analytical chemistry

Letter

pubs.acs.org/ac

Single Cell Total RNA Sequencing through Isothermal Amplification in Picoliter-Droplet Emulsion

Yusi Fu,[†] He Chen,[†] Lu Liu,[†,‡] and Yanyi Huang*[,†,‡,§]

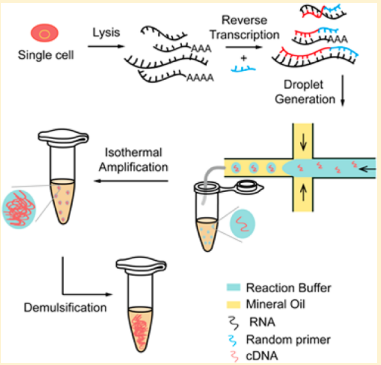
[†]Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, and Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing 100871, China

[‡]College of Engineering, Peking University, Beijing 100871, China

[§]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

Supporting Information

ABSTRACT: Prevalent single cell RNA amplification and sequencing chemistries mainly focus on polyadenylated RNAs in eukaryotic cells by using oligo(dT) primers for reverse transcription. We develop a new RNA amplification method, "easier-seq", to reverse transcribe and amplify the total RNAs, both with and without polyadenylate tails, from a single cell for transcriptome sequencing with high efficiency, reproducibility, and accuracy. By distributing the reverse transcribed cDNA molecules into 1.5×10^5 aqueous droplets in oil, the cDNAs are isothermally amplified using random primers in each of these 65-pL reactors separately. This new method greatly improves the ease of single-cell RNA sequencing by reducing the experimental steps. Meanwhile, with less chance to induce errors, this method can easily maintain the quality of single-cell sequencing. In addition, this polyadenylate-tail-independent method can be seamlessly applied to prokaryotic cell RNA sequencing.



RNA sequencing (RNA-seq) has become a routine method to quantitatively assess the whole transcriptome landscape in cells or tissues.[1] However, for those samples that contain scarce amount of RNAs, such as single cells, the cDNA must be amplified after reverse transcription (RT) in order to yield a large enough amount of identical copies of each cDNA fragments for the construction of sequencing libraries.[2] In order to deal with the various sources of error induced through cDNA amplification, the chemistry of single-cell RNA-seq process has been continuously evolving in the past decade along with the rapid development of next-generation sequencing technology.[2−6]

Recently a few single-cell RNA-seq chemistries have been reported,[2−9] and greatly extended our understanding of the biological complexity and heterogeneity of gene expression at the single cell level.[10,11] These chemistries mainly focus on detecting mRNA with polyadenylate (poly(A)) tails at their 3′-ends, with the use of oligo(dT) primers for RT to avoid amplifying highly abundant rRNA (rRNA).[12] An inevitable result of such an approach is the loss of information from the nonpolyadenylated RNA molecules, which include many long noncoding RNA, some mRNA, and circular RNA in mammalian cells. The incapability to amplify the non-polyadenylated RNA also makes these methods not applicable to prokaryotic cells, in which the mRNA molecules are not polyadenylated.

A feasible way to capture the total RNA, both with and without poly(A) tails, from a single cell is to use random primers for RT. With this approach, however, cDNA will be overwhelmed by the RT product from abundant rRNA and the consequent amplification will cause a loss of the relatively lower expressed mRNA and other RNA species. Another approach is preamplification rRNA depletion, which has been proved effective for bulk RNA samples[13] but not for single-cell samples due to increased bias and the random loss of mRNA molecules during the depletion process, creating difficulties in detecting low-abundance mRNA as well.

Here we present a novel method, named "emulsion-based amplification of sequence independent evenly transcribed RNA sequencing (easier-seq)", to amplify and sequence total RNA of single cells. This method uses random hexamers (N_6) as RT primers to capture total RNA, independent of their sequence composition and with high uniformity across the whole length of transcript. Following RT, a microfluidic device is employed to distribute cDNA product into 1.5×10^5 65-pL droplets for isothermal amplification. By segregating individual amplification reactions, this method improves the even amplification of low-abundance RNA, demonstrates even coverage along the whole-length transcripts, and thus effectively suppresses the amplification bias across RNA molecules. The application of isothermal multiple displacement amplification (MDA) not only guarantees high gain of cDNA amplification in emulsion but also decreases the difficulty of single cell whole-transcriptome analyses due to the simplified hands-on operations.

Received: July 7, 2016
Accepted: October 26, 2016
Published: October 26, 2016

DOI: 10.1021/acs.analchem.6b02581
Anal. Chem. 2016, 88, 10795−10799

The experimental process (Figure 1) of easier-seq, to our knowledge, is the simplest among all available chemistries of

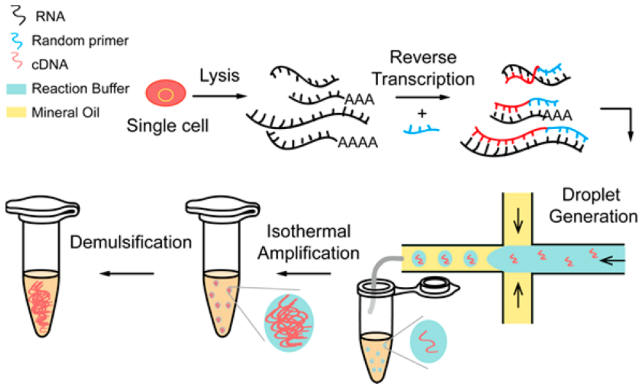


Figure 1. Experiment process of easier-seq.

single-cell RNA-seq. It contains only three steps of operation: cell lysis (~5 min), reverse transcription (~55 min), and isothermal emulsion amplification (>6 h). In our experiments, we picked single cells manually through mouth-pipetting and used PBS-BSA buffer to wash the cells at least three times before lysis to avoid possible extracellular contamination. Lysis buffer contained NP-40, RNase inhibitor, and N_6. Lysis was performed at 70 °C for 2 min with a total volume no more than 5 μL, and then the whole lysate was cooled to 4 °C for priming. To improve the capture sensitivity during RT, we slowly ramped the annealing temperature from 25 to 60 °C after adding the reverse transcriptase.

In the easier-seq method, N_6 oligonucleotide is a key reagent since it serves as a universal primer for both RT and MDA reactions. Although the oligo(dT) primer has been successfully used as RT primers in most single-cell RNA-seq chemistries, we have indicated that the primer with partially random sequences could significantly expanded the coverage to nonpoly(A)-tailed RNA species in single cells.[14] We also learned from single-cell emulsion whole genome amplification (eWGA) that N_6 primers efficiently and evenly amplified DNA fragments in separated picoliter reactors.[15] With easier-seq, we expected that replacing PCR with MDA, sharing the N_6 primers used in the poly(A)-independent RT step, can eliminate the need of two sets of primers that are commonly required by existing single-cell RNA sequencing chemistries.

During the RT step of easier-seq, the N_6 primers provide sufficient diversity to capture most RNA transcripts at various positions and hence produce polyA-independent first-strand cDNA molecules. In the isothermal amplification step, after a second round of heating and cooling, excess N_6 primers can bind to the cDNA without the need of RNA digestion. After priming, we added phi29 DNA polymerase into the reaction solution, at low temperature to prevent the amplification reaction from starting and then immediately split this solution into monodispersed microdroplets (50 μm in diameter, 65 pL in volume) with a microfluidic device. We collected this water-in-oil emulsion in a 2 mL centrifuge tube and then kept it at 30 °C for at least 6 h for amplification. According to our previous experience with single-cell eWGA,[15] we expected this easier-seq method to benefit from the emulsion-based compartmentalization of amplification reactions. Independent amplification of each molecule inside their respective droplet enabled the drastically reduced disturbance from other molecules while preserving the high gain of MDA.[16,17] Therefore, emulsion MDA in easier-seq is an extremely efficient for amplification of cDNA fragments. For each single cell, postamplification demulsification and purification of DNA resulted in a sufficient amount of amplified product, typically over a few micrograms, for construction of sequencing libraries and other downstream analysis.

