



RESEARCH

Open Access

# RNA-seq based identification and mutant validation of gene targets related to ethanol resistance in cyanobacterial *Synechocystis* sp. PCC 6803

Jiangxin Wang<sup>1,2</sup>, Lei Chen<sup>1,2</sup>, Siqiang Huang<sup>1,2</sup>, Jie Liu<sup>1,2</sup>, Xiaoyue Ren<sup>1,2</sup>, Xiaoxu Tian<sup>1,2</sup>, Jianjun Qiao<sup>1,2</sup> and Weiwen Zhang<sup>1,2\*</sup>

## Abstract

**Background:** Fermentation production of biofuel ethanol consumes agricultural crops, which will compete directly with the food supply. As an alternative, photosynthetic cyanobacteria have been proposed as microbial factories to produce ethanol directly from solar energy and CO<sub>2</sub>. However, the ethanol productivity from photoautotrophic cyanobacteria is still very low, mostly due to the low tolerance of cyanobacterial systems to ethanol stress.

**Results:** To build a foundation necessary to engineer robust ethanol-producing cyanobacterial hosts, in this study we applied a quantitative transcriptomics approach with a next-generation sequencing technology, combined with quantitative reverse-transcript PCR (RT-PCR) analysis, to reveal the global metabolic responses to ethanol in model cyanobacterial *Synechocystis* sp. PCC 6803. The results showed that ethanol exposure induced genes involved in common stress responses, transporting and cell envelope modification. In addition, the cells can also utilize enhanced polyhydroxyalkanoates (PHA) accumulation and glyoxalase detoxication pathway as means against ethanol stress. The up-regulation of photosynthesis by ethanol was also further confirmed at transcriptional level. Finally, we used gene knockout strains to validate the potential target genes related to ethanol tolerance.

**Conclusion:** RNA-Seq based global transcriptomic analysis provided a comprehensive view of cellular response to ethanol exposure. The analysis provided a list of gene targets for engineering ethanol tolerance in cyanobacterium *Synechocystis*.

**Keywords:** Ethanol, Tolerance, Transcriptomics, *Synechocystis*

## Background

Ethanol currently constitutes 99% of all biofuels in the United States. E-10 Unleaded, a blend of 10% ethanol and 90% ordinary gasoline, has been used in the U.S. for more than 25 years. Additionally, a blend of 85% ethanol and 15% ordinary gasoline (known as E-85) is rapidly growing in popularity [1]. The 3.4 billion gallons of ethanol blended into gasoline in 2004 amounted to about 2% of all gasoline sold by volume and 1.3% (2.5 x 10<sup>17</sup> J) of its energy content [1]. Greater quantities of ethanol are

expected to be used as a motor fuel in the future because of the federal policies, such as the “Twenty-in-Ten” program that proposes to cut gasoline consumption and greenhouse gas emissions from motor vehicles by 20 percent over the next 10 years. Large-scale ethanol production utilizes yeast or bacteria, such as *Saccharomyces cerevisiae* and *Zymomonas mobilis* to ferment sugar syrups [2]. The process has seen significant progress in recent years: inhibitor sensitivity, product tolerance, ethanol yield and specific ethanol productivity have been improved in modern industrial strains to the degree that up to 20% (v/v) of ethanol can be produced from starch-derived glucose [3]. However, since the large-scale ethanol fermentation consumes significant

\* Correspondence: wwzhang8@tju.edu.cn

<sup>1</sup>School of Chemical Engineering & Technology, Tianjin University, Tianjin 300072, People's Republic of China

<sup>2</sup>Key Laboratory of Systems Bioengineering, Ministry of Education, Tianjin 300072, People's Republic of China

amount of agricultural crops, which competes directly with the world food supply, and its increased production has been blamed for the food price increases in recent years.

Photosynthetic cyanobacteria have recently attracted significant attention as a 'microbial factory' to produce biofuels and fine chemicals due to their capability to utilize solar energy and CO<sub>2</sub> as sole energy and carbon sources, respectively [4]. By expressing a bacterial pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adh*) from the bacterium *Z. mobilis* in the cyanobacterium *Synechococcus* sp. PCC 7942, Deng and Coleman (1999) obtained a recombinant microorganism which can produce up to 230 mg/L ethanol directly from CO<sub>2</sub> within 4 weeks of growth [5]. More recently, a genome-scale metabolic network model of *Synechocystis* sp. PCC 6803 was used to improve cyanobacterial ethanol production up to 690 mg/L in a week [6]. Although still at very low productivity, these works clearly demonstrated that photoautotrophic cyanobacteria could potentially be engineered for a direct conversion of solar energy and CO<sub>2</sub> into biofuel products such as ethanol.

One of the key factors responsible for the low ethanol productivity is the low tolerance of photosynthetic systems to ethanol [7,8]. Ethanol can interfere with cell membrane's ability to act as a barrier, and interrupt key cellular processes such as protein biosynthesis, energy transduction and transport [8]. Although ethanol tolerance mechanism and application of ethanol-tolerant strains for enhanced production have been reported in native-producing yeasts and bacteria, current knowledge on ethanol tolerance in cyanobacteria is not enough to guide a rational engineering of more robust cyanobacterial hosts. To address this issue, we previously applied a quantitative iTRAQ LC-MS/MS proteomics approach to determine the responses of model cyanobacterial *Synechocystis* sp. PCC 6803 to ethanol [9]. The analysis showed that the *Synechocystis* cells employed a combination of induced common stress response, modifications of cell membrane and envelope, and induction of multiple transporters and cell mobility-related proteins as major protection mechanisms against ethanol toxicity [9]. To further decipher responses at transcriptional level, in this study, we applied a quantitative transcriptomics approach with a next-generation sequencing technology, combined with quantitative reverse-transcript PCR (RT-PCR) analysis, to reveal the global metabolic responses to ethanol in *Synechocystis* sp. PCC 6803 [10]. We then compared the transcriptomics data with proteomic data obtained previously to further confirm the targets related to ethanol tolerance [9]. Finally, we constructed several knockout mutants of ethanol-induced genes to validate their potential application as targets for engineering ethanol tolerance. The

RNA-seq transcriptomics analysis not only further confirmed the cellular responses revealed from previous proteomics analysis, but also showed that *Synechocystis* cells can also utilize enhanced PHA accumulation and glyoxalase detoxication pathway as means against ethanol stress. The study provided a list of gene targets for tolerance engineering in cyanobacterium *Synechocystis*.

## Results and discussion

### Ethanol effects on *Synechocystis* sp. PCC 6803

To make the transcriptomics data comparable with previous proteomics data, we used the identical sampling conditions for transcriptomics as our previous proteomics analysis [9]. As described before, the growth of *Synechocystis* sp. PCC 6803 supplemented with 0, 1.25, 1.50 and 2.00% ethanol was assessed to determine an appropriate ethanol concentration for proteomic studies. The results showed that the concentration of ethanol that caused a 50% growth decrease was found to be 1.50% (*v/v*) at 24 h (corresponding to middle-exponential phase), and was selected for the analysis in this study [9]. Cell morphology under ethanol-treated and control conditions was compared under microscope, and the results showed that visible aggregation of large number of cells was found after 24 h treatment even at a concentration of 1.50%, compared with the clearly individual cells in the control (data not shown). For transcriptomic analysis, two independent cultivations for both control (no ethanol) and 1.5% ethanol-treated experiments were conducted, and cells were collected by centrifugation (8,000 x *g* for 10 min at 4°C) at 24 h, 48 h and 72 h, resulting two biological replicates for each time point. The time points of sampling were corresponded to middle-exponential, exponential-stationary transition and stationary phases of the cell growth, respectively [9].

### Overview of transcriptomics analysis

A total of 112-million raw sequencing reads was obtained from the RNA-seq transcriptomics analysis of nine samples, with average reads of 12.5-million reads per sample. After a two-step data filtering process, first to eliminate reads with low-quality bases (such as multiple N) and reads shorter than 20 bp, and then to eliminate sequence reads mapped to non-coding RNA of *Synechocystis* sp. PCC 6803 [10], a total of 20.4-million qualified mRNA-based sequence reads were identified (Table 1). Except for the control sample at 24 h (Control-24 h) which has a genome mapping ratio of 57%, all other samples have mapping ratio larger than 60%, with the control sample at 72 h larger than 80%. Reproducibility between biological replicates of ethanol-treated samples at three time points was plotted (Figure 1), with correlation coefficient around 0.98-0.99, indicating the

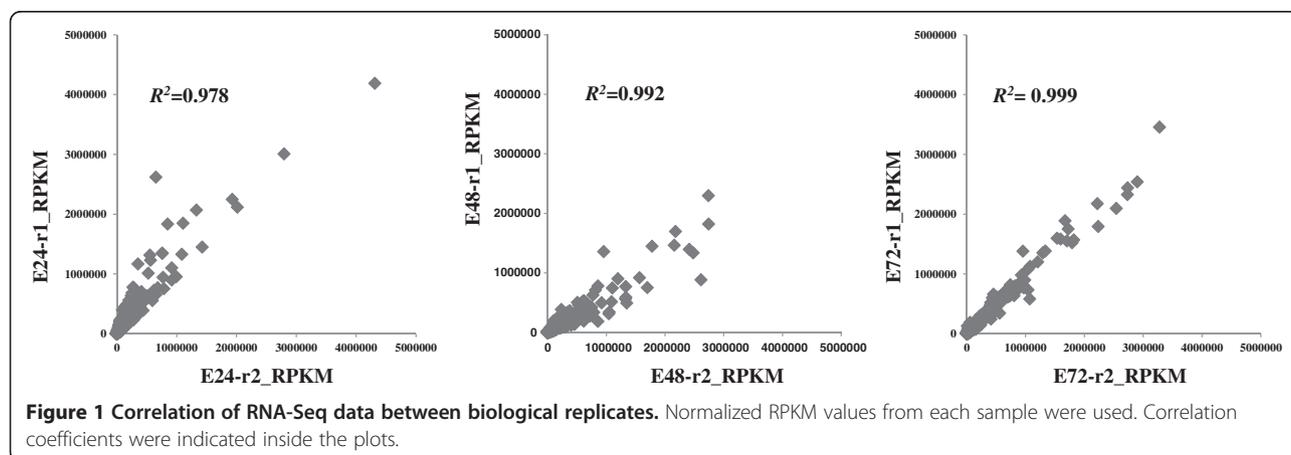
**Table 1 Statistics of RNA-Seq transcriptomics analysis**

| Sample ID       | Raw sequences reads | Qualified mRNA reads | Genome-mapped Reads | Mapping ratio |
|-----------------|---------------------|----------------------|---------------------|---------------|
| Control-24 h    | 18,859,596          | 5,552,572            | 3,177,168           | 57.22%        |
| Control-48 h    | 10,589,107          | 2,331,738            | 1,784,966           | 76.55%        |
| Control-72 h    | 11,249,884          | 3,026,360            | 2,455,571           | 81.14%        |
| Ethanol-24 h-r1 | 12,277,943          | 3,211,398            | 1,985,207           | 61.82%        |
| Ethanol-24 h-r2 | 13,736,430          | 3,755,422            | 2,323,029           | 61.86%        |
| Ethanol-48 h-r1 | 13,384,367          | 4,041,104            | 2,453,989           | 60.73%        |
| Ethanol-48 h-r2 | 10,306,514          | 2,934,288            | 1,772,957           | 60.42%        |
| Ethanol-72 h-r1 | 11,883,271          | 3,623,746            | 2,363,916           | 65.23%        |
| Ethanol-72 h-r2 | 9,814,306           | 3,160,336            | 2,083,580           | 65.93%        |

overall good quality of RNA sequencing. The sequence reads matched to all 3189 coding genes in *Synechocystis* sp. PCC 6803 genome (Additional file 1: Table S1), suggesting that the sequencing is deep enough to cover almost all species of transcripts in the cells. Abundance of the qualified mRNA-based raw sequence reads ranged from 1 to 341,135 for control samples, and from 1 to 154,326 for ethanol-treated samples, respectively, representing an expression dynamic range of  $10^5$ , which is higher than  $10^{3-4}$  of typical microarray-based analyses [11,12]. Using Reads Per Kilobase of Gene per Million Mapped Reads (RPKM) as an index of the normalized transcript abundance [13], we identified the top expressed genes under the control and ethanol-treated conditions through the growth time course (Table 2). The top 50 expressed genes were found involved mostly in energy metabolism, including genes coding photosynthesis-related phycocyanin alpha subunit, phycocyanin beta subunit and photosystem I subunit XI and genes coding multiple subunits of ATP synthase, followed by genes encoding proteins synthesis such as multiple 50S ribosomal proteins and elongation factor, and genes involved in CO<sub>2</sub> fixation such as ribulose biphosphate carboxylase genes, consistent well

with previous analysis on highly expressed genes in *Synechocystis* [14,15]. Interestingly, we also found several genes encoding hypothetical proteins (*ssl0483*, *slr0144*, *slr1470* and *slr0373*) were also among the top expressed genes, suggesting possible important physiological functions they may be responsible for. Although the exact function is still unknown, *slr0144* has been suggested to encode a PSII-associated protein [16], and *slr0373* forms an operon with *slr0374* which has been found responsive to various environmental stresses [17].

