Supplementary Information: Detecting the early-warning signal of the critical transition by hidden Markov model

Pei Chen, Rui Liu, Yongjun Li, Luonan Chen

Contents

A Theoretical background

A.1 Theoretical basis near critical transition point

For a complex dynamical system with multiple variables or network, assume that we measure the variables at different time points. In this section, we theoretically introduce several generic properties of such a dynamical network when the system approaches a critical transition point. Specifically, we derive the conditions to obtain dynamical network marker (DNM) in a general system or dynamical network biomaker (DNB) in a biological system, which can characterize the generic properties and predict the critical transition, based on bifurcation theory and center manifold theory (*6, 7*).

We consider the following discrete-time dynamical system that represents the dynamical evolution of a network:

$$Z(t+1) = f(Z(t); P), \tag{S1}$$

where $Z(t) = (z_1(t), ..., z_n(t))$ is an n-dimensional state vector or variables at time instant t, while $P = (p_1, ..., p_s)$ is a parameter vector or driving factors that represent slowly changing factors, e.g., genetic factors (SNP, CNV, etc.) and epigenetic factors (methylation, acetylation, etc.). $f : \mathbf{R}^n \times \mathbf{R}^s \to \mathbf{R}^n$ are generally nonlinear functions.

Furthermore, we assume that the following conditions hold for Eq.(S1).

1. $\bar{Z}$ is a fixed point in system (S1) such that $\bar{Z} = f(\bar{Z}; P)$.

2. There is a value P_c such that one or a pair of the eigenvalues of the Jacobian matrix $\left.\frac{\partial f(Z;P_c)}{\partial Z}\right|_{Z=\bar{Z}}$ is equal to 1 in the modulus.

3. When $P \neq P_c$, the eigenvalues of (S1) are not always equal to 1 in the modulus.

These three assumptions with other transverse conditions imply that the system undergoes a phase change at $\bar{Z}$ or a codimension-one bifurcation when P reaches the threshold P_c (*1*).

From a mathematical perspective, the bifurcation is generic, *i.e.* almost all of the bifurcations in a general system satisfy these conditions. It is notable that most of the systems described by differential equations can be generally discretized and transformed into Eq.(S1), e.g., using methods such as the Euler scheme and the Poincaré section. Thus, we focus on difference equations (S1) during our theoretical analysis in this section.

It is known that the dynamics of a nonlinear system is highly complex far before or after a sudden transition; therefore, the state equations of systems are generally constructed in a very high-dimensional space using a large number of variables and parameters (*1–4, 12, 14*). However, if a system driven by known or unknown parameters approaches a critical point, which is a very special phase during its dynamical evolution, it is theoretically guaranteed that the system will eventually be constrained to one- or two-dimensional space (i.e., the center manifold), which can be expressed in a simple form around a codimension-one bifurcation point (*5,6,12,14*). This is generally guaranteed by the bifurcation theory and the center manifold theory (*5–8*). Thus, we can detect the signal of any dynamical system only during this special phase and not during other periods (i.e., neither the before-transition state nor the after-transition state), which is part of the theoretical foundation of this study (*12, 14*).

For system (S1) near $\bar{Z}$ and before P reaches P_c, we assume that the system is at a stable fixed point $\bar{Z}$, so all of the eigenvalues are within $(0, 1)$ in the modulus. The parameter value P_c at which the state shift of the system occurs, is known as a bifurcation parameter value or a critical transition value.

This theoretical result was derived based on consideration of the linearized system or equations for Eq.(S1) and the small noise perturbations near $\bar{Z}$. Specifically, by introducing the new variables $Y(t) = (y_1(t), ..., y_n(t))$ and a transformation matrix S, *i.e.*

$$Y(t) = S^{-1}(Z(t) - \bar{Z}),$$

or

$$z_i(k) = s_{i1}y_1(k) + \cdots + s_{in}y_n(k) + \bar{z}_i, \quad k, i = 1, 2, ..., n, \tag{S2}$$

we have

$$Y(t+1) = \Lambda Y(t) + \zeta(t). \tag{S3}$$

where $\Lambda(P)$ is the diagonalized matrix of $\frac{\partial f(Z;P)}{\partial Z}\Big|_{Z=\bar{Z}}$. $\zeta(t) = (\zeta_1(t), ..., \zeta_n(t))$ are small Gaussian noises with zero means. Without any loss of generality, the diagonalized matrix $\Lambda(P) = diag(\lambda_1(P), ..., \lambda_n(P))$ for each $|\lambda_i|$ is between 0 and 1. Denote the dominant eigenvalue as the largest eigenvalue or eigenvalues (the case of multiple roots) in modulus. Here, the largest eigenvalue (in the sense of modulus) characterizes the system's rate of change around a fixed point and is known as the dominant eigenvalue. The before-transition state corresponds to a period when the dominant eigenvalue is smaller than 1 in modulus, whereas the pre-transition stage corresponds to the period with the dominant eigenvalue approaching to 1 in modulus. Actually, in view of the dominant eigenvalue, there are two typical cases during the diagonalization process (*12*), i.e., the dominant eigenvalue is real (including multiple real dominant eigenvalues), and the dominant eigenvalues are a pair of complex conjugate values.

When the modulus of the largest eigenvalue or eigenvalue pairs approaches 1, there are three generic codimension-one bifurcations of the system, corresponding to these cases, that is the saddle-node bifurcation, period-doubling bifurcation and Neimark-Sacker bifurcation. Specifically, when the dominant eigenvalue is real, the critical point is the saddle-node bifurcation (transcritical and pitchfork bifurcation) if the dominant eigenvalue approaches 1, while the critical point is the period-doubling (or flip) bifurcation if the dominant eigenvalue approaches -1. When the dominant eigenvalues are a pair of complex conjugate eigenvalues (including several pairs with the same modulus), for such a case, the critical point is the Neimark-Sacker bifurcation point (*1*).

According to our previous works (*12, 14*), we have the following results.

1. When the dominant eigenvalues are real, then there is only one dominant group related to the first variable y_1 in Eq.(S3), which is with standard deviation

$$\mathrm{SD}(z_i) = \sqrt{s_{i1}^2 \frac{\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^{n} s_{ik}^2 \frac{\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k\neq m}^{n} s_{ik}s_{im}\frac{\kappa_{km}}{1-\lambda_k\lambda_m}},$$

and Pearson's coefficient correlation

$$\mathrm{PCC}(z_i, z_j) = \\ = \frac{s_{i1}s_{j1}\frac{\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^{n} s_{ik}s_{jk}\frac{\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k\neq m}^{n} s_{ik}s_{jm}\frac{\kappa_{km}}{1-\lambda_k\lambda_m}}{\sqrt{\left(\frac{s_{i1}^2\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^{n}\frac{s_{ik}^2\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k\neq m}^{n}\frac{s_{ik}s_{im}\kappa_{km}}{1-\lambda_k\lambda_m}\right)\left(\frac{s_{j1}^2\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^{n}\frac{s_{jk}^2\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k\neq m}^{n}\frac{s_{jk}s_{jm}\kappa_{km}}{1-\lambda_k\lambda_m}\right)}},$$

where λ_1 represents the dominant eigenvalue while λ_k is other eigenvalue with $|\lambda_k| \leq |\lambda_1|$, the constant s_{jk} is the coefficient in Eq.(S2), and κ_{ij} is the covariances between Gaussian noises ζ_i and ζ_j in Eq.(S2).

2. When the dominant eigenvalues are a pair of complex conjugate eigenvalues, then there are two dominant groups respectively related to y_1 and y_2 in Eq.(S3). The standard deviation for the variables z_i is

$$\mathrm{SD}(z_i) = \sqrt{\frac{2b^2(s_{i1}^2 + s_{i2}^2)(\kappa_{11}+\kappa_{22})}{(1-a^2-b^2)((a-1)^2+b^2)((a+1)^2+b^2)} + K_i}$$

where K_i are bounded values, and Pearson's coefficient correlation

$$\mathrm{PCC}(z_i, z_j) = \\ = \frac{\left(\frac{2b^2(s_{i1}s_{j1}+s_{i2}s_{j2})(\kappa_{11}+\kappa_{22})}{(1-a^2-b^2)((a-1)^2+b^2)((a+1)^2+b^2)}\right) + C_{ij}}{\sqrt{\left(\frac{2b^2(s_{i1}^2+s_{i2}^2)(\kappa_{11}+\kappa_{22})}{(1-a^2-b^2)((a-1)^2+b^2)((a+1)^2+b^2)} + K_i\right)\left(\frac{2b^2(s_{j1}^2+s_{j2}^2)(\kappa_{11}+\kappa_{22})}{(1-a^2-b^2)((a-1)^2+b^2)((a+1)^2+b^2)} + K_j\right)}}.$$

where C_{ij}, K_i and K_j are bounded values.

Theorem 1 *We consider a stochastically perturbed linearized system for Eq.(S1). When P approaches the saddle-node or period-doubling bifurcation point, there is a dominant group, and the following results hold.*

- *If both z_i and z_j are in the dominant group, then*

$$|\mathrm{PCC}(z_i, z_j)| \to 1,$$

  while $\mathrm{SD}(z_i) \to \infty$ *and* $\mathrm{SD}(z_j) \to \infty$;

- *if z_i is in the dominant group but z_j is not, then*

$$\mathrm{PCC}(z_i, z_j) \to 0,$$

  while $\mathrm{SD}(z_i) \to \infty$*, and* $\mathrm{SD}(z_j)$ *approaches a bounded value;*

- *if neither z_i nor z_j is in the dominant group, then* $\mathrm{PCC}(z_i, z_j)$ *approaches a constant, while both* $\mathrm{SD}(z_i)$ *and* $\mathrm{SD}(z_j)$ *approach bounded values,*

where PCC is the Pearson's correlation coefficient and SD is the standard deviation.

This dominant group of variables or elements is DNM. This theorem is the part of the theoretical basis of detecting the pre-transition state for multi-variable systems with small noise. The three conditions in the theorem are actually the criteria to detect the DNM (*12*). For a nonlinear system Eq.(S1) approaching the bifurcation point (the saddle-node or period-doubling bifurcation point) or tipping point, we can observe directly obtain the properties of DNM based on Theorem 1 as the following remark.