We found that simply replacing the PCR with emulsion MDA in Tang09 protocol,[2] i.e., using oligo dT primers for RT and then performing MDA in a single tube (see Figure S1 in the Supporting Information), could significantly improve the final sequencing performance. One of the major improvements is the significantly lower discordance rate (from 0.74% to 0.07%) when applying MDA using Phi29 DNA polymerase (Figure 2a), which has very high fidelity due to its proofreading

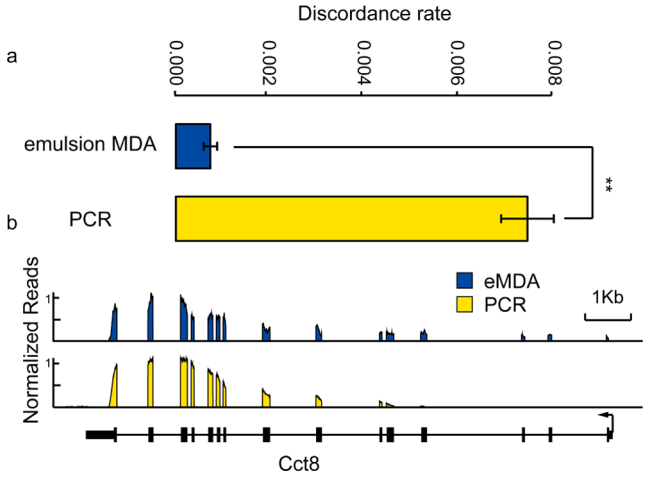


Figure 2. Effect of replacing PCR amplification with emulsion MDA. (a) The discordance rate of amplification is lowered when using Phi29 polymerase. (b) The coverage of 5′-end of Cct8 transcript amplified with eMDA is superior.

activity. Another notable advantage of using MDA over PCR is the better sequencing coverage of 5′-ends of transcripts (Figure 2b). The Taq DNA polymerase in PCR-based Tang09 protocol tends to lose the 5′-ends of many transcripts and causes strong 3′-enrichment in the sequencing reads.

We further simplified the whole process by replacing the RT primers with N_6, which not only made the RNA capture more complete but also allowed for skipping the primer digestion step that many other methods required (see Figure S1 in the Supporting Information). We tested the dose effect of N_6 in RT (Figure 3, and see Table S1 in the Supporting Information) and found that the optimal RT primer concentration was slightly dependent on the starting amount of RNA in a single cell. For large cells such as mouse embryonic stem cells (mESCs), a final N_6 concentration of ~60 μM was preferred; while for smaller cells such as 3T3L1 fibroblasts, a lower concentration (~25 μM) was sufficient (see Figure S2 in the Supporting Information). As higher concentration of primers could cause more nonspecific amplification of MDA, the lowest concentration of primers tested with successful amplification was used. We also verified that we could replace the RNase H digestion step with heat to separate the DNA-RNA hybrids (see Figure S3 in the Supporting Information).

The use of random primers in reverse transcription, combined with the high processivity of Phi29 DNA polymerase

DOI: 10.1021/acs.analchem.6b02581
Anal. Chem. 2016, 88, 10795−10799

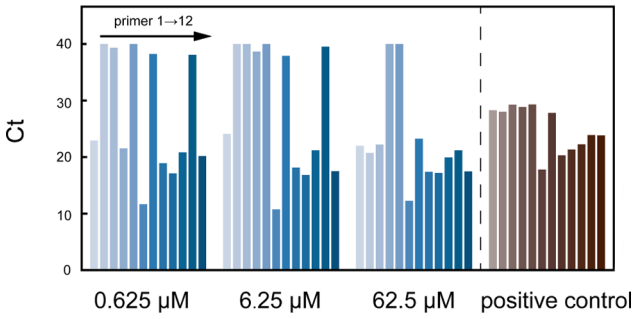


Figure 3. Dose effect of random primers concentration used in RT process of mESC single cell experiment compared to bulk RNA RT process as positive control using realtime PCR.

and the diversity of N_6 primers in MDA, resulted in further improvements in the uniformity and coverage of different transcripts (Figure 4a and see Figure S4 in the Supporting Information). Unlike the Tang protocol or similar approaches, which use poly-A tailing and oligo(dT) to prime the second strand synthesis and cause significant imbalanced coverage (the 3′-bias) across the transcript, easier-seq provides much evener coverage distribution along the full length of the transcripts. In addition, a higher coverage of the long transcripts (>4 kbp) can be sequenced using easier-seq (Figure 4b and see Figure S5 in the Supporting Information), while the coverage is comparable between different methods for the short transcripts (see Figure S6 in the Supporting Information).

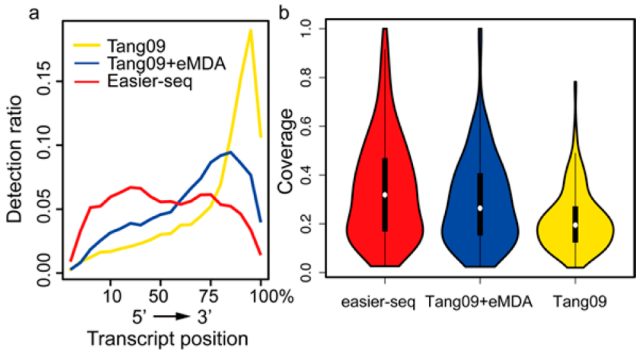


Figure 4. Easier-seq capture transcripts with higher completeness: (a) easier-seq amplified single cell sample have more uniform coverage distribution across transcripts with the length between 3 and 4 kbp and (b) easier-seq could detect higher coverage of long transcripts (>4 kbp) than other methods.

The coverage uniformity of easier-seq is very similar to that of Smart-seq, another popular single-cell RNA-seq chemistry, in which a template switching approach is used to add universal adaptors at the 5′-ends in order to amplify the full-length transcripts. However, Smart-seq suffers the partial loss of long transcripts when the universal adaptor fails to add and the tendency to amplify short transcripts during PCR, resulting in a lower detectable gene number compared to the Tang09 protocol. End-point quantification of amplified cDNA products confirmed that MDA outperformed PCR, with a typical amplification gain over 10^6 for single mammalian cells. But as MDA's low efficiency in amplifying very short templates, we found that easier-seq might cause slight loss of short transcripts (see Figure S7a in the Supporting Information). By separating different transcripts into emulsion droplets, we have improved the amplification evenness for lowly expressed genes (see Figure S7b,c in the Supporting Information).

With the above optimizations, easier-seq method now consists of only two major experimental operations, RT and emulsion MDA. As the simplest method of single-cell RNA-seq, it has greatly reduced the experimental labor and time (for a comparison, see Figure S1 in the Supporting Information) and significantly reduced the possibility of experimental errors.

We validated the reproducibility of easier-seq by processing four single mESCs and three 8-pg RNA input samples that were diluted from bulk mESC RNA. We chose these samples to validate easier-seq because mESCs are known to have low cell-to-cell transcriptome variability, and the 8-pg bulk RNA samples serves as a technical replicate which mimics the RNA content of a small-sized single cell.[9] The high correlation coefficients (Figure 5 and see Figure S8 in the Supporting Information, r = 0.885−0.996) between single cell samples confirmed the reliability of this new method. We identified ∼6 000 genes from single mESC and ∼4 000 genes from each 8-pg bulk RNA sample (FPKM ≥ 1, see Table S2 in the Supporting Information) with about 0.5 M mappable reads per sample. Between the replicates of the same cell type, those genes are highly correlated (see Figure S8 in the Supporting Information).

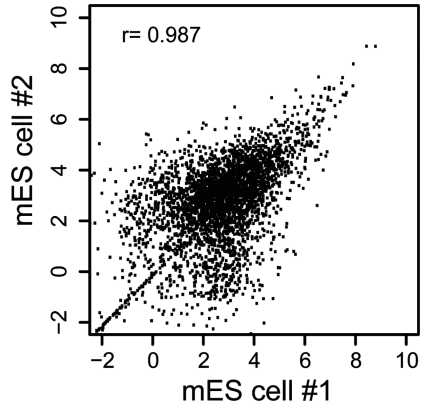


Figure 5. Reproducibility of easier-seq: The expression level of different transcripts are highly reproducible between two single cell experiments.

