BIOINFORMATICS ORIGINAL PAPER

Vol. 26 no. 8 2010, pages 1064–1072
doi:10.1093/bioinformatics/btq080

Systems biology

Advance Access publication March 1, 2010

Inferring dynamic gene networks under varying conditions for transcriptomic network comparison

Teppei Shimamura*, Seiya Imoto, Rui Yamaguchi, Masao Nagasaki and Satoru Miyano

Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Elucidating the differences between cellular responses to various biological conditions or external stimuli is an important challenge in systems biology. Many approaches have been developed to reverse engineer a cellular system, called gene network, from time series microarray data in order to understand a transcriptomic response under a condition of interest. Comparative topological analysis has also been applied based on the gene networks inferred independently from each of the multiple time series datasets under varying conditions to find critical differences between these networks. However, these comparisons often lead to misleading results, because each network contains considerable noise due to the limited length of the time series.

Results: We propose an integrated approach for inferring multiple gene networks from time series expression data under varying conditions. To the best of our knowledge, our approach is the first reverse-engineering method that is intended for transcriptomic network comparison between varying conditions. Furthermore, we propose a state-of-the-art parameter estimation method, relevance-weighted recursive elastic net, for providing higher precision and recall than existing reverse-engineering methods. We analyze experimental data of MCF-7 human breast cancer cells stimulated by epidermal growth factor or heregulin with several doses and provide novel biological hypotheses through network comparison.

Availability: The software NETCOMP is available at http://bonsai.ims.u-tokyo.ac.jp/~shima/NETCOMP/.

Contact: shima@ims.u-tokyo.ac.jp

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 2, 2009; revised on February 16, 2010; accepted on February 21, 2010

1 INTRODUCTION

Systems biology aims to understand a cellular process through quantitative mathematical models (Cantone *et al*., 2009). One of the important problems in systems biology is to reverse-engineer the dynamics of gene–gene interactions, called *gene network*, from time series expression data. Several techniques have been developed for the mathematical modeling of these dynamics from time series expression data, notably differential equation (Bansal *et al*., 2006; Chen *et al*., 1999), state space model (Hirose *et al*., 2008; Rangel *et al*., 2004), vector autoregressive (VAR) model (Opgen-Rhein and Strimmer, 2007; Shimamura *et al*., 2009) and dynamic Bayesian networks (Perrin *et al*., 2003; Tamada *et al*., 2009).

In this light, a key challenge in the application of these methods is to develop new strategies and theoretical frameworks for the comparative analysis of multiple gene networks, that is, to infer gene networks from multiple time series expression data under various biological conditions or external stimuli, for example, different doses of a chemical compound or a growth factor, and to identify subnetworks (interactions) that are conserved or not conserved among the conditions. Comparative topological analysis provides a very good opportunity to reveal the differences between regulatory structures and understand cellular responses to varying conditions. However, this is not an easy task because the number of candidate interactions is very large as compared to the number of true interactions. This problem becomes more serious for short time series data. As of October 2009, in time series experiments designed to comparison between multiple biological conditions from public databases such as Gene Expression Omnibus (GEO) and Array Express, >90% were short time series with fewer than 10 time points. In these cases, the application of existing reverse-engineering methods to these short time series data leads to poor performance in that a large number of false positives and false negatives are produced in network inference. Therefore, the noisy interactions inferred using existing methods make it difficult to identify critical differences and similarities between conditions and to produce useful results.

In this study, we propose a novel-integrated approach for inferring multiple gene networks from time series expression data under varying biological conditions. It should be noted that earlier algorithms aimed to infer a gene network under a condition of interest and they were not intended to compare multiple gene networks, that is, to infer each gene network independently based on only samples from a condition of interest, called *exact samples* (Hu, 1994). Our approach is different from these earlier approaches in that it aims at both network inference and comparison, and it uses exact samples along with samples from other conditions, called *relevant samples* (Hu, 1994). To the best of our knowledge, other than the proposed method, no other method is available for reverse-engineering multiple gene networks in order to compare transcriptomic responses under different conditions.

Figure 1 shows a conceptual view of the proposed method. Our key assumption is that large regions of the gene networks are conserved and that small regions are changed between conditions. This implies that a network model for generating a sample under

*To whom correspondence should be addressed.

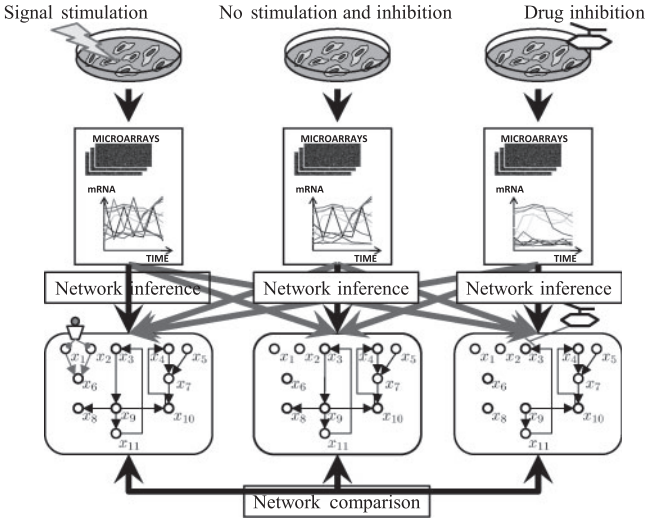

Fig. 1. Conceptual view of our method using an example of inferring gene networks under three different biological conditions (no stimulation and no perturbation, stimulation with receptor ligands and perturbation with small compounds) from time series microarray experiments. Traditional inference is based on exact sample information (black line). Our proposed inference is based on both exact sample information (gray line) and relevant sample information (black line).

a condition is partially related to the other network models for generating a sample under other conditions. We develop a new parameter estimation method called *relevance-weighted recursive elastic net* (RW-RENET) for inferring multiple network structures based on both exact sample information and relevant sample information. RW-RENET automatically infers conserved interactions from samples under all the relevant conditions and unconserved interactions from samples under each condition with a similarity based on the comparison between estimators. This enables us to achieve a higher prediction accuracy of inferring gene networks as compared to earlier methods without losing critical differences between gene networks under certain conditions. We test our approach using simulated data and experimental data of MCF-7 human breast cancer cells to assess its performance relative to that of the existing methods.

2 METHODS

2.1 Model description

We assume that time series expression data are obtained from m different biological conditions. More formally, for each condition ($l=1,\ldots,m$), we have $\mathbf{x}_1^{(l)},\ldots,\mathbf{x}_{n_l}^{(l)}$, where $\mathbf{x}_i^{(l)}=(x_1^{(l)}(t_i),\ldots,x_p^{(l)}(t_i))'$ represents the mRNA concentrations of p genes at time step i whose j-th element $x_j^{(l)}(t_i)$ indicates the concentration of the j-th gene transcript measured at time t_i, and n_l is the number of observed time points in the l-th condition.

