

Comprehensive Analysis of Gene-Environmental Interactions with Temporal Gene Expression Profiles in *Pseudomonas aeruginosa*

Kangmin Duan¹, William M. McCullough², Michael G. Surette², Tony Ware³, Jiuzhou Song^{4*}

¹ Department of Oral Biology; Department of Medical Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada, ² Department of Medicine, Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, Ontario, Canada, ³ Department of Mathematics and Statistics, University of Calgary, Calgary, Canada, ⁴ Department of Animal and Avian Science, Animal Science Center, University of Maryland, College Park, Maryland, United States of America

Abstract

To explore gene-environment interactions, based on temporal gene expression information, we analyzed gene and treatment information intensively and inferred interaction networks accordingly. The main idea is that gene expression reflects the response of genes to environmental factors, assuming that variations of gene expression occur under different conditions. Then we classified experimental conditions into several subgroups based on the similarity of temporal gene expression profiles. This procedure is useful because it allows us to combine diverse gene expression data as they become available, and, especially, allowing us to lay the regulatory relationships on a concrete biological basis. By estimating the activation points, we can visualize the gene behavior, and obtain a consensus gene activation order, and hence describe conditional regulatory relationships. The estimation of activation points and building of synthetic genetic networks may result in important new insights in the ongoing endeavor to understand the complex network of gene regulation.

Citation: Duan K, McCullough WM, Surette MG, Ware T, Song J (2012) Comprehensive Analysis of Gene-Environmental Interactions with Temporal Gene Expression Profiles in *Pseudomonas aeruginosa*. PLoS ONE 7(4): e35993. doi:10.1371/journal.pone.0035993

Editor: Baohong Zhang, East Carolina University, United States of America

Received: December 7, 2011; **Accepted:** March 26, 2012; **Published:** April 27, 2012

Copyright: © 2012 Song et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by a grant from the Canadian Cystic Fibrosis Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: songj88@umd.edu

Introduction

Current high throughput gene expression techniques, such as oligonucleotide and cDNA microarray, SAGE (series analysis gene expression), promoter array and RNA-seq [1,2,3,4] make it possible to quickly obtain vast amount of time series data in all kinds of organisms under various conditions. Gene expression can be measured simultaneously in a genome-wide manner. Temporal gene expressions under varying environmental conditions have, for instance, been measured during the cell cycle of the yeast *Saccharomyces cerevisiae* and *Bacillus subtilis* [5,6,7]. The massively abundant data prove to be invaluable for the possibility of determining the underlying various regulatory relationships among genes and their derivatives whereas the inference of genetic interactions remains to be one of the most challenging tasks of modern functional genomics [8,9,10,11,12,13].

The biological networks could in principle be divided into several types. The metabolic network is used to denote the network of proteins that synthesize and breakdown cellular molecules. It represents the enzymatic processes within the cell to transform nutrients into energy or into other molecules, i.e. biosynthesis. Protein interaction networks describe communication and signaling networks where the basic reaction is between two proteins or more. The genetic regulatory network is used to represent the general interaction of genes, gene products, and small molecules (i.e. from the DNA level, to the mRNA level, to the protein level). It describes the pathway of gene expression regulation as well as

decisions used to turn genes on/off. Deciphering interaction networks is an important task in the post-genomics era.

To build genetic networks, one of the hardest problems is the dimensionality issue, which is the exponential number of potential connections among genes [14,15]. Current solutions include clustering co-regulated genes via unsupervised analysis [16,17,18,19]. The computing methods involve choosing robust mathematical formalisms for inferring the causal connections between genes etc [20,21,22,23]. Bayesian methods [24] are excellent approaches to infer relationship between genes. They rely on prior information concerning genes, however, and it is difficult to analyze gene expression at the whole genome level due to the number of unknown genes. High throughput gene expression analysis involves many operations and at a not-insignificant cost, consequently there are not many datasets that have measured gene expression levels at a large number of time points. As a consequence, we believe that the current genetic network models generated based on few points provide limited information. Therefore, integrating diverse data types and exploring new ways to construct genetic networks are required. In this paper, to explore the interaction of gene and environmental factors, we assume that gene expression is a comprehensive process of gene and treatments. Because of the interaction, we can classify all experimental conditions into different subgroups based on the similarity of temporal gene expression profiles. Theoretically, these genes within each subgroup showing similar behaviors may share some regulatory mechanism and regulatory network. Finally, by

Expression Profiles of the 31 Genes in One Conditions

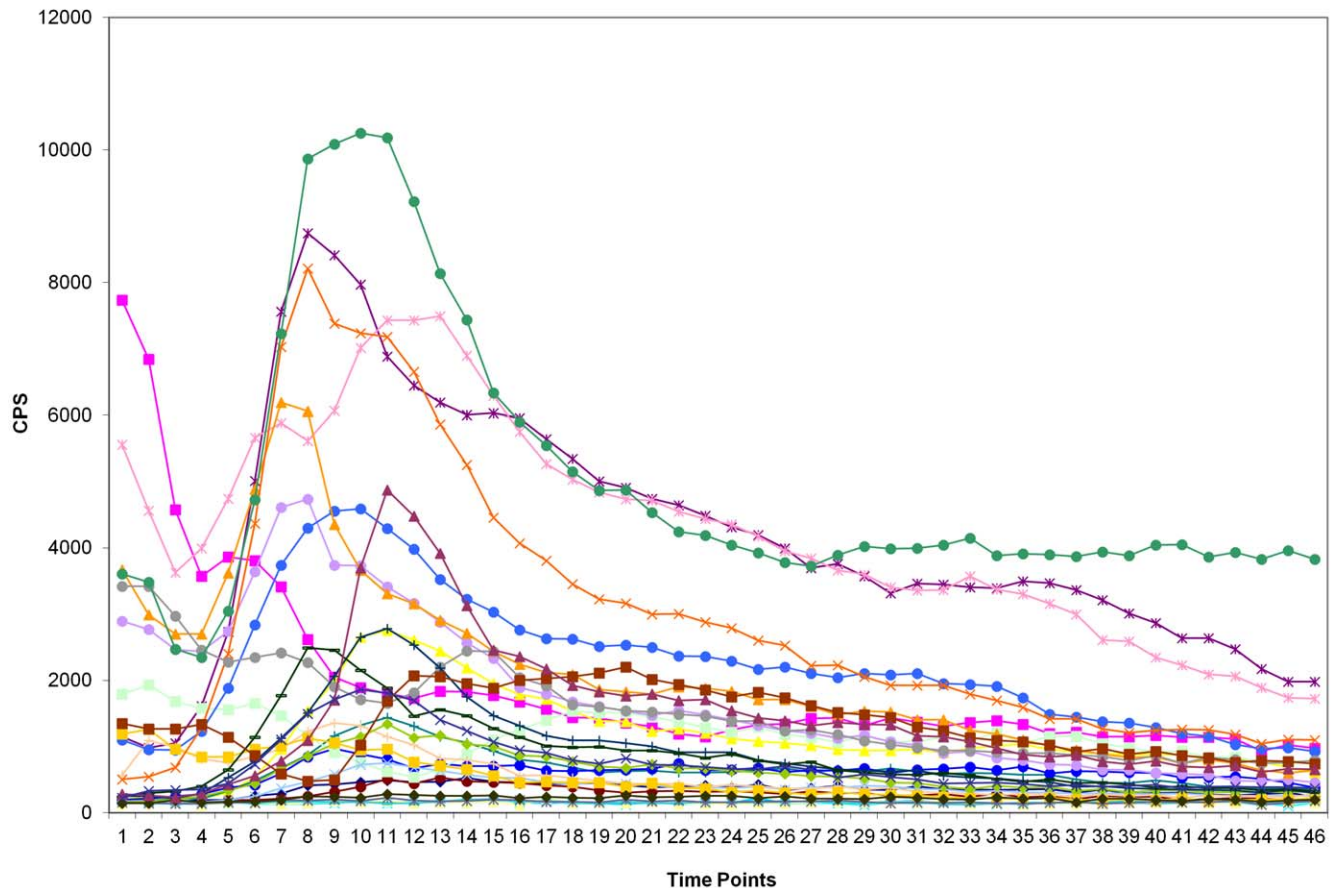


Figure 1. Expression Profiles of the 31 Genes in One Condition.
doi:10.1371/journal.pone.0035993.g001

combining all of the information, we estimated a consensus gene activation order within each subgroup. We illustrated our strategy with an example of a 31 gene set in *Pseudomonas aeruginosa*, which was expressed in 72 conditions and measured across 48 time points.

