Detecting shifts in gene regulatory networks during time-course experiments at single-time-point temporal resolution

Yoichi Takenaka*, Shigeto Seno and Hideo Matsuda
*Graduate School of Information Science and Technology
Osaka University
Yamadaoka 1-5, Suita, Osaka, Japan
takenaka@ist.osaka-u.ac.jp

Accepted 13 September 2015
Published 27 October 2015

Comprehensively understanding the dynamics of biological systems is one of the greatest challenges in biology. Vastly improved biological technologies have provided vast amounts of information that must be understood by bioinformatics and systems biology researchers. Gene regulations have been frequently modeled by ordinary differential equations or graphical models based on time-course gene expression profiles. The state-of-the-art computational approaches for analyzing gene regulations assume that their models are same throughout time-course experiments. However, these approaches cannot easily analyze transient changes at a time point, such as diauxic shift. We propose a score that analyzes the gene regulations at each time point. The score is based on the information gains of information criterion values. The method detects the shifts in gene regulatory networks (GRNs) during time-course experiments with single-time-point resolution. The effectiveness of the method is evaluated on the diauxic shift from glucose to lactose in *Escherichia coli*. Gene regulation shifts were detected at two time points: the first corresponding to the time at which the growth of *E. coli* ceased and the second corresponding to the end of the experiment, when the nutrient sources (glucose and lactose) had become exhausted. According to these results, the proposed score and method can appropriately detect the time of gene regulation shifts. The method based on the proposed score provides a new tool for analyzing dynamic biological systems. Because the score value indicates the strength of gene regulation at each time point in a gene expression profile, it can potentially infer hidden GRNs from time-course experiments.

Keywords: Gene regulatory network; network dynamics; time-course; diauxic shift.

1. Background

Comprehensively understanding the dynamics of biological systems is one of the greatest challenges in biology. Biological technologies have dramatically improved in recent years, providing us with a vast amount of information, including the DNA sequences of genes and genomes, amino acid sequences and the steric structures of

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 3.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

proteins, mRNA expression levels, and the relationships among gene regulations. These data have become freely available through public databases. Bioinformatics must now advance to discover novel knowledge, pertaining to the dynamics of biological systems.

Biological dynamics are typically investigated using time-course gene expression profiles and mathematical and informatics models such as ordinary differential equations and graphical models. These models describe dynamics using an equation and an edge. Suppose the conditions of an environment inhabited by an organism change. The gene regulatory networks (GRNs) in both environments are known, but the transitions that bridge the gap between them are unknown. Figure 1 illustrates a diauxic shift from glucose to lactose as the sole nutrient source. The dynamics are modeled by an ordinary differential equation model that evolves the expression of each gene. The bridge between the two environments is described by a graphical model. However, when the shift starts and ends cannot be determined using these approaches. Therefore, we propose a score-based method that detects the timing of the shift of a GRN.

1.1. Gene expressions

Gene expression is the transcription of DNA to mRNA within cells. The number of mRNA molecules per gene is determined using DNA arrays, chips, and sequencers. A set of gene expression values throughout an experiment is called a gene expression

Fig. 1. Stages of regulations of gene expression. The proposed method detects the time of a shift to a new stage. The nutrient source switches from glucose to lactose. The GRNs of both environments are known. The ordinary differential equation system models each regulated gene according to a single equation. The graphical model describes a GRN that bridges the two environments.

profile. Many such profiles have been registered in public databases such as the Gene Expression Omnibus (GEO)^{1,2} and Array Express.³ Special-purpose databases are also available; for instance, the Reference Expression Dataset (RefEx) archives mammalian tissue gene expression data,⁴ the Human Protein Atlas contains profiles of human tissues, cell lines, and cancers,⁵ and the Gene Expression Database (GXD) of Mouse Genome Informatics (MGI) specializes in mouse genomics.⁶ These databases contain the time-course gene expression profiles that researchers require for dynamic biological systems analysis.

1.2. Gene regulations

Gene expression is controlled primarily by the products of transcription factor (TF) genes. The process by which organisms control their gene expression levels is called gene regulation. Gene regulations have been studied in model organisms such as *Escherichia coli*, whose gene regulations are almost completely known and are archived in RegulonDB.^{7,8} The FANTOM projects have recorded the gene regulations of various mouse and human tissue types.^{9–11} Gene regulations have been extensively researched and are collected in pay-to-access databases such as TransFac^{12,13} and QIAGEN's Ingenuity Pathway Analysis.¹⁴ In the near future, gene regulation in the static state will be well known.