Remark 1 *For nonlinear case Eq.(S1) near a bifurcation point (the saddle-node or period-doubling bifurcation point), the dynamical behavior has the same tendency as that of the linearized case, that is, when the system is approaching to the bifurcation point, both indices*

SD *and* $|\mathrm{PCC}|$ *in the DNM increase sharply, while* $|\mathrm{PCC}|$ *between DNM and other non-DNM molecules decreases rapidly, i.e.,*

1. *If both* z_i *and* z_j *are in the DNM, then* $\mathrm{PCC}(z_i, z_j)$ *increases, while* $\mathrm{SD}(z_i)$ *and* $\mathrm{SD}(z_j)$ *drastically increase;*

2. *if* z_i *is in the DNM but* z_j *is not, then* $\mathrm{PCC}(z_i, z_j)$ *decreases, while* $\mathrm{SD}(z_i)$ *drastically increases, and there is no significant change for* $\mathrm{SD}(z_j)$*;*

3. *if neither* z_i *nor* z_j *is in the DNM, then there are no significant changes on* $\mathrm{PCC}(z_i, z_j)$, $\mathrm{SD}(z_i)$ *and* $\mathrm{SD}(z_j)$.

Theorem 2 *We consider a stochastically perturbed linearized system for Eq.(S1). When* P *approaches the Neimark-Sacker bifurcation point, the following results hold.*

- *If both* z_i *and* z_j *are in the same dominant group 1 or 2, then*

$$|\mathrm{PCC}(z_i, z_j)| \to 1,$$

  while $\mathrm{SD}(z_i) \to \infty$ *and* $\mathrm{SD}(z_j) \to \infty$;

- *if* z_i *is in dominant group 1 and* z_j *in dominant group 2, then*

$$\mathrm{PCC}(z_i, z_j) \to 0,$$

  while $\mathrm{SD}(z_i) \to \infty$*, and* $\mathrm{SD}(z_j) \to \infty$*;*

- *if* z_i *is in a dominant group (including dominant 1 or 2, common-dominant group) and* z_j *is in the non-dominant group, then*

$$\mathrm{PCC}(z_i, z_j) \to 0,$$

  while $\mathrm{SD}(z_i) \to \infty$ *and* $\mathrm{SD}(z_j)$ *approaches a bounded value;*

- *if z_i is in the common dominant group, and z_j is in dominant group k $(k = 1, 2)$ or the common dominant group, then $|\mathrm{PCC}(z_i, z_j)|$ approaches a constant less than 1, while $\mathrm{SD}(z_i) \to \infty$, and $\mathrm{SD}(z_j) \to \infty$;*

- *if neither z_i nor z_j is in the dominant group, then $|\mathrm{PCC}(z_i, z_j)|$ approaches a constant less than 1, while both $\mathrm{SD}(z_i)$ and $\mathrm{SD}(z_j)$ approach bounded values;*

where PCC is the Pearson's correlation coefficient and SD is the standard deviation.

Remark 2 *When both z_i and z_j are in the common dominant group, if their relations to y_1 are much more stronger than that to y_2, then generally z_i and z_j are considered in dominant group 1. Contrarily, if the relations of z_i and z_j to y_2 are much more stronger than that to y_1, then generally z_i and z_j are considered in dominant group 2.*

A.2 Three states during a critical transition

During a state transition, the dynamics of the system can be divided as three stages, as shown in Fig.S1. Before-transition state corresponds to a stable equilibrium. In bio-medical systems, it generally corresponds to a "normal state" or a stable period that the disease is under control. Pre-transition state is the limit of Before-transition state. In bio-medical systems, it represents a "pre-disease state" just before the critical transition to the disease state. After-transition state corresponds to another stable equilibrium. In bio-medical systems, it represents a badly ill stage or "disease state", and is usually difficult to return to the before-transition state even by big perturbations.

A.3 Identifying the switching point of stationary Markov process based on inconsistence index

In our assumption, the stage before the transition point is quite different from that after the transition point (Fig.S1). Therefore, we regard that the progression of a biological system in

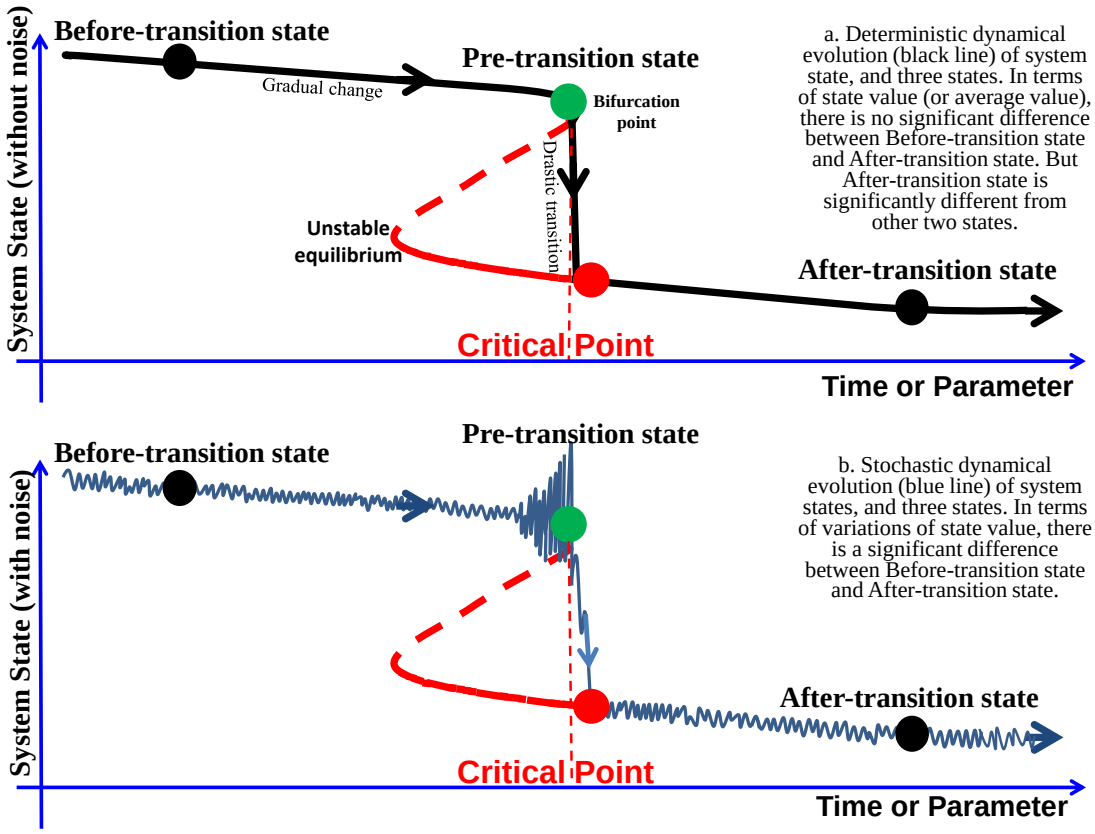

Figure S1: | **Definition of three states in a transition process.** Before-transition state is a stable equilibrium. Pre-transition state is the limit of the Before-transition state. After-transition state is another stable equilibrium. **(a)**. Deterministic dynamical evolution (black line) of system state with time or parameter. **(b)**. Stochastic dynamical evolution (blue line) of system states with time or parameter. In terms of state value (i.e., when it is a deterministic system or we only consider average value of a stochastic system), there is no significant difference between before-transition state and after-transition state. But after-transition state is significantly different from other two states. In terms of variations of state value (when it is fluctuated by noise or it is a stochastic system), there is a significant difference between the before-transition state and the after-transition state. In a real system, the system is always fluctuated by noise, and thus the process of a state transition can be characterized by three states. The green circle represents the pre-transition or critical state, while the red one is the state immediately after the transition. We aim to detect the pre-transition state or green circle, rather than the red circle.

the before-transition stage is a stationary Markov process, since during this stage the system is stable and insensitive to parameter changes (stationary feature). However, the progression of a biological system in the pre-transition stage is modelled as a time-varying Markov process, due

to the strong fluctuated dynamics during this stage, i.e., the system is unstable and so sensitive to the parameter changes that even a small change in the parameters may suffice to drive the system into collapse (time-varying feature). Therefore, to detect the onset of the pre-transition stage is equivalent to identify the changing period or the switching point from a stationary Markov process to a time-varying Markov process.

On the basis of above settings and hypothesis, we propose an inconsistence index (I-index) to measure the probability of a candidate time point as the changing or switching point from the stationary Markov process to the time-varying Markov process (Fig.S2). Specifically, we test each candidate time point based on a HMM derived from an unsupervised learning procedue.

A.3.1 Deriving the HMM based on an unsupervised learning procedure

In order to simplify the derivation of the HMM for of an n-variable system during its progression, a few notations are presented as follows .

- Denote the time variable as t. The progression of a system is along the time series $\{1, 2, ..., t-1, t, ...\}$.

- Denote the observed sequence up to time point t as $O = \{o_1, o_2, ..., o_{t-1}, o_t\}$, where o_t represents the sample set derived at time point t. Furthermore, if there are m samples for the system at time t, then the observed sample set o_t is

$$o_t = (Z^1(t), Z^2(t), ..., Z^m(t)) = \begin{pmatrix} z_1^1(t) & z_1^2(t) & \cdots & z_1^m(t) \\ z_2^1(t) & z_2^2(t) & \cdots & z_2^m(t) \\ \vdots & \vdots & \ddots & \vdots \\ z_n^1(t) & z_n^2(t) & \cdots & z_n^m(t) \end{pmatrix}_{n \times m},$$

  where $z_i^s(t)$ represents the expression or value of the i th variable in the s th sample at