Using a cutoff of 1.5-fold change in both biological replicates, we determined that 1874 and 274 genes were down- and up-regulated by ethanol, respectively. For the down-regulated genes, 1343, 596 and 830 genes were down-regulated at 24, 48 and 72 h, respectively. Among them 167 genes were down-regulated in all three time points (Additional file 2: Table S2). Analysis of the functional category of the down-regulated genes was shown in Figure 2. The results showed that the most affected functional categories were “hypothetical proteins” and “unknown function”, representing a total of more than 68% of all the down-regulated genes, consistent with the fact that nearly half of the *Synechocystis*

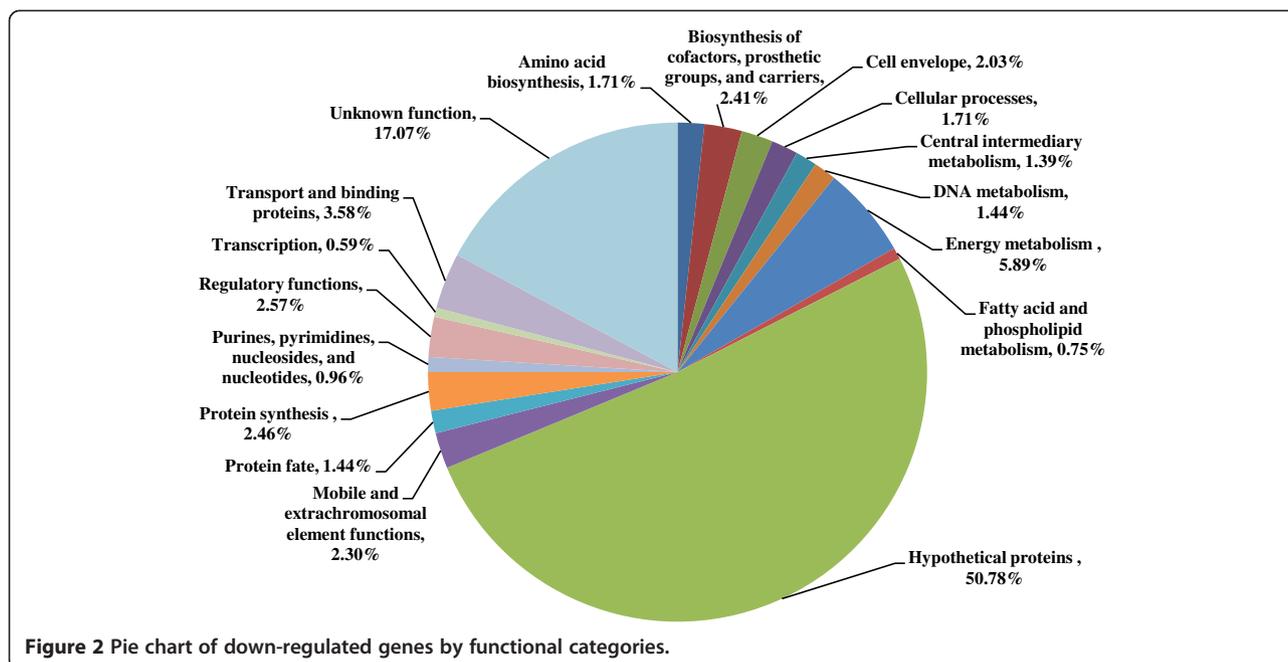


**Table 2 Top 50 expressed genes based on normalized expression level (RPKM values)**

| Gene ID        | Control-24h | Control-48h | Control-72h | Ethanol-24h-r1 | Ethanol-24h-r2 | Ethanol-48h-r1 | Ethanol-48h-r2 | Ethanol-72h-r1 | Ethanol-72h-r2 | Description  | Cellular role         |
|----------------|-------------|-------------|-------------|----------------|----------------|----------------|----------------|----------------|----------------|--|-----------------------|
| <i>sl11744</i> | 7.42E+05    | 8.96E+05    | 9.57E+05    | 6.14E+05       | 6.66E+05       | 6.65E+05       | 3.97E+05       | 7.24E+05       | 6.49E+05       | 50S ribosomal protein L1   | Protein synthesis     |
| <i>sl11745</i> | 3.42E+06    | 3.74E+06    | 3.97E+06    | 2.80E+06       | 3.01E+06       | 2.74E+06       | 1.82E+06       | 2.90E+06       | 2.54E+06       | 50S ribosomal protein L10  | Protein synthesis     |
| <i>sl11743</i> | 3.56E+05    | 5.73E+05    | 6.30E+05    | 3.00E+05       | 3.86E+05       | 4.17E+05       | 2.23E+05       | 4.90E+05       | 4.11E+05       | 50S ribosomal protein L11  | Protein synthesis     |
| <i>sl11746</i> | 6.76E+05    | 5.38E+05    | 7.26E+05    | 1.66E+05       | 2.49E+05       | 2.90E+05       | 1.78E+05       | 2.78E+05       | 2.58E+05       | 50S ribosomal protein L12  | Protein synthesis     |
| <i>sl10767</i> | 2.07E+06    | 9.55E+05    | 2.40E+06    | 9.92E+05       | 9.52E+05       | 7.76E+05       | 6.25E+05       | 7.37E+05       | 6.30E+05       | 50S ribosomal protein L20  | Protein synthesis     |
| <i>sl11801</i> | 3.47E+05    | 5.18E+05    | 7.09E+05    | 1.62E+05       | 2.61E+05       | 4.74E+05       | 2.32E+05       | 5.43E+05       | 5.32E+05       | 50S ribosomal protein L23  | Protein synthesis     |
| <i>sl11807</i> | 4.12E+05    | 6.41E+05    | 9.77E+05    | 1.72E+05       | 3.12E+05       | 5.68E+05       | 2.51E+05       | 6.14E+05       | 5.94E+05       | 50S ribosomal protein L24  | Protein synthesis     |
| <i>ssl1426</i> | 5.89E+06    | 2.35E+06    | 5.49E+06    | 4.31E+06       | 4.18E+06       | 2.74E+06       | 2.30E+06       | 2.22E+06       | 2.17E+06       | 50S ribosomal protein L35  | Protein synthesis     |
| <i>slr2067</i> | 1.00E+06    | 2.39E+06    | 4.47E+06    | 3.52E+05       | 1.16E+06       | 2.61E+06       | 8.80E+05       | 3.28E+06       | 3.45E+06       | Allophycocyanin alpha subunit  | Energy metabolism     |
| <i>slr1986</i> | 8.68E+05    | 1.50E+06    | 2.98E+06    | 2.53E+05       | 6.79E+05       | 1.35E+06       | 4.90E+05       | 1.72E+06       | 1.75E+06       | Allophycocyanin beta subunit   | Energy metabolism     |
| <i>slr1198</i> | 1.72E+06    | 1.22E+06    | 1.77E+06    | 1.08E+06       | 1.33E+06       | 1.10E+06       | 7.41E+05       | 1.08E+06       | 1.12E+06       | Ant ioxidant protein   | Unclassified          |
| <i>sl11322</i> | 2.11E+05    | 6.36E+05    | 8.15E+05    | 1.31E+05       | 4.50E+05       | 6.64E+05       | 3.23E+05       | 9.71E+05       | 7.65E+05       | ATP synthase A chain of CF(0)  | Energy metabolism     |
| <i>ssl2615</i> | 4.95E+05    | 1.22E+06    | 2.06E+06    | 2.69E+05       | 7.74E+05       | 1.33E+06       | 5.88E+05       | 1.82E+06       | 1.57E+06       | ATP synthase C chain of CF(0)  | Energy metabolism     |
| <i>sl11323</i> | 2.28E+05    | 5.71E+05    | 8.68E+05    | 1.17E+05       | 3.22E+05       | 6.10E+05       | 2.82E+05       | 8.06E+05       | 6.33E+05       | ATP synthase subunit b' of CF(0)   | Energy metabolism     |
| <i>sl11099</i> | 4.39E+05    | 9.53E+05    | 1.51E+06    | 1.94E+05       | 4.07E+05       | 7.88E+05       | 3.36E+05       | 9.11E+05       | 8.63E+05       | Elongation factor Tu   | Protein synthesis     |
| <i>ssl0020</i> | 2.52E+06    | 6.33E+05    | 7.00E+05    | 6.84E+05       | 7.65E+05       | 8.17E+05       | 7.10E+05       | 8.75E+05       | 7.41E+05       | Ferredoxin I, essential for growth                                       | Energy metabolism     |
| <i>sl10018</i> | 1.97E+05    | 7.10E+05    | 1.21E+06    | 1.36E+05       | 2.37E+05       | 7.34E+05       | 3.61E+05       | 1.04E+06       | 1.08E+06       | Fructose-bisphosphate aldolase, class II                                 | Unclassified          |
| <i>ssl0483</i> | 1.48E+06    | 5.07E+05    | 6.64E+05    | 6.22E+05       | 7.23E+05       | 6.12E+05       | 5.26E+05       | 6.88E+05       | 6.09E+05       | Hypothetical protein   | No Data               |
| <i>slr0144</i> | 8.77E+05    | 5.27E+05    | 5.68E+05    | 3.89E+05       | 4.55E+05       | 3.04E+05       | 2.59E+05       | 2.49E+05       | 2.15E+05       | Hypothetical protein   | Hypothetical proteins |
| <i>slr1470</i> | 6.22E+05    | 6.04E+05    | 8.56E+05    | 4.92E+05       | 6.35E+05       | 6.70E+05       | 4.85E+05       | 7.30E+05       | 6.91E+05       | Hypothetical protein   | Hypothetical proteins |
| <i>slr0373</i> | 4.60E+05    | 5.51E+05    | 1.65E+06    | 1.51E+05       | 3.63E+05       | 7.25E+05       | 2.74E+05       | 9.89E+05       | 8.96E+05       | Hypothetical protein   | No Data               |
| <i>slr0749</i> | 9.08E+05    | 6.46E+05    | 1.19E+06    | 7.58E+05       | 1.34E+06       | 7.50E+05       | 4.19E+05       | 4.46E+05       | 4.20E+05       | Light-independent protochlorophyllide reductase ironprotein subunit ChL  | Cofactor biosynthesis |
| <i>slr0749</i> | 9.08E+05    | 6.46E+05    | 1.19E+06    | 7.58E+05       | 1.34E+06       | 7.50E+05       | 4.19E+05       | 4.46E+05       | 4.20E+05       | Light-independent protochlorophyllide reductase iron protein subunit ChL | Energy metabolism     |
| <i>sl11342</i> | 3.21E+05    | 5.41E+05    | 7.43E+05    | 2.32E+05       | 3.29E+05       | 4.17E+05       | 2.38E+05       | 4.80E+05       | 5.42E+05       | NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase                | Energy metabolism     |
| <i>slr1834</i> | 1.46E+06    | 1.55E+07    | 3.92E+07    | 1.23E+06       | 5.28E+06       | 1.28E+07       | 3.37E+06       | 1.50E+07       | 1.66E+07       | P700 apoprotein subunit Ia   | Energy metabolism     |
| <i>slr1835</i> | 1.21E+06    | 5.21E+06    | 1.16E+07    | 6.51E+05       | 2.62E+06       | 5.15E+06       | 1.74E+06       | 5.93E+06       | 5.79E+06       | P700 apoprotein subunit Ib   | Energy metabolism     |
| <i>sl10819</i> | 1.20E+06    | 6.37E+05    | 6.10E+05    | 3.32E+05       | 2.99E+05       | 3.72E+05       | 2.59E+05       | 4.10E+05       | 3.91E+05       | Photosystem I reaction center subunit III precursor                      | Energy metabolism     |
| <i>slr0737</i> | 1.85E+06    | 1.15E+06    | 2.11E+06    | 9.16E+05       | 1.10E+06       | 1.20E+06       | 8.97E+05       | 1.34E+06       | 1.38E+06       | Photosystem I subunit II   | Energy metabolism     |

