How cyanobacteria pose new problems to old methods: Challenges in microarray time series analysis

Robert Lehmann[*1] , Rainer Machné[3] , Jens Georg[4] , Manuela Benary[1] , Ilka Axmann[1,2] , Ralf Steuer[1]

[1]Institute for Theoretical Biology, Humboldt University Berlin, Invalidenstraße 43, D-10115, Berlin, Germany
[2]Institute for Theoretical Biology, Charité Universitätsmedizin, Invalidenstraße 43, D-10115, Berlin, Germany
[3]Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090, Vienna, Austria
[4]Faculty of Biology and Freiburg Initiative in Systems Biology, University of Freiburg, Schänzlestraße 1, D-79104, Freiburg, Germany

Email: Robert Lehmann[*]- r.lehmann@biologie.hu-berlin.de; Rainer Machne - raim@tbi.univie.ac.at; Jens Georg - jens.georg@biologie.uni-freiburg.de; Manuela Benary - manuela.benary@cms.hu-berlin.de; Ilka M. Axmann - ilka.axmann@charite.de; Ralf Steuer - ralf.steuer@hu-berlin.de;

[*]Corresponding author
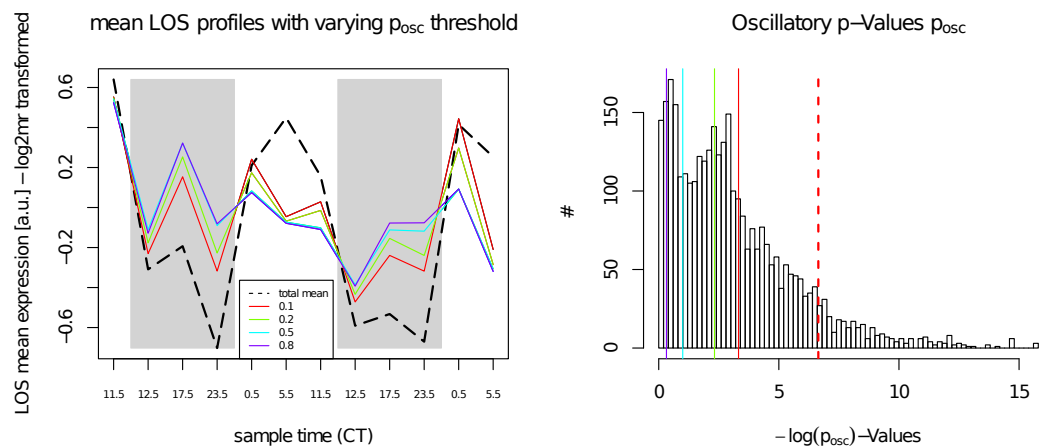
Figure 1: Global expression oscillation in unnormalized data (A, black dashed line). Mean profiles of gene sets used for LOS normalization using different cutoffs for p_{osc}. (B) The distribution of p_{osc} in the complete dataset with cutoffs indicated with identical colors to (A). The red dashed line marks the significance level of 0.001.

