BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 6 2014, pages 852–859
doi:10.1093/bioinformatics/btt620

Systems biology

Advance Access publication October 31, 2013

Prediction and early diagnosis of complex diseases by edge-network

Xiangtian Yu[1,2], Guojun Li[2] and Luonan Chen[1,3,*]

[1]Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, [2]School of Mathematics, Shandong University, Jinan 250100, China and [3]Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: In this article, we develop a novel edge-based network i.e. edge-network, to detect early signals of diseases by identifying the corresponding edge-biomarkers with their dynamical network biomarker score from dynamical network biomarkers. Specifically, we derive an edge-network based on the second-order statistics representation of gene expression profiles, which is able to accurately represent the stochastic dynamics of the original biological system (with Gaussian distribution assumption) by combining with the traditional node-network, which is based only on the first-order statistics representation of the noisy data. In other words, we show that the stochastic network of a biological system can be described by the integration of its node-network and its edge-network in an accurate manner.

Results: By applying edge-network analysis to gene expressions of healthy adults within live influenza experiment sampling at time points before the appearance of infection symptoms, we identified the edge-biomarkers (80 edges with 22 densely connected genes) discovered in edge-networks corresponding to symptomatic adults, which were used to predict the subsequent outcomes of influenza infection. In particular, we not only correctly predict the final infection outcome of each individual at an early time point before his/her clinic symptom but also reveal the key molecules during the disease progression. The prediction accuracy achieves ~90% under the leave-one-out cross-validation. Furthermore, we demonstrate the superiority of our method on disease classification and predication by comparing with the conventional node-biomarkers. Our edge-network analysis not only opens a new way to understand pathogenesis at a network level due to the new representation for a stochastic network, but also provides a powerful tool to make the early diagnosis of diseases.

Contact: lnchen@sibs.ac.cn

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

Received on September 12, 2013; revised on October 12, 2013; accepted on October 21, 2013

1 INTRODUCTION

Time-course data are increasingly used to study dynamical biological processes or disease progression, like drug treatment or virus infection that evolves in a stochastic and temporal fashion (Wu and Wu, 2013). Instead of a snapshot of gene expression, the time-course gene expression over several continuous time points, allows investigators to study or even predict dynamic behaviors of a biological system (Huang *et al*., 2011; Wu and Wu, 2013). Based on high-throughput time-course data, although there are extensive works to identify molecular biomarkers for diagnosing complex diseases, it is strongly demanded to develop a systematical framework by exploiting such dynamical and stochastic information to predict early signals of disease states and also their occurrence times from both theoretical and computational viewpoint, which is also crucial to achieve predictive and preventive medicine (Liu *et al*., 2013a, b). In particular, molecular network is widely used to analyze the molecular response (Oates and Mukherjee, 2012; Zhi *et al*., 2013) as well as biomarkers for distinguishing disease and normal samples. Traditionally, a molecular network with node (e.g. gene or protein) as basic element, i.e. node-network, is constructed mainly in the following two ways: one is to extract the conditional existence of known molecular interactions, which consist of a subnetwork induced from a given background network (Chuang *et al*., 2007), such as Weighted Correlation Network Analysis (Zhang and Horvath, 2005); the other is to directly infer *de novo* molecular associations, which represent a significant topological structure connecting molecules (He *et al*., 2012; Margolin *et al*., 2006; Zhang *et al*., 2012), such as ARACNE (Margolin *et al*., 2006), InferGRN (Wang *et al*., 2006) and NARROMI (Zhang *et al*., 2013). This network can represent associations or interactions among molecules but cannot directly describe the stochastic dynamics of a biological system.

Generally, a biological system at a molecular level can be described by stochastic dynamics modeled by a master equation (Chen *et al*., 2010; Van Kampen, 1992). As shown in Figure 1, with the linearization and Gaussian distribution assumption, the system can be exactly expressed by two sets of equations, i.e. one for the mean vector of molecules (used in first-order statistics representation) and another for the covariance matrix of molecules (used in second-order statistics representation). However, the traditional molecular network, e.g. gene network or protein interaction network, is based only on the equations of mean vector rather than the equations of covariance matrix, e.g. a set of linear equations for molecular concentrations, which cannot represent whole stochastic dynamics of the original system (i.e. it is the representation of a biological system without any stochastic fluctuation or with zero noise). Different from

*To whom correspondence should be addressed.

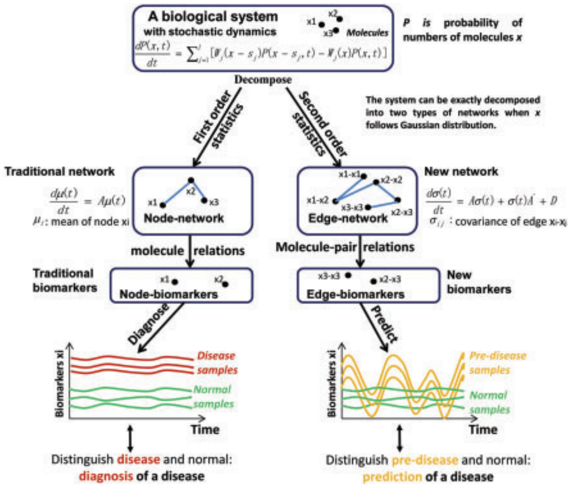
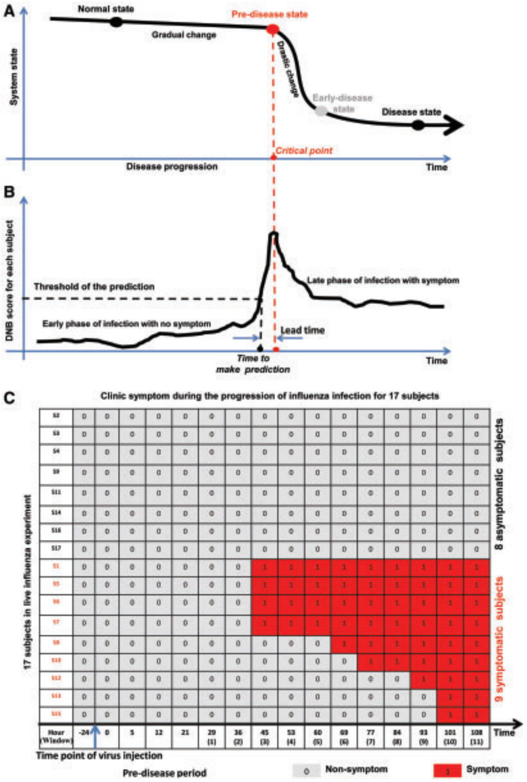