**Table 2 Top 50 expressed genes based on normalized expression level (RPKM values) (Continued)**

|                |          |          |          |          |          |          |          |          |          |   |                       |
|----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|---|-----------------------|
| <i>smr0004</i> | 4.02E+06 | 3.59E+06 | 5.24E+06 | 1.43E+06 | 1.45E+06 | 1.78E+06 | 1.44E+06 | 1.60E+06 | 1.58E+06 | Photosystem I subunit VIII                      | Energy metabolism     |
| <i>slr1655</i> | 1.08E+07 | 5.24E+06 | 5.95E+06 | 8.83E+06 | 8.55E+06 | 7.79E+06 | 5.87E+06 | 7.50E+06 | 6.75E+06 | Photosystem I subunit XI                        | Energy metabolism     |
| <i>slr0906</i> | 2.96E+05 | 5.56E+05 | 4.91E+05 | 1.78E+05 | 4.12E+05 | 5.30E+05 | 2.60E+05 | 7.12E+05 | 6.19E+05 | Photosystem II core light harvest ing protein   | Energy metabolism     |
| <i>sll0851</i> | 2.98E+05 | 1.21E+06 | 2.33E+06 | 1.71E+05 | 4.95E+05 | 1.05E+06 | 3.40E+05 | 9.30E+05 | 9.78E+05 | Photosystem II CP43 protein                     | Energy metabolism     |
| <i>slr1311</i> | 2.37E+06 | 1.75E+06 | 2.79E+06 | 8.49E+05 | 1.83E+06 | 2.41E+06 | 1.39E+06 | 2.73E+06 | 2.44E+06 | Photosystem II D1 protein                       | Energy metabolism     |
| <i>sll1867</i> | 1.22E+06 | 1.06E+06 | 1.62E+06 | 5.25E+05 | 1.01E+06 | 1.33E+06 | 7.68E+05 | 1.54E+06 | 1.59E+06 | Photosystem II D1 protein                       | Energy metabolism     |
| <i>sll0849</i> | 1.31E+05 | 1.09E+06 | 1.59E+06 | 9.21E+04 | 3.93E+05 | 1.05E+06 | 3.10E+05 | 1.21E+06 | 1.20E+06 | Photosystem II react ion center D2 protein      | Energy metabolism     |
| <i>slr0335</i> | 4.12E+05 | 1.38E+06 | 2.12E+06 | 2.41E+05 | 6.30E+05 | 1.09E+06 | 5.13E+05 | 1.30E+06 | 1.35E+06 | Phycobilisome core-membrane linker polypept ide | Energy metabolism     |
| <i>sll1580</i> | 3.09E+06 | 2.89E+06 | 2.91E+06 | 1.93E+06 | 2.24E+06 | 2.18E+06 | 1.69E+06 | 2.23E+06 | 1.79E+06 | Phycobilisome rod linker polypept ide           | Energy metabolism     |
| <i>sll1579</i> | 2.30E+06 | 2.95E+06 | 3.31E+06 | 1.33E+06 | 2.07E+06 | 2.16E+06 | 1.46E+06 | 2.54E+06 | 2.09E+06 | Phycobilisome rod linker polypept ide           | Unclassified          |
| <i>slr2051</i> | 1.44E+06 | 6.37E+05 | 7.37E+05 | 4.63E+05 | 5.46E+05 | 6.26E+05 | 5.26E+05 | 6.77E+05 | 6.51E+05 | Phycobilisome rod-core linker polypeptide       | Energy metabolism     |
| <i>ssl3093</i> | 2.10E+06 | 1.29E+06 | 1.36E+06 | 7.71E+05 | 9.40E+05 | 8.59E+05 | 7.79E+05 | 1.04E+06 | 7.32E+05 | Phycobilisome small rod linker polypeptide      | Unclassified          |
| <i>sll1578</i> | 2.59E+07 | 5.16E+07 | 8.62E+07 | 1.43E+07 | 2.65E+07 | 4.10E+07 | 2.17E+07 | 5.20E+07 | 5.40E+07 | Phycocyanin alpha subunit                       | Energy metabolism     |
| <i>sll1577</i> | 2.49E+07 | 5.74E+07 | 8.88E+07 | 1.88E+07 | 3.58E+07 | 5.52E+07 | 2.70E+07 | 7.35E+07 | 7.72E+07 | Phycocyanin beta subunit                        | Energy metabolism     |
| <i>sll1694</i> | 5.95E+06 | 8.44E+05 | 1.14E+06 | 2.01E+06 | 2.12E+06 | 9.58E+05 | 1.35E+06 | 1.07E+06 | 5.81E+05 | Pilin polypept ide PilA1                        | Cell envelope         |
| <i>sll0199</i> | 8.92E+05 | 6.80E+05 | 9.96E+05 | 2.02E+05 | 5.59E+05 | 9.24E+05 | 4.93E+05 | 8.49E+05 | 7.72E+05 | Plastocyanin                                    | Unclassified          |
| <i>slr0011</i> | 1.94E+06 | 3.06E+06 | 6.38E+06 | 1.11E+06 | 1.85E+06 | 2.48E+06 | 1.33E+06 | 2.73E+06 | 2.32E+06 | Possible Rubisco chaperonin                     | Hypothetical proteins |
| <i>slr1841</i> | 1.24E+06 | 2.09E+06 | 3.27E+06 | 5.53E+05 | 1.31E+06 | 1.70E+06 | 7.51E+05 | 1.67E+06 | 1.89E+06 | Probable porin; major outer membrane protein    | Unclassified          |
| <i>slr0009</i> | 4.48E+05 | 1.44E+06 | 2.53E+06 | 2.81E+05 | 7.69E+05 | 1.33E+06 | 5.61E+05 | 1.70E+06 | 1.55E+06 | Ribulose bisphosphate carboxylase large subunit | Energy metabolism     |
| <i>slr0012</i> | 5.54E+05 | 8.40E+05 | 2.10E+06 | 1.58E+05 | 2.52E+05 | 4.06E+05 | 2.72E+05 | 4.43E+05 | 3.79E+05 | Ribulose bisphosphate carboxylase small subunit | Energy metabolism     |
| <i>sll1338</i> | 9.93E+05 | 2.56E+06 | 4.45E+06 | 5.64E+05 | 1.23E+06 | 1.57E+06 | 9.14E+05 | 1.79E+06 | 1.52E+06 | Unknown protein                                 | No Data               |
| <i>sll1951</i> | 7.54E+05 | 8.37E+05 | 1.25E+06 | 3.14E+05 | 6.19E+05 | 7.53E+05 | 3.27E+05 | 7.44E+05 | 8.16E+05 | Unknown protein                                 | Unclassified          |



genome is still annotated as hypothetical up to now [10,18]. Other most affected functional categories included “Energy metabolism”, “Protein synthesis” and “Regulatory functions”. Down-regulation of the central metabolism is consistent with the overall slower growth upon ethanol stress [9]. For the up-regulated genes, 29, 114 and 161 genes were up-regulated at 24 h, 48 h and 72 h, respectively, among which 3 genes were up-regulated in all three time points (Table 3 and Additional file 3: Table S3). More genes up-regulated at late growth phases suggested that cells needed time to adjust their metabolism and initiate resistance responses.

#### Correlation with quantitative RT-PCR analysis

Based on their expression level and regulation patterns by ethanol, a subset of 12 genes was selected for quantitative RT-PCR validation. Among them, six genes were down-regulated (*i.e.* *sll0721*, *sll1796*, *slr1992*, *sll0248*, *sll1327*, *ssr1399*) and six genes were up-regulated (*i.e.* *sll1734*, *slr1761*, *slr1828*, *sll1091*, *slr0288*, *sll0057*) by ethanol, respectively according to the RNA-seq transcriptomics data. Under control condition, their expression levels varied from the normalized RPKM values 2529.6 for *sll0248* (encoding a flavodoxin) to 421749.3 for *ssr1399* (encoding ribosomal protein S18) (Additional file 4: Table S4). RT-PCR analysis was performed for the genes between the treated sample and control for all three time points (*i.e.* 24, 48 and 72 h). The results showed obvious positive correlation can be detected between qRT-PCR and RNA-Seq transcriptomics data (with correlation coefficient of 0.75-0.8) (Figure 3), suggesting a good quality of RNA-seq data.

#### Cells utilize multiple approaches to cope with ethanol stress

Our previous proteomic analysis found that the *Synechocystis* cells employed a combination of induced common stress response, modifications of cell membrane and envelope, and induction of multiple transporters and cell mobility-related proteins as protection mechanisms against ethanol toxicity [9]. At transcriptional level, a very similar response was also observed [9]. First, we found that common stress responses were induced: one gene encoding a heat-shock DnaK homolog (*slr0086*) was induced at 72 h. Multiple genes involved in resistance against reactive oxygen species (ROS), such as *slr2033* encoding a membrane-associated rubredoxin, *slr1109* encoding ankyrin homolog [19], *sll1545* encoding glutathione S-transferase [20], *slr0242* encoding a bacterioferritin comigratory protein [21] and *slr1379* encoding quinol oxidase subunit I [22] were up-regulated. In addition, consistent with findings from proteomic analysis, we found circadian rhythms of *Synechocystis* sp. PCC 6803 was also regulated by ethanol. It was reported that cyanobacterial circadian rhythms are controlled by a cluster of three genes, *kaiA*, *kaiB*, and *kaiC* [23]. Previous proteomic analysis showed that one of the key circadian clock proteins, KaiB (Slr0757), was induced [9]. RNA-Seq transcriptomic analysis showed that *kaiC* gene (*slr0758*) was also induced (Table 3). Transcriptomics analysis here complemented well with the proteomic analysis, further confirming that circadian rhythms are induced by ethanol treatment. The ethanol-induced genes were listed in Table 3, while the induced genes

**Table 3 Genes induced by ethanol exposure \***

| Gene ID        | Description   | Ratio-Ethanol-24h-r1 vs. Control-24h | Ratio-Ethanol-24h-r2 vs. Control-24h | Ratio-Ethanol-48h-r1 vs. Control-48h | Ethanol-48h-r2 vs. Control-48h | Ethanol-72h-r1 vs. Control-72h | Ratio-Ethanol-72h-r1 vs. Control-72h |
|----------------|---|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------|--------------------------------|--------------------------------------|
| <i>sll0034</i> | putative carboxypeptidase                             |                                      |                                      | <b>2.69</b>                          | <b>1.50</b>                    |                                |                                      |
| <i>sll0250</i> | pantothenate metabolism flavoprotein                  |                                      |                                      |                                      |                                | <b>3.00</b>                    | <b>1.91</b>                          |
| <i>sll0289</i> | septum site-determining protein MinD                  |                                      |                                      |                                      |                                | <b>4.56</b>                    | <b>3.89</b>                          |
| <i>sll0300</i> | riboflavin synthase alpha chain                       |                                      |                                      |                                      |                                | <b>4.33</b>                    | <b>3.33</b>                          |
| <i>sll0330</i> | sepiapterine reductase                                |                                      |                                      |                                      |                                | <b>7.09</b>                    | <b>29.09</b>                         |
| <i>sll0368</i> | uracil phosphoribosyltransferase                      |                                      |                                      |                                      |                                | <b>2.33</b>                    | <b>2.00</b>                          |
| <i>sll0374</i> | urea transport system ATP-binding protein             |                                      |                                      | <b>2.80</b>                          | <b>3.60</b>                    |                                |                                      |
| <i>sll0384</i> | Cations and iron carrying protein                     |                                      |                                      | <b>2.31</b>                          | <b>1.70</b>                    |                                |                                      |
| <i>sll0450</i> | cytochrome b subunit of nitric oxide reductase        | <b>1.62</b>                          | <b>4.26</b>                          |                                      |                                |                                |                                      |
| <i>sll0536</i> | probable potassium channel protein                    |                                      |                                      | <b>2.00</b>                          | <b>1.93</b>                    |                                |                                      |
| <i>sll0540</i> | phosphate-binding protein PstS homolog                |                                      |                                      |                                      |                                | <b>1.57</b>                    | <b>1.79</b>                          |
| <i>sll0613</i> | Holliday junction DNA helicase RuvB                   |                                      |                                      | <b>1.73</b>                          | <b>1.91</b>                    |                                |                                      |
| <i>sll0621</i> | putative c-type cytochrome biogenesis protein CcdA    |                                      |                                      |                                      |                                | <b>2.22</b>                    | <b>2.00</b>                          |
| <i>sll0629</i> | alternative photosystem I reaction center subunit X   |                                      |                                      | <b>1.94</b>                          | <b>1.72</b>                    |                                |                                      |
| <i>sll0671</i> | probable cation transporter                           | <b>4.00</b>                          | <b>4.00</b>                          |                                      |                                |                                |                                      |
| <i>sll0686</i> | probable cytochrome c-type biogenesis protein         |                                      |                                      |                                      |                                | <b>3.00</b>                    | <b>3.25</b>                          |
| <i>sll0687</i> | RNA polymerase ECF-type (group 3) sigma factor        |                                      |                                      | <b>3.00</b>                          | <b>6.00</b>                    |                                |                                      |
| <i>sll0759</i> | ABC transporter ATP-binding protein                   |                                      |                                      |                                      |                                | <b>2.13</b>                    | <b>2.09</b>                          |
| <i>sll0759</i> | ABC transporter ATP-binding protein                   |                                      |                                      |                                      |                                | <b>2.13</b>                    | <b>2.09</b>                          |
| <i>sll0792</i> | Zinc-responsive repressor ZiaR                        |                                      |                                      | <b>3.43</b>                          | <b>3.00</b>                    |                                |                                      |
| <i>sll0856</i> | RNA polymerase ECF-type (group 3) sigma-E factor      |                                      |                                      |                                      |                                | <b>1.89</b>                    | <b>1.61</b>                          |
| <i>sll1041</i> | similar to sulfate transport ATP-binding protein CysA |                                      |                                      | <b>2.63</b>                          | <b>1.63</b>                    |                                |                                      |
| <i>sll1051</i> | phycocyanin alpha-subunit phycocyanobilin lyase       |                                      |                                      |                                      |                                | <b>8.00</b>                    | <b>13.00</b>                         |
| <i>sll1051</i> | phycocyanin alpha-subunit phycocyanobilin lyase       |                                      |                                      |                                      |                                | <b>8.00</b>                    | <b>13.00</b>                         |
| <i>sll1170</i> | unknown protein                                       |                                      |                                      |                                      |                                | <b>2.25</b>                    | <b>1.50</b>                          |
| <i>sll1223</i> | diaphorase subunit of the bidirectional hydrogenase   | <b>1.55</b>                          | <b>2.18</b>                          |                                      |                                |                                |                                      |
| <i>sll1226</i> | hydrogenase subunit of the bidirectional hydrogenase  |                                      |                                      |                                      |                                | <b>1.55</b>                    | <b>1.89</b>                          |
| <i>sll1316</i> | cytochrome b6-f complex iron-sulfur subunit           |                                      |                                      |                                      |                                | <b>1.69</b>                    | <b>1.56</b>                          |