1.3. Dynamic analysis

Dynamic analysis follow the static-state gene regulations. To investigate the changes in gene regulations, researchers refer to time-course gene expression profiles. In fact, the FANTOM projects have measured time-course deepCAGE data^{9,15} and have attempted to reveal the dynamics of human and mouse systems.^{16,17} Dynamic analysis ranks among the most exciting research topics, not only in biology but also in bioinformatics and systems biology. To analyze the large number of time-course gene expression profiles in public databases, researchers have proposed various bioinformatic methods.

1.3.1. Ordinary differential equations

The dynamics of biological systems are frequently modeled by ordinary differential equations. The independent variable in these equations is the expression level of a gene.^{18,19} Reverse-engineering algorithms based on ordinary differential equations yield directed graphs from time-series expression profiles.^{20,21}

1.3.2. Graphical models

A graphical model describes the relationships among control and controlled genes as directed edges between nodes that represent the genes. The graphical Gaussian model defines an independence graph, in which the interdependences of any two genes are measured using partial correlations.²² Because this model can express the Markov property, it can analyze the dynamics of biological systems.²³

Given a measure of association between gene expression levels and a threshold value, the Gene Relevance Network generates a network from gene expression profiles. The association between two genes is measured using their mutual information, and the genes are connected by an edge if the mutual information exceeds the threshold.^{24–26} This model can also express time-course gene expression profiles.²⁷

Bayesian networks represent the probabilistic relationships among genes. The joint probability distribution of a gene expression profile is derived by information-theoretic approaches, such as the Bayesian information criterion and minimum description length.²⁸ Bayesian networks have been extended to dynamic Bayesian networks, which infer interactions with Markov properties.^{29,30}

1.4. *Single-time-point analyses*

Using the above methods, we can obtain the basic governing equation for each gene or GRN. These equations describe the behaviors observed in experiments. For example in Fig. 1, the ordinary differential equation models relate each gene to its parent, and the graphical models generate a GRN. These approaches assume a differential equation for each gene or a GRN throughout the experiment. This assumption was required in initial analyses of biological systems dynamics because the number of time points in gene expression profiles was very restricted. However, these circumstances have changed with the increasing availability of cheap gene chips and DNA sequencers. The number of time points in gene expression profiles has gradually increased, and some researchers have discarded the above assumption *et al.*^{31–34} Tamada *et al.*,³⁴ Jia and Huan³¹ and Nakayama *et al.*³² split the time-course gene expression profiles into two halves, from which they inferred two GRNs. Their methods are easily expanded to three or more networks by splitting the profiles into three or more sub-profiles. However, they cannot infer the network at a single time point because single-valued measurements are not suitable for this purpose. Song *et al.*³³ proposed a method to infer the network at a single time point. Their method penalizes gene expression data from distant time points more than those from close time point via a kernel re-weighted auto-regression. Because the method is sensitive to the kernel parameters and requires tuning for each experiment, it was not practical. Therefore, to infer the relations between genes at each time point, a new approach is required.

1.5. *Increase of known regulations*

The above methods have revealed novel regulations among genes, which have been reported in the literature and stored in databases. Most of the GRNs of *E. coli* are now known, and the networks of other organisms should be revealed in the near future. This information is invaluable for dynamic biological systems analysis, because rather than inferring the novel regulations, we need access only to the known regulations. Sufficient knowledge of *E. coli* is already available for this undertaking. Under these circumstances, we assume that the GRN of the target organism is given in the dynamic analysis.

Fig. 2. Stages of regulations of gene expression. The proposed method detects the time points of shifting to a new stage. Assume that there are two stages or hidden networks between GRNs that are active in the presence of glucose and lactose. Our method captures the shift at the temporal resolution of a single time point.

We propose a score and a method that times the gene regulation in time-course gene expression profiles. The temporal resolution of the proposed method is a single time point. Our concept is illustrated in Fig. 2. The method aims to detect the time points at which two hidden GRNs are activated during the experiment.

2. Methods

2.1. GRN model and the information criterion

To detect the temporal variation of the GRN, we propose a score that quantifies the strength at which the expression levels of a target gene are controlled at a specified point in time. From this score, we detect when the GRN shifts from the time-course expression profile. The genetic control relations are described by conditional probability and the degree of conformance between the model and gene expression profile is digitized using an information criterion. This model is used to infer Bayesian networks.^{29,35,36} The information criteria can be formulated in several ways. The Akaike, Bayesian, and other information criterion share a common property; the smaller the value, the more closely the expression profiles conform to the gene regulator network. The Bayesian approach searches for the network with the smallest information criterion value.