Results

The variation of gene expression profile

The large data set was from a unique gene expression experiment of the 31 promoter-reporter set in *Pseudomonas aeruginosa*, tested in 72 conditions, each with 48–60 time points. We first checked the expression profiles of the 31 genes in one condition as shown in Figure 1. We then used the *aprA* promoter-reporter as an example to show the variation in gene expression profiles. The figure shows the expression profile variations of *aprA* in different experimental conditions (Figure 2). From Figure 1 and 2, we can see not only the different behaviors of different genes, but profile differences even for each individual gene under different conditions with the maximum positions shifted among conditions. The profile types increase with condition number. Figure 3 shows the fluctuations of the mean, maximum and minimum for the *aprA* reporter at each time point for all conditions. The results clearly show the expression profiles and levels are condition-specific; they should be classified into several

subgroups based on the conditions. An attempt of building a comprehensive genetic network in all conditions is clearly unpractical even though the expression profiles of some genes do not change as dramatically in different treatment conditions as *aprA*. Alternative approaches need to be taken.

The constructed interaction networks with network motifs

To avoid conflicting gene connections in different experimental conditions and obtain the most popular genetic networks, we clustered all 72 conditions via clustering analysis based on the gene expression profiles (each gene has more than 1400 expression measurements). We used clustering result to guide the formation of environmental condition subgroups, based on the assumption that the condition-dependent expression profiles in each subgroup are similar, and that the genes in each cluster share similar expression pattern and regulatory mechanism. We calculated the transit relationship matrix of the each condition, identified the transit relationship with reference construct pMS402, and then obtained an inferred genetic network for each subgroup.

The five constructed interaction networks are shown in Figure 4. The direction of transit relationship is shown by the clockwise turn of the connecting line, and the thickness and color of each connection are proportional to its popularity and strength in the

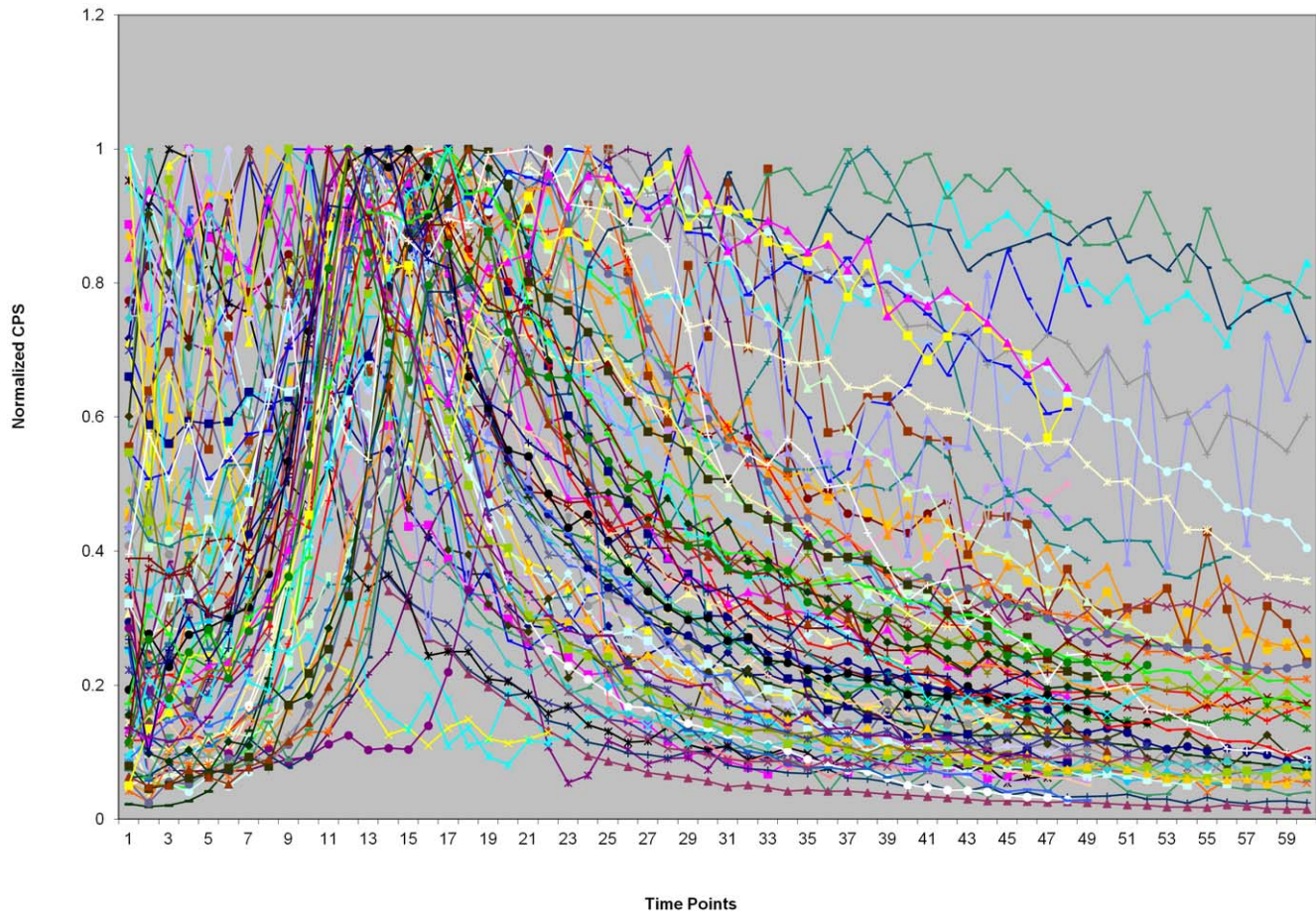


Figure 2. The *aprA* gene expression profiles in 72 conditions and 60 time points.
doi:10.1371/journal.pone.0035993.g002

subgroup. The deepest red indicates the strongest positive relationship, otherwise, the dark blue give us the most negative relationship. In the network A-E in Figure 4, we can easily identify the most popular regulation relationships via the thickness and color; for example, *migA* and *pilT* in the network A, gene *mr* and *fliC* in the network B, *hemo*, *aprA* and *plcH* in the network D, are the most popular positive transit relationships; while gene *adh* and *exoT* in the network A, *adh* and *migA* in network B and C, *pvcA* and *rhlR* in network D and *toxA*, *gacA*, and *PA4350* are the most popular negative relationship in the condition clusters. We also find the transitions are absolutely condition-specific, with the changing of the condition, the direction and strength of the interaction relationship among genes are modified, for example, the relationship between pKD203 and *fliA* in the network A and B is dramatically changed, a mild positive relationship in network A and a negative relationship in network B.

The connections among genes in network A-E (Figure 4) are neither uniformly distributed nor random, similar to that observed with genetic regulatory network motifs [27,28]. There are a lot of short paths between two genes and highly clustered connections, and several genes have more connections than others. Most of the short paths between two genes are conservative network motifs; these network motifs are very clear in the network architectures. The most popular network motif is analogous a single input module (SIM), i.e., a gene regulates simultaneously several genes; it exists in every network, for example, in network B, gene *gacA* has positive regulation relationship with gene *pvcA* and negative

association with *exoS*. The second popular network motif is analogous to dense overlapping regulons (DOR), a set of genes combinatorially control another set of genes, for example in network F, gene *gacA* and *rhlI* (pKD202) coordinately regulate gene *toxA*. The third common motif is analogous to the feed-forward loop in network B, gene *rpoS* has a mild negative connection with gene *plcH*, but *plcH* has a little stronger negatively feed-back on gene *rpoS*; such a loop is a common motif connection in network B.