**Table 3 Genes induced by ethanol exposure \* (Continued)**

|                |  |             |             |             |             |             |              |
|----------------|--|-------------|-------------|-------------|-------------|-------------|--------------|
| <i>sll1330</i> | two-component system response regulator OmpR subfamily   |             |             | <b>2.47</b> | <b>2.33</b> |             |              |
| <i>sll1370</i> | mannose-1-phosphate guanylyltransferase                  |             |             |             |             | <b>3.25</b> | <b>1.88</b>  |
| <i>sll1423</i> | global nitrogen regulator                                |             |             | <b>1.57</b> | <b>1.55</b> |             |              |
| <i>sll1428</i> | probable sodium-dependent transporter                    | <b>7.00</b> | <b>6.00</b> |             |             | <b>5.00</b> | <b>2.00</b>  |
| <i>sll1440</i> | pyridoxamine 5'-phosphate oxidase                        |             |             |             |             | <b>3.00</b> | <b>1.50</b>  |
| <i>sll1471</i> | phycobilisome rod-core linker polypeptide                | <b>3.90</b> | <b>2.77</b> | <b>3.56</b> | <b>3.07</b> |             |              |
| <i>sll1473</i> | a part of phytochrome-like sensor histidine kinase gene  |             |             | <b>1.59</b> | <b>1.79</b> |             |              |
| <i>sll1483</i> | periplasmic protein                                      |             |             |             |             | <b>2.00</b> | <b>8.40</b>  |
| <i>sll1545</i> | glutathione S-transferase                                |             |             | <b>2.00</b> | <b>1.60</b> |             |              |
| <i>sll1612</i> | folylpolyglutamate synthase                              | <b>2.00</b> | <b>2.67</b> |             |             |             |              |
| <i>sll1679</i> | periplasmic protease HhoA                                | <b>3.35</b> | <b>2.42</b> |             |             |             |              |
| <i>sll1682</i> | alanine dehydrogenase                                    |             |             |             |             | <b>2.91</b> | <b>2.27</b>  |
| <i>sll1723</i> | probable glycosyltransferase                             |             |             |             |             | <b>1.53</b> | <b>4.80</b>  |
| <i>sll1724</i> | probable glycosyltransferase                             |             |             |             |             | <b>2.11</b> | <b>4.89</b>  |
| <i>sll1994</i> | porphobilinogen synthase (5-aminolevulinate dehydratase) |             |             |             |             | <b>1.97</b> | <b>1.83</b>  |
| <i>sll1998</i> | putative transposase [ISY100d: 1623697-1624643]          |             |             |             |             | <b>1.67</b> | <b>4.00</b>  |
| <i>slr0018</i> | fumarase   |             |             |             |             | <b>2.27</b> | <b>1.67</b>  |
| <i>slr0051</i> | periplasmic beta-type carbonic anhydrase                 |             |             | <b>1.55</b> | <b>1.60</b> |             |              |
| <i>slr0070</i> | methionyl-tRNA formyltransferase                         |             |             |             |             | <b>1.62</b> | <b>1.69</b>  |
| <i>slr0086</i> | similar to DnaK protein                                  |             |             |             |             | <b>1.50</b> | <b>2.50</b>  |
| <i>slr0089</i> | gamma-tocopherol methyltransferase                       |             |             |             |             | <b>4.60</b> | <b>3.80</b>  |
| <i>slr0091</i> | aldehyde dehydrogenase                                   |             |             |             |             | <b>7.00</b> | <b>11.00</b> |
| <i>slr0242</i> | bacterioferritin comigratory protein homolog             |             |             | <b>1.58</b> | <b>1.65</b> |             |              |
| <i>slr0328</i> | low molecular weight phosphotyrosine protein phosphatase |             |             |             |             | <b>2.75</b> | <b>2.00</b>  |
| <i>slr0381</i> | lactoylglutathione lyase                                 |             |             |             |             | <b>5.14</b> | <b>5.00</b>  |
| <i>slr0502</i> | cobalamin synthesis protein cobW homolog                 | <b>1.75</b> | <b>1.88</b> |             |             | <b>1.83</b> | <b>1.67</b>  |
| <i>slr0574</i> | cytochrome P450  |             |             |             |             | <b>2.34</b> | <b>2.41</b>  |
| <i>slr0585</i> | argininosuccinate synthetase                             |             |             |             |             | <b>2.32</b> | <b>2.42</b>  |
| <i>slr0618</i> | cobyric acid synthase                                    |             |             | <b>1.69</b> | <b>1.56</b> |             |              |
| <i>slr0678</i> | biopolymer transport ExbD like protein                   |             |             |             |             | <b>2.19</b> | <b>2.26</b>  |
| <i>slr0721</i> | malic enzyme   |             |             |             |             | <b>1.67</b> | <b>2.29</b>  |
| <i>slr0724</i> | HtaR suppressor protein homolog                          |             |             |             |             | <b>2.00</b> | <b>1.50</b>  |
| <i>slr0741</i> | transcriptional regulator                                |             |             |             |             | <b>2.13</b> | <b>2.63</b>  |
| <i>slr0758</i> | circadian clock protein KaiC homolog                     |             |             | <b>2.16</b> | <b>1.52</b> |             |              |
| <i>slr0819</i> | apolipoprotein N-acyltransferase                         |             |             |             |             | <b>2.17</b> | <b>1.67</b>  |
| <i>slr0898</i> | ferredoxin-nitrite reductase                             |             |             | <b>1.65</b> | <b>1.53</b> |             |              |
| <i>slr0903</i> | molybdopterin (MPT) converting factor, subunit 2         | <b>3.00</b> | <b>2.50</b> |             |             |             |              |
| <i>slr0940</i> | zeta-carotene desaturase                                 |             |             | <b>1.87</b> | <b>2.00</b> | <b>2.03</b> | <b>1.70</b>  |
| <i>slr0942</i> | alcohol dehydrogenase [NADP+]                            |             |             | <b>1.76</b> | <b>1.58</b> |             |              |

**Table 3 Genes induced by ethanol exposure \* (Continued)**

|                |  |             |             |             |              |             |
|----------------|--|-------------|-------------|-------------|--------------|-------------|
| <i>slr0946</i> | arsenate reductase   |             |             |             | <b>4.50</b>  | <b>3.50</b> |
| <i>slr0947</i> | response regulator for energy transfer from phycobilisomes to photosystems |             |             |             | <b>2.56</b>  | <b>1.74</b> |
| <i>slr0949</i> | Integral membrane protein of the ABC-type Nat permease NatD                |             |             |             | <b>14.00</b> | <b>7.00</b> |
| <i>slr1093</i> | 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase        |             | <b>1.73</b> | <b>1.94</b> |              |             |
| <i>slr1109</i> | similar to ankyrin   |             |             |             | <b>2.09</b>  | <b>1.53</b> |
| <i>slr1120</i> | type 4 prepilin-like proteins leader peptide processing enzyme             | <b>8.00</b> | <b>3.00</b> |             |              |             |
| <i>slr1185</i> | cytochrome b6-f complex alternative iron-sulfur subunit                    |             |             |             | <b>8.00</b>  | <b>5.00</b> |
| <i>slr1197</i> | SMF protein  |             |             |             | <b>1.73</b>  | <b>1.53</b> |
| <i>slr1204</i> | protease   |             | <b>9.52</b> | <b>1.96</b> | <b>2.45</b>  | <b>6.65</b> |
| <i>slr1205</i> | similar to chlorobenzene dioxygenase, ferredoxin component                 |             |             |             | <b>4.00</b>  | <b>4.00</b> |
| <i>slr1225</i> | serine/threonine kinase  |             |             |             | <b>2.70</b>  | <b>2.50</b> |
| <i>slr1291</i> | NADH dehydrogenase subunit 4   |             |             |             | <b>2.44</b>  | <b>2.44</b> |
| <i>slr1300</i> | similar to 2-octaprenyl-6-methoxyphenol hydroxylase                        |             |             |             | <b>1.59</b>  | <b>1.59</b> |
| <i>slr1350</i> | acyl-lipid desaturase  |             |             |             | <b>1.72</b>  | <b>1.57</b> |
| <i>slr1379</i> | quinol oxidase subunit I   | <b>1.56</b> | <b>1.70</b> |             |              |             |
| <i>slr1418</i> | dihydroorotate dehydrogenase   |             |             |             | <b>2.50</b>  | <b>2.63</b> |
| <i>slr1452</i> | sulfate transport system substrate-binding protein                         | <b>1.83</b> | <b>2.00</b> |             |              |             |
| <i>slr1596</i> | a protein in the cytoplasmic membrane                                      |             |             |             | <b>2.31</b>  | <b>2.08</b> |
| <i>slr1626</i> | dihydroneopterin aldolase  |             | <b>1.80</b> | <b>3.30</b> |              |             |
| <i>slr1805</i> | two-component sensor histidine kinase                                      |             |             |             | <b>1.60</b>  | <b>1.98</b> |
| <i>slr1828</i> | ferredoxin, petF-like protein  |             | <b>2.00</b> | <b>2.33</b> | <b>5.50</b>  | <b>9.00</b> |
| <i>slr1848</i> | histidinol dehydrogenase   |             |             |             | <b>2.15</b>  | <b>2.06</b> |
| <i>slr1848</i> | histidinol dehydrogenase   |             |             |             | <b>2.15</b>  | <b>2.06</b> |
| <i>slr1853</i> | carboxymuconolactone decarboxylase   | <b>1.80</b> | <b>1.70</b> | <b>2.71</b> | <b>1.59</b>  | <b>2.56</b> |
| <i>slr1854</i> | unknown protein  |             |             | <b>2.10</b> | <b>1.84</b>  | <b>1.88</b> |
| <i>slr1874</i> | D-alanine-D-alanine ligase   |             |             | <b>1.84</b> | <b>1.88</b>  |             |
| <i>slr1877</i> | 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase                              |             |             | <b>2.00</b> | <b>1.92</b>  |             |
| <i>slr1884</i> | tryptophanyl-tRNA synthetase   |             |             |             | <b>2.09</b>  | <b>1.72</b> |
| <i>slr1910</i> | probable N-acetylmuramoyl-L-alanine amidase                                |             |             |             | <b>1.71</b>  | <b>2.00</b> |
| <i>slr1933</i> | dTDP-4-dehydrorhamnose 3,5-epimerase                                       |             |             |             | <b>3.50</b>  | <b>5.50</b> |
| <i>slr1938</i> | putative translation initiation factor EIF-2b subunit 1                    |             | <b>2.45</b> | <b>2.00</b> |              |             |
| <i>slr1962</i> | probable extracellular solute-binding protein                              |             | <b>1.79</b> | <b>1.50</b> |              |             |
| <i>slr1993</i> | PHA-specific beta-ketothiolase   |             |             |             | <b>1.93</b>  | <b>2.00</b> |
| <i>slr1994</i> | PHA-specific acetoacetyl-CoA reductase                                     | <b>6.00</b> | <b>9.00</b> | <b>5.00</b> | <b>2.00</b>  | <b>2.57</b> |
| <i>slr2033</i> | membrane-associated rubredoxin   | <b>1.85</b> | <b>3.38</b> |             | <b>2.23</b>  | <b>1.56</b> |
| <i>slr2114</i> | perosamine synthetase  |             |             |             | <b>7.00</b>  | <b>6.00</b> |

