Genome Biology

Quartz-Seq: a highly reproducible and sensitive single-cell RNA-Seq reveals non-genetic gene expression heterogeneity

Genome Biology 2013, **14**:R31 doi:10.1186/gb-2013-14-4-r31

Yohei Sasagawa (sasagawayohei@gmail.com)
Itoshi Nikaido (dritoshi@gmail.com)
Tetsutaro Hayashi (t-hayashi88@cdb.riken.jp)
Hiroki Danno (hdanno@cdb.riken.jp)
Kenichiro D Uno (uno-ken@cdb.riken.jp)
Takeshi Imai (imai@cdb.riken.jp)
Hiroki R Ueda (uedah-tky@umin.ac.jp)

ISSN 1465-6906

Article type Method

Submission date 21 December 2012

Acceptance date 11 April 2013

Publication date 17 April 2013

Article URL http://genomebiology.com/2013/14/4/R31

Quartz-Seq: a highly reproducible and sensitive single-cell RNA-Seq reveals non-genetic gene expression heterogeneity

Yohei Sasagawa[1,7,8], Itoshi Nikaido[1,7,8], Tetsutaro Hayashi[2], Hiroki Danno[3], Kenichiro D Uno[1], Takeshi Imai[4,5] and Hiroki R Ueda[1,3,6,*]

[1]Functional Genomics Unit, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
[2]Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
[3]Laboratory for Systems biology, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
[4]Laboratory for Sensory Circuit Formation, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
[5]JST, PRESTO, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
[6]Laboratory for Synthetic Biology, Quantitative Biology Center, RIKEN, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
[7]These authors contributed equally to this work
[8] Present address: Bioinformatics Research Unit, Advanced Center for Computing and Communication, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.
*Corresponding author

Yohei Sasagawa: sasagawayohei@gmail.com
Itoshi Nikaido: dritoshi@gmail.com
Tetsutaro Hayashi: t-hayashi88@cdb.riken.jp
Hiroki Danno: hdanno@cdb.riken.jp
Kenichiro D. Uno: uno-ken@cdb.riken.jp
Takeshi Imai: imai@cdb.riken.jp
Hiroki R. Ueda: uedah-tky@umin.ac.jp

Abstract

A highly reproducible and sensitive single-cell RNA-Seq method will facilitate the understanding of the biological roles and underlying mechanisms of the non-genetic cellular heterogeneity. In this study, we report a novel single-cell RNA-Seq method called Quartz-Seq that has a simpler protocol and higher reproducibility and sensitivity compared to previously developed methods. We demonstrate that single-cell Quartz-Seq quantitatively detects various kind of non-genetic cellular heterogeneity. The method detects different cell types and different cell-cycle phases of a single-cell type. Moreover, this method can comprehensively reveal gene expression heterogeneity between single-cells of the same cell type at the same cell-cycle phase.

Keywords

single-cell, RNA-Seq, transcriptome, sequencing, bioinformatics, cellular heterogeneity, cell biology

Background

Non-genetic cellular heterogeneity at the mRNA and protein levels has been observed within cell populations in diverse developmental processes and physiological conditions [1-4]. However, the comprehensive and quantitative analysis of this cellular heterogeneity and its changes in response to perturbations has been extremely challenging. Recently, several researchers reported the quantification of gene expression heterogeneity within genetically identical cell populations as well as its biological roles and underlying mechanisms [5-8]. Although gene expression heterogeneities have been quantitatively measured for several target genes using single-molecule imaging or single-cell qPCR, comprehensive studies on the quantification of gene expression heterogeneity are limited [9] and are thus being developed. Because global gene expression heterogeneity may contain biological information (e.g., cell fate, culture environment and drug response), the question of how to comprehensively and quantitatively detect the heterogeneity of mRNA expression in single cells and extract biological information from those data remains to be addressed.

Single-cell RNA-Seq analysis has been demonstrated to be an effective approach for the comprehensive quantification of the gene expression heterogeneity that reflects the cellular heterogeneity at the single-cell level [10, 11]. To understand its biological roles and underlying mechanisms, an ideal single-cell transcriptome analysis would provide a simple, highly reproducible and sensitive method for measuring the gene expression heterogeneity of cell populations. In addition, this method should significantly distinguish the gene expression heterogeneity from experimental errors.

Single-cell transcriptome analyses, which have been achieved through the use of various platforms, such as microarrays, massively parallel sequencers and bead arrays [12-17], are able to identify cell-type markers and/or rare cell types in tissues. These platforms require nanogram quantities of DNA as the starting material. However, a typical single cell has approximately 10 pg of total RNA and often contains only 0.1 pg of polyadenylated RNA. To obtain the amount of DNA starting material that is required by these platforms, it is necessary to perform whole-transcript amplification (WTA).

Previous WTA methods for single cells fall into two categories, according to the modifications that are introduced into the first-strand cDNAs in the PCR-based methods. One approach is based on the poly-A tailing reaction, and the other method type is based on the template-switching reaction. In principle, the goal of poly-A tailing is to obtain both full-length first-strand cDNAs and truncated cDNAs. The aim of template switching is to obtain first-strand cDNAs that have reached the 5’ ends of the RNA templates. These modified cDNAs are amplifiable by subsequent PCR enrichment methods.

Kurimoto et al. reported a quantitative whole-transcript amplification method based on the poly-A tailing reaction for single-cell microarrays [12]. Using this single-cell transcriptome analysis, these authors published initial validation data for technical replicates, each of which utilized 10 pg of total RNA. The Pearson correlation for the reproducibility of this method using 10 pg of total RNA per reaction was approximately 0.85 [12]. Tang et al. performed single-cell RNA-Seq using a method similar to the one used by Kurimoto et al. When Tang et al. applied their method to a single mouse oocyte (~1 ng total RNA), these researchers were capable of detecting a greater number of genes than those identified through the microarray approach [13]. However, these methods are complicated because they require multiple PCR tubes for a single

cell and gel purification for the removal of unexpected byproducts [18, 19]. Furthermore, the detailed quantitative performance of Tang et al.'s single-cell RNA-Seq method, e.g., its reproducibility and sensitivity, has not been analyzed.

Two single-cell RNA-Seq methods based on the template-switching reaction have been reported. Islam et al. described STRT-Seq, which is a highly multiplexed single-cell RNA-Seq that can detect the restricted 5' ends of mRNAs [14]. In addition, Ramsköld et al. developed Smart-Seq, which exhibits a greater read coverage across transcripts than previously developed methods [16]. The Pearson correlations for the reproducibility of the methods developed by Islam et al. and Ramsköld et al. using 10 pg of total RNA were approximately 0.7. Recently, Hashimshony et al. described CEL-Seq, which is an IVT (In vitro transcription)-based method but not a PCR-based method. CEL-Seq is a highly multiplexed single-cell RNA-Seq that can detect the 3' end of mRNA [17]. CEL-Seq detected significantly more genes in mouse ES single-cell compared with STRT-Seq. The performance of these reported methods is sufficient for the identification of cell-type markers. However, their specifications for WTA did not validate whether the methods were sufficient to quantitatively assess the global gene expression heterogeneity that is indicative of cellular heterogeneity. Because the Pearson correlation for reproducibility is greater than 0.95 in conventional non-WTA RNA-Seq, it is desirable to further improve the reproducibility and sensitivity of the previously developed single-cell RNA-Seq methods.

To comprehensively and quantitatively detect gene expression heterogeneity, we developed a simple and highly quantitative single-cell RNA-Seq approach (Quartz-Seq). Through this study, we identified some defective factors that are responsible for simplifying the experimental procedures and improving the quantitative performance. In particular, to maintain the simplicity and enhance the quantitative performance of WTA, we improved 3 critical points: 1) we achieved a robust suppression of the synthesis of byproducts, 2) we identified a robust PCR enzyme that allows the use of a single-tube reaction, and 3) we determined the optimal conditions of reverse transcription and second-strand synthesis for the capturing mRNA and the first-strand cDNA. We also performed a quantitative comparison between our method and previously developed methods using 10 pg of total RNA as the starting material. The reproducibility and sensitivity of the Quartz-Seq method was improved compared to those of the previously developed methods.

When used in the global expression analysis of real single cells, the single-cell Quartz-Seq approach successfully detected gene expression heterogeneity even between cells of the same cell type and at the same cell-cycle phase. This observed gene expression heterogeneity was found to be highly reproducible in two independent experiments and can be distinguished from experimental errors, which were measured through technical replicates of pooled samples. We also showed that single-cell Quartz-Seq is able to more easily discriminate between different cell types and/or between different cell-cycle phases. Therefore, single-cell Quartz-Seq is a useful method for the comprehensive identification and quantitative assessment of cellular heterogeneity.

Results

Whole-transcript amplification for single-cell Quartz-Seq and Quartz-Chip

The whole-transcript amplification for Quartz-Seq and Quartz-Chip consists of five main steps (Figure 1). The first step is a reverse transcription with an RT primer to generate the first-strand cDNAs from the target RNAs. The second step is a primer digestion with exonuclease I; this is one of the key steps that prevent the synthesis of byproducts. The third step is the addition of a poly-A tail to the 3' ends of the first-strand cDNAs. The fourth step is the second-strand synthesis using a Tagging primer, which prepares the substrate for subsequent amplification. The fifth step is a PCR enrichment reaction with a suppression PCR primer to ensure that a sufficient quantity of DNA is obtained for the massively parallel sequencers or microarrays. All of the steps are completed in a single PCR tube without any purification. The amplified cDNA contains WTA adaptor sequences from the RT primer and the Tagging primer.

The amplified cDNA was then used in an Illumina sequencer (Quartz-Seq) and a microarray (Quartz-Chip) system. For the Quartz-Seq, the amplified cDNA was fragmented using the Covaris shearing system. The fragmented cDNA was ligated to Illumina adaptors, which enable the multiplex production of paired-end sequences. The DNA sequencing library was analyzed using an Illumina sequencer. For the Quartz-Chip, we synthesized labeled cRNA from the amplified cDNA through in vitro transcription. The labeled cRNA was used for the microarray analysis.

Performance improvements of whole-transcript amplification

In previous WTA methods based on the poly-A tailing reaction, an excessive amount of byproducts are produced (see Additional file 1, Figure S1). These byproducts (approximately < 200 bp DNA) are derived from the reverse transcription primer. The reverse transcription primer was modified by terminal deoxynucleotidyl transferase similarly to the first-strand cDNA. The modified reverse transcription primer causes the synthesis of the byproducts [18]. The amplified byproducts should then be removed by gel purification [18, 19] (see Additional file 2, Supplementary Note). This gel purification step for the removal of the byproducts increases the complexity of the method. The byproducts contain WTA adaptor sequences. We found that the byproducts cause early saturation of the PCR amplification and reduce the molar ratio between the objective cDNA and the byproducts. The contamination rate from the WTA adaptor sequence was dramatically increased in the results obtained through the Illumina sequencing method (see Additional file 1, Figure S2e and f).

To overcome the contamination of the byproducts, the byproduct syntheses were completely eliminated through the combination of an exonuclease I treatment, a restricted poly-A tailing and an optimized suppression PCR (see Additional file 1, Figure S3 and Figure S4). We successfully eliminated the synthesis of byproducts using three parameters and eliminated the need for gel purification.

The first parameter is the adjustment of the RT primer concentration after the reverse transcription step. We used the minimum primer concentration for the reverse transcription. Moreover, we removed the reverse transcription primer through the treatment with exonuclease I, which digests single-strand DNAs, such as primers. This exonuclease I digestion suppressed the synthesis of byproducts (see Additional file 1, Figure S2a). However, the primer removal was not complete, which is in agreement with the results of a previous study ([18]. see Additional file

1, Figure S1). The molar ratio between the single-cell-level mRNA (0.1 pg, Ensembl Mouse Transcript, 1817 bp average size) and the reverse transcription primer is greater than 190,000. A complete removal solely by exonuclease I digestion was difficult. The remaining primers were then modified by terminal deoxynucleotidyl transferase, similarly to the first-strand cDNA; these modified primers caused the production of byproducts.

