Integrative Biology

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/ibiology

Insight Box for manuscript ID IB-TIN-03-2015-000075:

Recent studies showed that even isogenic populations of exponentially growing microorganisms carried substantial cell-to-cell heterogeneities at both cellular and molecular levels that are an order of magnitude greater than previously thought. It is increasingly recognized that by using averaged molecular or phenotypic measurements of a whole population for describing cell behaviors, conclusions could be biased as the unique patterns related to specific or distinct functional sub-populations cannot be revealed. Here we report the first bacterial single-cell RNA-seq method for whole transcriptome analysis of single bacterial cells. The study provided a new tool for studying gene-expression heterogeneity in a genetically identical bacterial population, and could be valuable in the areas of microbial ecology and environmental sciences.

RNA-seq based transcriptomic analysis of single bacterial cells

Jiangxin Wang [1, 2, 3], Lei Chen [1, 2, 3], Zixi Chen [1, 2, 3], Weiwen Zhang [1, 2, 3, *]

[1] *Laboratory of Synthetic Microbiology, School of Chemical Engineering & Technology, Tianjin University, Tianjin 300072, P.R. China;* [2] *Key Laboratory of Systems Bioengineering, Ministry of Education of China, Tianjin 300072, P.R. China;* [3] *Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, P.R. China*

* Correspondence to:

Prof. Dr. Weiwen Zhang

Laboratory of Synthetic Microbiology

School of Chemical Engineering & Technology

Tianjin University, Tianjin 300072, P.R. China

Phone: 0086-22-2740-6394

Fax : 0086-22-2740-6364;

Email: wwzhang8@tju.edu.cn

Running title: Transcriptomic analysis of single bacterial cells

Abstract

Gene-expression heterogeneity among individual cells determines the fate of a bacterial population. Here we report the first bacterial single-cell RNA sequencing (RNA-seq), BaSiC RNA-seq, a method integrating RNA isolation, cDNA synthesis and amplification, and RNA-seq analysis of the whole transcriptome of single cyanobacterium *Synechocystis* sp. PCC 6803 cells which typically contain approximately 5-7 femtogram total RNA per cell. We applied the method to 3 *Synechocystis* single cells at 24 h and 3 single cells at 72 h after a nitrogen-starvation stress treatment, as well as their bulk-cell controls for the same conditions, to determine the heterogeneity upon an environmental stress. With 82-98% and 31-48% of all putative *Synechocystis* genes identified in single cells of 24 and 72 h, respectively, the results demonstrated the method could achieve good identification of the transcripts in single bacterial cells. In addition, the preliminary results from nitrogen-starved cells also showed a possible increasing gene-expression heterogeneity from 24 h to 72 h after nitrogen starvation stress. Moreover, preliminary analysis of single-cell transcriptomic datasets revealed that genes from "Mobile elements" functional category has the most significant increase of gene-expression heterogeneity upon stress, which was further confirmed by single-cell RT-qPCR analysis of gene expression in 24 randomly selected cells.

Keywords: single-cell analysis; BaSiC RNA-seq; heterogeneity; *Synechocystis*

Author Summary

Recent studies showed that even isogenic populations of exponentially growing microorganisms carried substantial cell-to-cell heterogeneities at both cellular and molecular levels with an order of magnitude greater than previously thought. Heterogeneity in isogenic cell populations could arise from the intrinsically stochastic processes of the expression of individual genes and such stochasticity, once amplified to certain amplitude threshold, could result in heterogeneity at the cellular level, and eventually lead to different fates for a microbial population. It is increasingly recognized that by using averaged molecular or phenotypic measurements of a whole population to describe cell behaviors, conclusions could be biased by the expression profiles of outliers; meanwhile, these unique patterns could be distinctive functional behaviors at a given place and time and therefore important to analyze and describe. Here we report the first bacterial single-cell RNA-seq method for RNA isolation, cDNA synthesis and amplification, and next-generation sequencing of the whole transcriptome of single cyanobacterium *Synechocystis* sp. PCC 6803 cells which contain approximately 5-7 femtogram total RNA per cell. Using the method, we determined the transcriptomic profiles of *Synechocystis* single cells under nitrogen-starvation stress. The results revealed significant gene-expression heterogeneity between single cells. The study provided a new single bacterial cell transcriptomic method that could be valuable in studying gene-expression heterogeneity in a genetically identical bacterial population under environmental stress conditions.

.

Microbiologists typically assume that microbial cells growing under the same conditions are uniform and thus can be described by average values of their physiologic, phenotypic, genotypic or other measurements [1]. While this approach has been applied to generate current knowledge of microbial biology successfully, it was recently found that even isogenic populations of exponentially growing microorganisms showed substantial cell-to-cell heterogeneities at both cellular and molecular levels with an order of magnitude greater than previously thought [2]. Heterogeneity in isogenic cell populations could arise from the intrinsically stochastic processes of the expression of individual genes [3, 4], and such stochasticity, once amplified to certain amplitude threshold, will result in heterogeneity at the cellular level, and eventually lead to different fates for a microbial population [5, 6]. Such cell-cell heterogeneities could be further magnified in natural environments when complicated chemical, physical and biological factors are involved [7]. It is now increasingly recognized that by using conventional averaged molecular or phenotypic measurements of a whole population for describing cell behaviors, conclusions could be biased as the unique patterns related to distinct functional sub-populations cannot be revealed.

Another major drive to pursue single-cell analysis stems from the fact that more than 99% of microbes from environmental samples are “unculturable” in laboratory and thus cannot be studied using the conventional microbiological methods [8]. Meanwhile, many of these 'unculturable' microbes represent new phylotypes, families and divisions in domain bacteria and archaea, and could be of direct relevance to various valuable biological processes involved in bioremediation, global warming, alternative energy, and the basic sciences of global cycling of carbon, nitrogen and metals [9]. Although several

"culture-independent" single-cell genomics technologies have been established and successfully applied in microbial ecology studies [10-13], they only provide information related to genetic architecture and metabolic potential of cells, and do not reveal metabolic functionality and gene expression dynamics related to various environmental parameters.

Due to the challenging facts that most bacterial cells are small and difficult to lyse efficiently, and contain extremely low amount and short half lived mRNA [14], the first transcriptomic analysis of single bacterial cells was reported only recently [15]. In the study, the whole transcriptome of single *Burkholderia thailandensis* cells was amplified using the φ29 polymerase multiple displacement amplification (MDA) of circularized cDNA, and then analyzed by a DNA microarray, the results showed lower fold-change bias (less than two-fold difference and *Pearson* correlation coefficient $R \sim 0.87–0.89$) and drop-outs (4%–6% of 2842 detectable genes) as compared with the data obtained from non-amplified RNA samples. However, the estimated total RNA in a single *B. thailandensis* cell is about 2 picograms (*p*g), which is several orders of magnitude higher than the estimated total RNA amount of a typical bacterial cell, such as 3~9 femtograms (*f*g) in *Escherichia coli* and marine planktons [16-18]. In this study, we reported a new bacterial single-cell RNA-sequencing method (known as BaSiC RNA-seq) to amplify whole transcriptome of single cyanobacterial cells with total RNA of ~5 *f*g, and to analyze the single-cell transcriptome through a next-generation RNA-seq sequencing technology. We applied the BaSiC RNA-seq protocol to obtain high-coverage transcriptomic profiles of single cells of *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) grown under a common environmental stress, nitrogen-starvation condition [19], as well as the bulk-cell

controls for the same treatments (**Fig. 1**). The results showed significant gene-expression heterogeneity among single cells, especially for cells after long-term stress treatments. In addition, the results revealed varying heterogeneity among transcripts of different functional categories upon stress.