To represent the dynamics of gene–gene interactions across time under the l-th biological condition, we assume a Gaussian process of the form

$$x_k^{(l)}(t_{i+1})\sim f_k^{(l)}=\phi(x_k^{(l)}(t_{i+1})|\mathbf{x}_i^{(l)}(t_i);\mu_k^{(l)}(t_{i+1}),\sigma_k^{(l)}), \tag{1}$$

where $f_k^{(l)}$ is a conditional density function of $x_k^{(l)}(t_{i+1})$ described by a normal density function as follows:

$$\phi(x_k^{(l)}(t_{i+1})|\mathbf{x}_i^{(l)}(t_i);\mu_k^{(l)}(t_{i+1}),\sigma_k^{(l)})$$

$$\propto\exp\left(-\frac{(x_k^{(l)}(t_{i+1})-\mu_k^{(l)}(t_{i+1}))^2}{2\sigma_k^{(l)^2}}\right),$$

with mean $\mu_k^{(l)}(t_{i+1})$ and SD $\sigma_k^{(l)}$. The mean $\mu_k^{(l)}(t_{i+1})$ is given as

$$\mu_k^{(l)}(t_{i+1})=b_{0,k}^{(l)}+\sum_{j=1}^{p}b_{j,k}^{(l)}x_j^{(l)}(t_i),$$

where $b_{0,k}^{(l)}$ and $b_{j,k}^{(l)}$ define the y-intercept and the influence of the j-th gene transcript on the k-th gene transcript under the l-th condition, respectively. We expect most gene transcripts to be regulated by only a small subset of a total of p gene transcripts, and thus assume that a large number of $b_{j,k}^{(l)}$ ($j=1,\ldots,p$ and $k=1,\ldots,p$) will be zero. The estimates $\hat{b}_{j,k}^{(l)}$ for all j and k provide a network topology among gene transcripts under the l-th biological condition; nodes represent gene transcripts and an edge $j\rightarrow k$ exists if and only if $\hat{b}_{j,k}^{(l)}\neq 0$.

2.2 Estimation

Without loss of generality, we explain the problem of estimating the parameters $b_{1,k}^{(1)},\ldots,b_{p,k}^{(1)}$ and $\sigma_k^{(1)}$ under the first condition. We assume that the density functions $f_k^{(2)},\ldots,f_k^{(m)}$ might be related to the density function of the first condition $f_k^{(1)}$. In this case, $\mathbf{x}_1^{(1)},\ldots,\mathbf{x}_{n_1}^{(1)}$ are called the exact samples and $\mathbf{x}_1^{(l)},\ldots,\mathbf{x}_{n_l}^{(l)}$ ($l=2,\ldots,m$) are called the relevant samples (Hu, 1994). Since the conditional density function of $x_k^{(1)}(t_{i+1})$ depends on the parameters $\mathbf{b}_k^{(1)}$ and $\sigma_k^{(1)}$, we rewrite the conditional density as $\phi(x_k^{(l)}(t_{i+1})|\mathbf{x}^{(l)}(t_i);\mathbf{b}_k^{(l)},\sigma_k^{(l)})$.

Classical maximum likelihood estimators of $b_{j,k}^{(1)}$ ($j=1,\ldots,p$) and $\sigma_k^{(1)}$ are obtained from $\mathbf{x}_1^{(1)},\ldots,\mathbf{x}_{n_1}^{(1)}$ alone. To incorporate the information from the relevant samples $\mathbf{x}_1^{(l)},\ldots,\mathbf{x}_{n_l}^{(l)}$ ($l=2,\ldots,m$), we use the weighted log-likelihood (Hu, 1994; Hu and Zidek, 2002; Wang, 2001) as

$$\mathrm{WL}_k^{(1)}=\sum_{l=1}^{m}\sum_{i=1}^{n_l-1}\rho_k^{(1,l)}\log\phi(x_k^{(l)}(t_{i+1})|\mathbf{x}^{(l)}(t_i);\mathbf{b}_k^{(1)},\sigma_k^{(1)}) \tag{2}$$

where $\rho_k^{(1,l)}$ is an weight parameter called *likelihood weight* for down-weighting the relevant samples according to the degree of similarity between $f_k^{(1)}$ and $f_k^{(l)}$. Note that although maximizing the weighted log-likelihood (2) provides estimators of $b_{j,k}^{(1)}$, these estimators are not sparse. Therefore, we need to further investigate whether $b_{j,k}^{(1)}$ is zero or non-zero, which is a structure learning problem. In many cases, this task is computationally hard (Chickering *et al.*, 2004).

For simultaneous parameter estimation and structure learning, we propose a new parameter estimation method called RW-RENET to maximize the following regularized weighted log-likelihood:

$$\mathrm{RWL}_k^{(1)}=\mathrm{WL}_k^{(1)}-\lambda_{1,k}^{(l)}\sum_{j=1}^{p}w_{j,k}^{(1)}|b_{j,k}^{(1)}|-\frac{\lambda_2^{(1)}}{2}\sum_{j=1}^{p}b_{j,k}^{(1)^2}, \tag{3}$$

where $\lambda_{1,k}^{(1)}$ and $\lambda_2^{(1)}$ are regularization parameters of the L_1- and L_2-penalty terms, respectively, and $w_{j,k}^{(1)}$ is an importance weight of $b_{j,k}^{(1)}$ that enables a different amount of shrinkage for each coefficient in the L_1-penalty term. The L_1-penalty shrinks some coefficients to zero (Tibshirani, 1996), which gives a sparse structure of graphical models. In contrast, the L_2-penalty encourages the grouping effect in that highly correlated variables will be in or out of the model and it stabilizes the solution (Zou and Hastie, 2005). The importance weights $w_{j,k}^{(1)}$ ($j=1,\ldots,p$) allow for different tuning parameters for different coefficients $b_{j,k}^{(1)}$, respectively. If $w_{j,k}^{(1)}$ takes a large value, an estimator $\hat{b}_{j,k}^{(1)}$ tends to be zero. In contrast, if $w_{j,k}^{(1)}$ takes a small value nearly equal to zero, $\hat{b}_{j,k}^{(1)}$ tends to be non-zero. These weights encourage a more sparse structure of the network than lasso and elastic net. For practical implementation,

these weights can be estimated by $w_{j,k}^{(1)}=1/(|\bar{b}_{j,k}^{(1)}|+\epsilon)$ where $\bar{b}_{j,k}^{(1)}$ is a plug-in estimator and ϵ is a positive constant to avoid dividing zeros (Shimamura *et al*., 2009). The degree of sparsity of $\hat{b}_{j,k}^{(1)}$ with RW-RENET depends on the values of $\rho_k^{(1,l)}$, $\lambda_{1,k}^{(1)}$, $\lambda_2^{(1)}$ and $w_{j,k}^{(1)}$ $(j=1,\ldots,p)$. Therefore, the problem of learning the network structure leads to the problem of selecting the values of these parameters, as discussed in a latter section.

2.3 Similarity between densities

To obtain reasonable estimates for $\rho_k^{(1,l)}$, we present a method based on the comparison between the estimated parameters. We expect that if the density $f_k^{(l)}$ is similar to the density $f_k^{(1)}$, the parameters $b_{j,k}^{(l)}$ $(j=1,\ldots,p)$ are also similar to the parameters $b_{j,k}^{(1)}$, respectively. We then define the likelihood weights $\rho_k^{(1,l)}$ by

$$\rho_k^{(1,l)}=\frac{S(c_k^{(1,l)};\alpha,\beta)}{\sum_{l=1}^{m}S(c_k^{(1,l)};\alpha,\beta)}, \tag{4}$$

where $S(c_k^{(1,l)};\alpha,\beta)$ is the logistic function

$$S(c_k^{(1,l)};\alpha,\beta)=\frac{1}{1+\exp\{-\alpha(c_k^{(1,l)}-\beta)\}}, \tag{5}$$

with hyperparameters α and β. The $c_k^{(1,l)}$ is the Pearson's correlation coefficient between the estimators $b_{j,k}^{(1)}$ and $b_{j,k}^{(l)}$ $(j=1,\ldots,p)$. In the proposed likelihood weights (4), if the estimators $\hat{b}_{j,k}^{(l)}$ are similar to the estimators $\hat{b}_{j,k}^{(1)}$, the likelihood weight of $\rho_k^{(1,l)}$ approaches that of $\rho_k^{(1,1)}$. Otherwise, $\rho_k^{(1,l)}$ approaches zero. The likelihood weights (4) depend on two parameters α and β, as discussed in Section 2.4.

