

# Identifying regulational alterations in gene regulatory networks by state space representation of vector autoregressive models and variational annealing

Kaname Kojima<sup>1</sup>, Seiya Imoto<sup>1\*</sup>, Rui Yamaguchi<sup>1</sup>, André Fujita<sup>2</sup>, Mai Yamauchi<sup>1</sup>, Noriko Gotoh<sup>1</sup>, Satoru Miyano<sup>1</sup>

From The Tenth Asia Pacific Bioinformatics Conference (APBC 2012)  
Melbourne, Australia. 17-19 January 2012

## Abstract

**Background:** In the analysis of effects by cell treatment such as drug dosing, identifying changes on gene network structures between normal and treated cells is a key task. A possible way for identifying the changes is to compare structures of networks estimated from data on normal and treated cells separately. However, this approach usually fails to estimate accurate gene networks due to the limited length of time series data and measurement noise. Thus, approaches that identify changes on regulations by using time series data on both conditions in an efficient manner are demanded.

**Methods:** We propose a new statistical approach that is based on the state space representation of the vector autoregressive model and estimates gene networks on two different conditions in order to identify changes on regulations between the conditions. In the mathematical model of our approach, hidden binary variables are newly introduced to indicate the presence of regulations on each condition. The use of the hidden binary variables enables an efficient data usage; data on both conditions are used for commonly existing regulations, while for condition specific regulations corresponding data are only applied. Also, the similarity of networks on two conditions is automatically considered from the design of the potential function for the hidden binary variables. For the estimation of the hidden binary variables, we derive a new variational annealing method that searches the configuration of the binary variables maximizing the marginal likelihood.

**Results:** For the performance evaluation, we use time series data from two topologically similar synthetic networks, and confirm that our proposed approach estimates commonly existing regulations as well as changes on regulations with higher coverage and precision than other existing approaches in almost all the experimental settings. For a real data application, our proposed approach is applied to time series data from normal Human lung cells and Human lung cells treated by stimulating EGF-receptors and dosing an anticancer drug termed Gefitinib. In the treated lung cells, a cancer cell condition is simulated by the stimulation of EGF-receptors, but the effect would be counteracted due to the selective inhibition of EGF-receptors by Gefitinib. However, gene expression profiles are actually different between the conditions, and the genes related to the identified changes are considered as possible off-targets of Gefitinib.

**Conclusions:** From the synthetically generated time series data, our proposed approach can identify changes on regulations more accurately than existing methods. By applying the proposed approach to the time series data on

\* Correspondence: imoto@ims.u-tokyo.ac.jp

<sup>1</sup>Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan  
Full list of author information is available at the end of the article

normal and treated Human lung cells, candidates of off-target genes of Gefitinib are found. According to the published clinical information, one of the genes can be related to a factor of interstitial pneumonia, which is known as a side effect of Gefitinib.

## Background

Gene network estimation from time series gene expression data is a key task for elucidating cellular systems. Thus far, wide variety of approaches have been proposed based on the vector autoregressive (VAR) model [1,3], the state space model [4-6], and the dynamic Bayesian network [7,8]. Recently, time series gene expression data on multiple conditions aiming at analyzing effects of cell treatment such as drug dosing and heat shock are available. We here assume that some gene regulations are disrupted but many of the gene regulations do not change due to some treatment of interest, and try to find a small number of changes on regulations as keys for elucidating effects by the treatment.

A possible way for finding changes on regulations is to estimate networks from two data sets separately and then compare their structures. However, due to the limited length of time series data (usually less than 10 time points) and unignorable measurement noise, networks are estimated with high error rates and the estimation errors cause the serious failure on identifying changes on regulations. Thus, approaches using two time series data in an efficient manner are strongly demanded. Also, widely used statistical methods such as the VAR model and dynamic Bayesian network assume equally spaced time points in time series data. However, observed time points on usually available time series

data are not equally spaced [5,6,9], and approaches that can handle unequally spaced time series data in a theoretically correct way should be considered.

We propose a new statistical model that estimates gene networks on two different conditions in order to identify changes on regulations between the conditions. As the basis of the proposed model, we employ the state space representation for VAR model (VAR-SSM), in which observation noise is considered between the measured or observed gene expressions and the true gene expressions in observation model and gene regulations between true gene expressions are considered in the system model [10]. The VAR-SSM can handle unequally spaced time series data by ignoring observation model on the non-observed time points. For considering the changes on regulations, we introduce hidden variables to the VAR-SSM in order to indicate the presence of regulations in each condition. If hidden binary variables on two conditions indicating the presence of a regulation are both estimated as one, the regulation is considered as a commonly existing regulation. On the other hand, if only one of the hidden binary variables for the regulation is estimated as one, the regulation is considered as a condition specific regulation. We also introduce a potential function between the hidden binary variables that is designed to take high probability if the hidden binary variables on two conditions take the same value. From the design of the potential function, the similarity of networks on two conditions is automatically considered. Since the time series data on both conditions are used for estimating commonly existing regulation due to the use of the hidden binary variables, an efficient data assignment is achieved. In addition, from the more accurate estimation of commonly existing regulations by the efficient data assignment, accurate identification of changes on regulations is induced.

The hidden binary variables are estimated by searching the configuration of binary variables that maximizes the marginal likelihood of the model. However, searching the optimal configuration is computationally intractable. Thus, as an alternative approach, we derive a new variational annealing method based on [11] in order to estimate the hidden binary variables. We also give a proof for the effectiveness of the variational annealing compared to other candidate alternatives, the variational annealing and the EM algorithm, in order to show the validity of using the variational annealing.

**Table 1 Comparison of the variation annealing (Proposed) and EM algorithm (EM) based on the proposed model**

(a)						
# of time points	50			25		
	# TP	# FP	PRE	# TP	# FP	PRE
Proposed	295.9	41.7	0.88	238.4	71.6	0.77
EM	294.9	119.2	0.71	196.8	66.9	0.75

(b)						
# of time points	50			25		
	# TP	# FP	PRE	# TP	# FP	PRE
Proposed	39.8	13.2	0.75	23.4	20.8	0.53
EM	39.9	39.4	0.5	11.5	10.0	0.53

(a) The number of true positives (# TP) and false positives (# FP) of estimated regulations in two network models by the proposed approach and EM for equally spaced time series data. PRE denotes the precision of the results. Regulations in two networks are 305 in total. (b) The number of true positives (# TP) and false positives (# FP) of changes on regulations between two network models estimated by the proposed approach and EM for equally spaced time series data. The regulations changed in two networks are in total 47.

For the performance evaluation, we generate two regulatory networks in such a way that most of the regulations commonly exist and some exist only on one of the networks. We then apply our proposed approach and existing var model based and dynamic Bayesian network based approaches to two equally spaced time series data drawn separately from the generated networks. From the comparisons of true positive rates and false positive rates of these approaches, we confirm the effectiveness of our approach. We also generate unequally spaced time series data from these networks, and show that our approach works correctly on unequally spaced time series data while the performance of the existing approaches assuming equally spaced time points is drastically worsened.

Our proposed approach is used to analyze changes on regulations in gene networks between normal Human lung cells and Human lung cells treated by stimulating EGF-receptors and dosing an anticancer drug termed Gefitinib. A lung cancer condition is simulated by the stimulation of EGF-receptors in the treated cells. Since Gefitinib is known as a selective inhibitor of EGF-receptors, the stimulation of EGF-receptors would be counteracted by Gefitinib, and hence the treated cells are expected to be the same condition as normal cells. However, gene expression profiles from normal and treated cells are actually different, and off-targets of Gefitinib causing unexpected positive or negative effects are implied. We focus on genes with changes on regulations between the networks estimated by our approach and find possible off-target genes of Gefitinib. According to the published clinical information, one of the possible off-target genes is suggested as one of factors of interstitial pneumonia, which is known as a side effect of Gefitinib.

## Methods

### Vector autoregressive model and its state space representation

#### Vector autoregressive model

Given gene expression profile vectors of  $p$  genes during  $T$  time points  $\{y_1, \dots, y_T\}$ , the first order vector autoregressive (VAR(1)) model at time point  $t$  is given by

$$y_t = Ay_{t-1} + \varepsilon_t,$$

where  $A$  is a  $p \times p$  autoregressive coefficient matrix, and  $\varepsilon_t$  is observation noise at time  $t$  and follows  $\mathcal{N}(0, \text{diag}[\sigma_1^2, \dots, \sigma_p^2])$ , a normal distribution with mean 0 and variance  $\text{diag}[\sigma_1^2, \dots, \sigma_p^2]$ . The  $(i, j)$ th element of  $A$ ,  $A_{ij}$ , indicates a temporal regulation from the  $j$ th gene to the  $i$ th gene, and if  $A_{ij} \neq 0$ , regulation from the  $j$ th gene to the  $i$ th gene is considered. By examining whether  $A_{ij}$  is zero or not for all  $i$  and  $j$ , a gene network is

constructed. Since equally spaced time points are assumed in the VAR model, it has difficulty on handling unequally spaced time series data.