A biological system
with stochastic dynamics
x1
x2
x3
Molecules
$\frac{dP(x,t)}{dt} = \sum_{j=1}^{J}[W_j(x-s_j)P(x-s_j,t) - W_j(x)P(x,t)]$
P is probability of numbers of molecules x
Decompose
First order statistics
Second order statistics
The system can be exactly decomposed into two types of networks when x follows Gaussian distribution.
Traditional network
$\frac{d\mu(t)}{dt} = A\mu(t)$
μ_i: mean of node xi
Node-network
New network
x1-x1
x2-x2
x1-x2
x3-x3
x2-x3
Edge-network
$\frac{d\sigma(t)}{dt} = A\sigma(t) + \sigma(t)A' + D$
σ_{ij} : covariance of edge xi-xj
molecule relations
Molecule-pair relations
Traditional biomarkers
Node-biomarkers
Edge-biomarkers
New biomarkers
Diagnose
Predict
Biomarkers xi
Disease samples
Normal samples
Pre-disease samples
Time
Distinguish disease and normal: diagnosis of a disease
Distinguish pre-disease and normal: prediction of a disease

Fig. 1. Node-network and edge-network for a biological system. A biological system at a molecular level can be modeled as a master equation, where *P* is the probability of *x*; W is the propensity function; *x* is the numbers or concentrations of molecules following Gaussian distribution with a mean vector μ and a covariance matrix σ and *s* is the changes of molecules. By liberalizing the master equation, we have two sets of the equations, i.e. one is linear equations corresponding to the traditional molecular network or node-network, which is based only on the first-order statistics information or average values of molecules, and another is Lyapunov differential equations corresponding to our edge-network, which is based on the second-order statistics information or covariance between molecules. Theoretically, the information from those two-level statistics can fully recover the stochastic dynamics of the original system. The first-order statistics information or traditional node-network is used to distinguish the disease and normal samples for disease diagnosis by identifying molecular biomarkers or node biomarkers, whereas the second-order statistics information or edge-network is able to distinguish the predisease and normal samples by identifying edge biomarkers or DNBs, thereby achieving the early diagnosis or the prediction of the disease

those conventional node-networks, we propose a new edge-based network, i.e. edge-network, to exploit higher-order statistics information among molecules, where a node represents a pair of the connecting nodes, i.e. an edge in the traditional node-network. As indicated in Figure 1, the edge-network is based on the covariance matrix of molecules governed by Lyapunov differential equation (Ichikawa *et al.*, 2009). In an edge-network, a node is not a molecule but a pair of molecules (i.e. an edge), and a link represents the relationship between two molecule pairs (i.e. between two edges) rather than between two molecules as in a node-network. Clearly, an edge-network reflects the second-order statistics information, and therefore theoretically with Gaussian distribution assumption on each molecule' expression it is able to recover stochastic dynamics of the original biological system by combining with the node-network (first-order statistics).

Recent studies (Chen *et al.*, 2012; Li *et al.*, 2013; Liu *et al.*, 2012, 2013a) show that the second-order statistics information can be used to predict the predisease state (the state of an individual before the appearance of clinical symptom) and thus achieve the early diagnosis of a disease by detecting its dynamical network biomarker (DNB), in contrast to the molecular network



Hour (Window)	-24	0	5	12	21	29 (1)	36 (2)	45 (3)	53 (4)	60 (5)	69 (6)	77 (7)	84 (8)	93 (9)	101 (10)	108 (11)
S2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
S5	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
S6	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
S7	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
S8	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
S10	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
S12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
S13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
S15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Fig. 2. Disease progression, DNBs and time-course gene expression profiles of 17 healthy adults within live influenza experiment. (**A**) The progression of the influenza infection can be considered to have three stages, i.e. (i) normal stage with the gradual progression of the disease, (ii) predisease stage that is considered as the limit of the normal stage just before the symptom appears and (iii) disease or infection stage after the symptom appears. (**B**) Edge-biomarkers or DNBs are able to identify the predisease stage due to dynamical and higher-order statistics information, and therefore predict the outcome of the influenza infection before the symptom appears. (**C**) The biological time-course expression data contains 17 subjects challenged with influenza H3N2/Wisconsin, for which, 9 subjects are infected (Sx), whereas 8 subjects stay healthy (Asx) finally. Gene expression profiles were obtained and measured on whole peripheral blood drawn from all subjects at an interval of 8 h post inoculation (hpi) through 108 hpi. In all, 268 gene microarrays were obtained for all subjects at 16 time points including baseline (−24 hpi). For the purpose of the prediction, we only use the data with non-symptom to identify the edge-biomarkers

that is mainly used to identify molecular biomarkers or node-biomarkers for the diagnosis of a disease. Thus, one major advantage of the edge-network is its predictive power for early diagnosis of a disease, which can not only predict the future occurrence of a disease but also estimate the critical time when the change from a normal to a disease state happens. As shown in Figure 2A, the progression of the disease progression, e.g. influenza infection can be considered to have three stages (or states), i.e. (i) normal stage possibly with the gradual progression of the disease, (ii) predisease stage that is considered as the limit of the normal stage just before the disease symptom appears and (iii) disease or infection stage after the disease symptom appears (Chen *et al.*, 2012; Liu *et al.*, 2012, 2013a). Our edge-biomarkers derived from the edge-network are able to identify the predisease stage due to its dynamical and additional covariance information, and therefore predict the outcome of the complex disease (e.g. influenza infection) before the clinic symptom appears (Fig. 2B).

Specifically, by applying edge-network analysis to 268 gene expression profiles of 17 healthy adults within 16 time points across the whole live influenza experiment (Huang *et al.*, 2011), we discovered 80 edges (with 22 densely connected genes) involved in edge-networks of most symptomatic adults before the appearance of clinic symptom, which were used as edge-biomarkers to predict the subsequent outcomes of influenza infection for each individual. Our results indicated that those edge-biomarkers in this case have similar dynamical features to DNB (Chen *et al.*, 2012; Liu *et al.*, 2012, 2013a, 2013b). In particular, the results show that these edge-biomarkers can predict the outcomes of influenza infection with 90% accuracy under leave-one-out cross-validation (LOOCV), i.e. not only predict the final infection outcome of each individual or subject but also estimate the early time point of the subsequent infection. Furthermore, we compare the results with the conventional biomarkers and methods, and also conduct the functional analysis on the edge-biomarkers, which all demonstrate the superiority of our method on disease classification and prediction. We also investigate the molecular mechanism of the disease development after virus infection by analyzing the identified key molecules at the critical time points. In all, our edge-network analysis opens a new way to deeply understand disease progression, e.g. influenza virus-induced pathogenesis from dynamical features and higher-order statistics information of big biological data, and also provides a powerful tool to prevent disease occurrence or make the early diagnosis of a disease for each individual.