FlowClust BIC curves for data processings
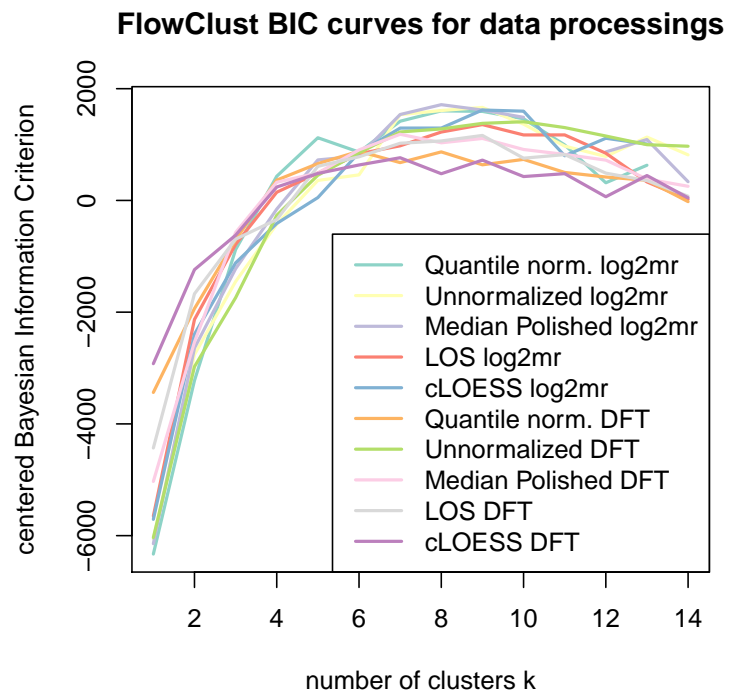
Figure 2: Number of Clusters in the Dataset. Bayesian information criterion for clusterings of data with different normalization and transformation as provided by FlowClust.
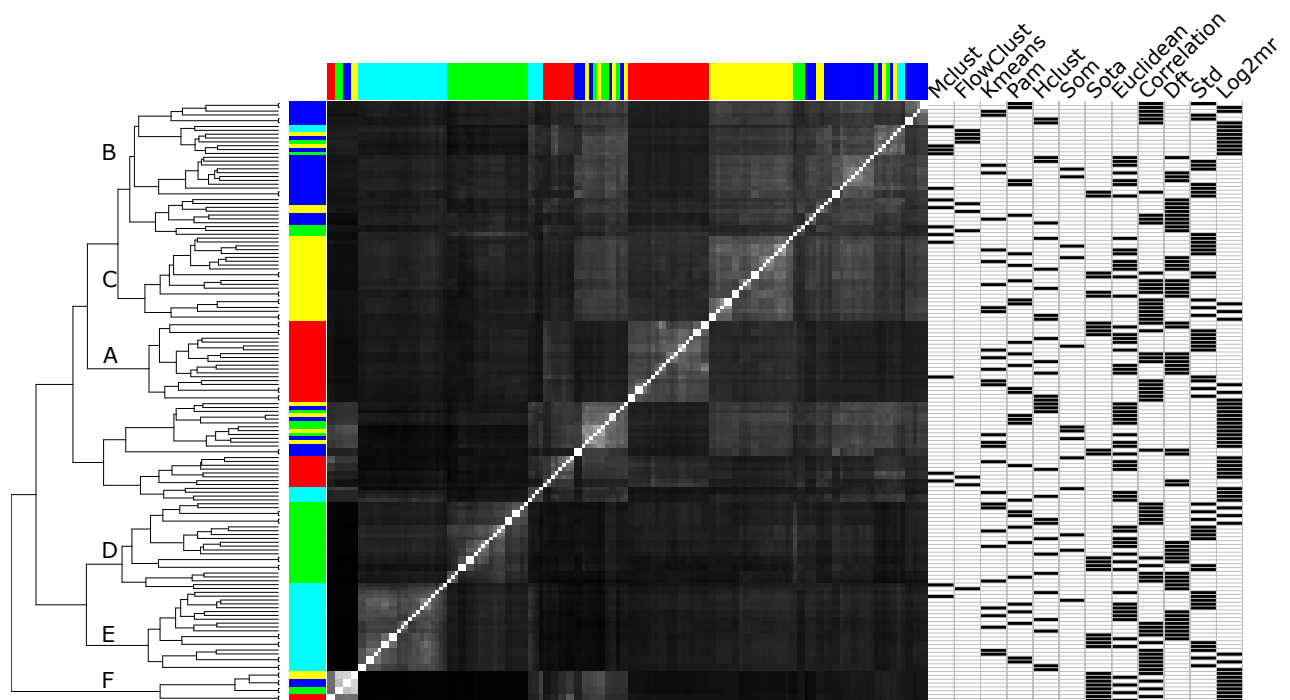
Figure 3: Comparing Clustering Results with the normalized Variation of Information between all clustering combinations with k=8. The visualization is identical to Figure 3.
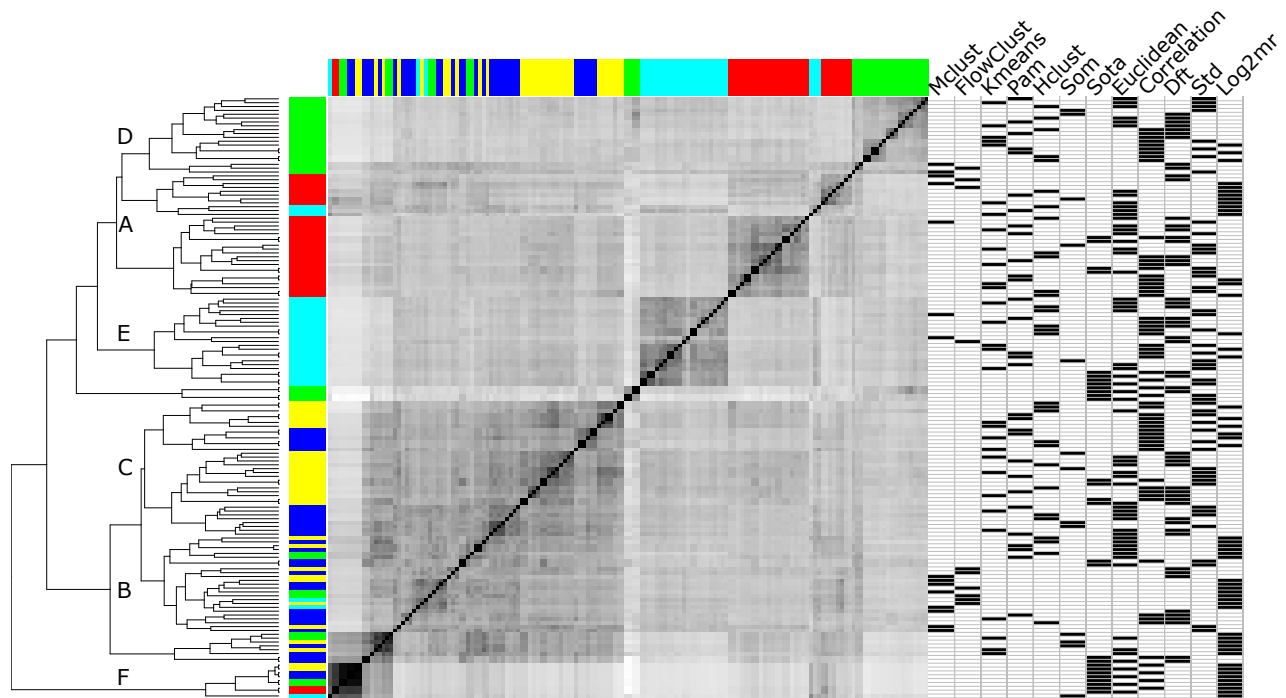
Figure 4: Comparing Clustering Results with the adjusted Rand index between all clustering combinations with k=8. The visualization is identical to Figure 3. The dendrogram of clusterings varies in comparison to the similarity matrices using mutual information (Fig.3) and normalized Variation of Information (Fig.S3). Subgroup A contains only a small subbranch of three cLOESS normalized clusterings. The remaining clusterings of this subbranch are moved into subgroup B, which is consequently more diverse.