2.4 Model selection criterion

The regularization parameters $\lambda_{1,k}^{(1)}$ and $\lambda_2^{(1)}$, the importance weights $w_{j,k}^{(1)}$ $(1\le k\le p)$ with the hyperparameters α and β in the logistic function (5) affect the sparsity of $b_{j,k}^{(1)}$, that is, they determine the network structure. Therefore, the problem of choosing these parameters is a crucial issue in practical situations. We investigate this problem from the viewpoint of statistical model evaluation and derive an information criterion for choosing these parameters theoretically.

Hens *et al*. (2006) derived a bias-corrected weighted Akaike information criterion for evaluating models estimated by the weighted maximum likelihood. To extend their methodology to the framework of evaluating models estimated by the regularized weighted maximum likelihood, we derive an unbiased estimate of the degrees of freedom for the regularized weighted log-likelihood (3). Degrees of freedom can be used to quantify the model complexity estimated by L_1-regularization (Zou *et al*., 2007). For simplicity of explanation, we introduce the notations

$$\tilde{\mathbf{X}}=\begin{pmatrix}(\rho_k^{(1,l)})^{1/2}\mathbf{X}^{(1)}\\(\rho_k^{(1,2)})^{1/2}\mathbf{X}^{(2)}\\\vdots\\(\rho_k^{(1,m)})^{1/2}\mathbf{X}^{(m)}\end{pmatrix}$$

and

$$\mathbf{X}^{(l)}=\begin{bmatrix}x_1^{(l)}(t_1)&\cdots&x_p^{(l)}(t_1)\\x_1^{(l)}(t_2)&\cdots&x_p^{(l)}(t_2)\\\vdots&\ddots&\vdots\\x_1^{(l)}(t_{n_l-1})&\cdots&x_p^{(l)}(t_{n_l-1})\end{bmatrix}.$$

Using the result of a previous study (Shimamura *et al*., 2009), an unbiased estimate of the degrees of freedom for the regularized weighted likelihood (3) in the Gaussian density (1) is given by

$$\hat{\mathrm{df}}_k^{(1)}=\mathrm{tr}\left[\tilde{\mathbf{X}}_{\hat{\mathcal{A}}_k^{(1)}}(\tilde{\mathbf{X}}_{\hat{\mathcal{A}}_k^{(1)}}^T\tilde{\mathbf{X}}_{\hat{\mathcal{A}}_k^{(1)}}+\lambda_2^{(1)}\mathbf{I})^{-1}\tilde{\mathbf{X}}_{\hat{\mathcal{A}}_k^{(1)}}^T\right], \tag{6}$$

where $\hat{\mathcal{A}}_k^{(1)}=\{j;\hat{b}_{j,k}^{(1)}\neq 0\}$, $\tilde{\mathbf{X}}_{\hat{\mathcal{A}}_k^{(1)}}$ is the submatrix of $\tilde{\mathbf{X}}$ whose columns correspond to indices of non-zero estimators $\hat{\mathcal{A}}_k^{(1)}$ and $\mathbf{I}$ is the identity matrix. Substituting the total number of estimated parameters in the bias-corrected weighted Akaike information criterion proposed by Hens *et al*. (2006) by the degrees of freedom (6), we obtain a modified version of the bias-corrected weighted Akaike information criterion (mWAICc$_k$)

$$\mathrm{mWAICc}_k^{(1)}=-2\hat{\mathrm{WL}}_k^{(1)}+\frac{2N_k^{(1)}(\hat{\mathrm{df}}_k^{(1)}+1)}{N_k^{(1)}-\hat{\mathrm{df}}_k^{(1)}-2}, \tag{7}$$

where $N_k^{(1)}=\sum_{l=1}^{m}\rho_k^{(1,l)}(n_l-1)$ and $\hat{\mathrm{WL}}_k^{(1)}$ is the weighted log-likelihood (2) with the parameters $\hat{b}_{1,k}^{(1)},\ldots,\hat{b}_{p,k}^{(1)}$ estimated by the regularized weighted log-likelihood (3) and $\hat{\sigma}_k^{(1)}$. The estimation of $\hat{\sigma}_k^{(1)}$ is described in Supplementary Material 1. In a practical situation, we choose the model with the smallest $\mathrm{mWAICc}_k^{(1)}$ among the candidate models.

2.5 Learning algorithm

In this section, we propose iterative algorithms for inferring gene networks from time series gene expression data based on RW-RENET; an iterative learning algorithm for inferring a gene network under the l-th condition and an iterative learning algorithm for inferring multiple gene networks under m conditions. Supplementary Materials 1 and 2 describe the details of the two algorithms and their pseudocodes, respectively.

These algorithms automatically update $\rho_k^{(l,r)}$, $\lambda_{1,k}^{(l)}$, $\lambda_2^{(l)}$ and $w_{j,k}^{(l)}$ based on the modified version of the bias-corrected weighted Akaike information criterion (7) and estimate $\sigma_k^{(l)}$, $b_{0,k}^{(l)}$ and $b_{j,k}^{(l)}$ $(1\le j,k\le p$ and $1\le l,r\le m)$. Note that the algorithms do not guarantee an increase in the regularized weighted log-likelihood (3); however, they guarantee a decrease in the sum of $\mathrm{mWAICc}_k^{(l)}$, $\sum_{k=1}^{p}\mathrm{mWAICc}_k^{(l)}$. In these algorithms, the likelihood weights $\rho_k^{(l,r)}$ $(1\le l,r\le m$ and $1\le k\le p)$ are first initialized and then updated by $\mathrm{mWAICc}_k^{(l)}$. The selection of the initial values of $\rho_k^{(l,r)}$ is described in Supplementary Material 3.

3 REVERSE-ENGINEERING SYNTHETIC NETWORKS

To evaluate the performance of RW-RENET, we performed a comparison with other traditional reverse-engineering approaches with L_1-regularization, namely, lasso (Tibshirani, 1996) and elastic net (Zou and Hastie, 2005), on simulated time series gene expression data from synthetic networks. It should be noted that elastic net is equivalent to RW-RENET when we apply an uniform importance weight $w_{j,k}^{(l)}=1$ and an uniform likelihood weight

$$\rho_k^{(l,r)}=\begin{cases}0, & l\neq r\\1, & l=r\end{cases},$$

and lasso is a special case of elastic net when we set $\lambda_2^{(l)}=0$. Therefore, we can estimate gene networks with these traditional approaches using the same algorithm based on RW-RENET. It should also be noted that lasso, elastic net and RW-RENET estimate gene networks with different numbers of edges. We then controlled the numbers of edges identified using lasso and elastic net to be equal to that using RW-RENET by deleting all the edges whose $|\hat{b}_{j,k}|$ lies below a given threshold T. Here, lasso and elastic

net using such a threshold are described as lasso-T and elastic net-T, respectively.