2 METHODS

We first describe the biological data used to study influenza infection; then provide the mathematical basis of edge-network; next display our computational method of edge-network analysis shown in Figures 1 and 2; finally illustrate the prediction results as well as the comparison between our edge-network and traditional node-network (i.e. molecular network). Note that an edge in this article means a pair of two connecting molecules.

2.1 Experimental data

The biological data GSE30550 (Huang *et al.*, 2011) contains 17 subjects (or adults) challenged with influenza H3N2/Wisconsin, for which, 9 subjects are actually infected (with the clinical infection symptom), whereas 8 subjects stay healthy (without the clinical infection symptom) finally. Gene expression profiles were obtained and measured on whole peripheral blood drawn from all subjects at an interval of 8 h post-inoculation (hpi) through 108 hpi. Totally, 268 gene microarrays were obtained for all subjects at 16 time points including baseline (24 h before subjects were injected with influenza virus, e.g. −24 hpi) (Huang *et al.*, 2011). Because our method is designed for predicting the influenza infection, we only use the predisease gene expression data (i.e. the time-course data before the appearance of clinical symptom of influenza infection or the data shown as Non-symptom in Fig. 2C) instead of the whole gene expression profiles (i.e. the data before and after the symptom appearance). Because there is no baseline information for subject-13, we did not take account of the baseline data of any subject. As shown in Figure 2C, we chose the predisease gene expression data for 17 subjects according to the clinical index provided in original article (Huang *et al.*, 2011). We divided subjects into two groups according to the clinical symptom chart based on the standardized symptom scoring (Dowling *et al.*, 1958): symptomatic (Sx) group with 9 subjects (subjects 1,5,6,7,8,10,12,13,15) and asymptomatic (Asx) group with 8 subjects (subjects 2,3,4,9,11,14,16,17).

(1) For symptomatic group, the lengths of used expression data for the predisease periods are different due to the different clinical outcomes of these pathogen subjects. For subjects 1, 5, 6 and 7, we all used the data before 39 h when they were diagnosed to have virus infection. Similarly, for subjects 8, 10, 12, 13 and 15, the data obtained before 62, 74, 86, 98 and 98 h were used, respectively.

(2) For asymptomatic group, we used the data before 39 h, i.e. the first six time points without baseline point, although we can use more data. This is because we expect to get the accurate prediction of influenza infection by our method at the earliest time before a medical doctor can give the clinical diagnostic results.

2.2 Theoretical basis of edge-network

A biological system at a molecular level can generally be described by stochastic dynamics, which can be modeled by a master equation (Chen *et al.*, 2010; Van Kampen, 1992), i.e. Equation (1).

$$\frac{dP(x(t),t)}{dt}=\sum\nolimits_{j=1}^{m}\left[W_j(x(t)-s_j)P(x(t)-s_j,t)-W_j(x(t))P(x(t),t)\right] \quad (1)$$

where the system is composed of n molecular species $x=(x_1,\ldots,x_n)$ and m reactions, and $P(x,t)$ is the time evolution of the probability in state x at time t. s_{ij} is the change of x_i by the reaction-j with $s_j=(s_{1j},\ldots,s_{nj})$, and W_j is the propensity function, which is the transition probability. Note that x_i is the number of molecule-i. As shown in Figure 1, with the linearization and Gaussian distribution assumption of Equation (1), the biological system can be exactly expressed by two sets of equations (Ichikawa *et al.*, 2009), i.e. Equation (2) for the mean vector of molecules (or the first-order statistics information) and Equation (3) for the covariance matrix of molecules (or the second-order statistics information) (Ichikawa *et al.*, 2009).

$$\text{Node}-\text{network dynamics}:\frac{d\mu(t)}{dt}=A(t)\mu(t) \quad (2)$$

$$\text{Edge}-\text{network dynamics}:\frac{d\sigma(t)}{dt}=A(t)\sigma(t)+\sigma(t)A'(t)+D(t) \quad (3)$$

where μ is the mean vector of x, and σ is an $n\times n$ covariance matrix of x. Clearly, Equation (2) is linear differential equations and represents the traditional molecular network (e.g. gene network or protein interaction network) or so-called node-network, where A is the network connection matrix or adjacent matrix and a node is a molecule (i.e. μ_i), which is based only on first-order statistics information. *A'* is the transpose of A. On the other hand, Equation (3) is the Lyapunov differential equations, which are based on the second-order statistics information. It constructs the covariance network or edge-network, where a node is a pair of molecules (i.e. σ_{ij}) in contrast to a molecule in a node-network, and a link in an edge-network describes the relationships between two molecule pairs in contrast to two molecules in a node-network. Clearly, Equations (2 and 3) can fully recover the stochastic dynamics of the biological system with an appropriate assumption (Ichikawa *et al.*, 2009). For instance, by simulating Equations (2 and 3), we can obtain stochastic dynamics of the original biological system Equation (1). Also Equations (2 and 3) are independent of each other in the form, i.e. node-network and edge-network can be analyzed separately. Although the adjacent matrix of the edge-network can be obtained by using the elements of A shown in Equation (3), it can be also numerically inferred by the covariance matrix introduced in Equation (5) belows.