To prevent the amplification of the modified primers, we used suppression PCR technology. Suppression PCR is very effective in the suppression of the amplification of small-size DNA that contains complementary sequence at both ends of the template DNA [20]. In suppression PCR, these complementary sequences can bind each other. The self-bound template DNA forms a “pan-like” structure. In addition, the DNA is not amplified by PCR because the PCR primer cannot bind to the template DNA (see Additional file 1, Figure S4). The target DNA size of suppression PCR depends on the end of the complementary sequences (the length, GC content, and so on). We also identified a good primer sequence for the suppression PCR. The B-primer effectively suppressed the synthesis of byproducts (see Additional file 1, Figure S2c and Figure S4b).

To shorten the length of the remaining primers modified by terminal transferase and thus make them targets of suppression PCR, we restricted the reaction time of the terminal transferase. This restriction suppressed the synthesis of byproducts (see Additional file 1, Figure S2b).

In addition, we found that topoisomerase V could suppress byproduct synthesis (see Additional file 1, Figure S2d). However, the mechanism of byproduct suppression by topoisomerase V is not known. Through the combination of the three previously discussed parameters, we successfully suppressed the synthesis of byproducts completely in a single-tube reaction (see Additional file 1, Figure S2e); this suppression was achieved without topoisomerase V.

Furthermore, we selected a robust PCR enzyme that was optimal for the single-tube reaction. The use of the MightyAmp DNA polymerase (also marketed as Terra PCR Direct Polymerase) improved the yield of cDNA (see Additional file 1, Figure S2c and Figure S5) and the reproducibility of the WTA replication (see Additional file 1, Figure S2d and Figure S5). In addition, by employing the MightyAmp DNA polymerase, the number of PCR cycles in WTA was decreased.

Moreover, we improved the efficiencies of the reverse transcription and second-strand synthesis steps to counter the lowered reproducibility of the WTA that occurred due to the variable efficiencies of these steps. We found an optimal annealing temperature that reduced the variability in the efficiencies of these steps (see Additional file 1, Figure S2a and b). Our simplified method enabled us to consistently obtain highly reproducible cDNA that was optimized for RNA-Seq (see Additional file 1, Figure S6).

Reproducibility and sensitivity of single-cell Quartz-Seq

We performed single-cell Quartz-Seq with 10 pg of diluted total ES-cell (embryonic stem cell) RNA to validate the reproducibility of technical replicates. We prepared a multiplex, paired-end DNA sequencing library from the amplified cDNA produced from the 10 pg of total RNA. The DNA sequencing library was analyzed using an Illumina HiSeq 1000/2000 sequencer.

Pairwise comparisons of the products of triplicate amplifications were used to quantify the reproducibility of the protocol based on the Pearson correlation coefficients (PCCs, Figure 2a) in log10-transform FPKM values. The PCCs of these comparisons were approximately 0.93. We counted highly reproducible expressed transcripts that were larger than 1.0 in FPKMs (fragments per kilobase of transcript per million fragments sequenced) and exhibited less than two-fold expression changes between technical replicates. The single-cell Quartz-Seq method was capable of reproducibly detecting 8,110.3 (± 100.8) of 9,872.6 (± 54.4) transcripts (82.1 ± 0.6 %, Figure 2b). All of the pairwise plots are presented in Additional file 3, Figure S7. All result of linear regression and correlation analysis are presented in Additional file 4, Table S1.

To evaluate the sensitivity of Quartz-Seq, we compared the results of conventional RNA-Seq (non-WTA) and Quartz-Seq method. The PCCs of these comparisons were approximately 0.89, Figure 2a). We also counted highly sensitive transcripts that had an FPKM value greater than 1 and exhibited less than two-fold expression changes between technical replicates. The single-cell Quartz-Seq method was capable of sensitively detecting 6,605 (± 139.9) of 9,686 (± 63.7) transcripts (68.1 ± 0.5 %, Figure 2b).

To evaluate the over-representation of the sequences derived from the WTA and the library preparation, we searched for the WTA adaptor sequence in all of the sequence reads using sequence similarity (see details in Materials and methods). Using Smart-Seq, 21.1 ± 3.05 % of the sequences were identified as WTA adaptor (see Additional file 1, Figure S8, “Smart-Seq, ES, 10 pg (PE, 30 M), n = 4”) whereas 7.68 ± 0.66 % of the sequences were identified as WTA adaptors by Quartz-Seq (see Additional file 1, Figure S8, “Quartz-Seq, ES, 10 pg (PE, 60 M), n = 3”).

We also evaluated the number of reads required for the method to detect mRNAs. From 0.01, 0.1, 0.5 1.0, 5.0, 10, 30, and 45 million reads (uniquely mapped reads, single-ended, 50 bp), we counted the number of detected genes and calculated the Pearson correlation coefficients between different samples of the same origin (10 pg of total RNA). We found that a Quartz-Seq result from more than 1 million reads showed correlations of greater than 0.9 and detected more than 7,642 ± 40 (73 ± 0.19 %) transcripts (see Additional file 1, Figure S9).

Furthermore, we compared the cDNA lengths resulting from the Quartz-Seq and conventional RNA-Seq methods (see Additional file 1, Figure S10). The median of the read coverage across the expressed transcripts (FPKM ≥ 10) was 53.8 % (705 bp) when Quartz-Seq was used, whereas the median of the read coverage obtained using the conventional RNA-Seq approach was 84.8 % (1,326 bp).

Comparison between Quartz-Seq and other methods

We carefully compared the quantitative performance of Quartz-Seq/Chip and three reported methods for single-cell transcriptome analysis. To compare the reproducibility of Quartz-Seq and Smart-Seq, four Smart-Seq data sets from the first Smart-Seq paper published by Ramsköld et al. were downloaded from NCBI GEO (GSE38495). We calculated the Pearson correlation coefficient (PCC) between pairs of samples using 10 pg of total RNA (see Additional file 3 and Additional file 4, Figure S7 and Table S1). The read numbers for the Quartz-Seq and Smart-Seq methods were adjusted to approximately 30 million reads of single-end sequences of 50 bp. The PCC values for Quartz-Seq and Smart-Seq were approximately 0.93 and approximately 0.7, respectively (Figure 3a).

We counted highly reproducible transcripts that had an FPKM or RPKM (Reads Per Kilobase of exon Model per million mapped) value greater than 1 and exhibited less than two-fold expression changes between technical replicates. Quartz-Seq detected 8,110.3 (± 100.8) of 9,872.6 (± 54.4) transcripts (82.1 ± 0.6 %), whereas Smart-Seq achieved 2,745.5 (± 87.2) of 5,286.6 (± 92.0) transcripts (51.9 ± 0.8 %)(human brain), 2,312 of 4,202 transcripts (55.0 %)(mouse brain) and 3,706.0 (± 173.5) of 7,388.4 (± 256.2) transcripts (50.1 ± 1.5 %)(universal human reference RNA) detection. To confirm the experimental reproducibility, we performed additional Quartz-Seq and Smart-Seq procedures using 10 pg of diluted total RNA from an ES cell (n = 4). The read numbers for Quartz-Seq and Smart-Seq were adjusted to approximately 30 million reads of paired-end sequences of 50 bp. In this comparison, the PCC of Quartz-Seq was approximately 0.93, whereas the PCC of Smart-Seq was approximately 0.72 (Figure 3a). Quartz-Seq detected 7,739 (± 38.5) of 9,506.1 (± 22.7) transcripts (81.4 ± 0.5 %), whereas Smart-Seq achieved 2,320.6 (± 34.7) of 3675 (± 100.6) transcripts (63.1 ± 1.3 %) detection.

We also compared the sensitivity of Quartz-Seq and Smart-Seq. To evaluate the sensitivity of each method, we compared the results of conventional RNA-Seq (non-WTA) and each method from same pooled total RNA. The PCCs of these comparisons were approximately 0.88 in Quartz-Seq. On the other hand, the PCCs of these comparisons were approximately 0.7 in Smart-Seq (Figure 3b). We also counted highly sensitive transcripts that had an FPKM value greater than 1 and exhibited less than two-fold expression changes between technical replicates. The Smart-Seq with 10 pg ES total RNA was capable of detecting 2,906 (± 87.4) of 6,263 (± 149.1) transcripts (46.3 ± 0.5 %). Although, the Quartz-Seq with 10 pg ES total RNA was capable of sensitively detecting 6,191.2 (± 58.0) of 9,342 (± 134.7) transcripts (66.2 ± 0.6 %).

Next, we compared the quantitative performance between Quartz-Seq and CEL-Seq. CEL-Seq data sets from the original paper published by Hashimshony et al. were downloaded from NCBI SRA (SRP014672). We calculated the PCC between pairs of samples using *Caenorhabditis elegans* 10 pg of total RNA (see Additional file 3 and Additional file 4, Figure S7 and Table S1). The PCCs of these comparisons were approximately 0.72. We counted highly reproducible expressed transcripts that were larger than 1.0 in tpm (tag per million) and exhibited less than two-fold expression changes between technical replicates. CEL-Seq was capable of reproducibly detecting 2,564 (± 183.8) of 5,196.8 (± 364.9) transcripts (49.3 ± 1.7 %). Moreover, we reanalyzed CEL-Seq data with mouse ES cell. We counted highly reproducible expressed transcripts that were larger than 1.0 in tpm using the data. CEL-Seq detected 4,070.3 ± 332.4 transcripts in mouse ES single-cell (n = 9). On the other hand, Quartz-Seq detected 6,069.1 ± 854.9 transcripts in mouse ES single-cell (n = 35).

Subsequently, we compared the quantitative performances of our method and other method based on poly-A tailing reaction. The detailed quantitative performance of Tang et al.'s single-cell RNA-Seq method has not been analyzed using 10 pg total RNA. Therefore, we evaluated the performance (reproducibility and sensitivity) of the Quartz-Chip and Kurimoto et al. methods using a GeneChip with 10 pg of diluted total RNA. For this comparison, we reanalyzed the original data from Kurimoto et al. and compared the results with the Quartz-Chip data. We first compared the technical duplicates to quantify the reproducibility of both protocols (Figure 3c, upper panels). In this analysis, we counted a transcript that had RMA expression larger than 7.0 and exhibited less than two-fold expression changes between technical duplicates. The Quartz-

Chip method was capable of reproducibly detecting 7,520 of 9,622 transcripts (78.2 %) in the technical duplicates. However, the Kurimoto et al. method detected 5,224 of 6,976 transcripts (74.9 %) in the technical duplicates. In addition, 69.8 % of the transcripts detected by the Kurimoto method were detected by Quartz-Chip.

We then compared the results of conventional non-WTA GeneChip with the Quartz-Chip and the Kurimoto et al. methods to quantify the sensitivity of both protocols (Figure 3c, lower panels). Similarly to the reproducibility analysis, we counted a transcript that had an RMA expression larger than 7.0 and exhibited less than two-fold expression changes between the non-WTA GeneChip and the WTA samples (either the Quartz-Chip or the Kurimoto et al. methods). The Quartz-Chip method was capable of sensitively detecting 7,557 of 9,849 transcripts (77 %), whereas the Kurimoto et al. method detected 4,686 of 7,704 transcripts (60.8 %). In addition, 73.7 % of the transcripts detected by the Kurimoto et al. method were also detected by Quartz-Chip.

Limitations of Quartz-Seq

We investigated whether there are specific structures of transcripts that are noisier or under-represented. We calculated and compared the GC content and cDNA lengths of the amplified and unamplified isoforms obtained by each single-cell RNA-Seq method (see Additional file 1, Figure S11). As expected, we found that the unamplified isoforms from Quartz-Seq showed a higher GC content (average: 52.1 %) compared with the amplified isoforms from Quartz-Seq (average: 50.2%). Nevertheless, the unamplified isoforms from Quartz-Seq showed a higher GC content (average: 52.1 %) compared with those from Smart-Seq (average: 51.5%). In the analysis of cDNA lengths, we found that the unamplified (or non-detected) isoforms from Quartz-Seq showed a shorter cDNA length (average: 1,684.0 bp) compared with the amplified (or detected) isoforms from Quartz-Seq (average: 2558.6 bp).

