

RESEARCH

Open Access



Automated transition analysis of activated gene regulation during diauxic nutrient shift in *Escherichia coli* and adipocyte differentiation in mouse cells

Yoichi Takenaka^{1,2,3*}, Kazuma Mikami⁴, Shigeto Seno² and Hideo Matsuda²

From The Sixteenth Asia Pacific Bioinformatics Conference
Yokohama, Japan. 15-17 January 2018

Abstract

Background: Comprehensively understanding the dynamics of biological systems is among the biggest current challenges in biology and medicine. To acquire this understanding, researchers have measured the time-series expression profiles of cell lines of various organisms. Biological technologies have also drastically improved, providing a huge amount of information with support from bioinformatics and systems biology. However, the transitions between the activation and inactivation of gene regulations, at the temporal resolution of single time points, are difficult to extract from time-course gene expression profiles.

Results: Our proposed method reports the activation period of each gene regulation from gene expression profiles and a gene regulatory network. The correctness and effectiveness of the method were validated by analyzing the diauxic shift from glucose to lactose in *Escherichia coli*. The method completely detected the three periods of the shift; 1) consumption of glucose as nutrient source, 2) the period of seeking another nutrient source and 3) consumption of lactose as nutrient source. We then applied the method to mouse adipocyte differentiation data. Cell differentiation into adipocytes is known to involve two waves of the gene regulation cascade, and sub-waves are predicted. From the gene expression profiles of the cell differentiation process from ES to adipose cells (62 time points), our method acquired four periods; three periods covering the two known waves of the cascade, and a final period of gene regulations when the differentiation to adipocytes was completed.

Conclusions: Our proposed method identifies the transitions of gene regulations from time-series gene expression profiles. Dynamic analyses are essential for deep understanding of biological systems and for identifying the causes of the onset of diseases such as diabetes and osteoporosis. The proposed method can greatly contribute to the progress of biology and medicine.

Keywords: Gene regulatory network, Network dynamics, Time course, Cell differentiation, Adipocyte

*Correspondence: takenaka@kansai-u.ac.jp

¹Faculty of Informatics, Kansai University, Ryosenji 2-1-1, Takatsuki, Osaka, Japan

²Graduate School of Information Science and Technology, Osaka University, Yamadaoka 1-5, Suita, Osaka, Japan

Full list of author information is available at the end of the article

Background

Acquiring a comprehensive understanding of biological systems dynamics is among the biggest challenges in biology and medicine. The processes of all organisms (including humans) are time-variant. Cells cycle and divide, change their internal states and differentiate into other cell types when stimulated from the outside. Abnormal cell differentiation causes diseases and cancer. Dynamic systems are universal in organisms.

Biological systems dynamics can now be understood through advanced biological technologies. On-chip hybridization and DNA sequencing has enabled the simultaneous measurement of the expression levels of many genes. Many of the accumulated data are registered in public databases such as GEO (Gene Expression Omnibus) [1, 2], Array Express [3] and Sequence Read Archive [4]. Special-purpose databases are also accessible. The references of mammalian tissue gene expressions are stored in RefEx (Reference Expression Dataset) [5], whereas the data from human tissues, cell lines and cancers are distributed through the Human Protein Atlas [6]. Mouse data are archived in the specialized database GXD (Gene expression database) of the Mouse Genome Informatics resource [7]. These databases are freely accessible to all users.

The above databases store the time-course movements of the gene expression levels in multiple organisms. For example, a search for “*time-course*” (2017/7/12) returned 913 datasets in GEO and 7623 entries in SRA. Based on these data, researchers have revealed the biological mechanisms of processes such as cell cycling [8] and diauxic shifts [9]. The adipogeneses of mice and humans have also been compared using time-course profiles [10]. These studies cover a broad range, from biology, through medicine, to pharmacology.

Numerous systems operate in living cells. Examples are metabolism, genetic and environmental information processing, cellular processes, and organismal systems [11]. Concurrent functioning of these systems in vivo maintains the life activities of the cell. Systems have been extensively researched and summarized as pathways in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [11]. Control of gene expression levels is another life-sustaining system. Gene regulation defines the relation between a controlling agent and its target gene. The gene regulatory system can be represented as a network of nodes (genes) connected by directed edges (regulation of a downstream gene by an upstream gene). The gene regulatory networks of model organisms are available in databases such as RegulonDB [12, 13], which contains the nearly-complete gene regulation of *Escherichia coli*, and the yeast network databases YEASTRACT [14–17]. The FANTOM projects have attempted to exhaustively reveal the relations between mouse and human in many tissues

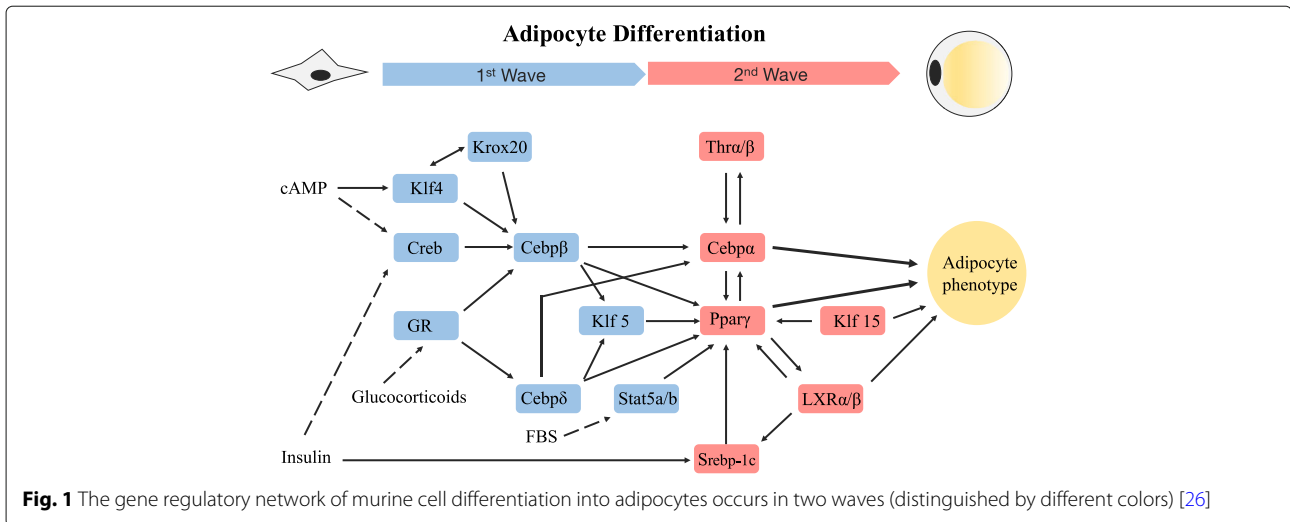
[18–22]. Gene regulatory networks are also provided in commercial databases such as TransFac [23, 24] and Quiagen’s Ingenuity Pathway Analysis [25]. The regulatory relations between genes, at least in the model organisms, will be revealed exhaustively in the near future. However, although databases provide numerous gene regulations, the relationships are basically static. Each database entry supplies a gene regulatory relation and the experimental environment in which the relation was activated. Very few entries include the chronological order of the activations, despite our increasing knowledge of signaling pathway cascades. When the chronological order is given, the experimental environment includes the cell cycles and diauxic shift of the nutrient source. Adipocyte differentiation is accompanied by a rough sketch of the transition of active expression controls [26] (see Fig. 1). The differentiation involves a single transition point and two waves of activated gene regulations, although more than two waves have been suggested [26]. The dynamics of gene-expression control is a new research topic, with many unknowns at present.