Results

Quantitation of total RNA in single *Synechocystis* cells

It was estimated that a typical single bacterial cell contains ca 1.5~100 *f*g of total RNA, comparing with 10~30 *p*g total RNA per eukaryotic cell [20]. We first performed a bulk cells-based analysis to estimate total RNA in single *Synechocystis* cells. Total RNA was isolated from ~10^7 *Synechocystis* cells. The cell number was determined by directly cell counting under microscope. The analysis showed that the total RNA content was 7.5 ± 0.8, 7.1 ± 1.2 and 5 ± 0.6 *f*g per single *Synechocystis* cell for the control cells, the cells treated with nitrogen starvation for 24 and 72 h, respectively (**Suppl. Fig. S1**). This estimation was in the similar range as the previous estimation for a single *E. coli* cell (3.4 *f*g) [14, 16] and marine bacterial plankton cells (1.5~9.0 *f*g) [17, 18].

Whole-transcriptome amplification and quality control

Synechocystis single cells were isolated using an Olympus single cell manipulator. Single live bacterial cells were picked randomly from the samples. We immediately isolated the total RNA from the single *Synechocystis* cells using a Zymo RNA Isolation Kit that has been verified to yield high-quality total RNA without DNA contamination from both single eukaryotic and bacterial cells in our previous study [14]. The lysis and RNA isolation

procedures have been successfully conducted in *Desulfovibrio vulgaris* with *Methanosarcina barkeri*[21], *E. coli* [14], diatom[22] and human single cells[23] [24]. Then the purified total RNA was used as template for amplification using a modified One-Direct RNA Amplification System to generate high quality cDNAs. Concentration and purity of the amplified cDNAs from both bulk- and single-cells were firstly evaluated with Nanodrop as the ratio of absorbance readings at 260 and 280 nm, and then on standard 1.5% agarose gel electrophoresis. The ratios of A_{260}:A_{280} of amplified products obtained were around 1.8-1.9. Since the amplified cDNA product from One-Direct RNA Amplification System contained both sense and antisense cDNA strands, it was recommended by the manufacturer to use the single stranded cDNA conversion factors (1 A_{260} unit = 33 ng/μL of cDNA) for quantitation. Based on this calculation, we were able to obtain 6.85~16.78 μg cDNAs after amplification form single-cell RNA samples (**Table 1**). Although no-template control (NTC) also generated around 2.8 μg cDNA, gel analysis showed that they were at significantly lower DNA sizes (*i.e.*, 150-1,000 bp in NTC *vs.* 200-5,000 bp in single-cell samples), suggesting they were probably random non-specific cDNA products. The median length of the amplified cDNA varied from 200 to 295 bp, with peak length around 150-180 bp measured by Agilent Bioanalyzer using the DNA 100 size markers (**Suppl. Fig. S2)**. A blunt-end clone library of the amplified cDNAs from a single *Synechocystis* cell was constructed, and a total of 30 clones were randomly selected for DNA sequencing. The BlastN results showed that all inserts were matched to the *Synechocystis* genome, indicating the successful amplification of the target transcripts, and no RNA contamination from other organisms. Meanwhile,

sequences obtained from NTC found no homolog in GenBank, indicating a random non-specific amplification during the RNA amplification in NTC.

RNA-seq analysis of single-cell transcriptome

We used the Encore® Rapid Library Systems to construct libraries for next-generation sequencing on the Illumina NGS platforms. Libraries for 9 cDNA samples, including 3 bulk-cells samples (one each from 0, 24 h and 72 h of nitrogen-starved cultures, respectively) and 6 single cells (biological triplicate cell-1, 2 and 3 at 24 h, and biological triplicate cell-4, 5 and 6 at 72 h after nitrogen-starvation treatment, respectively) were prepared and sequenced on Illumina HiSeq with 1 × 50 bp configuration. The analysis produced 332~938 Mbase yields and 6.5~18.4 M reads for bulk-cells samples, 344~3,283 Mbases and 6.7~64.3 M reads for single cells. Analysis of the sequence reads showed that the mean Q scores of reads were 38.66~39.15, with 98.08~99.22% of reads having Q scores ≥ 30, and 90.1% of reads within 38~40 Q scores, suggesting overall good quality of sequencing (**Suppl. Table S1**). The expression level of the genes was normalized using a standard RPKM (Reads Per Kilobase of Gene Per Million Mapped Reads) method before they were subjected to further analysis. The RPKM values of all transcripts detected in the samples were provided in **Suppl. Table S3**.

The single cell transcriptomic mapping was then conducted following standard protocols for RNA-seq data. The sequence reads were pre-processed using the FASTX toolkit to remove low-quality bases and reads shorter than 20 bp, and then mapped to the *Synechocystis* genome using Bowtie (version 2.0.0) with default parameters. Analysis of cDNAs detected in single *Synechocystis* cells showed that 82.9, 98.6 and 98.6% of the

putative transcripts in the *Synechocystis* genome (3,620 genes) were detected in single cell-1, 2 and 3 from the 24 h nitrogen-starved cultures, respectively. Meanwhile, 33.5, 38.4 and 50.6% of the putative transcripts were detected in single cell-4, 5 and 6 from the 72 h nitrogen-starved culture, respectively (**Table 2**). The low transcript coverage at 72 h was probably due to the increasing dormant status caused by the extended nitrogen starvation stress, as similar trends were also observed in the bulk-cell samples, 99.4% putative transcripts detected in 0 h sample, while only 92.8% and 63.2% in bulk-cell samples of 24 h and 72 h after nitrogen starvation treatment, respectively (**Table 2**).

Hierarchical clustering of both single-cell and bulk-cell transcriptomic data was conducted on basis of the *Z*-scores of the log2 signals of the robust multiarray analysis (RMA) using the 3,117 genes with RPKM > 0 in at least one single *Synechocystis* cell (**Fig. 2A**), and 424 genes detected in all 6 single-cells and 3 bulk-cell transcriptomic profiles (**Fig. 2C**). The clustering results showed the similar trends that profiles of single cells at 24 h or 72 h were grouped together with those of bulk-cells of the same time points. Further principal component analysis (PCA) of the same data sets showed that: *i*) the changes along the nitrogen-starvation time were clearly observed for the transcriptomic profiles of bulk-cells at 0, 24 and 72 h, suggesting treatment time is a important factor of separating the transcriptomic profiles; *ii*) although none of the single-cell transcriptomic profiles were identical with the corresponding bulk-cells profile, transcriptomic profiles of single cells of 24 and 72 h were clustered together correspondingly with profiles of their bulk-cells, while the group of three single cells at 24 h were well separated from the group of 72 h; *iii*) compared with the single-cell profiles of 24 h, higher degree of gene-expression heterogeneity was observed among

three single-cell transcriptomic profile of 72 h (**Fig. 2B and 2D**). To further confirm the trends, the correlation between transcriptomic profiles of single and bulk cells of the same time points was plotted (**Fig. 3**). The results showed that the transcriptomic profiles of three single cells shared higher correlation with the bulk-cell transcriptomic profile at 24 h (R^2 = 0.77, 0.85 and 0.85 for single cell-1, 2 and 3, respectively), compared with very low correlation of the transcriptomic profiles of three single cells of 72 h with their corresponding bulk-cell transcriptomic profile (R^2 = 0.014, 0.028 and 0.088 for single cell-4, 5 and 6, respectively). While the high correlation of three single-cell transcriptomic profiles with the bulk-cell profile at 24 h is indicative of good reliability of the BaSiC method for single-cell bacterial analysis, the low correlation of three single-cell transcriptomic profiles with the bulk-cell profile at 72 h suggested that a increased gene-expression heterogeneity among single cells may have occurred for the *Synechocystis* population under the nitrogen-starvation condition for an extended period of time. The same results were also obtained when we plotted the gene-expression variation among single cells (*i.e.*, standard deviation divided by the median measurement of three single cells) with the median gene-expression measurement of all genes or genes with expression level greater than RPKM values of 500 (**Suppl. Fig. S3**).