2.2. The information criterion with one sample

The conditional probability of the model is derived from the expression profile. In particular, the value of the information criterion is calculated from the expression profile. Furthermore, the calculation requires only part of the profile. For example, the expression profile can be subdivided into two half-profiles, from which two GRNs can be inferred. As mentioned above, this approach was adopted by Tamada *et al.*³⁴

and Nakayama *et al.*³² However, the information criterion value cannot be calculated from a profile with only one sample point. In this case, because the information amount of the conditional probability is zero, the information criterion value becomes a constant. To avoid this problem, a conceptual breakthrough is required. Here we define the information criterion value of a single time point from the gene expression profile with the target point removed. The information criterion value without the time point is then subtracted from the average of the information criterion values obtained by removing each of the single time points. We consider that the higher the value, the stronger the control over the expression level of the target gene at that time point.

2.3. Information gain

The information criterion computes a relative value. To make sense of the value computed by a model, we must compare it with the values of other models. Here we assume two gene models. The first assumes that gene expression is controlled by other genes and the second assumes that gene expression is independent of other genes. The information criterion value of both models is assumed to be small, but it is larger in the first method than in the second one. Under these assumptions, the gene is more likely to be independent of other genes than controlled by other genes. The difference between the values of the two models indicates the strength at which other genes control the expression of the target gene. This difference is called information gain.

2.4. Definitions

The strength at which the expression levels of a target gene are controlled by other genes at a given time point is described by the score. To define the score, we require the following notations.

GRN(V, E)	Gene regulatory network, where V is a set of genes and E is a set of directed edges ($v_1, v_2 v_1, v_2 \in V$),
EXP	Time-course gene expression profile with T time points,
EXP _{A}	Subset of EXP containing a set of genes A ,
EXP _{A, t}	EXP _{A} excluding time point t ,
EXP _{$*, t$}	EXP excluding time point t ,
IC(EXP, $g A$)	Information criterion value of the probability of EXP _{$\{g\}$} under the condition EXP _{A} , where g is a gene and A is a set of genes.

Equation (1) defines the information criterion of a gene g under the condition of A at time point t with an expression profile EXP.

$$\text{IC}_t(g|A) = \text{IC}(\text{EXP}_{*, t}, g|A). \tag{1}$$

Equation (2) defines the information gain of a gene g against a gene regulatory network $\text{GRN}(V, E)$. The information gain is the difference between the information values in the presence and absence of conditions.

$$\text{IG}_{g,t} = \text{IC}_t(g) - \text{IC}_t(g|\{v|(v, g) \in E\}). \quad (2)$$

Finally, the score of a gene g at time point t is given by Eq. (3):

$$\text{Score}_{g,t} = \text{average}_{t \in T}(\text{IG}_{g,t}) - \text{IG}_{g,t}. \quad (3)$$

The score equation is illustrated in Fig. 3, where $\text{Score}_{g,t}$ is defined as the differences between the average $\text{IG}_{g,t}$ and the information gain at time point t' $\text{IG}_{g,t'}$.

2.5. Shifting point detection

Our proposed procedure detects the time at which the GRN shifts from the time-course expression profile. Below, we define the input and describe the steps of the procedure.

INPUT

EXP : a time-course expression profile with T time points

GRN(V, E) : a gene regulatory network

Step 1:

for each gene g that is an innode in GRN

for each time point t of EXP

calculate $\text{IG}_{g,t}$ and $\text{Score}_{g,t}$.

Step 2:

Delete gene g' from the Score matrix,

if $\text{average}_{t \in T}(\text{IG}_{g,t}) + \max_{t \in T}(\text{Score}_{g,t})$ is negative.

Step 3:

for each gene g ,

find the local maximum in $\text{Score}_{g,t}$.

At time point t , the regulation shifts toward gene g .

Fig. 3. Relationship between $\text{IG}_{g,t}$ and $\text{Score}_{g,t'}$.

2.6. Markov process

When analyzing GRN from time-series gene expression profiles, a Markov property is often assumed in the adopted regulatory model of gene expression. A Markov process is a stochastic process in which the conditional probability distribution depends only on the present state. In other words, the expression level of a gene at time t is affected by the expression levels of genes at time $t - 1$. For example, the Markov property is assumed in dynamic Bayesian networks.^{37,38} Our method also operates by a Markov process. In this case, we simply override the definition of the information criterion value with the following notations.