Gene regulation activity can be dynamically analyzed from gene expression profiles. Given time-course data, researchers have sought the transformations in the gene regulatory network through models, such as ordinary differential equations and graphical models. Ordinary differential equations can explain the activity level of a gene regulation. The explanatory and objective variables are the expression levels of the upstream genes and the downstream genes, respectively [27–29]. Graphical models describe the control relations in the gene expression as a network of nodes (genes) with directed edges (controls). The dynamics of biological systems can be analyzed by a graphical Gaussian model with a Markov property [30, 31], which measures the independence of any two genes using partial correlations.

Other graphical models are the gene relevance network and the Bayesian network. The former generates a gene regulatory network from the mutual information, and measures the degree of association between gene expression levels [32–34]. This model can also manage time-course gene expression profiles [35].

In the Bayesian network, a directed edge represents the probabilistic relationship between two genes by information-theoretic approaches. The plausibility of the network is measured by the Bayesian information criterion and the minimum description length [36, 37]. Bayesian networks have been extended to time-course data by introducing Markov properties into the joint probability distribution. The time-dependent Bayesian network is called a dynamic Bayesian network [38, 39].

The above-described methods have analyzed not only the gene expression profiles, but also general data. Researchers have obtained meaningful results from all



data except time-course gene expression profiles, which are most severely limited by the small number of time points. Figure 2 shows the distribution of time points in the GEO database. There are 913 datasets with time-series expression profiles. The average, mean and standard deviation of the time points are 4.66, 4, and 2.82, respectively. The number of time points is far below the numbers of other common data. To clarify this fact, Table 1 lists the number of data points in benchmark problems for inferring causal relationship networks [37, 40–44]. All of these benchmarks are included in bnlearn of the R package [44]. Comparing Fig. 2 and Table 1, we observe that the number of time points is two orders of magnitude lower in the time-series expression profile than in the benchmarks. This data limitation not only prevents the estimation of correct relationships, but also stifles the dynamic analyses.

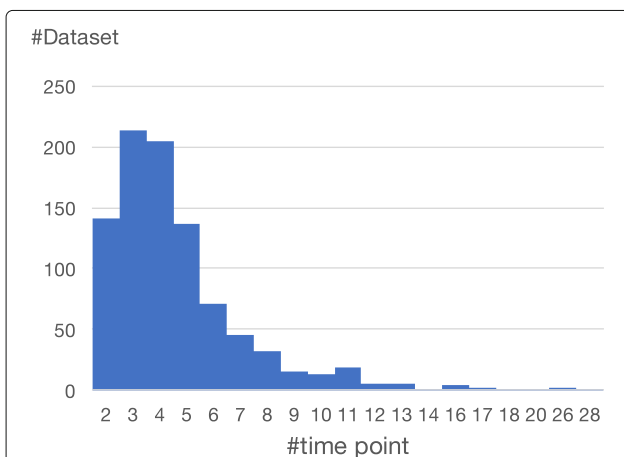
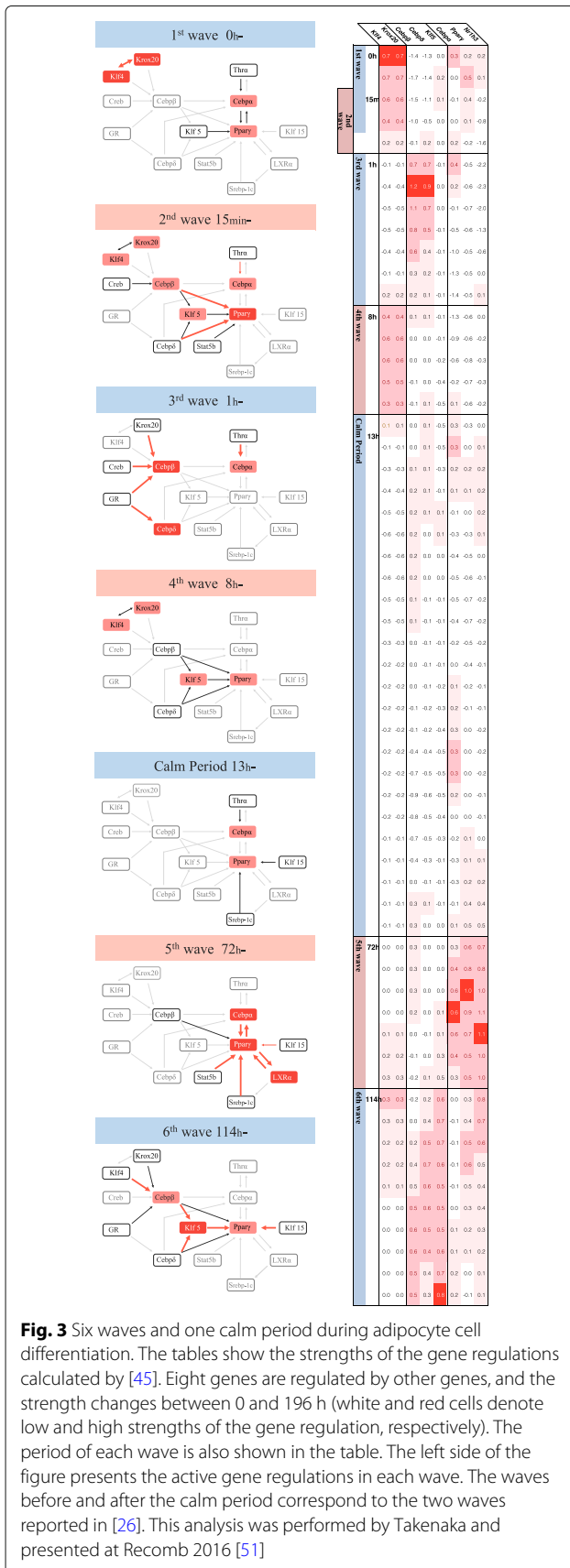