We then represented the detailed comparison of technical noise between different polymerases. We performed Quartz-Seq with Ex Taq DNA polymerase instead of MightyAmp DNA polymerase because Ex Taq DNA polymerase is often used in previous methods (Kurimoto et al. and of Tang et al.) that were based on the poly-A tailing reaction. We calculated the GC contents and cDNA lengths of the amplified and unamplified isoforms (see Additional file 1, Figure S11). The unamplified isoforms from Quartz-Seq with MightyAmp showed a higher GC content (average: 52.1%) than those from Quartz-Seq with Ex Taq (average: 51.7%). The unamplified isoforms from Quartz-Seq with MightyAmp showed a shorter cDNA length (average: 1,684.0 bp) than those from Quartz-Seq with Ex Taq (average: 2,481.1 bp).

Single-cell Quartz-Seq detects different cell types

Heterogeneous cell populations, such as cultured cell lines and tissues, are composed of various types of single cells, which have different gene expression patterns. We therefore tested whether single-cell Quartz-Seq can distinguish between different cell types and whether it can detect the differentially expressed genes that are characteristic of each cell type. We performed single-cell Quartz-Seq with 12 mouse embryonic stem (ES) cells and 12 primitive endoderm (PrE) cells, which are directly differentiated from ES cells [21]. We collected single cells directly into PCR tubes using FACS sorting, as previously reported [22]. In this system, all sorted cells were readily discerned as single cells (see Additional file 1, Figure S13). We successfully

obtained amplification products from almost all of these single-cells (98 %, n = 50, see Additional file 1, Figure S13). The ES and PrE cells were collected during the G1 phase of the cell cycle (see Additional file 1, Figure S12). Although the cell population that combines cells from all cell-cycle phases contained an average of approximately 10 pg of total RNA per cell, the cells at the G1 phase contained only approximately 6 pg of total RNA per single cell (see Additional file 1, Figure S14).

We first performed a cluster analysis of all the transcripts from all of the samples. The global expression patterns of the ES and PrE cells were clearly divided into two clusters (Figure 4a). A heatmap of the ES and PrE marker genes and the non-differentially expressed genes is shown in Figure 4b. We detected 1,620 and 1,436 differentially expressed genes in the ES and PrE cells, respectively. These differentially expressed genes included the ES marker genes (e.g., *Nanog*, *Pou5f1*, and *Fgf4*) and the PrE marker genes (e.g., *Gata4*, *Gata6*, and *Dab2*). In addition, these marker genes were clearly differentially expressed between the ES and PrE cells. In contrast, the non-differentially expressed genes, such as *Gnb1 l* and *Eef1b2*, did not exhibit a differential expression pattern (Figure 4b).

We then validated the expression patterns of the ES and PrE marker genes and the non-differentially expressed genes using an amplification-free single-cell qPCR method. To avoid any amplification bias, we directly detected the gene expression from single cells without amplification (see details in Materials and methods). The results show expression patterns for the ES and PrE cell markers and the non-differentially expressed genes that were highly correlated with the single-cell RNA-Seq data. The gene expression levels of *Pou5f1* and *Zfp42* were dramatically decreased in the majority of the single PrE cells. However, the gene expression levels of *Pou5f1* and *Zfp42* remained high in a small number of single PrE cells (Figure 4c). This trend was observed in both the single-cell Quartz-Seq and the amplification-free single-cell qPCR methods.

Single-cell Quartz-Seq detects the different cell-cycle phases of a single cell type

In addition to the gene expression heterogeneity that is caused by different cell types, gene expression heterogeneity can also be derived from different cell-cycle phases. To investigate the performance limits of the single-cell Quartz-Seq method, we tested whether single-cell Quartz-Seq is able to distinguish the cell-cycle-dependent heterogeneity among ES cells.

We performed single-cell Quartz-Seq with ES cells from different cell-cycle phases (G1, S, and G2/M) and then used principal components analysis (PCA) to analyze the results. The single cells PrE and all of the single ES cells formed two clearly divided clusters, i.e., the ES and PrE clusters (see Additional file 1 and Additional file 5, Figure S15a and Supplementary movie 1). If the single PrE cells are excluded, the single ES cells from each cell-cycle phase formed three different clusters, although a few single cells from the G1 and S phases were close to the G2/M cluster (see Additional file 1 and Additional file 6, Figure S15b and Supplementary movie 2). As expected, the differences among the cell-cycle phases were smaller than the difference between the stem cells and the more-differentiated cells. Despite these smaller differences, the single-cell Quartz-Seq method is able to detect the different cell-cycle phases within a single cell type.

Single-cell Quartz-Seq reveals the gene expression fluctuations in a single cell type at the same cell-cycle phase

If the differences associated with the different cell types and cell-cycle phases are excluded, there still remains a small amount of heterogeneity due to the fluctuations of gene expression among single cells.

To test whether single-cell Quartz-Seq can detect these gene expression fluctuations, we calculated and plotted the standard deviations from two independently amplified sets of single-cell Quartz-Seq data (n = 12 and n = 8) of ES cells at the G1 phase (Figure 5a). The Pearson correlation coefficient of these standard deviations from two independent sets was approximately 0.85 (Figure 5a). We performed an F-test of the equality of the two Quartz-Seq data variances to identify reproducible gene expression fluctuations (FDR > 0.6). The variances of 17,064 gene expressions were reproducibly observed. These results suggest that the gene expression fluctuations observed by single-cell Quartz-Seq are highly reproducible and thus not due to experimental errors.

To further validate the observed gene expression fluctuations using an independent experimental method, we used amplification-free single-cell qPCR to assess the gene expression of 9 genes (Figure 5b), which were selected according to their gene expression levels. To assess both the gene expression fluctuations and the experimental errors, we prepared samples from single cells (for the assessment of the gene expression fluctuations) and single-cell-sized samples of pooled cells (for the assessment of the experimental errors). We expected that, if the gene expression fluctuations observed by single-cell Quartz-Seq were not solely due to experimental errors, the gene expression fluctuations detected by the single-cell qPCR of the single-cell samples would be larger than the experimental errors detected from the pooled samples. As expected, we observed that the gene expression fluctuations detected by single-cell qPCR for the single-cell samples was, in fact, significantly larger than the experimental errors detected from the pooled samples (F-test, $p < 0.001$, Figure 5b). This result indicates single-cell qPCR clearly distinguish the gene expression fluctuations from the experimental errors.

If the single-cell Quartz-Seq method quantitatively detected the gene expression fluctuations, we can also expect that the gene expression fluctuations would be highly correlated between the single-cell Quartz-Seq and the qPCR methods. To confirm this, we compared the single-cell Quartz-Seq data with the single-cell qPCR data. To compare that two different platforms (single-cell Quartz-Seq and qPCR), we used a relative measure, such as the coefficient of variation (CV), rather than an absolute measure, such as the standard deviation. As expected, we found that the CVs detected by the single-cell Quartz-Seq approach are highly correlated with those obtained with single-cell qPCR (Figure 5c, Pearson correlation for the CVs was 0.992), which suggests that the single-cell Quartz-Seq method quantitatively detects gene expression fluctuations.

To observe functional features of fluctuated genes, we performed overrepresentation analysis of Gene ontology and Reactome pathway database for single-cell Quartz-Seq with 20 mouse ES cells in G1 phase (see Additional file 1, Figure S16). First of all, we performed clustering using by principle component analysis to collect groups of similar fluctuated genes. Each principle component was calculated hypergeometric test with Gene ontology and Reactome pathway database. We found that “Chromosome Maintenance”, “G1/S−Specific

Transcription”, and “RNA Polymerase II Transcription” pathways are observed significantly overrepresentation in principle component 1.

Discussion

In this study, we established a novel whole-transcript amplification method that was optimized for single-cell RNA-Seq and detects gene expression heterogeneity among individual cells. This whole-transcript amplification method for single-cell Quartz-Seq is substantially easier to perform than other previously developed methods that are based on the poly-A tailing reaction [18, 19] (see Additional file 1, Figure S1). For example, the Kurimoto et al. method requires approximately 17 PCR tubes and 11 reaction steps for a single cell [18].The single-cell Quartz-Seq amplification method, however, requires only one PCR tube and 6 reaction steps per single cell: all of the steps are completed in a single PCR tube without any purification. These improvements, which drastically simplify the single-cell Quartz-Seq method, will be useful for the high-throughput production of single-cell preparations.

In addition to its simplicity, the single-cell Quartz-Seq method exhibited a highly quantitative performance (Figure 2). We validated the performance of single-cell Quartz-Seq using 10 pg samples of purified total RNA prepared from pooled cell populations. The quantitative performance of single-cell Quartz-Seq was better than that of previously developed single-cell methods (Figure 3). Moreover, Quartz-Seq was useful for the analysis of cell subpopulations (50 cells containing 300-350 pg of total RNA) with a highly quantitative performance (R = 0.99, see Additional file 1, Figure S17).

Any method based on PCR amplification would have difficulty amplifying transcripts with an extremely high GC content, which would thus be underrepresented in the Quartz-Seq. We performed a detailed comparison of technical noise between different polymerases. However, Quartz-Seq is more robust against a high GC content because MightyAmp DNA polymerase is preferable for the amplification of GC-rich sequences compared with Ex Taq DNA polymerase.

In the analysis of cDNA lengths in each method, we found that the unamplified (or undetected) isoforms from Quartz-Seq showed a shorter cDNA length (average: 1,684.0 bp) compared with the amplified (or detected) isoforms from Quartz-Seq (average: 2558.6 bp). This is counter-intuitive but can be explained by the principle of next-generation sequencing with WTA, in which a longer cDNA generates more reads and therefore can be detected more sensitively in comparison with a shorter cDNA (see Additional file 1, Figure S11d). The non-detected isoforms from Quartz-Seq also showed a shorter cDNA length (average: 1,684.0 bp) compared with that of all of the Ensembl Mouse Transcripts (average: 1,817 bp). The unamplified isoforms from Quartz-Seq showed a significantly shorter cDNA length (average: 1,684.0 bp) compared with those from Smart-Seq (average: 2,382.0 bp), suggesting that Quartz-Seq is more robust against a shorter cDNA length. We also found that Smart-Seq could not amplify 3,924 (SD: 124.5) isoforms, whereas Quartz-Seq could not amplify only 1,614 (SD: 88.9) isoforms.

As a result of its higher reproducibility and sensitivity, single-cell Quartz-Seq can distinguish not only different cell types but also different cell-cycle phases of the same cell type. In addition, this method can also comprehensively detect gene expression fluctuations within the same cell type and cell-cycle phase; these fluctuations were highly reproducible in two independent experiments (Figure 5a, c) and were distinguished from experimental errors measured from technical replicates of pooled samples (Figure 5b). Therefore, our method is capable of comprehensively and quantitatively revealing gene expression fluctuations. These gene expression fluctuations can be generated by both the intrinsic stochastic nature of gene

expression and the extrinsic environmental differences among cells [5-8]. In fact, it has been reported that individual cells in a population of ES cells exhibit mRNA and protein expression fluctuations under the same culture conditions (e.g., as reported for *Nanog*, *Zfp42 Whsc2, Rhox9, and Zscan4*) that might be associated with different cellular phenotypes [1, 23-25]. Hence, the single-cell Quartz-Seq approach will be useful for the analysis of the roles and mechanisms of non-genetic cellular heterogeneity.

Conclusions

Single-cell Quartz-Seq is a simplified protocol relative to previously established methods based on the poly-A tailing reaction. All of the steps are completed in a single PCR tube without any purification. The reproducibility and sensitivity of Quartz-Seq were improved relative to other single-cell RNA-Seq methods. The use of Quartz-Seq in technical replicates with 10 pg each of total RNA produced a Pearson correlation coefficient of approximately 0.93. The reproducibility of previous methods is approximately 0.7. To evaluate the sensitivity of Quartz-Seq, we compared the performance of conventional RNA-Seq and single-cell RNA-seq methods with 10 pg total RNA. The Pearson correlation coefficient is approximately 0.88 in the Quartz-Seq as against approximately 0.7 in other methods. When used in the global expression analysis of real single cells, the single-cell Quartz-Seq approach successfully detected gene expression heterogeneity even between cells of the same cell type and/or between different cell-cycle phases. This observed gene expression heterogeneity was found to be highly reproducible in two independent experiments and can be distinguished from experimental errors, which were measured through technical replicates of pooled samples. Therefore, single-cell Quartz-Seq is a useful method for the comprehensive identification and quantitative assessment of cellular heterogeneity.