We then processed four single 3T3L1 fibroblasts using easier-seq. The titration of RT primer concentration (see Figure S2 in the Supporting Information) supported our previous observation that an overdose of RT primer would not benefit the amplification performance but resulted in shorter cDNA product and interfered with the subsequent MDA reaction. In each 3T3L1 single cell, we detected about 4 000 genes. This number is smaller than that of mESC because fibroblasts are smaller in size and contain less RNA molecules in each cell. Between single 3T3L1 cells, the identified genes strongly overlapped, and the correlation between expression profiles was high (see Figure S8 in the Supporting Information). Principal component analysis (PCA) of these samples clearly showed the cell-type dominant clustering (see Figure S9 in the Supporting Information). Using the differentially expressed genes identified through PCA, we further illustrated the intersample variation in a heat-map (Figure 6). The clear distinction between fibroblast and mESC highlights the genes that are specifically expressed in each cell type, while

DOI: 10.1021/acs.analchem.6b02581
Anal. Chem. 2016, 88, 10795−10799

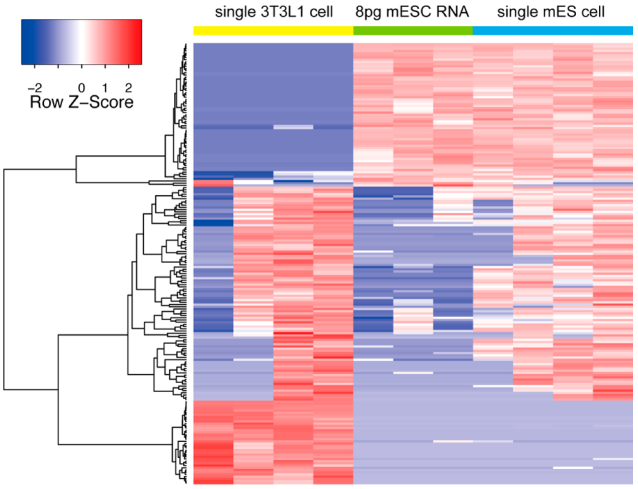


Figure 6. Unsupervised hierarchical clustering of different samples amplified by easier-seq with differentially expressed genes identified through PCA.

within the mESC samples the single cells and 8-pg bulk RNA samples exhibit very similar transcriptome patterns.

We then checked the sequencing data and focused on the nonpoly(A)-tailed RNA species captured by N_6 RT primers. With easier-seq, there is a larger portion of rRNA sequences in the final products (6−88%, sample-dependent, details shown in Table S3 in the Supporting Information; while using oligo-dT RT primers a typical rRNA ratio is less than 15%). More investigations are required in the future to facilitate the understanding of rRNA ratio difference between samples. Despite this nonideal ratio of rRNA reads, easier-seq enables detection of "dark matter" RNAs in single cells or samples with low-input material, which are neglected by other methods. We compiled a selected list of high confidence transcripts without poly(A) tails by screening the published data sets[18] (see experimental methods in the Supporting Information), and then plotting the detection state of these transcripts in single cell experiment using different RT primers (Figure 7). Using of N_6 indeed captured more such RNA molecules that might offer critical information to better understand gene regulation at single-cell level.

Since easier-seq is poly(A)-independent, we have further tested if this method is suitable for sequencing picogram-level

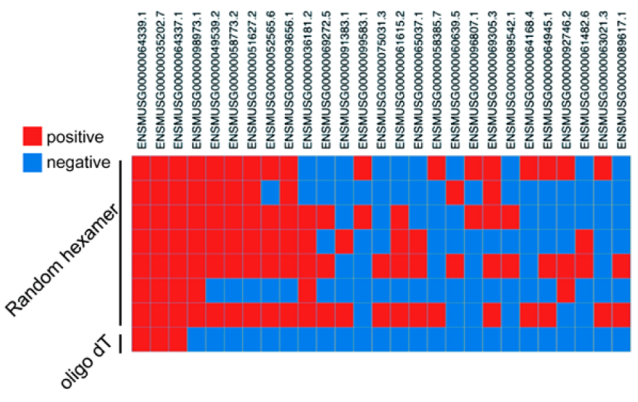


Figure 7. Detectability of high confidence transcripts (with relatively higher expression in the total RNA sequencing samples than in the rRNA depleted ones) without poly(A) tails using different RT primers for single cell samples.

prokaryotic mRNA, which does not contain poly(A) tails and hence becomes a great challenge in low-input RNA-seq. We extracted the RNA from bulk amount of *E. coli* and picked 10 pg (equivalent to ∼100 *E. coli* cells) to perform easier-seq for transcriptome analysis. To validate the success of amplification, two groups of control experiments using bulk amount of cDNA, reverse transcribed from 2 ng of *E. coli* RNA using random hexamers, are performed. In control 1, 5% of the bulk cDNA product was used for each qPCR experiment; while in control 2, 0.5% of the bulk cDNA product was first amplified through emulsion MDA and then 5% of amplified product was used for each qPCR experiment (Figure 8, also see

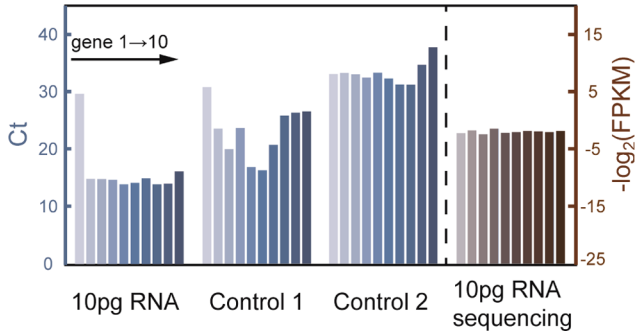


Figure 8. Realtime PCR and sequencing result demonstrates the feasibility of easier-seq to amplify *E. coli* RNA; control 1, nonamplified cDNA; control 2, eMDA amplified 10-pg *E. coli* cDNA; $-\log_2$(FPKM) is used for easy comparison.

experimental methods in the Supporting Information). Both real-time PCR (see Table S4 in the Supporting Information) and sequencing analyses demonstrated that easier-seq provided uniform amplification across the whole transcriptome, indicating that it has potential to become a powerful technique to unveil the transcriptional networks in small prokaryotic populations.

In conclusion, we have developed a simple yet powerful method, easier-seq, to amplify and sequence total RNA from a single cell using DNA random hexamer as primers for reverse transcription and picoliter emulsion MDA. Through easier-seq, RNA transcripts are more uniformly and accurately amplified with significantly improved coverage of 5′-ends. More kinds of transcripts have been captured by this poly(A)-independent method, making easier-seq applicable to studying nonpoly(A)-tailed mRNA and noncoding RNA, as well as the mRNA in prokaryotes. This feature could eventually open an avenue to probe the transcriptome interplay and dynamics of complex biosystems such as the host−pathogen interaction at single-cell levels *in situ*.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b02581.

Detailed experimental process, bioinformatic analysis, Figures S1−S10, and Tables S1−S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: yanyi@pku.edu.cn.

DOI: 10.1021/acs.analchem.6b02581
Anal. Chem. 2016, 88, 10795−10799

Author Contributions

Y.F and Y.H conceived the project. Y.F and L.L conducted the experiment. Y.F and H.C performed the bioinformatic analysis. Y.F and Y.H wrote the manuscript.

Funding

This work was supported by Ministry of Science and Technology of China (Grant 2015AA0200601) and National Natural Science Foundation of China (Grants 21327808 and 21525521).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Yun Zhang and the Peking University High-throughput Sequencing Center for the experimental assistance.