A node-network (traditional molecular network with node μ_i or x_i) or adjacent matrix A in Equation (2) is generally inferred by analyzing the correlations of molecules based on expression data. On the other hand, in theory, the edge-network (with node σ_{ij}) can also be directly constructed from high-throughput data based on its definition and correlations. Specifically, the links of an edge-network could be approximately inferred by using the corresponding correlations between molecule pairs. Generally, a link in an edge-network is a fourth-order statistics (or the fourth-order moment) due to its relationship between two molecule pairs. The Pearson correlation coefficient (PCC) is a second-order statistics, which reflects the relationship between two molecules, i.e.

$$PCC(x_i, x_j) = \frac{C(x_i, x_j)}{\sqrt{V(x_i)V(x_j)}} = \frac{E((x_i - \mu_i)(x_j - \mu_j))}{\sqrt{E(x_i - \mu_i)^2 E(x_j - \mu_j)^2}} \tag{4}$$

which is the basis of many methods to construct the traditional node-network. On the other hand, for four molecules, i.e. two molecule pairs, we define the similar measurement as follows:

$$PCC(x_i, x_j, x_k, x_l) = \frac{C(x_i, x_j, x_k, x_l)}{\sqrt[4]{V(x_i)V(x_j)V(x_k)V(x_l)}}$$
$$= \frac{E((x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_j)(x_l - \mu_l))}{\sqrt[4]{E(x_i - \mu_i)^4 E(x_j - \mu_j)^4 E(x_k - \mu_k)^4 E(x_l - \mu_l)^4}}$$

According to Isserlis' theorem, we have

$$\begin{aligned}
&E((x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)) \\
&= E((x_i - \mu_i)(x_j - \mu_j))E((x_k - \mu_k)(x_l - \mu_l)) \\
&+ E((x_i - \mu_i)(x_k - \mu_k))E((x_j - \mu_j)(x_l - \mu_l)) \\
&+ E((x_i - \mu_i)(x_l - \mu_l))E((x_j - \mu_j)(x_k - \mu_k)) \\
&= C(x_i, x_j)C(x_k, x_l) + C(x_i, x_k)C(x_j, x_l) + C(x_i, x_l)C(x_j, x_k)
\end{aligned}$$

Because of $E(x_i - \mu_i)^4 = 3V(x_i)^2$, then the fourth-order correlation coefficient is

$$\begin{aligned}
&PCC(x_i, x_j, x_k, x_l) \\
&= \frac{C(x_i, x_j)C(x_k, x_l) + C(x_i, x_k)C(x_j, x_l) + C(x_i, x_l)C(x_j, x_{k)}}{\sqrt[3]{V(x_i)V(x_j)V(x_k)V(x_l)}}
\end{aligned} \tag{5}$$

where $C(x_i, x_j) = E((x_i - \mu_i)(x_j - \mu_j))$ and $V(x_i) = E(x_i - \mu_i)^2$. Clearly, we can calculate the correlation between two molecule pairs i.e. $x_i - x_j$ and $x_k - x_l$ from biological data based on the aforementioned Equation (5) provided that there are time-course data or multisample data, and then, we can construct the network of molecule pairs, e.g. edge-network, by Equation (5).

2.3 Computational algorithm of edge-network analysis

Based on the aforementioned theoretical basis, we carried edge-network analysis on the biological data GSE30550 downloaded from NCBI GEO. To predict live influenza infection, the edge-network analysis includes several steps as below.

(0) In the preprocess of original data, we choose 1188 and 2909 genes, respectively, for two groups of subjects (or individuals) by using fold change selection (1.2 for asymptomatic and 1.3 for symptomatic in this study). In details, for each subject, his/her differentially expressed genes (DEGs) include those genes selected at different time points, and the DEGs at a given time point are genes whose expression fold change (the ratio between expressions at this time and the first time) larger than the aforementioned threshold. Then, the DEGs of subjects from the same group (Sx or Asx) are united together. Note that, these thresholds are determined by the change of gene number with the fold-change value (Supplementary Fig. S1A and B), so that genes are chosen as many as possible before the number of genes exceeds an half of all genes.

(1) Now, we use PCC to construct co-expression networks (a form of node-network) for 9 Sx subjects, respectively. The correlations of any gene pair (i.e. edge in node-network, or correlation of any two genes) are different in nine networks and they form a nine dimension vector by Equation (4). If the absolute mean value of such correlation vector is >0.8 and its standard deviation value <0.1 (Supplementary Fig. S1C), we select this gene pair. When this threshold is >0.1, the number of the selected edges increases rapidly, and thus we determine these standard deviation thresholds by the effect of edge filtering (Supplementary Fig. S1C). In such a way, we ensure that the selected gene pairs/edges are consistently significant in nine networks, and they will be used to construct the edge-network in following steps. Different from a general node-network, these pre-selected edges will be the background 'nodes' of the final edge-networks. In other words, we actually choose the edges/gene pairs with high correlations in all nine networks as the candidates of biomarkers, which represent the common correlated gene associations among 9 Sx subjects.

(2) Next, for each Sx subject, we carry out the fourth-order correlation coefficient estimation for each edge pair (i.e. a pair of gene pairs) by Equation (5). The threshold of absolute value of such correlation is set as 0.97 (Supplementary Fig. S1D) to choose the meaningful association between two edges (i.e. four molecules). The rule for this threshold selection is based on the ratio between number of the selected gene pairs and number of all possible gene pairs (i.e. complete graph) under the same number genes (Supplementary Fig. S1D), which is on purpose to keep the scarcity of associations among genes. Note that, during this step, we only compute the correlations between preselected edges from aforementioned step; i.e. we just consider the edges/gene pairs that are consistently significant in the original node-networks. In such a manner, we can reduce the computation time and memory space drastically. Then, we will get nine edge-networks for Sx subjects correspondingly, and those edges (i.e. molecule pairs) presenting in at least seven Sx edge-networks are thought to be closely related to disease development as the edge-biomarkers of Sx group.

(3) As the same as the construction method of edge-networks for 9 Sx subjects, we can also build edge-networks for 8 Asx subjects, respectively. The thresholds used to construct Asx edge-networks are the same as above values used in the construction of Sx edge-networks. Finally, the edge-biomarkers of Asx group as the selected common edges are in at least 5 Asx edge-networks, i.e. from more than a half of the Asx subjects due to the indistinctive common edges in Asx group.

(4) Actually, we can use the differential gene pairs of two groups (i.e. differential edge relations in Sx subjects and Asx subjects) as novel edge-biomarkers to distinguish symptomatic (Sx) and asymptomatic (Asx) groups with influenza infection. For example, the correlation values of edges in these edge-biomarkers by Equation (4) can be used for hierarchical clustering. To compare the contribution of these edge-biomarkers from differential analysis, we also examine the edge-biomarkers induced from two groups (Sx and Asx), respectively.