### State space representation of VAR model (VAR-SSM)

Let  $\mathcal{T}$  be the set of equally spaced entire  $T$  time points and  $\mathcal{T}_{obs}$  the set of time points where gene expressions are observed. Note that  $\mathcal{T}_{obs} \subseteq \mathcal{T}$  holds. VAR-SSM is comprised of two models: system model and observation model. Let  $x_t$  be hidden variable vector representing true gene expression at time  $t$ . The system model is given as the VAR model of  $x_t$ :

$$x_t = Ax_{t-1} + \eta_t, \quad t \in \mathcal{T},$$

where  $\eta_t$  is the system noise normally distributed with mean 0 and variance  $H = \text{diag}[h_1, \dots, h_p]$ . The observation model represents measurement error of observed gene expression  $y_t$  and true gene expression  $x_t$  at observed time point  $t \in \mathcal{T}_{obs}$ :

$$y_t = x_t + \rho_t, \quad t \in \mathcal{T}_{obs},$$

where  $\rho_t$  is the observation noise normally distributed with mean 0 and variance  $R = \text{diag}[r_1, \dots, r_p]$ . Unequally spaced time series data are handled by ignoring observation model at non-observed time points.

### Joint model of VAR-SSM for two time series data

Let  $\{y_t^{(c)}\}_{t \in \mathcal{T}_{obs}^{(c)}}$  be time series gene expression data on cell condition  $c$ , where  $\mathcal{T}_{obs}^{(c)}$  is the set of observed time points on cell condition  $c$ . We also let  $\mathcal{T}^{(c)}$  be the set of time points from 1 to  $T^{(c)}$ , where  $T^{(c)} = \max\{t \in \mathcal{T}_{obs}^{(c)}\}$ . Given time series data on two types of cell conditions  $c = 1$  and 2, we propose a new VAR-SSM model to estimate gene networks in the two conditions as well as identify changes on regulations between them. The model is comprised of the following two equations:

$$x_t^{(c)} = A \circ E^{(c)} x_{t-1}^{(c)} + \eta_t^{(c)}, \quad t \in \mathcal{T}^{(c)},$$

$$y_t^{(c)} = x_t^{(c)} + \rho_t^{(c)}, \quad t \in \mathcal{T}_{obs}^{(c)},$$

where  $\circ$  denotes the Hadamard product,  $E^{(c)}$  is a  $p \times p$  binary matrix, and  $\eta_t^{(c)}$  and  $\rho_t^{(c)}$  are respectively system and observation noises from  $\mathcal{N}(0, H)$  and  $\mathcal{N}(0, R)$ . In this model, the  $(i, j)$ th element  $E_{ij}^{(c)}$  takes one if regulation from gene  $j$  to gene  $i$  exists on condition  $c$  and zero otherwise, i.e., the presence of regulations is controlled by  $E^{(c)}$  and the AR coefficient matrix  $A$  is commonly used in conditions 1 and 2. Changes on regulations are identified when regulations exist only in a condition.

The complete likelihood of our model,  $P(Y, X, \Theta, E)$ , where  $Y, X, \Theta$ , and  $E$  are respectively the sets of  $y_t^{(c)}, x_t^{(c)}$ , parameters, and  $E^{(c)}$ , is given by the following equation:

$$P(Y, X, \Theta, E) = \prod_{c=1}^2 \prod_{i \in T^{(c)}} \frac{|H|^{-1/2}}{\sqrt{2\pi^r}} \exp \left\{ -\frac{1}{2} (x_i^{(c)} - A \circ E^{(c)} x_{i-1}^{(c)})' H^{-1} (x_i^{(c)} - A \circ E^{(c)} x_{i-1}^{(c)}) \right\} \\ \times \prod_{i \in T_{\text{obs}}^{(c)}} \frac{|R|^{-1/2}}{\sqrt{2\pi^p}} \exp \left\{ -\frac{1}{2} (y_i^{(c)} - x_i^{(c)})' R^{-1} (y_i^{(c)} - x_i^{(c)}) \right\} P(\Theta, E),$$

where the prior distribution  $P(\Theta, E)$  is assumed to be factorized as

$$P(\Theta, E) = \prod_i P(h_i) P(r_i) \prod_j P(A_{ij}) P(E_{ij}^{(1)}, E_{ij}^{(2)}, z_{ij}).$$

Here,  $z_{ij}$  is a parameter for a potential function of  $E_{ij}^{(1)}$  and  $E_{ij}^{(2)}$  defined later. The prior distributions of  $A_{ij}$  is given by

$$P(A_{ij} | h_i, E_{ij}^{(1)}, E_{ij}^{(2)}) = \mathcal{N}(A_{ij}; 0, h_i \cdot \alpha_1)^{F_{ij}} \mathcal{N}(A_{ij}; 0, h_i \cdot \alpha_0)^{1-F_{ij}},$$

where  $\alpha_0$  and  $\alpha_1$  are parameters controlling the shrinkage of coefficients  $A$  and  $F_{ij}$  is a binary variable that takes 1 if  $E_{ij}^{(1)}$  or  $E_{ij}^{(2)}$  takes 1 and 0 otherwise, i.e.,  $F_{ij}$  is given by  $1 - \prod_{c=1}^2 (1 - E_{ij}^{(c)})$ .  $\alpha_1$  is set to a large value, while  $\alpha_0$  is set to smaller than  $\alpha_1$ . From the design of the prior for  $A_{ij}$ , if  $E_{ij}^{(1)}$  or  $E_{ij}^{(2)}$  takes one, i.e., there exists regulation from gene  $j$  to  $i$  in condition 1 or 2, weaker prior  $\mathcal{N}(A_{ij}; 0, h_i \cdot \alpha_1)$  is selected and the shrinkage of the coefficients are avoided. Otherwise stronger prior  $\mathcal{N}(A_{ij}; 0, h_i \cdot \alpha_0)$  is selected and the sparsity of the network structures is promoted. The prior distributions of  $h_i$ , and  $r_i$  are given by

$$P(h_i) = \mathcal{IG}(h_i; u_0, k_0), \\ P(r_i) = \mathcal{IG}(r_i; v_0, l_0),$$

where  $\mathcal{IG}$  represents the density function of inverse gamma distribution,  $u_0$  and  $v_0$  are the shape parameters, and  $k_0$  and  $l_0$  are the inverse scaling parameters. Under the assumption that a small number of regulations change between two conditions, we design the prior distribution for  $E_{ij}^{(1)}$  and  $E_{ij}^{(2)}$ ,  $P(E_{ij}^{(1)}, E_{ij}^{(2)}, z_{ij})$  by using the following potential function between  $E_{ij}^{(c)}$  for  $c = 1$  or 2:

$$\frac{1}{2} \phi(E_{ij}^{(1)}, E_{ij}^{(2)}; z_{ij}) = \begin{cases} z_{ij} & \text{if } E_{ij}^{(1)} \neq E_{ij}^{(2)} \\ 1 - z_{ij} & \text{otherwise} \end{cases}.$$

In this setting, if  $z_{ij}$  is small,  $E_{ij}^{(1)}$  and  $E_{ij}^{(2)}$  tend to take the same value and thus most of the regulations exist in both two conditions. We also introduce a prior distribution for  $z_{ij}$  by beta distribution with parameters  $\zeta_{i0}$  and  $\zeta_{i1}$ :

$$\mathcal{B}(z_{ij}; \zeta_{i0}, \zeta_{i1}) = \frac{1}{B(\zeta_{i0}, \zeta_{i1})} z_{ij}^{\zeta_{i0}-1} (1 - z_{ij})^{\zeta_{i1}-1},$$

where  $B(\cdot)$  is the beta function. Thus, the prior distribution of  $E_{ij}^{(c)}$  is given by

$$P(E_{ij}^{(1)}, E_{ij}^{(2)}, z_{ij}) = \phi(E_{ij}^{(1)}, E_{ij}^{(2)}) \mathcal{B}(z_{ij}; \zeta_{i0}, \zeta_{i1}).$$

Figure 1 shows a graphical representation of the proposed model, where dependency of the parameters and variables are indicated. The hyperparameters are omitted and the observed data  $y_t$  is represented by gray nodes. In the observed data  $y_t^{(1)}$  and  $y_t^{(2)}$  are propagated mainly via hidden variables  $E_{ij}^{(1)}$  and  $E_{ij}^{(2)}$ . Due to the data propagation, more accurate estimation is expected in the proposed model than the approaches considering data on two conditions independently.

For the parameter estimation, we search the configuration of  $E$  maximizing the following marginal likelihood:

$$\hat{E} = \arg \max_E \int dX \int d\Theta P(Y, X, \Theta, E). \quad (1)$$

Finding the optimal configuration of  $E$  is computationally intractable, and heuristics approaches such as the EM algorithm and the variational method are used in practice. Here, we use the variational annealing, an extension of the deterministic annealing for discrete variables [11]. In the next section, we give a small explanation of the variational annealing and show its effectiveness compared to the EM algorithm and the variational method.