Materials and methods

Cell culture

We used EB5 ES cells for the preparation of the total RNA. This cell line is derived from E14tg2a ES cells in which a blasticidin resistance gene disrupts one endogenous *Pou5f1* allele. We used 5G6GR ES cells for the single-cell Quartz-Seq. This cell line was generated through the random integration of the linearized Gata6-GR-IRES-Puro vector into EB5 ES cells [21]. These cells were cultured on gelatin-coated dishes, in the absence of feeder cells and in Glasgow minimal essential medium (GMEM, Sigma) supplemented with 10 % fetal calf serum, 1000 U/ml Leukemia Inhibitory Factor (ESGRO, Invitrogen), 100 μM 2-mercaptoethanol (Nacalai Tesque), 1X non-essential amino acids (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 2 mM L-glutamine (Nacalai Tesque), 0.5X penicillin/streptomycin (Invitrogen), and 10 μg/ml blasticidin (Invitrogen). For the culture of the 5G6GR ES cells, 0.5 μg/ml puromycin (Sigma) was also added to the culture. For the differentiation of the 5G6GR ES cells into PrE cells, the cells were seeded into medium supplemented with 100 mM dexamethasone instead of blasticidin and cultured for 72 h. After this 72-h culture, the 5G6GR cells had completely differentiated into PrE cells.

RNA preparation

The total RNA was purified from the ES cells (EB5 cell line) using TRIzol (Life Technologies) and the RNeasy Mini kit (Qiagen). The amount of total RNA from an ES cell was quantified using an ND-1000 absorptiometer (LMS). The length distribution of the total RNA was measured using a bioanalyzer nano kit (Agilent), which produced an RNA integrity number of 10 for the total RNA. The spike RNAs were synthesized using the pGIBS-LYS, pGIBS-DAP, pGIBS-PHE, and pGIBS-THR plasmids (ATCC) and the MEGAscript T3 kit (Ambion), as previously reported (Kurimoto et al. 2006). The spike RNAs were added to the total RNA from the ES cells as follows (per 10 pg of total RNA): *Lys*, 1000 copies; *Dap*, 100 copies; *Phe*, 20 copies; and *Thr*, 5 copies. The total RNA containing the spike RNAs was diluted to 25 pg/μl using single-cell lysis buffer (0.5 % NP40, Thermo) immediately prior to its amplification.

Single-cell collection using FACS

The cultured cells were dissociated using trypsin-EDTA at 37 °C for 3 min. The resulting cells were subsequently washed with PBS buffer. A total of 0.5×10^6 cells were stained with 1 ml of PBS buffer containing 10 μg/ml Hoechst 33342 at 37 °C for 15 min. The stained cells were sorted as previously reported according to the Hoechst 33342-stained cell area of the FACS distribution [22]. To increase the amplification success rate, we added several bubbles to the single-cell lysis buffer using a micropipette (see Additional file 1, Figure S13). We sorted each single cell into a 0.4-μl aliquot of lysis buffer with a bubble; the buffer was pre-chilled to 0 °C using an IsoFreeze rack. Subsequently, we performed WTA with each single-cell lysis sample. All of the samples used are listed in Additional file 7, Table S2.

Whole-transcript amplification for single-cell Quartz-Seq

To reduce the risk of RNase contamination, the workbench environment and all experimental equipment were cleaned using the RNase removal reagent RNase Out (Molecular BioProducts).

We used low-retention single PCR tubes or sets of 8-linked PCR tubes for single-cell amplifications (TaKaRa). The cells and 10-pg total RNA samples were dissolved in 0.4 µl of single-cell lysis buffer (0.5 % NP40) on an aluminum PCR rack at 0 °C and transferred to ice. These solutions were mixed using a bench-top MixMate mixer (Eppendorf) at 2,500 rpm and 4 °C for 15 sec and then at 3000 x *g* and 4 °C for 10 sec. Immediately after the second centrifugation, 0.8 µl of priming buffer (1.5X PCR-buffer with $MgCl_2$ (Takara), 41.67 pM of the RT primer (HPLC purified, TATAGAATTCGCGGCCGCTCGCGATAATACGACTCACTATAGGGCGTTTTTTTTTTTTTTTTTTTTTTTT), 4 U/µl of RNasin Plus (Promega), and 50 µM dNTPs) were added to each tube. The solutions were mixed at 2,500 rpm and 4 °C for 15 sec. The denaturation and priming were performed at 70 °C for 90 sec and 35 °C for 15 sec using a thermal cycler (C1000 and S1000 BioRad), and the reaction tubes were placed into an aluminum PCR rack at 0 °C. Subsequently, 0.8 µl of RT buffer (1X PCR buffer, 25 U/µl SuperScript III (Life technologies), and 12.5 mM DTT) was added to each tube. The reverse-transcription was performed at 35 °C for 5 min and 45 °C for 20 min, and the reactions were heat-inactivated at 70 °C for 10 min. The reaction-tubes were then placed into an aluminum PCR rack at 0 °C. We consistently used the latest available lots of SuperScript III for the single-cell amplification. After centrifugation at 3000 x *g* and 4 °C for 10 sec, 1 µl of the exonuclease solution (1X Exonuclease buffer (Takara) and 1.5 U/µl exonuclease I (Takara)) was added to each tube. The primer digestion was performed at 37 °C for 30 min, and the reactions were heat-inactivated at 80 °C for 10 min. The reaction tubes were placed into an aluminum PCR rack at 0 °C. After centrifugation at 3,000 x *g* and 4 °C for 30 sec, 2.5 µl of poly-A tailing buffer (1X PCR-buffer, 33.6 T/µl terminal transferase (Roche), and 0.048 U RNase H (Invitrogen)) was added to each tube in the aluminum PCR rack at 0 °C. The reaction tubes were mixed at 2500 rpm and 4 °C for 15 sec. Immediately after centrifugation at 3000 x *g* and 0 °C for 10 sec, the reaction tubes were placed into a thermal cycler block, which was pre-chilled to 0 °C. Subsequently, the poly-A tailing reaction was performed at 37 °C for 50 sec and heat-inactivated at 65 °C for 10 min. The reaction tubes were then placed into an aluminum PCR rack at 0 °C. After centrifugation at 3000 x *g* and 4 °C for 30 sec, the reaction tubes were placed into an aluminum PCR rack at 0 °C. We then added 23 µl of the second-strand buffer (1.09X MightyAmp Buffer v2 (marketed as Terra PCR Direct Buffer, TaKaRa-Clontech), 70 pM Tagging primer (HPLC purified, TATAGAATTCGCGGCCGCTCGCGATTTTTTTTTTTTTTTTTTTTTTT), and 0.054 U/µl MightyAmp DNA polymerase (marketed as Terra PCR Direct Polymerase Mix, TaKaRa-Clontech)) to each tube. The reaction tubes were mixed at 2500 rpm and 4 °C for 15 sec. After centrifugation at 3000 x *g* and 4 °C for 10 sec, the second-strand synthesis was performed at 98 °C for 130 sec, 40 °C for 1 min, and 68 °C for 5 min. The reaction tubes were then immediately placed into an aluminum PCR rack at 0 °C, and 25 µl of PCR buffer (1X MightyAmp Buffer v2 and 1.9 µM suppression PCR primer (HPLC-purified, (NH_2)-GTATAGAATTCGCGGCCGCTCGCGAT)) was added. The reaction tubes were mixed at 2500 rpm and 4 °C for 15 sec. After centrifugation at 3000 x *g* and 4 °C for 10 sec, the PCR enrichment was performed using the following conditions per cycle for a total of 21 PCR cycles: 98 °C for 10 sec, 65 °C for 15 sec, and 68 °C for 5 min. After the PCR step, the reaction tubes were incubated at 68 °C for 5 min. The reaction tubes were then placed into an aluminum PCR

rack at 25 °C. The amplified cDNA was purified using a MinElute PCR purification column (Qiagen) or an Agencourt Ampure XP system (Beckman). The obtained amplified cDNA was used for subsequent detection by each platform.

For the Smart-Seq analysis, we amplified cDNA from 10 pg of the total RNA from ES cells using the SMARTer Ultra Low RNA Kit for Illumina sequencing (Clontech). After 19 cycles of PCR enrichment from the 10 pg of total ES-cell RNA, 2-3 ng of amplified cDNA were obtained.

LIMprep for single-cell Quartz-Seq

To prepare a library for conventional RNA-Seq (with non-WTA), we used the TruSeq RNA sample kit (Illumina). The library preparation was performed according to the Illumina protocol with the exception of the PCR enrichment, for which we used the Kapa HiFi DNA polymerase (Kapa Biosystems).

To prepare the library for Quartz-Seq (with WTA) and Smart-Seq (with WTA), we prepared a DNA sequencing library using our optimized library preparation method, which is termed LIMprep (Ligation-based Illumina Multiplex library Preparation). In LIMprep, we used the Kapa library preparation kit (Kapa Biosystems) and self-produced the TruSeq adaptors and PCR primers. For Quartz-Seq, 20 ng of amplified cDNA was diluted in 130 μl of TE buffer. The solutions were transferred into snap-cap microtubes (Covaris). The amplified cDNAs in the microtubes were fragmented using an S220 focused ultrasonicator (Covaris). The ultrasonication process was configured as follows: duty factor, 10 %; peak incident power, 175; cycles per burst, 100; and time, 600 sec. The fragmented cDNA was purified into 10 μl of nuclease-free water using a DNA Clean and Concentrator-5 column (Zymo Research). Subsequently, 40 μl of the End Repair Reaction mix (1.25X End Repair Buffer and 1.25X End Repair Enzyme Mix) was added to 10 μl of the fragmented cDNA solution. The end-repair reaction was performed at 20 °C for 30 min. The end-repaired DNA was purified into 12.5 μl of EB1/10 buffer (1 mM Tris-HCl pH 8.0) using a DNA Clean and Concentrator-5 column. Subsequently, 12.5 μl of the A-tailing mix (2X A-tailing Buffer and 2X A-tailing Enzyme) was added to 12.5 μl of the end-repaired DNA solution. The A-tailing reaction was performed at 30 °C for 30 min. The A-tailed DNA was purified into 12.5 μl of EB1/10 buffer using a DNA Clean and Concentrator-5 column. Subsequently, 12.5 μl of the adaptor ligation mix (2X Ligation Buffer, 2X DNA Ligase, and 10 pmol of each self-produced adaptor) was added to 12.5 μl of the A-tailed DNA solution at 4 °C. The adaptor ligation was performed at 20 °C for 15 min. For the adaptor ligation, we used 10 pmol of the self-produced TruSeq adaptor per sample. Each self-produced TruSeq adaptor was prepared using the following HPLC-purified primers (Hokkaido System Science):

TRSU:
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
TRSI-2: (5'-phosphate)GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTT*G
TRSI-4: (5'-phosphate)GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTT*G