**Table 3 Genes induced by ethanol exposure \* (Continued)**

|                |  |      |      |      |      |
|----------------|--|------|------|------|------|
| <i>slr2131</i> | RND multidrug efflux transporter                                     |      |      | 3.45 | 3.44 |
| <i>slr2143</i> | L-cysteine/cystine lyase   | 1.67 | 2.56 |      |      |
| <i>smr0003</i> | cytochrome b6-f complex subunit PetM                                 |      |      | 2.40 | 3.40 |
| <i>smr0009</i> | photosystem II PsbN protein  |      | 1.56 | 1.78 |      |
| <i>ssl0563</i> | photosystem I subunit VII  |      |      | 1.76 | 1.56 |
| <i>ssl0707</i> | nitrogen regulatory protein P-II                                     |      | 2.00 | 1.64 |      |
| <i>ssl2153</i> | probable ribose phosphate isomerase B                                |      | 4.50 | 2.50 |      |
| <i>ssl2296</i> | pterin-4a-carbinolamine dehydratase                                  |      | 1.75 | 1.86 |      |
| <i>ssl2542</i> | high light-inducible polypeptide HliA, CAB/<br>ELIP/HLIP superfamily |      |      | 4.00 | 5.00 |
| <i>ssl3580</i> | putative hydrogenase expression/formation<br>protein HypC            |      | 3.15 | 2.05 |      |
| <i>ssr1176</i> | putative transposase   |      | 1.50 | 1.83 |      |
| <i>ssr1480</i> | putative RNA-binding protein   |      | 2.42 | 1.92 |      |

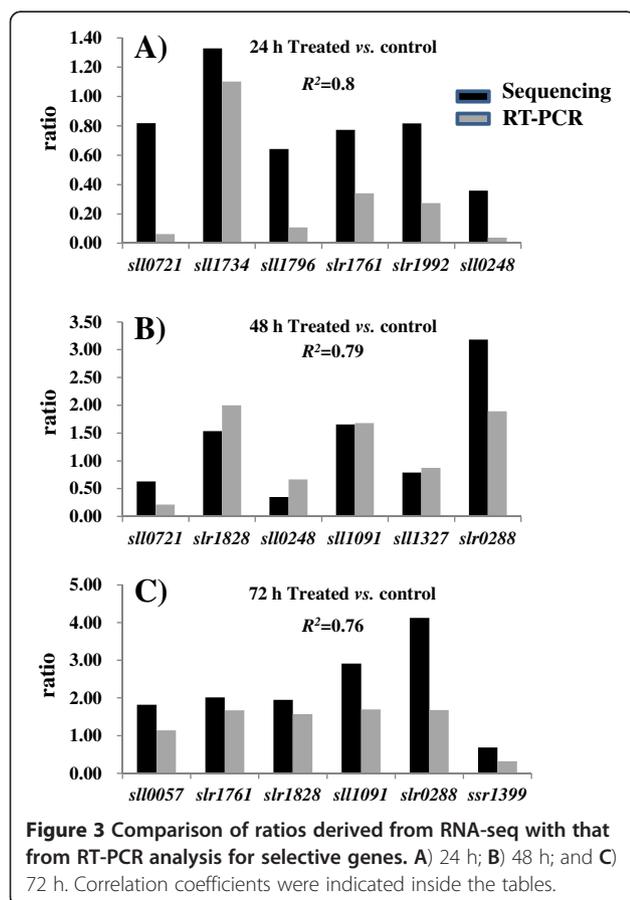
\* Induced genes encoding hypothetical proteins are provided in Additional file 3: Table S3.

encoding hypothetical proteins were provided in Additional file 3: Table S3.

Cross-membrane transporters for small molecules have been suggested as one important mechanism against ethanol toxicity in the early studies with yeast [24,25]. In cyanobacteria, transporters were also involved in tolerance to many different types of stresses, such as arsenate, Cu<sup>2+</sup>,

salinity and heavy metals [26-30]. Our quantitative proteomic analysis also identified 5 putative transporters with different substrate specificity induced by ethanol exposure [9]. RNA-Seq based transcriptomics found 12 transporters were induced by ethanol at varying growth phases. Similarly, these transport proteins were also with a wide range of putative functions and substrate specificity: *sll0759* encoding an ABC transporter ATP-binding protein, *slr0949* encoding an integral membrane protein of the ABC-type Nat permease, *sll0540* encoding a phosphate-binding protein PstS homolog, *sll0671* encoding a probable cation transporter, *sll0536* encoding a probable potassium channel protein, *sll1428* encoding a probable sodium-dependent transporter, *slr2131* encoding a RND multidrug efflux transporter, *sll0384* encoding a cation and iron carrying protein, *sll1041* encoding a sulfate transport ATP-binding protein CysA, *sll0374* encoding a urea transport system ATP-binding protein, and *slr0678* encoding a bio-polymer transport ExbD like protein, and *slr1452* encoding a sulfate transport system substrate-binding protein. Interestingly, they represented a totally different set of ethanol-induced transporters when compared with transporters revealed by proteomics analysis [9], although they shared some similarity in terms of substrate specificity as two of previously identified transporters, Sll0689 as a sodium-dependent transporter and Slr1295 as an iron transporter.

Early studies have found that many microbes can modify their cell membrane and envelope to increase tolerance to ethanol [24,31]. One well described change is the shift from *cis* to *trans* unsaturated fatty acids to decrease membrane fluidity, resulting in a corresponding increase in solvent tolerance [8]. RNA-seq transcriptomics analysis showed that *slr1350* encoding acyl-lipid desaturase was up-regulated at 72 h. In a previous study, the acyl-lipid desaturase (*desA*) gene from *Synechocystis* sp. PCC6803 was expressed in prokaryotic (*E. coli*) and



eukaryotic (*Solanum tuberosum*) cells, which led to an enhanced cold tolerance due to increased unsaturated fatty acid concentration in their lipids [32]. Several genes encoding cell envelope proteins were found induced by ethanol exposure (Table 3). The *slr0819* gene encoding apolipoprotein N-acyltransferase was induced 2.17 and 1.67 fold in both biological replicates at 72 h. Apolipoprotein N-acyltransferase is able to transfer an acyl group from sn-1-glycerophospholipid to the free alpha-amino group of the N-terminal cysteine of apolipoproteins, resulting in mature triacylated lipoprotein which plays important role in bacterial survival in mice for *Staphylococcus aureus* [33,34]. The *sll1370* gene encoding a mannose-1-phosphate guanylyltransferase was induced 3.25 and 1.88 fold in both biological replicates at 72 h. Mannose-1-phosphate guanylyltransferase is involved in lipopolysaccharide biosynthesis which has been found necessary for adaptation to high external NaCl stress in *Rhizobium tropici* [35]. The *slr1910* gene encoding a probable N-acetylmuramoyl-L-alanine amidase was induced 1.71 and 2.00 fold in both biological replicates at 72 h. N-acetylmuramoyl-L-alanine amidase has been suggested involved in degradation and reconstruction of the cell peptidoglycan layer in *Anabaena* sp. strain PCC 7120 [36]. Up-regulation of these cell envelope proteins by ethanol exposure could contribute to strengthening cell wall and extracellular matrix for stress resistance, although the mechanism still needs more investigation.

Polyhydroxyalkanoates (PHAs) are highly reduced bacterial storage compounds that are accumulated in most bacteria during unbalanced growth conditions [37]. Accumulation and degradation of PHAs endow bacteria with enhanced survival, competition abilities, and stress tolerance, increasing fitness in changing environments [37,38]. RNA-seq analysis identified two genes involved in PHA biosynthesis, *slr1994* encoding a PHA-specific acetoacetyl-CoA reductase and *slr1993* encoding a PHA-specific beta-ketothiolase were up-regulated. Genetic analysis suggested that these two genes were probably located in the same operon [39]. Among them, *slr1994* encoding PHA-specific acetoacetyl-CoA reductase was up-regulated significantly at all three time points (*i.e.* 24, 48 and 72 h) with 6.0 and 9.0 fold increase in both biological replicates at 24 h (Table 3). Although PHA accumulation has been reported for many natural stress conditions [38], it is the first time to report that this pathway is also responsive to organic solvents and biofuels.

One factor that may affect the long-term survival of bacterial cells in a population is the level of damage incurred by macromolecules via the nonenzymatic process of glycation, which is responsible for the formation of several compounds identified as advanced glycation end

products (AGEs) [40]. Many biochemical pathways produce reactive dicarbonyl intermediates, such as glyoxal and methylglyoxal (MG), which can further react with DNA, proteins, or other biomolecules to form AGEs [40]. In *E. coli*, it has been found that the predominant MG detoxification system consisted of glyoxalase enzyme I which converts MG to S-lactoyl glutathione [41]. In plant, the level of MG is enhanced upon exposure to different abiotic stresses and overexpression of glyoxalase pathway genes can support survival and growth of transgenic plants under various abiotic stresses [42]. RNA-Seq analysis of the ethanol-treated cells showed that lactoylglutathione lyase (also called as glyoxalase enzyme I) was up-regulated significantly by 5.14 and 5.0 fold in both biological replicates at 72 h, suggesting that glyoxalase pathway may play important roles in resistance to ethanol stress in *Synechocystis*.

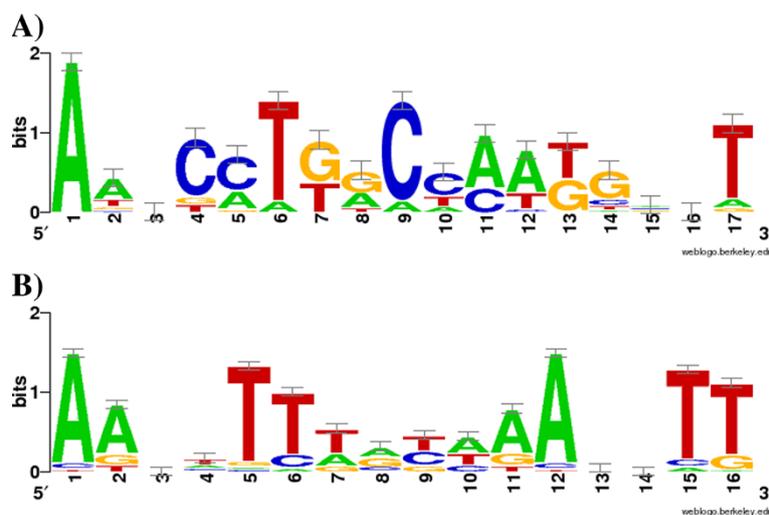
In the previous proteomics analysis, we unexpectedly discovered that many proteins involved in multiple aspects of photosynthesis activity (*i.e.* photosystem I and II, cytochrome, ferredoxin) were up-regulated even when the cell growth was slow down. We further confirmed the results by comparatively measuring chlorophyll a concentration in cells [9]. Based on our results we proposed that ethanol treatment might enhance photosynthesis in *Synechocystis* to generate more ROS which will trigger oxidative stress response [9]. RNA-seq transcriptomics analysis showed very similar results, although cell growth was slow, and genes involved in energy metabolism and protein synthesis were mostly down-regulated (Figure 2 and Additional file 2: Table S2), the genes involved in photosystem I and II, light collection and electron transfer, such as *ssl0563* encoding photosystem I subunit VII, *smr0009* encoding photosystem II PsbN protein, *sll1051* encoding phycocyanin alpha-subunit phycocyanobilin lyase, and *sll1471* encoding a phycobiosome rod-core linker polypeptide, and *slr1828* encoding a ferredoxin were up-regulated. Among them, *sll1051* encoding phycocyanin alpha-subunit phycocyanobilin lyase was increased significantly by 8.0 and 13.0 folds in both biological replicates at 72 h (Table 3). In addition, up-regulation of multiple cytochromes, such as *slr1185* encoding cytochrome *b6-f* complex alternative iron-sulfur subunit, *sll1316* encoding cytochrome *b6-f* complex iron-sulfur subunit, *sll0450* encoding cytochrome *b* subunit of nitric oxide reductase, *smr0003* encoding cytochrome *b6-f* complex subunit PetM were also up-regulated. The results further confirmed this unique phenomenon of cyanobacteria under stress of biofuels.