### Parameter estimation by variational annealing

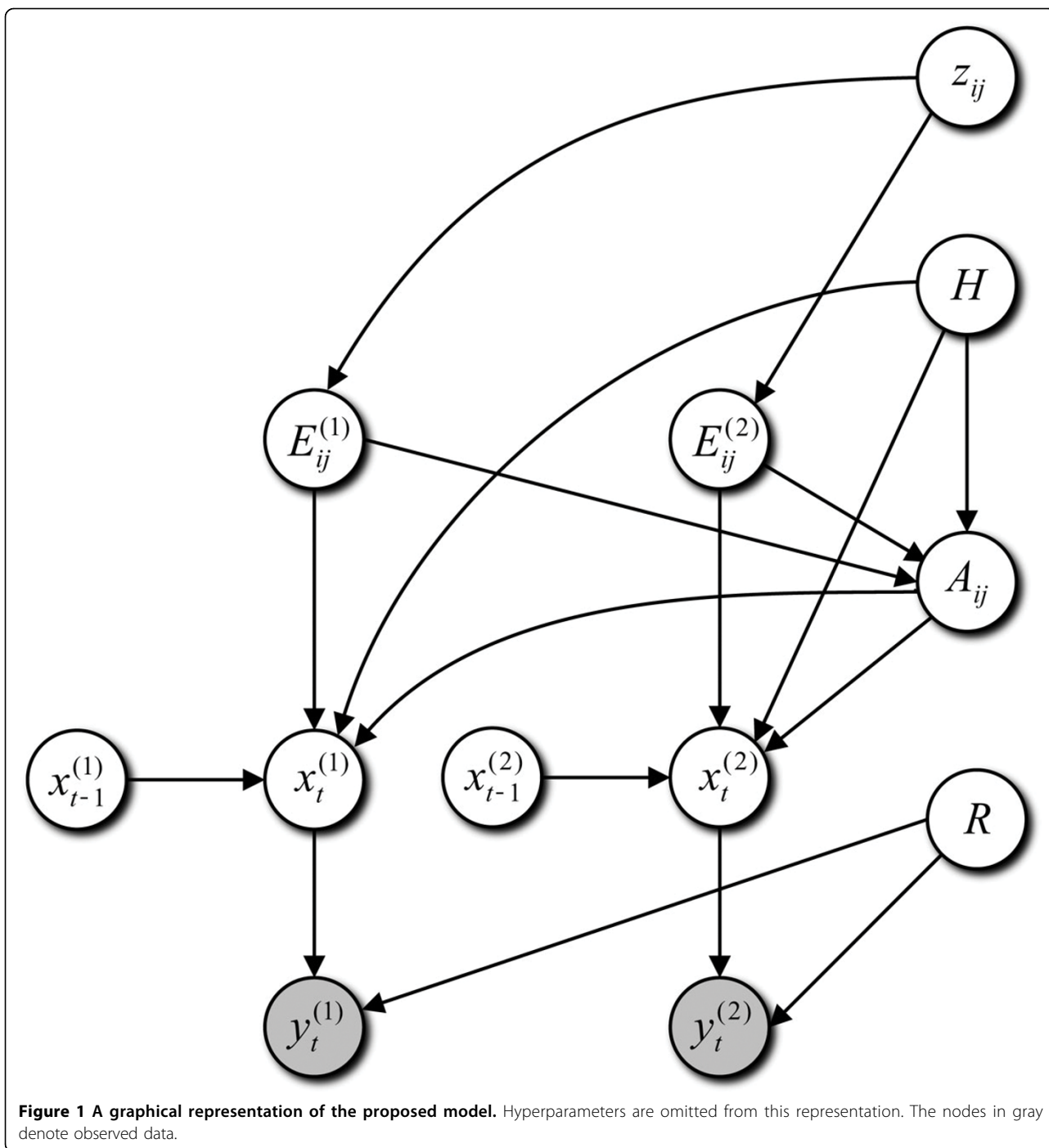
In the deterministic annealing, optimization problem is solved while gradually changing temperature in a some schedule, and maximum likelihood estimator is obtained like the EM algorithm [12-14]. Yoshida and West proposed to use the deterministic annealing to find the configuration of the binary variables that maximizes the likelihood of factor models with sparseness priors [11]. We derive a new variational annealing method by extending Yoshida and West's approach to find the configuration of the binary variables on marginal likelihood function, which can be applied for searching  $E$  that maximizes Equation (1).

Let  $E$ ,  $X$ , and  $\Theta$  be  $p$  dimensional binary variables, unobserved variables, and parameters, respectively, and consider to search  $E$  maximizing the following marginal likelihood:

$$\max_{E \in \{0,1\}^p} \log \int dX \int d\Theta P(X, \Theta, E). \quad (2)$$

The maximum of the marginal likelihood on  $E$  is bounded by the following formula:

$$\max_{E \in \{0,1\}^p} \log \int dX \int d\Theta P(X, \Theta, E) \geq \tau \log \int_{\{0,1\}^p} dE \left[ \int dX \int d\Theta P(X, \Theta, E) \right]^{1/\tau}. \quad (3)$$



where  $\tau$  is called temperature and the equality holds for  $\tau \rightarrow +0$ . Hereafter, the integral range of  $E$  is omitted if no confusion occurs. Let  $Q(E)$  be a normalized non-negative function, i.e.,  $Q(E) \geq 0$  and  $\int Q(E)dE = 1$ .

From the Gibbs inequality, the right side of Equation (3) is also bounded:

$$\tau \log \int dE \left[ \int dX \int d\Theta P(X, \Theta, E) \right]^{1/\tau} \geq \int dE Q(E) \log \frac{\int dX \int d\Theta P(X, \Theta, E)}{Q(E)^\tau}.$$

Here,  $Q(E)$  is considered as an approximation function of  $\frac{P(E)^{1/\tau}}{\int P(E')^{1/\tau} dE'}$ . Under the assumption that  $P(X, \Theta, E) \propto Q(X)Q(\Theta)$ , where  $Q(X)$  and  $Q(\Theta)$  are normalized non-negative functions, we have the following inequality

$$\int dE Q(E) \log \frac{\int dX \int d\Theta P(X, \Theta, E)}{Q(E)^\tau} \geq \int dE \int dX \int d\Theta Q(X) Q(\Theta) Q(E) \log \frac{P(X, \Theta, E)}{Q(X) Q(\Theta) Q(E)^\tau}. \quad (4)$$

Thus, as an approximation of  $E$  maximizing Equation (2), we try to find  $\hat{E} = \arg \max_E Q(E)$ , where  $Q(E)$  is the function maximizing the lower bound in Equation (4). Since higher values on  $P(E)$  are weighed more in the approximation by  $Q(E)$  for  $\tau < 1$ , the better approximation is expected for the higher values. This property on the limited case is shown in Proposition 1. As is in the variational method and the EM algorithm, the maximum of the lower bound in Equation (4) is searched by a hill climbing from high temperature  $\tau > 1$ .

In the hill climbing,  $Q(X)$ ,  $Q(\Theta)$ , and  $Q(E)$  are alternately updated from the following equations:

$$\begin{aligned} Q(E) &\propto \exp \left( \frac{1}{\tau} \int dX \int d\Theta Q(X) Q(\Theta) \log P(E|X, \Theta) \right), \\ Q(X) &\propto \exp \left( \int dE \int d\Theta Q(E) Q(\Theta) \log P(X|\Theta, E) \right), \\ Q(\Theta) &\propto \exp \left( \int dE \int dX Q(E) Q(X) \log P(\Theta|X, E) \right). \end{aligned}$$

Gradually converging  $\tau$  to 0, local optimum of the lower bound and corresponding  $Q(E)$ ,  $Q(X)$ , and  $Q(\Theta)$  are obtained.

#### Effectiveness of variational annealing

As alternatives of the variational annealing, we may consider the variational method and the EM algorithm where  $X$  is the set of hidden variables and  $E$  is handled as the set of parameters to be maximized. We show the effectiveness of variational annealing compared to the variational method and the EM algorithm under the following conditions:  $P(X, \Theta, E)$  is factorized into  $P(X, E)P(\Theta, E)$ , and  $P(E_i|X, \Theta, E \setminus \{E_i\})$  is given as a binomial distribution, where  $E_i$  is the  $i$ th element of  $E$ . If factorization of  $P(X, \Theta, E) = Q(X)Q(\Theta)Q(E)$  is assumed in the calculation of the variational method,  $\arg \max_E Q(E)$  is not the optimal solution of Equation (2) in general. In the EM algorithm, by allowing  $E$  to move around a  $p$ -dimensional continuous space  $(0, 1)^p$ ,  $\hat{E}_{EM} = \arg \max_{E \in (0, 1)^p} P(E)$  can be calculated. Let  $\hat{E}_{EM, i}$  be the  $i$ th element of  $\hat{E}_{EM}$ . Usually,  $\hat{E}_{EM, i}$  is mapped to 1 if  $\hat{E}_{EM, i} > 0.5$  and 0 otherwise for discretizing  $\hat{E}_{EM}$  to the  $p$ -dimensional binary space  $\{0, 1\}^p$ , but such a mapping is not guaranteed to provide  $\arg \max_{E \in \{0, 1\}^p} P(E)$ . Although other mappings can be considered, to the best of our knowledge, no mapping is guaranteed to provide the optimal solution in polynomial time of  $p$ .

In the following, we prove a proposition in order to show that the variational annealing possibly give the optimal solution of Equation (2) even if the factorization of  $Q(E) = \prod_{i=1}^p Q(E_i)$  is additionally considered.

