SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
COMPUTATIONAL BIOLOGY
BIOINFORMATICS
BIOPHYSICS
CANCER MODELS

Received
22 September 2011

Accepted
1 March 2012

Published
29 March 2012

Correspondence and requests for materials should be addressed to L.C. (lnchen@sibs.ac.cn)

Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers

Luonan Chen[1,2], Rui Liu[2], Zhi-Ping Liu[1], Meiyi Li[1] & Kazuyuki Aihara[2]

[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, China, [2]Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan.

Considerable evidence suggests that during the progression of complex diseases, the deteriorations are not necessarily smooth but are abrupt, and may cause a critical transition from one state to another at a tipping point. Here, we develop a model-free method to detect early-warning signals of such critical transitions, even with only a small number of samples. Specifically, we theoretically derive an index based on a dynamical network biomarker (DNB) that serves as a general early-warning signal indicating an imminent bifurcation or sudden deterioration before the critical transition occurs. Based on theoretical analyses, we show that predicting a sudden transition from small samples is achievable provided that there are a large number of measurements for each sample, e.g., high-throughput data. We employ microarray data of three diseases to demonstrate the effectiveness of our method. The relevance of DNBs with the diseases was also validated by related experimental data and functional analysis.

It has been identified that a sudden change of a system state exists widely in ecosystems[1,2], climate systems[3,4], economics and global finance[5,6]. Such a change often occurs at a critical threshold, or the so-called "tipping point", at which the system shifts abruptly from one state to another. This is well known in dynamical systems theory as a bifurcation that results in a qualitative transition in states or attractors[7,8]. Recently, evidence has been found suggesting that the similar phenomena exist in clinical medicine, that is, during the progression of many complex diseases, e.g., in chronical diseases such as cancer, the deterioration is not necessarily smooth but abrupt[9–13]. In other words, there exists a sudden catastrophic shift during the process of gradual health deterioration that results in a drastic transition to a disease state. In order to describe the underlying dynamical mechanism of complex diseases, their evolutions are often modeled as time-dependent nonlinear dynamical systems, in which the abrupt deterioration is viewed as the phase transition at a bifurcation point, e.g., for prostate cancer[14], asthma attacks[9] and epileptic seizures[15]. We are particularly interested in the complex diseases with sudden deterioration phases or critical transition points during their progressions. Depending on the progression level of such illnesses, we divide the process into three stages: *i.e.*, a normal state, a pre-disease state (or a critical state), and a disease state (Figs. 1 a and b–d). The normal state is a steady state, representing a relatively healthy stage during which the disease is under control, in an incubation period or in a chronic inflammation period. The pre-disease state has been discussed in many previous works (e.g., Achiron et al.[16]). It is usually defined as the limit of the normal state immediately before the tipping point is reached. In this pre-disease stage (see Figs. 1 a and c), the process is usually reversible to the normal state if appropriately treated, implying instability of the pre-disease state. However, it usually becomes irreversible to the normal state if the system passes the critical point and enters another stable state, or the unique attractor at the disease stage (Figs. 1a and d). Hence, it is crucial to detect the pre-disease state so as to prevent qualitative deterioration by taking appropriate intervention actions.

Recently, topics on predicting the tipping point are increasingly receiving much attention in many scientific fields[17], e.g., in ecological systems[1,18–20], climate systems[3,21,22] and physiological systems[10]. Many such predictions are mainly based on "critical slowing down"[23], a generic dynamical phenomenon occurring in the vicinity of a bifurcation point where the system becomes increasingly slow when recovering from small perturbations back to