RNA-seq transcriptomics analysis identified ten signal transduction proteins induced upon ethanol exposure, including two histidine kinases (*sll1473*, *slr1805*), two response regulators (*slr0947*, *sll1330*) of bacterial two-component system (TCS), one serine/threonine kinase

(*slr1225*) and three transcriptional regulators (*slr0741*, *sll0792*, *sll1423*, *ssl0707*) (Table 3). *sll1473* encoding a phytochrome-like sensor histidine kinase, was up-regulated at 48 h. Phytochromes are red/far-red photoreceptors that bear linear tetrapyrrole (bilin) chromophores attached to an N-terminal sensory module, and have been identified in many prokaryotes, including cyanobacteria [43,44]. In a study, the *cikA* gene of the cyanobacterium *Synechococcus elongatus* PCC 7942, encoding a phytochrome-related histidine kinase, was found involved in signal perception for resetting the circadian clock in response to environmental cues [45]. Although still needs more proof, the up-regulation of *sll1473* gene may be consistent with the enhanced expression of *kaiC* gene (*slr0758*) related to circadian rhythms. *slr0947* encoding a response regulator for energy transfer from phycobilisomes to photosystems was up-regulated at 72 h after ethanol exposure. Early study has found that RpaB response regulator (Slr0947) can bind to the upstream region of the high light (HL)-inducible genes in *Synechocystis* sp. PCC 6803 to cope with the potentially damaging effects of high light [46]. *slr0741* encoding transcriptional regulator was up-regulated at 72 h. The gene was previously found involved in transduction of the phosphate-limitation signal in *Synechocystis* [47]. *sll1330* encoding a two-component system response regulator OmpR subfamily was induced by ethanol at 48 h. A recent study found that expression of *sll1330* can be enhanced by nitrogen depletion under the control of NtcA, which then activates transcript accumulation of sugar catabolic genes during nitrogen starvation [48]. *slr1805* encoding a two-component sensor histidine kinase was up-regulated,

which was previous found participating in the perception and transduction of salt-stress and hyperosmotic-stress signals [49].

Considering many signal transduction genes were involved in the ethanol induced responses, we speculated that some of the ethanol-responsive genes may be under direct control of the response regulators or transcriptional regulators. To seek evidence to this hypothesis, we performed a promoter DNA-binding motif searching using 500 bp sequences extracted from upstream region of all the up-regulated genes using the Gibbs Motif Sampler software [50,51]. This analysis showed that the top conserved motifs identified were two palindromes containing 16 and 17 total sites with the DNA sequence “AXXCCTGGCCAAGGXXT” and “AAXXTTXXAAAXTT”, respectively (Figure 4) [52]. Both motif models have several conserved positions with information bits greater than 0.5 and are highly likely to be significant [50]. The genes associated with the first motif included *slr0086* encoding a DnaK protein and *slr0942* encoding an alcohol dehydrogenase [NADP<sup>+</sup>] which have been confirmed in ethanol resistance in *Clostridium* [53] and *sll1330* encoding a OmpR subfamily response regulator which was shown Sll1330 to control the expression of glycolytic genes in *Synechocystis* sp. PCC 6803 [54]. The genes associated with the second motif included *slr1109* encoding an ankyrin, *slr1828* encoding a ferredoxin, *slr1994* encoding a PHA-specific acetoacetyl-CoA reductase, *slr2033* encoding a membrane-associated rubredoxin and *slr0940* encoding a zeta-carotene desaturase, which were all involved in stress response in various microbes. Functions of these motifs may worth further investigation.



**Figure 4** Putative regulatory module identified upstream of common-responsive genes. The motif is represented by a sequence logo generated by the WebLogos software [52].

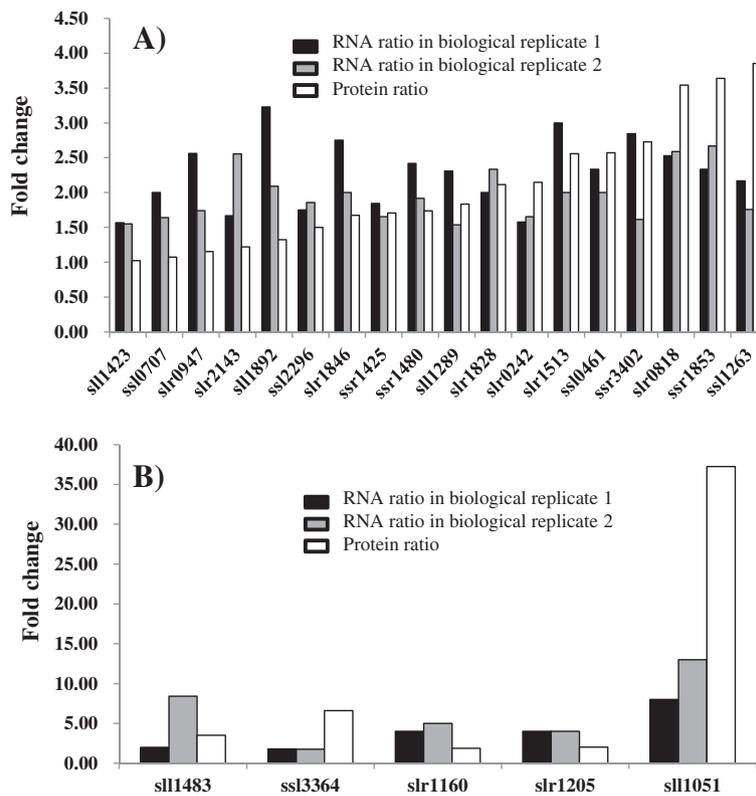
### Correlation of transcriptomic and proteomic analyses

While it is well-known that RNA expression and protein abundance are not always correlated well [55,56], we have presented evidences above that overall cellular responses identified from transcriptomics and proteomics are very similar: responses such as induction of common stress response, transporters, cell envelope proteins and photosynthesis were observed in both proteomic and transcriptomic datasets. To further compare the proteomic and transcriptomic datasets quantitatively, twenty-three common genes/proteins up-regulated in both transcriptomics and proteomics datasets were plotted together (Figure 5). The results also showed very similar trends of up-regulation, with only five genes up-regulated in transcriptomic data, but almost no change in proteomics dataset (*i.e.* *sll1423*, *ssl0707*, *slr0947*, *slr2143* and *sll1892*). However, no gene/protein with opposite regulation direction was found. In *Saccharomyces cerevisiae*, it has been proposed that there are three potential reasons for the lack of a strong correlation between transcriptomic and proteomic datasets: *i*) translational regulation, *ii*) difference in protein half-lives *in vivo* and *iii*) significant levels of experimental error, including

differences with respect to the experimental conditions being compared [57,58]. The inconsistency between transcriptomic and proteomic datasets also highlighted that it may not be enough to analyze biological systems only at a single level.

### Validation of the potential resistance targets by mutant strains

Two genes, *slr0724* and *sll1392* which were found induced by ethanol exposure at 72 h for 1.5-2.0 and 4.0-5.0 folds, respectively (Table 3, Additional file 3: Table S3), were selected for construction of knockout mutants and for validation of their involvement in ethanol resistance. *slr0724* encodes a HtaR suppressor protein homolog (*sohA*, or *prlF*) according to CYORF Cyanobacteria Gene Annotation Database (<http://cyano.genome.ad.jp/>), and *sll1392* encodes a regulatory gene, designated as *pfsR* (photosynthesis, Fe homeostasis and stress-response regulator) [59]. After confirmation by PCR and sequencing analysis, the mutants were grown in parallel with wild type *Synechocystis* sp. PCC 6803 in both normal BG11 medium and the BG11 medium supplemented with 1.5% ethanol. Comparative analysis showed that although there is no visible difference in terms of growth



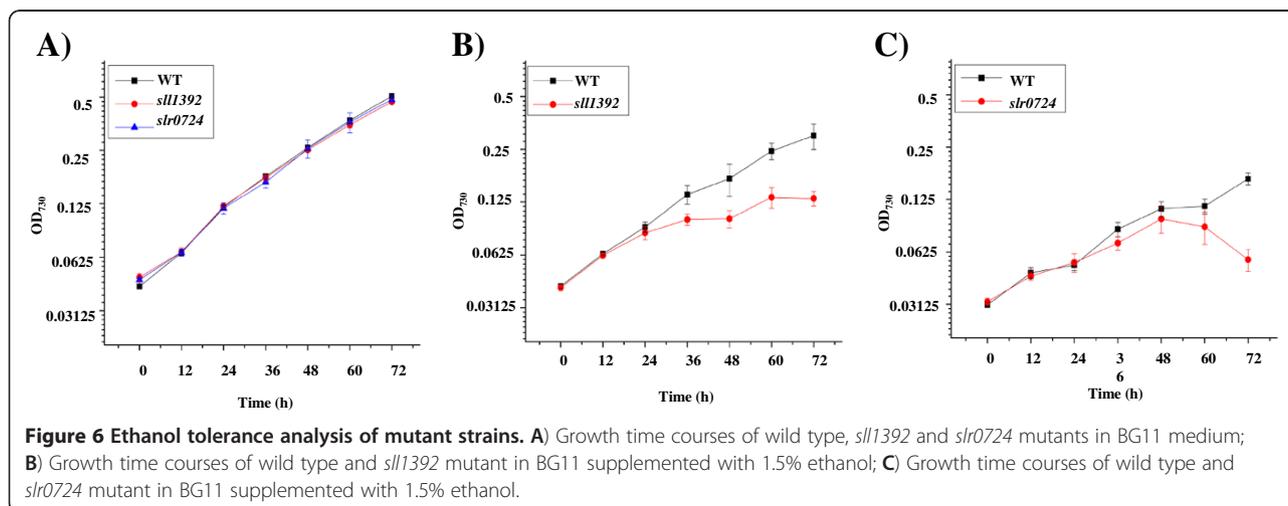
**Figure 5** Comparison of ratios derived from RNA-Seq based transcriptomics with those from proteomics for up-regulated genes/proteins by ethanol. **A)** Responsive genes/proteins with fold change smaller than 5.0; **B)** Responsive genes/proteins with fold change greater than 5.0.

patterns between the wild type and the mutants in BG11 medium (Figure 6A), the *slr0724* and *sll1392* mutants grew slower than the wild type under 1.5% ethanol (Figure 6BC), suggesting that the mutants are more sensitive to ethanol, and the gene *slr0724* and *sll1392* may be involved in ethanol resistance. In addition, the results also showed that the growth difference between the wild type and the mutants became more significant at the late growth phases (*i.e.* 60–72 h), consistent the transcriptomic results that both genes were up-regulated only at 72 h (Table 3, Additional file 3: Table S3). According to NCBI annotation (NCBI accession ID: NP\_439991.1), the *slr0724* gene could be involved in protein secretion and it induces growth defect when overproduced or mutated; however, under our growth condition, no difference in terms of growth was observed between the mutant and the wild type strains (Figure 6A). In addition, the PrIF mutation was found to induce the activity of the Lon protease. In prokaryotic cells the ATP-dependent proteases Lon are involved in the turnover of misfolded proteins and the degradation of regulatory proteins, and depending on the organism, these proteases contribute variably to stress tolerance [60,61]. Early studies have shown that *lon* mutants of *Campylobacter jejuni* grow poorly at high temperature [60] and Lon protease is involved in the control of the SOS response, acid tolerance and nutritional deprivation in *Escherichia coli* [61]. It still needs more proof whether the similar biological process was also functional in PCC6803 against ethanol. An early study has found that the *sll1392* (*pfsR*) deletion mutants were less sensitive to iron limitation under low light conditions and to suffer less lipid peroxidation following exposure to high light, suggesting a critical role of PfsR in regulation of iron homeostasis and stress response [59]. It may worth further investigation of the relationship

between ethanol stress and iron homeostasis in *Synechocystis* sp. PCC 6803.

## Conclusions

To fully elucidate microbial metabolism and its responses to ethanol, it is necessary to include functional characterization and accurate quantification of all levels of gene products, mRNA, proteins and even metabolites [54]. While high-throughput ‘omics’ approaches to analyze molecules at different cellular levels are rapidly becoming available, it is also becoming clear that any single ‘omics’ approach may not be sufficient to characterize the complexity of biological systems. To provide confirmation to previous proteomic analysis and also to reveal more responses at transcriptional level, in the study, we applied a quantitative RNA-Seq based transcriptomics approach combined with quantitative reverse-transcript PCR (RT-PCR) analysis to reveal the global transcriptomic responses to ethanol in *Synechocystis* sp. PCC 6803. The results showed that *Synechocystis* probably employed multiple and synergistic resistance mechanisms in dealing with ethanol stress. In addition, we found that the overall cellular responses inferred from transcriptomic and proteomic analyses were very similar, although the responsive genes were not always the same. By constructing knockout mutants and analyzing their ethanol tolerance, we have provided preliminary validation that the targets identified by the study could be used to obtain ethanol-tolerant cyanobacterial hosts by genetic engineering in *Synechocystis* sp. PCC 6803. Finally, our results showed that gene knockout of the potential targets individually caused only partial loss of the ethanol tolerance, consistent with the early conclusion that microbes tend to employ multiple resistance mechanisms in dealing with stress of single biofuel product [7,8]. With the ethanol-tolerance gene targets discovered from this study and previously



proteomic analysis [9], it may be possible to engineer multiple gene targets from different cellular functional categories simultaneously to achieve high-tolerance hosts in the future.