(5) Furthermore, a criterion Equation (6) based on DNB is used to indicate a sudden deterioration before the disease. This composite criterion for DNB in the predisease state is defined by combining three conditions (Chen *et al.*, 2012; Liu *et al.*, 2012; 2013b):

$$CI =: \frac{SD_d \cdot PCC_d}{PCC_o} \tag{6}$$

where PCC_d is the average PCC of the expressions of genes in the dominant group or DNB (e.g. a group of marker genes) in absolute value; PCC_o is the average PCC between the expressions of the dominant group genes and other genes in absolute value; and SD_d is the average standard deviation of the expressions of the dominant group genes (Chen *et al.*, 2012; Liu *et al.*, 2012, 2013b). This criterion can also be applied to quantify our edge-biomarkers and used to reflect molecular network rewiring before disease occurrence. Actually, from the biological viewpoint, the selected edge-biomarkers can be used to predict influenza infection. This is because there is also a critical transition from virus infection to disease, which can be indicated by the DNB score of edge-biomarkers. In this study, the dominant group is the molecules or genes of edge-biomarkers from Sx group (22 genes related to disease) and others (not in dominant group) are the genes for studying Sx and Asx, respectively, i.e. the union of 1188 and 2909 genes. To calculate such a criterion, we predefine the time window for expression correlation calculation. The window length is five time points, i.e. he DNB score including expression data during any consecutive five time points. Thus, each Sx or Asx subject has 11 time windows and the corresponding DNB scores to represent his/her diagnostic score over time. When DNB score becomes significantly larger than a given prediction threshold, this time window will be regarded as the early warning point/period of the disease. By using DNB score quantifying edge-biomarkers to predict influenza infection in practice, the Receiver-Operating Characteristic curve (ROC) can be drawn along with the change of the prediction threshold from 0 to 3 with interval as 0.01, and the corresponding Area Under ROC Curve (AUC) should be used to evaluate the accuracy of such prediction.

(6) Finally, to confirm the robustness of our edge-biomarkers and avoid overfitting, the LOOCV is applied for Sx subjects (our discriminative edge-biomarkers are finally selected not from Asx subjects but from Sx subjects, so that, LOOCV only removes Sx subjects one by one). In detail, we rerun above edge-network analysis nine times. In each time, an Sx subject is kept and other eight subjects are used to infer the edge-network biomarkers. Then these biomarkers are used to predict that the remaining Sx subject and other Asx subjects. After nine times, all the prediction results based on the threshold change of edge-biomarkers are summarized to draw ROC and calculate AUC.

Therefore, by implementing the aforementioned algorithm, we actually have following conclusions: (i) for each Sx subject, the early warning signal of the symptom was found before clinical diagnosis; (ii) for each Sx subject, the critical time point was detected; and (iii) the edge-biomarkers are significantly related to the disease progression and development (e.g. virus infection).

3 RESULTS

Although there already have many elegant experiments of the study of host response to invading pathogens (Fenner *et al.*, 2006; Ichinohe *et al.*, 2009; Proud *et al.*, 2008; Ryo *et al.*, 2008; Zhu *et al.*, 2008), how to predict the outcome of the influenza infection from the observed data before the appearance of disease symptom and what are the key molecules to result in the transition from a healthy state to a disease state still remain unclear. Thus, we carried edge-network analysis on the biological data GSE30550 downloaded from NCBI GEO, which tried to predict live influenza infection.

3.1 Edge-networks from non-symptom or predisease expression data can distinguish influenza symptomatic (Sx) and asymptomatic (Asx) subjects

(1) The gene composition in different edge-biomarkers. Figure 2C shows the time-course gene expression profiles and clinical data of 17 healthy adults within live influenza experiment (Huang *et al.*, 2011), for which 8 adults have non-symptom, i.e. asymptomatic (Asx) during whole period, whereas 9 adults have the symptom, i.e. symptomatic (Sx) after the virus infection but at different time points. By analyzing these expression profiles (Huang *et al.*, 2011), we constructed the corresponding 17 edge-networks based on the computational algorithm of edge-network analysis (see Section 2), and further identified edge-biomarkers as well as the DNB scores during infection of influenza A. Specifically, these two groups of edge-networks were constructed from the non-symptom (i.e. predisease state before the infection symptom appears) gene expression data corresponding to 9 symptomatic (Sx) and 8 asymptomatic (Asx) adults, respectively, and two sets of common genes (or gene pairs) were further extracted from respective symptomatic and asymptomatic edge-networks (Supplementary Tables S1 and S2). Note that for the purpose of infection prediction, we only use the data with non-symptom for Sx subjects and with non-symptom period for Asx subjects in Figure 2C to identify the edge-biomarkers. We have found 80 edges with 22 common genes (IFI44, IFI44L, DDX58, GBP1, TDRD7, IFI35, IFIT2, IFIT1, IFIT3, MX1, OAS1, OAS2, OAS3, LAP3, HERC5, PLSCR1, EIF2AK2, IFIH1, SERPING1, OASL, RSAD2 and ISG15) appearing in many symptomatic edge-networks (corresponding to nine symptomatic subjects), which are totally different from 34 edges with 41 genes commonly observed in the eight asymptomatic edge-networks. These results provide obvious evidence that different networks or edges observed before the symptom appearance lead to divergent disease outcomes after virus infection, thereby implying that they can be used to predict the disease outcomes.

