subsequently, we performed reverse transcription at 23ºC for
915 10 min and 50ºC for 10 min. The reverse transcription was stopped at 80ºC for 10 min. The
916 resulting solution was diluted with qPCR solution (10 mM Tris-HCl pH 8.0, 0.05% Tween-20).
917 The obtained diluted solution was then used for qPCR detection with QuantiTect SYBR Green
918 PCR Master Mix and the LightCycler480 system. For primer sets for each gene, see Additional
919 file 3: Table S2.
920
921 **Drop-seq experiments and data analysis**
922 Drop-seq was performed as reported previously [6] and in line with an online protocol
923 (http://mccarrolllab.com/dropseq/), but with the following modifications: flow rates for oil and
924 aqueous suspensions were 14,000 μl/h and 4,000 μl/h, respectively. The diameter of droplets
925 was 95–100 μm. Microfluidic devices were fabricated by Fluidware Technologies (Japan). The

926 lot number of barcoded beads was 051415 (ChemGenes). Data analysis for Drop-seq was
927 performed as described online (http://mccarrolllab.com/dropseq/). The versions of the software
928 and databases were as follows: STAR: v2.5.1b; mouse genome: mm10; genome annotation:
929 gencode GRCm38.p4 vM9; and Drop-seq tools: v1.11.
930
931 **Quartz-Seq2 read alignment and generation of digital expression data**
932 The structure of the sequence library and data processing was designed based on those of
933 Drop-seq ([6] and online as referenced above). BCL files generated by Illumina NextSeq500
934 were converted to fastq files by bcl2fastq2 (v2.17.1.14) with demultiplexing pool barcodes. The
935 --mask-short-adapter-reads parameter was set to 20. If needed, fastq reads were randomly
936 downsampled by seqtk software (version: sgdp). The last two bases of Read2, the sequence
937 quality of which was relatively low, were removed by FASTX-Toolkit (version: 0.0.14). Fastq files
938 for Read1 and Read2 were converted to a bam file by FastqToSam of Picard tools (version:
939 1.134). Extracting cell barcodes and UMI (equal to molecular barcode) and filtering out reads
940 with low barcode quality were performed using Drop-seq tools (version: 1.11). The resulting bam
941 files were re-converted to fastq files by SamToFastq of Picard tools, and mapped to the mouse
942 genome (mm10) using STAR (version: 2.5.1b). After sorting the resulting bam files by SortSam
943 of Picard tools, the unaligned bam and aligned bam were merged by MergeBamAlignment of
944 Picard tools. Then, gene names were assigned to each record using TagReadWithGeneExon of
945 Drop-seq tools and gtf file (version: gencode GRCm38.p4 vM9). For the correction of errors of
946 cell barcodes considering the Sequence–Levenshtein distance, a custom-made python program
947 was used, which enabled the correction of up to two errors of substitution, insertion, or deletion.
948 This program used python2 (version: 2.7.11+), PypeR (version: 1.1.2), R (version: 3.2.3),
949 Bioconductor (version: 3.2), and Bioconductor package DNABarcodes (version: 1.0.0). Finally,
950 the UMI for each gene for each cell was counted using DigitalExpression of Drop-seq tools,
951 which generated the digital expression matrix. To generate a non-UMI-filtered matrix, a
952 custom-made python program that counts reads for each gene for each cell was used.
953
954 **Dimensionality reduction, clustering, and term analysis**

955 Cells with low detected gene counts were removed for further analysis (Quartz-Seq2 on ES/PrE
956 mixture: 4,000 genes; Quartz-Seq2 on SVF: 500 genes). Total counts of each cell were
957 normalized to 10,000 or the mean of total UMI counts for cells. UMI counts had 1 added to them
958 and were then log-transformed (natural logarithm), after which they were scaled to have mean
959 and variance for each gene of 0 and 1, respectively. For PCA, UMI counts for all detected genes
960 (for ES/PrE mixture) or genes with highly variable expression (for SVF) were used. For t-SNE,
961 the top 10 to 40 principal components of matrices produced by PCA were used. For clustering,
962 the DBSCAN algorithm was used. The values of the parameter epsilon were 0.69 for
963 Quartz-Seq2 on the 4,500 mouse ES/PrE mixture, 5 for Quartz-Seq2 on the 384 mouse ES/PrE
964 mixture, 3 for Drop-seq on the 500 mouse ES/PrE mixture, and 2.2 for SVF analysis.
965 Significantly enriched GO terms with the top principal components were calculated using the
966 GO-PCA package [43].
967
968 **Identification of differentially expressed genes for each cluster**
969 Marker genes for each cluster were identified based on a generalized linear model. The
970 difference in deviance between two models, in which the gene was or was not differentially
971 expressed between two clusters, was calculated for the genes. To filter out noise, pseudocount 1
972 was added to the averaged gene expression in clusters and only genes with fold change of 2 or
973 more between two clusters were further analyzed. P-values were calculated as 1 − cumulative
974 density function of a chi-squared continuous random variable for difference in deviance. After
975 corrections for multiple testing, genes with FDR of less than 0.05 were identified as differentially
976 expressed genes for the cluster. For comparison between differentially expressed genes
977 identified by scRNA-seq methods and those identified by bulk RNA-seq, this pseudocount was
978 not added.
979
980 **Identification of cell cycle-associated genes and cell cycle-phase independent variable**
981 **genes**
982 To identify genes for which the expression changes depending on the phase of the cell cycle, we
983 used the intensity of Hoechst 33342 staining measured using a cell sorter. First, cells were

984 discretized into 40 equal-sized buckets based on the rank of Hoechst 33342 staining intensity.
985 Then, the CV of averaged UMI counts for a gene in each bin was calculated. After z-scaling of
986 this, genes with high z-scores were identified. To identify genes whose expression fluctuates in a
987 manner not dependent on the cell cycle phase, we calculated the CV of UMI counts for a gene in
988 each cell and the CV of averaged UMI counts for a gene in each bin. After z-scaling of such data,
989 the genes of which difference between two scaled CV was large were identified as cell
990 cycle-phase independent variable genes.
991
992 **Enrichment analysis on pathway and Gene Ontology (GO) terms**
993 Pathways that were particularly enriched for the differentially expressed genes were calculated
994 using the ReactomePA package of R Bioconductor (version: 1.14.4, [44]) and Metascape
995 (http://metascape.org/ [45]). The cut-off parameter for the q value was 0.05. Terms that were
996 enriched for genes with highly variable expression were calculated using DAVID, the Database
997 for Annotation, Visualization and Integrated Discovery (version: 6.8, [46,47]). Ontology used for
998 calculation was GOTERM_BP_FAT, GOTERM_MF_FAT, and GOTERM_CC_FAT. The terms
999 with FDR < 0.05 were identified as enriched terms.
1000
1001 **Data resources**
1002 Raw and processed data files for Quartz-Seq2 and Drop-seq experiments are available under
1003 GEO: GSE99866 for single-cell RNA-seq data and DRA002954 for bulk-RNA-seq data.
1004
1005
1006
1007
1008
1009
1010
1011
1012

1013 **List of abbreviations**

1014

1015 UMI: unique molecular identifier

1016 PCR: polymerase chain reaction

1017 SCC: Spearman's rank correlation coefficient

1018 CV: coefficient of variation

1019 ES cell: embryonic stem cell

1020 PrE cell: primitive endoderm cell

1021 SVF: stromal vascular fraction

1022 RT: reverse-transcription

1023 Dex: dexamethasone

1024 PCA: principal component analysis

1025 MSC: mesenchymal stem cell

1026 TdT: terminal deoxynucleotidyl transferase

1027 GO: Gene Ontology

1028 FDR: false discovery rate

1029 t-SNE: t-distributed stochastic neighbor embedding

1030 nt: nucleotide

1031 bp: base pair

1032 SSC: side scatter

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042 **Declarations**

1043

1044 **Ethics approval and consent to participate**

1045 Not applicable.

1046

1047 **Consent for publication**

1048 Not applicable.

1049

1050 **Availability of data and material**

1051 The data have been deposited under GEO accession number GSE99866 and DRA accession

1052 number: DRA002954.