Pattern Matching in Temporal Gene Expression Data

We were also interested in the general pattern matching issue in the temporal dataset. Given an arbitrary set of multivariate temporal data, how can similar patterns be located together? Here we used a novel pattern matching methodology on unsupervised learning and multivariate statistical techniques (KRZANOWSKI 1979). We obtained an original similarity matrix from the PCA similarity analysis, as shown in Figure 5A; the deep red on diagonal is similarity of 1, itself, then the redder the color, the higher the similarity. The reorganized similarity matrix based on clustering analysis is shown in Figure 5B. It illustrates the quality of cluster analysis: the clearer the block, the better the cluster analysis. It is worth noting that the PCA similarity analysis is not only for evaluating the quality of cluster analysis, but also for unknown pattern mapping. For example, the expression pattern in the complex media of sputum extracts looks most like minimal media growth conditions (Figure 6). The clustering analysis of the expression data for the 72 conditions can yield groups of

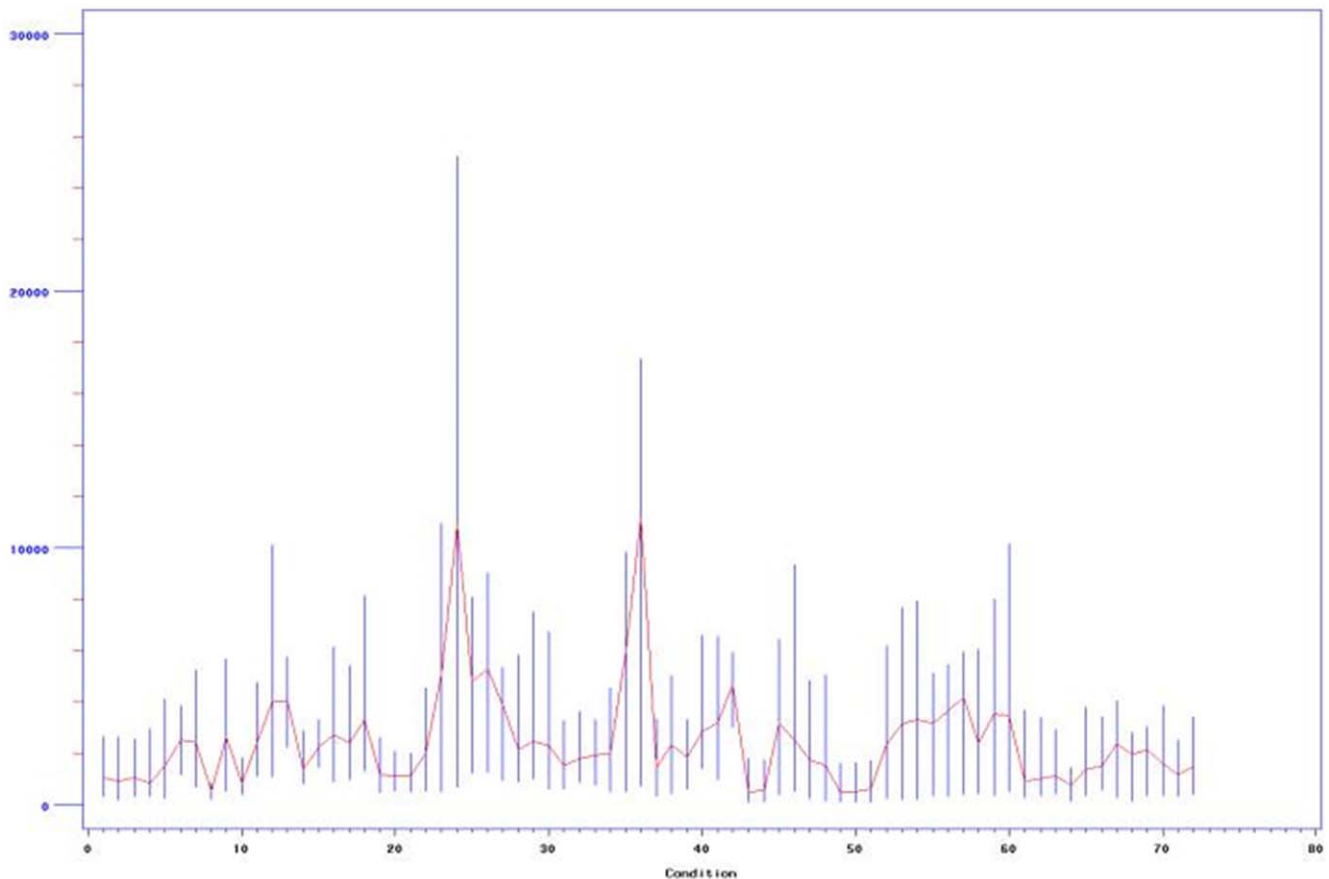


Figure 3. The fluctuation of standard deviation of *aprA* gene in different conditions and time series.
doi:10.1371/journal.pone.0035993.g003

conditions with similar expression profiles, which can be used for pattern mapping of unknown condition based on expression pattern mapping.

The Determination of Gene Activation Order

To understand the mechanism of gene expression and gene activation order, we estimated the gene turning-on and turning-off time points via the least square method, i.e., the half positions of the prior & subsequent of the maximum, as shown in the Figure 7. The visualization tool, as shown in Figure 8, reads and parses the gene expression data into an easily accessible array. The data given were 31 genes in 72 conditions; the green bars and red bars represent the gene turned on and turned off positions, respectively. We were able to sort all the data based on such information and obtain the gene order of turning off and turning on in different networks as shown in Table 1. From the analysis we find that some genes always had the same activity order; they were clearly expressed in stable patterns. For example, gene *lasI*, *rhlI*, *PA4350*, and *znuA* were in quite stable order which was not influenced by different treatment conditions, which is in agreement with the fact that *lasI* and *rhlI* are in a hierarchic order in the bacterium.

Discussion

The main purpose of exploring gene-environmental interaction is to provide indications about regulation mechanisms of the genes in response of environmental changes. The abundance of genome sequences and high throughput gene expression data is providing

input for reverse engineering of genetic networks [29,30,31,32,33,34,35]. The critical issue is how to use all kinds of information which include both genome and gene expression information to infer the relationships between genes and their activities. The strategy presented here offers new capability of extracting fundamental interaction network from expression data, and identifying the most popular regulatory relationships and gene activity order.

The most powerful feature of our strategy is the intensive temporal gene expression profile information. Because gene expression varies with the conditions, the process of building an interaction network is a difficult one. We assume that there should be a virtual network in the cell process, the connection among genes should be stable no matter observable or not. Diverse experimental conditions may enhance or repress transcription from DNA to mRNA, and lead to the changes in mRNA levels in different experimental conditions. In the past, many gene regulatory studies involved only a limited number of genes in some given conditions. It is difficult to infer genetic regulatory networks or predict connections among genes when using Bayesian theory and advanced data mining method for large and diverse prior information or expression data. Here, our key step is to cluster all conditions into several subgroups, where each of the genes in the sub-cluster set has similar expression profile and thus may share a common biochemical regulatory mechanism. Hence, the synthetic regulatory networks built are for specific condition clusters. The most popular connections indicate the fundamental regulatory relationship in each specific condition