We generated two different scale-free graphs with $p=120$ nodes according to the Barabási–Albert model, where the distribution of the degree k follows the power law $P(k)\propto k^{-1.1}$. For each of the two graphs, five edges were randomly removed and 15 edges were randomly added, and we then generated eight synthetic networks $(N_1,\ldots,N_8)$, where N_1 and N_5 were the first two scale-free graphs, and $\{N_2,\ldots,N_4\}$ and $\{N_5,\ldots,N_8\}$ were generated from N_1 and N_5, respectively. Therefore, the structures of the four networks $\{N_{j+k};k=0,\ldots,3\}$ $(j=1$ or $2)$ were similar to each other but those of $\{N_1,\ldots,N_4\}$ were different from those of $\{N_5,\ldots,N_8\}$. The coefficient $b_{j,k}^{(l)}$ corresponding to the non-zero edge was simulated uniformly and randomly in the ranges $[-0.8$ to $-0.4]$ and $[0.4$ to $0.8]$. We simulated time series data with 10 time points from each of the eight networks based on the Gaussian process model (1). These settings are similar to the real data that we analyze in Section 4. We also simulated time series data of different sample size with $n=20,30,40$ and 50 in order to assess whether RW-RENET gained better performances as the number of observations was increased. We set a grid of parameters ($\{0.1,0.3,0.5\}$ $(n=10)$ or $\{0.5,1.0\}$ $(n=20,30,40$ and $50)$ for $\lambda_2^{(l)}$, $\{100,300,500\}$ for α and $[0.8:0.05:0.9]$ for β) and then infer gene networks using RW-RENET. The reason for setting different grid values for $\lambda_2^{(l)}$ is that it depends on the number of observations.

The performance of each algorithm was evaluated from two viewpoints: network selection and prediction for new observations. In terms of network selection, the performance was evaluated on how many overall, conserved and unconserved interactions were recovered. Here, conserved and unconserved interactions between N_i and N_j are defined by $\mathcal{C}_{i,j}=\{\mathcal{E}_i\cap\mathcal{E}_j\}$ and $\bar{\mathcal{C}}_{i,j}=\{(\mathcal{E}_i\cup\mathcal{E}_j)\setminus(\mathcal{E}_j\cap\mathcal{E}_i)\}$, where $\mathcal{E}_i$ and $\mathcal{E}_j$ indicate the edge sets of N_i and N_j, respectively. As a performance measure of the network inference, we used F-score (Meyer *et al.*, 2007b), which is the harmonic mean of the precision (PR) and the recall (RC), that is, $F=2\times\mathrm{PR}\times\mathrm{RC}/(\mathrm{PR}+\mathrm{RC})$. Here, PR and RC are calculated by $|\hat{\mathcal{S}}\cap\mathcal{S}|/|\hat{\mathcal{S}}|$ and $|\hat{\mathcal{S}}\cap\mathcal{S}|/|\mathcal{S}|$, where $\mathcal{S}$ is a true set of overall interactions $\mathcal{E}_i$, conserved interactions $\mathcal{C}_{i,j}$ and unconserved interactions $\bar{\mathcal{C}}_{i,j}$; $\hat{\mathcal{S}}$ is an inferred set of $\mathcal{S}$ by a reverse-engineering algorithm; and $|\mathcal{S}|$ represents the size of $\mathcal{S}$. The F-score is a performance measure because it tries to realize a balance between the PR and the RC, and it is high only if both PR and RC are high. In terms of prediction, the performance was assessed on whether the method can predict new observations from the true network rather than overfitting observations used for reverse-engineering networks. Here, 50 new observations were generated from each of the eight networks using different initial values of $X_j^{(l)}(t_0)$ and the prediction error of an estimator for l-th condition is defined using newly generated observations by

$$\mathrm{PE}^{(l)}=\sum_{k=1}^{p}\sum_{i=1}^{49}(\dot{x}_k^{(l)}(t_{i+1})-\sum_{j=1}^{p}\tilde{b}_{j,k}\dot{x}_j^{(l)}(t_i))^2,$$

where $\dot{x}_k^{(l)}(t_i)$ $(1\le i\le 50)$ are standardized observations with mean zero and unit variance that are not used for reverse-engineering networks and $\tilde{b}_{j,k}^{(l)}$ $(1\le j,k\le p)$ are the standardized estimated coefficients.

Figure 2a–d shows the performance of five methods (lasso, elastic net, lasso-T, elastic net-T and RW-RENET) when reverse-engineering eight synthetic networks where the number of time series data for each condition is 10 through 100 simulations. Each point indicates the average value of F-scores or prediction errors for 100 simulations. The F-scores for lasso-T and elastic net-T were larger than those for lasso and elastic net, respectively. Whereas, the prediction errors of lasso and elastic net were smaller than those of lasso-T and elastic net-T, respectively. We found that RW-RENET outperformed lasso, elastic net, lasso-T and elastic net-T in all cases. The values of F-score with RW-RENET represent increases of approximately 1.9, 3.6 and 1.7 times as compared with lasso, approximately 1.9, 3.5 and 1.6 times as compared with elastic net, approximately 1.9, 2.5 and 2.0 times as compared with lasso-T and approximately 1.9, 2.0 and 2.0 times as compared with elastic net-T in (a), (b) and (c), respectively. While, the values of prediction error with RW-RENET show decreases of $\sim$7, 6 , 21 and 21 % as compared to lasso, elastic net, lasso-T and elastic net-T in (d), respectively.

The lower four panels of Figure 2 show the performance of the five methods when the number of observations is increased by 10 from 10 to 50. As can be seen, all methods gain better performances with increased number of observations. Although the differences between RW-RENET and the other method in performance are smaller as increasing the number of observations, RW-RENET was superior to lasso, elastic net, lasso-T and elastic net-T in terms of both network selection and prediction for new observations. Therefore, the results of simulated data clearly indicate that RW-RENET has significant advantages in various situations when comparing multiple networks.

4 ANALYSIS OF MCF-7 BREAST CANCER CELLS STIMULATED WITH EPIDERMAL GROWTH FACTOR AND HEREGULIN

In this section, we investigated the following using time series microarray data of MCF-7 human breast cancer cells upon stimulation with four doses of two different ErbB ligands, epidermal growth factor (EGF) and heregulin (HRG) (Nagashima *et al.*, 2007): whether RW-RENET can recover known biological relationships and whether the conserved or non-conserved interactions estimated by RW-RENET support biological hypotheses and provide new findings. We considered eight dose conditions (stimulation with 0.1, 0.5, 1.0 and 10.0 nM of either EGF or HRG) and measured the expression values at eight time points (0, 5, 10, 15, 30, 45, 60 and 90 min) using Affymetrix GeneChip U133A-2. These datasets are available from GEO (Accession Number GSE6462). The expression values with 10.0 nM at 60 min were not observed, and were instead estimated by linear interpolation. These datasets were also normalized by faster cyclic loess (Ballman *et al.*, 2004).