To further decipher the gene-expression heterogeneity, the overrepresentation of gene functional categories was analyzed for the 424 transcripts detected in all 6 single-cell and 3 bulk-cell profiles. Based on *Spearman*'s correlation analysis, the functional category "Mobile elements" was detected as significantly correlated ($p < 0.001$) among all 3 single cells under both 24 and 72 h nitrogen starvation (**Suppl. Table S2**). Anova *F*-test was performed between single cells against each functional category to determine

whether a given functional category is more different from the others at each of the time points (*i.e.*, 24 or 72 h). The functional category of "Mobile elements" showed significant difference from other functional categories among 3 single cells at 24 h using BaSiC RNA-seq data, and the coefficient of variation (CV) values at both 24 and 72 h time points (standard deviation/average from single cell-1, 2, 3 and single cell-4, 5, 6 for 24 h and 72 h respectively, either with 424 genes detected in all single cells or all genes detected in each single cells), while "Protein synthesis" functional category was significantly distinguished from others only at 24 h (**Table 3**). The results suggested that transcripts from different functional categories could contribute differently to the overall gene-expression heterogeneity, among which "Mobile elements" category contributed the most for the increased heterogeneity caused by the nitrogen-starvation stress.

RT-qPCR verification of selected genes

Since the statistic analysis based on only limited single-cell RNA-seq transcriptomic data (in this case, 3 cells for each condition) may be not enough to calibrate the possible bias during single call analysis. We further applied a single-cell RT-qPCR analysis to 24 single *Synechocystis* cells randomly selected from each cultivation condition. To verify the heterogeneity revealed by the BaSiC RNA-seq transcriptomic analysis of single *Synechocystis* cells, one transposase encoded gene (*slr1684*) belonged to "Mobile elements" functional category and one asparaginyl-tRNA synthetase related gene (*sll0495*) belonged to "Protein synthesis" functional category were selected for single-cell gene expression measurements using RT-qPCR method established previously [14]. In the single-cell transcriptomic data, these two genes were found with increasing heterogeneity

after nitrogen-starvation treatment (**Suppl. Fig. S4**). As the RT-qPCR involves no cDNA pre-amplification, we expected it provided more reliable measurements to validate results from the single-cell transcriptomics. Since there is no suitable internal reference gene available for single cell RT-qPCR [21], the raw Ct values of all single cells were presented in this study. The RT-qPCR results confirmed that the gene expression levels of 16S rRNA, *slr1684* and *sll0495* were slightly down-regulated at 72 h when compared with 24 h after nitrogen-starvation stress, consistent with the increased degradation of cyanobacteria total RNA after longer nitrogen starvation observed in this study (**Suppl. Fig. S1**). More importantly, remarkable differences in terms of single cell gene-expression distribution patterns of these selected genes between different time-points after a nitrogen-starvation treatment were also observed: bigger *sigma* coefficient factors (σ) of curve fitting were detected for both *slr1684* and *sll0495* at 72 h than those at 24 h, 1.46 *vs.* 0.76 and 1.48 *vs.* 1.15 respectively with $R^2 > 0.92$. In general, smaller *sigma* values indicate a narrower distribution. Therefore, the population at 72 h displayed considerable heterogeneity in transcript expression levels compared to that of 24 h for *slr1684* and *sll0495* genes, while no such heterogeneity increase was observed for 16s rRNA gene (**Fig. 4**). The gene-expression heterogeneity of selected genes was increased at 72 h after nitrogen-starvation treatment of a longer time, suggesting that increased gene-expression heterogeneity may be one of the responses of a genetically identical bacterial population towards challenging environmental conditions.

Discussion

Several approaches have been developed in the past to define gene-expression heterogeneity at the level of single bacterial cells, such as reporter gene, fluorescent *in situ* hybridization (FISH), *in situ* PCR combined with *in situ* reverse transcription (*in situ* RT-PCR), RT-qPCR, and various imaging-based gene expression measurements [7], leading to an increasingly recognized view that gene-expression heterogeneity was common phenomena even in isogenic bacterial population. However, most of these previous methods can measure only a small number of genes in each single cell. Transcriptomic analysis of single bacterial cells has obvious advantages in deciphering heterogeneity at a whole-genome level and its underlying biological significance. In recent years, methods to perform transcriptomic profiling of single eukaryotic cells with average ~10 *p*g total RNA per cell have been established and applied to the developing pancreas, retina and olfactory systems, stem cells, cancer cells and fungi [25, 26]. However, since all these methods for eukaryotic single cell transcriptomics employed poly(T) as the primers for cDNA synthesis, none could be adapted directly for transcriptomic analysis of single bacterial cells whose messager RNA have no unique ploy(A) structures. In the NuGen RNA amplification systems we utilized in this study, the usage of both poly(T) and hexamer primers provides the potential of prokaryotic mRNA amplification [27, 28]. The first bacteria single-cell RNA amplification protocol was also recently established with estimated starting RNA of 2 *p*g using random DNA hexamers as primers, and assessed by a DNA microarray-based analysis [15]. In this study, using NuGen RNA amplification system as the core, we developed the first RNA-seq based whole transcriptomics method for single bacterial cells, BaSiC RNA-seq, and validated it on *Synechocystis* cells

containing only ~5 *f*g total RNA per cell. The single *Synechocystis* cells under nitrogen-starvation condition for 24 and 72 h were randomly isolated and subjected to global transcriptomic profiling using this method. The results demonstrated that the BaSiC RNA-seq method has achieved up to 82~98% coverage of the *Synechocystis* genome which is at the same or higher level of coverage as several other single-cell transcriptomic methods developed for animal cells (61.4-96%) [26], fungi (4-7%) [29] and bacteria (94-96%) [15].

As the first RNA-seq transcriptomic method for bacterial single cells, the BaSiC RNA-seq method includes several other advantages: *i*) high accessibility: it utilizes a combination of commercially available kits with minor modifications for all procedures from RNA isolation, amplification to sequencing library preparation, no special reagents or instruments was needed so that the method can be accessed by most of the molecular biology laboratories. In addition, its utilization of commercial kits also provided good extra quality controls across the laboratories so that the transcriptomic data can be more comparable; *ii*) user friendly and less time consuming: core parts of the method, RNA isolation and amplification processing took only 10 min and ~5.5 h, respectively, which allows preparation of several dozens of single-cell samples for sequencing within 1~2 days; *iii*) sensitivity and reliability: it can be used to sequence transcriptome from a total RNA less than a few *f*g, which makes it suitable for most of prokaryotic cells. For the future application of environmental samples, we expect single-cell isolation using either micromanipulation or other higher throughput technologies will allow capture of clear single cells out of environmental samples. In addition, necessary modification and improvement may still needed if we want to extend the application to environmental

samples; *iv*) generality: the key component used for RNA amplification in the BaSiC RNA-seq method, the One-Direct RNA Amplification System, was also applied in single-cell transcriptome of fungi [29] and prostate cancer cells [30] recently, since both poly(T) and random primers are used for the first stranded cDNA synthesis during RNA amplification, it is thus expected that with little modification, the BaSiC RNA-seq protocol can also be applied to transcriptomic profiling of eukaryotic single cells.

Considering the greater challenges related to single bacterial cells (*i.e.*, low abundance of RNA and difficult lysis), the results derived from the single-cell transcriptomics needed to be verified by other approaches. To this end, some methodologies have been recently introduced, such as pre-addition of unique molecular identifiers (UMIs) [31-34] and spike-in control RNA [35] during RNA amplification or single cell FISH,[31] and single cell RT-qPCR [22, 23]. Like most of other single-cell transcriptomics protocols established previously [25, 26], BaSiC RNA-seq produces ds-cDNA, therefore cannot retain strand specificity for noncoding RNAs such as antisense RNA detection. However, the issue can be addressed by coupling the BaSiC RNA-seq analysis with single-cell RT-qPCR analysis [14], which can be employed to validate results from single-cell bacterial transcriptomic analysis. It is difficult to split the RNA from a single bacterial cell for both RT-qPCR and amplification, which makes it impossible to validate the linear amplification with RT-qPCR at both pre- and post-amplification of a few selected genes. In our study, to exclude effects of possible amplification bias, we validated the expression patterns of the genes belonging to "Mobile elements" and "Protein synthesis" functional categories using an amplification-free single-cell RT-qPCR method, the preliminary results showed that the gene-expression levels and the

distribution patterns of the selected mobile elements and protein synthesis related genes were highly correlated between the single-cell transcriptomic data and single-cell RT-qPCR data. Together with various noise validation models such as spike-in control RNAs, various chip-level RT-qPCR instruments become more available in the future, the single-cell bacterial transcriptomics method described here can be further validated.