**Proposition 1.**  $P(X, \Theta, E)$  is factorized into  $P(\Theta, E)P(X, E)$ , and  $P(E_i|X, \Theta, E \setminus \{E_i\})$  is given as a binomial distribution. Let  $\hat{Q}(E_i)$  be  $Q(E_i)$  maximizing the lower bound of the variational annealing for  $\tau \rightarrow +0$  given by

$$\int dE_1 \dots \int dE_p \int dX \int d\Theta Q(X) Q(\Theta) \prod_i Q(E_i) \log \frac{P(X, \Theta, E)}{Q(X) Q(\Theta) \prod_i Q(E_i)^\tau}. \quad (5)$$

Then, the set of  $E_i \in \{0, 1\}$  maximizing  $\hat{E}_{EM}$  is  $\arg \max_{E \in \{0, 1\}^p} P(E)$ .

For the proof of the proposition, see Section 1 in Additional file 1. From Proposition 1, if the factorization  $P(X, \Theta, E) = P(\Theta, E)P(X, E)$  is satisfied and optimal  $Q$  functions are found, the variational annealing is guaranteed to provide the optimal solution of Equation (2) while the variational method and the EM algorithm are not. Although the factorization is not a generally satisfied property, the factorization is often assumed in approaches based on the variational method, and the assumption usually works as good approximations. Thus, the variational annealing is expected to provide the better performance than the variational method and the EM algorithm even if the factorization is not satisfied exactly.

#### Procedures of variational annealing on proposed model

In the variational annealing on the proposed model, we calculate  $Q$  functions for hidden variables  $X$ , parameters  $\Theta$ , and binary variables  $E$  iteratively while cooling temperature  $\tau$  to zero gradually at each iteration cycle. In the following, we show the calculation procedures of  $Q(X)$ ,  $Q(\Theta)$ , and  $Q(E)$  on the proposed model as variational E-step, variational M-step, and variational A-step, respectively. More details of the procedures are given in Additional file 1. For the notational brevity, we denote the expectation of a value  $x$  with a probability distribution  $Q(y)$  as  $\langle x \rangle_{Q(y)}$ .

##### Variational E-step

Parameters of  $Q(X)$  are mean of  $\mathbf{x}_t$ , variance of  $\mathbf{x}_t$ , and cross time variance of  $\mathbf{x}_{t-1}$  and  $\mathbf{x}_t$ . These parameters can be calculated via variational Kalman filter by using following terms expected with  $Q(\Theta)Q(E)$ :  $\langle E^{(c)} \rangle_{Q(\Theta)Q(E)}$ ,  $\langle A \rangle_{Q(\Theta)Q(E)}$ ,  $\langle H^{-1} A \circ E^{(c)} \rangle_{Q(\Theta)Q(E)}$ , and  $\langle (A \circ E^{(c)})' H^{-1} A \circ E^{(c)} \rangle_{Q(\Theta)Q(E)}$ . For the details of variational Kalman filter, see [4]. From the parameters of  $Q(X)$ , expectations of the following terms with  $Q(X)$  required in other steps are calculated:  $\langle \mathbf{x}_t^{(c)} \rangle_{Q(X)}$ ,  $\langle \mathbf{x}_t^{(c)} (\mathbf{x}_t^{(c)})' \rangle_{Q(X)}$ , and  $\langle \mathbf{x}_{t+1}^{(c)} (\mathbf{x}_t^{(c)})' \rangle_{Q(X)}$ .

##### Variational M-step

$Q(\Theta)$  is factorized into  $\prod_i Q(A_i|h_i) Q(h_i) Q(r_i) \prod_j Q(z_{ij})$ , where  $A_i$  is a vector given by  $(A_{i1}, \dots, A_{ip})'$ . From the design of the proposed model,  $Q(A_i|h_i)$ ,  $Q(h_i)$ ,  $Q(r_i)$ , and  $Q(z_{ij})$  are given in the following form:

$$\begin{aligned} Q(A_i|h_i) &= \mathcal{N}(A_i; \boldsymbol{\mu}_{A_i}, h_i T_{A_i}^{-1}), \\ Q(h_i) &= \mathcal{IG}(h_i; u_i, k_i), \\ Q(r_i) &= \mathcal{IG}(r_i; v_i, l_i), \\ Q(z_{ij}) &= \mathcal{B}(\zeta_{ij,0}; \zeta_{ij,1}). \end{aligned}$$

Parameters for the above functions are calculated by using  $\langle \mathbf{x}_t^{(c)} \rangle_{Q(X)}$ ,  $\langle \mathbf{x}_t^{(c)} (\mathbf{x}_t^{(c)})' \rangle_{Q(X)}$ ,  $\langle \mathbf{x}_{t+1}^{(c)} (\mathbf{x}_t^{(c)})' \rangle_{Q(X)}$ , and  $\langle E_{ij}^{(c)} \rangle_{Q(E)}$ .

#### Variational A-step

For the calculation of  $Q(E)$ , we assume the factorization of  $Q(E)$  to  $\prod_c \prod_{ij} Q(E_{ij}^{(c)})$  in order to make the computation tractable.  $Q(E_{ij}^{(c)})$  follows a binomial distribution that takes one with probability  $e_{ij}^{(c)}$  and zero with probability  $1 - e_{ij}^{(c)}$ , and thus the expectation of  $E_{ij}^{(c)}$  with  $Q(E)$  is given by  $e_{ij}^{(c)}$ . For the preparation, we calculate  $\langle A \rangle_{Q(\Theta)}$ ,  $\langle H^{-1}A \rangle_{Q(\Theta)}$ ,  $\langle A'H^{-1}A \rangle_{Q(\Theta)}$ ,  $\langle \mathbf{x}_t^{(c)} \rangle_{Q(X)}$ ,  $\langle \mathbf{x}_t^{(c)} (\mathbf{x}_t^{(c)})' \rangle_{Q(X)}$ , and  $\langle \mathbf{x}_{t+1}^{(c)} (\mathbf{x}_t^{(c)})' \rangle_{Q(X)}$ .  $Q(E_{ij}^{(c)})$  is then iteratively calculated by using these expected terms as well as  $\langle E_{ik}^{(c)} \rangle_{Q(E)}$  for  $k \neq j$ . A few iterations are enough for the convergence.

#### Update and selection of hyperparameters

The proposed model contains  $u_0$ ,  $k_0$ ,  $v_0$ ,  $l_0$ ,  $\zeta_{i0}$ ,  $\zeta_{i1}$ ,  $\alpha_0$ , and  $\alpha_1$  as hyperparameters.  $\alpha_0$  and  $\alpha_1$  should be  $\alpha_0 < \alpha_1$  as in the model setting, but this condition can be violated in the update step of the variational method. Thus, we select  $\alpha_0$  and  $\alpha_1$  by cross validation, and update other hyperparameters as in the variational method.

We first consider update of hyperparameters  $u_0$ ,  $k_0$ ,  $v_0$ ,  $l_0$ ,  $\zeta_{i0}$ , and  $\zeta_{i1}$  to increase the lower bound of marginal probability.  $u_0$  and  $k_0$  are updated by maximizing the following equation:

$$\begin{aligned} (\hat{u}_0, \hat{k}_0) &= \arg \max_{(u_0, k_0)} \sum_i \int Q(h_i) \log \mathcal{IG}(h_i; u_0, k_0) dh_i \\ &= \arg \max_{(u_0, k_0)} (u_0 - 1) \frac{\sum_i (\log h_i)_{Q(h_i)}}{p} + u_0 \log k_0 - k_0 \frac{\sum_i (1/h_i)_{Q(h_i)}}{p} - \log \Gamma(u_0). \end{aligned}$$

$\hat{u}_0$  and  $\hat{k}_0$  are obtained by numerical optimization methods such as the Newton-Raphson method.  $v_0$  and  $l_0$  are also updated in a similar manner to  $u_0$  and  $k_0$ .  $\zeta_{i0}$  and  $\zeta_{i1}$  are updated by solving the following equation:

$$\begin{aligned} (\hat{\zeta}_{i0}, \hat{\zeta}_{i1}) &= \arg \max_{\zeta_{i0}, \zeta_{i1}} \sum_j \int Q(z_{ij}) \log \mathcal{B}(z_{ij}; \zeta_{i0}, \zeta_{i1}) dz_{ij} \\ &= \arg \max_{\zeta_{i0}, \zeta_{i1}} (\zeta_{i1} - 1) \frac{\sum_j (\log(1 - z_{ij}))_{Q(z_{ij})}}{p} + (\zeta_{i0} - 1) \frac{\sum_j (\log z_{ij})_{Q(z_{ij})}}{p} - \log B(\zeta_{i0}, \zeta_{i1}). \end{aligned}$$

$\hat{\zeta}_{i0}$  and  $\hat{\zeta}_{i1}$  can also be obtained by the Newton-Raphson method.

For the selection of  $\alpha_0$  and  $\alpha_1$ , we set  $\alpha_1$  to some large value and select  $\alpha_0$  by a leave one out cross validation procedure. For condition  $c \in \{1, 2\}$  and time point  $t \in \mathcal{T}_{obs}^{(c)}$ , we remove  $\mathbf{y}_t^{(c)}$  from data set  $\mathbf{y}$  and use the data

set to train the model. We calculate square sum of residues  $r_{t,c}^2$  between  $\mathbf{y}_t^{(c)}$  and the prediction of  $\mathbf{x}_t^{(c)}$  estimated from the variational Kalman filter on the trained model given by  $r_{t,c}^2 = (\mathbf{y}_t^{(c)} - \mathbf{x}_t^{(c)})'(\mathbf{y}_t^{(c)} - \mathbf{x}_t^{(c)})$ . By grid search on parameter space of  $\alpha_0$ , we select  $\alpha_0$  that minimizes  $\sum_{c=1}^2 \sum_{t \in \mathcal{T}_{obs}^{(c)}} r_{t,c}^2$ .