Fig. 2 Number of time points in datasets returned by a “time course or time-course” query in the Gene Expression Omnibus. The average, mean and standard deviation of the distribution are 4.66, 4 and 2.82, respectively

Takenaka et al. [45] proposed a method that captures the dynamic changes in gene expression control from time-course expression profiles with a small number of time points. Their method quantifies the intensity at which genes control the expression levels of other genes. The temporal resolution of intensity is one time point (the observation unit of the time-course gene expression profile). Single time-point resolution is finer than the resolutions of other methods [46–49]. Assuming a profile with N time points, the method generates N sub-profiles, each omitting one time point. The effectiveness of the method was confirmed in the diauxic shift of *E.coli* [9] and the adipocyte cell differentiation in mouse tissue [45, 50, 51]. The results detected the time of the diauxic shift in *E. coli*, and two waves of active regulations during adipocyte cell differentiation (Fig. 3). The authors concluded that the method effectively reveals the dynamics of gene regulations. But the method does not have any procedure how to determine the number of breakpoints and when the breaks happen. Therefore, finding the breakpoint of active gene expression control by this approach is somewhat subjective, and is undeniably influenced by already known biological findings. In this research, we propose a mechanical method that eliminates this subjectivity.

Table 1 Numbers of samples and variables in benchmark problems for inferring causal relationship networks

| Benchmark | #Samples | #Variables |
|-----------------|----------|------------|
| Learning.test | 5000 | 6 |
| Cigaussian.test | 5000 | 8 |
| Asia | 5000 | 8 |
| Alarm | 20000 | 37 |
| Insurance | 20000 | 27 |
| Hailfinder | 20000 | 56 |



Method

We propose a method that analyzes the dynamics of gene regulations from time-series gene expression profiles. Here, the dynamics of gene regulations take two meanings. The first interpretation considers the dynamics of each gene regulation; whether the control of gene expression level by other genes is activated or inactivated at each time point in the time-series data. The second considers the dynamics of the gene regulatory network; the periods and pairs of their start and end times. The periods (sometimes called states, waves, or phases) are often biologically meaningful. The most well-known period is the cell cycle, which has two states; interphase and cell division. The interphase is divided into three sub-states called phases (Gap1, Synthesis, and Gap2), each with sub-phases. The cell cycle has four phases in total. Under these dynamics, the number of periods in a cell cycle is not fixed. The same situation occurs in other biological processes. Therefore, determining the number of periods in a biological process is a difficult task, and the dynamics of the gene regulatory network are difficult to analyze, especially in a manual analysis.

Notations

To analyze the dynamics of gene regulations in the two interpretations by an automated approach, we formulate the analysis as an optimization problem. The notations of the formulation are listed below.

- Exp** : time-course gene expression profiles with $|T|$ time points and $|G|$ genes.
- T** : a set of time points in *Exp*. If $i < j$, then $t_i \in T$ is earlier than $t_j \in T$.
- G** : a set of genes in *Exp*.
- M** : a two-dimensional matrix $|G| \times |T|$, where M_{ij} represents the intensity at which the expression of gene $g_i \in G$ is controlled by others at time $t_j \in T$. *M* is calculated from *Exp* by the method proposed in [45].

In our method, the activity of gene regulation is represented by a binary number, with 1 and 0 denoting active and inactive, respectively. The notations of the gene-regulation dynamics are explained below.

- a_g** : a binary vector of length $|T|$ representing the activity of gene $g \in G$ at time $t_j \in T$.
- Act** : a binary matrix of size $|G| \times |T|$, where Act_i equals a_{g_i} .
- Tra** : a subset of $|T|$ representing the breaks between periods. For example, if $Tra = \{t_3\}$, the system has two periods; the first ending at t_3 and the second starting at t_4 . If *Tra* is an empty set, no transition of the gene regulation occurs throughout the period.

Score function

As previously mentioned, we analyze the dynamics of gene regulations by solving the optimization problem; that is, by finding the most plausible dynamic model of the gene regulations. This model is conditional on the activity of each gene and the model complexity. Regarding the first condition, a positive M_{ij} indicates high likelihood of activating the gene g_i at time point t_j . The second conditional expression describes the penalty term of the Bayesian information criterion. Here, the complexity of the model is defined as the number of periods in the model, which deals with the trade-off between the goodness of fit to the model and the tractability of the model.

To find a plausible model for M , we must optimize the following score:

$$\text{Score}(Act) = \sum_{i=1}^{|G|} \sum_{j=1}^{|T|} M_{ij} \cdot Act_{ij} - |Tra| \cdot \ln(|G| \cdot |T|), \quad (1)$$

where Tra is calculated from Act . A time point $t_i \in T$ is a member of Tra if $\exists j Act_{ij} \neq Act_{i(j+1)}$.

Algorithm

The algorithm that finds the most plausible dynamic model of gene regulation is described below.

Input a Matrix M calculated from the gene expression profile and a gene regulatory network by the method in [45].

Output a Matrix Act , a set Tra , and a $score$.

Measure Maximize the score function $\text{Score}(Act)$.

Variables

k : The integer k represents the number of breaks between the periods in the model.