Bacterial cells have developed significant abilities to survive and propagate in highly diverse and changing environments by evolving phenotypic heterogeneity [36]. Single-cell analysis of gene-expression patterns during nutrient starvation in *Bacillus subtilis* reveals large phenotypic variation [37], which is considered as a risk spreading strategy and can ensure survival of at least a subpopulation of cells in unpredictably fluctuating environments. In the study, preliminary single-cell transcriptomics based on 6 *Synechocystis* single cells showed a possibility of increased gene–expression heterogeneity after long-term nitrogen-starvation stress. As phenotypic heterogeneity could be resulted from the gene-expression heterogeneity, the increased gene-expression heterogeneity could be a benefit to the survival of bacterial population. In addition, the single-cell transcriptomics and single-cell RT-qPCR analyses showed that not all functional categories contributed equally to the overall increase of gene-expression heterogeneity, genes belonged to the functional categories such as “Mobile elements” seemed to play more roles. Mobile elements are DNA sequences such as plasmids, transposons, insertion sequences, and miniature inverted repeat transposable elements [38], among which transposons have been suggested as an important mean of adaptive variation and evolution at the genomic level in many microbes under stress conditions [39-41]. Comparative genomics analysis showed an extremely high and hierarchical diversity

of mobile elements present in cyanobacteria [42, 43], and transposase genes are involved in genomic rearrangements associated with differentiation process, formation of filaments and the regulation of cell-cell interactions in the colonial way of life in cyanobacteria [43]. Although more proof is still needed, the preliminary results based on a limited number of single cells implied a possibility that the increased heterogeneity of gene expression of selected functional categories such as "Mobile elements" could be one of the important means to increase phenotypic variations, which will eventually enhance the survival possibilities of a bacterial population under nutrient starvation and other adverse stress conditions.

As a final remark, due to significant amplification and possible bias involved in this single-cell transcriptomic method, for example, DNA amplification also observed in negative controls, caution should be observed when interpreting the single-cell transcriptomic data quantitatively before the method is further evaluated and improved. In this study we only studied RNA-seq in 3 cells per condition and perhaps the low sampling was not ideal in order to discriminate between a heterogeneous stochastic behavior and genuine different expression profiles of subset of cells within the population. This is also the major reason for which we did not come to statistic analyses for details about what genes are differentially expressed between 24 and 72 h treatments for a given pair of cells. What's more, spike-in RNA with known copy numbers should also be added at the single cell lysis step, to validate the amplification bias from cell to cell. However, given the experimental limitations, we showed that heterogeneity can be detected with confidence and so as the first RNA-seq based transcriptomic method for

single bacterial cells, the method may be valuable in many research areas, such as microbial ecology and environmental sciences.

Methods

Bacterial growth conditions

Synechocystis sp. PCC 6803 was grown in BG11 medium (pH 7.5) under a light intensity of approximately 50 μmol photons m^{-2} s^{-1} in a shaking incubator with 130 rpm speed at 30°C (HNY-211B Illuminating Shaker, Honour, China). Cell density was measured on a UV-1750 spectrophotometer (Shimadzu, Japan). For nitrogen-starvation treatment, 10 mL fresh cultures grown in BG11 medium at exponential phase (OD_{730} = 0.8) were collected by centrifugation (8000 × *g*), washed twice by ddH_2O, and then were inoculated into 25 mL fresh BG11 liquid medium without nitrogen source in a 250-mL flask cultivated in the same condition described above. Growth experiments were repeated at least three times to confirm the growth patterns. Cell samples were collected at 0, 24 and 72 h.

Single cell harvesting and RNA isolation

To eliminate the potential exogenous DNA or bacterial contamination during the BaSiC RNA-seq, multiple super-clean incubators with ultraviolet lights were employed, especially individual incubator was used for each step of RNA amplification. All the related kits and reagents were put in the incubators in advance and treated with ultraviolet light for at least 30 min before manipulation.

Single cell harvesting was conducted using an Olympus single cell manipulator (Olympus Inc, Japan) with the Inverted microscope IX71 and a micromanipulator for ICSI/Injection (Narishige Group, Tokyo Japan) with ICSI 10 micrometer diameter micropipette (Origio, Charlottesville, VA). Briefly, cells were centrifuged at 12,000 × *g* for 2 min at room temperature. After removing the supernatant, cells were immediately resuspended with 5 mL 1× phosphate buffer (PBS) (pH 7.0) containing 20% RNALaterTM (*v*/*v*) (Ambion, Austin, TX). The cell suspension was further diluted with PBS-RNALaterTM solution if needed, and then pipetted onto a microscope slide. Under light microscope (40X objective), cells were located and aspirated into a micropipette, visually confirmed and then loaded into a 0.2 ml PCR tube preloaded with 40 μL 1 × PBS containing 20% RNALater (*v*/*v*). Isolation of total RNA from the harvested single cells was performed immediately using ZYMO RNA MicroPrep Isolation Kit (ZYMO, Irvine, CA), following the manufacturer's instruction with a minor modification to omit the step of IIIC column due to limited amount of lyzed cell residue derived from a single cell [14].

Estimation of total RNA content in single cell

We used diluted bulk cells to estimate the total RNA content of a single *Synechocystis* cell under different conditions, *i.e.*, normal growth after 0, 24 and 72 h after nitrogen starvation treatments. After cultivated for designated time, varying number of cells were collected by centrifugation at 12,000 × *g* for 2 min. After removing the supernatant, cells were resuspended with 5 mL RNALaterTM (Ambion, Austin, TX). RNA isolation was performed using ZR RNA MicroPrep Kit (ZYMO, Irvine, CA). RNA concentrations in each sample was determined immediately using Nanodrop 2000 (Thermal Scientific,

Waltham, MA), with at least 3 analytical replicates included. The linear regression of the total RNA content to the cell number was conducted and the slope was designed as the total RNA content in a single *Synechocystis* cell.

Single-cell RNA linear RNA amplification

Single-cell RNA amplification was conducted with NuGen WT-Ovation One-Direct RNA Amplification System (NuGen, San Carlos, CA) following the manufacturer's instruction, with the following modifications: *i*) instead of using 2 μL lysis buffer provided in the kit to lyze the single bacterial cells, we started from the isolated single-cell total RNA using the ZYMO RNA isolation kit; *ii*) the input volumes of primers for the first- and second-strand cDNA synthesis were reduced due to the low amount of started RNA from single cells; and *iii*) for each step, cDNA synthesis, Single Primer Isothermal Amplification (SPIA), and post-SPIA modification, was performed in a separated PCR workstation to avoid the possible contamination during the synthesis and amplification process. The no-template negative controls containing only RNase free water were used in parallel with total RNA samples isolated from single cells. Single cell harvesting, RNA isolation and RNA amplification were conducted immediately one after another to avoid possible RNA degradation and cross-contamination. The whole process took approximately 8 h in our laboratory. Amplified cDNA from single cells was purified using Qiagen MinElute Spin Column (Qiagen, Shanghai) following the manufacturer's instruction.

Quality control of single-cell transcriptome

Two approaches were used to determine the quality of the amplified single-cell transcriptome: *i*) cDNA product yield and purity: as recommended by the manufacturer, a standard NonoDrop measurement was used to quantify the amplified cDNA products using the single stranded cDNA conversion factor (1 A_{260} unit = 33 ng/μL of cDNA) suggested by NuGen One-Direct Kit; *ii*) size distribution of cDNA: cDNA generated by the WT-Ovation One-Direct System was measured by Agilent Bioanalyzer 2000 (Agilent, Santa Clara, CA). To verify the amplified cDNA product were derived from the target bacterial cells instead of contaminated sources, blunt-end clonal library (End-It DNA end repair Kit, Epicentre Biotechnologies, Madison, WI) and SMART Blunt Cloning Kit (Lucigen Corporation, Middleton, WI) were used to construct a clone library in Escherichia coli for the amplified cDNA products. Thirty randomly selected *E. coli* clones were subjected to sequencing analysis by standard ABI3700 (Life Science Inc, Carlsbad, CA).