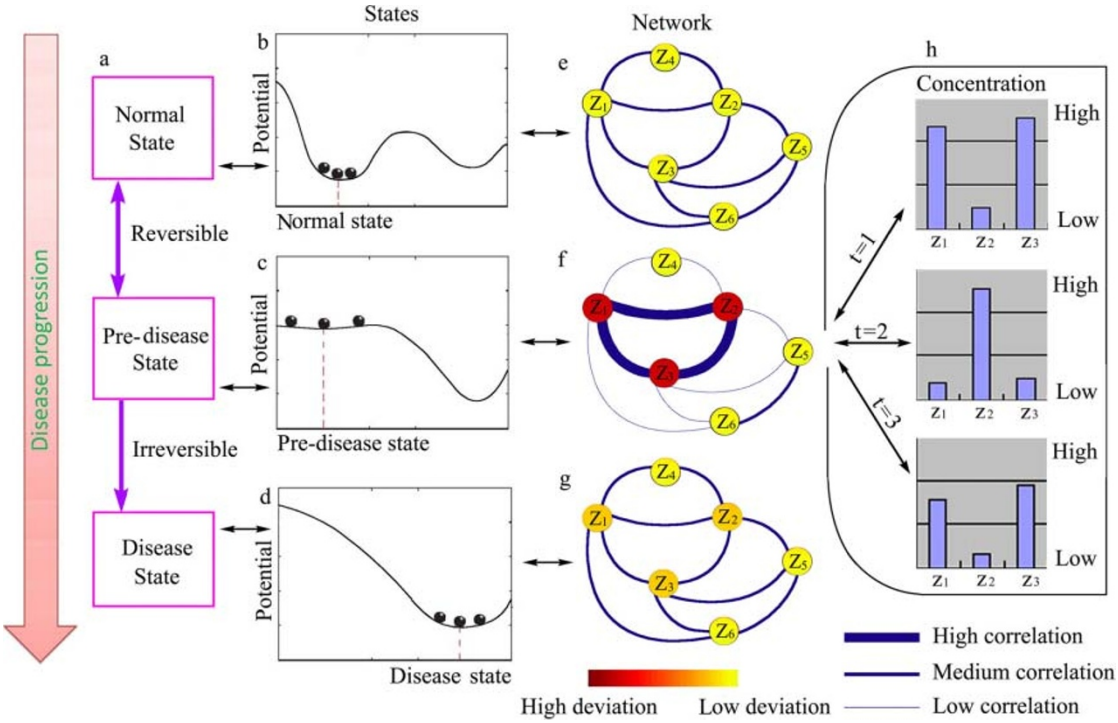

Figure 1 | Schematic illustration of the dynamical features of disease progression from a normal state to a disease state through a pre-disease state. (a) Deterioration progress of disease. (b) The normal state is a steady state or a minimum of a potential function, representing a relatively healthy stage. (c) The pre-disease state is situated immediately before the tipping point and is the limit of the normal state but with a lower recovery rate from small perturbations. At this stage, the system is sensitive to external stimuli and still reversible to the normal state when appropriately interfered with, but a small change in the parameters of the system may suffice to drive the system into collapse, which often implies a large phase transition to the disease state. (d) The disease state is the other stable state or a minimum of the potential function, where the disease has seriously deteriorated and thus the system is usually irreversible to the normal state. (e)–(g) The three states are schematically represented by a molecular network where the correlations and deviations of different species are described by the thickness of edges and the colors of nodes respectively. When the system approaches the pre-disease state, the deviations of (z_1, z_2, z_3) increase drastically, and the correlations among (z_1, z_2, z_3) also increase drastically whereas their correlations with other nodes (z_4, z_5, z_6) decrease drastically ((e) and (f)). We call (z_1, z_2, z_3) as the dominant group or the DNB. (g) At the disease stage, the system settles down in another steady state, i.e., the disease state, with lowered deviations and correlations for the DNB. (h) Graphs show an example of the dynamical fluctuations of the molecular concentrations for the DNB in the pre-disease state, which dynamically change with strong temporal deviations but are closely (positively or negatively) correlated.

its equilibrium state[24,25]. Considerable efforts have been made in successfully finding general early-warning signals, ranging from theoretical derivation to real data analysis, which have shown that the increases in variance[24], autocorrelation[21], skewness[26] and spatial correlation[27,28] are generally expected to occur as a system approaches to a critical point.

However, in the case of complex diseases, it is notably hard to predict such critical transitions for the following reasons. First, because a pre-disease state is a limit of the normal state, the state of the system may show little apparent change before the tipping point is reached. Thus, the diagnosis by traditional biomarkers and snapshot static measurements may not be effective to distinguish those two states (Figs. 1b, c). Second, despite considerable research efforts, no reliable disease model has been developed to accurately detect the early-warning signals. In particular, deterioration processes may be considerably different even for the same subtype of a disease, depending on individual variations, which makes model-based prediction methods fail for many cases. Third and most importantly, detecting the pre-disease state must be an individual-based prediction, however, usually there are only a few of samples available for each individual, unlike many other complex systems that are measured over a long term with a large number of samples.

On the other hand, rapid advancements on high-throughput technologies have enabled us to observe gene expressions or even protein expressions at the genome-wide scale[29], *i.e.,* with over thousands of measurements in one sample. Such high-dimensional data not only provides a global view with rich information on the concerned system, but also represents the accumulated effects of its long-term dynamics[30], which implies that even a small number of samples may characterize the dynamical features of a living organism. In this work, by using time-course high-throughput omic data and based on the principle of critical slowing down or a bifurcation process, we proposed a novel model-free (network-based) method to detect early-warning signals of complex diseases even with a small number of samples. Specifically, we theoretically showed that a dynamical network biomarker (DNB) can serve as a general early-warning signal indicating an imminent bifurcation or sudden deterioration before the critical transition occurs. We successfully identified critical transitions as well as DNBs for lung injury disease, liver cancer, and lymphoma cancer based on their microarray data. The relevance of the identified DNBs with the diseases was also validated by pathway enrichment analysis, functional analysis, bootstrap analysis and related experimental data. It should be note that we aimed to detect the pre-disease state by finding its associated DNB, rather than by distinguishing between the disease state (e.g., disease samples) and the normal state (e.g., control samples) by finding its biomarkers.

Results

Criteria for detecting early-warning signal. In particular, we theoretically derived network-based dynamical criteria that serve

as a general early-warning signal indicating the imminent bifurcation before a critical transition occurs. Based on theoretical analysis, we proved that the following generic properties hold when the system reaches the pre-disease state (see the derivation of DNBs in Methods, Figs. 1e–h, and Supplementary Information A for details).

- There exists a group of molecules, *i.e.,* genes or proteins, whose average Pearson's correlation coefficients (PCCs) of molecules drastically increase in absolute value.
- The average PCCs of molecules between this group and any others (*i.e.,* between molecules inside this group and any other molecules outside this group) drastically decrease in absolute value.
- The average standard deviations (SDs) of molecules in this group drastically increase.

If all of these three conditions are satisfied simultaneously, we call this group a dominant group of the system, whose change will reflect a transition of the system to the disease state. Actually, each of the three conditions represents a criterion, and their combination is naturally expected to be a strong signal or an indicator for the pre-disease state. Because the dominant group characterizes dynamical features of the underlying system and the molecules in the group are also strongly and dynamically correlated in the pre-disease state, the molecules in the group are expected to form a subnetwork from a network viewpoint. Hence, we regard it as a dynamical network of biomarkers, or a dynamical network biomarker (DNB). Unlike the traditional molecular biomarkers used in medicine, whose expressions reflect the presence or severity of the disease state and are required to have consistent (or constant) values that are different in the respective disease and normal states, the DNB is a strongly correlated molecular subnetwork where the concentrations of molecules, however, dynamically change without keeping constant values in the pre-disease state as shown in Fig. 1h. In other words, the concentrations of molecules in the DNB tend to increasingly fluctuate when the system approaches to the pre-disease state, although they behave dynamically in a strongly collective manner, which is a key feature of the DNB. This is why it can be used to detect the early signal of a complex disease in the early stage, which is not otherwise possible using traditional biomarkers or methods. Hence, the existence of the DNB implies that the system is in the pre-disease state. Note that each individual may have a different DNB even for the same disease due to individual variations. Therefore, in contrast to traditional biomarkers, a DNB is not necessarily composed of a fixed bunch of molecules even for the same disease but may have different members depending on individual features that are detected from individual high-throughput data.

Composite index for pre-disease state. We then describe the theoretical background of the DNB for model-free prediction, and we derive a general quantitative index as an early-warning signal for the pre-disease state regardless of the detailed differences among systems.