■ REFERENCES

(1) Wang, Z.; Gerstein, M.; Snyder, M. *Nat. Rev. Genet.* **2009**, *10* (1), 57−63.
(2) Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B. B.; Siddiqui, A.; Lao, K.; Surani, M. A. *Nat. Methods* **2009**, *6* (5), 377−382.
(3) Kang, Y.; Norris, M. H.; Zarzycki-Siek, J.; Nierman, W. C.; Donachie, S. P.; Hoang, T. T. *Genome Res.* **2011**, *21* (6), 925−935.
(4) Ramsköld, D.; Luo, S.; Wang, Y.-C.; Li, R.; Deng, Q.; Faridani, O. R.; Daniels, G. A.; Khrebtukova, I.; Loring, J. F.; Laurent, L. C.; Schroth, G. P.; Sandberg, R. *Nat. Biotechnol.* **2012**, *30* (8), 777−782.
(5) Hashimshony, T.; Wagner, F.; Sher, N.; Yanai, I. *Cell Rep.* **2012**, *2* (3), 666−673.
(6) Sasagawa, Y.; Nikaido, I.; Hayashi, T.; Danno, H.; Uno, K. D.; Imai, T.; Ueda, H. R. *Genome Biol.* **2013**, *14* (4), R31.
(7) Picelli, S.; Björklund, Å. K.; Faridani, O. R.; Sagasser, S.; Winberg, G.; Sandberg, R. *Nat. Methods* **2013**, *10* (11), 1096−1098.
(8) Islam, S.; Zeisel, A.; Joost, S.; La Manno, G.; Zajac, P.; Kasper, M.; Lonnerberg, P.; Linnarsson, S. *Nat. Methods* **2013**, *11* (2), 163−166.
(9) Streets, A. M.; Zhang, X.; Cao, C.; Pang, Y.; Wu, X.; Xiong, L.; Yang, L.; Fu, Y.; Zhao, L.; Tang, F.; Huang, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (19), 7048−7053.
(10) Buganim, Y.; Faddah, D. A.; Cheng, A. W.; Itskovich, E.; Markoulaki, S.; Ganz, K.; Klemm, S. L.; van Oudenaarden, A.; Jaenisch, R. *Cell* **2012**, *150* (6), 1209−1222.
(11) Deng, Q.; Ramskold, D.; Reinius, B.; Sandberg, R. *Science* **2014**, *343* (6167), 193−196.
(12) Kolodziejczyk, A. A.; Kim, J. K.; Svensson, V.; Marioni, J. C.; Teichmann, S. A. *Mol. Cell* **2015**, *58* (4), 610−620.
(13) Cui, P.; Lin, Q.; Ding, F.; Xin, C.; Gong, W.; Zhang, L.; Geng, J.; Zhang, B.; Yu, X.; Yang, J.; Hu, S.; Yu, J. *Genomics* **2010**, *96* (5), 259−265.
(14) Fan, X.; Zhang, X.; Wu, X.; Guo, H.; Hu, Y.; Tang, F.; Huang, Y. *Genome Biol.* **2015**, *16* (1), 236.
(15) Fu, Y.; Li, C.; Lu, S.; Zhou, W.; Tang, F.; Xie, X. S.; Huang, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (38), 11923−11928.
(16) Nakamura, K.; Iizuka, R.; Nishi, S.; Yoshida, T.; Hatada, Y.; Takaki, Y.; Iguchi, A.; Yoon, D. H.; Sekiguchi, T.; Shoji, S.; Funatsu, T. *Sci. Rep.* **2016**, *6*, 22259.
(17) Sidore, A. M.; Lan, F.; Lim, S. W.; Abate, A. R. *Nucleic Acids Res.* **2016**, *44* (7), e66−e66.
(18) Zhang, X.-O.; Yin, Q.-F.; Chen, L.-L.; Yang, L. *Genomics Data* **2014**, *2*, 237−241.

DOI: 10.1021/acs.analchem.6b02581
Anal. Chem. 2016, 88, 10795−10799

Supporting Information

Single cell total RNA sequencing through isothermal amplification in picoliter-droplet emulsion

Yusi Fu[†], He Chen[†], Lu Liu[†‡], and Yanyi Huang[*†‡ǁ]

[†]Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, and Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing 100871, China

[‡]College of Engineering, Peking University, Beijing 100871, China

[ǁ]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.

* Email: yanyi@pku.edu.cn

This Supporting Information file includes detailed experimental process and bioinformatic analysis, Figures S1-S10, Tables S1-S4.

Materials and methods

Cell culture

The 3T3L1 cells were purchased from the Cell Bank of the Chinese Academy of Sciences (http://www.cellbank.org.cn) and cultured at 37℃ in a humidified incubator containing 5% CO_2. These cells were passaged every 3 days. To passage, adherent 3T3L1 cells were detached by 0.25% trypsin with 0.1% EDTA (Invitrogen) and washed by PBS twice to flush out the dead ones. After centrifugation at 1200 rpm for 3 min, the supernatant was discarded and re-suspended cells would be available for continued culturing or single cell isolation for downstream experiment. Mouse embryonic stem cells (mES cells) were kindly provided by Prof. Fuchou Tang in the School of Life Sciences at Peking University.

Single cell isolation, lysis, and reverse transcription

Cells were diluted using PBS and pipetted gently to obtain a single cell suspension. About 10μl PBS was separated into several drops in a petri dish and then ~2μl diluted cell suspension was pipetted into one of the drops. With only a few cells in one drop, each single cell was identified under a microscope and then picked using a mouth pipet. After repeated washing (at least four times) with clean PBS drops, we used a new mouth-pipetting capillary to transfer a single cell to the lysis buffer. We changed the mouth pipet tip as often as possible to make sure only one cell was picked. The success of transfer was confirmed under stereoscope to ensure that there was no cell left in the PBS drop. By this stringent process, we were able to make sure only one cell had been picked into the lysis buffer. Lysis buffer (4.05 μl) consists of 1×PCR Reaction buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.05% NP-40, 5 mM DTT (Invitrogen), 0.045 μl Rnase Inhibitor (Ambion), 0.045 μl SUPERase In (Ambion), 22.2 μM N_6 primers, 0.05 mM dNTP. Lysis was carried out at 70 ℃ for 90 s, and followed by reverse transcription by adding 0.33 μl Superscript Ⅲ reverse transcriptase (Invitrogen), 0.05 μl Rnase Inhibitor (Ambion) and 0.07 μl T4 gene 32 protein (Roche), with the appropriate amount of N_6 primers. The whole RT process was performed at 25 ℃ for 10 min, then 50 ℃ for 30 min and finally 70 ℃ for 15 min.

Easier-seq amplification

MDA reaction buffer was added to reverse transcribed cDNA solution, which contains 1×Phi29 buffer (NEB), 50 μM N_6 primers (Invitrogen), and 1 mM dNTP (NEB). After heating to 95 ℃ for 5 min to separate the cDNA and the RNA, the tube was put on ice for at least for 20 min for N_6 primers to bind to the cDNA. We then added 8 units of Phi29 polymerase (NEB) and briefly centrifuged. Then proceed to the droplet generation step.

Droplets were generated by commercially available microfluidic chips (Dolomite, UK) as shown in Supplementary Figure S10. We use mineral oil (SIGMA M3516), supplemented with 4.5% Span80 (SIGMA S6760), 0.4% Tween80 (SIGMA P8074) and 0.05% Triton X-100 (Beyotime ST795) in volume, as continuous phase. Stable compressed air was used to push the reaction buffer and mineral oil forward, which went through the cross section of the device to generate w/o droplets with ~50 μm in diameter. To guarantee full recovery of DNA, reaction buffer was followed by 2 μl mineral oil into a 200-μl collecting tube to guarantee all the buffer were emulsified. During the whole process, an ice box or ice bag was used to keep the reaction buffer at around 4 ℃ both before and after the droplet generation, so that the amplification reaction wouldn't start in advance. The droplet generation step is described in detail in our previous paper[1].

Tang protocol single cell RNA amplification

Tang protocol was carried out as described in the original paper[2].

Demulsification and DNA purification

The stability of emulsion was visually confirmed under a microscope in order to guarantee that droplets wouldn’t merge after the 10-hour-30℃-incubation. 65℃-heat inactivation was carried out to stop amplification. We then added 700 μL isobutanol into tubes following ~30s fierce vortexing until the solution was clear to complete demulsification. To obtain purified DNA samples, 70 μL Binding Buffer (Zymo Reaearch) was added with centrifugation at 1700 g for 3 min. After discarding the upper layer, the remaining solution was purified by DNA (PCR) Clean-up &

Concentration kit (Zymo Research) following the manufacture's recommended protocol and was finally eluted with 20 μL water.