RNA-seq analysis of single-cell transcriptome

1 μg purified double-stranded cDNA was subjected to library preparation using the NuGen Encore Rapid DR Multiplex System 1-8 (NuGen, San Carlos, CA), through a four-step protocol including end repairing, adaptor ligation, final repair and library purification, following the manufacturer's instruction. To determine the quality of the libraries, a KAPPA Library Quantification Kit (KAPPA Biosystems, Woburn, MA) was first used to determine the DNA concentration of the libraries, and then Agilent Technologies 2000 Bioanalyzer (Agilent, Santa Clara, CA) was used to determine the product size of the libraries, with good libraries typically around 300 bp. The products

were then used directly for cluster generation using Illumina's Solexa Sequencer (Illumina, San Diego, CA) according to the manufacturer's instructions. For next-generation sequencing, RNA 1 × 50 bp paired-end sequencing was performed using Illumina's Solexa Genome Analyzer II using the standard protocol. Multiple cDNA libraries from 4 single cell samples were loaded onto a single lane of an Illumina flow cell. The image deconvolution and calculation of quality value were performed using Goat module (Firecrest v.1.4.0 and Bustard v.1.4.0 programs) of Illumina pipeline v.1.4. Sequenced reads were generated by base calling using the standard Illumina sequencing pipeline. For transcriptomic data analysis, reads were pre-processed using FASTX Toolkit (Version: 0.0.13) to remove low-quality bases and reads shorter than 20 bp. Genome sequences (including *nc*RNA sequences) and annotation information of *Synechocystis* were downloaded from NCBI and the Comprehensive Microbial Resource (CMR) of TIGR (http://www.tigr.org/CMR) (Downloaded on August 22, 2013). After successful RNA amplification and labeled with normal RNA-seq preparation kit, the single cell transcriptomic mapping was conducted following standard protocol for RNA-seq data. The sequence reads were pre-processed using the FASTX toolkit to remove low-quality bases and reads shorter than 20 bp, and then mapped to the *Synechocystis* genome using Bowtie (version 2.0.0) with default parameters. For gene expression determination, we performed a standard calculation of Reads Per Kilobase of Gene Per Million Mapped Reads (RPKM). Principal component analysis (PCA) and hierarchical clustering of single-cell gene expression profiles were performed on the *Z*-scores (equal to value-average/standard deviation) derived from the log2 RPKM data, using SIMCA-P

11.5 software and *R* software, respectively. For the overrepresentation of various function categories, ANOVA analysis using SAS software package was conducted.

Gene expression analysis by single bacterial cell real-time RT-qPCR

To verify the gene-expression heterogeneity obtained from single-cell transcriptomic analysis, three selected genes from "Mobile elements" and "Protein synthesis" functional categories and internal control 16S rRNA gene were analyzed for their expression level by single-cell RT-qPCR. The optimized primer sets [14] and their corresponding gene targets were as follows: *slr1684* (forward primer, 5'-GTAAGGTCATGGATAGAATATGGATTTAAG-3'; reverse primer, 5'-TCCCACCATTTTTGAATTTGCT-3'; with an amplicon of 101 bp); *sll0495* (forward primer, 5'-AGGGATGGTTACGCACTAAACG-3'; reverse primer, 5'-ACTTCCATCCAGCACCACTTG-3'; with an amplicon of 101 bp); 16s rRNA (forward primer, 5'-TGTAGCGGTGAAATGCGTAG-3'; reverse primer, 5'-CCACGCCTAGTATCCATCGT-3'; with an amplicon of 152-bp). A total of 24 individual cells were collected from each time point for the analysis. Primers for single-cell RT-qPCR were designed using Primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/) and evaluated as described previously, briefly, 10 pairs of qPCR primers were designed and evaluated for the efficiency at single-cell level for each of the three genes [14,34]. Briefly, total RNA isolated from a single cell was reverse transcribed into cDNA using the VILO cDNA Synthesis Kit (Invitrogen, Carlsbad, CA). The resulted cDNA in a final volume of 10 μL was diluted 2.5 folds and used as template for SYBR green based RT-qPCR analysis on the StepOne thermal cycler (Life Technology, Carlsbad, CA). At

least three technical replicates were included for each qPCR analysis. Statistical analyses for single cell RT-qPCR data, such as distribution significance tests, were conducted using the OriginPro software package (v8, OriginLab, Northampton, MA).

Data access

The raw data reported in this study is available as supplementary tables online.

Acknowledgements

The research was supported by grants from the National Natural Science Foundation of China (project No. 31170043, No. 31270086 and No. 31370115), the National Basic Research Program of China ("973" program, project No. 2014CB745100, 2011CBA00803 and No. 2012CB721101), and the National High-tech R&D Program ("863" program, project No. 2012AA02A707). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would also like to thank Lianju Gao, Xingrui Li, Yaxing Wang and Guangsheng Pei in our laboratory for their assistance with sampling and statistic analyses.

Author contribution

JW and WZ designed the research and experiments, JW, ZC and LC performed the experiments and data analysis, JW and WZ wrote the main manuscript. All authors reviewed and approved the manuscript.

Competing financial interests

The author(s) declare no competing financial interests.

References

1. B. F. Brehm-Stecher and E. A. Johnson, *Microbiol Mol Biol Rev*, 2004, 68, 538-559.
2. M. E. Lidstrom and D. R. Meldrum, *Nat Rev Microbiol*, 2003, 1, 158-164.
3. A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce and R. Brent, *Nature*, 2005, 437, 699-706.
4. A. Becskei, B. B. Kaufmann and A. van Oudenaarden, *Nat Genet*, 2005, 37, 937-944.
5. J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi and J. S. Weissman, *Nature*, 2006, 441, 840-846.
6. T. J. Strovas, L. M. Sauter, X. Guo and M. E. Lidstrom, *J Bacteriol*, 2007, 189, 7127-7133.
7. P. S. Stewart and M. J. Franklin, *Nat Rev Microbiol*, 2008, 6, 199-210.
8. E. Cardenas and J. M. Tiedje, *Curr Opin Biotechnol*, 2008, 19, 544-549.
9. W. Fenical and P. R. Jensen, *Nat Chem Biol*, 2006, 2, 666-673.
10. K. Zhang, A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm and G. M. Church, *Nat Biotechnol*, 2006, 24, 680-686.
11. D. Wang and S. Bodovitz, *Trends Biotechnol*, 2010, 28, 281-290.
12. D. Parkhomchuk, V. Amstislavskiy, A. Soldatov and V. Ogryzko, *Proc Natl Acad Sci U S A*, 2009, 106, 20830-20835.
13. K. G. Lloyd, L. Schreiber, D. G. Petersen, K. U. Kjeldsen, M. A. Lever, A. D. Steen, R. Stepanauskas, M. Richter, S. Kleindienst, S. Lenk, A. Schramm and B. B. Jorgensen, *Nature*, 2013, 496, 215-228.
14. W. Gao, W. Zhang and D. R. Meldrum, *J Microbiol Methods*, 2011, 85, 221-227.
15. Y. Kang, M. H. Norris, J. Zarzycki-Siek, W. C. Nierman, S. P. Donachie and T. T. Hoang, *Genome Res*, 2011, 21, 925-935.
16. A. Schmid, H. Kortmann, P. S. Dittrich and L. M. Blank, *Curr Opin Biotechnol*, 2010, 21, 12-20.
17. S. Lee and P. F. Kemp, *Limnology and oceanography*, 1994, 39, 869-879.
18. S. Lee, C. Malone and P. F. Kemp, *Marine Ecology-Progress Series*, 1993, 101, 193-193.
19. C. Richaud, G. Zabulon, A. Joder and J. C. Thomas, *J Bacteriol*, 2001, 183, 2989-2994.
20. A. J. Westermann, S. A. Gorski and J. Vogel, *Nat Rev Microbiol*, 2012, 10, 618-630.
21. Z. Qi, G. Pei, L. Chen and W. Zhang, *Sci. Rep.*, 2014, 4.
22. X. Shi, W. Gao, S. H. Chao, W. Zhang and D. R. Meldrum, *Appl Environ Microbiol*, 2013, 79, 1850-1858.
23. J. Wang, X. Shi, R. H. Johnson, L. Kelbauskas, W. Zhang and D. R. Meldrum, *PLoS One*, 2013, 8, e75365.
24. J. Zeng, J. Wang, W. Gao, A. Mohammadreza, L. Kelbauskas, W. Zhang, R. H. Johnson and D. R. Meldrum, *Anal Bioanal Chem*, 2011, 401, 3-13.
25. J. J. Goetz and J. M. Trimarchi, *Nat Biotechnol*, 2012, 30, 763-765.