1053

1054 **Competing interests**

1055 The authors declare that they have no competing interests.

1056

1057 **Funding**

1058 This work was supported by the Projects for Technological Development, Research Center

1059 Network for Realization of Regenerative Medicine by Japan, the Japan Agency for Medical

1060 Research and Development (AMED), and the Japan Science and Technology Agency (JST).

1061 The sequence analyses of bulk RNA-seq were supported by MEXT KAKENHI (No. 221S0002)

1062 and JSPS KAKENHI Grant Number JP24651218 and JSPS KAKENHI Grant number

1063 JP15K16329. The sequence operations of Quartz-Seq2 using HiSeq2500 were supported by the

1064 Platform Project for Supporting Drug Discovery and Life Science Research (Platform for Drug

1065 Discovery, Informatics, and Structural Life Science) from AMED.

1066

1067 **Author's contributions**

1068 YS, HD and IN designed the study. YS designed and performed experiments of single-cell

1069 RNA-seq. HD designed and performed data analysis. HT and AK designed and performed

1070 experiments of mouse stromal vascular fraction. ME assisted the experiments. TH assisted the

development of spin-down collection system. YS, HD, HT, AK and IN wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We greatly appreciate the Drop-seq team for useful comments on our technical introduction to Drop-seq. We thank Mr. Akihiro Matsushima and Mr. Manabu Ishii for assistance with the infrastructure for the data analysis. We thank Dr. Yutaka Suzuki for assistance with the bulk RNA-seq. We are also grateful to Dr. Mika Yoshimura and Dr. Haruka Ozaki (RIKEN) for assistance with the bulk RNA-seq analysis. Thanks also go to Sony's cell sorter team for special customization of the 384-well sorting mode of SH800.

1100 **References**

1101

1102 1. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel
1103 digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.
1104 2. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A Single-Cell
1105 Transcriptome Atlas of the Human Pancreas. Cell Syst. 2016;3:385–394.e3.
1106 3. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al.
1107 CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016;17:892.
1108 4. Fan X, Zhang X, Wu X, Guo H, Hu Y, Tang F, et al. Single-cell RNA-seq transcriptome
1109 analysis of linear and circular RNAs in mouse preimplantation embryos. Genome Biol.
1110 2015;16:236.
1111 5. Nakamura T, Yabuta Y, Okamoto I, Aramaki S, Yokobayashi S, Kurimoto K, et al. SC3-seq: a
1112 method for highly parallel and quantitative measurement of single-cell gene expression. Nucleic
1113 Acids Res. 2015;43:e60–0.
1114 6. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel
1115 Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell.
1116 2015;161:1202–14.
1117 7. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet Barcoding for
1118 Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell. 2015;161:1187–201.
1119 8. Fan HC, Fu GK, Fodor SPA. Combinatorial labeling of single cells for gene expression
1120 cytometry. Science. 2015;347:1258367–7.
1121 9. Bose S, Wan Z, Carr A, Rizvi A, Vieira G, Pe'er D, et al. Scalable microfluidics for single-cell
1122 RNA printing and sequencing. Genome Biol. 2015;16:120.
1123 10. Streets AM, Zhang X, Cao C, Pang Y, Wu X, Xiong L, et al. Microfluidic single-cell
1124 whole-transcriptome sequencing. Proc. Natl. Acad. Sci. 2014;111:7048–53.
1125 11. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively
1126 Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types.
1127 Science. 2014;343:776–9.

1128 12. Soumillon M, Cacchiarelli D, Semrau S, van Oudenaarden A, Mikkelsen TS. Characterization
1129 of directed differentiation by high-throughput single-cell RNA-Seq. 2014.
1130 13. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell
1131 RNA-seq with unique molecular identifiers. Nat. Methods. 2013;11:163–6.
1132 14. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for
1133 sensitive full-length transcriptome profiling in single cells. Nat. Methods. 2013;10:1096–8.
1134 15. Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T, et al. Quartz-Seq: a highly
1135 reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic
1136 gene-expression heterogeneity. Genome Biol. 2013;14:R31.
1137 16. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed
1138 Linear Amplification. Cell Rep. 2012;2:666–73.
1139 17. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq
1140 from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol.
1141 2012;30:777–82.
1142 18. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, et al. Characterization of the
1143 single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21:1160–
1144 7.
1145 19. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq
1146 whole-transcriptome analysis of a single cell. Nat. Methods. 2009;6:377–82.
1147 20. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell
1148 transcriptomics. Nat. Methods. 2014;11:637–40.
1149 21. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al.
1150 Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell. 2017;65:631–4.
1151 22. Buschmann T. DNABarcodes: an R package for the systematic construction of DNA sample
1152 tags. Bioinformatics. 2017;:btw759.
1153 23. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification
1154 of low quality cells from single-cell RNA-seq data. Genome Biol. 2016;17:57.
1155 24. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee S-L, et al. Zscan4 regulates
1156 telomere elongation and genomic stability in ES cells. Nature. 2010;464:858–63.

1157 25. Pak J, Lee JH, Park KS, Park M, Kang L-W, Lee SH. Current use of autologous adipose
1158 tissue-derived stromal vascular fraction cells for orthopedic applications. J. Biomed. Sci.
1159 2017;24:1496.
1160 26. Trounson A, McDonald C. Stem Cell Therapies in Clinical Trials: Progress and Challenges.
1161 Cell Stem Cell. 2015;17:11–22.
1162 27. Baer PC, Geiger H. Adipose-Derived Mesenchymal Stromal/Stem Cells: Tissue Localization,
1163 Characterization, and Heterogeneity. Stem Cells Int. 2012;2012:1–11.
1164 28. Lindroos B, Suuronen R, Miettinen S. The Potential of Adipose Stem Cells in Regenerative
1165 Medicine. Stem Cell Rev. 2010;7:269–91.
1166 29. Kwon HM, Hur S-M, Park K-Y, Kim C-K, Kim Y-M, Kim H-S, et al. Multiple paracrine factors
1167 secreted by mesenchymal stem cells contribute to angiogenesis. Vascular Pharmacology.
1168 2014;63:19–28.
1169 30. Madrigal M, Rao KS, Riordan NH. A review of therapeutic effects of mesenchymal stem cell
1170 secretions and induction of secretory modification by different culture methods. J Transl Med.
1171 2014;12:71.
1172 31. Rennert RC, Januszyk M, Sorkin M, Rodrigues M, Maan ZN, Duscher D, et al. Microfluidic
1173 single-cell transcriptional analysis rationally identifies novel surface marker profiles to enhance
1174 cell-based therapies. Nat Commun. 2016;7:11945.
1175 32. Baer PC. Adipose-derived mesenchymal stromal/stem cells: An update on their phenotype in
1176 vivo and in vitro. World J Stem Cells. 2014;6:256.
1177 33. Mabuchi Y, Houlihan DD, Akazawa C, Okano H, Matsuzaki Y. Prospective Isolation of
1178 Murine and Human Bone Marrow Mesenchymal Stem Cells Based on Surface Markers. Stem
1179 Cells Int. 2013;2013:1–7.
1180 34. Dromard CC, Barreau C, Andr M, Berger-M ller S, Casteilla L, Planat-Benard V. Mouse
1181 adipose tissue stromal cells give rise to skeletal and cardiomyogenic cell sub-populations. Front
1182 Cell Dev Biol. 2014;2:697.
1183 35. Heneidi S, Simerman AA, Keller E, Singh P, Li X, Dumesic DA, et al. Awakened by Cellular
1184 Stress: Isolation and Characterization of a Novel Population of Pluripotent Stem Cells Derived
1185 from Human Adipose Tissue. Ljubimov AV, editor. PLoS ONE. 2013;8:e64752.