V : A set of binary vectors of length $|T|$. Each element of V represents the activity of a gene regulation.

step 1 Initialize $k \leftarrow 0$ and $Tra \leftarrow \{\}$

step 2 Generate a set of vectors V by the following sub-steps.

- (1) add a zero vector to V
- (2) add a one vector to V

step 3 $score_0 \leftarrow 0$

step 4 for each $g_i \in G$, select a vector $v \in V$ that maximizes the g_i score in the following sub-steps.

- (1) select a vector v in V that maximizes $M_i \cdot v^t$, where M_i is the i -th row of M .
- (2) $score_0 += M_i \cdot v^t$
- (3) insert vector v into the i -th row of Act ;
 $Act_i \leftarrow v$.

step 5 $score_{best} \leftarrow score_0$; $Act_{best} \leftarrow Act$; $Tra_{best} \leftarrow Tra$

step 6 $k += 1$; If $k = |T|$, goto **Step 9**.

step 7 for each $Tra = \{t | t \in T\}$, where $|Tra| = k$,

step 7-1 generate a set of vectors V as follows:

- (1) for each period i of Tra , generate a set V_i with two elements; a zero vector and a unit vector with lengths equal to that of the period.
- (2) generate V as the direct product of all V_i .

step 7-2 $score_k \leftarrow -2 \cdot k \cdot \ln(|T| \cdot |G|)$

step 7-3 for each $g_i \in G$, select a vector $v \in V$ that maximizes the score by the following sub-steps.

- (1) select a vector v in V that maximizes $M_i \cdot v^t$, where M_i is the i -th row of M .
- (2) $score_k += M_i \cdot v^t$
- (3) insert vector v into the i -th column of Act ;
 $Act_i \leftarrow v$.

step 7-4 if $score_k > score_{best}$, $score_{best} \leftarrow score_k$;
 $Act_{best} \leftarrow Act$; $Tra_{best} \leftarrow Tra$.

step 8 if $score_k > score_{k-1}$, goto **Step 6**.

step 9 output Act_{best} , Tra_{best} and $score_{best}$.

The above algorithm is a brute-force search with poor computational efficiency. However, as the gene expression profile contains very few time points, the above algorithm satisfies our requirements.

For larger datasets, the calculation time can be shortened by using a hash function from a gene and a period to the sub-score. Meanwhile, the computational complexity can be reduced by dynamic programming.

Results and discussion

The effectiveness of the proposed method was tested on two biological datasets. The results of our method were also compared with a manual analysis of the periods in the time series. The first dataset contains the time-course gene expression profiles during the diauxic shift from glucose to lactose metabolism in *E. coli*. In this test, the method was assessed by its ability to detect the time of the diauxic shift.

The second dataset reports the adipocyte cell differentiation in the house mouse *Mus musculus*. Two periods (called waves in [26]) of gene regulation have been identified during this differentiation, and more waves are expected. This test evaluated whether the method can detect three or more plausible periods.

Diauxic shift

Material

The first dataset is the time-course gene expression profile GSE7265 in the GEO dataset. The diauxic shift was observed in *E. coli* MG1655 and isogenetic mutants

cultured in a medium containing glucose and lactose [9]. The expression levels of the genes were measured in oligonucleotide microarrays containing 70-base oligonucleotide probes. The wild-type profile contains 17 time points from 780 min. to 1089 min. after inoculation. At each time point, the triplicated expression values were averaged to obtain the representative expression level.

The time points are divided into two phases; the first lasting from 780 min. to 939 min., the second from 969 min. to 1089 min. The diauxic shift from glucose to lactose has been recognized in the first half, but information on the second half is completely lacking. Although the GS2765 gene expression profile contains 17 time points, the original paper [9] trimmed the profile to 10 time points (830-939 min. post-inoculation), and reported the growth ratio in the first half only. Therefore, the full dynamics of the diauxic shift can only be surmised.

The gene regulatory network of the diauxic shift comprises 31 genes and 50 gene regulations [45]. Fourteen enzymes are related to the glycolytic and lactose metabolic pathways. According to RegulonDB [12, 13], the expression levels of the enzymes are controlled by 14 transcription factors (TFs), and 50 TF–enzyme interactions are known. The present study adopts the gene list and the gene regulatory network used in [45].

The proposed method was implemented in R, and the matrix *M* representing the intensities of the gene regulations was calculated by an R package from GitHub <https://github.com/takenakayoichi/tacs>. The computed matrix *M* was input to the proposed method.

Results

The proposed method detected five periods during the diauxic shift of wild type *E. coli*. The maximum score during each period is tabulated in Table 2. Although the proposed algorithm identified 6 periods, the maximum score was recorded over 8 periods for reference. The score

Table 2 Maximum scores in each period of diauxic shift in *E. coli* (central column) and differentiation to murine adipocytes (right column) (maximum scores are in bold)

| #Period | Diauxic shift | Adipocyte |
|---------|---------------|--------------|
| 1 | 16.42 | 1.50E-13 |
| 2 | 23.57 | 50.38 |
| 3 | 26.23 | 70.30 |
| 4 | 28.00 | 74.76 |
| 5 | 28.44 | 73.41 |
| 6 | 26.81 | 72.02 |
| 7 | 25.02 | 69.25 |
| 8 | 23.14 | 67.87 |

peaks during Period 5 and the between-period breakpoints are 878 min., 908 min., 999 min. and 1049 min. The first, second and third periods belong to the first half of the diauxic shift; the remainder belong to the second half. The periods are also indicated by the vertical lines in Fig. 4.

Takenaka et al. [45] manually detected three periods in the GS2765 gene expression profile, with breakpoints at approximately 888 – 908 min. and 1070 – 1089 min. The breakpoints were expressed as ranges because they were difficult to pinpoint by the manual approach. In their analysis, the first and second periods belonged to the first half of the diauxic shift, and the third period belonged to the second half.

First half

Our method divided the first half of the experiment into three periods. Figure 4 shows the growth ratios in each period. During the first and third periods, the *E. coli* exist in the logarithmic growth phase and are consuming glucose and lactose, respectively. The second period is consistent with cessation of cell proliferation.

In comparison, the manual curation identified only two periods in the first half. The division point was exactly centered in the growth arrest period. From the difference between our results and the manually curated results, we can identify whether a mechanism that regulates the gene expression level exists independently in the logarithmic and growth-arrest phases.