TRSI-5: (5'-phosphate)GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGCTT*G
TRSI-6: (5'-phosphate)GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTT*G
TRSI-7: (5'-phosphate)GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTGCTT*G
TRSI-12: (5'-phosphate)GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGTCTTCTGCTT*G.
The asterisks indicate phosphorothioate bonds. Each primer was dissolved in the adaptor buffer (10 mM Tris-HCl pH 7.8, 0.1 mM EDTA pH 8.0, and 50 mM NaCl) to a concentration of 100 μM. Equal amounts of 100 μM TRSU and 100 μM of each TRSI primer were added to the PCR tubes. After mixing, these primers were incubated at 95 °C for 2 min. The primer annealing was then performed (95 °C for 2 min, followed by a temperature decrease of -0.5 °C per cycle for 170 cycles). Subsequently, the reaction tubes were incubated at 4 °C for 5 min. The resulting adaptors were diluted with adaptor buffer to a concentration of 10 μM. We prepared 1-μl aliquots with 10 μM of each adaptor, which were stored at -80 °C until use. The removal of the adaptor dimer was performed as follows. A 25-μl volume of binding support buffer (1 M NaCl, 20 mM $MgCl_2$, and 20 mM Tris-HCl, pH 7.8) and 60 μl of Agencourt Ampure XP bead solution was added to 25 μl of the adaptor-ligated DNA solution. After 15 min of incubation at 25 °C, the beads were separated using a magnetic stand for at least 10 min. The beads were washed twice with 80 % ethanol for 1 min. The adaptor-ligated DNA was eluted with 25 μl of EB1/10. The purification step for the removal of the adaptor dimer was repeated. Finally, the adaptor-ligated DNA was eluted with 20 μl of EB1/10. A volume of 30 μl of the PCR solution (1.666X Kapa HiFi DNA polymerase ready mix, 17.5 pmol TPC1 primer, and 17.5 pmol TPC2 primer) was added to 20 μl of the adaptor-ligated DNA. The primer sequences were as follows: TPC1, AATGATACGGCGACCACCGA*G; and TPC2, CAAGCAGAAGACGGCATACGA*G.
The asterisks indicate phosphorothioate bonds. Each primer was dissolved in the adaptor buffer (10 mM Tris-HCl pH 7.8, 0.1 mM EDTA pH 8.0, and 50 mM NaCl) to a concentration of 100 μM. Prior to the PCR enrichment, the reaction tubes were incubated at 98 °C for 45 sec. The PCR enrichment was then performed (98 °C for 15 sec, 60 °C for 30 sec and 72 °C for 30 sec per cycle). Typically, 10-12 PCR cycles were used. After the PCR enrichment, the DNA sequencing library was purified using Ampure XP beads. The concentrations of the DNA sequencing library, including the TruSeq Index sequence, were estimated using the Kapa library quantification kit (Kapa). The Pearson correlation coefficient for three technical replicates of the LIMprep protocol was 0.9976 ± 0.0005.

GeneChip

The cDNA was synthesized from 0.25 μg of total RNA using random 6-mers (Promega) and SuperScript II Reverse Transcriptase (Invitrogen) according to Illumina's standard protocol. The

cDNA synthesis, cRNA labeling reactions and hybridization to the Affymetrix high-density oligonucleotide arrays for *Mus musculus* (Mouse Genome 430 Array) were performed according to the instructions detailed in the Expression Analysis Technical Manual (Affymetrix). For single-cell Quartz-Chip, 10 ng of amplified cDNA was used in the cRNA labeling reactions. The expression values of the Kurimoto et al. and the Quartz-Chip methods were quantified using the RMA (Robust Multi-array Averaging) method. All of the data were normalized using the quantile normalization method to compare the expression values between the different microarrays [26].

Bioinformatics analysis

All raw sequencing reads were trimmed using the Trimmomatic software to remove the sequencing and WTA primers. All of the trimmed sequence reads were mapped to the mouse reference genome (mm9) using TopHat 2.0.3 [27] with the default parameters. The FPKMs were calculated using Cufflinks 2.0.1 [28] with a transcriptome reference (Ensembl Mouse Transcript). The sample clustering of the single-cell RNA-Seq data was performed using R and the pvclust package [29] with 1000 bootstrap resamplings and Ward distance functions. All of the data were visualized using R, the ggplot2 package and the cummeRbund software.

To identify the significant differential expression between two differentiated cell states from single-cell RNA-Seq data, we performed the Wilcoxon rank test and calculated the mutual information between the gene-expression distributions of the embryonic stem cells and the primitive endoderm cells using an empirical Bayes estimator.

Although our novel single-cell RNA-Seq method is highly sensitive, cost effective, and easy to perform, it is not completely without amplification bias. To identify the differentially expressed genes from the single-cell RNA-Seq method with small sample datasets, we inferred mutual information between the gene expression distributions of the two types of cells using an empirical Bayes estimator: the mutual information MI(X, Y) for pairs of cell states X and Y, where X and Y may, for example, represent expression levels of the two cell states. The mutual information is considered the Kullback-Leibler distance from the joint probability density to the product of the marginal probability densities:

$$MI(X,Y) = E_{f(x,y)}\left\{\log \frac{f(x,y)}{f(x)f(y)}\right\}. \qquad (1)$$

The mutual information (MI) is always non-negative, symmetric and equal to zero only if X and Y are independent. The MI can be represented as a summation of entropies:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y). \qquad (2)$$

To infer the entropy from gene expression data with a small sample size, we applied an empirical Bayes approach, i.e., the so-called James-Stein estimator [30]. First, the gene expression data were made discrete using the Freedman and Diaconis algorithm to determine the number of bins and the width of the histograms. We then estimated the K2 cell frequencies of the K × K contingency table for each cell-state pair X and Y using the James-Stein estimator. Finally, we calculated H(X), H(Y), H(X, Y) and MI(X, Y).

To define the reproducibility of the variation in the measured expression levels among the single cells analyzed using Quartz-Seq, we used FDR to apply the F-test to two independent Quartz-Seq datasets with multi-testing adjustments.

All of the results of our analyses are available for download at http://quartz-seq.org/.

Accession codes

All data are available on GEO [GSE42268].

qPCR validation of reverse-transcribed cDNA and quantification of amplified cDNA

To validate the quantification of the reverse-transcribed cDNA and the amplified cDNA, we detected the gene expression using the Power SYBR Green PCR master mix and qPCR primers according to the methods described in a previous study [12]. The qPCR primer sets for each gene are shown in Additional file 8, Table S3.

The non-WTA sample was prepared by reverse-transcription using 200 ng of total ES-cell (EB5 cell line) RNA containing the spike RNAs (*Lys*, *Dap*, *Phe,* and *Thr*). A sample of 200 ng total RNA was dissolved in 10 μl of single-cell lysis buffer (0.5 % NP40) on an aluminum PCR rack at 0 °C and placed on ice. Immediately after, 20 μl of priming buffer (1.5X PCR buffer with $MgCl_2$ (TaKaRa), 41.67 pM of RT primer (HPLC purified, TATAGAATTCGCGGCCGCTCGCGATAATACGACTCACTATAGGGCGTTTTTTTTTTTTTTTTTTTTTTT), 4 U/μl RNasin-plus (Promega), and 50 μM dNTPs) was added to each tube. The solution was mixed at 2500 rpm and 4 °C for 15 sec. The denaturation and priming were performed at 70 °C for 90 sec and 35 °C for 15 sec using a thermal cycler (C1000 and S1000 BioRad), and the reaction tubes were placed in an aluminum PCR rack at 0 °C. A 20-μl volume of RT buffer (1X PCR-buffer, 25 U/μl SuperScript III (Life Technologies), and 12.5 mM DTT) was added to each tube. The reverse transcription was performed at 35 °C for 5 min and 45 °C for 18 min and then heat-inactivated at 70 °C for 10 min. Subsequently, the reaction tubes were placed in an aluminum PCR rack at 0 °C.

Amplification-free single-cell qPCR

We used a semi-skirted 96-well PCR plate for the amplification-free single-cell qPCR. Single cells in the G1 cell-cycle phase were individually collected into 1 μl of lysis buffer (0.25 % NP40 and 1 U/μl RNasin plus RNase Inhibitor) at 4°C in an IsoFreeze rack. After centrifugation at 2000 x *g* and 4°C for 2 min, these solutions were mixed at 2000 rpm for 15 sec. To detect the variability of the reverse-transcription of each gene, we prepared an “averaged” single-cell pooled sample. For the pooled sample, 200 single cells in the G1 phase were collected into 200 μl of lysis buffer. After mixing, 1 μl of the pooled lysis solution was dispensed into each well of the 96-well PCR plate. Subsequently, 2 μl of RT buffer (1.5X VILO Reaction Buffer and 1.5X SuperScript Enzyme Mix) was added to each well. The VILO Reaction Buffer contains random primers. These plates were incubated as follows: 25°C for 10 min, 42°C for 60 min, 85°C for 10 min. Then, 15 μl of the qPCR-dilution solution (0.0001 % NP40) was added and mixed well. Mouse genomic DNA was used to normalize the qPCR quantification. Volumes of 7 μl of the qPCR solution (1.4X QuantiTect SYBR Green PCR Master Mix, 5 pmol forward primer and 5 pmol reverse primer) and 3 μl of diluted solution were added to a 384-well qPCR plate. The qPCR plate was incubated at 95 °C for 15 min. Subsequently, qPCR was performed for 45

cycles, which consisted of 95 °C for 15 sec, and 60 °C for 1 min. The data were collected at 60 °C. The primer sets for each gene are shown in Additional file 8, Table S3.

Abbreviations

cDNA: complementary DNA; cRNA: complementary RNA; CV: Coefficient of variation; ES cell: Embryonic stem cell; FDR: False discovery rate; FPKM: Fragment per kilobase of transcript per million fragments sequenced; IVT: In vitro transcription; PCA : Principle component analysis; PCC: Pearson correlation coefficient; PCR: Polymerase chain reaction; PE: Paired-end; PrE: Primitive endoderm; RMA: Robust Multi-array Averaging; RPKM: Reads Per Kilobase of exon Model per million mapped reads; RT: Reverse transcription; S.D.: Standard deviation; SE: Single-end; WTA: Whole-transcript amplification

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

YS, IN and HRU. designed and configured the various approaches used in this study. YS performed the majority of the experiments. IN analyzed the data and developed the bioinformatics methods and tools. TH performed the cell sorting and assisted with the amplification-free single-cell qPCR. HD assisted with the cellular experiments. KDU. assisted with the GeneChip experiments. T.I. assisted with the discussion on the amplification of small amounts of RNA. YS, IN and HRU. prepared the figures and wrote the manuscript. All authors read and approved the final manuscript.

Description of additional data files

Additional file 1: Figure S1: Schematic of the whole-transcript amplification methods based on the poly-A tailing reaction. Figure S2: Improvement parameters of whole-transcript amplification for Quartz-Seq. Figure S3: Key steps for robust suppression of byproducts. Figure S4: Optimization of suppression PCR for Quartz-Seq. Figure S5: Optimal DNA polymerase for whole-transcript amplification. Figure S6: Quality check of the library preparation for single-cell Quartz-Seq. Figure S8: Percentage of sequence reads of the suppression PCR primer or rRNA. Figure S9: Relationship between the read number and the reproducibility. Figure S10: Optimization of cDNA length in technical development for single-cell Quartz-Seq. Figure S11: Trend of un-amplified isoforms in each single-cell RNA-Seq method. Figure S12: Amplified cDNA lengths resulting from single-cell RNA-Seq methods. Figure S13: Success rate of whole-transcript amplification from FACS-sorted single-cell. Figure S14: Amount of total RNA from a single cells at each cell-cycle phase. Figure S15: Principle component analysis (PCA) of single cells from different cell types at different cell-cycle phases. Figure S16: Overrepresentation analyses for Principle component (PC) of single cells from same cell types at same cell-cycle phases (G1). Figure S17: Scatter plots of conventional RNA-Seq and Quartz-Seq using 50 ES cells at the G1 phase of the cell cycle and Quartz-Seq using 10 pg of total ES RNA. Figure S18: Effect of carried over buffer for PCR efficiency.

Additional file 2: Supplementary note.

Additional file 3: Figure S7: All scatter plots

Additional file 4: Table S1. All result of linear regression and correlation analysis

Additional file 5: Supplementary movie 1. PCA with single-cell Quartz-Seq data of ES and PrE single-cell.

Additional file 6: Supplementary movie 2. PCA with single-cell Quartz-Seq data of ES from different cell-cycle phase.

Additional file 7: Table S2. Sequencing information.

Additional file 8: Table S3. Primer information.