26. F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao and M. A. Surani, *Nat Methods*, 2009, 6, 377-382.
27. M. J. Filiatrault, *Curr Opin Microbiol*, 2011, 14, 579-586.
28. S. R. Head, H. K. Komori, G. T. Hart, J. Shimashita, L. Schaffer, D. R. Salomon and P. T. Ordoukhanian, *Biotechniques*, 2011, 50, 177-181.
29. C. de Bekker, O. Bruning, M. J. Jonker, T. M. Breit and H. A. Wosten, *Genome Biol*, 2011, 12, R71.
30. C. J. Welty, I. Coleman, R. Coleman, B. Lakely, J. Xia, S. Chen, R. Gulati, S. R. Larson, P. H. Lange, B. Montgomery, P. S. Nelson, R. L. Vessella and C. Morrissey, *BMC Mol Biol*, 2013, 14, 6.
31. A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park and A. Regev, *Nature*, 2013, 498, 236-240.
32. S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lonnerberg and S. Linnarsson, *Nat Methods*, 2014, 11, 163-166.
33. D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay and I. Amit, *Science*, 2014, 343, 776-779.
34. D. Grun, L. Kester and A. van Oudenaarden, *Nat Methods*, 2014, 11, 637-640.
35. S. Katayama, V. Tohonen, S. Linnarsson and J. Kere, *Bioinformatics*, 2013, 29, 2943-2945.
36. I. G. de Jong, J. W. Veening and O. P. Kuipers, *J Bacteriol*, 2010, 192, 2053-2067.
37. I. G. de Jong, J. W. Veening and O. P. Kuipers, *Environ Microbiol*, 2012, 14, 3110-3121.
38. H. J. Kazazian, *Science*, 2004, 303, 1626-1632.
39. Z. Zhang and M. J. Saier, *J Mol Microbiol Biotechnol*, 2011, 21, 59-70.
40. G. Feng, Y. E. Leem and H. L. Levin, *Nucleic Acids Res*, 2013, 41, 775-789.
41. F. Maumus, A. E. Allen, C. Mhiri, H. Hu, K. Jabbari, A. Vardi, M. A. Grandbastien and C. Bowler, *BMC Genomics*, 2009, 10, 624.
42. S. Lin, S. Haas, T. Zemojtel, P. Xiao, M. Vingron and R. Li, *Gene*, 2011, 473, 139-149.
43. L. E. Mikheeva, E. A. Karbysheva and S. V. Shestakov, *Russian Journal of Genetics: Applied Research*, 2013, 3, 91-101.

Table 1. Summary of RNA amplification from single bacterial cells.

	Concentration (µg/µL)	Total (µg)	A260/280
Single cell-1	0.36	9.00	1.82
Single cell-2	0.30	7.40	1.82
Single cell-3	0.57	14.18	1.84
Single cell-4	0.63	15.78	1.90
Single cell-5	0.67	16.78	1.88
Single cell-6	0.38	9.45	1.82
Bulk cells-0 h	0.34	8.60	1.81
Bulk cells-24 h	0.27	6.85	1.92
Bulk cells-72 h	0.28	6.90	1.84
NTC	0.11	2.80	1.81

Table 2: Summary of transcriptome identified.

	0 h	24 h				72 h				
	Bulk cells	Bulk cells	Single cell-1	Single cell-2	Single cell-3	Bulk cells	Single cell-4	Single cell-5	Single cell-6	Genes detected in all samples
Total genes identified	3117	2937	3117	3102	2615	1961	1132	1521	982	
Percentage of the genome (%)	98	92	98	98	82	62	36	48	31	
Amino acid metabolism	75	73	75	75	64	57	25	38	18	6
Biosynthesis of cofactors, prosthetic groups, and carriers	82	81	82	81	71	47	23	34	21	6
Cell envelope	69	69	69	69	63	46	31	36	23	12
Cellular process	91	84	91	90	79	64	28	48	31	10
Central intermediary metabolism	57	53	57	57	48	35	20	33	14	6
DNA metabolism	50	49	50	50	44	35	23	38	21	10
Energy metabolism	269	246	269	267	217	159	90	118	80	26
Fatty acid and phospholipid metabolism	35	33	35	35	32	26	15	18	8	1
Hypothetical protein	1275	1193	1275	1272	1027	753	437	582	387	180
Mobile and extrachromosomal element functions	81	70	81	77	65	62	54	51	42	28
Protein fate	62	61	62	62	57	46	29	30	21	12
Protein synthesis	98	84	98	96	67	43	20	34	16	5
Purines, pyrimidines, nucleosides, and nucleotides	42	42	42	41	38	26	13	24	15	6
Regulatory functions	140	135	140	140	126	102	62	82	55	21
Transcription	24	22	24	24	15	11	4	11	8	4
Transport and binding proteins	166	164	166	166	158	118	70	98	66	16
Unclassified	501	478	501	500	444	331	188	246	156	87

Table 3: Schematic representation of heterogeneity in functional categories "Mobile elements" and "Protein synthesis" as reveled by single cell transcriptomics (*Anova* analysis)

BaSiC RNA-seq	F-test Value	*P* value	Significance category**
24 h cell-1	2.41	0.0018	10*** (9, 18, 15, 7, 6)
24 h cell-2	2.8	0.0003	10 (9, 18, 15, 7, 6, 12, 14)
24 h cell-3	3.24	<0.0001	10 (17, 9, 18, 15, 3, 12, 11, 7, 14, 5, 6, 1)
72 h cell-4	0.71	0.7882	none
72 h cell-5	0.62	0.8641	none
72 h cell-6	0.9	0.575	none
24 h CV * [1]	3.57	<0.0001	10 (9, 15, 18, 17, 7, 12, 4, 6)
72 h CV * [1]	4.22	<0.0001	10 (3, 4, 5, 7, 9, 15, 17, 18)
24 h CV * [2]	6.49	<0.0001	10 (7, 9, 1, 4, 2, 5, 18, 14, 3, 8, 12, 15, 6, 17) 13 (7, 9, 18, 3, 12, 15, 6, 17)
72 h CV * [2]	4.81	<0.0001	10 (1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18)

* CV is equal to standard deviation/average from cell-1,2,3 and cell-4,5,6 for 24 h and 72 h, respectively, "1" indicated that we used the data of 424 transcripts detected in all single cells; "2" indicated that we used the data of 3117 transcripts detected in at least one single cell.

** The functional category shows significantly different from some other categories (in parentheses)

*** Function categories according to KEGG: 1-Amino acid biosynthesis; 2-Biosynthesis of cofactors, prosthetic groups, and carriers; 3-Cell envelope; 4-Cellular processes; 5-Central intermediary metabolism; 6-DNA metabolism; 7-Energy metabolism; 8-Fatty acid and phospholipid metabolism; 9-Hypothetical protein; 10-Mobile and extrachromosomal element functions; 11-No Data; 12-Protein fate; 13-Protein synthesis; 14-Purines, pyrimidines, nucleosides, and nucleotides; 15-Regulatory functions; 16-Transcription; 17-Transport and binding proteins; 18-Unclassified.