A

Expression [a.u.]

1, 420
2, 288
3, 624
4, 286
5, 84
6, 66
7, 728
8, 262
9, 401
10, 188

11.5_r1 12.5_r1 17.5_r1 23.5_r1 0.5_r1 5.5_r1 11.5_r2 12.5_r2 17.5_r2 23.5_r2 0.5_r2 5.5_r2   CT

B

	1	2	3	4	5	6	7	8	9	10
Ribosomal proteins: synthesis and modif.	1 / 1	51 / 1.9e−64	2 / 1	0 / 1	0 / 1	3 / 0.051	5 / 1	0 / 1	6 / 1	1 / 0.96
Translation	16 / 0.88	65 / 3.7e−41	33 / 0.27	16 / 0.23	0 / 1	4 / 0.29	20 / 1	1 / 1	1 / 1	5 / 0.95
ATP synthase	0 / 1	10 / 3.5e−16	0 / 1	0 / 1	1 / 1	0 / 1	2 / 1	1 / 1	1 / 1	1 / 3
RNA synthesis, modification, and DNA tr.	2 / 0.73	9 / 8.8e−07	2 / 0.92	1 / 0.82	1 / 0.31	0 / 1	2 / 0.96	2 / 0.42	1 / 0.93	3 / 0.056
Transcription	1 / 0.87	10 / 2.2e−06	5 / 0.58	1 / 0.9	1 / 0.41	0 / 1	5 / 0.74	2 / 0.59	1 / 0.97	2 / 0.12
Protein modification and translation fa.	1 / 0.97	8 / 0.00016	10 / 0.0068	5 / 0.034	0 / 1	0 / 1	2 / 0.99	0 / 1	1 / 0.97	2 / 0.36
Photosynthesis and respiration	1 / 1	22 / 0.00037	18 / 0.97	12 / 0.44	11 / 2.8e−05	46 / 3.7e−72	6 / 1	20 / 0.00069	3 / 1	2 / 1
Transposon–related functions	5 / 0.72	0 / 1	19 / 0.39	32 / 1.8e−16	0 / 1	0 / 1	17 / 0.88	0 / 1	13 / 1	15 / 1.3e−05
Other categories	34 / 1	9 / 1	58 / 0.31	51 / 1.5e−09	2 / 1	1 / 1	76 / 0.03	19 / 0.84	32 / 0.75	23 / 0.027
Photosystem II	1 / 1	0.89 / 1	1 / 1	0.6 / 1	4 / 0.00045	17 / 1.9e−33	1 / 1	1 / 0.86	1 / 0.96	1 / 0.73
Phycobilisome	1 / 0.87	1 / 1	1 / 0.83	1 / 0.33	0 / 1	13 / 7.4e−28	0 / 1	1 / 1	0 / 1	0 / 1
Photosystem I	0 / 1	1 / 0.68	3 / 0.47	1 / 0.67	3 / 0.00074	8 / 7.9e−15	0 / 1	0 / 1	0 / 1	0 / 1
Cytochrome b6/f complex	1 / 1	2 / 0.083	0 / 1	0 / 1	0 / 1	5 / 4e−10	2 / 1	1 / 0.39	1 / 0.58	7 / 1
Other	23 / 0.11	6 / 0.99	33 / 0.11	13 / 0.45	0 / 1	0 / 1	49 / 0.00015	11 / 0.57	13 / 0.92	7 / 0.7
NADH dehydrogenase	0 / 1	1 / 0.17	1 / 0.73	2 / 0.45	3 / 0.3	1 / 1	0.84 / 1	10 / 6.1e−09	0 / 1	0 / 1
WD repeat proteins	1 / 0.36	0 / 1	0 / 1	1 / 1	1 / 4	0 / 1	0 / 1	1 / 0.00037	0.34 / 1	1 / 6
Transport and binding proteins	28 / 0.14	19 / 0.2	36 / 0.45	4 / 1	0.64 / 1	1 / 1	30 / 0.99	12 / 0.8	57 / 3.8e−16	6 / 0.96
DNA replication, restriction, modificat.	7 / 0.55	2 / 0.96	6 / 0.97	3 / 0.87	0 / 1	1 / 0.61	22 / 0.00078	1 / 1	17 / 2e−05	3 / 0.57