4.1 Evaluating performances of inferring gene networks using the cell system markup language reaction database

The evaluation of the prediction performances of network inference under each condition requires topology information of real networks that are still unknown. Therefore, as a gold standard of real networks, we used biological interactions from the system markup language (cell system markup language, CSML) reaction

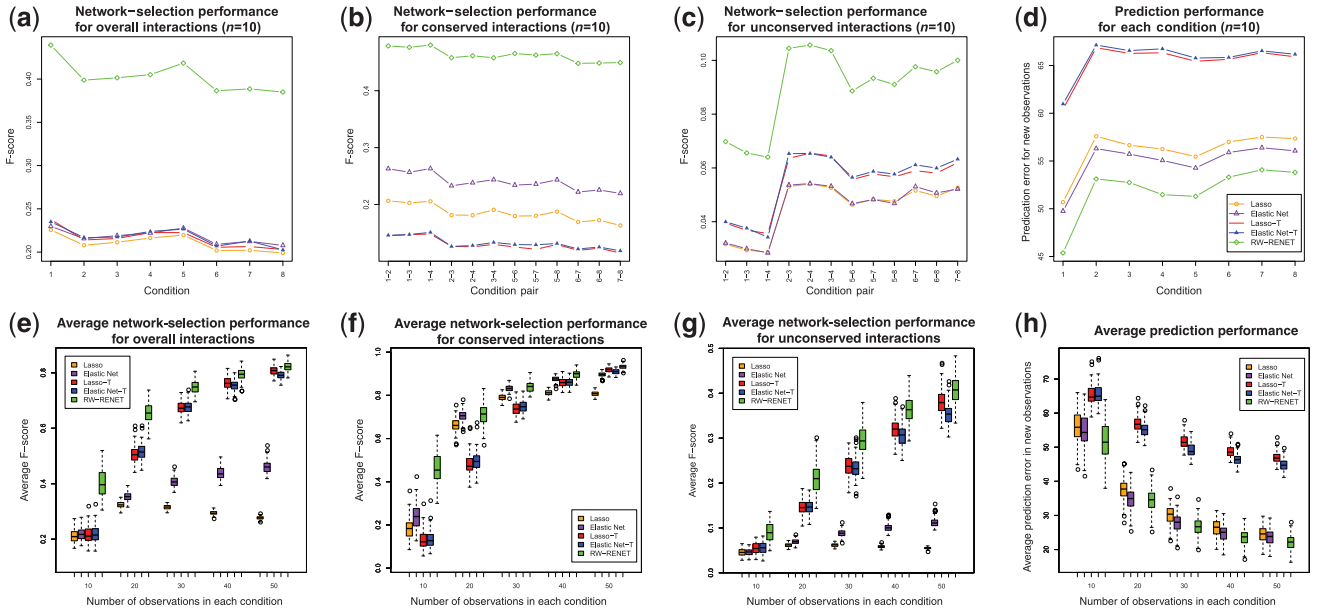

Fig. 2. Performance of five approaches (lasso, elastic net, lasso-T, elastic net-T and RW-RENET) when reverse-engineering eight synthetic networks through 100 simulations. F-scores for overall interactions (**a**), conserved interactions (**b**) and unconserved interactions (**c**), and prediction errors for each condition (**d**) using the different reverse-engineering methods when the number of observations for each condition is 10. The orange circle, purple triangle, red-filled circle , blue-filled triangle and green diamond marks indicate the results of lasso, elastic net, lasso-T, elastic net-T and RW-RENET, respectively. Each point stands for the average value of F-scores or prediction errors for 100 simulations. F-scores averaged for all combinations of overall interactions (**e**), conserved interactions (**f**) and unconserved interactions (**g**), and prediction errors averaged for eight conditions (**h**) when the number of observations is increased by 10 from 10 to 50.

database (Nagasaki *et al.*, 2008), which is a systematically reconstructed database of the manually curated high-quality pathway database TRANSPATH (Krull *et al.*, 2006), using CSML (http://www.csml.org/). The database contains information about more than 100 000 direct gene regulations and signal transduction regulations in mammals.

In addition to lasso and elastic net, we further performed a comparison with different network inference algorithms: mutual information networks with ARACNE (Basso *et al.*, 2005), CLR (Faith *et al.*, 2007) and MRNET (Meyer *et al.*, 2007a); graphical Gaussian model (GGM) with an empirical Bayes approach (Schäfer and Strimmer, 2006); VAR model with the James–Stein shrinkage (Opgen-Rhein and Strimmer, 2007); and dynamic Bayesian networks using first order conditional dependencies (Lébre, 2009). For estimating gene networks with these approaches, we used the following softwares: R-package minet (Meyer *et al.*, 2007b), R-package GeneNet (http://strimmerlab.org/), R-package G1DBN (Lébre, 2009) and R-code for the shrinkage estimation of the VAR network (http://strimmerlab.org/). Although mutual information approaches such as ARACNE, CLR and MRNET intend to infer static relationships between genes rather than time-lagged relationships, we considered time series data as static data and applied them. Following the same idea of Shimamura *et al.* (2009), we also considered a time-lagged data matrix whose (i,j)-element indicates $x_j^{(l)}(t_{i+1}) - x_j^{(l)}(t_i)$ and used it as an input of ARACNE, CLR and MRNET in order to increase their capabilities of recovering the time-lagged relationships. These modified versions of ARACNE, CLR and MRNET were described as ARACNE-D, CLR-D and MRNET-D, respectively. To calculate mutual information for inferring mutual networks, a Spearman-based mutual information was used. Here, mutual information approaches and GGM infer gene networks as undirected graphs, while the other approaches infer gene networks as directed graphs. To compare different types of graphs, we converted the real and the inferred directed networks to undirected ones and then performed comparisons. Although VAR model using the James–Stein shrinkage, dynamic Bayesian networks using G1DBN and Gaussian process model using lasso, elastic net and RW-RENET can identify self-feedback loops, these inferred self-feedback loops were not used to ensure a fair comparison between algorithms.

Nagashima *et al.* (2007) identified 254 probes corresponding to 240 genes as EGF- and HRG-regulated genes. Among the 240 genes, each of the 120 genes has no connection with the other 239 genes within a path length of 2 in the CSML reaction database, and therefore these 120 genes were removed. As a result, all the methods were run on time series expression profiles with 119 probes corresponding to the remaining 120 genes for inferring gene networks under eight conditions. We calculated the shortest path matrix among all combinations of the 120 genes, and 62 direct interactions except for 113 self-loops and 1072 interactions mediated through one gene (shortest path = 2) were considered as true interactions of the gold standard model.

It should be noted that the results are not perfect owing to the limited knowledge of the real network (Bansal *et al.*, 2006). Thus, the performance of each algorithm was assessed on the number of true positives (TP) and the PR (the proportion of the TP against all

Table 1. Results of the application of the 12 reverse-engineering algorithms (ARACNE, CLR, MRNET, ARACNE-D, CLR-D, MRNET-D, GGM, VAR model with the James-Stein shrinkage (VAR-JS), G1DBN, lasso, elastic net and RW-RENET) to the experimental data of MCF-7 breast cancer cells under eight biological conditions