Assume that the progression of a disease can be expressed by the following dynamical system

$$Z(k+1) = f(Z(k); P). \tag{1}$$

$Z(k) = (z_1(k), \ldots, z_n(k))$ represent observed data, *i.e.,* concentrations of molecules (e.g., gene expressions or protein expressions) at time k ($k = 0, 1, \ldots$), e.g., hours or days, which are the variables describing the dynamical state of the system. P are parameters representing slowly changing factors, including genetic factors (e.g., SNP and CNV) and epigenetic factors (e.g., methylation and acetylation), which drive the system from one state (or attractor) to another (see Supplementary Fig. S1). $f = (f_1, \ldots, f_n)$ are generally nonlinear functions of $Z(k)$ with a fixed point $Z = \bar{Z}$, such that $\bar{Z} = f(\bar{Z}; P)$. Assume that there is a value P_c such that at least one of the eigenvalues of the Jacobian matrix $\left.\frac{\partial f(Z;P_c)}{\partial Z}\right|_{Z=\bar{Z}}$ equals 1 in modulus, which implies that the system undergoes a bifurcation at $\bar{Z}$ when P reaches the threshold P_c. By analyzing the nonlinear dynamics near the bifurcation point, we can theoretically derive the above three criteria to detect the dominant group or DNB (see Methods and Supplementary Information A). To obtain a strong signal in the pre-disease state, we further combine the three criteria together to construct a composite index I as follows:

$$I =: \frac{\mathrm{SD}_d \cdot \mathrm{PCC}_d}{\mathrm{PCC}_o}, \tag{2}$$

where PCC_d is the average PCC of the dominant group in absolute value; PCC_o is the average PCC between the dominant group and others in absolute value; and SD_d is the average SD of the dominant group.

Although the expression of each z_i may stochastically change at any time instant due to the fluctuation, the composite index is expected to increase sharply whenever the system approaches to a critical transition point, and therefore, it can serve as an effective early-warning signal to identify the pre-disease state.

According to the derivation, the warning signal is generated by the interdependent molecules in the dominant group, through their dynamical interactions near the critical transition. Because the molecules in the dominant group are strongly and dynamically correlated, they behave as a dynamical subnetwork or a module in the pre-disease state, forming the DNB, which will move to the disease state first when the system undergoes further perturbations of parameters into the disease stage. These properties should be generic and lie in many complex diseases with sudden deterioration phenomena. Clearly, analysis based on such a signal is applicable to diseases with or without a model. The algorithm to numerically identify the DNB from high-throughput data is given in detail in Supplementary Information C. Note that the purpose of this work is not to identify the parameters or factors that first drive the system into the disease state, but to identify the responsive DNB which first moves into the disease state driven by any known or unknown factors (or first reflects such a critical transition). Some molecules related to the driving factors (e.g., SNPs or CNVs) may or may not be in the DNB.

Numerical experiments. We simulated an example of a five-node gene regulatory network (see Fig. 2a) illustratively to demonstrate the DNB and the composite index. The detailed descriptions of the network represented by a set of stochastic differential equations are provided in Supplementary Information B, and numerical simulations are shown in Fig. 2, which demonstrates the predictive power of the DNB.

Application to complex diseases. We further conducted the prediction for three real diseases using high-throughput experimental data, which are respectively the microarray data of a lung injury with carbonyl chloride inhalation exposure (or acute lung injury)[31], the hepatic lesion by chronic hepatitis B (or HBV induced liver cancer)[32], and the B-cell lymphomagenesis of transgenic murine (or lymphoma)[33]. The detailed algorithm and data description are described in Supplementary Information C and D, respectively. For the diseases of phosgene inhalation lung injury and hepatic lesion, Fig. 3 shows that the composite indices based on DNBs during the disease progression are consistent with the observed biological phenotypes. The analysis for the B-cell lymphomagenesis is presented in Supplementary Information D. To explore the biological implication, functional analysis was implemented on these identified DNBs from the gene expression datasets respectively. Figures 4a–d show the DNBs with the functional interactions (protein-protein interactions and TF-target regulations) for the diseases. In particular, we take the acute lung injury as a concrete example to present the calculation procedure in