(2) The biological function and disease relevance of edge-biomarkers. We also analyzed the functional enrichment of 22 common genes for Sx (Supplementary Table S3). The enriched pathways are consistent with our expectations. A total of 8 of 22 genes (DDX58, MX1, OAS1, OAS2, OAS3, EIF2AK2, IFIH1 and RASD2) are observed in Influenza A pathway (Supplementary Fig. S2), and 7 genes (DDX58, MX1, OAS1, OAS2, OAS3, EIF2AK2 and IFIH1) are also in Measles pathway, which possibly share the similar influence on interferon-alpha signaling pathway as influenza infection. In addition, another three significantly enriched pathways include hepatitis C with 6 genes (DDX58, IFIT1, OAS1, OAS2, OAS3 and EIF2AK2), herpes simplex infection with 7 genes (DDX58, IFIT1, OAS1, OAS2, OAS3, EIF2AK2 and IFIH1) and RIG-I-like receptor signaling pathway with 3 genes (DDX58, IFIH1 and ISG15). Particularity, MX1 [myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)] is an important gene associated with influenza infection (Engelhardt *et al.*, 2004; Garber *et al.*, 1991; Salomon *et al.*, 2007) and it is also relevant with infection of hepatitis C (Knapp *et al.*, 2003). Furthermore, other genes undiscovered in influenza A pathway or other pathways are also tightly related to influenza according to their reported roles in biological functions and pathogen mechanisms. For example, IFI44, IFI44L, GBP1, IFI35, IFIT2, IFIT3 and HERC5 are all interferon-induced proteins. It has been reported that ISG15 conjugation inhibits influenza A virus gene expression and replication and targets on the viral NS1 protein in virus-infected cells (Hsiang *et al.*, 2009; Zhao *et al.*, 2010). HERC5 is found to attenuate influenza A virus by catalyzing ISGylation of viral NS1 protein (Tang *et al.*, 2010). In contrast, 41 genes obtained from Asx show nothing significant with the influenza in pathway enrichment analysis, and they even did not contain well-known genes related to diseases. Thus, the identified genes in edge-biomarkers are reasonably explained on the biological functions and networks corresponding to the divergent outcomes of two different subject groups. Note that, the edge-biomarkers indicate the new roles (associations) of genes (which can be used for disease classification and prediction), although many genes are already known as disease genes.

(3) The comparison of edge-biomarkers determined in predisease stage and advanced-disease stage. To illustrate the gene and function differences between predisease and disease states during influenza A infection, we also carried out the same analysis by using the whole time-course gene expression data (see Supplementary Information). And according the analyzed result (see Supplementary Information), we still use the genes involved in edge-biomarkers from Sx subjects and their DNB score for early diagnosis or predicting influenza infection.

3.2 Common genes of edge-networks as edge-biomarkers can accurately predict time and outcome of influenza infection before the disease

(1) The prediction of influenza infection is not only to judge the final outcome of different subjects but also estimate the time when the subsequent infection occurs. In aforementioned analysis, the 80 edge-biomarkers with 22 genes from Sx subjects can distinguish Sx and Asx groups well, especially when using the gene expression data on all available time points. However, to use these new biomarkers to predict influenza infection for early diagnosis, a key question is if or not these biomarkers can predict the phenotypes of disease candidates before the appearance of their disease symptoms. Thus, different from traditional biomarkers to distinguish disease and normal samples, we have shown that common genes of edge-networks as edge-biomarkers can (i) correctly predict the outcome of influenza infection among healthy adults with live influenza (H3N2/Wisconsin) before the disease and (ii) also predict the critical time with the symptom of influenza infection, which can tell us if and when disease will occur for a specific subject. Using the same criterion as that in Chen *et al.* (2012), i.e. Equation (6), we calculated the DNB score in each time point (time window) during the disease progression after influenza injection, as shown in Figure 3A.

(2) The accuracy of edge-biomarker on the prediction of final outcome of virus infection. We used DNB score quantifying edge-biomarker to predict the subsequent result after virus infecting on each subject (see Section 2) as shown in Figure 2B. For each subject, if in any time window, its DNB score is larger than a given prediction threshold, we consider that this subject is now in a disease state and will become symptomatic. To study the threshold robustness of biomarkers, we used AUC to evaluate the total performance of prediction accuracy based on the DNB score of the edge-biomarker. We obtained the impressive result with 0.93 AUC (Fig. 3B) to classify Sx and Asx subjects. From Figure 3A, clearly, the DNB scores drastically increase for Sx after virus infection, but not for Asx (except the subject 16). That means that our edge-biomarkers can detect the relevant critical transitions during the progression of influenza infection. In particular, when a subject approaches the critical transition point (i.e. the time just before the symptom appears), DNB score drastically increases, thereby indicating the imminent influenza infection or symptom (Sx). Otherwise, for Asx, DNB score is always small due to non-occurrence of critical transition after the influenza injection. Therefore, edge-biomarkers with the DNB scores can correctly predict the outcome of influenza infection for each individual.

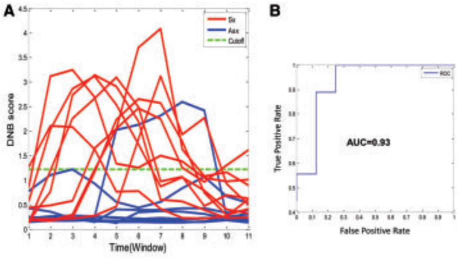


Fig. 3. Prediction of outcomes after influenza injection by edge-biomarkers (or DNB scores). (**A**) The DNB scores of edge-biomarkers for all subjects during the disease progression (22 genes from the Sx edge-network) after the influenza injection. The grey curves are DNB scores for Sx subjects and black ones for Asx subjects. The dotted line is a cutoff for distinguishing two subject groups. Note that one time window is five consecutive time points as indicted in Figure 2C, e.g. window-2 is the time period of 5–36 h. (**B**) The ROC curve of prediction performance based on DNB score by only using non-symptom data (i.e. data before the influenza infection symptom appears). Its AUC is about 0.93

(3) The accuracy of edge-biomarker on the prediction of occurred time of effective virus infection. We investigated whether edge-biomarker can reflect system state or network change before the disease occurrence, and we show the DNB score of the edge-biomarker for each Sx subject (Fig. 4) and each Asx subject (Supplementary Fig. S4). According to the cutoff of DNB score for judging the occurrence of the disease during the progression of influenza infection, clearly we can correctly predict the symptom for each subject (star label in Fig. 4) before the influenza infection was diagnosed with standardized symptom scoring record (circular label in Fig. 4), except one case, i.e. subject 5, which we failed to predict the outcome earlier than the clinical diagnosis but we still correctly predict the disease outcome of this subject. Thus, edge-biomarker can effectively predict the time of onset influenza infection.