	Dataset															
	Dose level of EGF								Dose level of HRG							
	0.1 nM		0.5 nM		1.0 nM		10.0 nM		0.1 nM		0.5 nM		1.0 nM		10.0 nM	
Methods	TP	PR	TP	PR	TP	PR	TP	PR	TP	PR	TP	PR	TP	PR	TP	PR
RAND	7.37	0.09	8.36	0.09	8.35	0.09	7.37	0.09	9.26	0.09	9.44	0.09	10.60	0.09	10.33	0.09
ARACNE	7	0.09	10	0.11	**14**	**0.15**	**15**	**0.18**	**17**	**0.17**	13	0.12	13	0.11	14	0.12
CLR	5	0.06	9	0.10	10	0.11	10	0.12	**16**	**0.16**	11	0.10	9	0.08	11	0.10
MRNET	6	0.07	10	0.11	**14**	**0.15**	10	0.12	**19**	**0.18**	10	0.10	7	0.06	**17**	**0.15**
ARACNE-D	8	0.10	13	0.14	12	0.13	6	0.07	13	0.13	6	0.06	11	0.09	11	0.10
CLR-D	7	0.09	6	0.06	**15**	**0.16**	8	0.10	12	0.12	8	0.08	13	0.11	9	0.08
MRNET-D	6	0.07	7	0.08	12	0.13	5	0.06	13	0.13	9	0.09	9	0.08	11	0.10
GGM	9	0.11	7	0.08	6	0.06	8	0.10	11	0.11	9	0.09	4	0.03	12	0.10
VAR-JS	12	0.15	8	0.09	9	0.10	12	0.15	8	0.08	9	0.09	6	0.05	15	0.13
G1DBN	6	0.07	9	0.10	8	0.09	12	0.15	7	0.07	**15**	**0.14**	10	0.08	**21**	**0.18**
Lasso	**14**	**0.17**	11	0.12	10	0.11	7	0.09	12	0.12	14	0.13	**26**	**0.22**	**20**	**0.17**
Elastic Net	12	0.15	9	0.10	11	0.12	9	0.11	9	0.09	12	0.11	**21**	**0.18**	**21**	**0.18**
RW-RENET	**13**	**0.16**	13	0.14	12	0.13	**14**	**0.17**	13	0.13	**19**	**0.18**	**26**	**0.22**	**23**	**0.20**

The RAND refers to the expected performance of the algorithm that randomly selects the same number of gene–gene pairs estimated by RW-RENET through 1 000 000 simulations. Numbers in bold indicate the algorithms that performed significantly better (P-value < 0.05) than the random model that selected the same number of significant edges randomly. The P-value was calculated by calculating a null distribution based on the random model through 1 000 000 simulations.

the identified edges) that are relative criteria. We chose the thresholds for the other methods so that the numbers of interactions inferred by the other algorithms are equal to that by RW-RENET. We also note that, since all the methods have the same numbers of inferred interactions, the largest TP means the largest . The setting of the tuning parameters of RW-RENET was also as same as that described in Section 3.

As a result, the proposed method inferred eight gene networks under stimulation with different dose levels of EGF (0.1, 0.5, 1.0 and 10.0 nM) and HRG (0.1, 0.5, 1.0 and 10.0 nM), which include 82, 94, 93, 82, 103, 105, 119 and 115 directed edges, respectively. Supplementary Figures 1–4 illustrate the four inferred gene networks under EGF stimulation and Supplementary Figures 5–8 illustrate the four inferred gene networks under HRG stimulation, respectively. After converting the direct graphs to undirected ones, the numbers of undirected edges were 82, 93, 93, 82, 103, 105, 118 and 115, respectively.

Table 1 lists the results of the analysis for the eight time series datasets. In Table 1, RAND refers to the expected performance of the algorithm that randomly selects the same number of gene–gene pairs as that of RW-RENET through 1 000 000 simulations. The results can be summarized as follows: ARACNE, CLR, MRNET, ARACNE-D, CLR-D, MRNET-D, GGM, VAR-JS and G1DBN were better than random selection for 7, 6, 6, 6, 4, 3, 4, 4 and 4 datasets, respectively. Interestingly, we found that the usual ARACNE, CLR and MRNET were better than their time-lagged versions.

All algorithms expect for ARACNE and MRNET have a small number of P-values <0.05 for the hypothesis that the difference in terms of the number of TP was significant as compared to that with random selection. In contrast, L_1-regularization approaches with lasso, elastic net and RW-RENET were better than random selection for 7, 7 and 8 datasets, respectively. In particular, RW-RENET outperformed lasso and elastic net in 7 datasets and it had the largest number of P-values <0.05 among all algorithms (five P-values <0.05). These results indicate that the proposed method could recover the known biological interactions more stably than the existing algorithms.

4.2 Comparing similarities between eight conditions

We investigated whether we can support recent biological hypotheses by comparing the conserved interactions between varying conditions. Nagashima *et al*. (2007) hypothesized that transcription that is directly modulated by upstream pathways should be controlled in a dose-dependent manner of EGF and HRG. Among the 120 genes, 1 gene was regulated by only EGF, 78 by only HRG, and 41 by both EGF and HRG. If their hypothesis is correct, we expect that a certain amount of interactions in gene networks for different dose levels of either EGF or HRG would be conserved. We calculated how many edges were conserved between the two networks for all combinations of the eight biological conditions.

Figure 3 shows 2D histograms that indicate the results of these calculations for each of the 12 inference algorithms. For clear visualization, the numbers of conserved interactions between the same conditions were set to 0. The complete details of the numbers of conserved interactions for the 12 algorithms are provided in Tables 1–12 of Supplementary Material, respectively. Clearly, all methods except for RW-RENET identified a small amount of conserved interactions, contrary to the hypothesis of