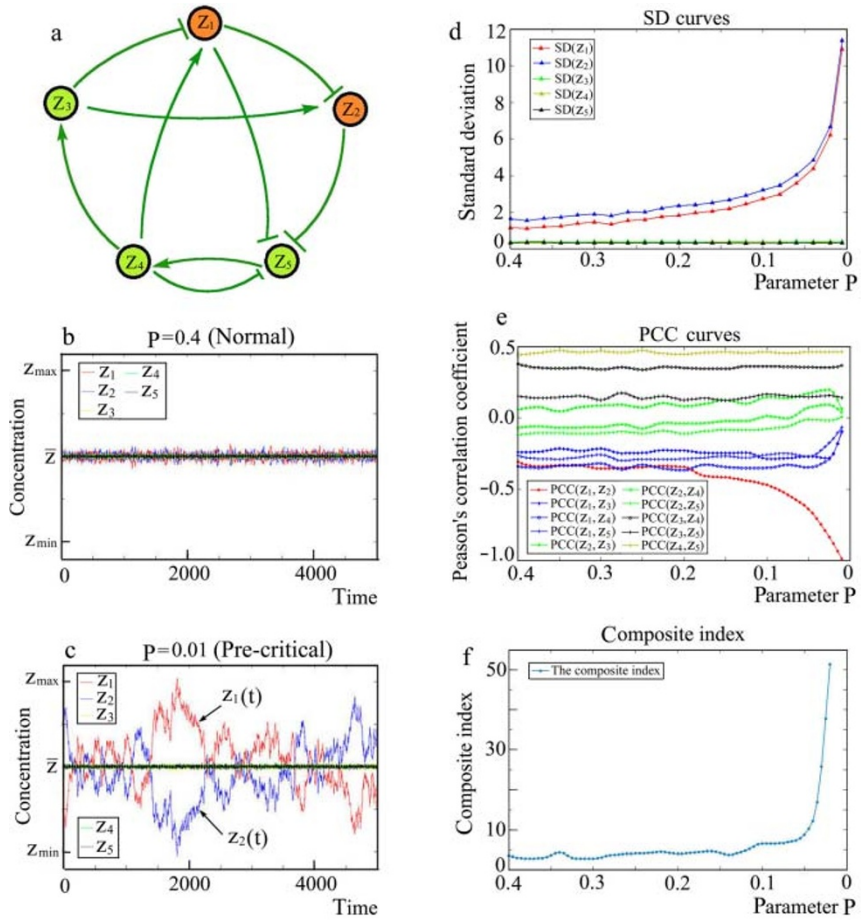

Figure 2 | Numerical validation of theoretical results. (a) A five-gene model for a DNB and an early-warning signal. The network model and detailed background are described in Supplementary Information B. The tipping point is at $P = 0$ in the theoretical model, at which the system undergoes a critical transition or a bifurcation detected by z_1 and z_2. (b)–(c) When the system approaches the tipping point ($P = 0$), z_1 and z_2 become closely correlated with increasingly strong deviations from $P = 0.4$ to $P = 0.01$. (d)–(e) Figures show the curves of SDs and PCCs for the variables against the parameter P, which clearly indicate the tendency of z_1 and z_2, *i.e.*, their fluctuations (SD(z_1) and SD(z_2)) and correlation ($|\mathrm{PCC}(z_1, z_2)|$) increase drastically whereas their correlations with other nodes ($|\mathrm{PCC}(z_1, z_3)|$), $|\mathrm{PCC}(z_1, z_4)|$, $|\mathrm{PCC}(z_1, z_5)|$, $|\mathrm{PCC}(z_2, z_3)|$, $|\mathrm{PCC}(z_2, z_4)|$, and $|\mathrm{PCC}(z_2, z_5)|$) decrease drastically when the system approaches the tipping point, which satisfies all three criteria for the DNB. (f) The curve shows the clear tendency of the composite index near the tipping point for the DNB composed of (z_1, z_2), which can be used as the early-warning signal for predicting the imminent change in the concerned system.

detail (see Supplementary Information D). The dynamical change in the network structure and expression variations for the identified DNB are shown in Figs. 4e–h. Clearly, the DNB forms a strongly correlated subnetwork to provide the significant warning signal near the critical state (8 h). The dynamical changes in the DNBs over whole periods are shown in Supplementary Fig. S8.

To further compare the DNB dynamics with other molecules, we also graphically demonstrate the dynamics of the whole mouse molecular network (protein-protein interactions and TF-target regulations) including the DNB in Figs. 4i–l (also see the whole mouse network for other periods in Supplementary Fig. S9), which clearly shows the significance of the DNB in terms of expression variations and network structures near the critical state (8 h). In particular, before the disease state, there are no significant differences between the members of the DNB and other genes during all periods except at 8 h, when the members of the DNB behave in a considerably different manner in terms of their expression variations and network connections by satisfying our three criteria. However, after the system is driven into the disease state, interestingly the members of the DNB appear to behave in a manner similar to other genes again (e.g., Fig. 4l).

In order to verify the biological significance of the identified DNBs, we carried out bootstrap analysis respectively for the three diseases, as described in Supplementary Information E. For the phosgene inhalation lung injury (or acute lung injury) and hepatic lesion (or HBV induced liver cancer), some enriched GO functions and dysfunctional pathways underlying the DNBs are listed in Table 1. Part of the identified DNBs are also shown in Table 1 (see Supplementary Table 'Identified DNBs' for complete lists). The details and full lists are presented in Supplementary Information F.

The functional analysis of these cases shows that the DNBs are closely relevant to the corresponding complex diseases, thereby validating the effectiveness and advantage of our method. In the lung injury study, we found that the dysfunctions of the DNB are consistent with the corresponding mechanisms of the phosgene-induced injury (or acute lung injury), especially for antioxidant reactions, dysfunctional metabolic processes, inflammatory regulation, and regulations of cell proliferation and death[31]. From pathway enrichments, we found that pathways, such as pathways of repairing DNA damage (e.g., p53 signaling pathways), pathways of inflammatory (e.g., acute myeloid leukemia) and pathways of reducing oxidation (e.g., glutathione metabolism), are highly related to the dysfunctional mechanisms of the aggressive disease[31]. In particular, some genes with the most significant changes are related to the antioxidant reflections and the response of reducing DNA damage. Some of them are involved in the related glutamine-metabolic activities (e.g., *ASNS* and *GCLC*) and participate in oxidation reduction (e.g., *SRXN1*, *PGD*, *TXNL1*, *HMOX1*, *CYP51* and *CH25H*). Moreover, owing to protection against the denaturation of proteins and lipoids caused by