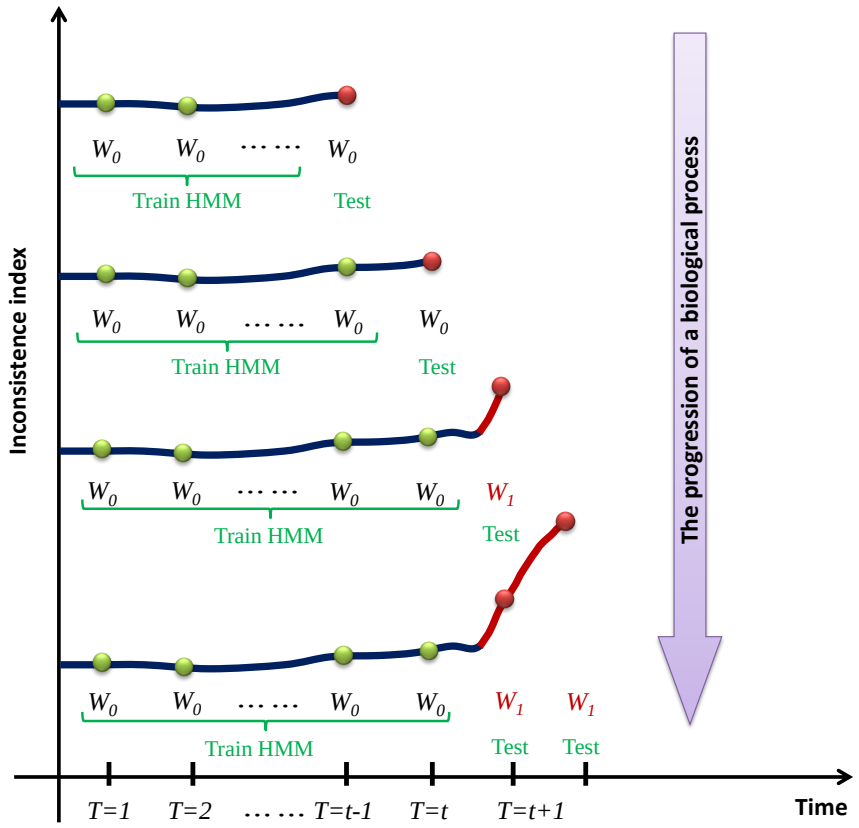

Figure S2: | **The dynamical change of the inconsistence index based on HMM** The sketch illustrates that the inconsistence index remains steady when the progression of a biological process is in a before-transition stage (state W_0), and increases suddenly when the process is in a pre-transition stage (state W_1). Specifically, based on the samples from time point $1, 2, ..., t-1$, the HMM θ_{t-1} is trained. We then calculate the inconsistence index, or HMM-based probability P_t measuring the inconsistence of sample o_t derived at time point t and the HMM θ_{t-1}.

time t with $i \in \{1, 2, ..., n\}_{variable}$ and $s \in \{1, 2, ..., m\}_{sample}$, $Z^s(t-1) = \begin{pmatrix} z_1^s(t) \\ z_2^s(t) \\ \vdots \\ z_n^s(t) \end{pmatrix}$

represents the s th sample of the system.

- Denote the state sequence up to time point t as $\{s_1, s_2, ..., s_{t-1}, s_t\}$, i.e., the state of the

system is s_t at time point t, or equivalently, $s_t = State(o_t)$.

- Denote the unobserved (hidden) states as W_0 and W_1, where W_0 represents the system state before the transition point, and W_1 stands for any system state that is not identical with W_0, i.e., the state not in the before-transition stage. Therefore, each s_i from the state sequence could be either $s_t = W_0$ if the system is at the before-transition stage at time t, or $s_t = W_1$ if the system is not any more at the before-transition stage at time t.

- For given HMM θ and observation O, denote $\gamma_t(i)$ as the probability of the system being at state W_i at time t, i.e.,

$$\gamma_t(i) = P(s_t = W_i \,|\, O,\, \theta) = \frac{P(s_t = W_i,\, O \,|\, \theta)}{P(O \,|\, \theta)} \tag{S4}$$

  where $i \in \{0, 1\}$.

- For given HMM θ and observation O, denote $\xi_t(i, j)$ as the probability of the system being at state W_i at time $t-1$ and being at state W_j at time t, i.e.,

$$\xi_t(i, j) = P(s_{t-1} = W_i, s_t = W_j \,|\, O,\, \theta) = \frac{P(s_{t-1} = W_i, s_t = W_j,\, O \,|\, \theta)}{P(O \,|\, \theta)} \tag{S5}$$

  where $i, j \in \{0, 1\}$.

We train a hidden Markov model (HMM) $\theta_{t-1} = (A, B, \pi)$ based on the basis of an observed sequence $\{o_1, o_2, ..., o_{t-1}\}$, i.e., the first $t-1$ sets of samples from time points $1, 2, ..., t-1$, where the subscript $t-1$ of θ represents that the HMM is derived from the training samples up to time point $t-1$, A is a state transition matrix, B is an emission matrix, and π is a probability vector for the initial state. The training process based on an unsupervised learning procedure, namely, Baum-Welch algorithm, is provided as the following four steps.

Step 1. Estimate the distribution at a former time point $(t-2)$**.**

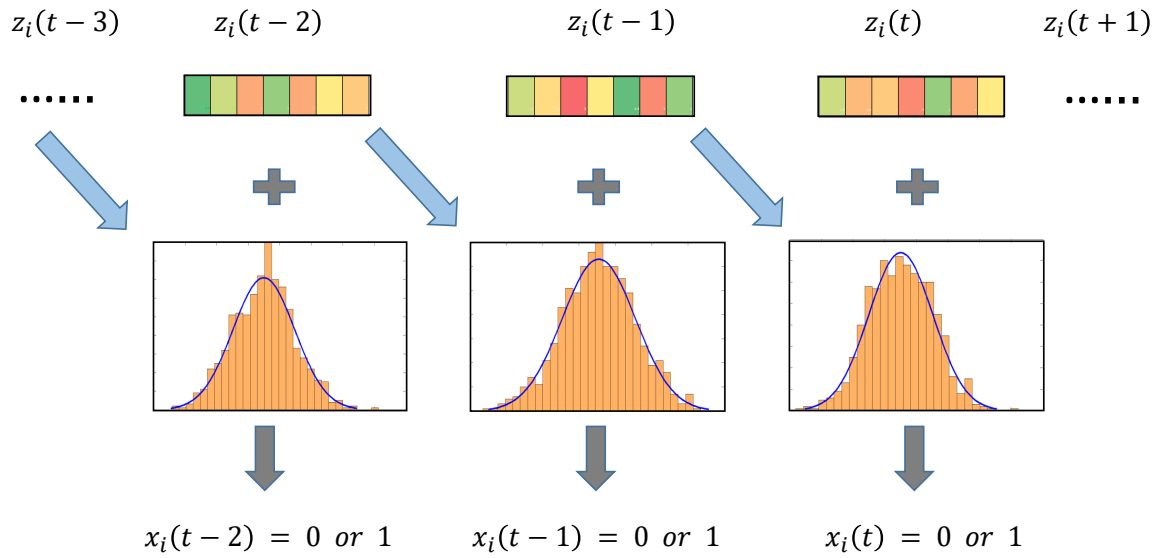

Figure S3: | **The Markov process of a system in the before-transition state.** The sketch shows the Markov process in a before-transition state.

Under the assumption that each variable follows Gaussian distribution, we obtain the estimation of the distribution for each variable k at time point $t-2$, i.e., based on samples $z_k^1(t-2), z_k^2(t-2), ..., z_k^m(t-2)$, we estimate the mean $\mu_k(t-2)$ and standard deviation $\sigma_k(t-2)$ for each variable k, i.e., the distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$ (see Fig.S3).

Step 2. Determine the consistence vector for each variable at $(t-1)$**.**

Let an index $x_k^s(t-1) \in \{0,1\}$ describe whether a sample $z_k^i(t-1)$ of the kth variable is consistent comparing with its former distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$, that is, whether the appearance of sample $z_k^i(t-1)$ at time t is with large probability in the distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$. For each sample $z_k^s(t-1)$ of variable k at time point $t-1$, we have

$$x_k^s(t-1) = \begin{cases} 0, & \text{if } z_k^s(t-1) \in [\mu_k(t-2) - \sigma_k(t-2), \mu_k(t-2) + \sigma_k(t-2)] \\ 1, & \text{if } z_k^s(t-1) \in (-\infty, \mu_k(t-2) - \sigma_k(t-2)) \cup (\mu_k(t-2) + \sigma_k(t-2), +\infty) \end{cases}. \tag{S6}$$

Obviously, $x_k^s(t-1) = 0$ represents that the sample $x_k^s(t-1)$ is consistent with the former distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$, while $x_k(t-1) = 1$ represents that sample $x_k^s(t)$ is

inconsistent with the former distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$ (see Fig.S3). Thus, for each sample $Z^s(t-1) = (z_1^s(t-1), z_2^s(t-1), ..., z_n^s(t-1))$, the vector $X^s(t-1) = (x_1^s(t-1), ..., x_n^s(t-1))$ is the consistence vector at time $t-1$.

Let $\#0(t-1)$ and $\#1(t-1)$ respectively denote the number of value 0 and that of value 1 in an inconsistence vector $X^s(t)$ at $t-1$. Obviously, $\#0(t-1) + \#1(t-1) = n$, where n is the number of variables in the system, among which there are $\#0(t-1)$ variables consistent with the former distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$, while $\#1(t-1)$ variables inconsistent with the former distribution $N(\mu_k(t-2), \sigma_k^2(t-2))$.

According to above settings, we actually transform the observed sample set $o_t = (Z^1(t), Z^2(t), ..., Z^m(t))$ into the corresponding consistence vector $o_t = (X^1(t), X^2(t), ..., X^m(t))$.

Step 3. Training the HMM at $(t-1)$ based on Baum-Welch algorithm.

In this step, we need to identify the state transition matrix A and the emission matrix B at $(t-1)$, that is, training the HMM $\theta_{t-1} = (A(t-1), B(t-1), \pi)$ based on the basis of an observed sequence $\{o_1, o_2, ..., o_{t-1}\}$.

There are two possible states W_0 and W_1 in time point $t-1$. Then, we calculate the possibilities for each possible state transition and thus obtain the state transition matrix $A(t-1) = (a_{ij}(t-1))_{2\times 2}$, where

$$a_{ij}(t-1) = P(s_{t-1} = W_i \,|\, s_{t-2} = W_j), \tag{S7}$$

with $i, j \in \{0, 1\}$.

Besides, for the emission matrix $B(t-1) = (b_{jk}(t-1))_{2\times(n+1)}$ where $b_{jk}(t-1)$ is the probability of the kth possible observation under the assumption that the system state is W_j at time $t-1$, i.e.,

$$b_{jk}(t-1) = P(\#1(t-1) = k \,|\, s_{t-1} = W_j), \tag{S8}$$

where $j \in \{0, 1\}$ and $k \in \{0, 1, 2, ..., n\}$. Obviously, there are $n + 1$ possible observable cases for any sample at $t - 1$, i.e., case $\#1(t - 1) = 0$, case $\#1(t - 1) = 1$, ... , case $\#1(t - 1) = n$. In the case of an n-molecules biological system, case $\#1(t - 1) = k$ reflects that there are k molecules differentially expressed in one observation (i.e., one sample) at $t - 1$ comparing with their former expressions.