Figure 5: A) Unnormalized data clustered using flowClust with DFT transformation and ten clusters. The clusters are sorted by phase and profiles plotted using log_2 mean ratio transformed data. In each cluster plot, the cluster index followed by the number of genes in the corresponding cluster are shown in the top left corner. Subjective night is marked by a gray background. The sample times are provided in circadian time (CT) for cluster ten and are identical for all clusters. B) Functional enrichment result for each cluster in panel A. On top of the matrix the cluster indices are shown together with a color mark that corresponds to the color in A. On the left of the matrix the enriched functional categories are shown, sorted by cluster. Only significantly enriched categories are shown ($p < .05$). The p-value is shown on the bottom of every cell, the number of genes which are in the corresponding cluster and associated with the corresponding function is shown on top of each cell. Additionally, the enrichment p-value is color-coded in the background of each cell with black marking highest significance and white as lowest.
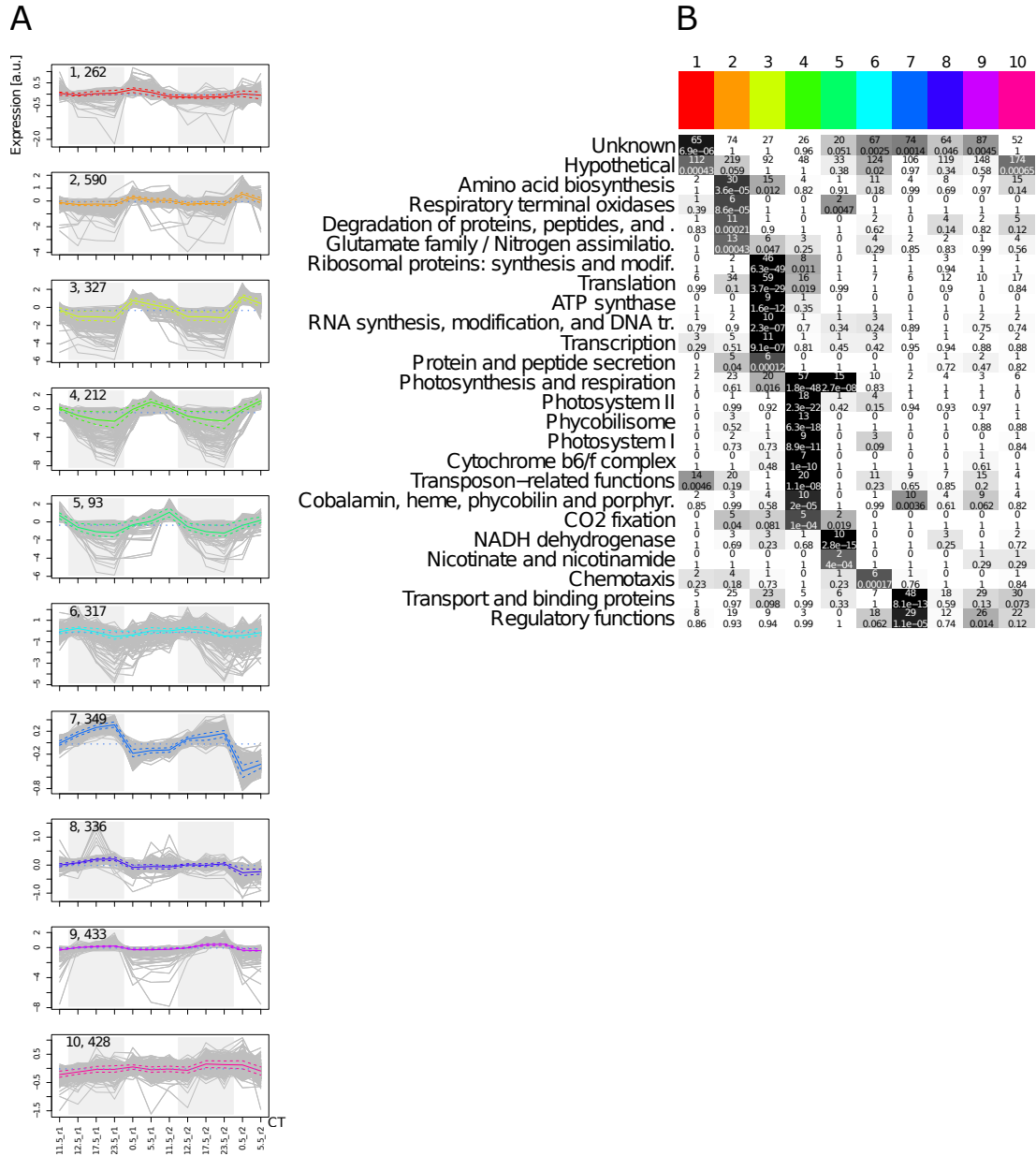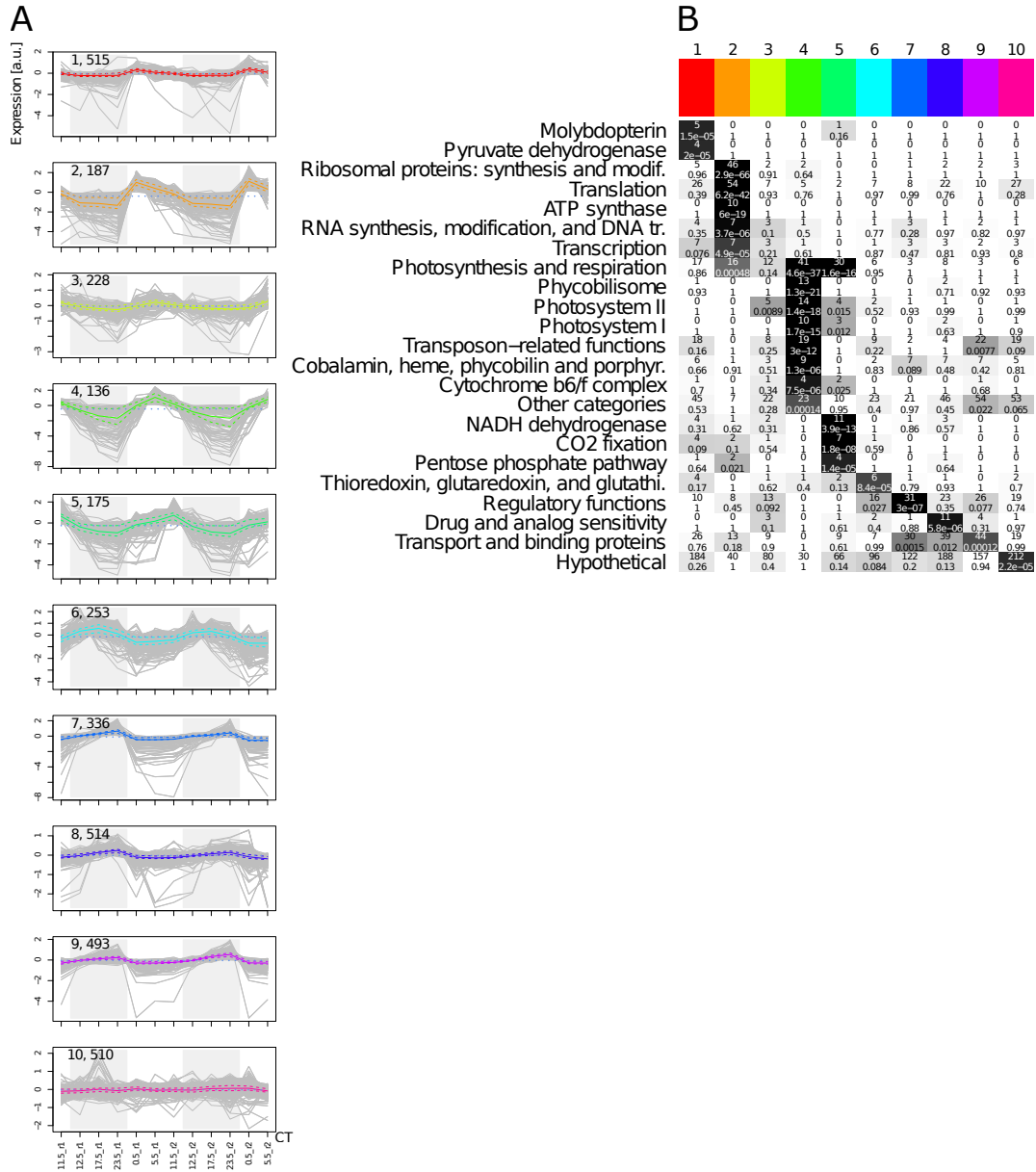
5