#### Summary of procedures

The procedures for estimating parameters in the proposed model are summarized as follows:

1. Set  $\tau$  to some large value. Also set  $\alpha_0$  to a small value and  $\alpha_1$  to a large value satisfying that  $\alpha_0 < \alpha_1$ .
2. Initialize other hyperparameters and hidden variables.
3. Perform the following procedures:
  - (a) Calculate variational M-step.
  - (b) Update hyperparameters.
  - (c) Calculate variational E-step.
  - (d) Calculate variational A-step.
  - (e) Go back to step (a) until some convergence criterion is satisfied.
4. Divide  $\tau$  by some value  $> 1$  such as 1.05.
5. Go back to step 3 if  $\tau$  is larger than some very small value  $> 0$ .

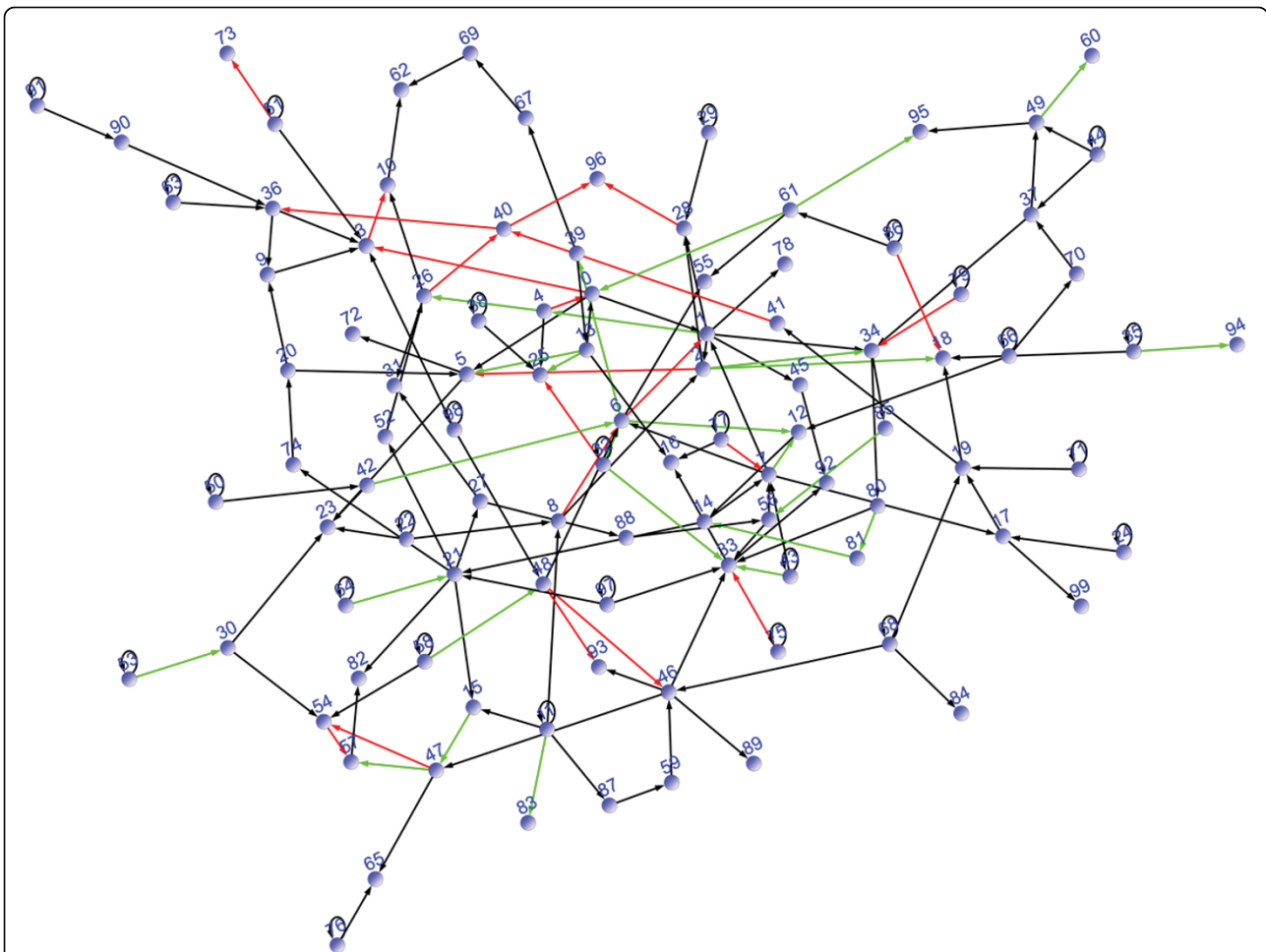
In our setting,  $\alpha_1$  is set to 1,000. For the initialization of  $\tau$  and other hyperparameters, we use the following settings:  $\tau = 2.5$ ,  $E_{ij}^{(c)} = 0.5$ ,  $u_i = 1$ ,  $k_i = 1$ ,  $v_i = 1$ ,  $l_i = 1$ ,  $\zeta_{i0} = 10$ , and  $\zeta_{i1} = 10$ . If  $\mathbf{y}_t^{(c)}$  is observed, we initialize  $\mathbf{x}_t^{(c)}$  with  $\mathbf{y}_t^{(c)}$ . Otherwise, we use the linearly interpolated one.

## Results and discussion

### Performance evaluation by Monte Carlo experiments

For the evaluation of the proposed approach, we generate two linear regulatory network models with similar topological structures  $G_1$  and  $G_2$  based on a linear regulatory network model  $G_0$ .  $G_0$  is prepared in the following manner: (i) a scale free network of 100 nodes and 150 edges is generated; (ii) edge directions are assigned randomly; (iii) autoloop edges are added to root nodes of the directed network; and (iv) AR coefficients for the directed edges are chosen randomly from  $\{-0.9, -0.8, -0.7, -0.6, -0.5, 0.5, 0.6, 0.7, 0.8, 0.9\}$ . We then generate  $G_1$  and  $G_2$  from  $G_0$  as follows: (i) autoloop edges and 70% of non-autoloop edges in  $G_0$  are used for commonly existing edges in  $G_1$  and  $G_2$ ; and (ii) the other 30% of non-autoloop edges are randomly assigned as either  $G_1$  or  $G_2$  specific edges. Note that AR coefficients on edges of  $G_1$  and  $G_2$  are preserved, i.e., if a regulation from gene  $j$  to gene  $i$  exists in  $G_1$  or  $G_2$ , then its coefficient is the same as that of the regulation from gene  $j$  to gene  $i$  in  $G_0$ .

Figure 2 gives graph structure of  $G_1$  and  $G_2$ , where commonly regulations,  $G_1$  specific regulations, and  $G_2$



**Figure 2** The graph structures of  $G_1$  and  $G_2$ . Commonly regulations,  $G_1$  specific regulations, and  $G_2$  specific regulations are represented with black, red, and green arrows, respectively.

specific regulations are represented with black, red, and green arrows, respectively.

From each of  $G_1$  and  $G_2$ , we obtain two equally spaced time series data of 25 time points and 50 time points. For system noise and observation noise, normally distributed values with mean 0 and standard deviation 1 and mean 0 and standard deviation 0.1 and 1 are used, respectively. The signal-noise ratio for system noise with standard deviation 1 and observation noise with standard deviation 0.1 is 0.03dB, and the signal is bit stronger than the noise. On the other hand, for observation noise with standard deviation 1, the signal-noise ratio is -0.26dB. In the condition, the noise is stronger than the signal, and the noise level is quite high.

#### Comparison between variational annealing and EM algorithm

We first compare the performances of the proposed approach and the approach that is based on the proposed approach but uses the EM algorithm instead of the variation annealing using the equally spaced time

series data of 50 and 25 time points on the system noise with standard deviation 1 and observation noise with standard deviation 0.1. From the comparison, we verify the effectiveness of the variational annealing, compared to the EM algorithm. Table 1 summarizes the results of the proposed approach and the EM algorithm based approach. For the EM algorithm, the regulation from  $j$  to  $i$  on condition  $c$  is considered to exist if the estimated  $E_{ij}^{(c)}$  is more than 0.5.

From the comparison, the results of the proposed approach contain more true positives than those of the EM algorithm based approach except for identifying changes on regulations for time points 50. For identifying changes on regulations, the EM algorithm based approach estimates bit more true positives than the proposed approach, but the difference is so small that it can be ignored. On the other hand, the results of the EM algorithm based approach contain more false positives than those of the proposed approach, and hence the precision of the results by the EM algorithm is



worse than that of the proposed approach. Therefore, the effectiveness of the variational annealing is confirmed in the computational experiment as well.