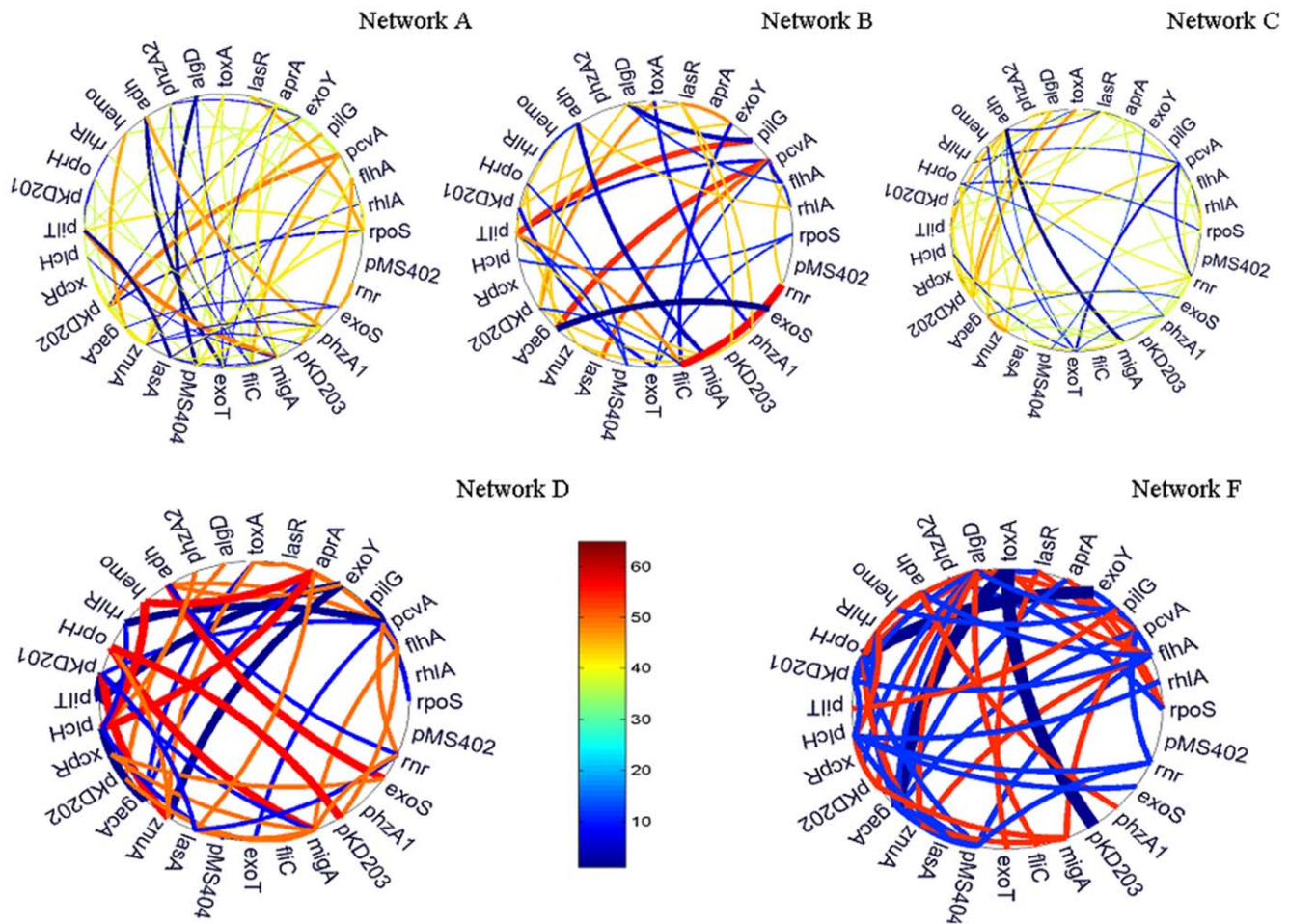


Figure 4. The networks of the five subgroups. The thickness and color of line indicate the popularity in each comprehensive genetic network. The direction of transit is clockwise. doi:10.1371/journal.pone.0035993.g004

subgroup. The main advantages of the method are simplifying the data by grouping and potentially avoiding the conflict of regulatory relationship in information combination.

Setting a reference gene is the second feature of our strategy. The threshold of genetic connections is a critical issue; it is related to the confidence and quality of the genetic regulatory network. In our research, pMS402 should not have any regulatory relationships with other genes. We used it as a control, and only picked up genes with stronger coefficients than that of pMS402. This procedure set a biological basis for the relationship matrix in each network. In future research, a statistical significance testing method will be provided for the regulations. In the predicted interaction networks, we also observed motifs analogous to the feed forward loop, single input module (SIM) and dense overlapping regulons (DOR) [27,36]. The network motifs are important for allowing multiple steady states of gene expression rates, and maintaining homeostasis of gene expression rates. Gene networks incorporated with loops, modules and regulons can readily produce oscillations and even more complex behaviors, such as quasiperiodic or chaotic variations in gene transcription rates [37].

The measuring of the turning on and turning off points via the half of maximum expression is another merit of the analysis strategy. Our strategy estimates the two time points, and pools all gene expression data in all conditions. It reveals pronounced

gene activation asymmetries, which emphasizes that gene expression during growth of bacteria is overall a strongly constrained and ordered process, and exposes the activation order of stable genes and environmental sensitive genes. The expression time courses analysis could reveal physiology state transitions in response to different environmental conditions if we have many conditions and enough time point measurements.

Overall, our computational framework adopts the principal idea that the gene expression level is the outcome of genetic regulations under specific experimental conditions, which allows classifying all experimental conditions into different subgroups based on their expression profiles, and combing more diverse gene expression data sets. The pattern matching methodology is generally applicable to a wide variety of pattern matching problems, including abnormal gene expression analysis, unknown pattern mapping and evaluation of temporal gene expression data. In addition, the estimation of activation points provides a new tool to understand the complex network of gene regulation.

Materials and Methods

Gene expression data

Promoter-reporter (*luxCDABE*) fusions for selected *P. aeruginosa* genes previously constructed [25,26] were used in this study.