$t_{\text{begin}}, t_{\text{end}}$	The beginning and end time points in EXP
$\text{EXP}_{A,\mathcal{T}}$	EXP excluding a set of time points $t \in \mathcal{T}$
$\text{Join}(\text{EXP}_A, \text{EXP}_B)$	join two profiles EXP_A and EXP_B in time, disregarding the names of the time points

$$\text{IC}_t(g|A) = \text{IC}(\text{EXP}_{M,i}, g|A),$$

$$\text{where } \text{EXP}_M = \text{Join}(\text{EXP}_{\{g\}, t_{\text{begin}}}, \text{EXP}_{A, t_{\text{end}}}). \quad (4)$$

3. Results

To verify the effectiveness of the proposed score-based procedure, we investigated whether the method properly detects the time of diauxic shift from glucose to lactose consumption by *E. coli*.

3.1. Material

3.1.1. Gene expression profile

The test profile was the expression profile reported in Ref. 39 and registered as GSE7265 in the GEO. *E. coli* MG1655 and isogenic mutants were cultured in a medium containing glucose and lactose, and the dynamics of the diauxic shift were observed. Gene expressions were measured in oligonucleotide microarrays containing 70-base oligonucleotide probes for all *E. coli* MG1655 genes. All measurements were performed in triplicate over 17 time points from 780 to 1089 min after inoculation in the wild-type case. At each time point, the triplicated expression values were averaged to obtain a representative expression level. The expression profile is presented in Fig. 6.

As the time of the diauxic shift is not obvious in the gene expression profile, we monitored bacterial growth by the absorbance at 600 nm. However, these data are not included in the GSE7265 entry, so we manually digitalized the growth from the relevant figure in Ref. 39. The nutrient source switched between 878 and 898 min, where the growth became low and negative.

3.1.2. Pathways

The diauxic shift occurs when the nutrient source switches from glucose to lactose, indicating a change in the active metabolic pathways. Therefore, we selected enzymes involved in the metabolic pathways from lactose to pyruvate via α -D-glucose from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.^{40,41} The pathways and enzymes are shown in Fig. 4.

3.1.3. Gene regulatory network

The enzyme expressions are controlled by TFs. The relationships between the genes and TF comprise the GRN. We used TF-gene interactions from RegulonDB 8.7,^{7,8} which cyclopaedically summarizes the gene regulations of *E. coli*. All of the TFs located upstream of the enzymes were used except *gpmA*, whose expression profile has 23 missing values. The GRN included 17 enzymes and 14 TFs with 50 edges that are shown in Fig. 5. In the figure, the green and yellow circles denote enzymes in the glycolysis and lactose metabolism pathways, respectively. The purple and blue circles denote the TFs that regulate the enzymes for glycolysis and lactose metabolism, respectively. The red circle indicates the common TF. The enzymes and their regulations were extracted from the KEGG^{40,41} and RegulonDB^{7,8} databases, respectively.

Fig. 4. Glycolytic and lactose metabolic pathways. The metabolic pathway from lactose to pyruvate is shown. Small circles and boxes indicate the substances and enzymes, respectively. The numbers in the squares are enzyme commission (EC) numbers. The names of the genes encoding the enzymes appear beside the boxes.

Fig. 5. GRN of *E. coli* related to the diauxic shift from glucose to lactose.

Only genes without missing values in the expression profile GSE7265 are shown in the network. The expression profiles of the enzymes and the wild-type TFs are plotted in Fig. 6. The expression ratios of the genes are shown in Fig. 5. The green and yellow lines indicate the enzymes involved in glycolytic and lactose metabolic pathways, respectively. The purple and blue lines indicate the TFs that regulate the enzymes of glycolysis and lactose metabolism, respectively. The red circle is the common TF. The gene expression profile was derived from wild-type data in the GEO^{1,2} GSE7265, and the growth ratio was derived from Ref. 39. The peak in the graph at 869–939 min after inoculation is consistent with the growth ratio extracted from the paper.

3.2. Experiment

This section describes the experimental conditions and presents the results of our method.