**Comparison between proposed approach and existing approaches**

We employ the elastic net based VAR model approach [2] and the dynamic Bayesian network based approach termed G1DBN [8] as existing approaches for estimating networks from time series data. For the experiments, two versions of these approaches are considered: ENet1 and ENet2 from the elastic net based VAR model approach, and G1DBN1 and G1DBN2 from G1DBN. These approaches are different in the following point: ENet1 and G1DBN1 estimate  $G_1$  and  $G_2$  independently, i.e., for the estimation of  $G_1$ , only time series data from  $G_1$  is used, while ENet2 and G1DBN2 assume that  $G_1$  and  $G_2$  have the same network structure and estimate a network by using two time series data. Thus, ENet2 and G1DBN2 are considered to more use data sample than ENet1 and G1DBN1 for network estimation, but changes on regulations between  $G_1$  and  $G_2$  are not considered. For selection of hyperparameters in ENet1 and ENet2, AICc is used [2], and for hyperparameters  $\alpha_1$  and  $\alpha_2$  of G1DBN1 and G1DBN2, a setting of  $\alpha_1 = 0.1$  and  $\alpha_2 = 0.0059$  considered in [8] is used.

For the comparison of these approaches, we focus on the following two points: the number of correctly

estimated regulations and the number of correctly estimated changes on regulations. The former is usually considered for evaluating the performance of gene network estimation methods. The numbers of true positives and false negatives of the estimated regulations are summarized in Table 2(a). The precisions of the results given by

$$\frac{\text{The number of true positives}}{\text{The number of true positives} + \text{The number of false positives}}$$

are also provided. The results are averaged on ten data sets. The number of regulations in the true network models of  $G_1$  and  $G_2$  are in total 305. For the latter point, we consider the estimated regulations existing only in one of two estimated networks as changed regulations, and check if they correctly exist only in the corresponding true network. The numbers of true positives and false negatives of the estimated changes on regulations and the precisions are summarized in Table 2(b). The results are also averaged on ten data sets. The number of true changes on regulations between in  $G_1$  and  $G_2$  are 47, i.e., the number of true positives on this case is at most 47.

For the estimation of the regulations in Table 2(a), the proposed approach outperforms other approaches in terms of true positives. The proposed approach contains more false positives than ENet1, G1DBN1, and

**Table 2 A summary of results for system noise with standard deviation 1 and observation noise with standard deviation 0.1**

(a)												
# of time points	Equally spaced						Unequally spaced					
	50			25			50			25		
	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE
Proposed	295.9	41.7	0.88	238.4	71.6	0.77	262.4	42.1	0.86	110.7	37.2	0.75
ENet1	246.3	119.7	0.67	109.2	67	0.62	84.7	140.6	0.38	20.3	70.4	0.22
ENet2	277.9	130.9	0.68	212.8	130	0.62	169.7	241.5	0.41	65.5	132.5	0.33
G1DBN1	223.7	48	0.82	99.9	46.2	0.68	65.1	83.1	0.44	19.3	72.7	0.21
G1DBN2	268.8	83.4	0.76	188.1	64.5	0.74	134.8	104.4	0.56	46.7	85.7	0.35

(b)												
# of time points	Equally spaced						Unequally spaced					
	50			25			50			25		
	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE
Proposed	39.8	13.2	0.75	23.4	20.8	0.53	31.2	16.5	0.65	5.5	15.9	0.26
ENet1	38.5	153.1	0.2	18.6	113	0.14	12.3	186.6	0.06	3.7	85	0.04
ENet2	-	-	-	-	-	-	-	-	-	-	-	-
G1DBN1	35.6	88.9	0.29	16.8	91.9	0.15	10.3	121.9	0.08	2.4	87.8	0.03
G1DBN2	-	-	-	-	-	-	-	-	-	-	-	-

(a) The number of true positives (# TP) and false positives (# FP) of estimated regulations in two network model by the proposed approach, ENet1, ENet2, G1DBN1, and G1DBN2 for equally and unequally spaced time series data. PRE denotes the precision of the results. Regulations in two networks are 305 in total. (b) The number of true positives (# TP) and false positives (# FP) of changes on regulations between two network models estimated by the proposed approach, ENet1, ENet2, G1DBN1, and G1DBN2 for equally and unequally spaced time series data. Since no changes are estimated by ENet2 and G1DBN2, their results are indicated by '-'. The regulations changed in two networks are in total 47.

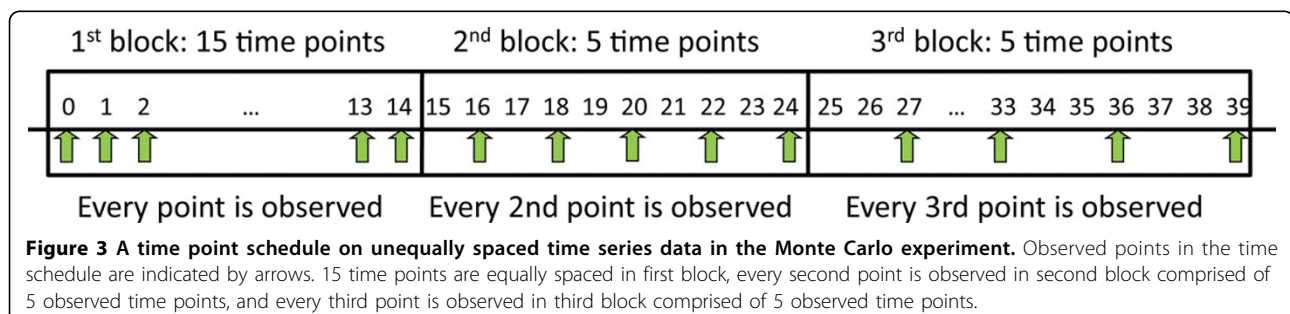
G1DBN2 on the data of 25 time points. We further consider these cases in terms of the precision. The precisions of the proposed approach, ENet1, G1DBN1, and G1DBN2 are given by 0.77, 0.62, 0.68, and 0.74, respectively. From this analysis, the proposed approach shows better performance than ENet2, G1DBN1, and G1DBN2 on the whole. More true positives are estimated by ENet2 than ENet1, and the precisions of ENet2 tend to be better than those of ENet1. This type of relationship is also observed between G1DBN1 and G1DBN2. Also, the elastic net based VAR model approaches estimate more true positives than the approaches from G1DBN. However, in the precision, the approaches from G1DBN are better than the elastic net based VAR model approaches. From the results in Table 2(b), we see that the proposed approach estimates more true changes on regulations than ENet1 and G1DBN1 on both data of 25 and 50 time points. In addition, the results of the proposed approach contain less false positives than those of ENet1 and G1DBN1. No changes on regulations are detected in ENet2 and G1DBN2 as topological differences are ignored in these approaches. Since the proposed approach considers differences on network structures as well as uses two time series data efficiently, it can provide better results than other approaches on both estimating regulations and identifying changes on regulations.

One may think it is strange that false positives in ENet1 and G1DBN1 in Table 2(b) is more than those in Table 2(a), but this case can occur from the following reason. If a regulation exists in both of the true network models of  $G_1$  and  $G_2$ , but is estimated only for  $G_1$ , then the case is not counted as a false positive in Table 2(a) while it is counted as a false positive in Table 2(b). Thus, the number of false positives in Table 2(b) can be greater than those in Table 2(a).

In order to show the performance in unequally spaced time series data, we generate unequally spaced time series data of 25 and 50 observed time points. For time series data of 25 observed time points, we first generate equally spaced time series data of 40 time points and divide it into three blocks: 15 time points, 10 time

points, and 15 time points. We then remove time points in the following manner: no time point is removed in the first block; one of every two time points are removed in the second block; and two of every three time points are removed in the third block. Figure 3 shows the time point schedule of time series data obtained in this process. For time series data of 50 observed time points, we first generate equally space time series data of 80 time points, divide it into three blocks: 30 time points, 20 time points, and 30 time points. Then, some time points are removed in a similar manner. We apply the proposed approach, ENet1, ENet2, G1DBN1, and G1DBN2 to the unequally spaced time series data. Results for the dataset are also summarized in Tables 2(a) and 2(b). From the comparison of results on equally and spaced time series data, the results of the proposed approach from unequally spaced time series data are worse than those from equally spaced one even with the same number of observed time points. However, results of ENet1, ENet2, G1DBN1, and G1DBN2 are worsened more than those of the proposed approach. This is probably because unequally spaced time points break their assumption, and their estimation process is misled.