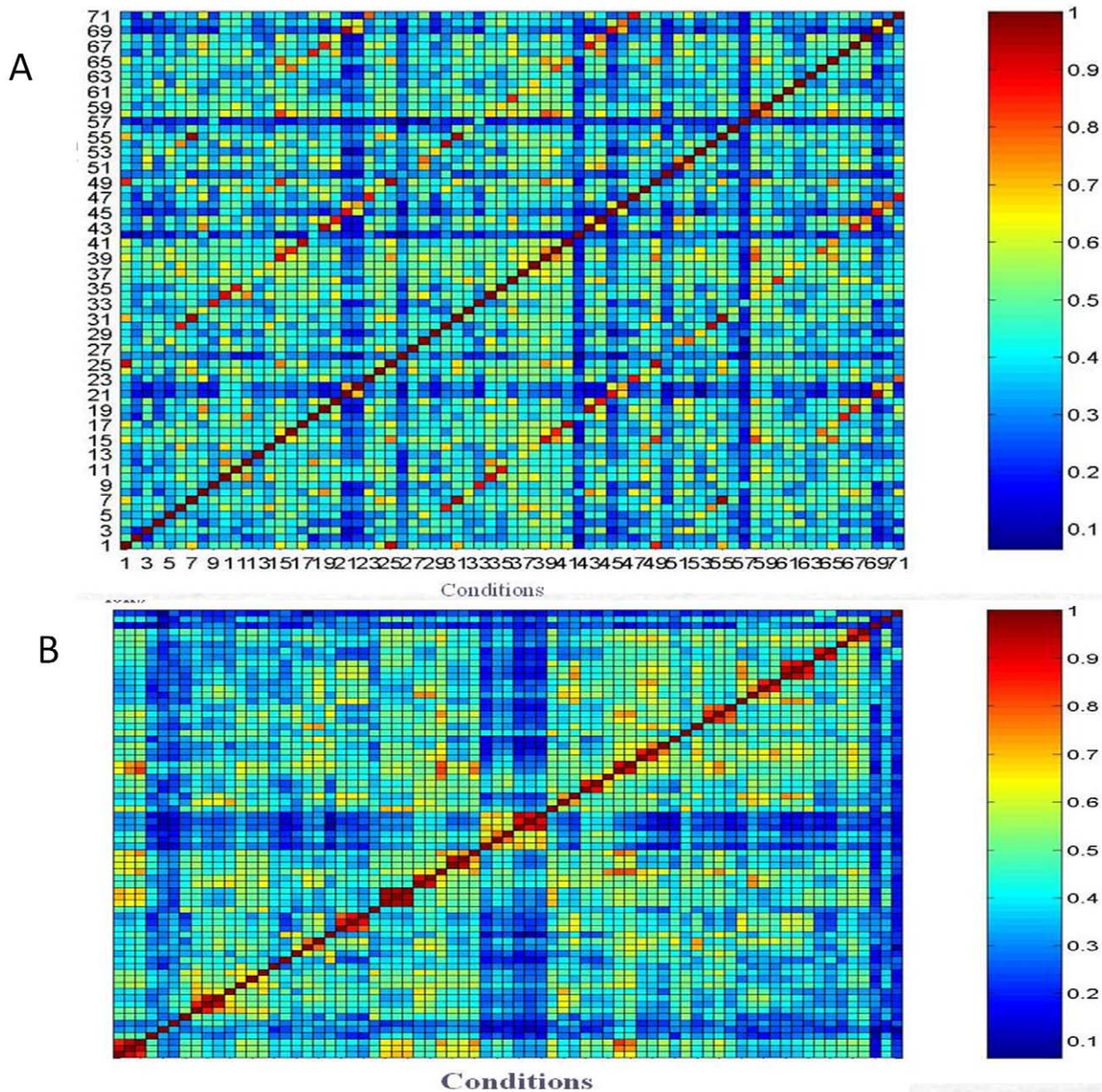


Figure 5. Pattern matching of temporal data. A. This is an original similarity matrix from PCA similarity analysis, the deep red on diagonal is similarity of itself, the similarity is 1. B. This is the reorganized similarity matrix based on clustering analysis.
doi:10.1371/journal.pone.0035993.g005

These genes are currently known virulence factors and genes that are associated with pathogenicity in *P. aeruginosa*. Reporters used in this study are listed in Table 2. Briefly, the promoter regions of selected *P. aeruginosa* virulence factors were amplified by PCR using oligonucleotide primers synthesized according to the PAO1 genome data and PAO1 chromosomal DNA as the template. The PCR amplified promoter regions were then cloned into the *XhoI*-*Bam*HI sites of pMS402 and transformed into PAO1 by electroporation. PCR, DNA manipulation and transformation were performed following general procedures. The promoterless *luxCDABE* operon in pMS402 enables the activity of the promoter fused upstream of the operon to be measured as counts per second (cps) of light production in a Victor² Multilabel Counter.

Initial cultures were grown in M9 minimal medium supplemented with casamino acid (0.1%), and glucose (0.5%) with trimethoprim added at 200 µg/ml. Overnight cultures of the reporter strains were diluted 1:200 in a 96-well microtiter plate and the promoter activity of the virulence factors in different conditions was measured every 30 minutes for 24 hours. Bacterial growth was monitored at the same time by measuring the optical density at 620 nm (OD₆₂₀) in the Victor² Multilabel Counter. All the expression assays were carried out at least twice. Growth conditions examined are listed in Table 3. These conditions include different growth media that are frequently used in the microbiology laboratories and conditions containing factors found in *P. aeruginosa* infection sites e.g. sputum extract from cystic fibrosis patients.

Gene Expression in CF sputum: Cluster with other conditions

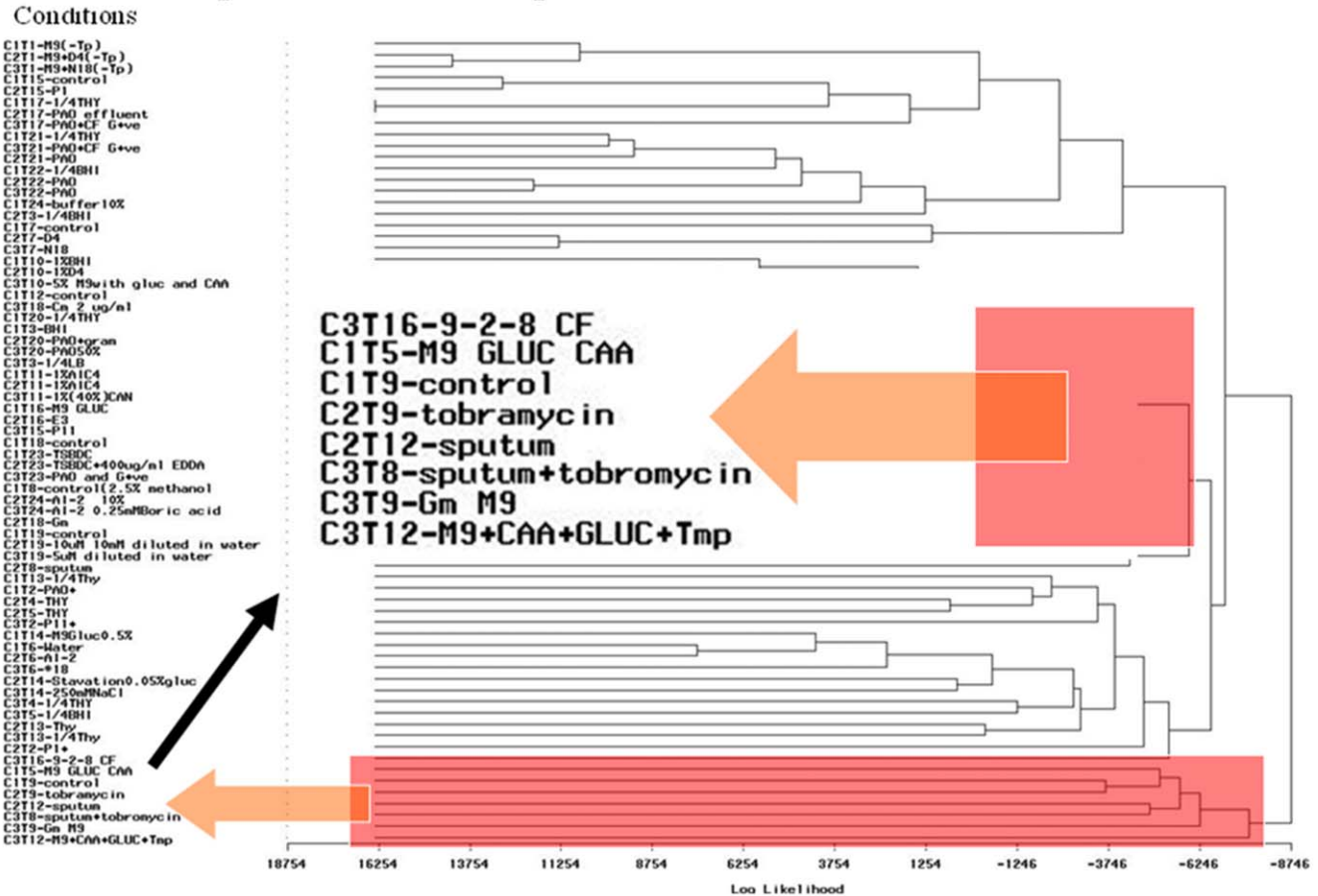


Figure 6. The mapping of unknown condition based on pattern matching of expression. For example, the expression pattern in the complex media of sputum extracts looks most like minimal media growth conditions. The clustering analysis of the expression data for the 72 conditions can yield groups of conditions with similar expression profiles, which can be used for pattern mapping of unknown condition based on expression pattern mapping.
doi:10.1371/journal.pone.0035993.g006

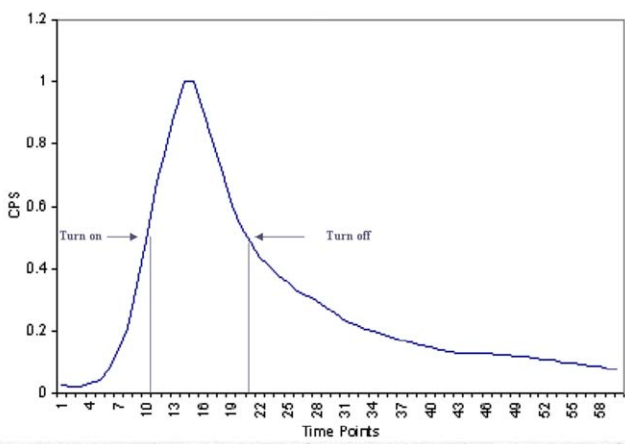


Figure 7. The expression profile of the rpoS gene. The turn point is the half position from lift maximum, the turn off point is the half position from the right maximum.
doi:10.1371/journal.pone.0035993.g007

Gene expression level model

In gene expression analysis, we assume that any gene expression level is a comprehensive result of gene effects and environmental effects. The simple formula is as follows:

$$Y = G + E + G \times E + \epsilon. \tag{1}$$

Here, Y is the column vector of expression level measurements of m genes in a specific treatment $m \times n$ matrix if there are n measurements at n time points; G is the gene effect. The effects could be single gene effects, or interaction effects among multiple genes, or a complicated genetic regulatory network for a set of genes or whole genes in a genome. The effects indicate the inner biochemical and physiological mechanisms; E is the environmental effect, it represents effects of different experimental treatments; ϵ is a random error.