The initial state distribution $\pi = \{\pi_1, \pi_2\}$ is defined at time $t - 2$, where

$$\pi_i = P(s_{t-2} = W_i), \tag{S9}$$

with $i \in \{0, 1\}$.

According to Baum-Welch algorithm, we build A, B, and π based on the training set $\{o_1, o_2, ..., o_{t-1}\}$, i.e., sample sets up to time $t - 1$. The training process at time $t - 1$ includes the following three steps, in which we omit time $t - 1$.

- **Initialization**

  For $h = 0$, set initial values for a_{ij}^0, b_{jk}^0, and π_i^0, we have the HMM $\theta^0 = (A^0, B^0, \pi^0)$.

- **Update**

  For $h = 1, 2, ...$, we have the update for a_{ij}^h, b_{jk}^h, and π_i^h by recursion

$$a_{ij}^h = \frac{\sum_{T=1}^{t-1} \xi_T(i, j)}{\sum_{T=1}^{t-1} \gamma_T(i)}, \tag{S10}$$

$$b_{jk}^h = \frac{\sum_{T=1, \#1(t-1)=k}^{t-1} \gamma_T(k)}{\sum_{T=1}^{t-1} \gamma_T(k)}, \tag{S11}$$

$$\pi_i^h = \gamma_1(i), \tag{S12}$$

where $\gamma_T(i)$ and $\xi_T(i,j)$ are respectively of form Eq.(S4) and Eq.(S5). For $\gamma_T(i)$ and $\xi_T(i,j)$, the HMM used in the prior knowledge is that updated from the preceding step. The observation sequence used in the prior knowledge is $O = \{o_1, o_2, ..., o_{t-1}\}$.

- **Ending**

  When $h = H$, i.e., the Hth-updating step, the recursion is terminated. Then

$$\theta_i^H = (A^H, B^H, \pi^H). \tag{S13}$$

  The HMM used in the testing process follows $\theta_{t-1} = \theta_i^H$.

A.3.2 Calculating the inconsistence index at a testing time point t

Under the assumption that the transition point is at t, or in other word, time point t is hypothesized as the end point of a stationary Markov process of the before-transition stage (see Fig.S2). The onset of a pre-transition stage is the end of the stationary Markov process in a before-transition stage.

At each time point t, we are in a position to test if time t is not any more in the before-transition stage, that is, the time point t is the switching point for the stationary Markov process described as HMM θ_{t-1}.

Therefore, at the testing time point t, we calculate the I-index, or HMM-based probability P_t measuring the inconsistence between sample $\{o_t\}$ derived at time point t and the HMM θ_{t-1}, is given as follows:

$$I(t) = P_t(s_t = W_1 \,|\, s_1 = W_0, s_2 = W_0, ..., s_{t-1} = W_0,\, \theta_{t-1},\, O), \tag{S14}$$

where the observation sequence $O = \{o_1, o_2, ..., o_{t-1}, o_t\}$ in each testing step. Obviously,

$$\begin{aligned} &P_t(s_t = W_1 \,|\, s_1 = W_0, s_2 = W_0, ..., s_{t-1} = W_0,\, \theta_{t-1},\, O) \\ &= 1 - Q_t(s_t = W_0 \,|\, s_1 = W_0, s_2 = W_0, ..., s_{t-1} = W_0,\, \theta_{t-1},\, O). \end{aligned}$$

where Q_t actually represents the probability of consistence between the system state of sample $\{o_t\}$ derived at time point t and the HMM θ_{t-1}. At each time point t, we first calculate the consistence probability based on HMM θ_{t-1} and observed sequence $O = \{o_{t-1}, o_t\}$

$$Q_t(s_t = W_0 \,|\, s_{t-1} = W_0,\, \theta_{t-1},\, O) = \frac{P(s_{t-1} = W_0, s_t = W_0 \,|\, \theta_{t-1},\, O)}{P(s_{t-1} = W_0 \,|\, \theta_{t-1},\, O)}. \tag{S15}$$

The numerator

$$P(s_{t-1} = W_0, s_t = W_0 \,|\, \theta_{t-1},\, O) = \frac{\beta_{t-1}(s_{t-1} = W_0) a_{00} b_{0k}}{\sum_{i=0}^{1} \beta_{t-1}(s_{t-1} = W_i) a_{ij} b_{jk}}, \tag{S16}$$

and the denominator

$$P(s_{t-1} = W_0 \,|\, \theta_{t-1},\, O) = \frac{\beta_{t-1}(s_{t-1} = W_0)}{\sum_{j=0}^{1} \beta_{t-1}(s_{t-1} = W_j)}, \tag{S17}$$

where a_{00} and a_{ij} is from the state transition matrix $A = (a_{ij})_{2\times 2}$ in Eq.(S7), b_{0k} and b_{jk} is from the emission matrix $B = (b_{jk})_{2\times(n+1)}$ in Eq.(S8) while $k = \#1(t)$ represents that for the sample set o_t there are k variables with consistence index 1 in average, β is the forward probability provided as below Eq.(S18). It should be noticed that in Eqs.(S16) and (S17) the backward probability is set to be 1, since samples $o_{t+1}, \cdots$ are not available when t is the testing time point.

According to above settings, given the HMM θ_{t-1}, the calculation of HMM probability Q_t (the consistence probability) at a time point t only relies on the samples from $t-1$ and t. Obtaining the HMM probability Q_t for every candidate time point, the time point $arg_T[min(Q_t)]_{T=1,2,...,t}$, or equivalently $arg_T[max(P_t)]_{T=1,2,...,t}$, is the transition point.

The HMM-based forward probability.

For given observation sequence $\{o_1, o_2, ..., o_t\}$, two possible states $\{W_0, W_1\}$, and HMM $\theta = (A, B, \pi)$ with $A = (a_{ij})_{2\times 2}$, $B = (b_{jk})_{2\times(n+1)}$ with b_{jk} related to observation o_t, and

$\pi = (\pi_i)_{i\in\{0,1\}}$, we calculate the HMM-based forward probability

$$\beta_t(i) = \beta_t(s_t = W_i) = P(o_1, o_2, ..., o_t, s_t = W_i \,|\, \theta), \quad i \in \{0, 1\}. \tag{S18}$$

as follows.

Initialization: For $t = 1$,

$$\beta_1(i) = \pi_i b_{ik_1}, \quad i = 0, 1, \tag{S19}$$

where k_1 is given in $\{1, 2, ..., n + 1\}$.

Recursion: For $t = 2, 3, ..., T - 1$

$$\beta_t(i) = \left[\sum_{j=0}^{1} \beta_{t-1}(j)a_{ji}\right] b_{ik_t}, \quad i = 0, 1. \tag{S20}$$

where $k_t = \#1(t - 1)$ with $k_t \in \{1, 2, ..., n + 1\}$.

Ending: When $t = T$,

$$\beta_T(0) = \beta_t(s_t = W_0) = \left[\sum_{j=0}^{1} \beta_{T-1}(j)a_{j0}\right] b_{0k_T},$$

$$\beta_T(1) = \beta_t(s_t = W_1) = \left[\sum_{j=0}^{1} \beta_{T-1}(j)a_{j1}\right] b_{1k_T},$$

and

$$P(O \,|\, \theta) = P(o_1, o_2, ..., o_t \,|\, \theta) = \sum_{i=0}^{1} \beta_T(i). \tag{S21}$$

B Numerical validation of HMM-based method

In this section, we use a ten-gene network (see Fig.S22 or Fig.3a in the main text) to conduct a numerical simulation and theoretically demonstrate the detection of early-warning signals through HMM-based method. These types of gene regulatory networks are often used to study transcription, translation, diffusion, and translocation processes that affect gene regulatory activities (*1, 4, 19–21*). The following six differential equations represent the gene regulation of

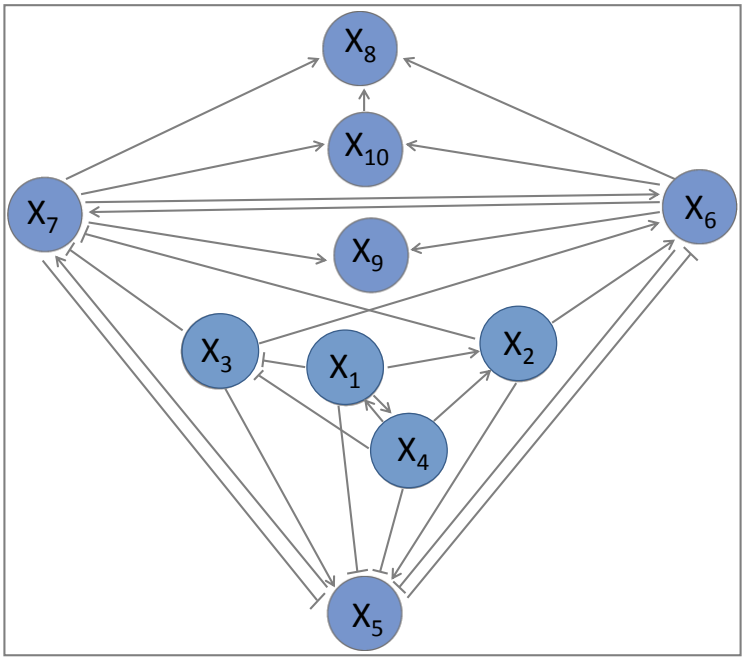

Figure S4: | **A model of a 10-molecular network.** In this sketch of a molecular network, there are 10 nodes whose dynamical regulatory relationships are given as stochastic system S22. The edges represent positive or negative regulations among nodes.

six genes in a network where gene regulation is represented in a Michaelis-Menten form, with the exception of the degradation rates, which are linearly proportional to the concentrations of the corresponding genes.