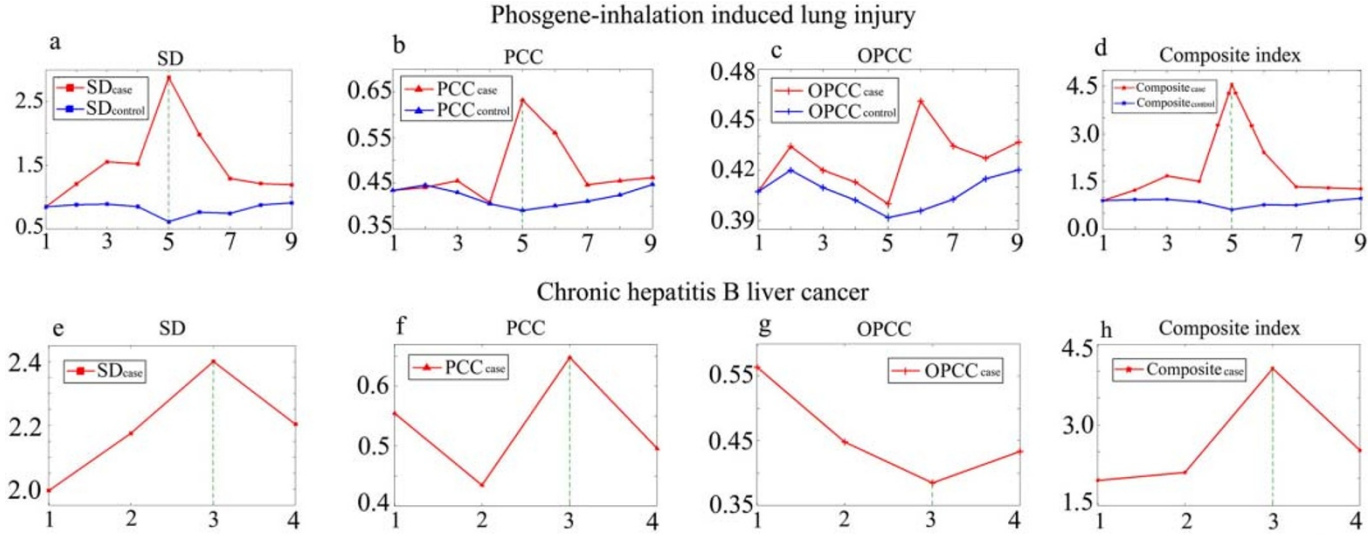

Figure 3 | Detecting early-warning signals for complex diseases. Detecting early-warning signals for diseases from two sets of high-throughput experimental data for the lung injury with carbonyl chloride inhalation exposure, i.e., acute lung injury ((a), (b), (c), and (d)), and the hepatic lesion by chronic hepatitis B, i.e., HBV induced liver cancer ((e), (f), (g), and (h)). In each sampling period, there are 2–5 samples for gene expressions. (a) and (e) represent the mean SDs in the DNB (i.e., SD_d in Eq.(2)), (b) and (f) are mean PCCs in the DNB (i.e., PCC_d in Eq.(2)), (c) and (g) represent the PCCs between the DNB and other molecules (i.e., OPCCs, or PCC_o in Eq.(2)), and (d) and (h) represent the composite index/in Eq.(2). The dotted green line indicates the period of the pre-disease state. Both cases demonstrate strong and significant early-warning signals before the diseases are eventually deteriorated. The results of these diseases show the effectiveness of our method to detect the early-warning signals using a small number of samples.

acylation of phosgene and tissue damage, heat shock protein activities (e.g., *HSPA1A*, *HSPA1b*, *DNAJC5*, *DNAJB4*) and regulation of apoptosis (e.g., *PMAIP1*, *GADD45G*, *EPHA2*) were identified in the DNB. Furthermore, 43 of the 262 genes and TFs (hypergeometric test, p-value $<$ 7.89e-36) were validated with significantly close relation with lung injury by phosgene[31] (Supplementary Table 'KEGG enrichment analysis').

In HBV infection and progression, the dysfunction of the regulation system causes the hepatitis B disease, which may lead to cirrhosis and hepatocellular carcinoma[32]. From the KEGG pathway enrichment results, the DNB members are included in many HBV-related pathways, such as pathways in cancer, hepatitis C, inflammatory activation (e.g., acute myeloid leukemia), and DNA repair (e.g., cytosolic DNA sensing pathway). The enriched GO functions underlying the DNB are particularly related to immune systems that are activated to protect against HBV and inordinate dysfunctions associated with the performance in the viral life cycle. In the DNB, *HLA-DMA* and *B2M*, which play crucial roles in antigen presentation, and *STAT6*, which is associated with the switch of antibody isotypes, undergo significant changes during the disease development[32]. Further, some genes in the basic cellular processes were expressed in a disorderly manner, e.g., *HDAC10* and *NAP1L1*, which are associated with the regulation of cell death and abnormal reaction in transcription and translation, and *GCC2* and *CDH1*, which are associated with transportation and localization. Moreover, *CDH1* has been identified as the target of HBV X proteins, and 17 out of 56 genes and TFs (p-value $<$ 8.51e−4) have been identified as DNB members (Supplementary Table 'KEGG enrichment analysis'). The findings are also highly related to the response of HBV infection *in vivo*, which provides more evidence of the vital functions of the DNB for the disease and further demonstrates the effectiveness of our method. The DNB of HBV sheds light on the progression of pathogen effects and also dynamically characterizes the disease progression in the liver.

Discussion

To make a model-free prediction of the critical transition of diseases using a small number of samples of high-throughput data, we proposed a composite index serving as a leading early-warning indicator based on the DNB. Different from the previous works, our method for predicting complex diseases is based on searching for a particularly interacted network or a DNB carrying reliable information about dynamical variations. From the viewpoint of nonlinear dynamics, this network is driven by certain factors or parameters to first move to (or reflect the transition to) the disease state. Therefore, the network can serve as a new type of biomarker to detect the pre-disease state in a dynamical manner (Figs. 1f and h), in contrast to the traditional gene- or protein-based biomarkers that evaluate the system in a rather static manner.

On the other hand, because of disease complexities and personal variations (e.g., genetic or epigenetic factors), each individual may progress to the same disease through different underlying networks, which hampers the discovery of effective molecule- or model-based biomarkers. To overcome this difficulty, we have developed this model-free method based on measured individual data, and therefore we can theoretically detect specific signals for each individual or potentially apply this method to personalized medicine. It should be noted that our prediction merely requires a few samples in each sampling period, in stark contrast to the consecutive time-series data over the entire period required by the traditional methods. Such early-warning signals can be applied to efficient prognosis as well as products of nichetargeting medicine. In addition to the theoretical conditions, we have also provided a computational algorithm to numerically identify the DNB based on high-throughput data. As the first step, this theoretical work is to challenge the complicated problem of detecting the pre-disease state by studying tissue based data. However, analyzing peripheral blood or plasma-based data is a practical approach to diagnosis at the molecular level, which we will consider in our future work. In addition, because different driving factors may result in different DNBs even for the same disease, the analysis of the difference and similarity of these DNBs is also an important future topic.

Methods

Derivation of DNB. The theoretical results were derived by considering the linearized equations[17,27] of Eq.(1) with the perturbations of noise near $\bar{Z}$. Namely, introducing new variables $Y(k) = (y_1(k), \ldots, y_n(k))$ and a transformation matrix S, *i.e.*, $Y(k) = S^{-1}(Z(k) - \bar{Z})$ (Supplementary Information A1), we have