## Methods

### Bacterial growth conditions and ethanol treatment

*Synechocystis* sp. PCC 6803 was grown in BG11 medium (pH 7.5) under a light intensity of approximately 50  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  in an illuminating incubator of 130 rpm at 30°C (HNY-211B Illuminating Shaker, Honour, China). Cell density was measured on a UV-1750 spectrophotometer (Shimadzu, Japan). For growth and ethanol treatment, 10 mL fresh cells at OD<sub>730</sub> of 0.5 collected by centrifugation and then were inoculated into 50 mL BG11 liquid medium in a 250-mL flask. Ethanol of varying concentration was added at the beginning of cultivation. 1 mL of culture samples were took and measured (OD<sub>730</sub>) every 12 h. Morphology of *Synechocystis* sp. PCC6803 control and ethanol-treated samples was observed using a BX43 fluorescence microscope (Olympus, Japan). Cells for transcriptomics analysis were collected by centrifugation at 8,000  $\times g$  for 10 min at 4°C.

### RNA preparation and cDNA synthesis

Approximately 10 mg of cell pellets were frozen by liquid nitrogen immediately after centrifugation and cell walls were broken with mechanical cracking at low temperature. Cell pellets were then resuspended in Trizol reagent (Ambion, Austin, TX) and mixed well by vortex. Total RNA extraction was achieved using a miRNeasy Mini Kit (Qiagen, Valencia, CA). Contaminating DNA in RNA samples was removed with DNase I according to the instruction in the miRNeasy Mini Manual (Qiagen, Valencia, CA). The RNA quality and quantity were determined using Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) and subjected to cDNA synthesis. The RNA integrity number (RIN) of every RNA sample used for sequencing was more than 8.0. For each sample, 500 ng total RNA were subjected to cDNA synthesis using a NuGEN Ovation<sup>®</sup> Prokaryotic RNA-Seq System according to manufacturer's protocol (NuGEN, San Carlos, CA). The resulting double-stranded cDNA was purified using the MinElute Reaction Cleanup Kit (Qiagen, Valencia, CA).

### RNA-seq library preparation

The double-stranded cDNA obtained was subjected to library preparation using the Illumina TruSeq<sup>™</sup> RNA Sample Preparation Kit (Illumina, San Diego, CA), through a four-step protocol of end repairing, adding adenylate 3' ends, adapter ligation, and cDNA template enrichment. Amplification program is: 98°C 30 s; 98°C 10 s,

60°C 30 s, 72°C 30 s for 15 cycles; 72°C for 5min, and then hold at 4°C. To determine the quality of the libraries, a Qubit<sup>®</sup> 2.0 Fluorometer and Qubit<sup>™</sup> dsDNA HS (Invitrogen, Grand Island, NY) were first used to determine the DNA concentration of the libraries, and then FlashGel DNA Cassette (Lonza, USA) or Agilent Technologies 2100 Bioanalyzer (Agilent, Santa Clara, CA) was used to determine the product size of the libraries, with good libraries typically around 300 bp. The product was used directly for cluster generation using Illumina's Solexa Sequencer according to the manufacturer's instructions.

### Next-generation sequencing

RNA 2 $\times$ 100 bp paired-end sequencing was performed using Illumina's Solexa Genome Analyzer II using the standard protocol. The cDNA library of each sample was loaded to 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 Illumina standard pipeline.

### Transcriptomics data analysis

Sequence reads were pre-processed using FASTX Toolkit (Version: 0.0.13) to remove low-quality bases, and reads shorter than 20 bp. The qualified sequence reads were then mapped to non-coding RNA (*ncRNA*) sequences using Bowtie (Version: 2.0.0) with default settings. Genome sequences (including *ncRNA* sequences) and annotation information of *Synechocystis* sp. PCC 6803 were downloaded from NCBI and the Comprehensive Microbial Resource (CMR) of TIGR (<http://www.tigr.org/CMR>) (Downloaded on April 22, 2012) [10]. Reads that mapped to *ncRNA* sequences were excluded from further analysis. For paired-end Illumina reads, both pairs were removed if either pair mapped to rRNA. Remaining reads were mapped to the *Synechocystis* sp. PCC 6803 genome using Bowtie (Version: 2.0.0) with the default parameters. For gene expression determination, we performed a standard calculation of Reads Per Kilobase of Gene Per Million Mapped Reads (RPKM) based on the following formula [13]:

$$\text{RPKM} = \frac{\text{transcription\_reads}}{\text{transcription\_length} \times \text{total\_assembly\_reads\_in\_run}} \times 10^9$$

in which "transcription\_reads" stands for the number of reads mapped to a given gene; transcription\_length stands for gene length; and "total\_mapped\_reads\_in\_run" stands for the total number of reads in a given measurement. For each time point, two biological replicates of

ethanol-treated samples and their control were analyzed and the corresponding gene expression ratios based on RPKM were calculated, the genes with 1.5 fold changes in both biological replicates were determined as differentially regulated genes.

#### Quantitative real-time RT-PCR analysis

The RNA samples were collected from cells grown under the same growth condition as described above for transcriptomic analysis. Approximately 10 mg of cell pellets were frozen by liquid nitrogen immediately after centrifugation and cell walls were broken with mechanical cracking at low temperature. Cell pellets were then resuspended in Trizol reagent (Ambion, Austin, TX) and mixed well by vortex. Total RNA extraction was achieved using a miRNeasy Mini Kit (Qiagen, Valencia, CA). First-strand cDNAs were synthesized using RevertAid™ Reverse Transcriptase (Fermentas, Glen Burnie, MD). cDNA was subjected to eight hundred fold dilutions, and 2 µl of each dilution was used as template for following qPCR reaction. The qPCR reaction was carried out in 20 µl reactions containing 10 µl of SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA), and 2 µl of each PCR primer at 2 mM, employing the StepOnePlus™ Real-Time PCR System (Applied Biosystems, Foster City, CA), under the following condition: 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Quantification of gene expression was determined according to standard process of RT-PCR which used serial dilutions of known concentration of chromosome DNA as template to make a standard curve. A total of 18 selected genes based on their differential expression patterns revealed by iTRAQ were selected for verification and the *rnpB* gene (*6803s01*) encoding RNase P subunit B was used as an internal control according to the previous publication [62]. Three technical replicates were performed for each gene. Data analysis was carried out using the StepOnePlus analytical software (Applied Biosystems, Foster City, CA). Briefly, the amount of relative gene transcript was normalized by that of *rnpB* in each sample (wild type or mutant), using the following method:

$$R_{\text{relative gene expression of gene } x} = \frac{2^{(\text{Ct}_{\text{control}} - \text{Ct}_{\text{treated}}) \text{ of } x}}{2^{(\text{Ct}_{\text{control}} - \text{Ct}_{\text{treated}}) \text{ of } rnpB}}$$

Then data was presented as ratios of the amount of normalized transcript in the treatment to that from the control. The gene ID and their related primer sequences used for real-time RT-PCR analysis were listed in Additional file 4: Table S4.

#### Promoter analysis and motif identification

The Gibbs Motif Sampler software from the Biometrics Laboratory of Wadsworth Center, (<http://www.bayesweb.wadsworth.org/gibbs/gibbs.html>), was used to identify matrix models describing DNA sequence motifs present upstream of genes responsive to ethanol treatment [50,51]. Regions representing approximately 500 base pairs of the DNA sequences upstream of the translational start site of genes responsive to ethanol stress were extracted from the NCBI genome database using the Regulatory Sequence Analysis Tools (RSAT) [63]. Both strands of each sequence were searched and possible motif locations were identified using the motif matrix score obtained from the Gibbs Motif Sampler software. The multilevel consensus sequence for each motif was then used to generate a sequence logo that is a graphical representation of nucleic acid multiple sequence alignment (<http://www.weblogo.berkeley.edu/>) [52].

#### Construction and analysis of gene knockout mutants

A fusion PCR based method was employed for the construction of gene knockout fragments [64]. Briefly, for the gene target selected, three sets of primers were designed to amplify a linear DNA fragment containing the chloramphenicol resistance cassette (amplified from a plasmid pACYC184) with two flanking arms of DNA upstream and downstream of the targeted gene. The linear fused PCR amplicon was used directly for transformation into *Synechocystis* sp. PCC 6803 by natural transformation. The chloramphenicol-resistant transformants were obtained and passed several times on fresh BG11 plates supplemented with 10 µg/ml chloramphenicol to achieve complete chromosome segregation (confirmed by PCR). Two genes, *slr0724* and *sll1392* that have been found differentially regulated by ethanol exposure, were selected for construction of gene knockout mutants. The mutants were confirmed by PCR and sequencing analysis. PCR primers for mutant construction and validation were listed in Additional file 4: Table S4. Comparative growth analysis of the wild type 6803 and the mutants were performed in 100-mL flasks each with 10 mL BG11 medium with or without 1.5% ethanol. Cultivation conditions are the same as described above. Growth analysis was performed in triplicates.

#### Additional files

- Additional file 1: Table S1.** Raw RNA-seq transcriptomics data.
- Additional file 2: Table S2.** Gene down-regulated by ethanol exposure.
- Additional file 3: Table S3.** Induced genes encoding hypothetical proteins.
- Additional file 4: Table S4.** Primers used in this study.

## Abbreviations

*Adh*: Alcohol dehydrogenase; AGES: Advanced glycation end products; CMR: Comprehensive Microbial Resource; *desA*: Acyl-lipid desaturase; HL: High light; iTRAQ: Isobaric tag for relative and absolute quantitation; LC-MS/MS: Liquid chromatography-tandem mass spectrometry; MG: Methylglyoxal; ncRNA: non-coding RNA; PHA: Polyhydroxyalkanoate; *pdC*: Pyruvate decarboxylase; PSII: Photosystem II; RPKM: Reads per kilobase of gene per million mapped reads; ROS: Reactive oxygen species; RSAT: Regulatory sequence analysis tools; RT-PCR: Reverse-transcript PCR; TCS: Two-component system.

## Competing interest

The authors declare that they have no competing interests.

## Authors' contributions

JW, LC and WZ conceived of the study. JW, LC, JQ and WZ drafted the manuscript. JW, LC, and XT carried out cultivation and transcriptomics analysis. LC, JL, XR and SH carried out the RT-PCR, mutant construction and phenotypic analysis. SH finished the promoter analysis. JW, LC and WZ finish the statistical analysis for transcriptomic data. All authors read and approved the final manuscript.

## Acknowledgements

The research was supported by grants from National Basic Research Program of China ("973" program, project No. 2011CBA00803 and No. 2012CB721101) and National High-tech R&D Program ("863" program, project No. SS2012AA023107). The authors would also like to thank Tianjin University and the "985 Project" of Ministry of Education for their generous supports in establishing the research laboratory.