Author’s information

[1]Functional Genomics Unit, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, Japan.
[2]Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, , 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, Japan.
[3]Laboratory for Systems biology, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, Japan.
[4]Laboratory for Synthetic Biology, Quantitative Biology Center, RIKEN, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, Japan.
[5]Laboratory for Sensory Circuit Formation, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, Japan.
[6]JST, PRESTO, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, Japan.
[7]These authors contributed equally to this work

Acknowledgments

We acknowledge the members of the Functional Genomics Unit and Genome Resource and Analysis Unit at the RIKEN Center for Developmental Biology (RIKEN CDB), particularly Junko Sakai, Chiharu Tanegashima and Kazu Itomi. We also thank Hitoshi Niwa (The Laboratory for Pluripotent Stem Cell Studies at RIKEN CDB) for providing the embryonic stem cells and the primitive endoderm cells. This work was supported by the Program for Innovative Cell Biology by Innovative Technology from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, the Leading Project for the Realization of Regenerative Medicine, and the Director's Fund 2011 of RIKEN CDB. The Special Postdoctoral Researchers Program from RIKEN also supported this work. Some of the calculations were performed using the RIKEN Integrated Combined Cluster (RICC) and super computer system in National Institute of Genetics (NIG), Research Organization of Information and Systems (ROIS)

References

1. Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H: **Identification and characterization of subpopulations in undifferentiated ES cell culture.** *Development* 2008, **135:**909-918.
2. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S: **Transcriptome-wide noise controls lineage choice in mammalian progenitor cells.** *Nature* 2008, **453:**544-547.
3. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R: **Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase.** *Cell* 2012, **150:**1209-1222.
4. Bumgarner SL, Neuert G, Voight BF, Symbor-Nagrabska A, Grisafi P, van Oudenaarden A, Fink GR: **Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment.** *Mol Cell* 2012, **45:**470-482.
5. Dunlop MJ, Cox RS, 3rd, Levine JH, Murray RM, Elowitz MB: **Regulatory activity revealed by dynamic correlations in gene expression noise.** *Nat Genet* 2008, **40:**1493-1498.
6. Eldar A, Elowitz MB: **Functional roles for noise in genetic circuits.** *Nature* 2010, **467:**167-173.
7. Kittisopikul M, Suel GM: **Biological role of noise encoded in a genetic network motif.** *Proc Natl Acad Sci U S A* 2010, **107:**13300-13305.
8. Munsky B, Neuert G, van Oudenaarden A: **Using gene expression noise to understand gene regulation.** *Science* 2012, **336:**183-187.
9. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS: **Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.** *Science* 2010, **329:**533-538.
10. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, Lao K, Surani MA: **Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis.** *Cell Stem Cell* 2010, **6:**468-478.
11. Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, Tuch BB, Heard E, Lao K, Surani MA: **Deterministic and stochastic allele specific gene expression in single mouse blastomeres.** *PLoS One* 2011, **6:**e21208.
12. Kurimoto K, Yabuta Y, Ohinata Y, Ono Y, Uno KD, Yamada RG, Ueda HR, Saitou M: **An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis.** *Nucleic Acids Res* 2006, **34:**e42.
13. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA: **mRNA-Seq whole-transcriptome analysis of a single cell.** *Nat Methods* 2009, **6:**377-382.
14. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S: **Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq.** *Genome Res* 2011, **21:**1160-1167.
15. Fan JB, Chen J, April CS, Fisher JS, Klotzle B, Bibikova M, Kaper F, Ronaghi M, Linnarsson S, Ota T, Chien J, Laurent LC, Nisperos SV, Chen GY, Zhong JF: **Highly parallel genome-wide expression analysis of single mammalian cells.** *PLoS One* 2012, **7:**e30794.
16. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R: **Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.** *Nat Biotechnol* 2012.
17. Hashimshony T, Wagner F, Sher N, Yanai I: **CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification.** *Cell Rep* 2012, **2:**666-673.

18. Kurimoto K, Yabuta Y, Ohinata Y, Saitou M: **Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis.** *Nat Protoc* 2007, **2:**739-752.
19. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, Lao K, Surani MA: **RNA-Seq analysis to capture the transcriptome landscape of a single cell.** *Nat Protoc* 2010, **5:**516-535.
20. Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA: **An improved PCR method for walking in uncloned genomic DNA.** *Nucleic Acids Res* 1995, **23:**1087-1088.
21. Shimosato D, Shiki M, Niwa H: **Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells.** *BMC Dev Biol* 2007, **7:**80.
22. Hayashi T, Shibata N, Okumura R, Kudome T, Nishimura O, Tarui H, Agata K: **Single-cell gene profiling of planarian stem cells using fluorescent activated cell sorting and its "index sorting" function for stem cell research.** *Dev Growth Differ* 2010, **52:**131-144.
23. Kalmar T, Lim C, Hayward P, Munoz-Descalzo S, Nichols J, Garcia-Ojalvo J, Martinez Arias A: **Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells.** *PLoS Biol* 2009, **7:**e1000149.
24. Carter MG, Stagg CA, Falco G, Yoshikawa T, Bassey UC, Aiba K, Sharova LV, Shaik N, Ko MS: **An in situ hybridization-based screen for heterogeneously expressed genes in mouse ES cells.** *Gene Expr Patterns* 2008, **8:**181-198.
25. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee SL, Stagg CA, Hoang HG, Yang HT, Indig FE, Wersto RP, Ko MS: **Zscan4 regulates telomere elongation and genomic stability in ES cells.** *Nature* 2010, **464:**858-863.
26. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** *Nucleic Acids Res* 2003, **31:**e15.
27. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25:**1105-1111.
28. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: **Improving RNA-Seq expression estimates by correcting for fragment bias.** *Genome Biol*, **12:**R22.
29. Shimodaira H: **Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling.** *Annals of Statistics* 2004, **32:**2616-2641.
30. Hausser J, Strimmer K: **Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks.** *Journal of Machine Learning Research* 2009, **10:**1469-1484.

Figure legends

Figure 1 Schematic of the single-cell Quartz-Seq and Quartz-Chip methods.
All of the steps of the whole-transcript amplification were executed in a single PCR tube. The first-strand cDNA was synthesized using RT primer, which contains oligo-dT24, the T7 promoter (T7) and the PCR target region (M) sequences. After the first-strand synthesis, the majority of the RT primer was digested by exonuclease I, although it was not possible to completely eliminate the RT primer. A poly-A tail was then added to the 3' ends of the first-strand cDNA and to any surviving RT primer. After the second-strand synthesis with the Tagging primer, the resulting cDNA and the byproducts from the surviving primers contained the WTA adaptor sequences, which include the RT primer sequence and the Tagging primer sequence. These DNAs were used for the suppression PCR, which used the suppression PCR primer. The enrichment of the short DNA fragments, such as the byproducts, was suppressed. After the enrichment, the high-quality cDNA, which does not contain any byproducts, was obtained. The amplified cDNAs have the T7 promoter sequence at the 3' ends of the DNA. These cDNAs were used for the Illumina sequencing and microarray experiments.

Figure 2 Reproducibility and sensitivity of single-cell Quartz-Seq.
(a) Representative scatterplot of the gene expression data from two replicate single-cell Quartz-Seq analyses of 10 pg of total ES-cell RNA (left panel). The blue line indicates a two-fold change. The red line is a linear regression. Scatterplot of single-cell Quartz-Seq (WTA) and conventional RNA-Seq (non-WTA) data using 1 μg of total ES-cell RNA (right panel). **(b)** Ratio of detected genes for three replicate Quartz-Seq analyses. A total of 82.1 % of the genes were detected by two different independent Quartz-Seq experiments (left panel). The right panel shows the ratio of the genes detected by Quartz-Seq and conventional RNA-Seq. More than 68.2 % of the genes were detected by single-cell Quartz-Seq (right panel).

Figure 3 Comparison of the performances of Quartz-Seq, Quartz-Chip and other methods.
(a) Box plot of Pearson correlations for the technical replication of Quartz-Seq and Smart-Seq with 10 pg diluted total RNA. We reanalyzed the following four original Smart-Seq datasets: MB, HB (Nx), UHRR and UHRR (Nx). HB: Human brain, MB: Mouse brain, UHRR: Universal human reference RNA, Nx: Nextera library preparation kit. The asterisk indicates the downsampling sequence reads from single-cell Quartz-Seq (PE, 60 M, n = 3). **(b)** Box plot of Pearson correlations between conventional RNA-Seq and single-cell RNA-Seq methods. **(c)** Comparison of our current and previous methods using 10 pg of total ES-cell RNA through the use of GeneChip. The left plots show the performance of Quartz-Chip, and the right plots show the performance of the Kurimoto et al. method. The Kurimoto et al. data were reanalyzed using the original sets. The bar plots in the right panels show the numbers of genes that were detected with each method. The gray bars show the total number of genes, and the blue bars indicate the number of detected genes.

Figure 4 Single-cell Quartz-Seq detects different cell types.
Results of single-cell Quartz-Seq with 12 embryonic stem cells and 12 primitive endoderm cells. **(a)** Clustering of all samples. **(b)** Heatmap of the marker genes for embryonic stem cells and primitive endoderm cells and the housekeeping genes. The bar plot in the right panel shows the

mutual information. A high degree of mutual information indicates a high differential expression between two cell states. **(c)** Verification of the expression pattern between single cells using amplification-free single-cell qPCR. The gene expression data for single cells correspond to ES (n = 96 single ES cells in the G1 phase of the cell cycle), PrE (n = 96 single PrE cells in the G1 phase), ES200 (n = 40 single-cell-sized samples from pooled lysis of ES cells in the G1 phase), and PrE200 (n = 40 single-cell-sized samples from pooled lysis of PrE cells in the G1 phase). The ES and PrE box plots represent the gene expression variability, which includes the biological variability and the experimental error. The ES200 and PrE200 box plots represent the gene expression variability due to experimental error.

Figure 5 Single-cell Quartz-Seq reveals fluctuations in global gene expression in a single cell type at the same cell-cycle phase.
(a) The X- and Y-axes represent the standard deviation (S.D.) of gene expression from the different datasets from single-cell Quartz-Seq with single ES cells in the G1 phase of the cell cycle (n = 12, n = 8). **(b)** We detected the expression of 9 genes by amplification-free single-cell qPCR with single ES cells in the G1 phase. We selected 9 genes from among the ES differentiated genes (ES, FPKM > 2; PrE, FPKM < 1): *Fn1*, *Zfp42*, *Sgk1*, *Tfrc*, *Utf1*, *Lefty1*, *Nanog*, *Sox2*, and *Spp1*. The plot shows the gene expression for the single-cell analysis of ES (n = 96 single ES cells in the G1 phase of the cell cycle) and ESA samples (n = 48 “averaged” single-cell samples from 300 pooled ES cells in the G1 phase). The ES single-cell box plots represent the gene expression variability that contains the biological variability and the experimental error. The ESA sample box plots represent the gene expression variability due to experimental error. **(c)** The X-axis represents the CV of the gene expression from single-cell Quartz-Seq data with single ES cells in the G1 phase of the cell cycle (n = 12). The Y-axis represents the CV of the gene expression from amplification-free single-cell qPCR with ES cells in the G1 phase of the cell cycle (n = 96). The CVs of the gene expression are plotted for 9 genes. The red lines represent regressions.