Which of these methods expresses more realistic biological properties? Manual curation suggests that lactose consumption begins immediately after glucose depletion. On the other hand, the proposed method suggests that once the glucose is depleted, the cells cease growing while seeking a new nutrient source, and resume growth when an alternative nutrient is found. We consider that the proposed method presents a more plausible dynamic model than manual curation. One expects that the *E. coli* cells cannot immediately access a new nutrition source once the original source is depleted. If this were possible, logarithmic growth would continue without interruption.

Second half

The second half of the experiment was continuous in the manual curation, but divided into two periods in the proposed method. Once the *E. coli* had consumed the glucose and lactose in their nutrient medium, they began seeking the next nutrition source. When no further source was found, the gene expression regulations changed sufficiently for detection by both methods. However, as the expression profile provides the only available information in the second half, the actual number of periods in the second half of the experiment is difficult to judge.

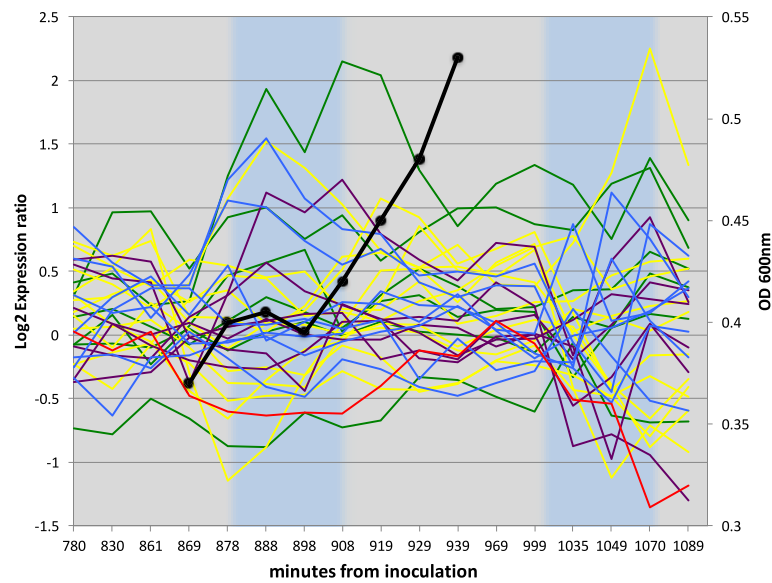


Fig. 4 Expression levels of 31 genes (colored plots) and growth ratio (solid black line) during diauxic shift in *E. coli*. Vertical lines delineate the periods of the dynamic model detected by the proposed method

Adipocyte differentiation

Materials

The second dataset was a time-course gene expression profile of mouse adipocyte differentiation. The RNAs in this dataset were collected from mouse ST2 bone marrow stroma cell-derived stem cells (RCB0224) obtained from the RIKEN BioResource Center (BRC, Tsukuba, Japan). ST2 cells were induced by changing the medium from RPMI1640 to DMEM supplemented with 10% FBS, 0.5 mM 3-isobutyl-1-methylxanthine (MIX), 0.25 μM DEX,

and with insulin-transferrin-selenium-X supplement containing 5 μg/ml of insulin and 1 μM rosiglitazone. After 48 hours, the differentiation medium was replaced with DMEM supplemented with 10% FBS. The RNAs collected at 62 time points during adipocyte differentiation were measured using an Affymetrix GeneChip Mouse Genome 430 2.0 Array. The time points were 0, 5, 15, 30 and 45 min., hourly from 1 to 30 h, and 6-hourly from 36 to 192 h after adipogenesis induction. Each datum was background-subtracted and normalized in the robust

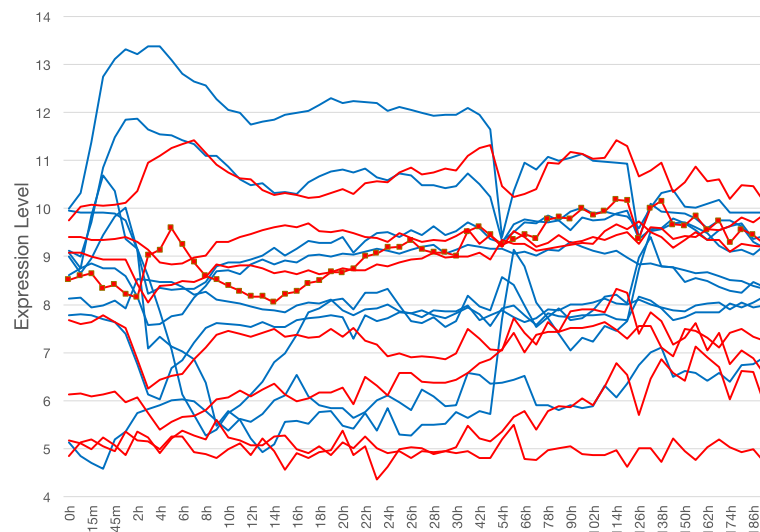


Fig. 5 Expression Levels of 14 genes during murine adipocyte differentiation. Blue and red lines indicate the expression levels of genes belonging to the first and second waves in Fig. 1, respectively. The red line with square markers plots the expression levels of pparg, a marker gene of adipose cells

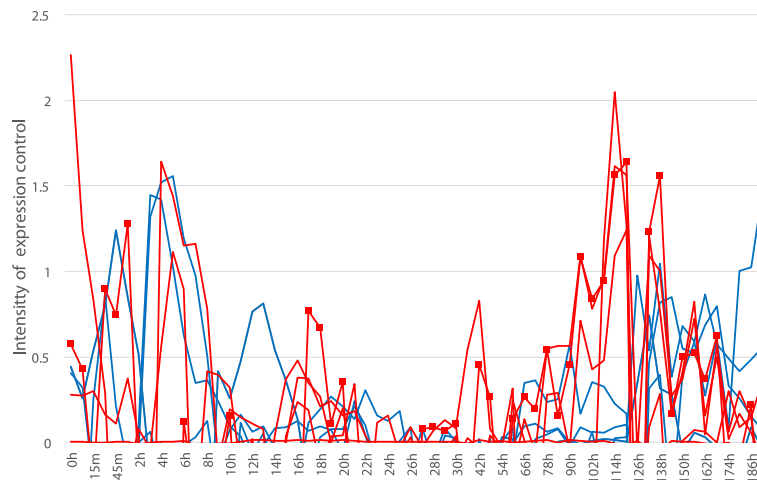


Fig. 6 Intensity of genes whose expression levels are controlled by other genes. Blue and red lines show the intensities of genes belonging to the first and second waves in Fig. 1, respectively. The red line with square markers plots the intensity of pparγ, a marker gene of adipose cells