3.2.1. Conditions

We used the Akaike Information Criterion as the information criterion in Eq. (1). The proposed procedures were implemented in R (version 3.1.2). We applied the materials to a Markov-type procedure executed on a Mac Pro (mid 2012) with a Mac OS X operating system (ver. 10.9.5), a 3.06-GHz 6-Core Intel Xeon, and 64 GB 1333 MHz DDR ECC. The calculation time was 1.71 s.

Fig. 6. Gene expression profiles and growth ratios during diauxic shift of wild-type *E.coli*.

3.3. Gene selection (Step 2)

Step 2 of the proposed algorithm calculates the average $\text{average}_{t \in T}(\text{IG}_{g,t}) + \max_{t \in T}(\text{Score}_{g,t})$ for each gene. The results are shown in Table 1. Only genes with parent genes in the GRN are shown, because genes with no parent genes are zero-valued. The table reveals 14 genes with positive values and six genes with negative values. The 14 positively valued genes are processed in Step 3.

3.4. Detecting network shifts

Figures 7 and 8 show bar charts of the $\text{Score}_{g,t}$ for the 14 genes selected in Step 2. Figure 7 displays the scores of two enzymes in the lactose metabolic pathway and two TFs in the glycolytic pathway. Figure 8 presents the scores of 10 enzymes in the glycolytic pathway. The growth rate is plotted as a solid black line in Fig. 6. The horizontal axes in Figs. 7 and 8 indicate the time from inoculation of the experiment, and the left and right vertical axes indicate the score and growth ratio, respectively. Step 3 of the proposed method finds the local maximum $\text{Score}_{g,t}$ for each gene. The two peaks at approximately 888 and 1070–1089 min indicate the shift of the GRN.

Table 1. Summary of information gain and score for each gene.

Enzymes in lactose metabolism						
	lacZ	galE	galT	galK	galM	pgm
ave. $IG_{g,t}$	10.80	-3.77	-6.97	-8.96	-8.29	-6.39
max. $Score_{g,t}$	5.28	13.77	5.99	6.15	5.07	2.43
sum	16.08	10.00	-0.98	-2.81	-3.22	-3.95
Enzymes in glycolysis						
	fbaB	pgi	pfkA	gapA	tpiA	pgk
ave. $IG_{g,t}$	13.17	10.23	7.22	7.89	6.02	-0.39
max. $Score_{g,t}$	3.61	2.94	4.88	3.88	3.38	7.86
sum	16.78	13.16	12.10	11.76	9.41	7.47
	glk	fbaA	gpmM	eno	pykF	
ave. $IG_{g,t}$	-1.38	0.91	1.64	1.43	-14.65	
max. $Score_{g,t}$	7.52	4.13	3.15	2.68	4.87	
sum	6.14	5.04	4.79	4.12	-9.77	
TFs of glycolysis						
	fur		soxS		ybiB	
IG	22.85		12.40		-4.88	
max_score	4.21		5.29		4.61	
sum	27.06		17.69		-0.27

Fig. 7. Scores of enzymes in the lactose metabolic pathway and the TFs of the glycolytic pathway. The bar graphs display the scores of the genes that have average $IG_{g,t}$ and $Score_{g,t}$ that sum to positive values. The line graph shows the growth ratio, with the peak indicating the diauxic shift.

Fig. 8. Scores of enzymes in the glycolytic pathway. The bar graphs present the scores of the enzymes that have average $IG_{g,t}$ and $Score_{g,t}$ that sum to positive values. The line graph shows the growth ratio, with the peak indicating the diauxic shift.

3.4.1. Former peak

The former peak agrees with the time of the diauxic shift when the *E.coli* ceases growing. Therefore, this peak detects the time at which the network switched from glucose to lactose metabolism. All the enzymes in the glycolytic pathway peaked at 888 min, implying that glycerol was depleted from the medium and the cells had begun consuming lactose. One of the two enzymes in the lactose metabolic pathway, *lacZ*, peaked at the same time as the glycolytic pathway enzymes; the other, *galE*, peaked 10 min earlier at 908 min. This difference is probably attributable to the position of the enzymes in the lactose metabolic pathway. *lacZ* is the first enzyme in the pathway, whereas *galE* metabolizes a side branch (Fig. 4) later in the pathway.

3.4.2. Latter peak

The latter peak appears at the end of the experiment. This time point was not discussed in the original paper on the gene expression profile.⁴⁰ Instead, the authors trimmed the expression profile and focused on 10 time points from 830 to 939 min. The actual profile GS2765 archived in GEO includes 17 time points from 780 to 1089 min. Furthermore, no information on the growth ratio is provided. Therefore, we can merely surmise that at the end of the experiment, lactose was also depleted from the medium and the *E.coli* cells searched for a new nutrient source. The later peak marked a change in the environment.