1186 36. Wachi H. Role of Elastic Fibers on Cardiovascular Disease. J. Health Sci. The
1187 Pharmaceutical Society of Japan; 2011;57:449–57.
1188 37. Kurimoto K. An improved single-cell cDNA amplification method for efficient high-density
1189 oligonucleotide microarray analysis. Nucleic Acids Res. 2006;34:e42–2.
1190 38. Yang L, Ma Z, Cao C, Zhang Y, Wu X, Lee R, et al. MR-Seq: measuring a single cell's
1191 transcriptome repeatedly by RNA-seq. Science Bulletin. 2017.
1192 39. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human
1193 haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol.
1194 2017;Complex Sy:1695.
1195 40. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq
1196 reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem
1197 cells. Genome Res. 2015;25:1860–72.
1198 41. Pauklin S, Vallier L. The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity.
1199 Cell. 2013;155:135–47.
1200 42. Shimosato D, Shiki M, Niwa H. Extra-embryonic endoderm cells derived from ES cells
1201 induced by GATA Factors acquire the character of XEN cells. BMC Dev. Biol. 2007;7:80.
1202 43. Wagner F. GO-PCA: An Unsupervised Method to Explore Gene Expression Data Using Prior
1203 Knowledge. Li X, editor. PLoS ONE. 2015;10:e0143196.
1204 44. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis
1205 and visualization. Mol. BioSyst. 2016;12:477–9.
1206 45. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and
1207 Orthogonal Integration of Influenza ?OMICs? Data Defines a Role for UBR4 in Virus Budding.
1208 Cell Host Microbe. 2015;18:723–35.
1209 46. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
1210 lists using DAVID bioinformatics resources. Nat Protoc. 2008;4:44–57.
1211 47. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the
1212 comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2008;37:1–13.
1213
1214

1215 **Figure titles and legends**

1216

1217 **Figure 1. Overview of Quartz-Seq2 experimental processes**

1218 Quartz-Seq2 consisted of five steps. (1) Each single cell in a droplet was sorted into lysis buffer
1219 in each well of a 384-well PCR plate using flow cytometry analysis data. (2) Poly-adenylated
1220 RNA in each well was reverse-transcribed into first-strand cDNA with reverse transcription
1221 primer, which had a unique cell barcode. We prepared 384 or 1,536 kinds of cell barcode (CB)
1222 with a unique sequence based on the Sequence–Levenshtein distance (SeqLv). The edit
1223 distance of SeqLv was 5. RT primer also had a UMI sequence for reduction of PCR bias (MB)
1224 and a poly-dT sequence for binding to poly-A RNA. (3) Cell barcode-labeled cDNAs from all 384
1225 wells were promptly collected by centrifugation using assembled collectors. (4) Collected
1226 first-strand cDNAs were purified and concentrated for subsequent whole-transcript amplification.
1227 In the poly-A tailing step, purified cDNA was extended with a poly-A tail by terminal
1228 deoxynucleotidyl transferase (TdT). Subsequently, second-strand cDNA was synthesized with a
1229 tagging primer, which had a poly-dT sequence. The resulting second-strand cDNA had a PCR
1230 primer sequence (M) at both ends of the cDNA. The cDNA was amplifiable at subsequent PCR
1231 amplification. (5) For conversion from amplified cDNA to sequence library DNA, we fragmented
1232 the amplified cDNA using the ultrasonicator Covaris. Such fragmented cDNA was ligated with a
1233 truncated Y-shaped sequence adaptor, which had an Illumina flow-cell binding sequence (P7)
1234 and a pool barcode sequence (PB). PB made it possible to mix different sets of
1235 cell-barcode-labeled cDNA. Ligated cDNA, which had CB and MB sequences, was enriched by
1236 PCR amplification. The resulting sequence library DNA contained P7 and P5 flow-cell binding
1237 sequences at respective ends of the DNA. We sequenced the cell-barcode site and the UMI site
1238 at Read1, the pool-barcode site at Index1, and the transcript sequence at Read2.

1239

1240 **Figure 2. Sequence performance of Quartz-Seq2 with molecular biological improvements**

1241 a) Improvement of poly-A tagging efficiency. Barplot represents the relative DNA yield in various
1242 poly-A tagging conditions using purified first-strand cDNA from 1 ng of total RNA. T55 buffer as
1243 the terminal deoxynucleotidyl transferase (TdT) buffer and the temperature condition “Increment”

1244 for the poly-A tagging step improved the cDNA yield of whole-transcript amplification. Buffer
1245 compositions were indicated in Additional file 4: Table S3. QuartzB represents use of a
1246 Quartz-Seq-like buffer as a positive control, in accordance with the approach described in the
1247 original Quartz-Seq paper. Finally, we quantified cDNA yield (300–9,000 bp) and byproduct DNA
1248 yield (50–300 bp) using Bioanalyzer (Agilent). (b) The improvement of cDNA yield with the
1249 combination of T55 buffer and "Increment" (T55+Inc) was confirmed by qPCR assay. We
1250 detected eight genes (*Eef1b2*, *Nanog*, *Pou5f1*, *Sox2*, *Utf1*, *Spp1*, *Ywhae*, *Tbp*) with amplified
1251 cDNA and nonamplified cDNA. We prepared nonamplified cDNA with 200 pg of total RNA and
1252 random hexamer. SCC stands for Spearman's rank correlation coefficient. c) Comparison
1253 between Quartz-Seq2 and Quartz-Seq-like conditions regarding sequence performance. We
1254 analyzed 384 wells with 10 pg of total RNA and used approximately 0.19 M fastq reads per well
1255 on average. We show the UMI count and gene count in barplots. We also show a diagrammatic
1256 depiction of the relationship between the cumulative number of detected genes and the gene
1257 expression level. d) Gene expression reproducibility between bulk poly-A-RNA-seq (1 μg of total
1258 RNA) and Quartz-Seq2 (10 pg of total RNA, averaged over 384 wells). e) The panels show a
1259 scatter plot between the mean of gene expression and the variability of gene expression with 10
1260 pg of total RNA in 384 wells. Red lines represent the theoretical variability of gene expression in
1261 the form of a Poisson distribution. The left and right panels represent the mean-CV plots before
1262 and after UMI filtering. The term "RT100" represents the condition of enzyme concentration in
1263 reverse transcription for Quartz-Seq2 (see Material and Methods).
1264
1265 **Figure 3. Cost reduction of Quartz-Seq2 upon improvement in quantitative performance**
1266 a) Reverse-transcription step has a highest share in single-cell experimental part of Quartz-Seq2.
1267 We present the different proportions of costs associated with each of the experimental steps.
1268 The reverse-transcription step accounts for approximately 65% of the total cost in the "RT100"
1269 condition. We thus established the "RT25" condition by reducing the enzyme concentration to
1270 25% in reverse transcription (for details, see Material and Methods). b) Barplots present the
1271 amount of amplified cDNA from three technical replicates of a 384-well plate with 10 pg of total
1272 RNA in all wells for the "RT100" and "RT25" conditions. The y-axis represents the yield of