Figure Legends:

Fig. 1: Scheme of the bacterial single-cell transcriptomics procedures. Single bacterial cells were isolated and total RNA from each single cell was purified, then RNA amplification was applied to obtain microgram-level cDNAs from single cells. After quality controls (*i.e.*, no-template negative control, gel electrophoresis, cDNA quantification and size distribution analysis, clone library construction and sequencing), RNA-seq library derived from the amplified single-cell cDNA was subjected for RNA-seq analysis.

Fig. 2: Hierarchal clustering and PCA analysis of the single-cell transcriptomic data. Hierarchal clustering of transcriptomic profiles of 3 bulk cell samples (*i.e.*, 0, 24 and 72 h) and transcriptomic profiles of 6 single *Synechocystis* cells (*i.e.*, 3 from 24 h and 3 from 72 h) using datasets containing 3117 genes which were identified in at least one profile (***A***) using datasets containing 424 genes that were identified in all profiles (***C***). PCA analysis of 9 transcriptomic profiles using datasets containing 3117 genes which were identified in at least one profile (***B***) and using datasets containing 424 genes which were identified in all profiles (***D***).

Fig. 3: Comparison of single-cell and bulk-cell transcriptomic profiles. X-axis represents gene expression level from the bulk cells at 24 h (**A**, **B** and **C**), and at 72 h (**D**, **E** and **F**), while Y-axis represents expression level in individual single cell at 24 h (**A**, **B** and **C**) and at 72 h (**D**, **E** and **F**). For 24 h, 2990, 2985 and 2602 genes detected were used for plotting of single cell-1, cell-2 and cell-3, respectively. For 72 h, 441 transcripts

detected were used for plotting of all three cells (cell-4, cell-5 and cell-6). Correlation efficient was indicated inside the plots.

Fig. 4. RT-qPCR analysis of the selected genes. Population distribution plots (horizontal axis represents Ct value, vertical axis represents percentage of total cell population) and fitting curves reveal that long term nitrogen-starvation (72 h) display considerable heterogeneity in transcript expression levels in comparison with that of short term treatment (24 h). A-C represents 16s rRNA, *slr1684* (transposase), and *sll0495* (asparaginyl-tRNA synthetase), respectively.