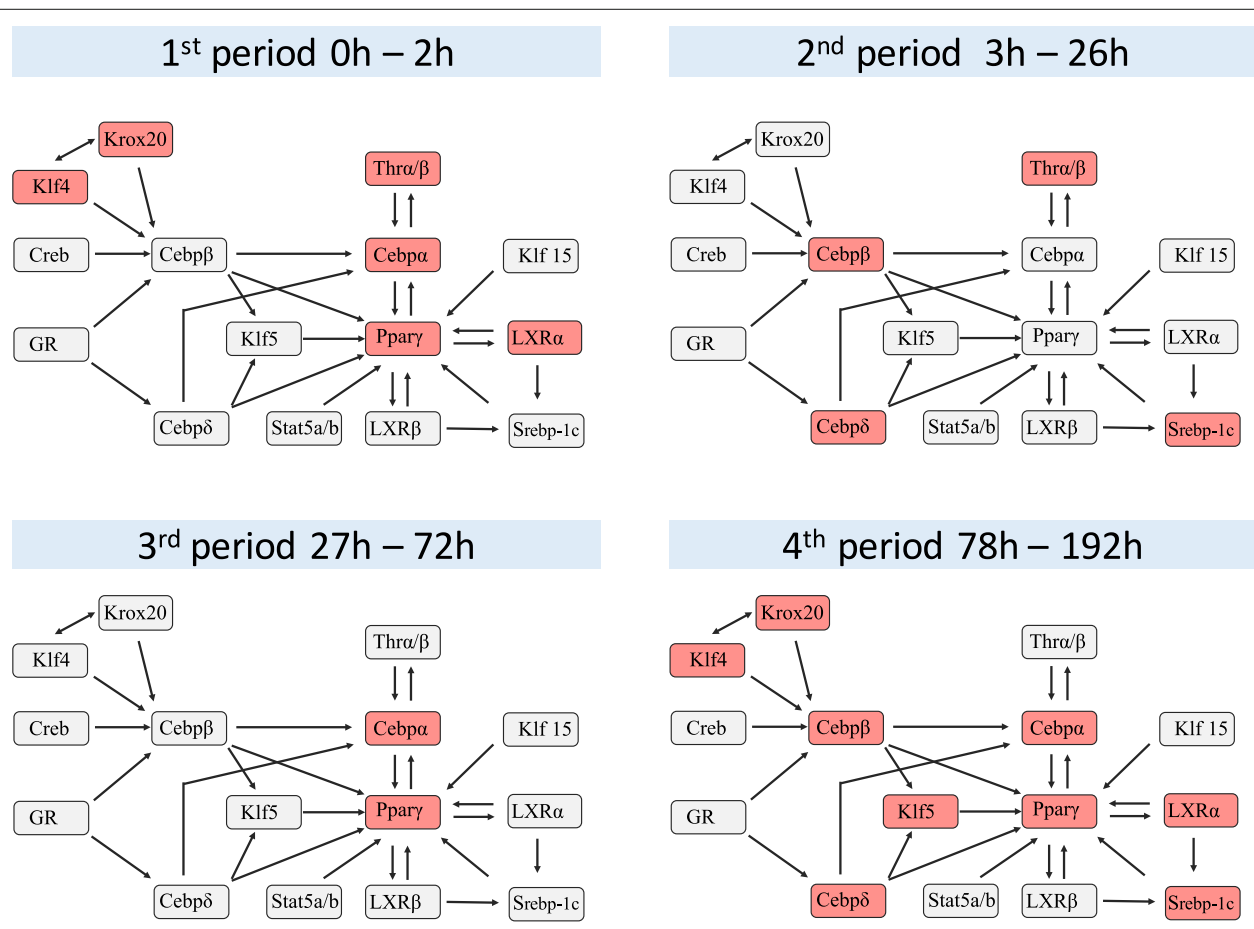


Fig. 7 The four periods of differentiation into murine adipocytes detected by the proposed method. Dark- and light-shaded boxes enclose genes whose expression level is controlled and not controlled by upstream genes, respectively

multi-array analysis (RMA). The gene expression profiles are available from the Genome Network Platform [50].

Previous research has unveiled the important regulators of adipocyte development and a two-wave cascade in this process [26] (see Fig. 1). The network includes 24 regulations between 14 genes, but undetected sub-waves have been predicted [26].

The expression levels of the 14 genes are shown in Fig. 5. Among them, the line with square indicates *pparg*, a known marker of adipocyte cells. The expression level of *pparg* increases from 2 to 6 h and then decreases. The decline is followed by a gradual rise after 14 h.

Figure 6 shows the intensities of the expression levels controlled by other genes, which constitute the intensity matrix M of our algorithm. The intensities were calculated from the gene expression profiles in Fig. 5 and the gene regulatory network in Fig. 1.

Result

The proposed method detected four periods of adipocyte differentiation. The maximum score in each period is listed in Table 2. The algorithm terminated after five periods, but 8 periods are included for referencing purposes. The score peaks during period 4, and the between-period breakpoints are 2 h, 26 h and 72 h. Figure 7 shows the dynamics of the gene regulatory network. In the figure, we classify the gene whose expression level during a period is controlled by upstream genes or not by the average of the intensity during the period is larger than zero or not.

The two waves of the known cascade in Fig. 1 appear in the four periods detected by our algorithm. During the first period of adipocyte differentiation, the expression controls of the most upstream genes (*Klf4* and *Knox20*) are activated. During the second period, the expressions of *Cebp β* and *Cebp δ* are controlled and the regulations of the second wave (*Thra β* and *Srebp-1c*) begin activating. The *pparg* and *cebp α* genes are regulated during the third period, and the most downstream genes are regulated during the fourth period. The four-period differentiation identified by our method includes the known waves, but appears to be more precise.