Prokaryotic cell RNA amplification and Quality control

ZR Fungal/Bacterial RNA MicroPrep™ kit (Zymo Research) was used to isolate E.coli RNA following the protocol provided along the kit. Then the RNA was diluted and 10 pg RNA was used as template for easier-seq. After amplification, 197.4 ng purified DNA was harvested. For each quantitative PCR experiment 0.987 ng DNA was used. To validate the success of amplification, we performed an emulsion MDA reaction using cDNA as the starting material for comparison. Random hexamers were used as primers to reverse transcribe 2 ng RNA to produce E.coli cDNA. The bulk cDNA was diluted and 10 pg cDNA was used as template for emulsion MDA amplification. After purification, 4.6 ng DNA was harvested in total. As the resulting DNA was very limiting, 0.23 ng product DNA was used for quantitative PCR quality control. The unamplified 2 ng cDNA was used as a positive control in realtime PCR with 100-pg input for each qPCR experiment.

Quality control qPCR

Purified DNA was first quantified by Qubit dsDNA HS Assay (Invitrogen) and then quantitative PCR was performed to examine the amplification bias. Briefly, qPCR was set up by adding 5 μL PCR SsoAdvanced SYBR Green Supermix (Biorad), 0.5 μL of 10 μM forward- and revers-primer, 1 μL template and 3 μL water. Then qPCR was carried out in the Illumina Eco thermocycler with melting curve analysis. We checked the melting curve to make sure the result was not due to the formation of primer dimers. We also performed negative controls (no DNA) and positive control (1 ng cDNA) along with experimental samples to ensure the lack of contamination and accuracy of qPCR. Amplification was considered successful when at least 70% of the primers pairs used can result in correct products (confirmed by melting curve analysis) and Ct value lower than 30.

Preparation of sequencing libraries

We used two library preparation protocols. For mES cell and the purified RNA sample, with each amplified sample, 100 ng DNA was used to build the sequencing library for Illumina platform using NEBNext Ultra DNA Library Prep Kit (NEB). For 3T3 cells, 20 ng DNA was used to build the sequencing library using TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme). The libraries were sequenced by Illumina Hiseq2500 or MiSeq sequencers. We did not find significant difference between these two methods.

Quality control and expression calculation

Sequencing data quality control was described in our previous paper[3]. Clean data was first mapped to the mm9 reference genome using Tophat2[4] and the expression level of different transcripts were calculated using Cufflinks[5] and annotation from GENCODE M2 annotation.

Coverage distribution calculation and discordance rate calculation

The coverage distribution of different transcripts were calculated from Tophat2 mapped data and the annotation data. Bedtools[6] was used to calculate the depth of each position of each transcripts, transcripts that overlapped with each other were deleted to avoid multiple calculation. Transcripts were divided into 20 even parts, the ratio of each part was calculated through divided the sum of base depth in each part to the total depth of transcript.

The discordance rate was calculated by sum of the depth the less abundance base composition in heterogeneous site divided by the total depth of all transcripts. It should be noticed that the real heterogeneous site could also be counted as error, which were neglected compared to the amplification error rate.

non-poly(A)-tailed RNA identification and ribosome RNA ratio calculation

A published RNA-seq dataset of mouse ES cell were downloaded and used for analysis (GSE53942). The datasets contain transcripts that were known to be with or without poly(A) tail. Reads were mapped to mm10 genome by STAR, and gene expression level was calculated by Cufflinks using GENCODE database.

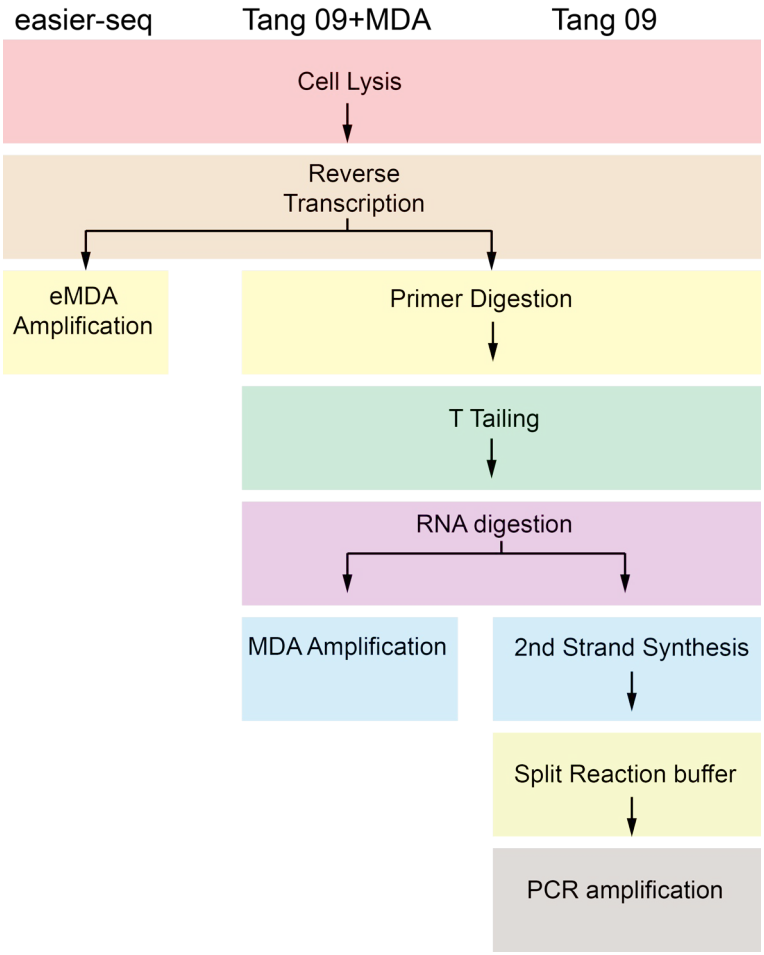
non-poly(A)-tailed genes were identified with the following three criteria: genes that expressed (FPKM>1) in all non-poly(A) samples; genes that have low expression level (FPKM<5) in all poly(A) tail samples; genes have 2-fold expression level in library without poly(A) tail samples than poly(A) tail samples. Finally 424 out of 15009 transcripts were picked out, it should be notice that the high confidence means a relative higher expression in the total RNA sequencing sample than the rRNA depleted ones. It does not necessarily mean that these transcripts can be detected with high confidence, especially in single cells.

To check the expression level of these genes in our easier–seq data, HTseq[7] were used to count the reads number in these non-poly (A) genes.

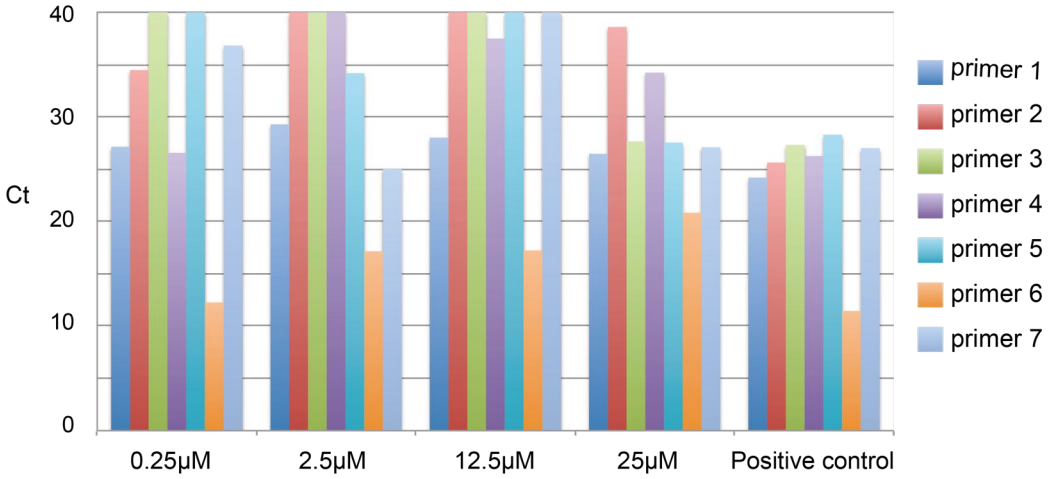
Ribosome RNA (rRNA) were calculated as described in our previous paper[8].