Received: 21 September 2012 Accepted: 4 December 2012

Published: 21 December 2012

## References

- Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM: **Ethanol can contribute to energy and environmental goals.** *Science* 2006, **311**:506–508.
- Dien BS, Cotta MA, Jeffries TW: **Bacteria engineered for fuel ethanol production: current status.** *Appl Microbiol Biotechnol* 2003, **63**(3):258–266.
- Antoni D, Zverlov VV, Schwarz WH: **Biofuels from microbes.** *Appl Microbiol Biotechnol* 2007, **77**:23–35.
- Machado IM, Atsumi S: **Cyanobacterial biofuel production.** *J Biotechnol* 2012;. [Epub ahead of print].
- Deng M, Coleman J: **Ethanol synthesis by genetic engineering in cyanobacteria.** *Appl Environ Microbiol* 1999, **65**:523–528.
- Fu P: **Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion.** *J Chem Technol Biotechnol* 2009, **84**:473–483.
- Nicolaou SA, Gaida SM, Papoutsakis ETA: **Comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation.** *Metab Eng* 2010, **12**:307–331.
- Dunlop MJ: **Engineering microbes for tolerance to next-generation biofuels.** *Biotechnol Biofuels* 2011, **4**:32.
- Qiao JJ, Wang JX, Chen L, Tian XX, Zhang WW: **Quantitative iTRAQ LC-MS/MS proteomics reveals metabolic response to biofuel ethanol in cyanobacterial *Synechocystis* sp. PCC 6803.** *J Proteome Res* 2012, **11**(11):5286–5300.
- Kaneko T, Nakamura Y, Sasamoto S, Watanabe A, Kohara M, Matsumoto M, Shimpo S, Yamada M, Tabata S: **Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803.** *DNA Res* 2003, **10**:221–228.
- Croucher NJ, Thomson NR: **Studying bacterial transcriptomes using RNA-seq.** *Curr Opin Microbiol* 2010, **13**(5):619–624.
- Shai RM: **Microarray tools for deciphering complex diseases.** *Front Biosci* 2006, **11**:1414–1424.
- Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**:R106.
- Mrázek J, Bhaya D, Grossman AR, Karlin S: **Highly expressed and alien genes of the *Synechocystis* genome.** *Nucleic Acids Res* 2001, **29**(7):1590–1601.
- Das S, Roymondal U, Chottopadhyay B, Sahoo S: **Gene expression profile of the cyanobacterium *Synechocystis* genome.** *Gene* 2012, **497**(2):344–352.
- Wegener KM, Welsh EA, Thornton LE, Keren N, Jacobs JM, Hixson KK, Monroe ME, Camp DG 2nd, Smith RD, Pakrasi HB: **High sensitivity proteomics assisted discovery of a novel operon involved in the assembly of photosystem II, a membrane protein complex.** *J Biol Chem* 2008, **283**(41):27829–27837.
- Singh AK, Sherman LA: **Characterization of a stress-responsive operon in the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *Gene* 2002, **297**(1–2):11–19.
- Furumichi M, Sato Y, Omata T, Ikeuchi M, Kanehisa M: **CYORF: community annotation of cyanobacteria genes.** *Genome Inform* 2002, **13**:402–403.
- Flint A, Sun YQ, Stintzi A: **CJ1386 is an ankyrin-containing protein involved in heme trafficking to catalase in *Campylobacter jejuni*.** *J Bacteriol* 2012, **194**(2):334–345.
- Xia JL, Wu S, Zhang RY, Zhang CG, He H, Jiang HC, Nie ZY, Qiu GZ: **Effects of copper exposure on expression of glutathione-related genes in *Acidithiobacillus ferrooxidans*.** *Curr Microbiol* 2011, **62**(5):1460–1466.
- Das PK, Bagchi SN: **Role of bacterioferritin comigratory protein and glutathione peroxidase-reductase system in promoting bentazone tolerance in a mutant of *Synechococcus elongatus* PCC7942.** *Protoplasma* 2012, **249**(1):65–74.
- Li H, Jubelirer S, Garcia Costas AM, Frigaard NU, Bryant DA: **Multiple antioxidant proteins protect *Chlorobaculum tepidum* against oxygen and reactive oxygen species.** *Arch Microbiol* 2009, **191**(11):853–867.
- Dvornyk V, Vinogradova O, Nevo E: **Long-term microclimatic stress causes rapid adaptive radiation of *kaiABC* clock gene family in a cyanobacterium, *Nostoc linckia*, from "Evolution Canyons" I and II, Israel.** *Proc Natl Acad Sci USA* 2002, **99**(4):2082–2087.
- Ding J, Huang X, Zhang L, Zhao N, Yang D, Zhang K: **Tolerance and stress response to ethanol in the yeast *Saccharomyces cerevisiae*.** *Appl Microbiol Biot* 2009, **85**(2):253–263.
- Stanley D, Bandara A, Fraser S, Chambers PJ, Stanley GA: **The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*.** *J Appl Microbiol* 2010, **109**(1):13–24.
- Thiel T: **Phosphate transport and arsenate resistance in the cyanobacterium *Anabaena variabilis*.** *J Bacteriol* 1988, **70**:1143–1147.
- Verma SK, Singh HN: **Evidence for energy-dependent copper efflux as a mechanism of Cu<sup>2+</sup> resistance in the cyanobacterium *Nostoc calcicola*.** *FEMS Microbiol Lett* 1991, **68**:291–294.
- Nomura M, Ishitani M, Takabe T, Rai AK, Takabe T: ***Synechococcus* sp. PCC7942 transformed with *Escherichia coli* bet genes produces glycine betaine from choline and acquires resistance to salt stress.** *Plant Physiol* 1995, **107**:703–708.
- Mikkat S, Hagemann M, Schoor A: **Active transport of glucosylglycerol is involved in salt adaptation of the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *Microbiol* 1996, **142**:1725–1732.
- Tong L, Nakashima S, Shibasaki M, Katsuhara M, Kasamo K: **A novel histidine-rich CPx-ATPase from the filamentous cyanobacterium *Oscillatoria brevis* related to multiple-heavy-metal cotolerance.** *J Bacteriol* 2002, **184**:5027–5035.
- Hermans MA, Neuss B, Sahn H: **Content and composition of hopanoids in *Zymomonas mobilis* under various growth conditions.** *J Bacteriol* 1991, **173**:5592–5595.
- Amiri RM, Yur'eva NO, Shimshilashvili KR, Goldenkova-Pavlova IV, Pchelkin VP, Kuznitsova EI, Tsydendambaev VD, Trunova TI, Los DA, Jouzani GS, Nosov AM: **J Expression of acyl-lipid Delta12-desaturase gene in prokaryotic and eukaryotic cells and its effect on cold stress tolerance of potato.** *J Integr Plant Biol* 2010, **52**(3):289–297.
- Buddelmeijer N, Young R: **The essential *Escherichia coli* apolipoprotein N-acyltransferase (Lnt) exists as an extracytoplasmic thioester acyl-enzyme intermediate.** *Biochemistry-US* 2010, **49**(2):341–346.
- Schmalzer M, Jann NJ, Götz F, Landmann R: **Staphylococcal lipoproteins and their role in bacterial survival in mice.** *Int J Med Microbiol* 2010, **300**(2–3):155–160.
- Nogales J, Campos R, BenAbdelkhalek H, Olivares J, Lluch C, Sanjuan J: ***Rhizobium tropici* genes involved in free-living salt tolerance are required for the establishment of efficient nitrogen-fixing symbiosis with *Phaseolus vulgaris*.** *Mol Plant Microbe In* 2002, **15**(3):225–232.

36. Zhu J, Jäger K, Black T, Zarka K, Koksharova O, Wolk CP: **HcWA, an autolysin, is required for heterocyst maturation in *Anabaena* sp. strain PCC 7120.** *J Bacteriol* 2001, **183**(23):6841–6851.
37. Ayub ND, Tribelli PM, López NI: **Polyhydroxyalkanoates are essential for maintenance of redox state in the Antarctic bacterium *Pseudomonas* sp. 14–3 during low temperature adaptation.** *Extremophiles* 2009, **13**(1):59–66.
38. Kadouri D, Jurkevitch E, Okon Y, Castro-Sowinski S: **Ecological and agricultural significance of bacterial polyhydroxyalkanoates.** *Crit Rev Microbiol* 2005, **31**:55–67.
39. Taroncher-Oldenburg G, Nishina K, Stephanopoulos G: **Identification and analysis of the polyhydroxyalkanoate-specific beta-ketothiolase and acetoacetyl coenzyme A reductase genes in the cyanobacterium *Synechocystis* sp. strain PCC6803.** *Appl Environ Microbiol* 2000, **66**(10):4440–4448.
40. Pepper ED, Farrell MJ, Nord G, Finkel SE: **Antiglycation effects of carnosine and other compounds on the long-term survival of *Escherichia coli*.** *Appl Environ Microbiol* 2010, **76**(24):7925–7930.
41. Ferguson GP: **Protective mechanisms against toxic electrophiles in *Escherichia coli*.** *Trends Microbiol* 1999, **7**(6):242–247.
42. Mustafiz A, Sahoo KK, Singla-Pareek SL, Sopory SK: **Metabolic engineering of glyoxalase pathway for enhancing stress tolerance in plants.** *Methods Mol Biol* 2010, **639**:95–118.
43. Mutsuda M, Michel KP, Zhang X, Montgomery BL, Golden SS: **Biochemical properties of CikA, an unusual phytochrome-like histidine protein kinase that resets the circadian clock in *Synechococcus elongatus* PCC 7942.** *J Biol Chem* 2003, **278**(21):19102–19110.
44. Hughes J, Lamparter T, Mittmann F, Hartmann E, Gärtner W, Wilde A, Börner T: **A prokaryotic phytochrome.** *Nature* 1997, **386**(6626):663.
45. Schmitz O, Katayama M, Williams SB, Kondo T, Golden SS: **CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock.** *Science* 2000, **289**(5480):765–768.
46. Kappell AD, van Waasbergen LG: **The response regulator RpaB binds the high light regulatory 1 sequence upstream of the high-light-inducible *hliB* gene from the cyanobacterium *Synechocystis* PCC 6803.** *Arch Microbiol* 2007, **187**(4):337–342. Epub 2007 Feb 10.
47. Juntarajumnong W, Hirani TA, Simpson JM, Incharoensakdi A, Eaton-Rye JJ: **Phosphate sensing in *Synechocystis* sp. PCC 6803: SphU and the SphS-SphR two-component regulatory system.** *Arch Microbiol* 2007, **188**(4):389–402.
48. Azuma M, Osanai T, Hirai MY, Tanaka K: **A response regulator Rre37 and an RNA polymerase sigma factor SigE represent two parallel pathways to activate sugar catabolism in a cyanobacterium *Synechocystis* sp. PCC 6803.** *Plant Cell Physiol* 2011, **52**(2):404–412.
49. Los DA, Zorina A, Sinetova M, Kryazhov S, Mironov K, Zinchenko W: **Stress Sensors and Signal Transducers in Cyanobacteria.** *Sensors-Basel* 2010, **10**(3):2386–2415.
50. Lawrence CE, Altschul SF, Bogouski MS, Liu JS, Neuwald AF, Wooten JC: **Detecting subtle sequence signals: a Gibbs Sampling Strategy for multiple alignment.** *Science* 1993, **262**:208–214.
51. Thompson W, Rouchka EC, Lawrence CE: **Gibbs recursive sampler: finding transcription factor binding sites.** *Nucleic Acids Res* 2003, **31**:3580–3585.
52. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**:1188–1190.
53. Brown SD, Guss AM, Karpinetz TV, Parks JM, Smolin N, Yang S, Land ML, Klingeman DM, Bhandiwad A, Rodriguez M Jr, Raman B, Shao X, Mielenz JR, Smith JC, Keller M, Lynd LR: **Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*.** *Proc Natl Acad Sci USA* 2011, **108**(33):13752–13757.
54. Tabei Y, Okada K, Tsuzuki M: **Slh1330 controls the expression of glycolytic genes in *Synechocystis* sp. PCC 6803.** *Biochem Biophys Res Co* 2007, **355**(4):1045–1050.
55. Nie L, Wu G, Culley DE, Scholten JC, Zhang W: **Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications.** *Crit Rev Biotechnol* 2007, **27**(2):63–75.
56. Zhang W, Li F, Nie L: **Integrating multiple 'omics' analysis for microbial biology: application and methodologies.** *Microbiology* 2010, **156**(Pt 2):287–301.
57. Greenbaum D, Colangelo C, Williams K, Gerstein M: **Comparing protein abundance and mRNA expression levels on a genomic scale.** *Genome Biol* 2003, **4**:117.1–117.8.
58. Nie L, Wu G, Brockman FJ, Zhang W: **Integrated analysis of transcriptomic and proteomic data of *Desulfovibrio vulgaris*: zero-inflated Poisson regression models to predict abundance of undetected proteins.** *Bioinformatics* 2006, **22**(13):1641–1647.
59. Jantaro S, Ali Q, Lone S, He Q: **Suppression of the lethality of high light to a quadruple HLI mutant by the inactivation of the regulatory protein PfsR in *Synechocystis* PCC 6803.** *J Biol Chem* 2006, **281**(41):30865–30874.
60. Cohn MT, Ingmer H, Mulholland F, Jørgensen K, Wells JM, Brøndsted L: **Contribution of conserved ATP-dependent proteases of *Campylobacter jejuni* to stress tolerance and virulence.** *Appl Environ Microbiol* 2007, **73**(24):7803–7813.
61. Langklotz S, Narberhaus F: **The *Escherichia coli* replication inhibitor CspD is subject to growth-regulated degradation by the Lon protease.** *Mol Microbiol* 2011, **80**(5):1313–1325.
62. Kloft N, Rasch G, Forchhammer K: **Protein phosphatase PphA from *Synechocystis* sp. PCC 6803: the physiological framework of PII-P dephosphorylation.** *Microbiology* 2005, **151**(Pt 4):1275–1283.
63. van Helden J: **Regulatory sequence analysis tools.** *Nucleic Acids Res* 2003, **31**:3593–3596.
64. Wang HL, Postier BL, Burnap RL: **Optimization of fusion PCR for *in vitro* construction of gene knockout fragments.** *Biotechniques* 2002, **33**(1):26, 28, 30.

doi:10.1186/1754-6834-5-89

Cite this article as: Wang et al.: RNA-seq based identification and mutant validation of gene targets related to ethanol resistance in cyanobacterial *Synechocystis* sp. PCC 6803. *Biotechnology for Biofuels* 2012 5:89.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