The four periods detected by our method differ from the two-wave differentiation in two ways. First, the regulations of *Pparg* and *Cebp α* are activated twice. Once at the first period, and re-activated in the second wave composed of third and fourth periods. *Pparg* is a marker of adipose cells and a known regulator of adipocyte differentiation. The two activations of the *pparg* gene are confirmed in Fig. 5. The expression level of *Pparg* peaks at 6 h and 114 h. Most of the gene expression regulations are activated in the fourth period. By the start time of this period (78 h), the cells will have produced lipid droplets [52]. Therefore, we consider that the differentiation into

adipocytes was completed during the fourth period, and was stabilized thereafter. It appears that many regulations of the adipocyte-related gene expression levels are activated in the fourth period.

Conclusions

We proposed a method that detects the dynamics of gene regulatory networks at the temporal resolution of single time points in the underlying gene expression profiles. The dynamics are modeled as periods of activated regulations. The plausibility of the model was quantified using the Bayesian information criterion. The problem constitutes a combinatorial optimization problem that find the highest-scoring model. The algorithm inputs the gene expression profiles and a gene regulatory network, and returns the detected regulations, divided into periods.

The effectiveness of the method was validated in two datasets; the diauxic shift from glucose to lactose in *E. coli* and adipocyte differentiation in the mouse. In both datasets, the proposed method detected more plausible dynamic models than existing models. We believe that the proposed method can precisely reveal the dynamics of biological systems.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers 15K00402 and 16H03261.

Funding

Funding for the publication of this article was provided by JSPS KAKENHI Grant Number 15K00402.

Availability of data and materials

The *E.coli* data analyzed in this research was available as GSE7265 Gene Data Set in Gene Expression Omnibus. The mouse data analyzed in this research was downloaded from Genome Network Platform [50]. The code used in this study is available in <https://github.com/takenakayoichi/tacs>.

About this supplement

This article has been published as part of *BMC Bioinformatics* Volume 19 Supplement 4, 2018: Selected articles from the 16th Asia Pacific Bioinformatics Conference (APBC 2018): bioinformatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-4>.

Authors' contributions

YT conceived and designed the algorithm and the experiments and wrote the manuscript. YT and KM implemented the algorithm and performed the analyses. SS and HM contributed by providing suggestions. All authors have read and approve the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publishers Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Faculty of Informatics, Kansai University, Ryousenji 2-1-1, Takatsuki, Osaka, Japan. ²Graduate School of Information Science and Technology, Osaka University, Yamadaoka 1-5, Suita, Osaka, Japan. ³Graduate School of Medicine, Osaka University, Yamadaoka 2, Suita, Osaka, Japan. ⁴Recruit Holdings Co. Ltd., Marunouchi 1-9-2, Chiyoda, Tokyo, Japan.

Published: 8 May 2018

References

- Edgar R, Domrachev M, Lash AE. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–10. <https://doi.org/10.1093/nar/30.1.207>.
- 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. *Nucleic Acids Res.* 2013;41(D1):991–5. <https://doi.org/10.1093/nar/gks1193>.
- 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. *Nucleic Acids Res.* 2015;43(Database issue):1113–6.
- Kodama Y, Shumway M, Leinonen R. The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Res.* 2012;40(D1):2011–3. <https://doi.org/10.1093/nar/gkr854>.
- Reference Expression Dataset. <http://refex.dbcls.jp>. Accessed 13 July 2017.
- 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, Szgyarto CA-K, 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.* 2015;347:6220. <https://doi.org/10.1126/science.1260419>.
- Ringwald M, Eppig JT, Begley DA, Corradi JP, McCright IJ, Hayamizu TF, Hill DP, Kadin JA, Richardson JE. The mouse gene expression database (gxd). *Nucleic Acids Res.* 2001;29(1):98–101. <https://doi.org/10.1093/nar/29.1.98>.
- Bähler J. Cell-Cycle Control of Gene Expression in Budding and Fission Yeast. *Annu Rev Genet.* 2005;39(1):69–94. <https://doi.org/10.1146/annurev.genet.39.110304.095808>.
- Traxler MF, Chang DE, Conway T. *Proc Natl Acad Sci of the United States of America.* 2006;103(7):2374–9.
- Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, Rosen ED. Comparative epigenomic analysis of murine and human adipogenesis. *Cell.* 2010;143(1):156–69. <https://doi.org/10.1016/j.cell.2010.09.006>.
- Kanehisa M, Goto S. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Huerta AM, Salgado H, Thieffry D, Collado-Vides J. Regulondb: A database on transcriptional regulation in escherichia coli. *Nucleic Acids Res.* 1998;26(1):55–9. <https://doi.org/10.1093/nar/26.1.55>.
- 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-Osorio G, Alquicira-Hernandez S, Alquicira-Hernandez K, Lopez-Fuentes A, Porrón-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. *Nucleic Acids Res.* 2013;41(D1):203–13. <https://doi.org/10.1093/nar/gks1201>.
- Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sá-Correia I. The yeasttract database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae. *Nucleic Acids Res.* 2006;34(suppl 1):446. <https://doi.org/10.1093/nar/gkj013>.
- Monteiro PT, Mendes ND, Teixeira MC, d'Orey S, Tenreiro S, Mira NP, Pais H, Francisco AP, Carvalho AM, Lourenço AB, Sá-Correia I, Oliveira AL, Freitas AT. Yeasttract-discoverer: new tools to improve the analysis of transcriptional regulatory associations in saccharomyces cerevisiae. *Nucleic Acids Res.* 2008;36(suppl 1):132. <https://doi.org/10.1093/nar/gkm976>.
- Abdulrehman D, Monteiro PT, Teixeira MC, Mira NP, Lourenço AB, dos Santos SC, Cabrito TR, Francisco AP, Madeira SC, Aires RS, Oliveira AL, Sá-Correia I, Freitas AT. Yeasttract: providing a programmatic access to curated transcriptional regulatory associations in saccharomyces cerevisiae through a web services interface. *Nucleic Acids Res.* 2011;39(suppl 1):136. <https://doi.org/10.1093/nar/gkq964>.
- Teixeira MC, Monteiro PT, Guerreiro JF, Gonçalves JP, Mira NP, dos Santos SC, Cabrito TR, Palma M, Costa C, Francisco AP, Madeira SC, Oliveira AL, Freitas AT, Sá-Correia I. The yeasttract database: an upgraded information system for the analysis of gene and genomic transcription regulation in saccharomyces cerevisiae. *Nucleic Acids Res.* 2014;42(D1):161. <https://doi.org/10.1093/nar/gkt1015>.
- 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.* 2010;140(5):744–52.
- 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. *Nat Genet.* 2009;41(5):553–62.
- Forrest A, Kawaji H, Rehli M, Baillie J, de HM, Haberer 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, Schmid 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, Drablos F, Edge A, Edinger M, Ekwall K, Endoh M, Enomoto H, Fagiolini M, Fairb. A promoter-level mammalian expression atlas. *Nature.* 2014;507:462. <https://doi.org/10.1038/nature13182>.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmid C, Suzuki T, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455–61.
- Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Rönnblad M, Hrydzusko O, Vitezic M, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science.* 2015;347(6225):1010–4.
- 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. *Nucleic Acids Res.* 2000;28(1):316–9.
- Wingender E. The transfac project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform.* 2008;9(4):326–32.
- QUIAGEN's Ingenuity Pathway Analysis. <https://www.qiagenbioinformatics.com/>. Accessed 14 July 2017.
- Siersbæk R, Nielsen R, Mandrup S. Transcriptional networks and chromatin remodeling controlling adipogenesis. *Trends Endocrinol Metab.* 2012;23(2):56–64.
- Tegner J, Yeung MS, Hasty J, Collins JJ. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc Natl Acad Sci.* 2003;100(10):5944–9.
- Gardner TS, Di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science.* 2003;301(5629):102–5.
- Kim S, Kim J, Cho KH. Inferring gene regulatory networks from temporal expression profiles under time-delay and noise. *Comput Biol Chem.* 2007;31(4):239–45.
- Schäfer J, Strimmer K. An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics.* 2005;21(6):754–64.
- Stark E, Drori R, Abeles M. Partial cross-correlation analysis resolves ambiguity in the encoding of multiple movement features. *J Neurophys.* 2006;95(3):1966–75.
- Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In: *Pac Symp Biocomput.* Singapore: World Scientific Pub Co Inc.; 2000. p. 418–29.