Prokaryotic cell RNA sequencing analysis

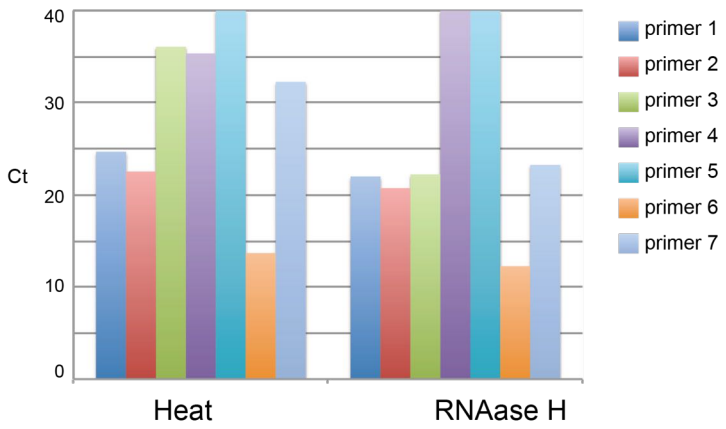
Reads were mapped to K12 genome (NC_000913.3) by bowtie2, at most one mismatch for each mapped reads was allowed. Only reads with MAPQ>20 were kept for further analysis. Of all the 1,584,514 reads sequenced, under this criteria, a total of 939,223 reads with 100bp in length were mapped to E.coli transcriptome. Then gene expression level was calculated by Cufflinks[5]. We found prokaryotic cells with quite low rRNA ratio (around 0.8%). We also processed the published data of bulk input with the same pipeline, and found quite large ratio range, from 0.7% to 36%. We have not fully understand the reason of this variation.



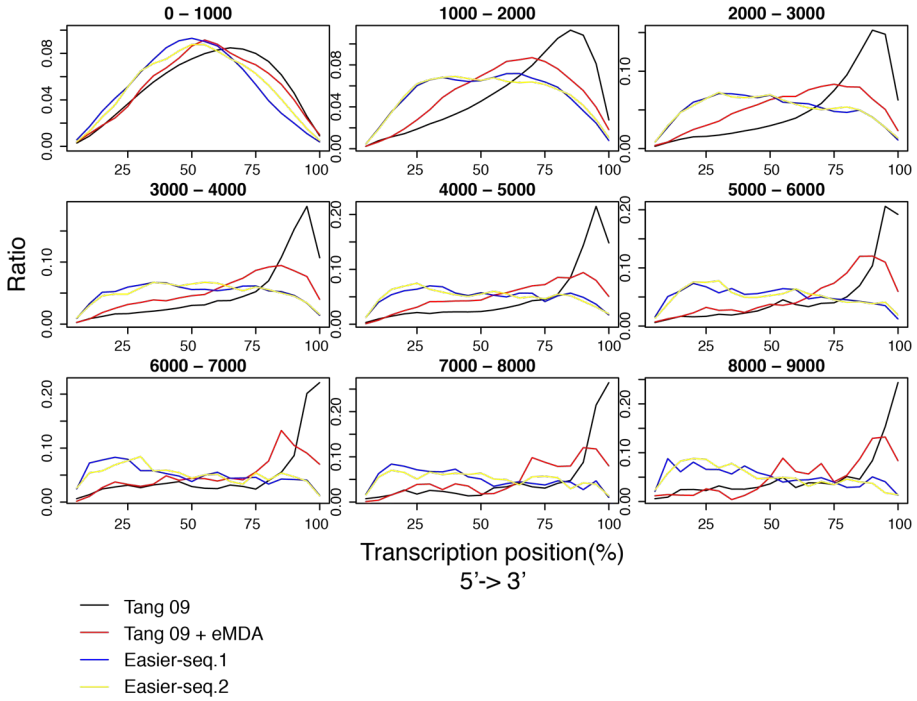
Supplementary Figure S1. The comparison of experimental steps between different single cell RNA-seq methods.



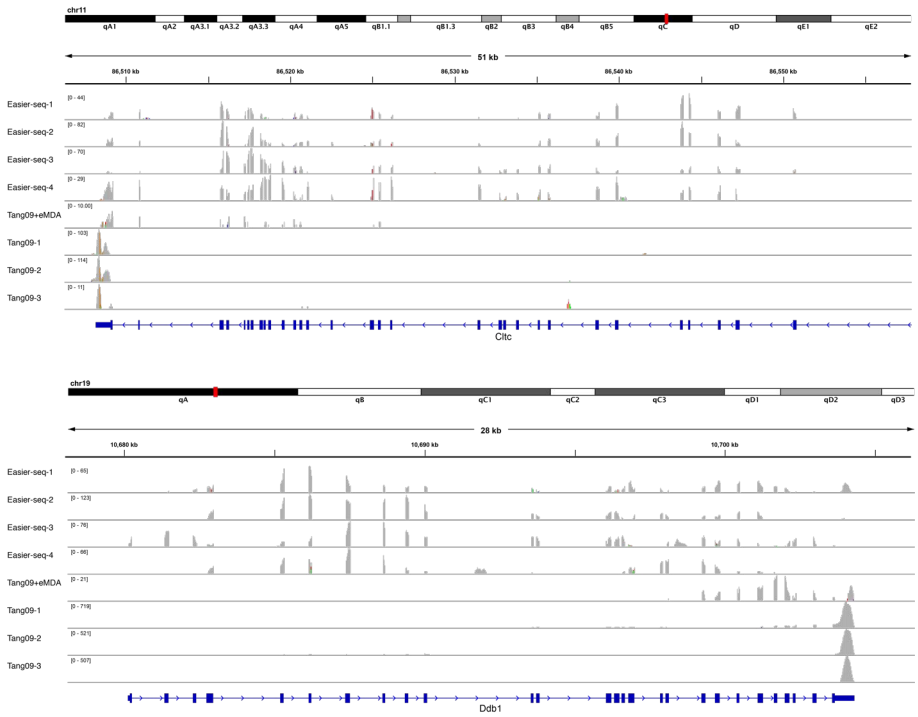
Supplementary Figure S2. The real-time PCR result of dose effect of the random primers used in RT process for 3T3L1 single cell experiment, positive control represents 3T3L1 unamplified cDNA and 0.2ng was used for each real-time PCR experiment. The amplification is considered successful when at least 70% of the primer pairs can be amplified with correct products through the melting curve analysis, and with Ct value lower than 30.



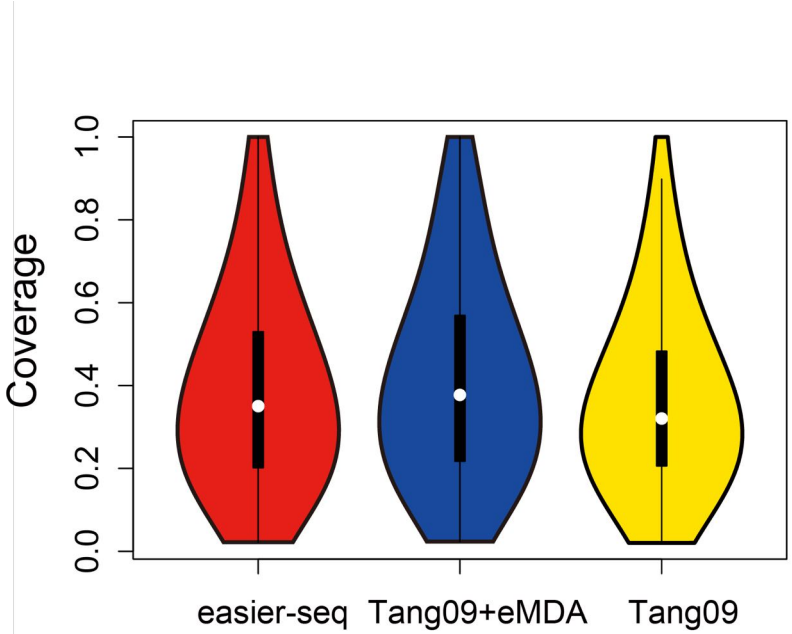
Supplementary Figure S3. The comparison of amplification result between heat and RNAse H digestion method to separate cDNA single strand from RNA-cDNA hybrid using real-time PCR result.