We also consider the time series data with the high level noise: system noise with standard deviation 1 and observation noise with standard deviation 1. The results for the case are summarized in Tables 3(a) and 3(b). The proposed approach shows the better performance on the number of true positives and precisions than other approaches except for the identification of the changes on regulations from the equally spaced time series data of 50 time points. For equally spaced time series data of 50 time points, the number of true positives on the changes on regulations estimated by the proposed approach is more than that of G1DBN1, but less than that of ENet1. However, the precisions on the estimated changes by both ENet1 and G1DBN are much worse than the proposed approach. Thus, overall, the proposed approach is more effective than other methods. Although the proposed approach provides the better performance than other methods, the results of all



**Table 3 A summary of results for system noise with standard deviation 1 and observation noise with standard deviation 1**

(a)												
# of time points	Equally spaced						Unequally spaced					
	50			25			50			25		
	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE
Proposed	190.2	122.8	0.61	88.1	121.0	0.42	132.1	675.5	0.66	52.1	108.9	0.33
ENet1	110.8	136.9	0.45	30.3	75.9	0.29	32.5	133.7	0.2	7.4	75.2	0.09
ENet2	189.8	218	0.47	85.8	136.2	0.39	75.9	180.7	0.3	23.5	123.3	0.16
GIDBN1	86.6	82.6	0.51	22.6	90.8	0.2	26.3	74	0.26	7.1	71.6	0.09
GIDBN2	163.9	105.7	0.61	54.4	99	0.35	66.2	91.2	0.42	17.4	92.8	0.16

(b)												
# of time points	Equally spaced						Unequally spaced					
	50			25			50			25		
	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE	# TP	# FP	PRE
Proposed	15.4	43.6	0.26	4.7	50.0	0.09	8.1	16.7	0.33	3.1	42.5	0.07
ENet1	16.9	184	0.08	3.8	95.4	0.04	5.2	155	0.03	1.2	81.4	0.01
ENet2	-	-	-	-	-	-	-	-	-	-	-	-
GIDBN1	14.5	125.1	0.1	3.9	105.7	0.04	4	91.9	0.04	1.7	76.8	0.02
GIDBN2	-	-	-	-	-	-	-	-	-	-	-	-

(a) The number of true positives (# TP) and false positives (# FP) of estimated regulations in two network model by the proposed approach, ENet1, ENet2, G1DBN1, and G1DBN2 for equally and unequally spaced time series data. PRE denotes the precision of the results. Regulations in two networks are 305 in total. (b) The number of true positives (# TP) and false positives (# FP) of changes on regulations between two network models estimated by the proposed approach, ENet1, ENet2, G1DBN1, and G1DBN2 for equally and unequally spaced time series data. Since no changes are estimated by ENet2 and G1DBN2, their results are indicated by '-'. The regulations changed in two networks are in total 47.

the approaches are worsened due to the high level noise, and the differences on the performance among the approaches get smaller, compared to the case of observation noise with standard deviation 0.1.

**Analysis of time series microarray data from Human small airway epithelial cells**

We apply the proposed approach to two time series microarray gene expression data from normal Human small airway epithelial cells (SAECs) and SAECs treated by stimulating EGF-receptors and dosing an anticancer drug termed Gefitinib. EGF-receptors are often overexpressed in lung cancer cells such as tumoral SAECs, and a lung cancer condition is simulated in the treated SAECs by stimulating EGF-receptors. Since Gefitinib is known as a selective inhibitor of EGF-receptors, the stimulation of EGF-receptors would be counteracted by Gefitinib, and the condition of treated SACEs should be the same as that of normal SAECs in theory. However, since some gene expression patterns are different between the two conditions in practice, some unknown effects by Gefitinib may be involved in the phenomenon. Thus, we focus on changed regulations between the gene networks estimated from gene expression data in these two conditions in order to find some insights on the unknown effects of Gefitinib.

For gene set selection, we first screen 500 genes from the ranking of the gene list sorted by coefficient

variation [5]. We then select 100 genes with highly varied expression profiles between normal SAECs and treated SAECs. The time series gene expression data from both types of cells are comprised of spaced 14 time points in 48 hours. The time schedule of 14 time points are {0 h, 6 h, 9 h, 12 h, 15 h, 18 h, 21 h, 24 h, 27 h, 30 h, 33 h, 36 h, 39 h, 48 h}. For the analysis on the proposed approach, we set interval on system model to three hours.

**Table 4 Changes on regulations between normal and treated SAECs**

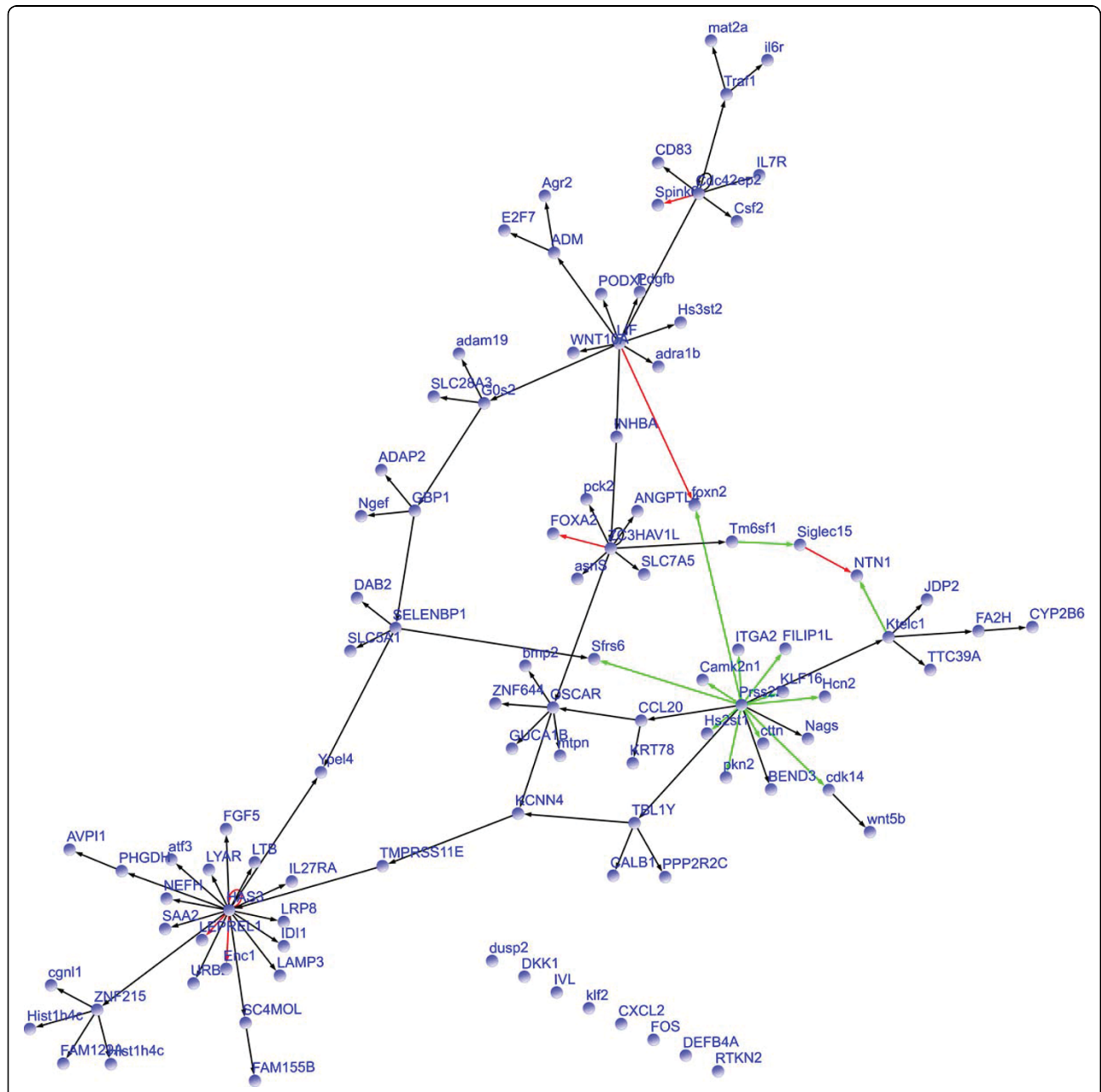
Normal SAECs	Treated SAECs
ZC3HAV1L → FOXA2	Prss22 → foxn2
LIF → foxn2	Prss22 → cdk14
Cdc42ep2 → Spink6	Prss22 → Camk2n1
Siglec15 → NTN1	Prss22 → cttn
HAS3 → HAS3	Prss22 → Sfrs6
HAS3 → Enc1	Prss22 → ITGA2
HAS3 → LEPREL1	Prss22 → pkn2
	Prss22 → Hs2st1
	Prss22 → FILIP1L
	Prss22 → Hcn2
	Prss22 → KLF16
	Ktelc1 → NTN1
	Tm6sf1 → Siglec15

Estimated regulations only in normal or treated SAECs are listed in the left side or right side, respectively.

The estimated networks from time series gene expression data in normal SAECs and treated SAECs are summarized and given in Figure 4. Black arrows, red arrows, and green arrows indicate regulations in both conditions, only in normal SAECs, and only in treated SAECs, respectively. Table 4 gives a list of the estimated regulations only in normal SAECs or in treated SAEC.

From Table 4, we see that Prss22 is involved in most of the regulations only in treated SAECs. Prss22 is a

trypsin, one of serine proteases, and its relationship with the airways is suggested by a report about its expression in the airways in a developmentally regulated manner. Trypsin is a potent mitogen of fibroblast [15], and it is reported that the increase and activation of fibroblast are promoted in lung cells under the condition of interstitial pneumonia. Interstitial pneumonia is known as a side effect of Gefitinib, and these findings suggest that Prss22 is an off-target of Gefitinib and is



**Figure 4** An estimated gene network by the proposed approach from time series gene expression data on normal SAECs and treated SAECs. In the estimated network, regulations in both conditions are in black, and regulations only in normal SAECs and treated SAECs are in red and green, respectively.

possibly related to the side effect of Gefitinib. Also, FILIP1L, a gene estimated as one of targets of Prss22 in treated SAECs, is reported as an inhibitor of cell proliferation of fibroblast [16]. This observation supports the relation between Prss22 and fibroblast as well.

We also focus on other several genes related to changes on regulations in normal and treated SAECs. LIF, leukemia inhibitory factor, is known to affect cell growth and development. Gefitinib is also known to be effective for acute myelogenous leukemia via Sky, which is an off-target gene of Gefitinib [17]. In addition, Wang *et al.* reported that LIF prolongs the cell cycle of stem cells on acute myelogenous leukemia lines [18]. Although, to the best of our knowledge, no direct influence from Gefitinib to LIF is reported, the above facts suggest some relation between Gefitinib and LIF, and support changes on regulations of LIF between normal and treated SAECs.