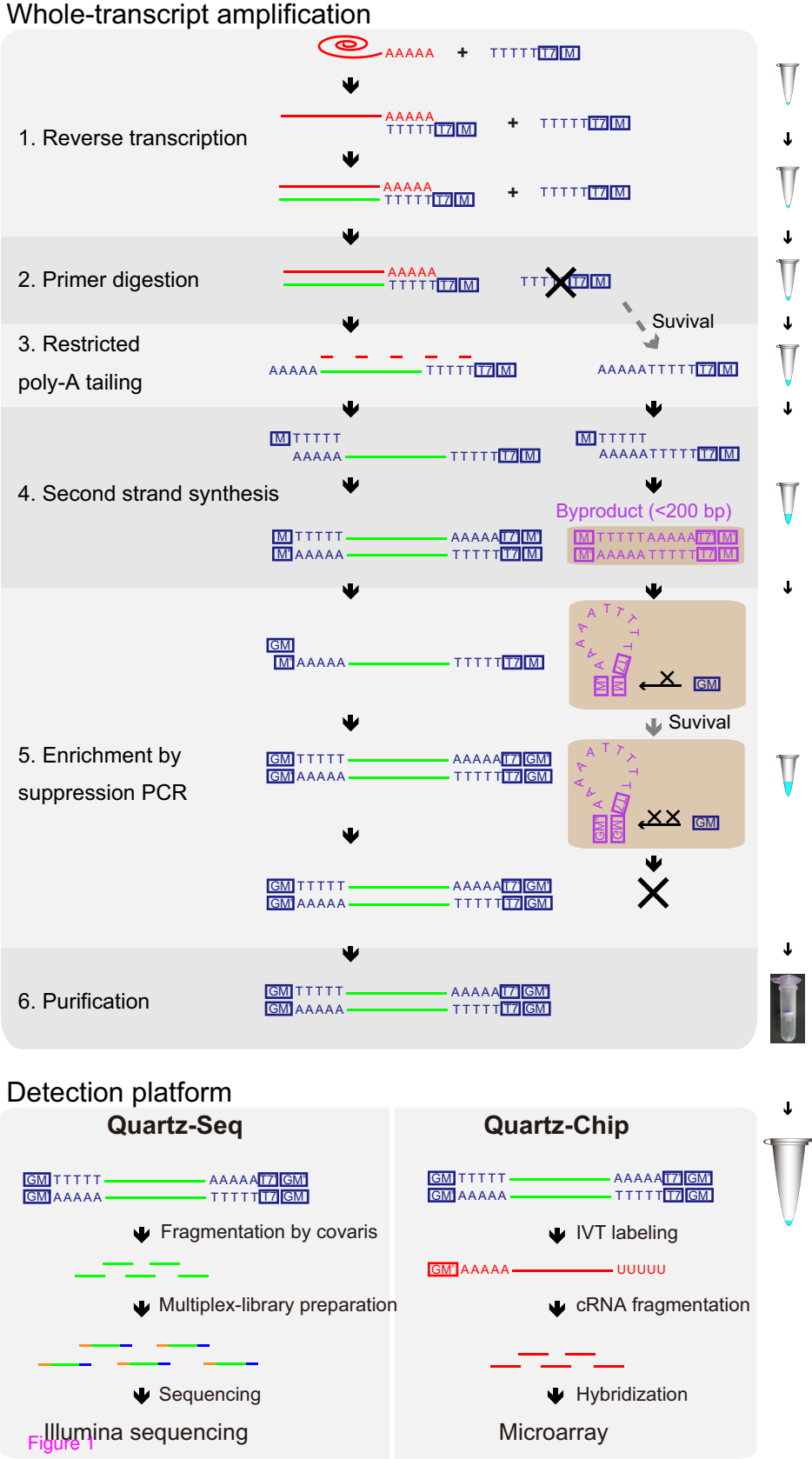
Whole-transcript amplification
1. Reverse transcription
2. Primer digestion
3. Restricted poly-A tailing
Suvival
4. Second strand synthesis
Byproduct (<200 bp)
5. Enrichment by suppression PCR
Suvival
6. Purification
Detection platform
Quartz-Seq
Fragmentation by covaris
Multiplex-library preparation
Sequencing
Illumina sequencing
Quartz-Chip
IVT labeling
cRNA fragmentation
Hybridization
Microarray

Figure 1

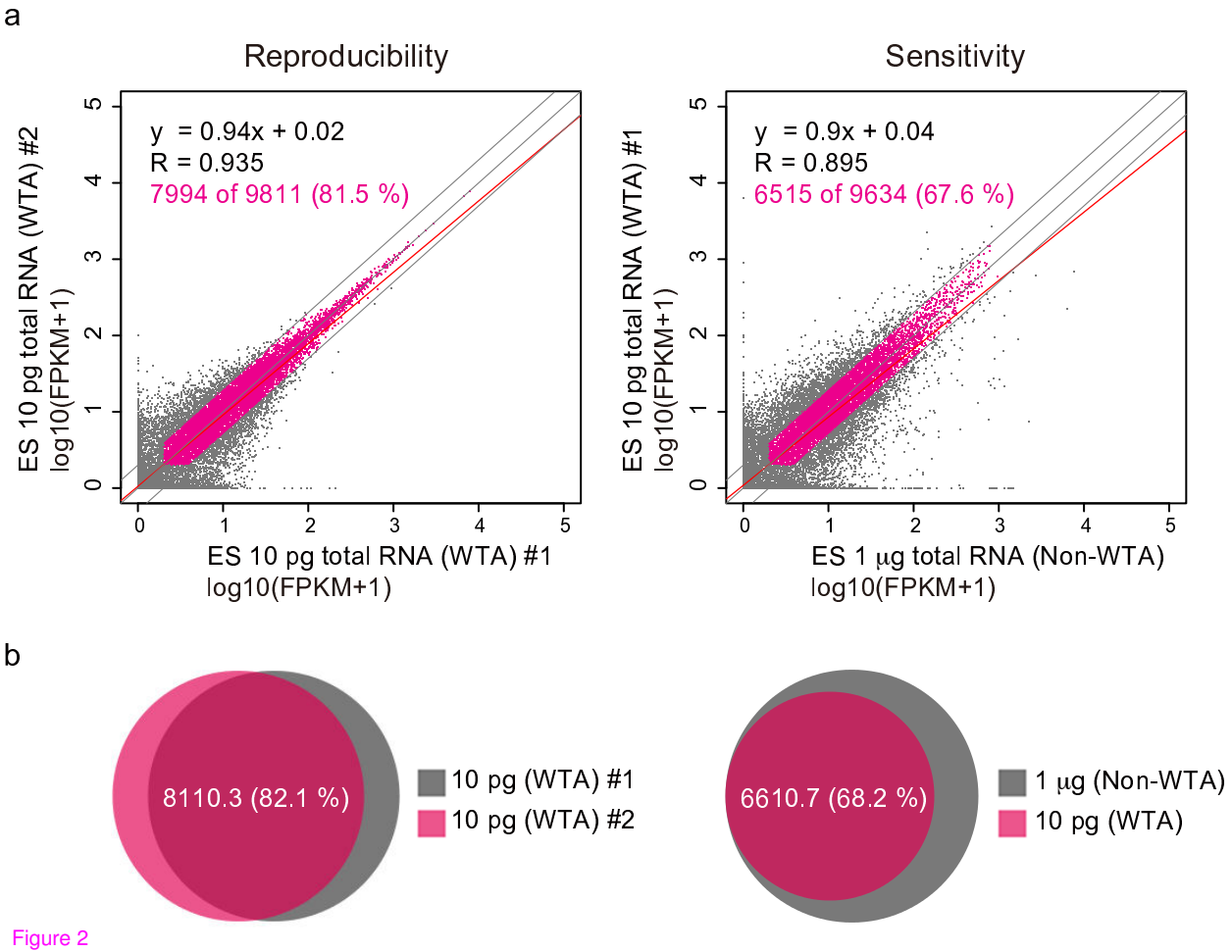
a
Reproducibility
y = 0.94x + 0.02
R = 0.935
7994 of 9811 (81.5 %)
ES 10 pg total RNA (WTA) #2
log10(FPKM+1)
ES 10 pg total RNA (WTA) #1
log10(FPKM+1)
Sensitivity
y = 0.9x + 0.04
R = 0.895
6515 of 9634 (67.6 %)
ES 10 pg total RNA (WTA) #1
log10(FPKM+1)
ES 1 μg total RNA (Non-WTA)
log10(FPKM+1)
b
8110.3 (82.1 %)
10 pg (WTA) #1
10 pg (WTA) #2
6610.7 (68.2 %)
1 μg (Non-WTA)
10 pg (WTA)

Figure 2

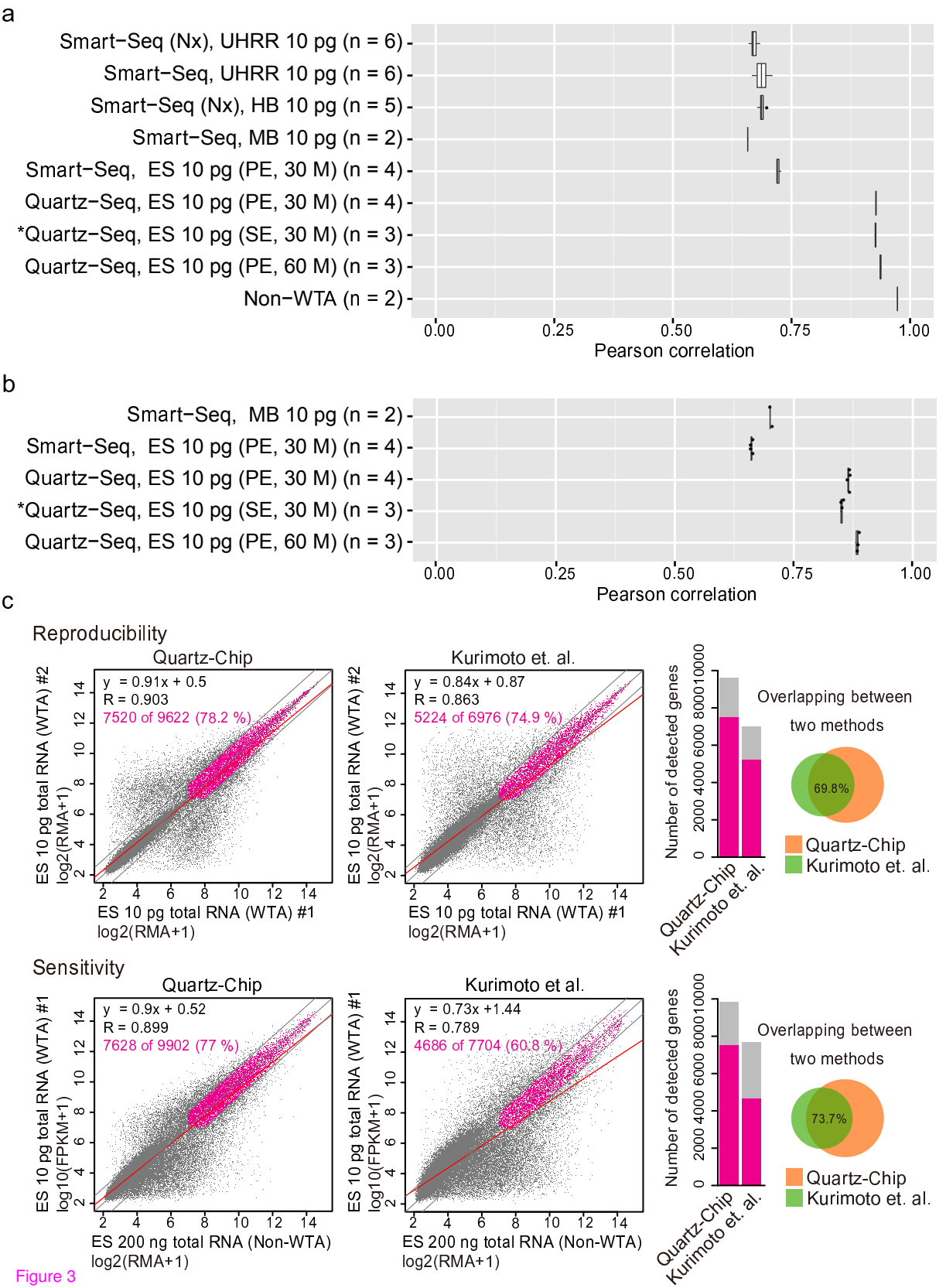
a
Smart−Seq (Nx), UHRR 10 pg (n = 6)
Smart−Seq, UHRR 10 pg (n = 6)
Smart−Seq (Nx), HB 10 pg (n = 5)
Smart−Seq, MB 10 pg (n = 2)
Smart−Seq, ES 10 pg (PE, 30 M) (n = 4)
Quartz−Seq, ES 10 pg (PE, 30 M) (n = 4)
*Quartz−Seq, ES 10 pg (SE, 30 M) (n = 3)
Quartz−Seq, ES 10 pg (PE, 60 M) (n = 3)
Non−WTA (n = 2)
0.00
0.25
0.50
0.75
1.00
Pearson correlation
b
Smart−Seq, MB 10 pg (n = 2)
Smart−Seq, ES 10 pg (PE, 30 M) (n = 4)
Quartz−Seq, ES 10 pg (PE, 30 M) (n = 4)
*Quartz−Seq, ES 10 pg (SE, 30 M) (n = 3)
Quartz−Seq, ES 10 pg (PE, 60 M) (n = 3)
0.00
0.25
0.50
0.75
1.00
Pearson correlation
c
Reproducibility
Quartz-Chip
y = 0.91x + 0.5
R = 0.903
7520 of 9622 (78.2 %)
ES 10 pg total RNA (WTA) #2
log2(RMA+1)
ES 10 pg total RNA (WTA) #1
log2(RMA+1)
Kurimoto et. al.
y = 0.84x + 0.87
R = 0.863
5224 of 6976 (74.9 %)
ES 10 pg total RNA (WTA) #2
log2(RMA+1)
ES 10 pg total RNA (WTA) #1
log2(RMA+1)
Number of detected genes
Overlapping between two methods
69.8%
Quartz-Chip
Kurimoto et. al.
Sensitivity
Quartz-Chip
y = 0.9x + 0.52
R = 0.899
7628 of 9902 (77 %)
ES 10 pg total RNA (WTA) #1
log10(FPKM+1)
ES 200 ng total RNA (Non-WTA)
log2(RMA+1)
Kurimoto et al.
y = 0.73x +1.44
R = 0.789
4686 of 7704 (60.8 %)
ES 10 pg total RNA (WTA) #1
log10(FPKM+1)
ES 200 ng total RNA (Non-WTA)
log2(RMA+1)
Number of detected genes
Overlapping between two methods
73.7%
Quartz-Chip
Kurimoto et. al.