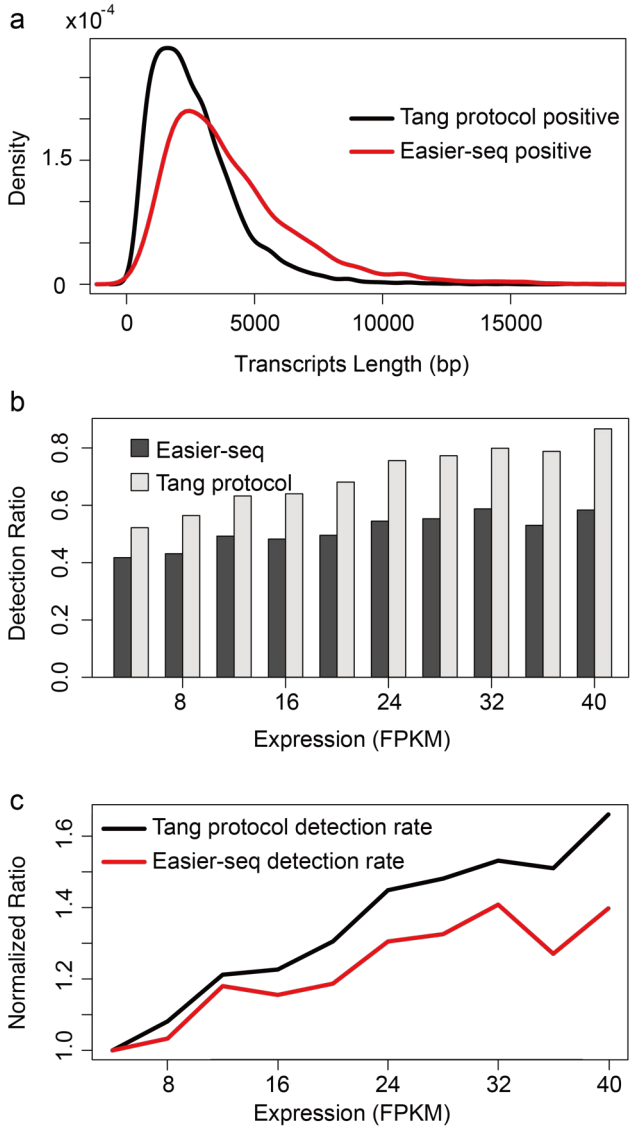
Supplementary Figure S4. The distribution of coverage of reads within different length range of transcripts. Transcripts of different length range were divided into 20 even parts, the ratio was calculated by dividing the depth of one part to total depth. The curves are mean values of the ratio.



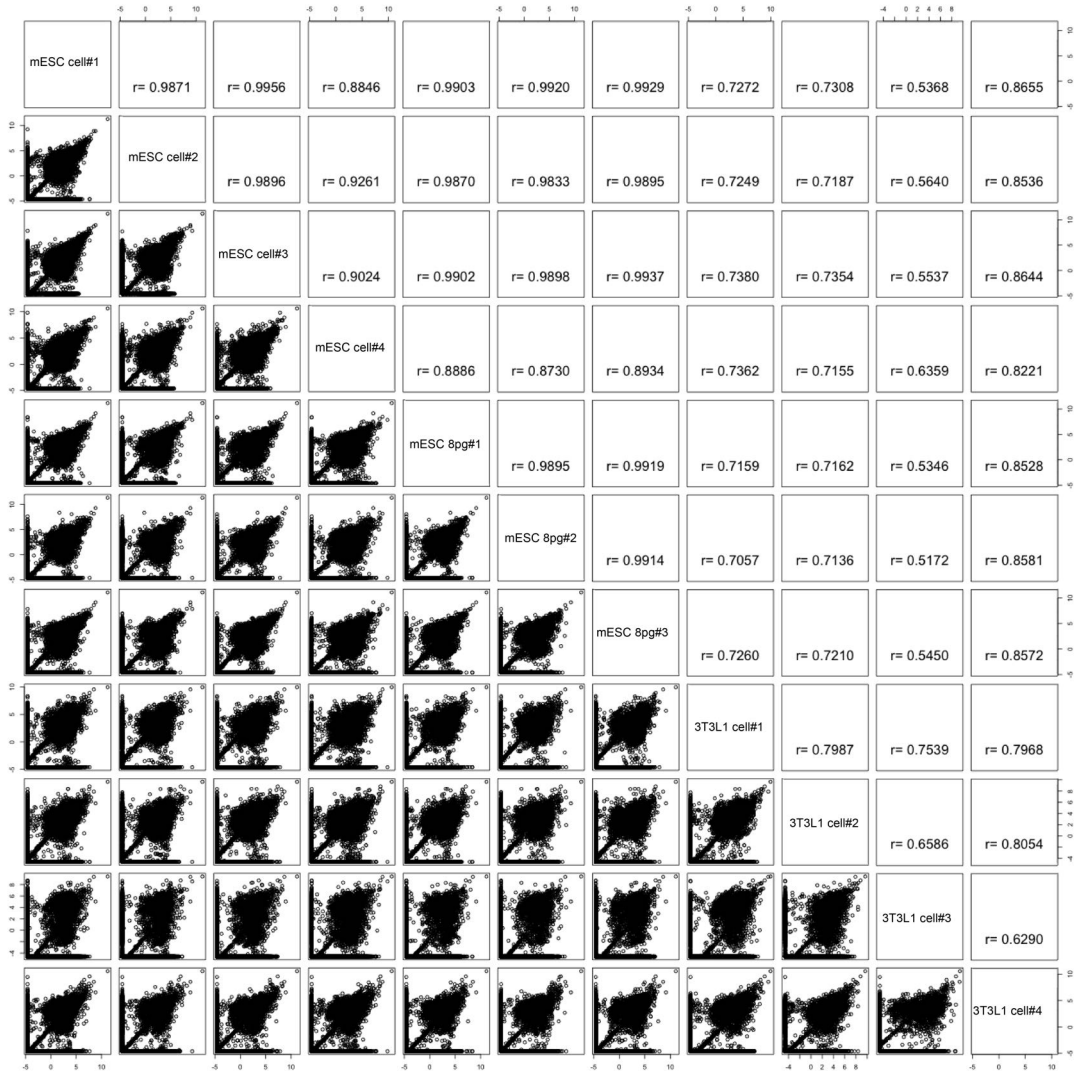
Supplementary Figure S5. More uniform coverage of transcripts for single cell samples amplified by easier-seq.



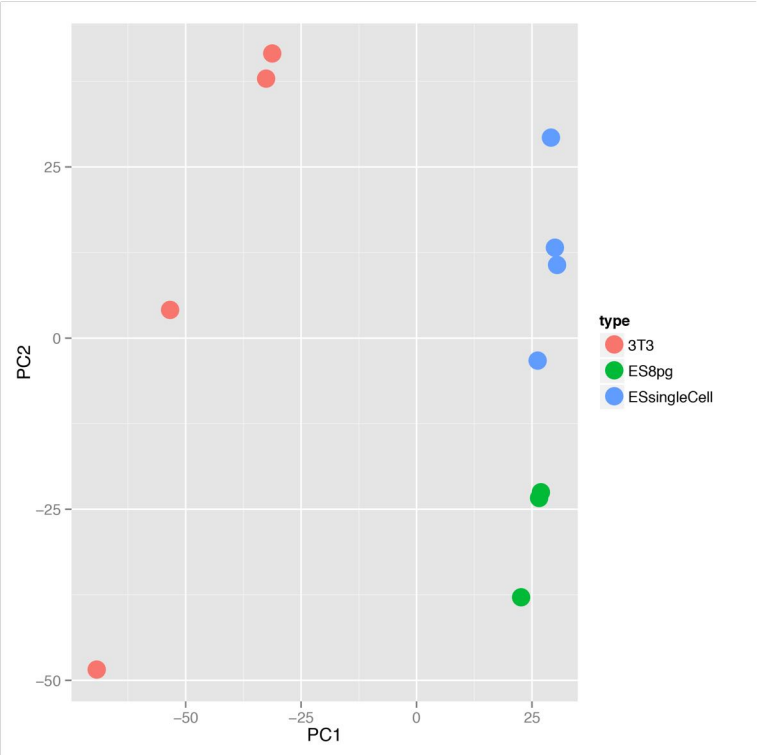
Supplementary Figure S6. The coverage detection ability is comparable between different methods for the short transcripts cut-off (2 kbp).