4. Discussion

We propose a score-based method that detects the times at which gene regulations change. The method was validated on experimental data for *E. coli*. To more concretely validate the model, we sought examples from other organisms, including eukaryotes and mammals, but achieved minimal success. In the method validation, the GRNs of the test organisms must be known and the exact times of the regulation shifts must be identified in the time-course gene expression profiles. Although gene regulations have been recorded for diverse organisms, *E. coli* is the only organism whose gene regulations are exhaustively available in databases. Pay-to-access databases, such as TransFac^{12,13} and QIAGEN's Ingenuity Pathway Analysis¹⁴ are largely unavailable for verifying bioinformatics methods because the examples are exclusive to third parties. Furthermore, the shift in gene regulations has been identified in very few time-course gene expression profiles. For these reasons, the present study discusses only one result. We expect that more data for verifying the proposed method will become available in the near future, enabling a more thorough assessment of the method's effectiveness. We also expect that our method will help researchers identify transitions in GRNs.

4.1. Analyzing eukaryotes

The proposed method assumes that the expression level of a gene is always regulated by the entire set of direct predecessor genes. This postulate is probably valid for organisms with simple GRNs such as *E. coli*. However, it may not hold for eukaryotes and higher organisms such as humans, whose networks are very complex. In general, only a subset of the direct predecessor genes regulates the target gene at any one time. Therefore, these complexities need to be considered when analyzing the biological networks of higher organisms.

4.2. Future work

The proposed method can detect the time of gene regulation shifts, providing insights into dynamic GRNs. In future work, we will develop methods that directly reveal the changes in GRNs. In particular, we need to reveal the hidden GRNs in Fig. 2. We postulate that our proposed score can meet this objective, but a more complete evaluation of the method will be possible only when adequate examples become available. This is another challenge to be tackled in the future.

5. Conclusion

We proposed a score-based method that detects the time of gene regulation shifts from a gene expression profile and a GRN. The score is computed from the information criterion values, which measure the conformance between gene regulations and gene expression profiles. These values are often used to infer GRNs in a Bayesian

framework. The method was verified via application to the diauxic shift of *E. coli*.³⁹ The input to the method was the gene expression profile GS7265 from GEO^{1,2} and the GRN from RegulonDB.⁸ The time of switching from glucose to lactose metabolism was consistent in our method, which was derived from the growth ratio.³⁹ We conclude that the method with its score will reveal the dynamics of other biological networks.

Availability of supporting data

The procedures were implemented in R and can be downloaded from <https://github.com/takenakayoichi/tacs/>. The software includes the data analyzed in this manuscript.

Acknowledgments

This research was partially supported by JSPS KAKENHI grant numbers 24650155, 22680023, 26280106, and 15K00402. This research used computational resources through the HPCI System Research Project (Project IDs: hp150077 and hp150232).

References

1. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A, NCBI GEO: Archive for functional genomics data sets-update, *Nucl Acids Res* **41**(D1):D991–D995, 2013, doi:10.1093/nar/gks1193.
2. Edgar R, Domrachev M, Lash AE, Gene expression omnibus: NCBI gene expression and hybridization array data repository, *Nucl Acids Res* **30**(1):207–210, 2002, doi:10.1093/nar/30.1.207.
3. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N, Brandizi M, Burdett T, Megy K, Pilicheva E, Rustici G, Tikhonov A, Parkinson H, Petryszak R, Sarkans U, Brazma A, Arrayexpress update—simplifying data submissions, *Nucl Acids Res* **43**(Database issue):D1113–D1116, 2015.
4. Reference Expression Dataset, Accessed on 20 March 2014.
5. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szegedy CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F, Tissue-based map of the human proteome, *Science* **347**(6220), 2015, doi:10.1126/science.1260419.
6. Ringwald M, Eppig JT, Begley DA, Corradi JP, McCright IJ, Hayamizu TF, Hill DP, Kadin JA, Richardson JE, The mouse gene expression database (gxd), *Nucl Acids Res* **29**(1):98–101, 2001, doi:10.1093/nar/29.1.98.
7. Huerta AM, Salgado H, Thieffry D, Collado-Vides J, Regulondb: A database on transcriptional regulation in escherichia coli, *Nucl Acids Res* **26**(1):55–59, 1998, doi:10.1093/nar/26.1.55.
8. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, Garcia-Sotelo JS, Weiss V, Solano-Lira H, Martinez-Flores I, Medina-Rivera A, Salgado-Orsorio G,