33. Hausser J, Strimmer K. Entropy inference and the James–Stein estimator, with application to nonlinear gene association networks. *J Mach Learn Res.* 2009;10:1469–84.
34. 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 Bioinformatics.* 2006;7(Suppl 1):7.
35. Zoppoli P, Morganella S, Ceccarelli M. TimeDelay-Aracne: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics.* 2010;11(1):154.
36. Jensen FV, Nielsen TD. *Bayesian Networks and Decision Graphs*, 2nd edn. New York: Springer; 2007.
37. Neapolitan RE. *Learning Bayesian Networks*. Upper Saddle River: Prentice-Hall, Inc.; 2003.
38. Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alche-Buc F. Gene networks inference using dynamic Bayesian networks. *Bioinformatics.* 2003;19(suppl 2):138–48.
39. Zou M, Conzen SD. A new dynamic Bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics.* 2005;21(1):71–9.
40. Agresti A. *Categorical Data Analysis*, 3rd edn. Wiley Series in Probability and Statistics. 111 River Street: Wiley-Interscience; 2012.
41. Lauritzen SL, Spiegelhalter DJ. Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). *J R Stat Soc Ser B Stat Methodol.* 1988;50(2):157–224.
42. Beinlich IA, Suermondt HJ, Chavez RM, Cooper GF. The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks. In: *Proceedings of the 2nd European Conference on Artificial Intelligence in Medicine*. Berlin: Springer-Verlag; 1989. p. 247–56.
43. Binder J, Koller D, Russell S, Kanazawa K. Adaptive Probabilistic Networks with Hidden Variables. *Mach Learn.* 1997;29(2–3):213–44.
44. Scutari M. Learning Bayesian Networks with the bnlearn R Package. *J Stat Softw.* 2010;35(3):1–22. Accessed 2 Aug 2017.
45. Takenaka Y, Seno S, Matsuda H. Detecting shifts in gene regulatory networks during time-course experiments at single-time-point temporal resolution. *J Bioinforma Comput Biol.* 2015;13(05):1543002. <https://doi.org/10.1142/S0219720015430027>.
46. Jia Y, Huan J. Constructing non-stationary dynamic Bayesian networks with a flexible lag choosing mechanism. *BMC Bioinformatics.* 2010;11(Suppl 6):27.
47. Nakayama T, Daiyasu H, Seno S, Takenaka Y, Matsuda H. *Bioinformatics and Computational Biology: The 2013 WorldComp International Conference Proceedings*. In: Arabnia HR, Tran Q-N, editors. Herndon: Mercury Learning and Information; 2014. p. 375–9.
48. Song L, Kolar M, Xing EP. Time-varying dynamic Bayesian networks. In: *Advances in Neural Information Processing Systems*. NY: Curran Associates, Inc.; 2009. p. 1732–40.
49. 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. In: Altman RB, Dunker AK, Hunter L, Murray T, Klein TE, editors. *Pacific Symposium on Biocomputing*. Singapore: World Scientific Pub Co Inc.; 2009. p. 251–63.
50. Adipocyte Differentiation of Mouse, Genome Network Platform. ST2 stem cell was induced DMEM with 10% FBS, Affymetrix GeneChip 430 2.0 Array, 62 time points. http://genomenetwork.nig.ac.jp/index_e.html. Accessed 2 Aug 2017.
51. Takenaka Y, Seno S, Matsuda H. Dynamic Analysis of Gene Regulations 734 using Leaving-One-Out Expression Profile. *Proceeding of The 20th Annual International Conference on Research in Computational Molecular Biology, RECOMB2016*, P48, 2016.
52. Qian SW, Li X, Zhang YY, Huang HY, Liu Y, Sun X, Tang QQ. Characterization of adipocyte differentiation from human mesenchymal stem cells in bone marrow. *BMC Dev Biol.* 2010;10:47. <https://doi.org/10.1186/1471-213X-10-47>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