Figure 3

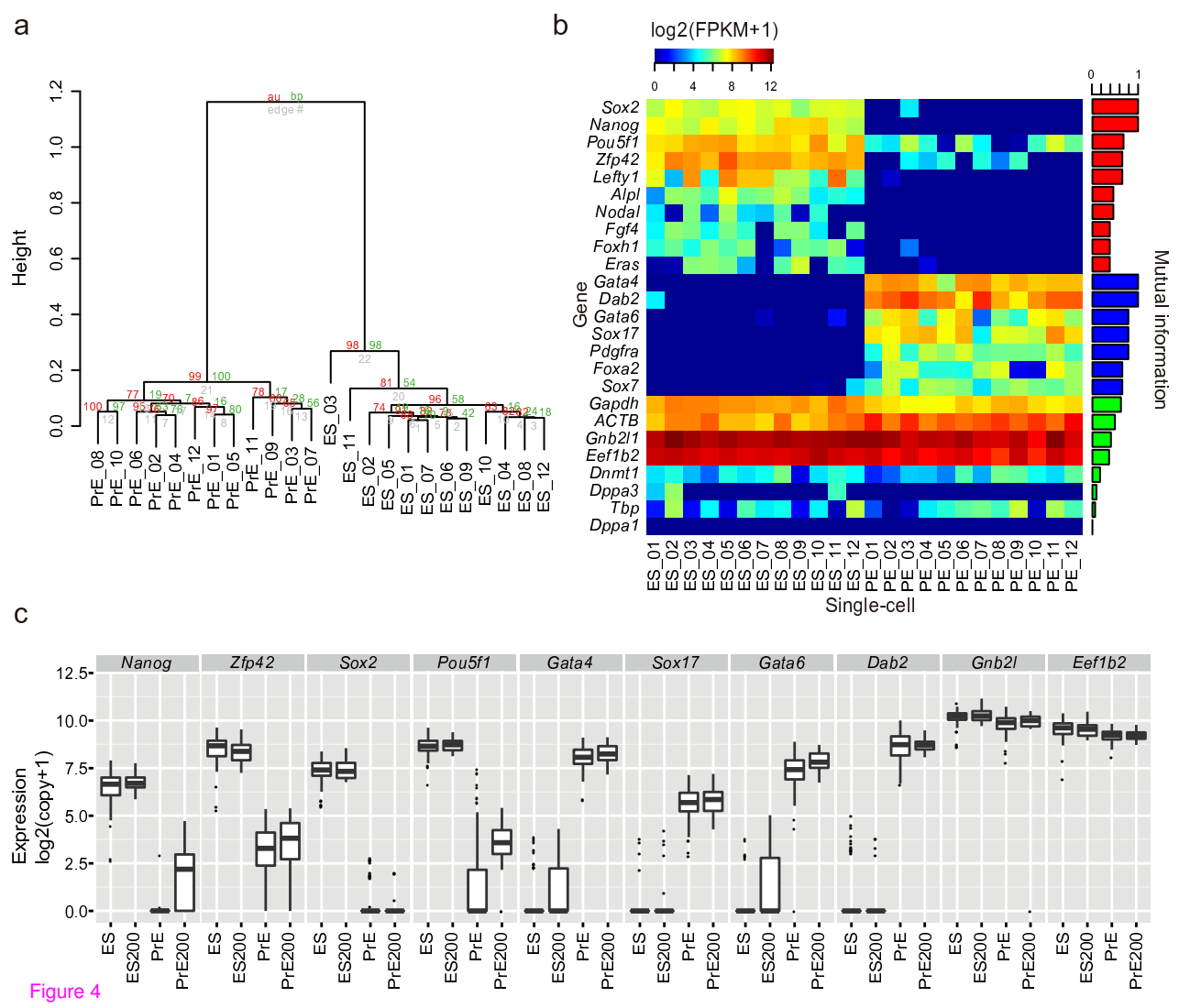
a
Height
au bp
edge #
PrE_08
PrE_10
PrE_06
PrE_02
PrE_04
PrE_12
PrE_01
PrE_05
PrE_11
PrE_09
PrE_03
PrE_07
ES_03
ES_11
ES_02
ES_05
ES_01
ES_07
ES_06
ES_09
ES_10
ES_04
ES_08
ES_12
b
log2(FPKM+1)
Sox2
Nanog
Pou5f1
Zfp42
Lefty1
Alpl
Nodal
Fgf4
Foxh1
Eras
Gata4
Dab2
Gata6
Sox17
Pdgfra
Foxa2
Sox7
Gapdh
ACTB
Gnb2l1
Eef1b2
Dnmt1
Dppa3
Tbp
Dppa1
Gene
Mutual information
Single-cell
c
Nanog
Zfp42
Sox2
Pou5f1
Gata4
Sox17
Gata6
Dab2
Gnb2l
Eef1b2
Expression
log2(copy+1)
ES
ES200
PrE
PrE200

Figure 4

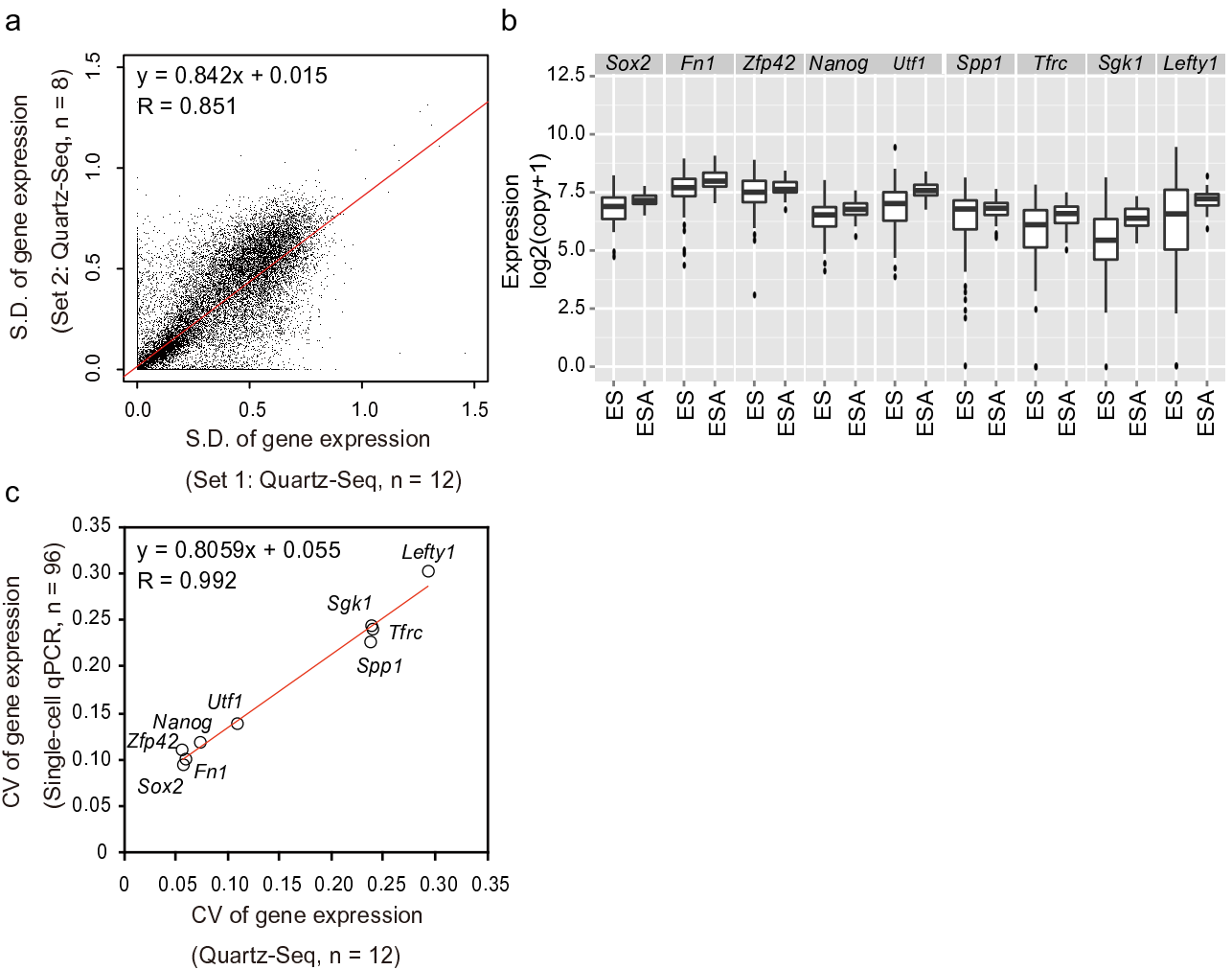
a
y = 0.842x + 0.015
R = 0.851
S.D. of gene expression
(Set 2: Quartz-Seq, n = 8)
S.D. of gene expression
(Set 1: Quartz-Seq, n = 12)
b
Sox2
Fn1
Zfp42
Nanog
Utf1
Spp1
Tfrc
Sgk1
Lefty1
Expression
log2(copy+1)
ES
ESA
c
y = 0.8059x + 0.055
R = 0.992
CV of gene expression
(Single-cell qPCR, n = 96)
CV of gene expression
(Quartz-Seq, n = 12)

Figure 5

Additional files provided with this submission:

Additional file 1: Additional file 1.pdf, 17477K
http://genomebiology.com/imedia/1624187081961370/supp1.pdf
Additional file 2: Additional file 2.docx, 32K
http://genomebiology.com/imedia/1620440569697286/supp2.docx
Additional file 3: Additional file 3.pdf, 3746K
http://genomebiology.com/imedia/1668360736969727/supp3.pdf
Additional file 4: Additional file 4.xls, 218K
http://genomebiology.com/imedia/1362131239969727/supp4.xls
Additional file 5: Additional file 5.gif, 2492K
http://genomebiology.com/imedia/1954857349697291/supp5.gif
Additional file 6: Additional file 6.gif, 2055K
http://genomebiology.com/imedia/7363608479697292/supp6.gif
Additional file 7: Additional file 7.xls, 44K
http://genomebiology.com/imedia/4706793219697277/supp7.xls
Additional file 8: Additional file 8.xls, 31K
http://genomebiology.com/imedia/8019602599697277/supp8.xls