1273 amplified cDNA. The cDNA yields from 10 pg of total RNA were 55.9 ± 5.5 ng for “RT100” and
1274 65.7 ± 2.5 ng for “RT25”. c) We analyzed amplified cDNA from 10 pg of total RNA by
1275 Quartz-Seq2 with various initial data sizes (approximately 0.27 M, 0.19 M, 0.171 M, 0.152 M,
1276 0.133 M, 0.114 M, 0.095 M, 0.076 M, 0.057 M, 0.038 M and 0.019 M fastq reads per 10 pg of
1277 total RNA on average) from NextSeq500. The Read2 length was 62 nt. In all panels, the x-axis
1278 represents the initial data size for one sample. We present the UMI count, gene count, and spike
1279 capture ratio in the upper panels. We also show the technical variability of the UMI count or gene
1280 count between wells in the lower panels. The y-axis presents the coefficient of variation of the
1281 UMI count or the gene count. d) Relationship between initial data size and the number of single
1282 cells for sequence analysis in NextSeq500 runs. Typically, one sequence run with NextSeq
1283 500/550 High Output v2 Kit reads out 400–450 M fastq reads. The x-axis represents the input
1284 cell number for one sequence run. The y-axis represents the initial data size (fastq reads) per
1285 cell on average. If we analyzed 4,000 single cells in one sequence run, approximately 0.1 M
1286 fastq reads on average would be assigned to one single cell. Red circle represents the typical
1287 range of shallow input read depth for a single cell. e) Capture ratio of ERCC spike mix I in the
1288 “RT25” condition. The capture ratio of ERCC spike-in mix I was 16.38±0.66% (n=3) at
1289 approximately 0.27 M fastq reads on average per well. The x-axis represents the copy number of
1290 input spike RNAs. The y-axis represents the averaged UMI count for ERCC spike RNAs.
1291
1292 **Figure 4. Quartz-Seq2 with high UMI conversion efficiency identified more marker genes**
1293 **and biological pathways**
1294 a) We defined the UMI counts divided by the raw fastq reads as the UMI conversion efficiency. b)
1295 UMI counts, gene counts, and UMI conversion efficiency for Quartz-Seq2 and Drop-seq2
1296 experiments. These values depended on the initial fastq reads. c) Venn diagram of genes that
1297 were differentially expressed between the ES cluster and the Dex-treated ES (PrE) cluster, as
1298 identified by Quartz-Seq2, Drop-seq, and bulk RNA-seq (left). The number of genes that differed
1299 in expression levels between ES and PrE cells by at least twofold was counted (FDR < 0.05). We
1300 also present a Venn diagram of the Reactome pathway (right). d) Venn diagram of GO terms for
1301 genes with highly variable expression, as identified by Quartz-Seq2 and Drop-seq.

1302

1303 **Figure 5. Quartz-Seq2 analysis of 4,484 cells from mouse embryonic stem cells and**
1304 **differentiated cells**

1305 a) We successfully analyzed 97.3% of 4,608 wells. The procedures for cell suspension as used
1306 in this assay are shown. Cells cultured under ES-maintenance and Dex-treatment conditions
1307 were separately dissociated into single cells, stained with Hoechst 33342 and/or Calcein-AM,
1308 and mixed evenly. Calcein-AM-positive and -negative cells were sorted to 384-well plates in a
1309 checkered pattern. White scale bars represent 100 μm. b) Clustering of 4,484 single cells
1310 according to the transcriptome. Plotting of cells on t-SNE space with color labeling for each
1311 cluster. The percentage indicates the proportion of cells for each cluster relative to all cells
1312 analyzed. Numbers in parentheses indicate the numbers of cells making up the cluster. c)
1313 Marker genes for each cluster identified by Quart-Seq2. Cluster-specific or cluster-enriched
1314 genes were calculated for each cluster, and their expression is displayed as a color in a heatmap.
1315 No more than 50 cells are shown for simplicity. d) Reconstructed distribution of Calcein-AM
1316 intensity for each cluster. The x-axis represents the intensity of Calcein-AM dye staining. e)
1317 Reconstructed distribution of Hoechst 33342 intensity for each cluster. The y-axis represents the
1318 density of cells. The x-axis represents the intensity of Hoechst 33342 dye staining.

1319

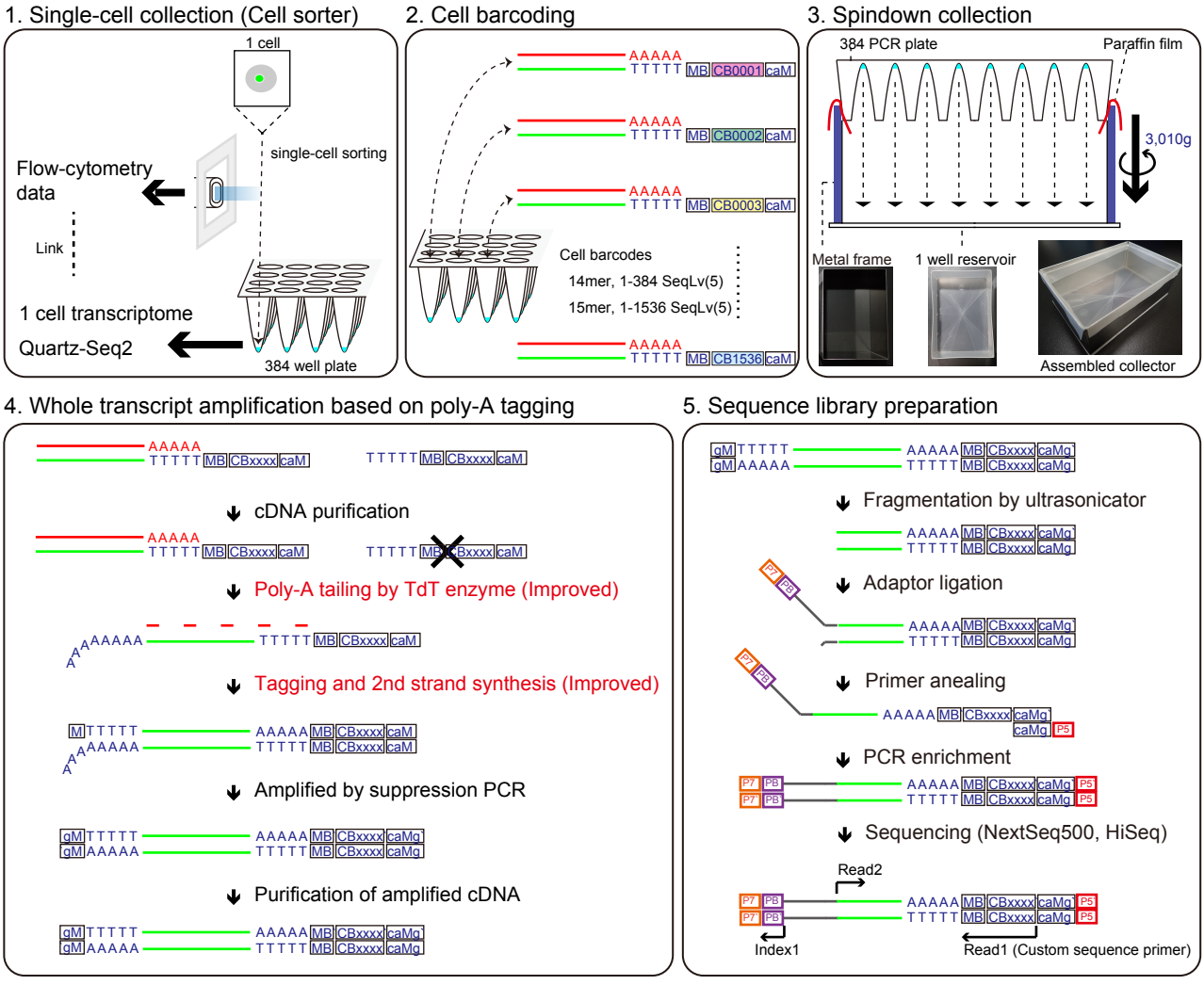
1320 **Figure 6 Quartz-Seq2 analysis on two types of cell state: cell cycle-dependent and**
1321 **-independent cellular heterogeneity**