$$
\begin{cases}
\frac{\mathrm{d}z_1(t)}{\mathrm{d}t} = \frac{(8-3\,|P|)z_4(t)}{20\,(1+z_4(t))} - \frac{8+3\,|P|}{20}\,z_1(t) + \zeta_1(t), \\
\frac{\mathrm{d}z_2(t)}{\mathrm{d}t} = \frac{(5-3\,|P|)z_1(t)}{20(1+z_1(t))} + \frac{(5-3\,|P|)z_4(t)}{20(1+z_4(t))} - \frac{1}{2}\,z_2(t) + \zeta_2(t), \\
\frac{\mathrm{d}z_3(t)}{\mathrm{d}t} = \frac{3\,|p|-6}{10} + \frac{6-3\,|P|}{20(1+z_1(t))} + \frac{6-3\,|P|}{20(1+z_4(t))} - \frac{3}{5}\,z_3(t) + \zeta_3(t), \\
\frac{\mathrm{d}z_4(t)}{\mathrm{d}t} = \frac{(8-3\,|P|)z_1(t)}{20(1+z_1(t))} - \frac{8+3\,|P|}{20}\,z_4(t) + \zeta_4(t), \\
\frac{\mathrm{d}z_5(t)}{\mathrm{d}t} = -\frac{1}{2} + \frac{1}{20(1+z_1(t))} + \frac{7z_2(t)}{10(1+z_2(t))} + \frac{3z_3(t)}{5(1+z_3(t))} + \frac{1}{20(1+z_4(t))} + \frac{1}{5(1+z_6(t))} \\
\qquad\qquad + \frac{1}{5(1+z_7(t))} - \frac{6}{5}\,z_5(t) + \zeta_5(t), \\
\frac{\mathrm{d}z_6(t)}{\mathrm{d}t} = -\frac{1}{5} + \frac{z_2(t)}{5(1+z_2(t))} + \frac{z_3(t)}{5(1+z_3(t))} + \frac{1}{5(1+z_5(t))} + \frac{z_7(t)}{5(1+z_7(t))} - \frac{6}{5}\,z_6(t) + \zeta_6(t), \\
\frac{\mathrm{d}z_7(t)}{\mathrm{d}t} = -\frac{2}{5} + \frac{1}{5(1+z_2(t))} + \frac{1}{5(1+z_3(t))} + \frac{z_5(t)}{5(1+z_5(t))} + \frac{z_6(t)}{5(1+z_6(t))} - \frac{6}{5}\,z_7(t) + \zeta_7(t), \\
\frac{\mathrm{d}z_8(t)}{\mathrm{d}t} = \frac{z_6(t)}{5(1+z_6(t))} + \frac{z_7(t)}{5(1+z_7(t))} + \frac{2z_{10}(t)}{5(1+z_{10}(t))} - \frac{8}{5}\,z_8(t) + \zeta_8(t), \\
\frac{\mathrm{d}z_9(t)}{\mathrm{d}t} = \frac{4z_6(t)}{5(1+z_6(t))} + \frac{4z_7(t)}{5(1+z_7(t))} - \frac{9}{5}\,z_9(t) + \zeta_9(t), \\
\frac{\mathrm{d}z_{10}(t)}{\mathrm{d}t} = \frac{z_6(t)}{1+z_6(t)} + \frac{z_7(t)}{1+z_7(t)} - 2z_{10}(t) + \zeta_{10}(t),
\end{cases}
\tag{S22}
$$

where P is a scalar control parameter and $\zeta_i(t)$ $(i = 1, 2, ..., 10)$ are Gaussian noises with zero means and covariances $\kappa_{ij} = \mathrm{Cov}(\zeta_i, \zeta_j)$. z_i $(i = 1, ..., 10)$ represent the concentrations of mRNA-i. In Eq.(S22), the degradation rates of mRNAs are $(\frac{8+3|P|}{20}, \frac{1}{2}, \frac{3}{5}, \frac{8+3|P|}{20}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5}, \frac{8}{5}, \frac{9}{5}, 2)$. There is a stable equilibrium point $\bar{Z} = (\bar{z_1}, \bar{z_2}, ..., \bar{z_{10}}) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. The differential equations Eq.(S22) can be transformed into the difference equations $Z(k+1) = f(Z(k), P)$ using the Euler scheme (*22*), ***i.e.,***

$$
\begin{cases}
z_1(k+1) = z_1(k) + \left[\frac{(8-3\,|P|)z_4(k)}{20\,(1+z_4(k))} - \frac{8+3\,|P|}{20}\,z_1(k) + \zeta_1(k)\right]\Delta t, \\
z_2(k+1) = z_2(k) + \left[\frac{(5-3\,|P|)z_1(k)}{20(1+z_1(k))} + \frac{(5-3\,|P|)z_4(k)}{20(1+z_4(k))} - \frac{1}{2}\,z_2(k) + \zeta_2(k)\right]\Delta t, \\
z_3(k+1) = z_3(k) + \left[(\frac{3\,|p|-6}{10} + \frac{6-3\,|P|}{20(1+z_1(k))} + \frac{6-3\,|P|}{20(1+z_4(k))} - \frac{3}{5}z_3(k) + \zeta_3(k)\right]\Delta t, \\
z_4(k+1) = z_4(k) + \left[\frac{(8-3\,|P|)z_1(k)}{20(1+z_1(k))} - \frac{8+3\,|P|}{20}z_4(k) + \zeta_4(k)\right]\Delta t, \\
z_5(k+1) = z_5(k) + \Big[-\frac{1}{2} + \frac{1}{20(1+z_1(k))} + \frac{7z_2(k)}{10(1+z_2(k))} + \frac{3z_3(k)}{5(1+z_3(k))} + \frac{1}{20(1+z_4(k))} + \frac{1}{5(1+z_6(k))} \\
\qquad\qquad + \frac{1}{5(1+z_7(k))} - \frac{6}{5}z_5(k) + \zeta_5(k)\Big]\Delta t, \\
z_6(k+1) = z_6(k) + \left[-\frac{1}{5} + \frac{z_2(t)}{5(1+z_2(k))} + \frac{z_3(k)}{5(1+z_3(k))} + \frac{1}{5(1+z_5(k))} + \frac{z_7(k)}{5(1+z_7(k))} - \frac{6}{5}\,z_6(k) + \zeta_6(k)\right]\Delta t, \\
z_7(k+1) = z_7(k) + \left[-\frac{2}{5} + \frac{1}{5(1+z_2(k))} + \frac{1}{5(1+z_3(k))} + \frac{z_5(k)}{5(1+z_5(k))} + \frac{z_6(k)}{5(1+z_6(k))} - \frac{6}{5}\,z_7(k) + \zeta_7(k)\right]\Delta t, \\
z_8(k+1) = z_8(k) + \left[\frac{z_6(k)}{5(1+z_6(k))} + \frac{z_7(k)}{5(1+z_7(k))} + \frac{2z_{10}(k)}{5(1+z_{10}(k))} - \frac{8}{5}\,z_8(k) + \zeta_8(k)\right]\Delta t, \\
z_9(k+1) = z_9(k) + \left[\frac{4z_6(k)}{5(1+z_6(k))} + \frac{4z_7(k)}{5(1+z_7(k))} - \frac{9}{5}\,z_9(k) + \zeta_9(k)\right]\Delta t, \\
z_{10}(k+1) = z_{10}(k) + \left[\frac{z_6(k)}{1+z_6(k)} + \frac{z_7(k)}{1+z_7(k)} - 2z_{10}(k) + \zeta_{10}(k)\right]\Delta t,
\end{cases}
\tag{S23}
$$

with a small time interval Δt. Note that $Z(k)$ is the vector of $Z(t)$ at the time instant $k\Delta t$.

We denote the Jacobian matrix of Eq.(S23) as $J = \left.\frac{\partial f(Z(k);p)}{\partial Z}\right|_{Z=\bar{Z}}$, where

$$
J = \mathrm{e}^{\Delta t \cdot A} \tag{S24}
$$

with

$$A=\begin{bmatrix} -\frac{8+3|P|}{20} & 0 & 1-\frac{2|P|}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{5-3|P|}{20} & -\frac{1}{2} & 0 & \frac{8-3|P|}{20} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{3|P|-6}{20} & 0 & -\frac{3}{5} & \frac{3|P|-6}{5} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{8-3|P|}{20} & 0 & 0 & -\frac{8+3|P|}{20} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{20} & 0 & 1-\frac{2|P|}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{5} & 0 & 1-\frac{2|P|}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{5} & -\frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & -\frac{6}{5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} & -\frac{8}{5} & 0 & \frac{2}{5} \\ 0 & 0 & 0 & 0 & 0 & \frac{4}{5} & \frac{4}{5} & 0 & -\frac{9}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -2 \end{bmatrix}.$$

From Eq.(S24), we obtain ten distinct eigenvalues $(0.74^{|P|}, 0.61, 0.55, 0.45, 0.37, 0.30, 0.25, 0.20, 0.17, 0.14)$ by taking $\Delta t = 1$. Thus, the equilibrium point $\bar{Z}$ is stable when $P \in (0, 1]$. Obviously, there is a critical value $P_c = 0$, where the system loses stability and undergoes a critical transition. We aimed to detect early warning signals that indicate the critical transition as a control parameter P approaches a critical value 0 from $P > 0$.

We then diagonalize the Jacobian matrix J using matrix S

$$S=\begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \tag{S25}$$

which satisfies $S^{-1}\, J\, S = \Lambda$ where $\Lambda = diag(0.74^{|P|}, 0.61, 0.55, 0.45, 0.37, 0.30, 0.25, 0.20, 0.17, 0.14)$ is a diagonal matrix. Based on the theoretical model, we collected time-course data of the ten-gene expressions. Then, applying the forward algorithm of HMM model to the time-course data, we identified the critical transition point shown in Fig.3 in the main text. We simulated the inconsistence index curves for the network, as shown in Fig.3 in the main text.

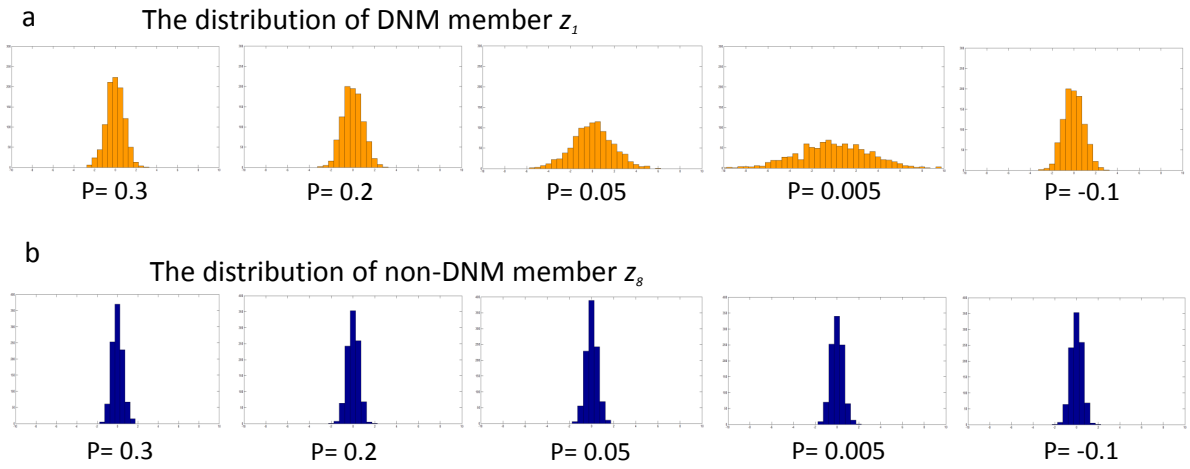

Figure S5: | **The distribution of DNM member z_1 and non-DNM member z_8.** **(a)** The distribution of DNM member z_1, which is directly related to the dominant eigenvector y_1. It can be seen that when the parameter P approaches the critical point $P = 0$, the distribution of z_1 changes significantly ($P = 0.005$). **(b)** The distribution of non-DNM member z_8, which is not related to the dominant eigenvector y_1. It can be seen that there are no significant changes in the distributions of z_8.