Clustering analysis of gene expression with different conditions

The large scale data consisted of gene expression measurements in 72 conditions over 48 time points, all measurements were corrected with OD value, and normalized in each condition. The

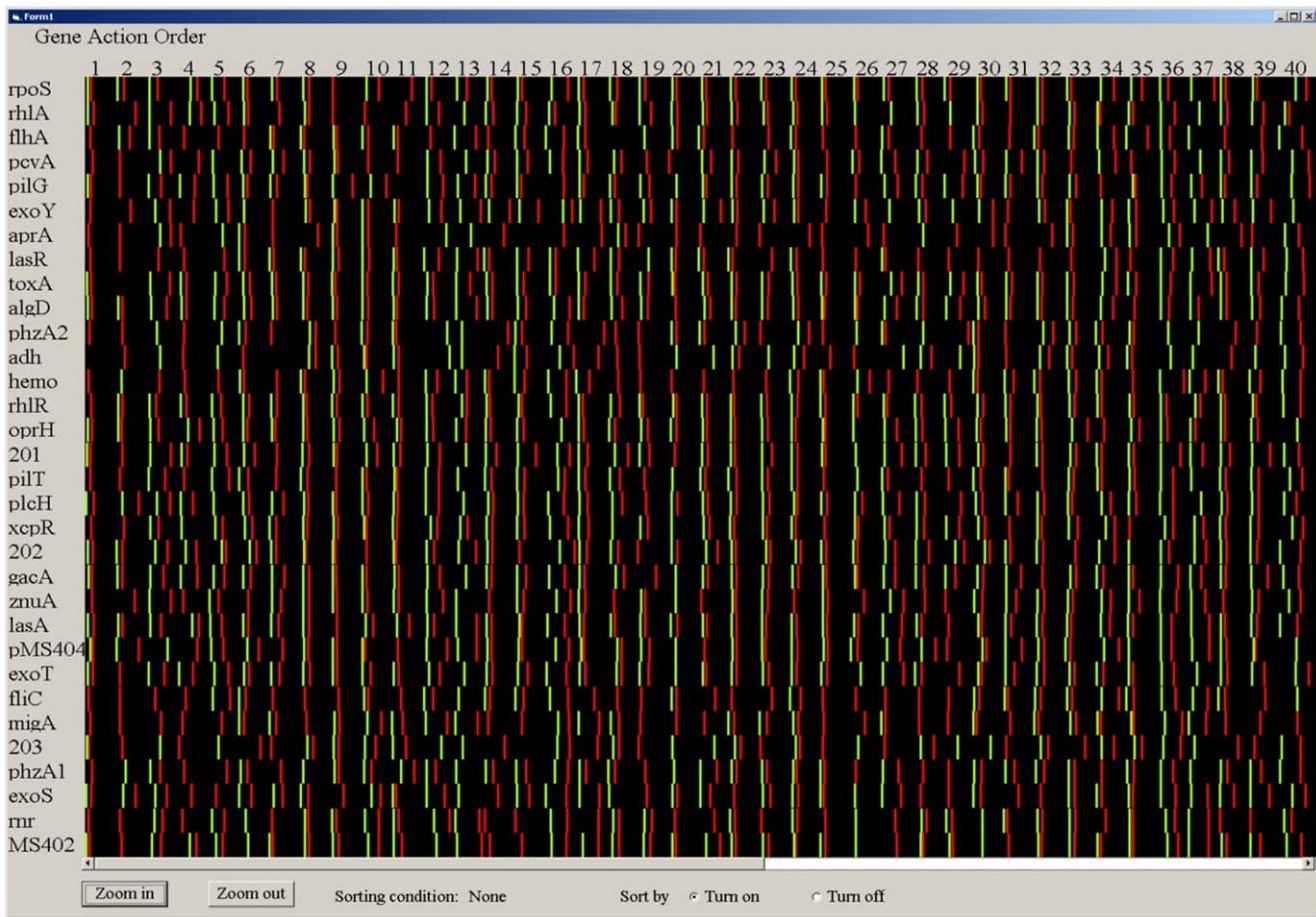


Figure 8. The visualization tool with Visual Basic. The red bar indicates the gene turning off point, the green bar indicates the gene turning on point. The gene order by genes and conditions can be obtained via sorting the data with turn on and turn off options. doi:10.1371/journal.pone.0035993.g008

distance between two clusters is defined by

$$D_{KL} = \frac{1}{N_K N_L} \sum_{i \in K} \sum_{j \in L} d(x_i, x_j).$$

$$D_{KL} = \|\bar{x}_K - \bar{x}_L\|^2 + \frac{W_K}{N_K} + \frac{W_L}{N_L}.$$

The combinational formula is

$$D_{JM} = \frac{(N_K D_{JK} + N_L D_{JL})}{N_M}.$$

If $d(x, y) = |x - y|^2$, then

Table 1. The Gene Activation Order in the Networks.

Network	Status	Activation Order (from left to right)
A	Turn-on	algD, adh, znuA, xcpR/toxA, pKD201/pKD202/pMS404, rhIA, D203, fliC, exoS, exoT
	Turn-off	RpoS, xcpR, migA, znuA, xcpR, pKD201/pKD202/pKD203, rhIA, rhIR, plcH, pMS404, phzA1/phzA2
B	Turn-on	pKD201, pKD203/pKD202, znuA, rpoS/rnr, xcpR, exoS, phzA1/phzA2,
	Turn-off	phzA1/xcpR, pKD201/pKD202/pKD203, rpoS/rnr, phzA1/phzA2, hemo, exoY, exoS.
C	Turn-on	Adh/aprA, exoY, toxA/xcpR/rpoS, pKD201/pKD202/pKD203, plcH, pilG, phzA1/phzA2, lasR, znuA, rnr
	Turn-off	plcH, rhIA/rhIR/toxA, pKD202/pKD201, exoS/lasR, rpoS, algD, phzA1/PhzA2, migA, pMS404, znuA, aprA, pKD203,
D	Turn-on	pKD202, pKD201, znuA, pKD203,
	Turn-off	PlcH, rpoS, pKD201, pKD202, pKD203, rnr, xcpR, toxA, algD, lasA, pilT
E	Turn-on	pMS404, rhIR, toxA
	Turn-off	pMS404, znuA, lasR, phzA1

doi:10.1371/journal.pone.0035993.t001

Table 2. The list of the gene reporters used in this study.