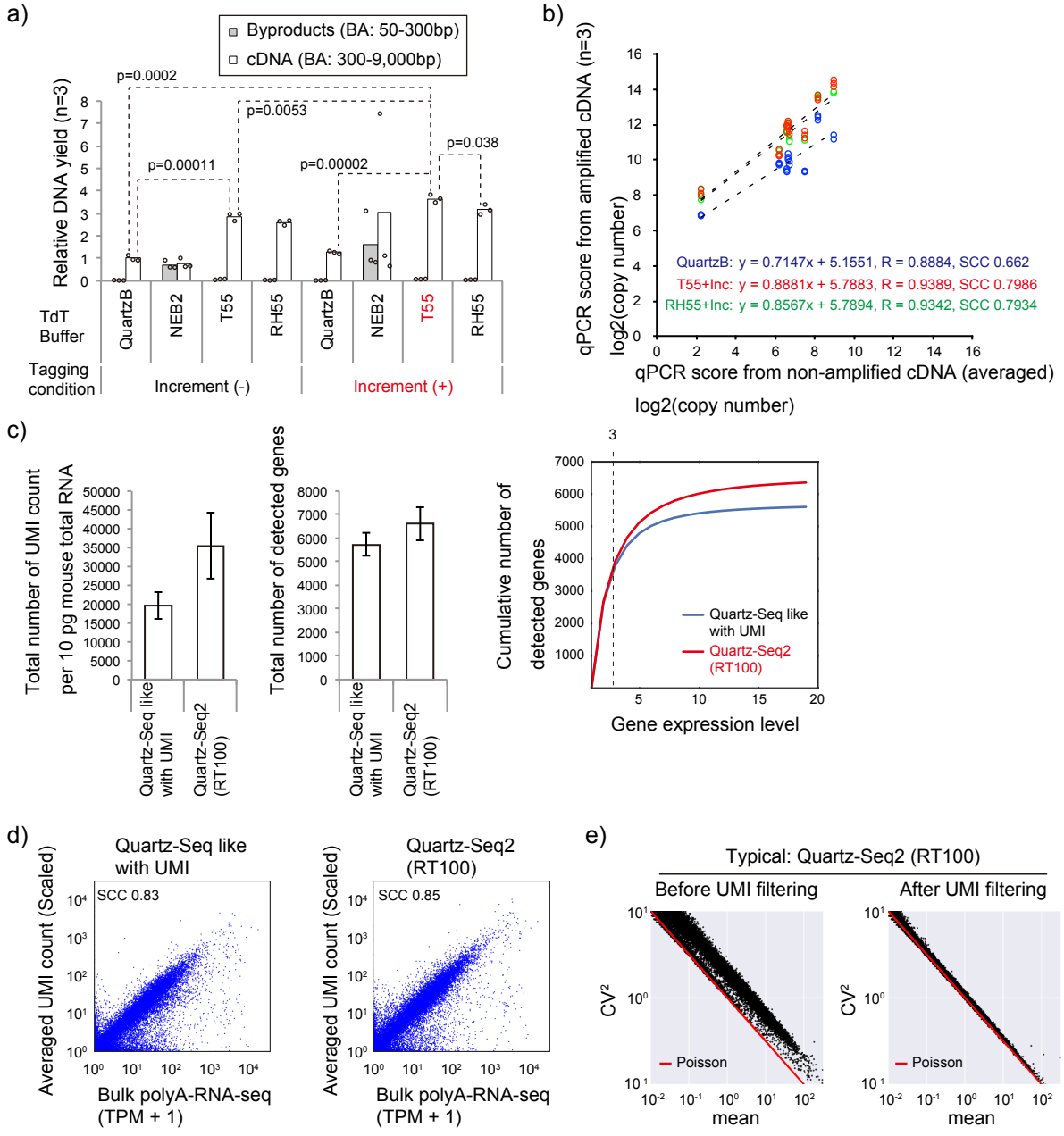
1322 a) Plotting cells on t-SNE space with the colors representing the intensity of Hoechst 33342
1323 staining. b) Dependency of expression of several genes on the intensity of Hoechst 33342
1324 staining [G2/M phase: *Ccnb1*; G1 phase: *Ccne1*; housekeeping gene: *Gnb2l1*; upper panels: ES
1325 main population (cluster 1); lower panels: PrE main population (cluster 4)]. c) Identification of cell
1326 cycle-associated genes. Average gene expression of cells in a bin is colored in the heatmap.
1327 Bins were arranged in order of Hoechst 33342 staining intensity. d) The number of cell
1328 cycle-associated genes identified for each z-threshold. e) Several examples for genes whose
1329 expression was variable, but not associated with the phase of the cell cycle. f) Single-cell qPCR
1330 detection of genes with variable expression.

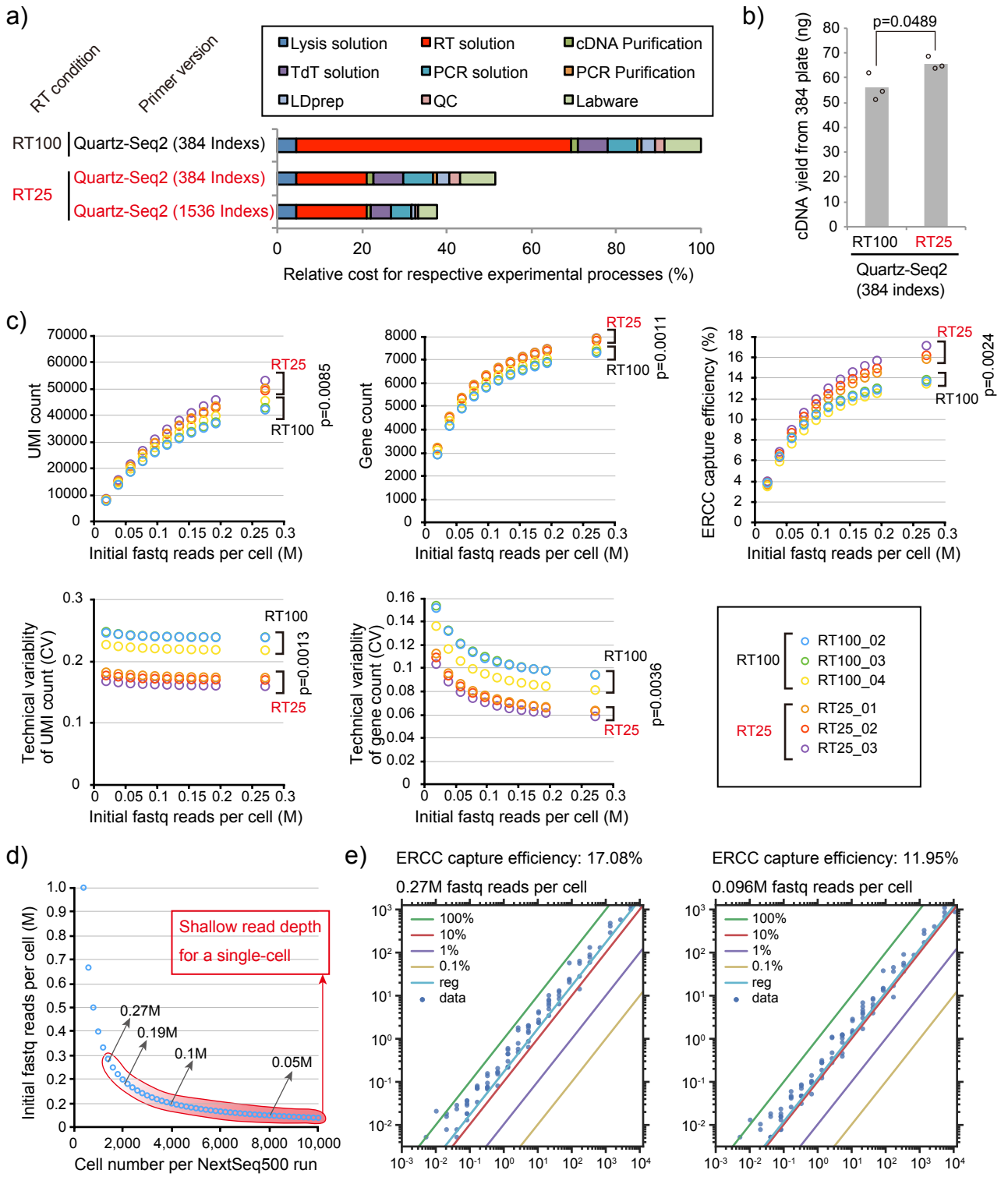
1331

1332 **Figure 7 Quartz-Seq2 analysis of stromal vascular fraction (SVF) from mouse adipose**
1333 **tissue**