Here, from the relationship

$$Y(k) = S^{-1}(Z(k) - \bar{Z})$$

or

$$
\begin{aligned}
z_1 - \bar{z_1} = & \quad -y_1 \quad + y_4, \\
z_2 - \bar{z_2} = & \quad -y_1 \quad - y_2, \\
z_3 - \bar{z_3} = & \quad y_1 \quad + y_3, \\
z_4 - \bar{z_4} = & \quad -y_1 \quad - y_4, \\
z_5 - \bar{z_5} = & \quad - y_2 + y_3 - y_5 + y_6, \\
z_6 - \bar{z_6} = & \quad y_5 - y_6 + y_7, \\
z_7 - \bar{z_7} = & \quad y_6 - y_7, \\
z_8 - \bar{z_8} = & \quad y_5 - y_8 + y_{10}, \\
z_9 - \bar{z_9} = & \quad y_5 - y_9, \\
z_{10} - \bar{z_{10}} = & \quad y_5 - y_{10},
\end{aligned}
$$

it is clear that among $(z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10})$, the four variables z_1, z_2, z_3 and z_4 are related directly to y_1, which corresponds to the dominant eigenvalue. Therefore, according to the theoretical results in Section B, $\{z_1, z_2, z_3, z_4\}$ constitute the dynamical dominant group or the DNM of the system when $P \in (0, 1]$ and this will reflect the breakdown of the system as $P \to 0$ (see Fig.S5). Thus, as shown in Fig.2c in the main text, when $P \to 0$, there are 4 variables having large-scale changes and the ratio of 4-changing-nodes increases a lot.

C Application to three real datasets

For the three datasets, that is, gene expression profiling dataset of acute lung injury induced by phosgene gas and the ecological data of eutrophic lake state, we identified the pre-transition states by using DNM.

The gene expression profiling dataset of acute lung injury was downloaded from the NCBI

GEO database (access ID: GSE2565, GSE13009, GSE6764) (www.ncbi.nlm.nih.gov/geo). In these datasets, probe sets without corresponding gene symbols were ignored during our analysis. The expression values of probe sets that are mapped to the same gene were averaged. The three datasets are described in Table S1. Besides, we described the data processing in details and conducted the functional analysis results (g:profiler: http://biit.cs.ut.ee/gprofiler/ and NOA: http://app.aporc.org/NOA/) (*24, 25*) for some important molecules.

Table S1: Descriptions of the three datasets

Experimental data	Description
Genomic data of lung injury due to carbonyl chloride inhalation exposure (GSE2565) (*28*)	
Sampling points	9 sampling points 0, 0.5, 1, 4, 8, 12, 24, 48, 72 (hours)
Number of observed objects	22690 genes
Groups	control group and case group
Case data	6 subjects
Control data	6 control samples
Genomic data of MCF-7 human breast cancer caused by heregulin (HRG) (GSE13009) (*32*)	
Sampling points	16 sampling points 1/4, 1/3, 1/2, 3/4, 1, 3/2, 2, 3, 4, 6, 8, 12, 24, 36, 48, 72 (hours)
Number of observations	22000 genes
Groups	control group and case group
Case data	2 subjects
Control data	3 control samples
Genomic data on hepatic lesions due to chronic hepatitis C (GSE6764) (*33*)	
Sampling points	7 sampling periods cirrhosis, low-grade dysplastic liver tissue, high-grade dysplastic liver tissue, very early HCC, early HCC, advanced HCC, very advanced HCC (period)
Number of observations	PPI network with 9513 cDNAs
Groups	control group and case group

C.1 Dataset 1. Genomic data of the lung injury with carbonyl chloride inhalation exposure (i.e., acute lung injury)

This dataset was obtained in an experiment on toxic gas-induced lung injury effects, i.e., pulmonary edema (*28*), and was downloaded from the NCBI GEO database (ID: GSE2565) (www.ncbi.nlm.nih.gov/geo). In this dataset, probe sets without corresponding gene symbols were not considered during our analysis. The expression values of probe sets mapped to the same gene were averaged. Genes for the this disease have been linked and correlated by the combined functional couplings among them from various databases of protein-protein interactions of STRING, FunCoup and BioGrid. In the disease dataset, the expression profiling information was mapped to the integrated networks individually for identifying corresponding DNM. We downloaded the biomolecular interaction networks from various databases, including BioGrid (www.thebiogrid.org), TRED (www.rulai.cshl.edu/cgi-bin/TRED/), KEGG (www.genome.jp/kegg), and HPRD (www.hprd.org). First, the available functional linkage information for *Mus musculus* was downloaded from these databases and combined. For instance, after removing the redundancy, we obtained 37950 linkages in 6683 mouse proteins/genes for acute lung injury. Next, the genes evaluated in these microarray datasets were mapped individually to these integrated functional linkage networks. The network information are employed in the post processing step for visualizing results (Figs.3a, 4a,4b in the main text) and functional analysis.

To study acute lung injury, a genomic approach was used to investigate the molecular mechanism of phosgene-induced lung injury. The experiments were conducted to determine the temporal effects of phosgene exposure on lung tissue antioxidant enzyme concentrations and the gene expression level, and these results were compared with those from air-exposed mice treated in a similar manner to assess the role of the GSH redox cycle in this oxidative lung injury model. To produce two groups of data, i.e., the control group data and case group data,

two groups of CD-1 male mice were exposed to air or phosgene, respectively. Lung tissues were collected from air- or phosgene-exposed mice at 0.5, 1, 4, 8, 12, 24, 48, and 72 hr after exposure. The details of the experiment are available in the original paper (*28*). We introduce some key background as well as our functional analysis as follows.

Phosgene gas is one of the most important and common chemical industry gases (*27*). Some pathogenic mechanisms of the acute lung injury induced by phosgene have been identified (*28*). According to the results of the former study (*14*), a major change in the entropy of the core dynamical network marker (DNM) occurs from 4 hr to 8 hr. Also, the pathway enrichment analysis and GO functional analysis from our previous result (*12*) showed that genes we identified in the DNM were closely related to the mechanism of disease progression (*28, 31*). Dysfunctions of glutathione metabolism and the chemokine signaling pathway related to inflammatory immune response were activated *in vivo*, which also reflected protection against the oxidant-like activity of phosgene. Pathways affected by the oxidant reaction became disordered, especially signal transduction via protein-modified activation, such as the MAPK signaling pathway and Wnt signaling pathway. The decrease in pH induced by the HCl-release reaction affected some pathways that were sensitive to intracellular conditions and related to communication or transport channels, e.g., gap junctions. Some signaling pathways may also be relevant to repair, survival, apoptosis, and reproduction, such as the GnRH signaling pathway, MAPK signaling pathway, and TGF-beta signaling pathway (*28, 31*). At the GO function level, some biological processes were also highly related to acute lung injury. For example, the expression profiles of some genes were related to abnormal changes in primary metabolic processes. This indicated the denaturation of lipoids, proteins, and nucleic acids that may have been oxidized by phosgene (*28, 31*). Some well-known genes that regulate or are directly involved in apoptosis were also included in the DNM, such as *JUN*, *NOTCH2*, and *MYC*. Some genes in the DNM were also related to inflammatory response, wounding induced by oxidant damage, and irritation, such as *IL1B*,

PTGS2, *CCL2*, and *MYD88*.

Briefly, investigators found that the main physiological effects occurred within the first 8 hours after exposure, resulting in common observations of enhanced BALF protein levels, increased pulmonary edema, and ultimately decreased survival rates (*28*). At the concentration delivered, 50%-60% mortality was routinely observed at 12 hours while 60%-70% mortality was observed at 24 hours (*28*). The detailed results are also available in the original paper (*28*). Early warning signals of lung injury based on the DNM we identified are shown in Fig. 5a of the main text, which showed that the pre-transition state may start around 4 hr, while the system may enter the after-transition state after 12 hr. Our prediction based on the DNM score agreed with the actual disease development.

To explain the how the HMM-based method applied in real dataset more clearly, we used acute lung injury as a concrete example to describe our computational procedure step by step. In GSE2565 data set, there are 22,690 original probe sets. We mapped them to the corresponding NCBI Entrez gene symbols by using the GEO annotation. Meanwhile, we screened out all probe sets with incorrect corresponding gene symbols while probe sets that detected the same genes were combined using the averaging method. After this procedure, there were 12,871 genes left.

Step 1 Choose differential expression genes from the high-throughput gene data for acute lung injury. At each sampling point (or period), there are 12,871 genes. Each gene has 6 case samples and 6 control samples. At the 0 h sampling point, the case samples are identical to the control samples. Each sampling time point is supposed as a candidate tipping point of the critical transition, based on which we apply the HMM-based algorithm and calculate the inconsistence index representing the probability of a candidate point being the end point of the stationary Markov process of the before-transition state.

At each sampling point, by the student t-test with statistical significance ($p < 0.05$)

and fold change, the genes are filtered. The numbers of differentially-expressed gene are $\{503, 840, 1635, 1692, 1393, 1337, 1620, 1399\}$ respectively at 0.5, 1, 4, 8, 12, 24, 48 and 72 hours. Then, at each sampling point, genes are ranked based on their P-values.