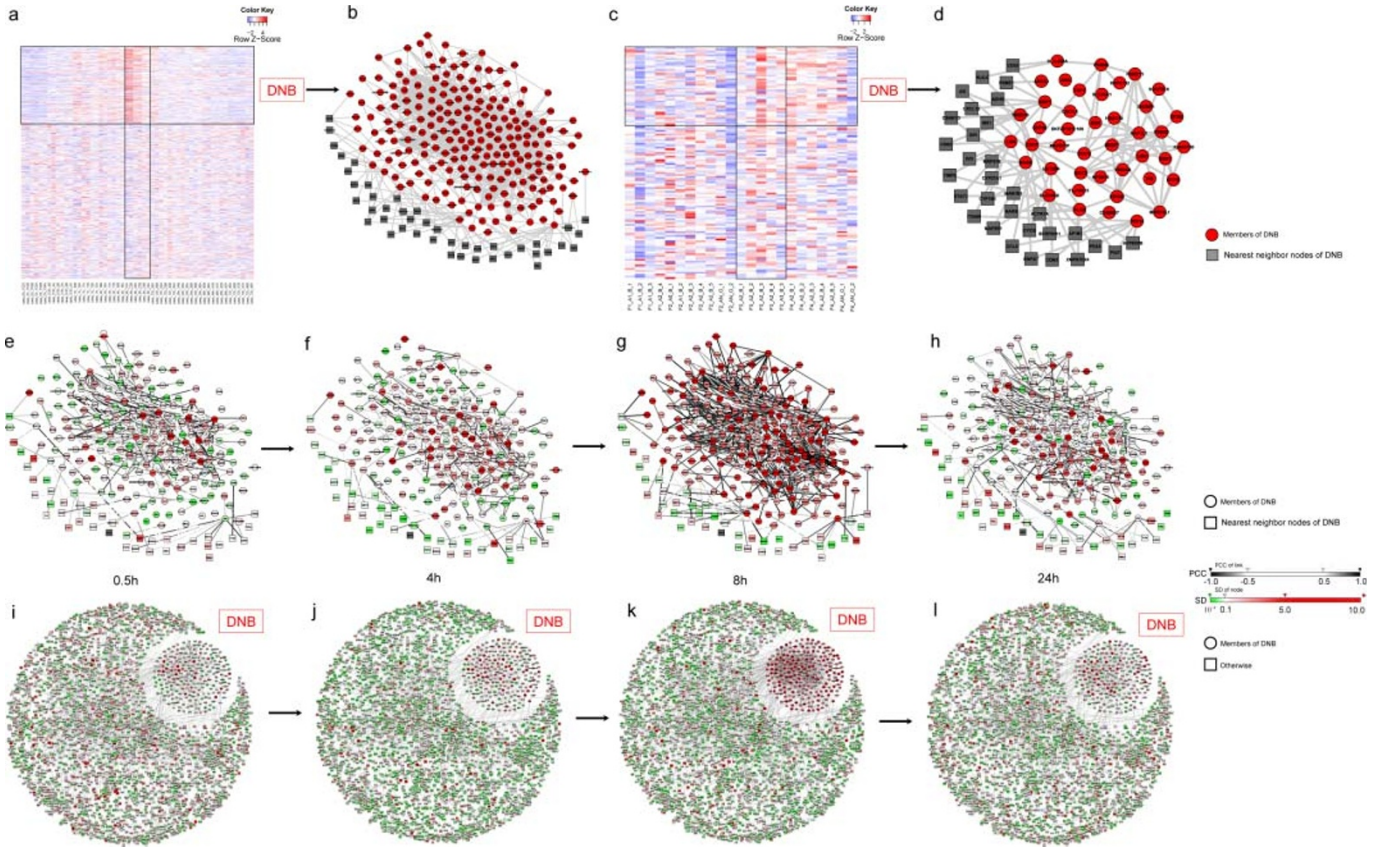

Figure 4 | DNBs for two complex diseases. (a) and (c) show expression profiles of the DNB genes and other genes (randomly selected genes with two times the size of the DNB) for the acute lung injury and the HBV induced liver cancer, respectively, which also indicate that genes in each DNB during the pre-disease period are correlated with strong deviation. Each horizontal part boxed by lines is the DNB, and each vertical part boxed by lines is the pre-disease period in (a) or (c) for the respective disease. The profiles for entire genes are described in Supplementary Information F. (b) and (d) show the identified DNBs for the acute lung injury and the HBV induced liver cancer, respectively. The DNB and whole mouse network are linked by the documented functional interactions from various databases (see Methods). Genes in each DNB are indicated in red and some of their nearest neighbors are indicated by grey nodes in (b) and (d). For acute lung injury, we also show the dynamic evolution of the network structure for the identified DNB (220 genes and 1167 links) and the whole mouse network (3452 genes and 9238 links) including the DNB. (e) The DNB at 0.5 h. (f) The DNB at 4 h. (g) The DNB at 8 h (the pre-disease state). (h) The DNB at 24 h. (i) The whole mouse network at 0.5 h. (j) The whole mouse network at 4 h. (k) The whole mouse network at 8 h (the pre-disease state). (l) The whole mouse network at 24 h. The dynamic evolution of the DNB for total 8 time points is shown in Supplementary Fig. S8 and the corresponding dynamics of the whole mouse network is shown in Supplementary Fig. S9.

$$Y(k+1) = \Lambda(P)Y(k) + \xi(k), \tag{3}$$

where $\Lambda(P)$ is the diagonalized matrix of $\frac{\partial f(Z;P)}{\partial Z}\Big|_{Z=\bar{Z}}$. $\xi(k) = (\xi_1(k), \ldots, \xi_n(k))$ are Gaussian noises with zero means and covariances $\kappa_{ij} = \mathrm{Cov}(\xi_i, \xi_j)$. Actually, there are typically three cases arising in the diagonalization process, and we present the detailed derivation in Supplementary Information A2 and here only illustrate the diagonal case with real eigenvalues for simplicity, *i.e.*, $\Lambda(P) = diag(\lambda_1(P), \ldots, \lambda_n(P))$ with each $|\lambda_i|$ between 0 and 1. Among the eigenvalues of Λ, the largest one in modulus, say λ_1, first approaches to 1 when parameter $P \to P_c$. This eigenvalue characterizes the system's rate of change around the fixed point and is called the dominant eigenvalue. Under the assumption that $Y(k)$ have zero means, the covariance and PCC are respectively represented as

$$\mathrm{Cov}(y_i, y_j) = \frac{k_{ij}}{1-\lambda_i\lambda_j},$$

$$\mathrm{PCC}(y_i, y_j) = \frac{k_{ij}}{\sqrt{k_{ii}k_{jj}}} \frac{\sqrt{\left(1-\lambda_i^2\right)\left(1-\lambda_j^2\right)}}{1-\lambda_i\lambda_j}.$$

Note that the variance $\mathrm{Var}(y_i) = \mathrm{Cov}(y_i, y_i)$. Hence when the dominant eigenvalue $|\lambda_1| \to 1$ because of the change in the parameter values, $\mathrm{Var}(y_1) \to +\infty$, but $\mathrm{Var}(y_i)$ is bounded for $i \neq 1$ because $0 < |\lambda_i| < |\lambda_1| < 1$ $(i = 2, \ldots, n)$. For any $i \neq j$, $\mathrm{Cov}(y_i, y_j)$ tends to have a positive bounded value.