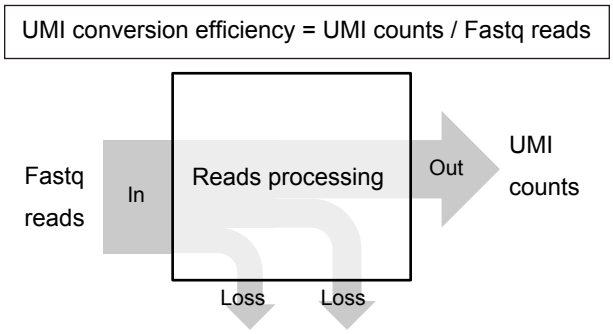
1334 a) Morphology of SVF cells. Adipose tissue from a cell suspension of SVF was prepared. Upper
1335 panels present a photograph of adipose tissues and dissociated SVF samples. Yellow scale bar
1336 represents 1 cm. White scale bar represents 10 μm. Lower panels represent the distribution of
1337 cell size information with different platforms (left: diameter of cell size using photography, right:
1338 flow cytometry information using a cell sorter). The diameter of cell size for SVF samples was
1339 6.43 ± 1.35 μm (n=200). b) Clustering of cells included in SVF. The transcriptome of
1340 approximately 1,000 cells was quantified by Quartz-Seq2 and clustering on t-SNE space was
1341 performed. In accordance with the genes and functional terms enriched in each cluster, the cell
1342 type was annotated. The percentage indicates the proportion of cells for each cluster relative to
1343 all cells analyzed. Numbers in parentheses indicate the numbers of cells constituting the cluster.
1344 c) Marker genes for each cluster were identified by Quart-Seq2. Cluster-specific or
1345 cluster-enriched genes were calculated for each cluster, with their expression being displayed as
1346 color in a heatmap. No more than 50 cells are shown for simplicity. d) The results of GO-PCA
1347 analysis. Functional terms enriched in the genes with high factor loadings of PCA were
1348 calculated and the enrichment is displayed as color in the heatmap. No more than 50 cells are
1349 shown for simplicity. e) Reactome pathway with genes differentially expressed between cluster 1
1350 and cluster 8.


1. Single-cell collection (Cell sorter)
1 cell
single-cell sorting
Flow-cytometry data
Link
1 cell transcriptome
Quartz-Seq2
384 well plate
2. Cell barcoding
AAAAA
TTTTT MB CB0001 caM
TTTTT MB CB0002 caM
TTTTT MB CB0003 caM
TTTTT MB CB1536 caM
Cell barcodes
14mer, 1-384 SeqLv(5)
15mer, 1-1536 SeqLv(5)
3. Spindown collection
384 PCR plate
Paraffin film
3,010g
Metal frame
1 well reservoir
Assembled collector
4. Whole transcript amplification based on poly-A tagging
TTTTT MB CBxxxx caM
cDNA purification
Poly-A tailing by TdT enzyme (Improved)
Tagging and 2nd strand synthesis (Improved)
Amplified by suppression PCR
Purification of amplified cDNA
5. Sequence library preparation
gM TTTTT
gM AAAAA
MB CBxxxx caMg
Fragmentation by ultrasonicator
Adaptor ligation
P7 PB
Primer anealing
P5
PCR enrichment
Sequencing (NextSeq500, HiSeq)
Read2
Index1
Read1 (Custom sequence primer)

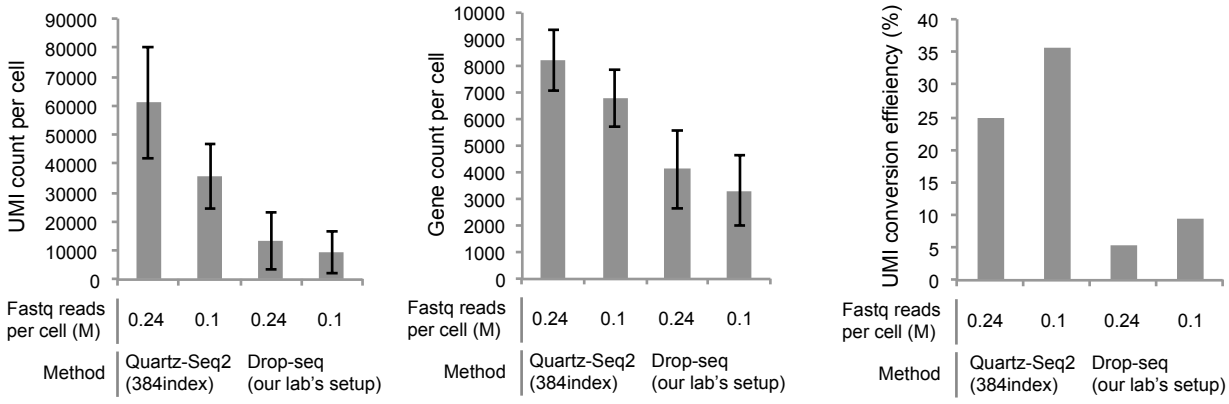

a)
Byproducts (BA: 50-300bp)
cDNA (BA: 300-9,000bp)
Relative DNA yield (n=3)
p=0.0002
p=0.0053
p=0.038
p=0.00011
p=0.00002
TdT Buffer
QuartzB
NEB2
T55
RH55
QuartzB
NEB2
T55
RH55
Tagging condition
Increment (-)
Increment (+)
b)
qPCR score from amplified cDNA (n=3)
log2(copy number)
QuartzB: y = 0.7147x + 5.1551, R = 0.8884, SCC 0.662
T55+Inc: y = 0.8881x + 5.7883, R = 0.9389, SCC 0.7986
RH55+Inc: y = 0.8567x + 5.7894, R = 0.9342, SCC 0.7934
qPCR score from non-amplified cDNA (averaged)
log2(copy number)
c)
Total number of UMI count per 10 pg mouse total RNA
Quartz-Seq like with UMI
Quartz-Seq2 (RT100)
Total number of detected genes
Quartz-Seq like with UMI
Quartz-Seq2 (RT100)
Cumulative number of detected genes
3
Quartz-Seq like with UMI
Quartz-Seq2 (RT100)
Gene expression level
d)
Quartz-Seq like with UMI
SCC 0.83
Averaged UMI count (Scaled)
Bulk polyA-RNA-seq (TPM + 1)
Quartz-Seq2 (RT100)
SCC 0.85
Averaged UMI count (Scaled)
Bulk polyA-RNA-seq (TPM + 1)
e)
Typical: Quartz-Seq2 (RT100)
Before UMI filtering
After UMI filtering
CV²
Poisson
mean
CV²
Poisson
mean


a)
RT condition
Primer version
Lysis solution
RT solution
cDNA Purification
TdT solution
PCR solution
PCR Purification
LDprep
QC
Labware
RT100
Quartz-Seq2 (384 Indexs)
RT25
Quartz-Seq2 (384 Indexs)
Quartz-Seq2 (1536 Indexs)
Relative cost for respective experimental processes (%)
b)
p=0.0489
cDNA yield from 384 plate (ng)
RT100
RT25
Quartz-Seq2 (384 indexs)
c)
UMI count
Initial fastq reads per cell (M)
p=0.0085
Gene count
p=0.0011
ERCC capture efficiency (%)
p=0.0024
Technical variablity of UMI count (CV)
p=0.0013
Technical variablity of gene count (CV)
p=0.0036
RT100_02
RT100_03
RT100_04
RT25_01
RT25_02
RT25_03
d)
Initial fastq reads per cell (M)
Cell number per NextSeq500 run
Shallow read depth for a single-cell
0.27M
0.19M
0.1M
0.05M
e)
ERCC capture efficiency: 17.08%
0.27M fastq reads per cell
ERCC capture efficiency: 11.95%
0.096M fastq reads per cell
100%
10%
1%
0.1%
reg
data