Based on the rank of the differentially-expressed genes, we select the top 500 genes with the most significant P-values at each candidate time point. For different candidate points, the identified differentially-expressed gene sets are distinct. Under the assumption that a time point T is a transition point, the HMM is trained based on the top 500 differentially-expressed genes selected at this time point T, and then to calculate the inconsistence index correspondingly. It should be noted that we found that the number of selected genes is not sensitive to the inconsistence index (see Note 3 below). If there are less than 500 significantly differential genes at a time point, we simply choose top 500 differential genes (i.e., not necessarily significant in terms of P-value).

Step 2 We carried out the HMM-based method and the forward algorithm.

We conducted the data normalization for all the variables (the 1st and 2nd moments).

$$A = \frac{D_{\text{case}} - \text{mean}(N_{\text{control}})}{\text{SD}(N_{\text{control}})}, \tag{S26}$$

where A denotes the normalized expression data for each variable in each case sample, D_{case} is the data for each variable in every case sample, while the mean(N_{control}) and SD(N_{control}) are the mean and standard deviation for each variable in all the control samples, respectively.

Step 3 The sharp increase of the inconsistence index indicates the impending critical transition and the peak of the inconsistence index demonstrates the most possible tipping point of the critical transition.

The obtained inconsistence index curve is consistent with the real experimental phenomenon,

and the inconsistence index start increasing sharply from the 4th time period (4 h) and reach peaks in the 5th time period (8 h, i.e., the pre-transition, see Fig.3a in the main text). This indices show that the pre-transition state starts near the 4th time period (4 h), and the system transitions to another state after the 5th time period (8 h). Our early-warning signals are coincident with the actual disease development that the most prominent physiological effects occur within the first 8 h after exposure, resulting in pulmonary edema and ultimately reducing survival rates (see the original paper (*28*)). Based on the dynamical information of the network, we have graphically illustrated the dynamical changes in the overall mouse PPI network (see Fig.4b in the main text), from which it can be seen that a critical transition occurs around 8 h sampling point (during 4h-8h period).

For the specific algorithm for this real dataset, we have the following three notations.

Note 1 During the progression of the disease, each time point is supposed as a candidate transition point. To validate a candidate point, we tested whether some genes show significant changes and behave dynamically in a strongly collective manner, i.e., the distribution of these genes change drastically comparing with their previous distribution. Therefore, we respectively investigate the set of genes differentially expressed at each candidate point, rather than the whole set of differential genes identified from all the time points.

Note 2 To compare the inconsistence index, it is required that the number of the selected genes is equal at all candidate time points. Specifically, in the HMM-training step at a candidate time point, the emission matrix B is $2 \times (n+1)$ dimension, where n denotes the number of variables. To unify the scale of the inconsistence index, so that the inconsistence indices are comparable, we only consider the same number of genes at each candidate point.

Note 3 The selection of top 500 differentially-expressed genes is mainly from the computa-

tional consideration. Actually, under the circumstances that we have to choose equal number of differentially-expressed genes at each time points, the number of selected genes is not sensitive to the inconsistence index (Fig.S6).

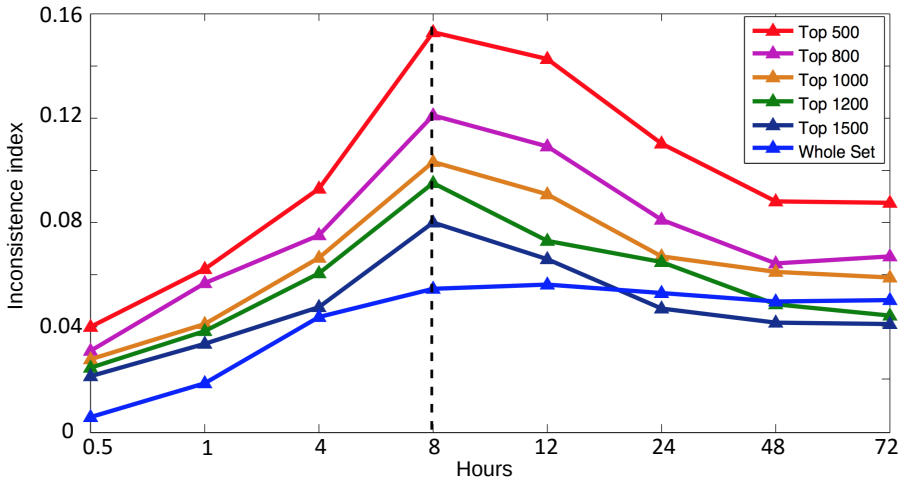

Figure S6: | **The inconsistence indices based on different number of top differentially-expressed genes.** To illustrate the significance of the results, the inconsistence indices are calculated based on control samples (the green curve) and 10 sets of randomly chosen genes from bootstrap (blue curves). It can be seen that the probability curves based on neither control samples nor randomized groups show significant signal to the critical transition in 8 h for 500-1500 genes. However, if we use all differential genes across whole periods, i.e., all 3982 differential genes during 0.5-72 hours (after removing the reduplicate genes), we cannot clearly identify the critical point.

To illustrate the significance of the results, a comparative figure with control samples and bootstrap is presented as Fig.S7, in which the green curve is based on the control samples of the top 500 differentially-expressed genes selected at 8 h, and each blue one is based on 500 randomly chosen genes. It can be seen that the original red curve shows significant signal for the upcoming critical transition between 4-8 h, while both the green curve (control) and the randomly-chosen genes fail to supply any signal.

We listed the top differential-expressed genes at the 5th time point in the Supplementary Table 'Top differential-expressed genes', among which Some well-known genes that regulate

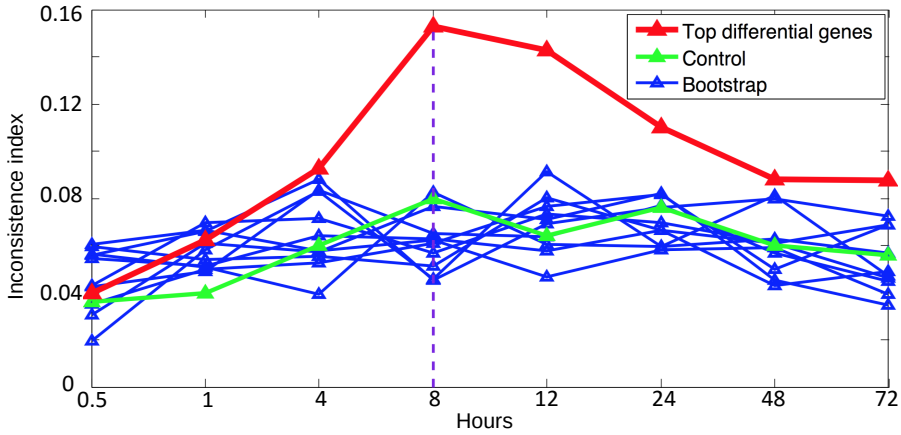

Figure S7: | **The inconsistence index based on control samples and randomly chosen genes from bootstrap.** To illustrate the significance of the results, the inconsistence indices are calculated based on control samples (the green curve) and 10 sets of randomly chosen genes from bootstrap (blue curves). It can be seen that neither the probability curve based on control samples nor on randomized groups show significant signal to the critical transition in 8 h.

or are directly involved in apoptosis were also included in the most significant group, such as JUN, NOTCH2, and MYC. Some genes in the most significant group were also related to the inflammatory response, wounding induced by oxidant damage, and irritation, such as IL1B, PTGS2, CCL2, and MYD88. and presented the dynamical changes in the network including the selected DNM during the progression in the Fig.3b of the main text.

C.2 Dataset 2: Genomic data of MCF-7 human breast cancer caused by heregulin (HRG)

Great attention is being paid to the study of breast cancer because it is one of the highest fatal-ratio diseases and increasing numbers of patients are suffering from breast cancer worldwide. The main cause of breast cancer is far from clear, but some of the mechanisms and phenomena that occur during the disease progression of HRG-induced breast cancer have been identified based on studies of its molecular mechanisms (*32, 34*) and pathogenesis (*35*).

Heregulin (HRG) induces dose-dependent transient and sustained intracellular signaling,

proliferation, and differentiation of MCF-7 breast cancer cells. This dataset was obtained in an experiment on MCF-7 cell line with HRG stimulation (GSE13009).

In the infected host, some metabolic pathways respond to these interruptions and become increasingly disordered. The following results show that some reported phenomena were consistent with our investigations, which also provided some novel insights. The identified DNB module is related to regulation of apoptotic process (GO:0042981), regulation of programmed cell death (GO:0043067) and regulation of cell death (GO:0010941) with significant P-value of gene enrichment by website tools of DAVID Bioinformatics Resource (*36*). By the pathway analysis in KEGG database, we found that 7 genes (*CEBPA*, *SMAD3*, *GSK3B*, *LAMC2*, *MMP1*, *PIK3R3*, *RXRA*) in this DNB module participate in the Pathways in cancer, and many genes of this module also take part in the other cancer related pathways, e.g. *Wnt* signaling pathway, *p53* signaling pathway, ECM-receptor interaction. Many genes in this DNB module have been proved to associate with cancer or tumor process, and some of these genes are associated with breast cancer. For example, *BCAS4* is an important gene for breast tumor development and progression (*37*), *ARID3B* is one of genes to regulate cell motility and actin cytoskeleton organization (*38*), and is found to associate with breast cancer onset (*39*). *TNFRSF21* encodes tumor necrosis factor receptor which can regulate the NF-kappaB and mediate apoptosis process (*40*). *LAMC2* encodes the gamma chain isoform laminin, which is involved in many biological process, and LAMC2 also is proved to be related to breast cancer process (*41, 42*). Therefore, DNB for HRG-induced breast cancer can mainly induce cancer by affecting the processes of regulation of apoptosis, regulation of programmed cell death and regulation of cell death.

C.3 Dataset 3: Genomic data on hepatic lesions due to chronic hepatitis C

Greater attention is being paid to the study of HCC because it is one of the highest fatal-ratio diseases and increasing numbers of patients are suffering from hepatitis C virus (HCV) infections worldwide. The main cause of HCC via HCV infection is far from clear, but some of the mechanisms and phenomena that occur during HCV infection and hepatitis C disease progression have been identified based on major studies of its molecular mechanisms and pathogenesis. When HCV invades the host cell, it recruits the necessary molecules to replicate itself, then translates and processes its proteins for packaging, before releasing its copies via host lipid transport systems. In the infected host, some metabolic pathways respond to these interruptions and become increasingly disordered. The following results show that some reported phenomena were consistent with our investigations, which also provided some novel insights.