Gene	Function	PA number
lasI (pKD201)	AHL synthase	PA1432
lasR	AHL dependent transcriptional regulator	PA1430
rhII (pKD202)	AHL synthase (rhIL)	PA3476
rhIR	AHL dependent transcriptional regulator	PA3477
lasA	protease (staphylolytic protease preproenzyme LasA)	PA1871
lasB (pMS404)	Elastase	PA3724
aprA	alkaline protease (alkaline metalloproteinase precursor)	PA1249
xcpP	xcp secretion pathway (different orientation from xcpR)	PA3104
xcpR	xcp (general secretion pathway protein E)	PA3103
rhIA	rhaminolipid (rhamnosyltransferase chain A)	PA3479
rpoS	stationary phase sigma	PA3622
gacA	transcriptional activator, response regulator	PA2586
pilT	Type IV pili (twitching motility protein PilT, pilT I followed by pilU)	PA0395
pilG	Type IV fimbrial (Part of the pilGHJKL gene cluster)	PA0408
algD	alginate (GDP-mannose 6-dehydrogenase AlgD), first of 18-kb alginate operon.	PA3540
plcH	hemolytic phospholipaseC (hemolysin) precursor	PA0844
toxA	exotoxinA	PA1148
exoS	exoenzymeS (ADP-ribosyltransferase)	PA3841
exoT	exoenzymeT (99% similar to ADP-ribosyltransferase (exoenzyme 53))	PA0044
exoY	adenylate cyclase	PA2191
PhzA1	pyocyanin synthesis (phenazine synthesis cluster)	PA4210
PhzA2	pyocyanin synthesis (phenazine synthesis cluster, first gene)	PA1899
pvcA	pyoverdine biosynthesis protein PvcA, first of four ORF cluster	PA2254
hem	putative hemagglutinins (43% identity to B. pertussis)	PA0041
nrn	exoribonuclease RNase R (virulence protein VacB) class2	PA4937
adh	probable adhesion protein	PA2407
znuA	probable adhesion	PA5498
fliC	flagellin	PA1092
flhA	flagellar biosynthesis protein	PA1452
migA	probable glycosyl transferase (mucin-inducible gene)	PA0705
oprH	PhoP/Q and low Mg ²⁺ inducible outer membrane protein H1 precursor	PA1178
PA4350 (pKD203)	Putative hemolysin	PA4350

doi:10.1371/journal.pone.0035993.t002

In average linkage the distance between two clusters is the average distance between pairs of observations, one in each cluster. Average linkage tends to join clusters with small variances, and is slightly biased toward producing clusters with the same variance.

Analysis methods

We deal with gene expression data from m genes over n time points t_1, \dots, t_n . The data are represented in an $m \times n$ array Y , and we assume that the gene expression levels at time t_{i+1} are determined by the expression levels at time t_i via the functional

$$Y(:, n+1) = f(W * X(:, n) + B(\cdot)) \sum_{i=1}^k f(W * X(:, n) + B(\cdot)),$$

$$Y(:, n+1) = f(W * Y(:, n) + \underline{b}), i = 1, \dots, n-1$$

where W is an $m \times m$ matrix, \underline{b} is a $m \times 1$ vector, f is some nonlinear switching function (for example a sigmoid centered at 0)

which acts on each element of the $m \times 1$ vector $W * Y(:, i) + \underline{b}$ to produce the $m \times 1$ vector $Y(:, i+1)$. The notation used here has the following meaning: if A is an array, then $A(:, i)$ denotes the i th column of A .

Equation above can be partially inverted to give

$$W * Y(:, n) + \underline{b} = f^{-1}(Y(:, i+1)) \quad i = 1, \dots, n-1,$$

and we are trying to use these equation to determine the matrix W and the vector \underline{b} . To do this most conveniently, we group the equations together, writing: $Y_{out} = [Y(:, 2), \dots, Y(:, n)]$ and $Y_{in} = [Y(:, 1), \dots, Y(:, n-1)]$ (note that these are $m \times (n-1)$ arrays). We also write

$$\tilde{W} = [W | \underline{b}] \quad \text{and} \quad \tilde{Y}_{in} = \begin{bmatrix} Y_{in} \\ \mathbf{1}^T \end{bmatrix}$$

where $\mathbf{1}^T$ is a row of $n-1$ ones. Then $\tilde{W} * \tilde{Y}_{in} = f^{-1}(Y_{out})$, and we

Table 3. The environmental conditions tested.

Condition number	Condition code	Description	Condition number	Condition code	Description
1	C1T1	M9 medium (-Tp)	37	C2T13	THY medium
2	C1T2	with PAO spent culture	38	C2T14	stivation M9+0.05%gluc
3	C1T3	BHI medium	39	C2T15	co-culture with P1 isolate
4	C1T4	LB medium	40	C2T16	co-culture with E3 isolate
5	C1T5	M9+0.05% CAA	41	C2T17	with 1% PAO biofilm effluent
6	C1T6	1% H2O	42	C2T18	with subinhibitory Gm
7	C1T7	control 1	43	C2T19	with 10 uM Al-2
8	C1T8	with 2.5% methanol	44	C2T20	with supernatant of PAO+gram
9	C1T9	control 2	45	C2T21	with supernatant of PAO
10	C1T10	M9+1% BHI	46	C2T22	with supernatant of PAO
11	C1T11	M9 with 1% AHL C4	47	C2T23	TSBDC+400 ug/ml EDDA
12	C1T12	control 3	48	C2T24	with Al-2
13	C1T13	LB medium	49	C3T1	co-culture with isolate N18(-Tp)
14	C1T14	M9 with 0.5% Gluc	50	C3T2	co-culture with isolate P11
15	C1T15	control 4	51	C3T3	1/4 diluted LB
16	C1T16	M9+1% Gluc	52	C3T4	1/4diluted THY
17	C1T17	1/4 diluted THY	53	C3T5	1/4 diluted BHI
18	C1T18	control 5	54	C3T6	with Al-2 analog #18
19	C1T19	control 6	55	C3T7	co-culture with isolate N18
20	C1T20	1/4 diluted THY	56	C3T8	with sputum extract and tobromycin
21	C1T21	1/4 dilute THY	57	C3T9	with subinhibitory Gm
22	C1T22	1/4 diluted BHI	58	C3T10	M9 with 5% gluc and CAA
23	C1T23	TSBDC medium	59	C3T11	1%(40% CAN
24	C1T24	PBS buffer 10%	60	C3T12	M9+CAA+Gluc+Tmp
25	C2T1	co-culture with isolate D4(-Tp)	61	C3T13	1/4 diluted THY
26	C2T2	co-culture with isolate P1	62	C3T14	250 mM NaCl
27	C2T3	1/4 diluted BHI	63	C3T15	co-culture with isolate P11
28	C2T4	THY medium	64	C3T16	co-culture with isolate 9-2-8
29	C2T5	THY medium	65	C3T17	with supernatant of PAO+CF G+ve isolate
30	C2T6	with 25uM Al-2	66	C3T18	with Cm 2 ug/ml
31	C2T7	co-culture with isolate D4	67	C3T19	5 uM Al-2 diluted in water
32	C2T8	with 1% sputum extract	68	C3T20	with supernatant of PAO 50%
33	C2T9	with subinhibitory tobramycin	69	C3T21	with supernatant of PAO+CF G+ve isolate
34	C2T10	1% D4 supernatant	70	C3T22	with supernatant of PAO
35	C2T11	1% AHL C4	71	C3T23	with supernatant of PAO and G+ve isolate
36	C2T12	with 2.5% sputum extract	72	C3T24	0.25 mM Boric acid

doi:10.1371/journal.pone.0035993.t003

want to solve this system for \tilde{W} . This is done (in a least-squares sense if the system is over or under-determined) by using the pseudoinverse of \tilde{Y}_{in} .

The thresholds of transit relationship matrix

Because pMS402 contains the promoterless *luxCDABE* operon, we set pMS402 as a reference. Theoretically, it does not have any relationships with other genes. To avoid meaningless regulation associations among the gene set, we took the absolute value of the coefficients of the gene pMS402 among the relationship matrix to be the threshold in each condition. If coefficients among two genes in the transit relationship matrix are greater than the threshold,

the coefficient is positive, and is given the value 1, indicating positive transit relation, otherwise it is given the value -1 , representing a negative transit relationship; any coefficient less than the threshold is given the value 0, representing no regulatory associations between the two genes.