- Alquicira-Hernandez S, Alquicira-Hernandez K, Lopez-Fuentes A, Porron-Sotelo L, Huerta AM, Bonavides-Martinez C, Balderas-Martinez YI, Pannier L, Olvera M, Labastida A, Jimenez-Jacinto V, Vega-Alvarado L, del Moral-Chavez V, Hernandez-Alvarez A, Morett E, Collado-Vides J, Regulondb v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more, *Nucl Acids Res* **41**(D1):D203–D213, 2013, doi:10.1093/nar/gks1201.
9. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T *et al.*, An atlas of active enhancers across human cell types and tissues, *Nature* **507**(7493):455–461, 2014.
10. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, Lennartsson A, Rönnerblad M, Hrydziusko O, Vitezic M *et al.*, Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells, *Science* **347**(6225):1010–1014, 2015.
11. Forrest A, Kawaji H, Rehli M, Baillie J, de HM, Haberle V, Lassmann T, Kulakovskiy I, Lizio M, Itoh M, Andersson R, Mungall C, Meehan T, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmidl C, Schaefer U, Medvedeva Y, Plessy C, Vitezic M, Severin J, Semple C, Ishizu Y, Young R, Francescato M, Alam I, Albanese D, Altschuler G, Arakawa T, Archer J, Arner P, Babina M, Rennie S, Balwierz P, Beckhouse A, Pradhan-Bhatt S, Blake J, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Burroughs A, Califano A, Cannistraci C, Carbajo D, Chen Y, Chierici M, Ciani Y, Clevers H, Dalla E, Davis C, Detmar M, Diehl A, Dohi T, Drabløs F, Edge A, Edinger M, Ekwall K, Endoh M, Enomoto H, Fagiolini M, Fairb, A promoter-level mammalian expression atlas, *Nature* **507**:462–470, 2014, doi:10.1038/nature13182.
12. Wingender E, The transfac project as an example of framework technology that supports the analysis of genomic regulation, *Brief Bioinform* **9**(4):326–332, 2008.
13. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F, Transfac: An integrated system for gene expression regulation, *Nucl Acids Res* **28**(1):316–319, 2000.
14. QUIAGEN's Ingenuity Pathway Analysis, Accessed on 19 March 2014.
15. Maeda N, Nishiyori H, Nakamura M, Kawazu C, Murata M, Sano H, Hayashida K, Fukuda S, Tagami M, Hasegawa A *et al.*, Development of a DNA barcode tagging method for monitoring dynamic changes in gene expression by using an ultra high-throughput sequencer, *Biotechniques* **45**(1):95, 2008.
16. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N *et al.*, An atlas of combinatorial transcriptional regulation in mouse and man, *Cell* **140**(5):744–752, 2010.
17. Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ *et al.*, The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line, *Nature Genet* **41**(5):553–562, 2009.
18. Gardner TS, Di Bernardo D, Lorenz D, Collins JJ, Inferring genetic networks and identifying compound mode of action via expression profiling, *Science* **301**(5629):102–105, 2003.
19. Tegner J, Yeung MS, Hasty J, Collins JJ, Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling, *Proc Nat Acad Sci USA* **100**(10):5944–5949, 2003.
20. Bansal M, Della Gatta G, Di Bernardo D, Inference of gene regulatory networks and compound mode of action from time course gene expression profiles, *Bioinformatics* **22**(7):815–822, 2006.