A

B

Figure 6: Clustering and functional enrichment for median polished data. The visualization is identical to Figure S5.

Figure 7: Clustering and functional enrichment for quantile normalized data. The visualization is identical to Figure S5.

A

1, 467

2, 699

3, 270

4, 399

5, 200

6, 296

7, 232

8, 251

9, 389

10, 144

Expression [a.u.]

CT

B

	1	2	3	4	5	6	7	8	9	10

Hypothetical
Aminoacyl tRNA synthetases and tRNA mod.
Purine ribonucleotide biosynthesis
Purines, pyrimidines, nucleosides, and .
Ribosomal proteins: synthesis and modif.
ATP synthase
Translation
Molybdopterin
Transport and binding proteins
Glycolate pathway
Transposon–related functions
Other categories
Unknown
Photosynthesis and respiration
Photosystem II
Cobalamin, heme, phycobilin and porphyr.
NADH dehydrogenase
CO2 fixation
Pentose phosphate pathway
Respiratory terminal oxidases
Photosystem I
Chaperones
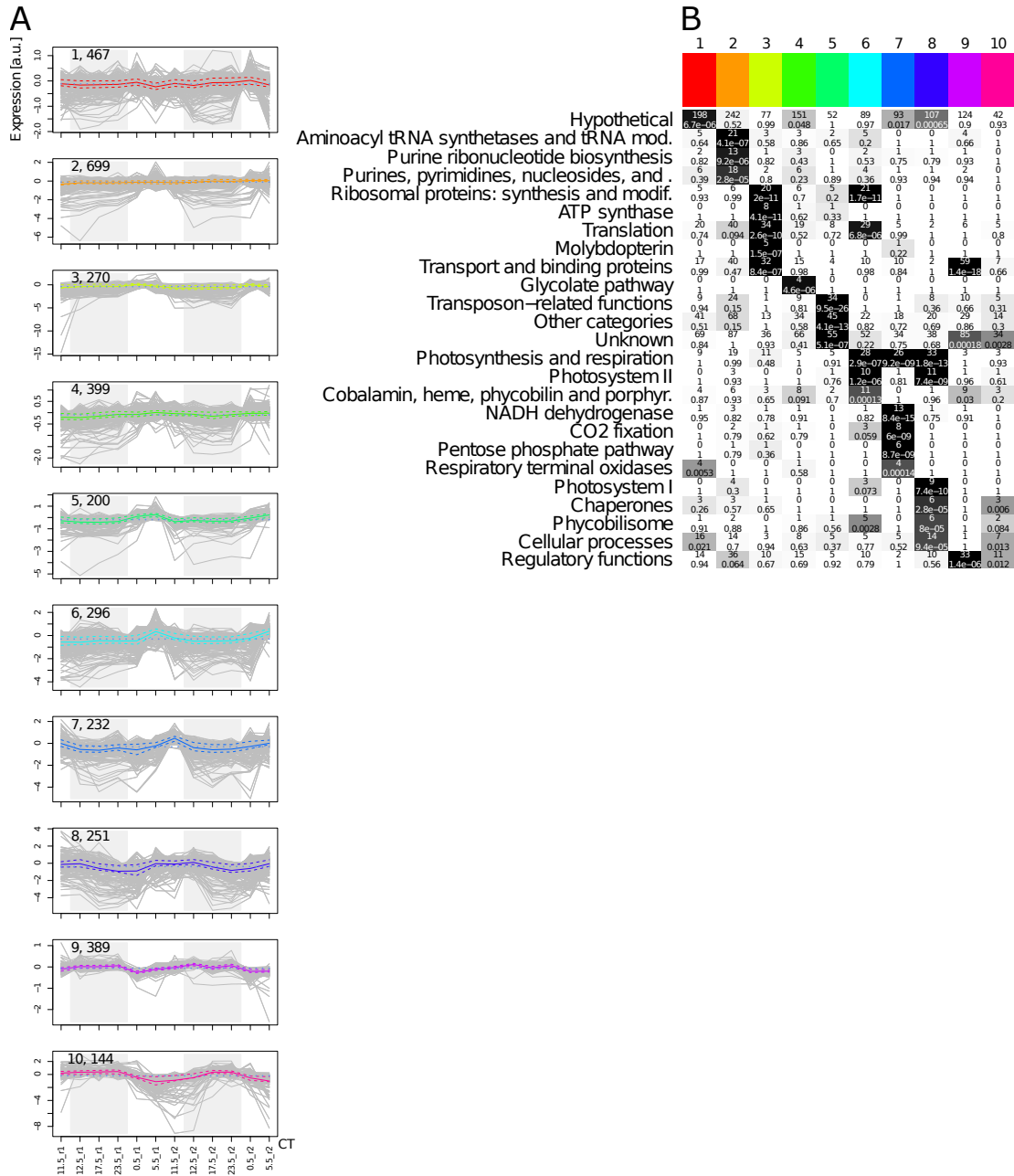Phycobilisome
Cellular processes
Regulatory functions

Figure 8: Clustering and functional enrichment for cyclic LOESS normalized data. The visualization is identical to Figure S5.

8