Pattern similarity analysis

Assuming that two data sets contain the same n variables but not necessarily the same number of measurements, for each data set we consider a principal component analysis (PCA) model containing k principal components (PC). The PC number, k , is chosen such that k principal components explain at least 95% of

the total variance of each data sets. The similarity between the two data sets is measured by comparing their principal components, defining as a single number. Let be the PC number describing at least 95% of the variance in data set S, and let as the PC number in data set H, which also describe 95% of its variance. If $k = \max(k_S, k_H)$, it ensures that k principal components explain 95% of the variance in each data set. The PCA similarity factor compares can be calculated from the angles between principal components:

$$S_{PCA} = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k \cos^2 \theta_{ij}$$

where θ_{ij} is the angle between the principal component i in data set S and principal component i in data set H. The similarity factor can also be expressed as

$$S_{PCA} = \frac{\text{trace}(L^T M M^T L)}{k}$$

here, L and M contain the k most significant principal components for S and H.

Visualization of gene turning-on and turning-off position

To describe gene effects mechanisms in regulatory networks, we defined expression prior to the maximum as the turning on section, with the half position of the maximum being the gene turning on point. Expression subsequent to the maximum is the gene turning off section, with the half position of the maximum

being the gene turning off point. We estimated the two positions via the least squares method. This visualization program reads and parses the gene expression data into an easily accessible array, and was created using Visual Basic. The data given were 31 genes in 72 conditions for 48 time points. When the gene was turned on there would be a 1, when the gene was turned off there would be a -1 and the rest of the time points would be zero. The screen was divided into 40 sections, where each section represented a particular condition. The section is divided into 60 different areas. When the gene was turned on, a green bar would be placed in one of these 60 areas, and when the gene was turned off a red bar would be placed into one of these areas. It was not unusual for a gene to be on before the experiment or on after the experiment was done, and in this case the appropriate bar was removed. The program also allowed you to limit the number of conditions that was displayed at any one time. For ease of viewing the upper limit was 40 conditions, and for ease of coding, the lower limit was put at 5. The display can be sorted four ways. The first is by the order of activation in one condition, or sorting by which gene was on first. The second is by which gene turned off first in a particular condition. The third is sorting the conditions based upon which one turned on a specific gene first. The last is sorting the conditions based upon which condition turned off a specific gene first.

Author Contributions

Conceived and designed the experiments: JZ. Performed the experiments: KM. Analyzed the data: TW. Contributed reagents/materials/analysis tools: MS WM. Wrote the paper: JZ KM.

References

- Xu R, Mao JH (2011) Gene transcriptional networks integrate microenvironmental signals in human breast cancer. *Integr Biol (Camb)* 3: 368–374.
- Wang L, Tang H, Thayanithy V, Subramanian S, Oberg AL, et al. (2009) Gene networks and microRNAs implicated in aggressive prostate cancer. *Cancer Res* 69: 9490–9497.
- Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, et al. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 292: 2080–2083.
- Heller MJ (2002) DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng* 4: 129–153.
- Logan MR, Nguyen T, Szapitel N, Knockleby J, Por H, et al. (2008) Genetic interaction network of the Saccharomyces cerevisiae type 1 phosphatase Glc7. *BMC Genomics* 9: 336.
- Lin YY, Qi Y, Lu JY, Pan X, Yuan DS, et al. (2008) A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation. *Genes Dev* 22: 2062–2074.
- DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686.
- Arda HE, Walhout AJ (2010) Gene-centered regulatory networks. *Brief Funct Genomics* 9: 4–12.
- Malcom JW (2011) Gene networks and metacommunities: dispersal differences can override adaptive advantage. *PLoS One* 6: e21541.
- Shimamura T, Imoto S, Yamaguchi R, Nagasaki M, Miyano S (2010) Inferring dynamic gene networks under varying conditions for transcriptomic network comparison. *Bioinformatics* 26: 1064–1072.
- Takahashi H, Morioka R, Ito R, Oshima T, Altaf-Ul-Amin M, et al. (2011) Dynamics of time-lagged gene-to-metabolite networks of Escherichia coli elucidated by integrative omics approach. *OMICS* 15: 15–23.
- Waaijenborg S, Zwinderman AH (2009) Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks. *BMC Bioinformatics* 10: 315.
- Zou M, Conzen SD (2005) A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* 21: 71–79.
- Huang MC, Wu JW, Luo YP, Petrosyan KG (2010) Fluctuations in gene regulatory networks as Gaussian colored noise. *J Chem Phys* 132: 155101.
- Bickel DR (2005) Probabilities of spurious connections in gene networks: application to expression time series. *Bioinformatics* 21: 1121–1128.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. *Nat Genet* 22: 281–285.
- Fabry-Asztalos L, Andonie R, Collar CJ, Abdul-Wahid S, Salim N (2008) A genetic algorithm optimized fuzzy neural network analysis of the affinity of inhibitors for HIV-1 protease. *Bioorg Med Chem* 16: 2903–2911.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863–14868.
- Banerjee N, Zhang MQ (2002) Functional genomics as applied to mapping transcription regulatory networks. *Curr Opin Microbiol* 5: 313–317.
- Weaver DC, Workman CT, Stormo GD (1999) Modeling regulatory networks with weight matrices. *Pac Symp Biocomput*. pp 112–123.
- D'Haeseleer P, Liang S, Somogyi R (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16: 707–726.
- De Jong H, Geiselmann J, Hernandez C, Page M (2003) Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics* 19: 336–344.
- Hasty J, McMillen D, Isaacs F, Collins JJ (2001) Computational studies of gene regulatory networks: in numero molecular biology. *Nat Rev Genet* 2: 268–279.
- Pe'er D, Regev A, Elidan G, Friedman N (2001) Inferring subnetworks from perturbed expression profiles. *Bioinformatics* 17 Suppl 1: S215–224.
- Shen L, Shi Y, Zhang D, Wei J, Surette MG, et al. (2008) Modulation of secreted virulence factor genes by subinhibitory concentrations of antibiotics in Pseudomonas aeruginosa. *J Microbiol* 46: 441–447.
- Duan K, Dammal C, Stein J, Rabin H, Surette MG (2003) Modulation of Pseudomonas aeruginosa gene expression by host microflora through interspecies communication. *Mol Microbiol* 50: 1477–1491.
- Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet* 31: 64–68.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. *Science* 298: 824–827.
- Hache H, Lehrach H, Herwig R (2009) Reverse engineering of gene regulatory networks: a comparative study. *EURASIP J Bioinform Syst Biol* 617281 p.
- Lin K, Husmeier D, Dondelinger F, Mayer CD, Liu H, et al. (2010) Reverse engineering gene regulatory networks related to quorum sensing in the plant pathogen Pectobacterium atrosepticum. *Methods Mol Biol* 673: 253–281.
- Summer G, Perkins TJ (2010) Functional data analysis for identifying nonlinear models of gene regulatory networks. *BMC Genomics* 11 Suppl 4: S18.
- Hempel S, Koseska A, Nikoloski Z, Kurths J (2011) Unraveling gene regulatory networks from time-resolved gene expression data – a measures comparison study. *BMC Bioinformatics* 12: 292.
- Ooi BN, Phan TT (2011) Insights gained from the reverse engineering of gene networks in keloid fibroblasts. *Theor Biol Med Model* 8: 13.

34. Taylor RC, Sanfilippo A, McDermott JE, Baddeley B, Riensche R, et al. (2011) Enriching regulatory networks by bootstrap learning using optimised GO-based gene similarity and gene links mined from PubMed abstracts. *Int J Comput Biol Drug Des* 4: 56–82.
35. Szallasi Z (1999) Genetic network analysis in light of massively parallel biological data acquisition. *Pac Symp Biocomput* 4: 5–16.
36. Thomas R, Thieffry D, Kaufman M (1995) Dynamical behaviour of biological regulatory networks—I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol* 57: 247–276.
37. Smolen P, Baxter DA, Byrne JH (2000) Modeling transcriptional control in gene networks—methods, recent results, and future directions. *Bull Math Biol* 62: 247–292.