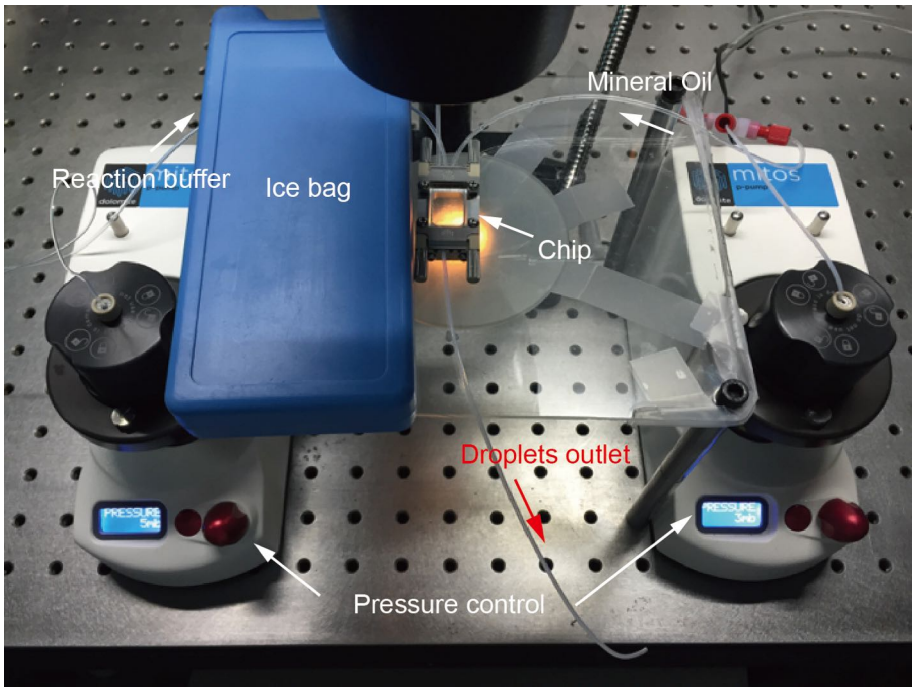
Supplementary Figure S7. The loss of transcripts in easier-seq is related to transcript length, not the abundance of them. a) The transcripts that can only be detected through Tang protocol are shorter than the transcripts that are specifically detected through easier-seq protocol. b) Lowly expressed transcripts (FPKM < 40) were split according to their expression level for each sample, the detection ratio is the percentage of the transcripts detected using one protocol compared to that using the other protocol. For Tang protocol amplified sample, higher detection ratio can be achieved with the increased expression level. However, the detection rate of easier-seq stays almost unchanged. c) We normalized the above detection rates according to the lowest expressed set of genes in each method for better comparison. All data are obtained from single cell experiments and the comparison was done with experiments that recovered medium number of genes for single cells.



Supplementary Figure S8. Correlation between different experiments using easier-seq.



Supplementary Figure S9. PCA analysis of different single cells amplified with easier-seq.



Supplementary Figure S10. Droplet generation device.

Supplementary Table S1. qPCR primers used for quality control of mouse cell RNA sample.

	sequence	amplified gene	primer name
forward primer	aagaaaagatacggcggctgac	rpap3 3' end	primer 1
reverse primer	taatgtacacagaggcggcaac		
forward primer	tgctgcaggagaccaaagagtt	uso1 3' end	primer 2
reverse primer	gcgatggtgctattggatgagt		
forward primer	aagagcaactgctcagggttca	uso1 middle	primer 3
reverse primer	ctgggaaagaatgttggtgcat		
forward primer	agtgaggatttgggaagccagt	uso1 5' end	primer 4
reverse primer	cgtggaaatcaaactcctccaa		
forward primer	cagcaaggacatttttgggaaa	rpap3 middle	primer 5
reverse primer	tcaaagtcttgcttggcctcat		
forward primer	cggggaggtagtgacgaaaaat	rRNA	primer 6
reverse primer	agggcctcgaaagagtcctgta		
forward primer	gcgagtaaggcagttgagttgc	rpap3 5' end	primer 7
reverse primer	ttcatgtccttctcccagtgct		

Primers 8-12 have been described in our previous paper[3]

Supplementary Table S2. Summary of sequencing data for easier-seq amplified samples.

	FPKM>0.1	FPKM>1	Mapped reads left(M)	Mapped reads right(M)
mESC cell#1	6225	5850	0.765357	0.732839
mESC cell#2	5024	4675	0.97244	0.93343
mESC cell#3	6419	5822	1.164742	1.121856
mESC cell#4	7620	7032	0.809718	0.794624
mESC 8pg#1	4487	4198	0.620791	0.593698
mESC 8pg#2	4276	3900	0.714769	0.678946
mESC 8pg#3	4694	4175	0.814054	0.782143
3T3L1 cell#1	3934	3844	0.303009	0.306061
3T3L1 cell#2	5473	5315	0.488006	0.496503
3T3L1 cell#3	3058	2618	2.883447	2.95965
3T3L1 cell#4	4715	4493	0.846756	0.832718

Supplementary Table S3. The rRNA ratio of different easier-seq amplified single cell sample.

Sample Name	rRNA ratio(%)
mESC cell#1	83.1
mESC cell#2	86.6
mESC cell#3	89.8
mESC cell#4	46.1
3T3L1 cell#1	11.4
3T3L1 cell#2	10.2
3T3L1 cell#3	6.0
3T3L1 cell#4	28.4

The rRNA ratio is calculated through dividing the number of reads mapped to rRNA by the number of reads mapped to reference genome.

Supplementary Table S4. qPCR primers used for quality control of E.coli RNA sample.

	sequence	primer name
forward primer	cggcgcttcccagaacatca	gene1 (GapA)
reverse primer	ttcggtgtagcccagaacgc	
forward primer	cagcttggcaccctcgacaa	gene2 (GroS)
reverse primer	ctcggtgcgcttgatgacga	
forward primer	aagcgcgtgaagttccagca	gene3 (rpsE)
reverse primer	tgcaccaccggcgatgatac	
forward primer	actgcctcgatcgtgcatgg	gene4 (thrS)
reverse primer	taacaacctgaaccggcgca	
forward primer	ggtttccaggcgttcgcaga	gene5 (tsf)
reverse primer	aacccagaacgtcgccttcc	
forward primer	ccgcaaatctccactggcga	gene6 (adk)
reverse primer	tctgcctgcggaatggtacg	
forward primer	tcccgtggtaacccgactgt	gene7 (eno)
reverse primer	tgagcgatcgggccgtttac	
forward primer	tgaccgcttcatgtccaccg	gene8 (ppa)
reverse primer	tcgcatcttcaccggcttcg	
forward primer	aaccagaccatcatcgcggg	gene9 (fusA)
reverse primer	accaccagactgtttcgcgt	
forward primer	tgagcgcggtaatggttgca	gene10 (mreB)
reverse primer	tcttcgctgaacaggtcgcc	

Reference

(1) Fu, Y.; Li, C.; Lu, S.; Zhou, W.; Tang, F.; Xie, X. S.; Huang, Y. *Proc Natl Acad Sci USA* **2015**, *112* (38), 11923–11928.

(2) Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B. B.; Siddiqui, A.; Lao, K.; Surani, M. A. *Nat Meth* **2009**, *6* (5), 377–382.

(3) Streets, A. M.; Zhang, X.; Cao, C.; Pang, Y.; Wu, X.; Xiong, L.; Yang, L.; Fu, Y.; Zhao, L.; Tang, F.; Huang, Y. *Proc Natl Acad Sci USA* **2014**, *111* (19), 7048–7053.

(4) Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S. L. *Genome Biol.* **2013**, *14* (4), R36.

(5) Roberts, A.; Trapnell, C.; Donaghey, J.; Rinn, J. L.; Pachter, L. *Genome Biol.* **2011**, *12* (3), R22.

(6) Quinlan, A. R.; Hall, I. M. *Bioinformatics* **2010**, *26* (6), 841–842.

(7) Anders, S.; Pyl, P. T.; Huber, W. *Bioinformatics* **2015**, *31* (2), 166–169.

(8) Fan, X.; Zhang, X.; Wu, X.; Guo, H.; Hu, Y.; Tang, F.; Huang, Y. *Genome Biol.* **2015**, *16* (1), 236.