21. Kim S, Kim J, Cho KH, Inferring gene regulatory networks from temporal expression profiles under time-delay and noise, *Comput Bio Chem* **31**(4):239–245, 2007.
22. Schäfer J, Strimmer K, An empirical bayes approach to inferring large-scale gene association networks, *Bioinformatics* **21**(6):754–764, 2005.
23. Stark E, Drori R, Abeles M, Partial cross-correlation analysis resolves ambiguity in the encoding of multiple movement features, *J Neurophysiol* **95**(3):1966–1975, 2006.
24. Butte AJ, Kohane IS, Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements, *Pac Symp Biocomput*, pp. 418–429, 2000.
25. Hausser J, Strimmer K, Entropy inference and the james-stein estimator, with application to nonlinear gene association networks, *J Mach Learn Res* **10**:1469–1484, 2009.
26. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A, Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, *BMC Bioinform* **7**(Suppl 1):S7, 2006.
27. Zoppoli P, Morganello S, Ceccarelli M, Timedelay-aracne: Reverse engineering of gene networks from time-course data by an information theoretic approach, *BMC Bioinform* **11**(1):154, 2010.
28. Neapolitan RE, *Learning Bayesian Networks*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.
29. Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J, d Alche-Buc F, Gene networks inference using dynamic bayesian networks, *Bioinformatics* **19**(Suppl 2):ii138–ii148, 2003.
30. Zou M, Conzen SD, A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data, *Bioinformatics* **21**(1):71–79, 2005.
31. Jia Y, Huan J, Constructing non-stationary dynamic Bayesian networks with a flexible lag choosing mechanism, *BMC Bioinformatics* **11**(Suppl 6):S27, 2010.
32. Nakayama T, Daiyasu H, Seno S, Takenaka Y, Matsuda H, Reconstruction of dynamic gene regulatory networks for cell differentiation by separation of time-course data, in: Arabnia HR, Tran QN (eds.), *Bioinformatics and Computational Biology: The 2013 WorldComp Int Conf Proc*, Mercury Learning and Information, Herndon, VA, USA, pp. 375–379, 2014.
33. Song L, Kolar M, Xing EP, Time-varying dynamic bayesian networks, *Advances in Neural Information Processing Systems*, pp. 1732–1740, 2009.
34. Tamada Y, Araki H, Imoto S, Nagasaki M, Doi A, Nakanishi Y, Tomiyasu Y, Yasuda K, Dunmore B, Sanders D, Humphreys S, Print CG, Charnock-Jones SD, Tashiro K, Kuhara S, Miyano S, Unraveling dynamic activities of autocrine pathways that control drug-response transcriptome networks, Altman RB, Dunker AK, Hunter L, Murray T, Klein TE (eds.), *Pacific Symp Biocomputing*, pp. 251–263, 2009.
35. Dean T, Kanazawa K, A model for reasoning about persistence and causation, *Comput Intell* **5**(3):142–150, 1989, doi:10.1111/j.1467-8640.1989.tb00324.x.
36. Jensen FV, Nielsen TD, *Bayesian Networks and Decision Graphs*, 2nd Edition, Springer Publishing Company, Incorporated, NYC, NY, USA, 2007.
37. Friedman N, Murphy K, Russell S, Learning the structure of dynamic probabilistic networks, *Proc Fourteenth Conf Uncertainty in Artificial Intelligence UAI'98*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 139–147, 1998, Available at <http://dl.acm.org/citation.cfm?id=2074094.2074111>.
38. Koller D, Friedman N, *Probabilistic Graphical Models: Principles and Techniques—Adaptive Computation and Machine Learning*, The MIT Press, Cambridge, MA, USA, 2009.

39. Traxler MF, Chang DE, Conway T, Guanosine 3',5'-bispyrophosphate coordinates global gene expression during glucose-lactose diauxie in escherichia coli, *Proc Nat Acad Sci USA* **103**(7):2374–2379, 2006.
40. Kanehisa M, Goto S, Kegg: Kyoto encyclopedia of genes and genomes, *Nucl Acids Res* **28**(1):27–30, 2000, doi:10.1093/nar/28.1.27.
41. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M, Data, information, knowledge and principle: Back to metabolism in kegg, *Nucl Acids Res* **42**(D1):D199–D205, 2014, doi:10.1093/nar/gkt1076.

Yoichi Takenaka received his M.Eng. and Ph.D. degrees in Computer Science from Osaka University, Japan, in 1997 and 2000, respectively. He was an Assistant Professor from 2000 to 2002 in Osaka University. He has been an Associate Professor in the Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, since 2002. His research interests include gene expression analysis, high-throughput sequence analysis, and legal informatics.

Shigeto Seno is an Assistant Professor of the Graduate School of Information Science and Technology, Osaka University. He received his B.E., M.E., and Ph.D. degrees from Osaka University in 2001, 2003, and 2006, respectively. He is a member of IPSJ, JSBi, and MBSJ.

Hideo Matsuda received his B.Sc. degree in Physics from Kobe University, Japan, in 1982, and he received his M.Eng. and Ph.D. degrees in Computer Science from Kobe University, Japan, in 1984 and 1987, respectively. He has been a Professor in the Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, since 2002. His research interests include genomic data analysis, gene regulatory networks, and gene expression analysis.