Fig. 1

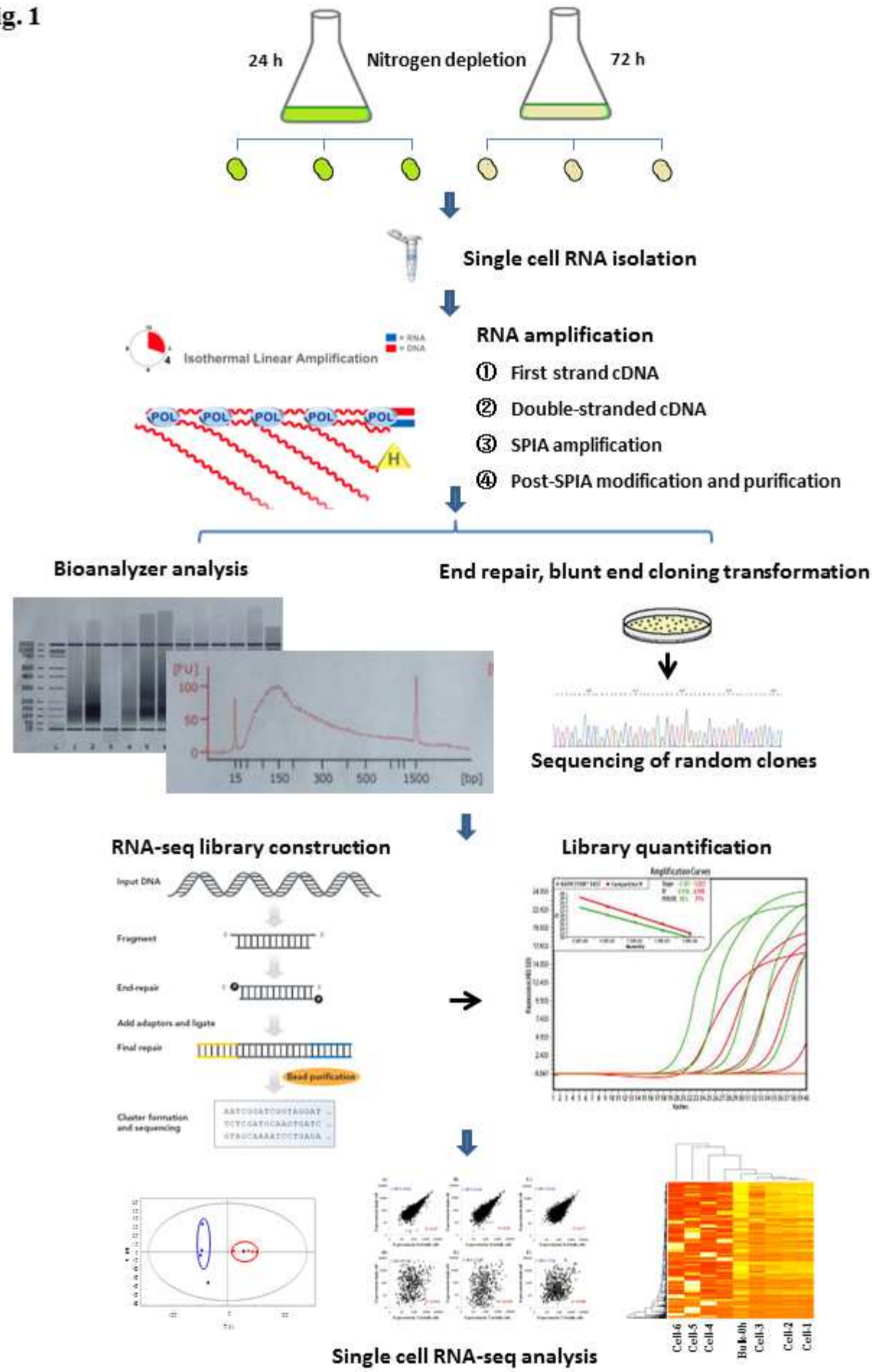


Fig. 2

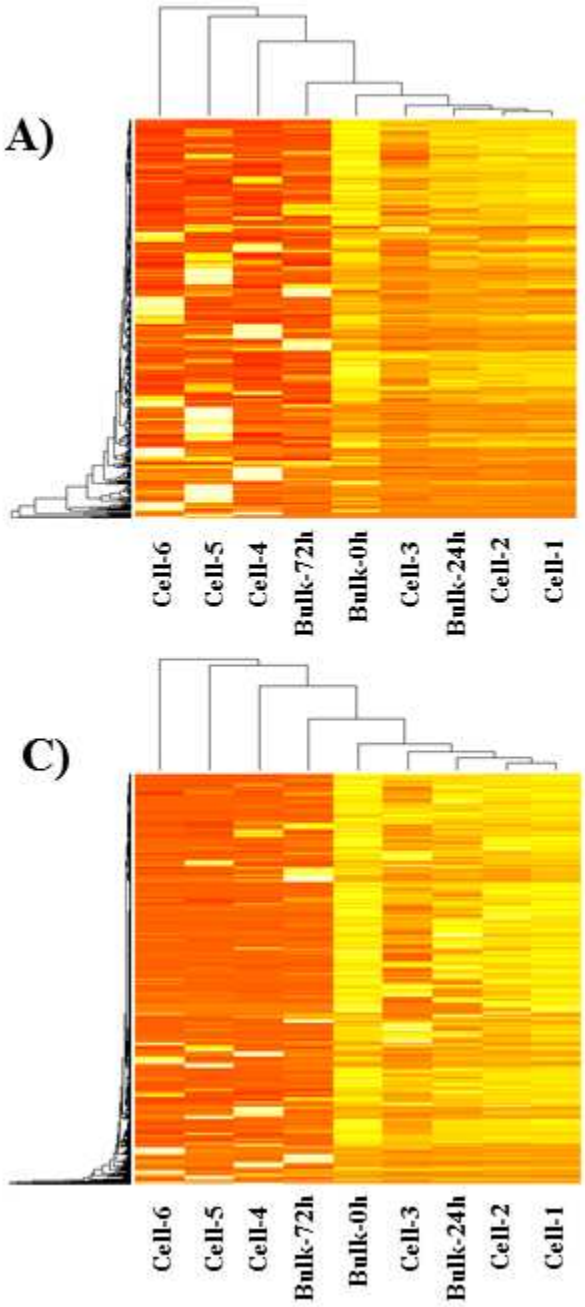


Fig. 3

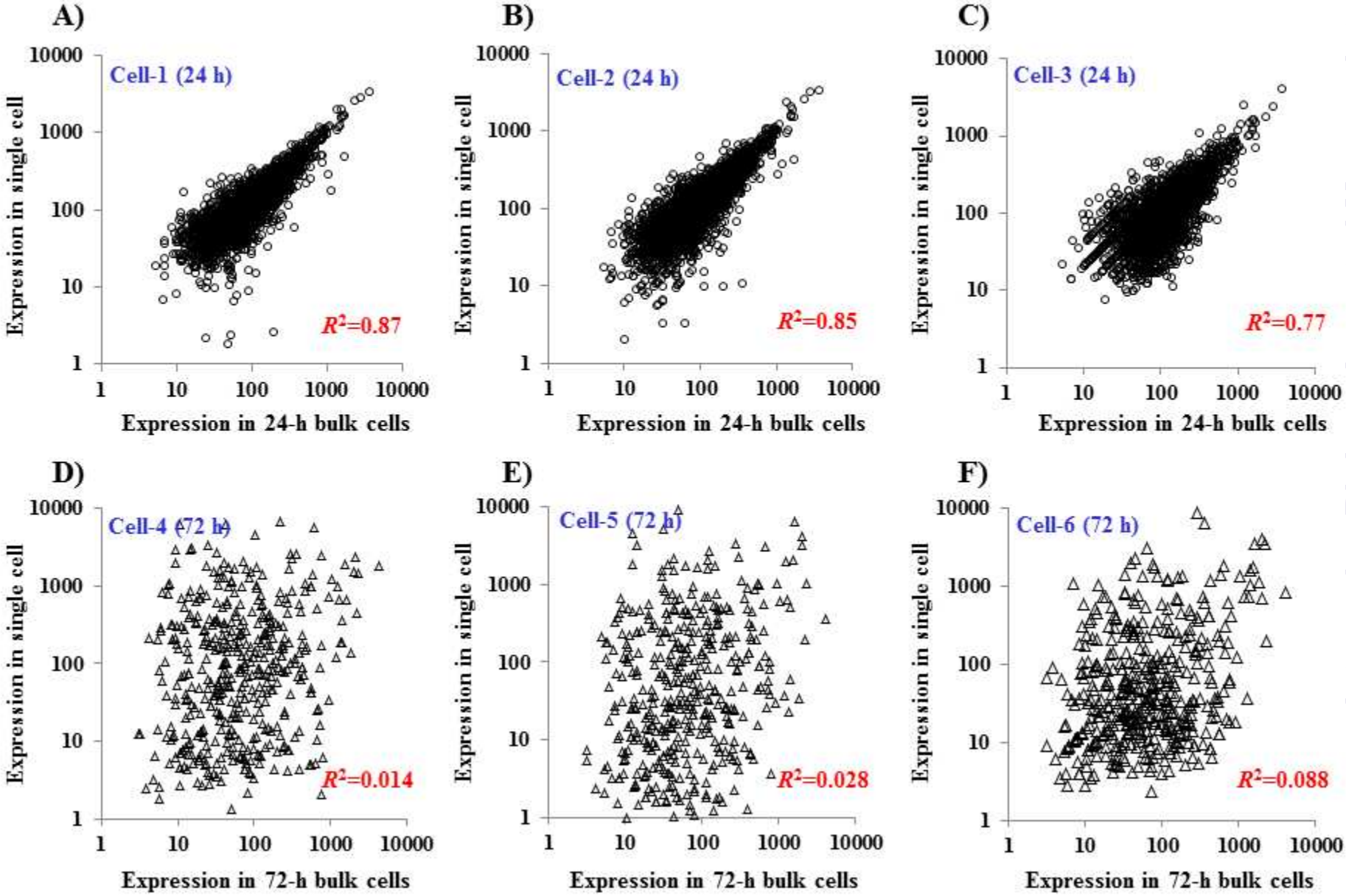
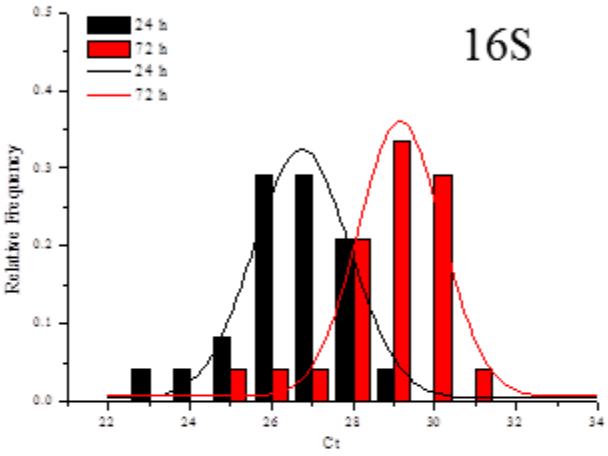
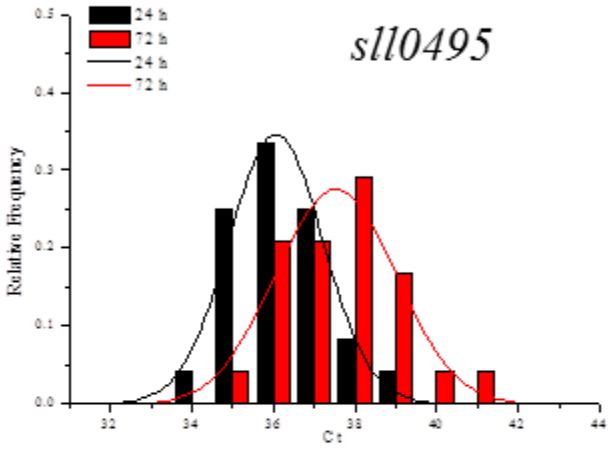


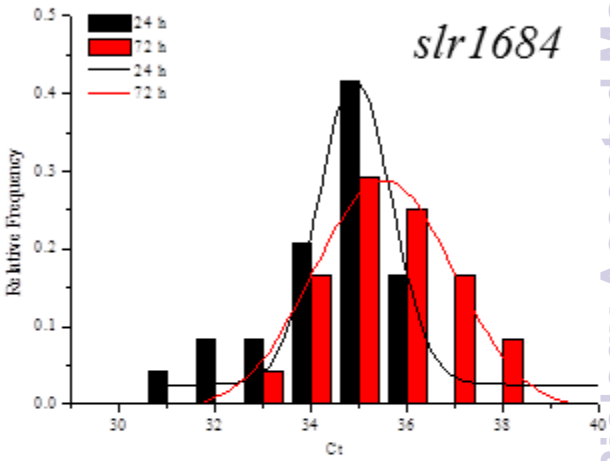
Fig. 4



	Adj. R-Square	sigma
24 h	0.95977	1.17723
72 h	0.95493	1.02529



	Adj. R-Square	sigma
24 h	0.98108	1.15355
72 h	0.92073	1.48065



	Adj. R-Square	sigma
24 h	0.92063	0.76276
72 h	0.97386	1.45617

A RNA-seq method for bacterial single-cell which not only shows reliable results but also reveals significant heterogeneity was firstly reported.