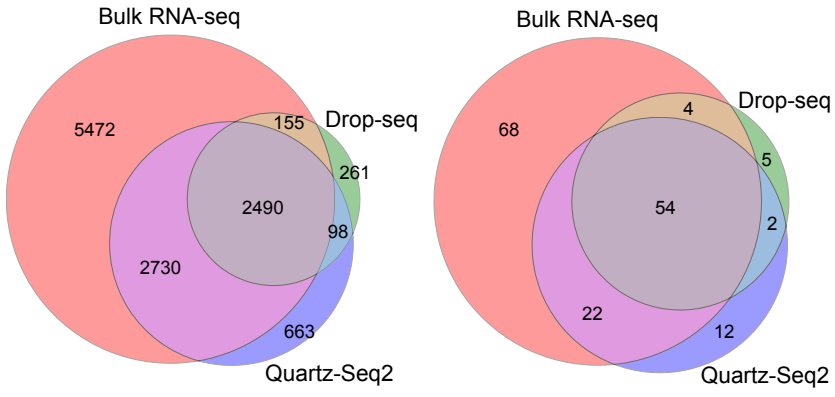
a)


UMI conversion efficiency = UMI counts / Fastq reads
Fastq reads
In
Reads processing
Out
UMI counts
Loss
Loss

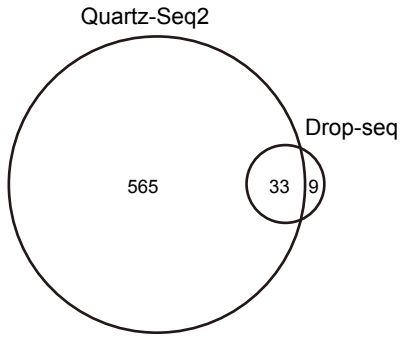
b)

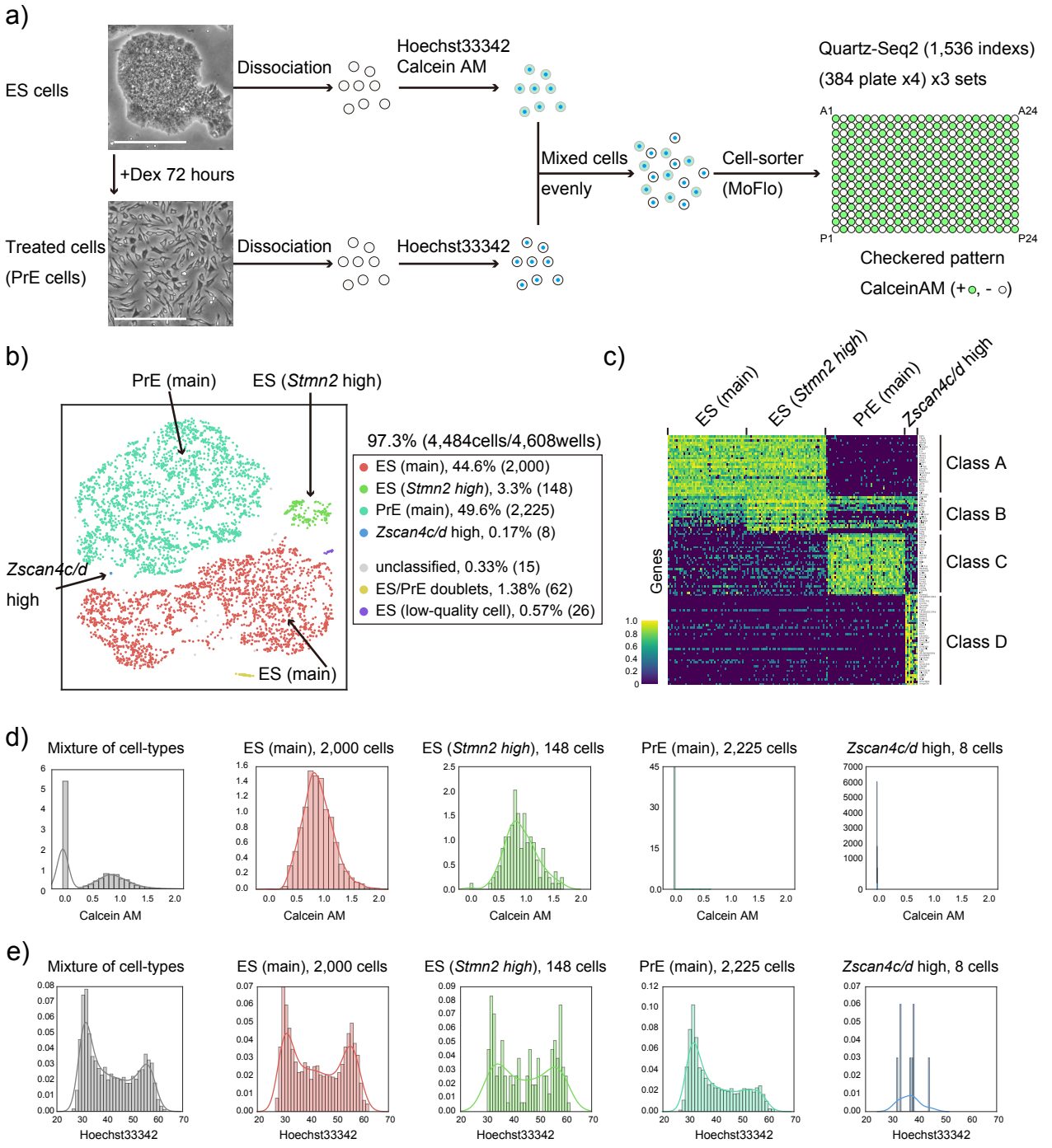

UMI count per cell
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
Gene count per cell
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
UMI conversion effieiency (%)
40
35
30
25
20
15
10
5
0
Fastq reads per cell (M)
0.24
0.1
0.24
0.1
Method
Quartz-Seq2 (384index)
Drop-seq (our lab's setup)