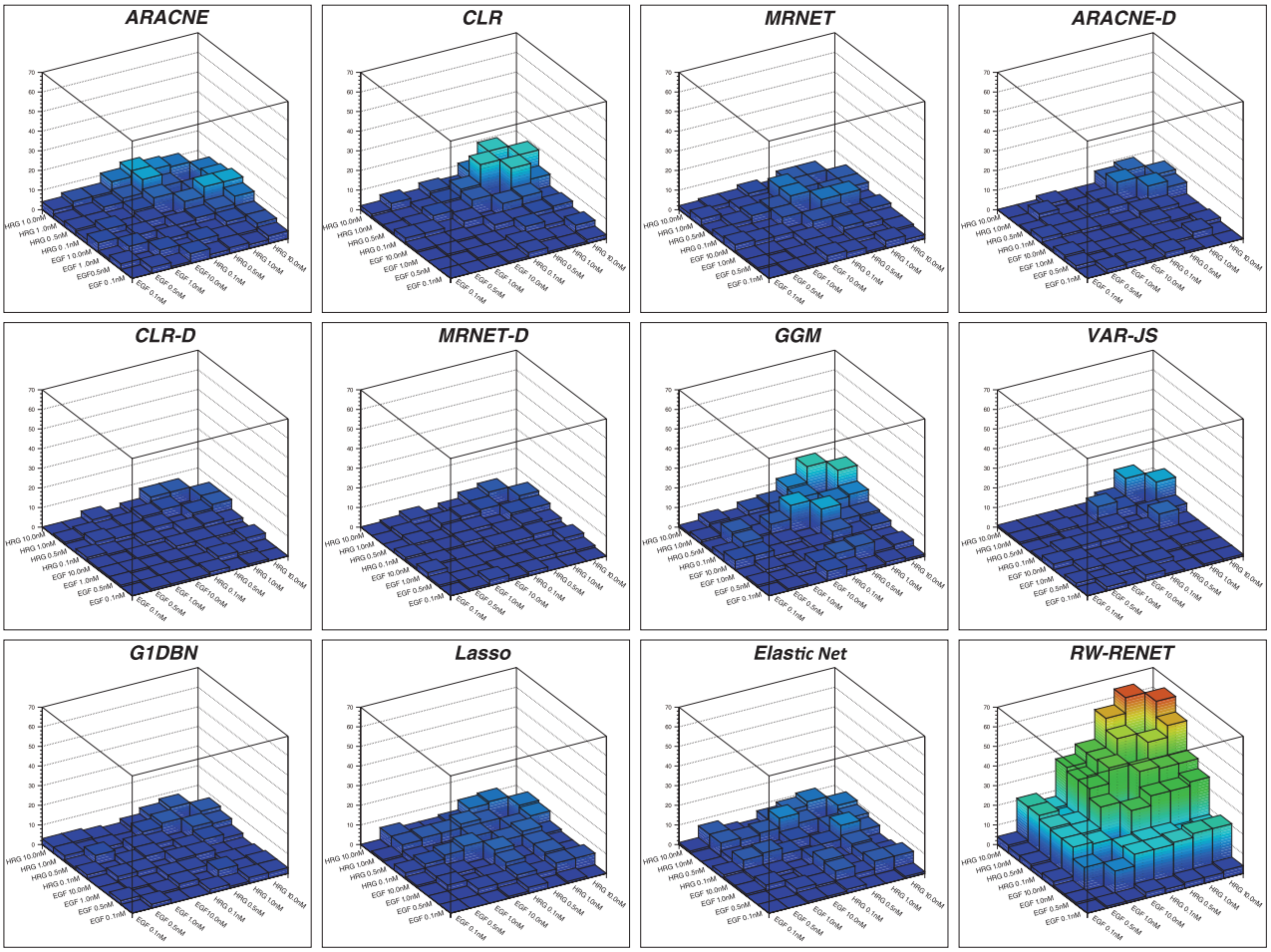

Fig. 3. The 2D histograms showing how many edges were conserved between the inferred networks for each of the 12 inference algorithms (ARACNE, CLR, MRNET, ARACNE-D, CLR-D, MRNET-D, GGM, VAR with the James-Stein Shrinkage (VAR-JS), G1DBN, lasso, elastic net and RW-RENET) under all combinations of the eight biological conditions (stimulation of EGF 0.1 , 0.5, 1.0 and 10.0 nM, and HRG 0.1, 0.5, 1.0 and 10.0 nM).

Nagashima *et al*. (2007). There is an extremely high possibility that the inferred networks include a large number of spurious interactions due to the small number of time points. This also implies that it is quite difficult to compare the similarity and dissimilarity between the inferred networks under different conditions using these existing methods. However, the proposed method succeeded in refining the performance of identifying the conserved interactions by taking advantage of the increasing numbers of relevant samples with the regularized weighted log-likelihood estimation. For example, in comparison between stimulation with 0.5 and 1.0 nM HRG, ARACNE, CLR, MRNET, ARACNE-D, CLR-D, MRNET-D, GGM, VAR-JS, G1DBN, LASSO and ENET identified 8, 19, 7, 10, 5, 5, 11, 5, 7, 7 and 7 conserved interactions including 1, 3, 1, 1, 1, 1, 0, 0, 1, 1 and 0 known biological interactions in the CSML reaction database, respectively. Whereas RW-RENET identified 50 conserved interactions with 10 known interactions. Supplementary Tables 13–24 list the number of known interactions in the CSML reaction database that were included in the conserved interactions estimated by the 12 algorithms, respectively. Therefore, only the result of a comparison between the conserved interactions based on RW-RENET in Figure 3 supports the hypothesis of Nagashima *et al*. (2007), at least for the HRG-induced networks.

4.3 Comparing differences between EGF and HRG stimulations

We also investigated whether we can provide a new biological hypothesis by comparing the unconserved interactions between EGF and HRG stimulations. We calculated *out-degree*, the number of directed edges that start from each node, for each inferred network using RW-RENET. Genes with a large out-degree are called *hub genes*, and they are considered to be essential in complex real networks (Albert and Barabási, 2002). Supplementary Table 24 lists out-degrees of the 120 genes in the eight gene networks under stimulation with different dosages of EGF and HRG. We found that *SIK1*, *JUN*, *KLF6*, *EGR2*, *DUSP1* and *MAP3K8* have different out degrees between EGF 10.0 nM and HRG 10.0 nM illustrated in Figure 4; 5, 0, 1, 22, 1 and 0 outdegrees for EGF 10 nM as compared with 27, 8, 9, 28, 6 and 4 outdegrees for HRG 10 nM. Interestingly, these six genes have *FOS* as upstream genes.

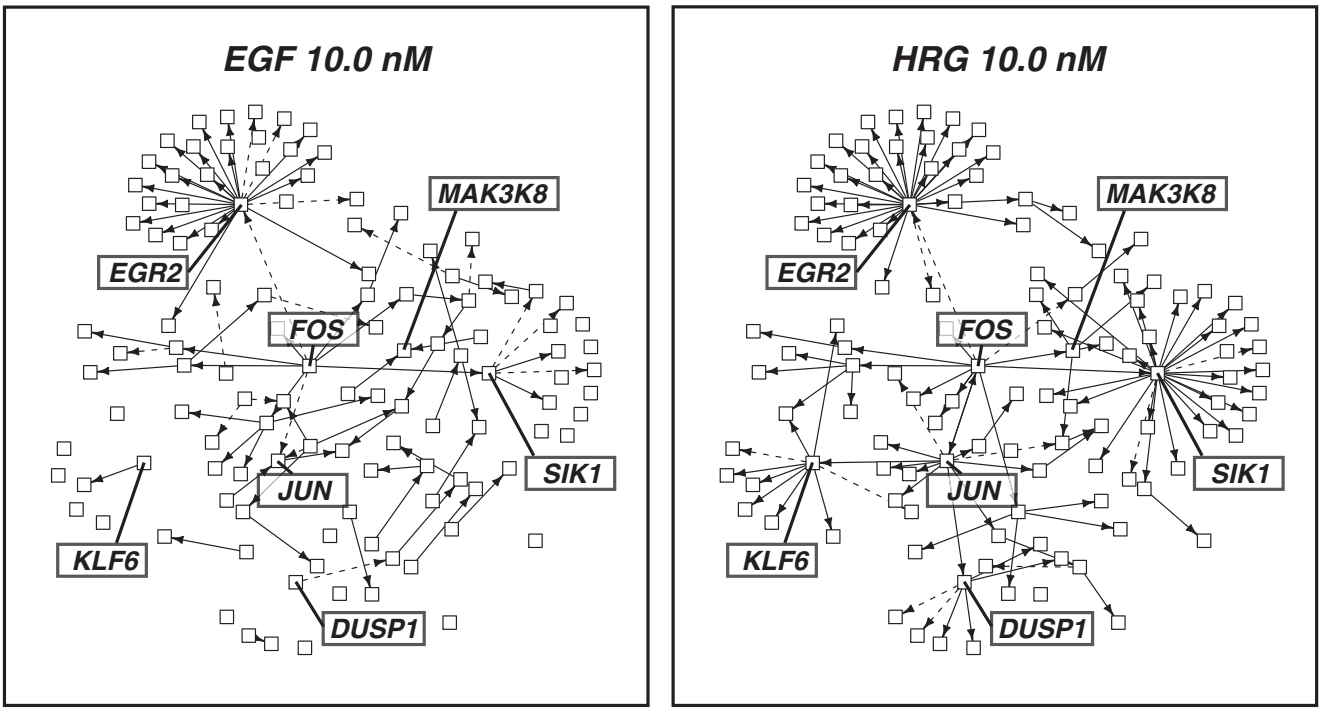

Fig. 4. Differences between the EGF- and the HRG-stimulated gene networks.

It is known that EGF and HRG stimulations lead to proliferation and differentiation of MCF-7 breast cancer cells, respectively (Nagashima *et al*., 2007). Our comparative analysis based on RW-RENET produced a new hypothesis that *FOS* might be a key gene that is strongly associated with this ligand-specific cell fate determination. Again, only the result of network comparison based on our method supported the recent biological hypothesis and provided a new hypothesis.