Returning to the original variables Z whose values are directly measured as high-throughput data, by $z_i = \sum_{j=1}^{n} s_{ij}y_j + \bar{z}_i$ where s_{ij} is the element of the linear transformation S, we have (Supplementary Information A3)

$$\mathrm{Cov}(z_i, z_j) = s_{i1}s_{j1}\mathrm{Var}(y_1) + \cdots + s_{in}s_{jn}\mathrm{Var}(y_n) + \sum_{k,m=1,k\neq m}^{n} s_{ik}s_{jm}\mathrm{Cov}(y_k, y_m),$$

$$\mathrm{PCC}(z_i, z_j) = \frac{\mathrm{Cov}(z_i, z_j)}{\sqrt{\mathrm{Var}(z_i)\mathrm{Var}(z_j)}}.$$

Notice that variable y_1 is related to the dominant eigenvalue λ_1. From the above equations, it is obvious that as $|\lambda_1| \to 1$, the SD, *i.e.*, $\sqrt{\mathrm{Var}(z_i)} = \sqrt{\mathrm{Cov}(z_i, z_i)}$ increases greatly, or $\mathrm{Var}(z_i) \to +\infty$ if s_{i1} is not vanishing, and $|\mathrm{PCC}(z_i, z_j)|$ approaches 1 drastically if both s_{i1} and s_{j1} are non-zero. In this case, variables z_i and z_j are directly affected by the dominant eigenvalue. A group composed of such variables is called the dominant group in the network. On the other hand, as $|\lambda_1| \to 1$, $|\mathrm{PCC}(z_i, z_j)|$ between the dominant group (e.g., including z_i) and others (e.g., including z_j, which does not belong to the dominant group) reduces to zero if $s_{i1} \neq 0$ but $s_{j1} = 0$. Hence, we theoretically derived the three criteria to detect the dominant group or the DNB (see Supplementary Information A for more details). Clearly, a DNB can be viewed as a temporally dynamical module (TDM), whose members are strongly correlated with each other (i.e., a module feature) but their expressions stochastically fluctuate without keeping constant values (i.e., a dynamical or stochastic feature) only during the critical transition period (i.e., a temporal or transient feature). Hence, the presence of the DNB can be observed by requiring the measurement of not only correlations but also dynamics. In particular, it appears in the critical transition period but disappears (or merges with other components) in other periods, thereby signaling the upcoming critical transition. In addition to the disease progression, our theoretical results based on the bifurcation and center manifold theory[23] may be widely applied to the detection of suddenly changing phenomena of various complex networks in a similar manner.

Table 1 | Functional enrichment of GO biological processes and KEGG pathway enrichment analysis in the identified DNBs of two diseases. Parts of the genes of DNBs are shown in this table (see Supplementary Table 'Identified DNBs' and Supplementary Information F for complete lists). Note that the identified DNBs are in the respective pre-disease states, rather than in the disease states

Disease	DNB	GO term	P-value	Description
Acute	{ GCLC; ASNS; PGD; EPHA2;	GO:0006979	7.6E-10	response to oxidative stress
Lung	TXNL1; SRXN1; CH25H;	GO:0009611	1.4E-6	response to wounding
Injury	HSPA1A;HSPA1B;CYP51;	GO:0050727	1.5E-8	oxidation reduction
	DNAJC5; DNAJB4; HMOX1;	GO:0006629	6.9E-6	lipid biosynthetic process
	GADD45G; PMAIP1; ... }	GO:0055114	2.7E-5	regulation of inflammatory response
HBV induced	{ B2M; GCC2; HDAC10;	GO:0010629	0.0029	antigen processing and presentation
liver	STAT6; CACH-1;	GO:0019882	0.0056	intracellular protein transport
cancer	HLA-DMA; TAP1;	GO:0006886	0.0189	chromosome organization
	CDH1; YWHAB;	GO:0006325	0.023	negative regulation of gene expression
	NAP1L1; ... }	GO:0045191	0.024	regulation of isotype switching

Acute lung injury		HBV induced liver cancer	
Pathway term	P-value	Pathway term	P-value
Pathway in cancer	1.54E-6	Acute myeloid leukemia	1.28E-4
MAPK signaling pathway	4.42E-6	Pathways in cancer	5.51E-4
p53 signaling pathway	6.57E-6	Leishmaniasis	1.43E-3
Chronic myeloid leukemia	4.08E-5	Pentose and glucuronate interconversions	3.51E-3
Hepatitis C	4.71E-5	Cytosolic DNA-sensing pathway	3.70E-3
Adipocytokine signaling pathway	7.95E-5	Adipocytokine signaling pathway	4.45E-3
Acute myeloid leukemia	5.12E-5	Hepatitis C	5.56E-3
Leishmaniasis	9.02E-5	Pancreatic cancer	6.17E-3
Cell cycle	1.14E-4	Toll-like receptor signaling pathway	6.80E-3

Functional analysis. Genes in the DNBs for the three diseases have been linked and correlated by the combined functional couplings among them from various databases of protein-protein interactions of STRING, FunCoup and BioGrid, transcriptional regulations of TRED, and metabolic pathways of KEGG (Supplementary Information F). Here, we considered the diseases of lung injury and HBV infection as examples to show the detailed computing procedure. We also compared the differences between our DNB-based method and the traditional methods, i.e., the fold-change method and the random gene selection method, based on respective gene sets (Supplementary Information E and G). Moreover, we also performed the pathway enrichment analysis of the identified DNBs (Supplementary Information F). The applicable computer program used for detecting the DNB from high-throughput data is described in the Supplementary Information of this paper.

1. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. &Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
2. Drake, M. J. & Griffen, D. B. Early warning signals of extinction in deteriorating environments. *Nature* **467**, 456–459 (2010).
3. Dakos, V. *et al.* Slowing down as an early warning signal for abrupt climate change. *Proc. Natl Acad. Sci. USA* **105**, 14308–14312 (2008).
4. Lenton, T. M. *et al.* Tipping elements in the earth's climate system. Proc. *Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
5. Kambhu, J., Weidman, S. & Krishnan, N. *New Directions for Understanding Systemic Risk: A Report on a Conference Cosponsored by the Federal Reserve Bank of New York and the National Academy of Sciences* (The National Academies Press, Washington D.C., 2007).
6. May, R. M., Levin, S. A. & Sugihara, G. Ecology for bankers. *Nature* **451**, 893–895 (2008).
7. Gilmore, R. *Catastrophe Theory for Scientists and Engineers* (Dover, New York, 1981).
8. Murray, J. D. *Mathematical Biology* (Springer, New York, 1993).
9. Venegas, J. G. *et al.* Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature* **434**, 777–782 (2005).
10. McSharry, P. E., Smith, L. A. & Tarassenko, L. Prediction of epileptic seizures: are nonlinear methods relevant? *Nature Med.* **9**, 241–242 (2003).
11. Roberto, P. B., Eliseo, G. & Josef, C. Transition models for change-point estimation in logistic regression. *Statist. Med.* **22**, 1141–1162 (2003).
12. Paek, S., *et al.* Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma. *Cancer* **104**, 580–590 (2005).
13. Liu, J. K., Rovit, R. L. & Couldwell, W. T. Pituitary Apoplexy. *Seminars in Neurosurgery* **12**, 315–320 (2001).
14. Tanaka, G., Tsumoto, K., Tsuji, S. & Aihara, K. Bifurcation analysis on a hybrid systems model of intermittent hormonal therapy for prostate cancer. *Physica D* **237**, 2616–2627 (2008).
15. Litt, B. *et al.* Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* **30**, 51–64 (2001).
16. Achiron, A. *et al.* Microarray analysis identifies altered regulation of nuclear receptor family members in the pre-disease state of multiple sclerosis. *Neurobiology of Disease* **38**, 201C209 (2010).
17. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
18. Carpenter, S. R. *et al.* Early warnings of regime shifts: a whole-ecosystem experiment. *Science* **332**, 1079–1082 (2011).
19. Carpenter, S. R. & Brock, W. A. Rising variance: a leading indicator of ecological transition. *Ecol. Lett.* **9**, 311–318 (2006).
20. Carpenter, S. R. Eutrophication of aquatic ecosystems: bistability and soil phosphorus. *Proc. Natl Acad. Sci. USA* **102**, 10002–10005 (2005).
21. Held, H. & Kleinen, T. Detection of climate system bifurcations by degenerate fingerprinting. *Geophys. Res. Lett.* **31**, L23207 (2004).
22. Kleinen, T., Held, H. & Petschel-Held, G. The potential role of spectral properties in detecting thresholds in the earth system: application to the thermohaline circulation. *Ocean Dyn.* **53**, 53–63 (2003).
23. Strogatz, S. H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering* (Addison-Wesley, Reading, Massachussets, 1994).
24. Van Nes, E. H. & Scheffer, M. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *Am. Nat.* **169**, 738–747 (2007).
25. Wissel, C. A universal law of the characteristic return time near thresholds. *Oecologia* **65**, 101–107 (1984).
26. Guttal, V. & Jayaprakash, C. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecol. Lett.* **11**, 450–460 (2008).
27. Dakos, V., Van Nes, E. H., Donangelo, R., Fort, H. & Scheffer, M. Spatial correlation as leading indicator of catastrophic shifts. *Theor. Ecol.* **3**, 163–174 (2010).
28. Donangelo, R., Fort, H., Dakos, V. & Scheffer, M. Early warning signals for catastrophic shifts in ecosystems: comparison between spatial and temporal indicators. *Int. J. Bifur. Chaos* **20**, 315–321 (2010).
29. Chen, L., Wang, R. & Zhang, X. *Biomolecular networks: methods and applications in systems biology* (John Wiley & Sons, Hoboken, New Jersey, 2009).
30. Chen, L., Wang, R., Li, C. & Aihara, K. *Modeling biomolecular networks in cells: structures and dynamics* (Springer, New York, 2010).
31. Sciuto, A. M. *et al.* Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation. *Chem. Res. Toxicol.* **18**, 1654–1660 (2005).
32. Honda, M., Kaneko, S., Kawai, H., Shirota, Y. & Kobayashi, K. Differential gene expression between chronic hepatitis B and C hepatic lesion. *Gastroenterology* **120**, 955–966 (2001).
33. Lenburg, M. E., Sinha, A., Faller, D. V. & Denis, G. V. Tumor-specific and proliferation-specific gene expression typifies murine transgenic B cell lymphomagenesis. *J. Biol. Chem.* **282**, 4803–4811 (2007).

Acknowledgements

We thank Drs. Jiarui Wu and Rong Zeng of Chinese Academy of Sciences (CAS) for helpful discussions, and also thank Drs. Masao Honda and Shuichi Kaneko of Kanazawa University for providing the data of reference 32. This work was supported by NSFC under No. 91029301, No. 61134013, No. 61072149 and No. 31100949, Shanghai Pujiang Program,

the Chief Scientist Program of SIBS of CAS with No. 2009CSP002, the Knowledge Innovation Program of SIBS of CAS with Grant No. 2011KIP203, the Knowledge Innovation Program of CAS with No. KSCX2-EW-R-01, National Center for Mathematics and Interdisciplinary Sciences of CAS, and JSPS/CSTP through the FIRST Program.

Author contributions

L.C. and K.A. conceived the research. L.C., R.L. and Z.P.L. designed the numerical simulation and real data analysis. R.L., M.L. and Z.P.L. performed the numerical experiments. Z.P.L. and M.L. contributed the functional analysis. All authors wrote the paper and approved the final manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing interests statement: The authors declare no competing financial interests.

How to cite this article: Chen, L., Liu, R., Liu, Z., Li, M. & Aihara, K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Rep.* **2**, 342; DOI:10.1038/srep00342 (2012).