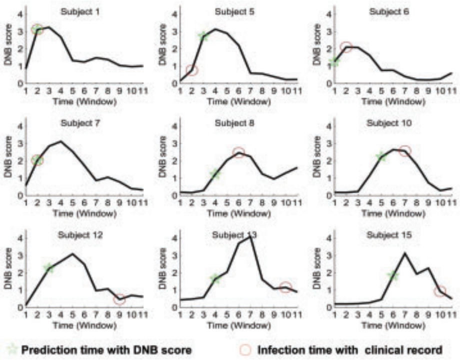


Fig. 4. Prediction performance of influenza subsequent infection time by DNB score of edge-biomarker. One time window is five time points as indicted in Figure 2C, e.g. window-2 is the time period of 5–36 h. Each sub-figure displays the DNB score of the edge-biomarker for each Sx subject during the progression of the influenza infection. The dotted mark indicates the predicted time by DNB score cutoff, whereas the circular mark shows the clinically diagnosed infection time for the corresponding subject. Obviously, almost all 9 Sx subjects can be accurately predicted by DNB score before actual clinic diagnosis, although our predicted time is a little delayed for subject 5

(4) The robustness of edge-biomarkers evaluated by cross-validation. We also repeat edge-network analysis with LOOCV to assess our edge-biomarkers with DNB scores and their prediction accuracy. The results are inspiring because the marker genes are significantly stable and AUC of prediction achieves 0.90 (Supplementary Fig. S9A). In all edge-biomarkers obtained in LOOCV, there are 17 genes (IFI44, IFI44L, GBP1, IFIT2, IFIT1, IFIT3, MX1, OAS1, OAS2, OAS3, LAP3, HERC5, PLSCR1, SERPING1, OASL, RSAD2 and ISG15) always appear, which are all found in our 22-genes edge-biomarkers. The total number of marker genes induced in LOOCV is 26, i.e. there are only 4 new genes (LAMP3, RTP4, TNFAIP6, TNFSF10) found during this cross-validation procedure. We found that these four genes are selected as biomarkers only when subject-5 is left out and TNFSF10 is in influenza A pathway. This fact might indicate that there is significant personal specificity with subject-5 so that the DNB score cannot predict it before disease occurs. Besides, we have calculated the F1 values of LOOCV (Supplementary Fig. S9B), the maximum score is 0.62 when the range of prediction threshold on DNB score is from 1.4 to 2.08.

3.3 Comparison between edge-biomarkers and node-biomarkers for predicting influenza infection

Different from previous biomarkers, we found that edge-biomarkers extracted from the edge-network without normal (Asx subjects) samples still have the ability to predict the phenotypes of diseases by exploiting the information of pathogen dynamical expressions, which can not only identify the influenza infection outcomes but also detect the actual infection time points. To compare the effectiveness with the conventional node-networks, we obtained node or molecular biomarkers by using the well-known method ARACNE in the same way as our method.

(1) For each subject in Sx group or Asx group, their gene expressions are used to infer node-network by ARACNE directly. The parameters in ARACNE are set to be their default values, and the *P*-value for mutual information threshold is set as 0.0001. Then the edges appearing in at least 7 Sx node-networks are selected, and their genes are used as node-biomarkers to compare with our edge-biomarkers.

(2) Similarly, the molecules or genes of node-biomarkers from Sx group belonging to the dominant group and others (non-dominant group) are used to study Sx and Asx, respectively. The time window for correlation calculation is predefined and the window length is five time points, which are all the same as our aforementioned DNB score evaluation in edge-biomarker analysis.

The identified marker genes based on ARACNE have 474 much more than 22 genes, which also include some in our edge-biomarkers. This fact shows that the conventional node-network would be hard to narrow down the ranked genes with pathogen relevance (or has higher false positive) due to the lack of the information for collective effects of molecules. Furthermore, these node-biomarkers show significantly less prediction power (AUC = 0.67) than our edge-biomarkers (see Supplementary Fig. S5) when they both use DNB score as a predictor of influenza infection. In fact, although ARACNE is an effective approach to reconstruct a gene regulatory network, it does not consider the dynamical and high-order statistical information in temporal expression data related to disease development. Thus, the genes in node-network from ARACNE may lack the ability on distinguishing predisease samples. This comparison validates the effectiveness of edge-biomarkers for early diagnosis of diseases.

In addition, we have also compared edge-biomarkers with traditional gene markers selected by student's *t*-test on the temporal gene expression data before the sixth time point. The selected genes are significantly differentially expressed from the control time point, i.e. the first time, for the Sx and Asx subjects, respectively ($P < 0.01$). There are 439 genes that are only significantly differentially expressed in Sx subjects. The hierarchical clustering of these genes shown in Supplementary Figure S6 illustrates no obvious expression patterns between Sx and Asx subjects. Furthermore, the DNB scores based on this gene group give a similar prediction performance (AUC = 0.64) as that of ARACNE (see Supplementary Fig. S7), which demonstrates again the superiority of edge-biomarkers on predicting influenza infection.

4 DISCUSSION

In contrast to conventional node-network focusing on association between nodes (e.g. genes), however, our edge-network aims at association between node-pairs (e.g. pairs of protein interactions). With appropriate assumptions, we can show that a biological system at a molecular level can be exactly modeled by the equations with first-order and second-order statistics, where the first-order equations correspond to the traditional molecular network or node-network, whereas the second-order equations correspond to the covariance network or edge-network. Node-network based on the average values of molecular concentrations is widely used to analyze the biological behaviors at a specific condition, e.g. a normal state or a disease state, but it cannot directly be applied to the analysis on the critical state before the drastic transition due to the requirement of the second-order statistics information. In contrast, as a complementary part, edge-network is based on covariance information among molecules, and thus it can be applied to predict the critical transition of the system by identifying edge-biomarkers from time-course (or stage-course) expression profiles. In this article, we have studied the time-course data of 17 subjects (healthy adults) with risk of influenza infection, and extracted the common edge-biomarkers from edge-networks in two groups of subjects with different clinical outcomes. The results show that the edge-biomarkers derived from the edge-networks are able to not only distinguish the Asx and Sx groups but also predict the outcomes of the symptoms before the clinical diagnosis. In this study, we only used one dataset to compare with other method, and thus the validation of this work seems insufficient. But with the accumulation of temporal expression data, we will implement more elaborate experiments to evaluate the effectiveness of our method. In addition, comparing with the traditional node-network, the size of the edge-network is much big and thus its analysis may be a computationally challenging task.

Our edge-network analysis to predict influenza infection is based on an assumption that subjects with same clinical outcomes tend to have similar biological responses on a higher-order network level after virus infection. This assumption ensures that our model is consistent with the actual clinical diagnosis. We constructed the edge-network for each adult based on the gene expression profile before the appearance of influenza symptom, and validated the prediction ability of edge-biomarkers quantified by DNB score (Chen *et al*., 2012; Liu *et al*., 2012, 2013a). The analysis results support that high-order statistics and dynamical information of biological data will be effective for mining pathogen genes closely related to disease occurrence and development even without normal control. Note that each edge-network is constructed not by different individuals but by the same person or species, and in contrast, traditional node-network and node-biomarkers are identified from groups of different individuals. Edge-network is also significantly different from the existing network-based methods and biomarkers (Liu *et al*., 2013a, b) including DNB, which are still based on node-network.