5 CONCLUSION

We addressed the problem of inferring and comparing multiple gene networks between varying conditions from time series gene expression data. Numerous approaches have been developed using mathematical models in order to reverse-engineer a gene network under a condition of interest from time series data. However, no studies have addressed the problem of gene network inference from the viewpoint of network comparison. We proposed a novel integrated approach for reverse-engineering multiple gene networks that is intended for transcriptomic network comparison. A state-of-the-art estimation method called RW-RENET automatically infers conserved interactions from samples under all the relevant information and unconserved interactions from samples under each condition with a similarity based on the comparison between the estimators; this method achieves higher PR and RC than the earlier methods in gene network inference and comparison. We also supported biological hypotheses and provided new biological findings in experimental data of MCF-7 breast cancer cells under EGF and HRG stimulations through comparative network analysis using RW-RENET.

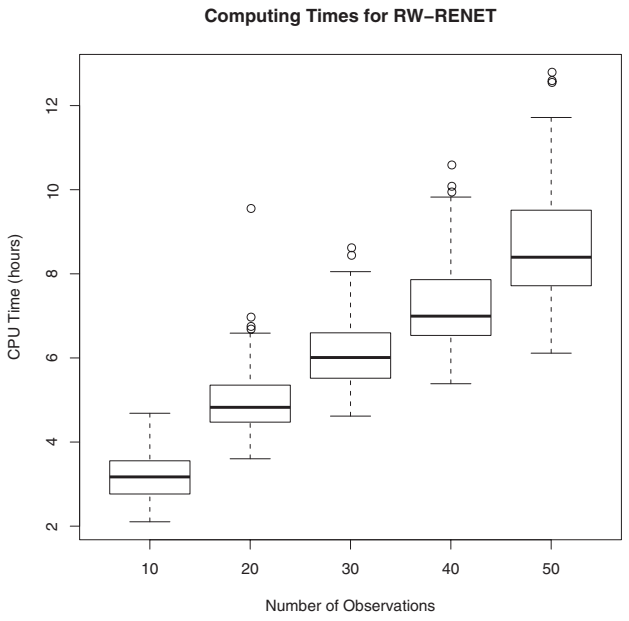

Fig. 5. Computing times for RW-RENET when reverse-engineering eight networks with 120 genes using one core of Intel Quad-Core Xeon Processor E5450 (3.0 GHz).

We would like to discuss about computational time of RW-RENET. Figure 5 presents the total calculation times using one core of Intel Quad-Core Xeon Processor E5450 (3.0 GHz) when reverse-engineering eight synthetic networks where we fixed the number

of variables to 120 and varied the number of observations from 10 to 50 by 10. We observe that the time complexity is nearly linear (more precisely it is more linear because of matrix calculations) with respect to the number of observations. For the fixed sample size, the time complexity should be also linear with respect to the number of variables (genes).

A limitation of our model comes from the assumption that time lag is stationary (time points are equidistant). This is not the case for many real applications where the sampling time points are unequally spaced and vary with different conditions. However, we believe that this problem can be solved by using a state space representation (Kojima *et al*., 2009), which is one of our future researches. Another limitation of our method comes from the high computational cost. As described above, the RW-RENET algorithm using single CPU is applicable only to medium-sized networks with several 100 genes. To estimate larger gene networks based on larger samples, super-computers should be useful, because the computation of RW-RENET can be performed completely in parallel for every gene.

With the rapid correction of high-dimensional and heterogeneous data of complex cellular dynamic processes under various types of biological conditions, we expect that the development of approaches for comparative network analysis is of increasing importance. We believe that the proposed method will serve as a valuable and timely contribution for better understanding of the differences between complex cellular dynamics and molecular diversities, and for easy management of these large and complex datasets.

ACKNOWLEDGEMENTS

We thank the reviewers for helpful and constructive comments. The computational resource was also provided by the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo.

Funding: Next-Generation Supercomputer Project, RIKEN Computational Science Research Program, MEXT, Japan.

Conflict of Interest: none declared.

REFERENCES

Albert,R. and Barabási,A.-L. (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.*, **74**, 47–97.

Ballman,K.V. *et al*. (2004) Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics*, **20**, 2778–2786.

Bansal,M. *et al*. (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, **22**, 815–822.

Basso,K. *et al*. (2005) Reverse engineering of regulatory networks in human B cells. *Nat. Genet.*, **37**, 382–390.

Cantone,I. *et al*. (2009) A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell*, **137**, 172–181.

Chen,T. *et al*. (1999) Modeling gene expression with differential equations. *Pac. Symp. Biocomput.*, **4**, 29–40.

Chickering,D.M. *et al*. (2004) Large-sample learning of Bayesian networks is NP-hard. *J. Mach. Learn. Res.*, **5**, 1287–1330.

Faith,J.J. *et al*. (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, **5**, e8.

Hens,H. *et al*. (2006) Model selection for incomplete and design-based samples. *Stat. Med.*, **25**, 2502–2520.

Hirose,O. *et al*. (2008) Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics*, **24**, 932–942.

Hu,F. (1994) Relevance weighted smoothing and a new bootstrap method. PhD Thesis, Department of Statistics, the University of British Columbia, British Columbia, Canada.

Hu,F. and Zidek,J. (2002) The weighted likelihood. *Can. J. Stat.*, **30**, 347–371.

Kojima,K. *et al*. (2009) A state space representation of VAR models with sparse learning for dynamic gene networks. *Genome Inform.*, **22**, 56–68.

Krull,M. *et al*. (2006) TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acid Res.*, **34**(Database issue), 546–551.

Lébre,S. (2009) Inferring dynamic genetic networks with low order independencies. *Stat. Appl. Genet. Mol. Biol.*, **8**, 9.

Meyer,P.E. *et al*. (2007a) Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J. Bioinform. Syst. Biol.*, **2007**, 79879.

Meyer,P.E. *et al*. (2007b) minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics*, **9**, 461.

Nagasaki,N. *et al*. (2008) Systematic reconstruction of TRANSPATH data into cell system markup language. *BMC Syst. Biol.*, **2**, 53.

Nagashima,T. *et al*. (2007) Quantitative transcriptional control of ErbB receptor signaling undergoes graded to biphasic response for cell differentiation. *J. Biol. Chem.*, **282**, 40.

Opgen-Rhein,R., and Strimmer,K. (2007) From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.*, **1**, 37.

Perrin,B.E. *et al*. (2003) Gene networks inference using dynamic Bayesian networks. *Bioinformatics*, **19** (Suppl. 2), ii138–148.

Rangel,C. *et al*. (2004) Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics*, **20**, 1361–1372.

Schäfer,J., Strimmer,K. (2006) An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, **21**, 754–764.

Shimamura,T. *et al*. (2009) Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Syst. Biol.*, **3**, 41.

Tamada,Y. *et al*. (2009) Unraveling dynamic activities of autoacine pathways that control drug-response transcriptome networks. *Pac. Symp. Biocomput.*, **14**, 251–263.

Tibshirani,R. (1996) Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B*, **58**, 267–288.

Wang,S.X. (2001) Maximum weighted likelihood estimation. PhD Thesis, Department of Statistics, the University of British Columbia, British Columbia, Canada.

Zou,H. and Hastie,T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B*, **67**, 301–320.

Zou,H. *et al*. (2007) On the "degrees of freedom" of the lasso. *Ann. Statist.*, **35**, 2173–2192.