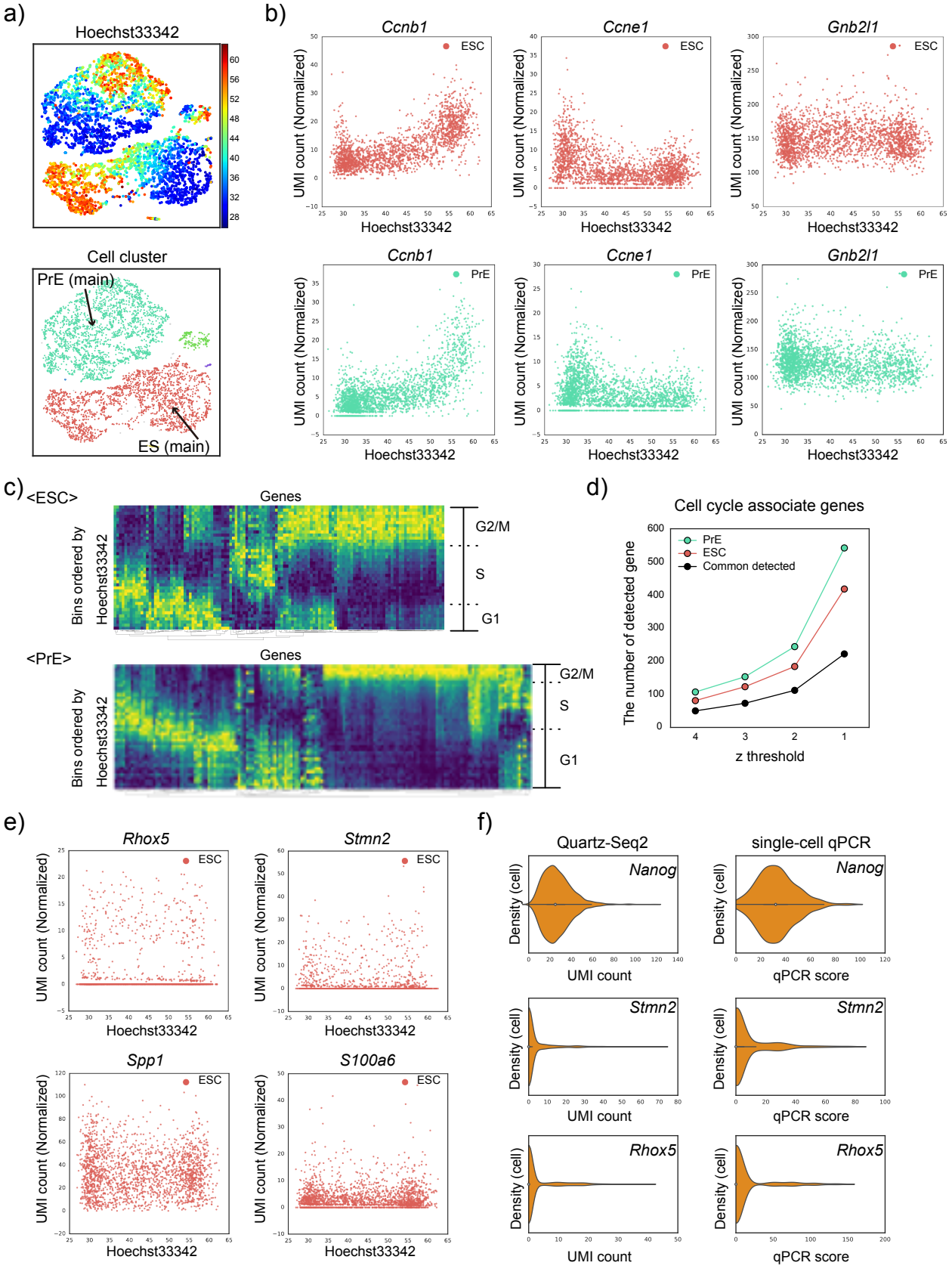
c)

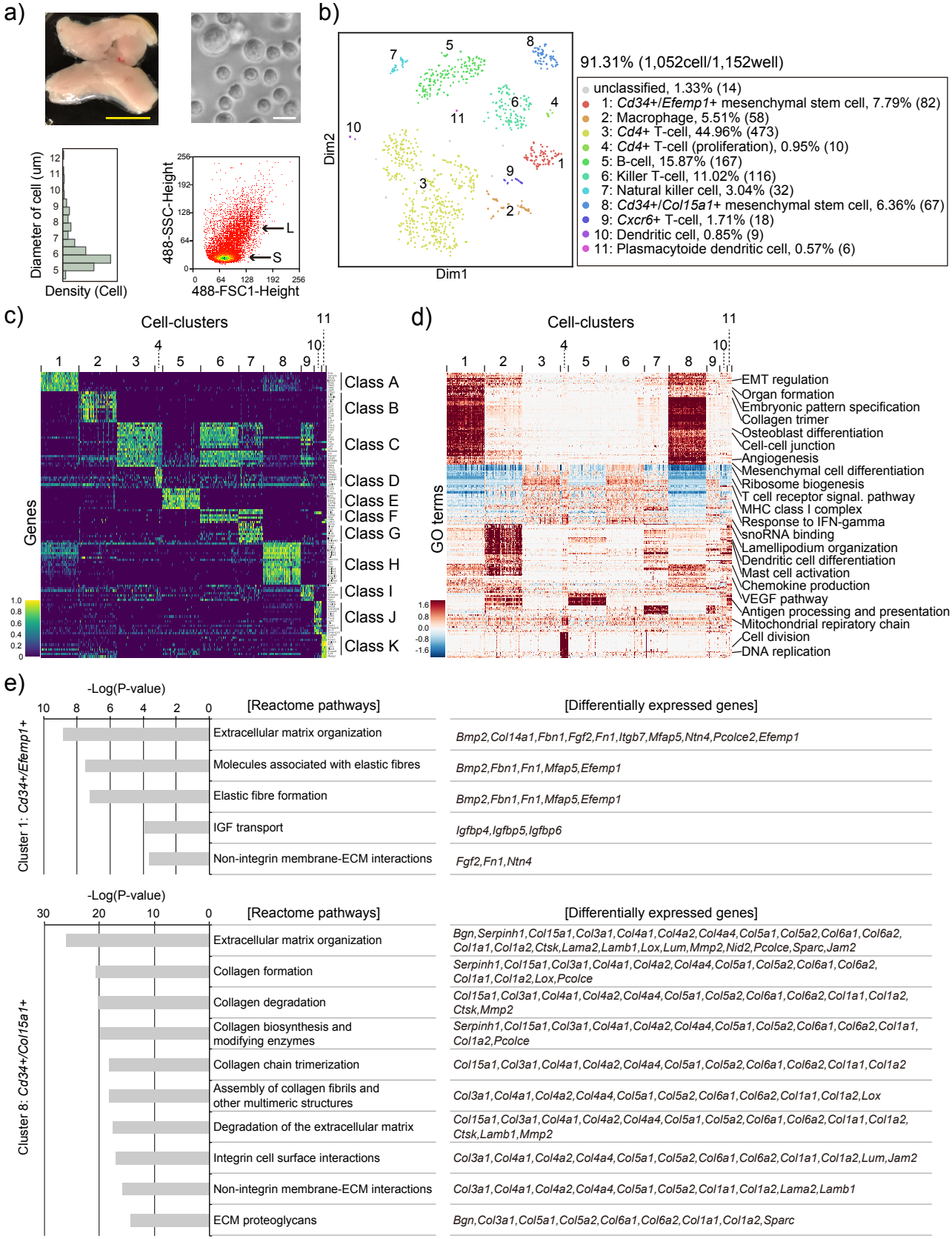

Differentially expressed genes
Bulk RNA-seq
Drop-seq
Quartz-Seq2
5472
155
261
2490
98
2730
663
Reactome pathway
Bulk RNA-seq
Drop-seq
Quartz-Seq2
68
4
5
54
2
22
12

d)


GO terms
on highly variable genes
Quartz-Seq2
Drop-seq
565
33
9


a)
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 indexs)
(384 plate x4) x3 sets
A1
A24
P1
P24
Checkered pattern
CalceinAM (+, -)
b)
PrE (main)
ES (Stmn2 high)
Zscan4c/d
high
ES (main)
97.3% (4,484cells/4,608wells)
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
Bins ordered by Hoechst33342
G2/M
S
G1
d)
Cell cycle associate 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% (1,052cell/1,152well)
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
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