From practical perspective, our method is attractive because it provides a powerful tool for disease prediction, which is crucial for the early treatment and prevention of the patients. The edge-network analysis does not depend on the specific details of a disease, and therefore it can be used to analyze other types of diseases for prevention, early diagnosis and treatment. The proposed method is a general way to represent a stochastic dynamical system, and thus in addition to disease progression, the proposed method can be used to study other biological processes in a similar manner. By combining with traditional methods such as Bayesian inference method, we may directly consider or characterize the stochastic dynamics of the system to detect effective edge-biomarkers. Also it is necessary to consider imprecise phenotype factors so as to achieve accurate diagnosis on complex diseases.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Tao Zeng for his suggestions, and members of Chen lab in SIBS, CAS.

Funding: National Natural Science Foundation of China (grant nos. 61134013, 91029301, 91130033, 11126282 and 61072149) (in part), the Chief Scientist Program of Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS) (No. 2009CSP002), the Knowledge Innovation Program of SIBS of CAS with Grant No. 2013KIP218 and the FIRST program from JSPS initiated by CSTP.

Conflict of Interest: none declared.

REFERENCES

Chen,L. *et al*. (2010) *Modeling Biomolecular Networks in Cells*. Springer-Verlag, London.

Chen,L. *et al*. (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Rep.*, **2**, 342.

Chuang,H.Y. *et al*. (2007) Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.*, **3**, 140.

Dowling,H.F. *et al*. (1958) Transmission of the common cold to volunteers under controlled conditions. III. The effect of chilling of the subjects upon susceptibility. *Am. J. Hyg.*, **68**, 59–65.

Engelhardt,O.G. *et al*. (2004) Mx1 GTPase accumulates in distinct nuclear domains and inhibits influenza A virus in cells that lack promyelocytic leukaemia protein nuclear bodies. *J. Gen. Virol.*, **85**, 2315–2326.

Fenner,J.E. *et al*. (2006) Suppressor of cytokine signaling 1 regulates the immune response to infection by a unique inhibition of type I interferon activity. *Nat. Immunol.*, **7**, 33–39.

Garber,E.A. *et al*. (1991) Avian cells expressing the murine Mx1 protein are resistant to influenza virus infection. *Virology*, **180**, 754–762.

He,D. *et al*. (2012) Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. *J. Mol. Cell Biol.*, **4**, 140–152.

Hsiang,T.Y. *et al*. (2009) Interferon-induced ISG15 conjugation inhibits influenza A virus gene expression and replication in human cells. *J. Virol.*, **83**, 5971–5977.

Huang,Y. *et al*. (2011) Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genet.*, **7**, e1002234.

Ichikawa,S. *et al*. (2009) Periodic Lyapunov differential equation for noise evaluation in oscillatory genetic networks. In: *Control Applications, Intelligent Control, 2009 IEEE*. pp. 83–88.

Ichinohe,T. *et al*. (2009) Inflammasome recognition of influenza virus is essential for adaptive immune responses. *J. Exp. Med.*, **206**, 79–87.

Knapp,S. *et al*. (2003) Polymorphisms in interferon-induced genes and the outcome of hepatitis C virus infection: roles of MxA, OAS-1 and PKR. *Genes Immun.*, **4**, 411–419.

Li,M. *et al*. (2013) Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: study of type 2 diabetes by cross-tissue analysis. *Brief Bioinform.*, [Epub ahead of print, doi: 10.1093/bib/bbt027, April 25, 2013].

Liu,R. *et al*. (2012) Identifying critical transitions and their leading biomolecular networks in complex diseases. *Sci. Rep.*, **2**, 813.

Liu,R. *et al*. (2013a) Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Med. Res. Rev.*, [Epub ahead of print, doi:10.1002/med.21293, June 17, 2013].

Liu,R. *et al*. (2013b) Dynamical network biomarkers for identifying critical transitions of biological processes. *Quant. Biol.*, **1**, 105–114.

Margolin,A.A. *et al*. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, **7 (Suppl. 1)**, S7.

Oates,C.J. and Mukherjee,S. (2012) Network inference and biological dynamics. *Ann. Appl. Stat.*, **6**, 1209–1235.

Proud,D. *et al*. (2008) Gene expression profiles during in vivo human rhinovirus infection: insights into the host response. *Am. J. Respir. Crit. Care Med.*, **178**, 962–968.

Ryo,A. *et al*. (2008) SOCS1 is an inducible host factor during HIV-1 infection and regulates the intracellular trafficking and stability of HIV-1 Gag. *Proc. Natl Acad. Sci. USA*, **105**, 294–299.

Salomon,R. *et al*. (2007) Mx1 gene protects mice against the highly lethal human H5N1 influenza virus. *Cell Cycle*, **6**, 2417–2421.

Tang,Y. *et al*. (2010) Herc5 attenuates influenza A virus by catalyzing ISGylation of viral NS1 protein. *J. Immunol.*, **184**, 5777–5790.

Van Kampen,N.G. (1992) *Stochastic Processes in Physics and Chemistry*. Elsevier Science Publishing Company, North-Holland.

Wang,Y. *et al*. (2006) Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics*, **22**, 2413–2420.

Wu,S. and Wu,H. (2013) More powerful significant testing for time course gene expression data using functional principal component analysis approaches. *BMC Bioinformatics*, **14**, 6.

Zhang,B. and Horvath,S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, **4**, e17.

Zhang,X. *et al*. (2012) Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics*, **28**, 98–104.

Zhang,X. *et al*. (2013) NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference. *Bioinformatics*, **29**, 106–113.

Zhao,C. *et al*. (2010) ISG15 conjugation system targets the viral NS1 protein in influenza A virus-infected cells. *Proc. Natl Acad. Sci. USA*, **107**, 2253–2258.

Zhi,W. *et al*. (2013) Network-based analysis of multivariate gene expression data. *Methods Mol. Biol.*, **972**, 121–139.

Zhu,Q. *et al*. (2008) Toll-like receptor ligands synergize through distinct dendritic cell pathways to induce T cell responses: implications for vaccines. *Proc. Natl Acad. Sci. USA*, **105**, 16260–16265.