Heikema *et al.* reported that Human Siglecs, Siglecs-14, -15, and -16 interact with transmembrane adaptor proteins containing the immunoreceptor tyrosine-based activation motif such as DAP12 [19], and therefore they potentially mediate the activation of intracellular signaling. Gefitinib is a selective inhibitor of tyrosine kinase, and the inhibition of tyrosine kinase is considered to affect the regulations around Siglec-15.

Although the stimulation of EGF-receptors in the treated SAECs is considered to be counteracted by Gefitinib, the expressions of some genes may be affected by the stimulation in practical conditions. HAS3 is related to synthesis of the unbranched glycosaminoglycan hyaluronic acid and is reported to be up-regulated by EGF [20]. The stimulation of EGF-receptors affects the amount of EGF taken into cells, and hence the stimulation is expected to cause the changes on the regulations around HAS3. Foxn2 is a member of family of Fox proteins. Fox proteins are known to play important roles on controlling the expressions of genes related to cell growth, proliferation, and differentiation. Some members of Fox family are related to EGF-receptors, e.g., the expression of Foxn1 is suppressed by EGF-receptor signaling [21] although no direct relation between EGF-receptors and Foxn2 is found. FOXA2 is also a member of family of Fox proteins. EGF-receptor signaling is known to decrease the expression of FOXA, which prevents the mucus production [22]. [23] reported the relation between EGF-receptors and FOXA2 in the airways of asthmatic patients. Thus, it appears that the change on the regulation related to FOXA2 is caused by the stimulation of EGF-receptors.

## Conclusions

We proposed the new computational model that is based on VAR-SSM and estimates gene networks from

time series data on normal and treated conditions as well as identifies changes regulations by the treatment. Unlike many of existing gene network estimation approaches assuming equally spaced time points, our approach can handle unequally spaced time series data. The efficient use of time series data is achieved by representing the presence of regulations on each condition with hidden binary variables. Since finding the optimal configuration of the hidden binary variables on the proposed model is computationally intractable, we derive the extended variational annealing method in order to address the problem as the alternative method.

In the Monte Carlo experiments, we use equally and spaced time series data from synthetically generated two regulatory networks whose structures are different in several regulations, and verified the effectiveness of the proposed model in both estimation of regulations and changes on regulations between the two conditions, compared to existing methods.

As the real data application, we use the proposed approach to analyze two time series data from normal SAECs and SAECs treated by stimulating EGF-receptors and dosing Gefitinib. From genes related to changes on regulations by the treatment, we find possible off-target genes of Gefitinib, and one of these genes is suggested to be related to a factor of interstitial pneumonia, which is known as a side effect of Gefitinib. In this study, we consider changes on regulations in two conditions, but the proposed approach can be extended to identifying changes among more than two conditions.

## Additional material

**Additional file 1: Proof of Proposition 1 and more details on the procedures of variational annealing.** A proof of Proposition 1 and more details on the procedures of variational annealing on the proposed model are described.

## Acknowledgements

We thank the anonymous reviewers for their constructive comments and suggestions, which improved the quality of this publication. This article has been published as part of *BMC Genomics* Volume 13 Supplement 1, 2012: Selected articles from the Tenth Asia Pacific Bioinformatics Conference (APBC 2012). The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2164/13?issue=S1>.

## Author details

<sup>1</sup>Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. <sup>2</sup>Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, São Paulo 05508-090, Brazil.

## Authors' contributions

KK, SI, RY, and SM designed the approach to identify the changes on regulations by the cell treatment to SAECs. KK, SI, and AF contributed to the statistical modeling for the approach, and devised the details of methodologies for estimating the proposed model. MY and NG carried out

the microarray experiment for measuring time series the gene expression data on normal and treated SAECs.

#### Competing interests

The authors declare that they have no competing interests.

Published: 17 January 2012

#### References

1. Opgen-Rhein R, Strimmer K: From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol* 2007, **1**:37.
2. Shimamura T, Imoto S, Yamaguchi R, Fujita A, Nagasaki M, Miyano S: Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Syst Biol* 2009, **3**:41.
3. Fujita A, Sato JR, Kojima K, Gomes LR, Nagasaki M, Sogayar MC, Miyano S: Identification of Granger causality between gene sets. *Journal of Bioinformatics and Computational Biology* 2010, **8**:679-701.
4. Beal MJ: Variational algorithms for approximate Bayesian inference. *PhD thesis* University College London, Gatsby Computational Neuroscience Unit; 2003.
5. Yamaguchi R, Imoto S, Yamauchi M, Nagasaki M, Shimamura T, Hatanaka Y, Ueno K, Higuchi T, Gotoh N, Miyano S: Predicting differences in gene regulatory systems by state space models. *Genome Inform* 2008, **21**:101-113.
6. Shiraishi Y, Kimura S, Okada M: Inferring cluster-based networks from differently stimulated multiple time-course gene expression data. *Bioinformatics* 2010, **26**(8):1073-1081.
7. Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alché Buc F: Gene networks inference using dynamic Bayesian networks. *Bioinformatics* 2003, **19**(Suppl 2):ii138-ii148.
8. Lèbre S: Inferring dynamic genetic networks with low order independencies. *Stat Appl Genet Mol Biol* 2009, **8**:Article 9.
9. Bar-Joseph Z, Gerber G, Simon I, Gifford DK, Jaakkola TS: Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proc Natl Acad Sci U S A* 2003, **100**(18):10146-10151.
10. Kojima K, Yamaguchi R, Imoto S, Yamauchi M, Nagasaki M, Shimamura T, Ueno K, Higuchi T, Gotoh N, Miyano S: A state space representation of VAR models with sparse learning for dynamic gene networks. *Genome Inform* 2010, **22**:56-68.
11. Yoshida R, West M: Bayesian learning in sparse graphical factor models via variational mean-field annealing. *J Mach Learn Res* 2010, **11**:1771-1798.
12. Rose K: Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE* 1998, **86**(11):2210-2239.
13. Ueda N: Deterministic annealing EM algorithm. *Neural Networks* 1998, **11**(2):271-282.
14. Itaya Y, Zen H, Nankaku Y, Miyajima C, Tokuda K, Kitamura T: Deterministic annealing EM algorithm in parameter estimation for acoustic model. *The 8th International Conference on Spoken Language Processing* 2004, 433-436.
15. Ruoss SJ, Hartmann T, Caughey GH: Mast cell tryptase is a mitogen for cultured fibroblasts. *J Clin Invest* 1991, **88**(2):493-499.
16. Kwon M, Hanna E, Lorang D, Quick JS, He M, Adem A, Stevenson C, Chung J, Hewitt SM, Zudaire E, Esposito D, Cuttitta F, Libutti SK: Functional characterization of filamin A interacting protein 1-like, a novel candidate for antivasular cancer therapy. *Cancer Res* 2008, **68**(18):7332-7341.
17. Downing JR: Can treating the SYK cell cure leukemia? *Cancer Cell* 2009, **16**(4):270-271.
18. Wang C, Lishner M, Minden MD, McCulloch EA: The effects of leukemia inhibitory factor (LIF) on the blast stem cells of acute myeloblastic leukemia. *Leukemia* 1990, **4**(8):548-552.
19. Heikema AP, Bergman MP, Crocker PR, Gilbert M, Samsom JN, van Wamel WJ, Endtz HP, van Belkum A: Characterization of the specific interaction between sialoadhesin and sialylated *Campylobacter jejuni* lipooligosaccharides. *Infect Immun* 2010, **78**(7):3237-3246.
20. Pasonen-Seppänen S, Karvinen S, Törrönen K, Hyttinen JM, Jokela T, Lammi MJ, Tammi MI, Tammi R: EGF upregulates, whereas TGF-beta downregulates, the hyaluronan synthases Has2 and Has3 in organotypic keratinocyte cultures: correlations with epidermal proliferation and differentiation. *J Invest Dermatol* 2003, **120**(6):1038-1044.
21. Mandinova A, Kolev V, Neel V, Hu B, Stonely W, Lieb J, Wu X, Colli C, Han R, Pazin MJ, Ostano P, Dummer R, Brissette JL, Dotto GP: A positive FGFR3/FOXP1 feedback loop underlies benign skin keratosis versus squamous cell carcinoma formation in humans. *J Clin Invest* 2009, **119**(10):3127-3137.
22. Zhen G, Park SW, Nguyenvu LT, Rodriguez MW, Barbeau R, Paquet AC, Erle DJ: IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production. *Am J Respir Cell Mol Biol* 2007, **36**(2):244-253.
23. Lai HY, Rogers DF: Mucus hypersecretion in asthma: intracellular signaling pathways as targets for pharmacotherapy. *Curr Opin Allergy Clin Immunol* 2010, **10**:67-76.

doi:10.1186/1471-2164-13-S1-S6

**Cite this article as:** Kojima et al.: Identifying regulational alterations in gene regulatory networks by state space representation of vector autoregressive models and variational annealing. *BMC Genomics* 2012, **13**(Suppl 1):S6.

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

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

Submit your manuscript at  
www.biomedcentral.com/submit