According to our algorithm, the critical point of the disease occurred between the high grade dysplastic stage and the very early cancer stage.

Interestingly, we found that many of genes in the identified leading network were consistent with the response to HCV infection in vivo, particularly the activation of the immune system and dysfunctions associated with basic cell metabolism in the hosts. We carried out functional analysis on the top 500 differential-expression gene in the early HCC sampling point. some enriched pathways were related to dysfunctions of basic cell metabolism that might be interrupted by the reproduction and release of HCV. Some pathways shared common characteristics with cancer, especially the signaling pathways involved in cell growth, such as transcriptional misregulation in cancer, purine metabolism, Wnt signaling pathway, TGF-beta signaling pathway, and others. These dysfunctional pathways indicated the cell status when HCV invades host cells and exploits the host resources for replication.

In the HCC dataset GSE6764, there were 20320 original probesets, which we mapped to the

PPI network. We screened out all probesets with incorrect corresponding gene symbols while probeset that detected the same genes were combined using the averaging method, leaving 9513 genes. Using the algorithm in Section E, we conducted the computation of inconsistence index and get the application result (Fig.S8), which suggests that the critical transition is around the early HCC stage.

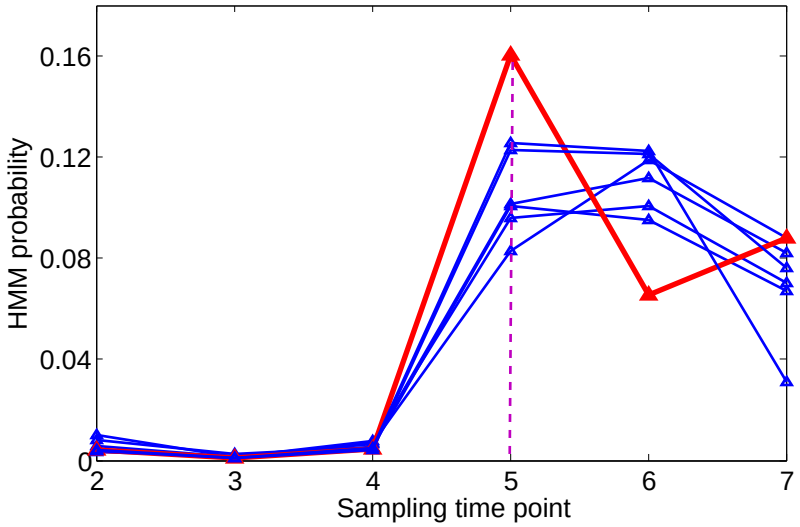

Figure S8: | **The application of HMM-based method on the dataset of HRG-induced breast cancer.** By applying HMM-based method to the microarray data of HRG-induced breast cancer, we show the inconsistence indices based on the top 500 differential-expression genes from each candidate transition time points. The red curve represents the inconsistence index calculated from the top 500 differential-expression genes which are selected at the 5th sampling point (early HCC stage), while the seven blue curves are that from other time points. It is can be seen that the most significant signal appear in the early HCC stage, which agrees with the experimental observation.

References

1. Chen, L., Wang, R., Li, C. & Aihara, K. *Modeling Biomolecular Networks in Cells: Structures and Dynamics*, (Springer, New York, 2010).

2. Voit, E.O. A systems-theoretical framework for health and disease: Inflammation and preconditioning from an abstract modeling point of view, *Math. Biosci.* **217**, 11–18(2009).

3. Hovinen, E., Kekki, M. & Kuikka, S. A theory to the stochastic dynamic model building for chronic progressive disease processes with an application to chronic gastritis, *J. Theor. Biol.* **57**, 131–152(1976).

4. Chen, L., Wang, R. & Zhang, X. *Biomolecular Networks: Methods and Applications in Systems Biology*, (John Wiley & Sons, Hoboken, New Jersey, 2009).

5. Guckenheimer, J. & Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, (Springer, 1983).

6. Arnol'd, V.I. *Dynamical systems V: bifurcation theory and catastrophe theory*, (Springer, 1994).

7. Murdock, J. *Normal forms and unfoldings for local dynamical systems*, (Springer, 2003).

8. Wiggins, S. *Global bifurcations and chaos: analytical methods*, (Springer, 1988).

9. Mlodinow, L. *The Drunkard's Walk*, (New York: Random House, 2008).

10. Cover, T. & Thomas, J. *Elements of information theory*, (Wiley, New Jersey, 2005).

11. Strogatz, S. H. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry And Engineering*, (Addison-Wesley, Reading, MA, 1994).

12. Chen,L. *et al.* (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers, *Scientific Reports*, **2**(342), 1-8.

13. Strogatz, S. H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. (Addison-Wesley, Reading, Massachussets, 1994).

14. Liu,R. *et al.* (2012) Identifying critical transitions and their leading networks for complex diseases, *Scientific Reports*, **2**(813), 1-9.

15. Liu,R. *et al.* (2014) Identifying critical transitions of complex diseases based on a single sample, *Bioinformatics*, **30**(11), 1579-1586.

16. Scheffer, M. *et al.* Early-warning signals for critical transitions, *Nature* **461**, 53–59(2009).

17. 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).

18. Wissel, C. A universal law of the characteristic return time near thresholds, *Oecologia* **65**, 101–107(1984).

19. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation, *Nature* **405**, 590–593(2000).

20. Chen, L. & Aihara, K. Stability of genetic regulatory networks with time delay, *IEEE Trans. Circuits Syst. I* **49**, 602–608(2002).

21. Li, C., Chen, L. & Aihara, K. Stability of genetic networks with SUM regulatory logic: Lur'e system and LMI approach, *IEEE Trans. Circuits Syst. I* **53**,

2451–2458(2006).

22. Kloeden, P. & Platen, E. *Numerical Solution of Stochastic Differential Equations*, (Springer, 1999).

23. Rencher, A. C. *Methods of Multivariate Analysis*, (Wiley, New York, 1995).

24. Reimand, J., Arak, T. & Vilo, J. g:Profiler – a web server for functional interpretation of gene lists (2011 update), *Nucleic Acids Res* **39**, W307–315(2011).

25. Wang, J. *et al.* (2011) NOA: a novel Network Ontology Analysis method, *Nucleic Acids Res*, **39**, e87.

26. Gene cards: http://www.genecards.org/.

27. Wolfgang, S. & Werner, D. *"Phosgene" in Ullmann's Encyclopedia of Industrial Chemistry Wiley-VCH*, (Weinheim, 2002).

28. Sciuto, A. M., *et al.* (2005) Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation, *Chem. Res. Toxicol.* **18**, 1654–1660.

29. Wang, R., Dearing, J.A., Langdon, P.G., Zhang, E., Yang, X., Dakos, V. & Scheffer. M. Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature* **492**, 419–422(2012).

30. R. Quax, D. Kandhai & P. Sloot. Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series, *Scientific Reports* **3**, 1–7(2013).

31. Wang, P. *et al.* (2011) Mechanism of acute lung injury due to phosgene exposition and its protection by cafeic acid phenethyl ester in the rat, *Exp. Toxicol Pathol.*, **24**.

32. Saeki, Y., *et al.* (2009) Ligand-specific sequential regulation of transcription factors for differentiation of MCF-7 cells, *BMC Genomics*, **20**, 545–552.

33. Wurmbach, E. *et al.* Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* **45**, 938–947(2007).

34. Suzuki, H., Okunishi, R., Hashizume, W., Katayama, S., Ninomiya, N., Osato, N., Sato, K., Nakamura, M., Iida, J., Kanamori, M. & Hayashizaki, H. Identification of region-specific transcription factor genes in the adult mouse brain by medium-scale real-time RT-PCR, *FEBS Lett* **573**, 214–218(2004).

35. Normanno, N., Ciardiello, F., Brandt, R. & Salomon, D. S. Epidermal growth factor-related peptides in the pathogenesis of human breast cancer, *Breast Cancer Research and Treatment* **29**, 11–27(1994).

36. Huang, D. W., Sherman, B. T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources, *Nature Protoc* **4**, 44–57(2009).

37. Barlund, M., Monni, O., Weaver, J. D., Kauraniemi, P., Sauter, G., Heiskanen, M., Kallioniemi, O. P. & Kallioniemi, A. Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo amplification, overexpression, and fusion in breast cancer, *Genes Chromosomes Cancer* **35**, 311–317(2002).

38. Casanova, J. C., Uribe, V., Badia-Careaga, C., Giovinazzo, G., Torres, M. & Sanz-Ezquerro, J. J. Apical ectodermal ridge morphogenesis in limb development is controlled by Arid3b-mediated regulation of cell movements, *Development* **138**, 1195–1205(2011).

39. Akhavantabasi, S., Sapmaz, A., Tuna, S. & Erson-Bensan, A. E. miR-125b targets ARID3B in breast cancer cells, *Cell Struct Funct* **37**(1), 27–38(2012).

40. Kasof, G.M., Lu, J.J., Liu, D., Speer, B., Mongan, K.N., Gomes, B.C. & Lorenzi, M.V. Tumor necrosis factor-alpha induces the expression of DR6, a

member of the TNF receptor family, through activation of NF-kappaB, *Oncogene*. **20**(55), 7965–7975(2001).

41. Sathyanarayana, U.G., Padar, A., Huang, C.X., Suzuki, M., Shigematsu, H., Bekele, B.N. & Gazdar, A.F. Aberrant promoter methylation and silencing of laminin-5-encoding genes in breast carcinoma, *Clin Cancer Res* **9**(17), 6389–6394(2003).

42. Koshikawa, N., Minegishi, T., Sharabi, A., Quaranta, V. & Seiki, M. Membrane-type matrix metalloproteinase-1 (MT1-MMP) is a processing enzyme for human laminin gamma 2 chain, *J Biol Chem* **280**(1), 88–93(2005).

D Supplementary Table 'Top differential-expressed genes'

See the Supplementary Table 'Top differential-expressed genes' attached after the References.

Supplementary Table `Top differential genes'

Identified genes of Lung injury

Spock2	6.00E-05	Hbegf	0.000201
Lrat	1.80E-05	Lrrc59	9.60E-05
Gbe1	0.000151	Tcf21	4.00E-06
Pex11a	0.000284	Mboat1	0
Gclm	0	Mt1	9.00E-06
Serpina3n	0.000149	Psmc6	0.000103
Slc7a10	2.70E-05	Gadd45g	1.60E-05
Txnl1	6.70E-05	Vamp5	0.000418
Tnfrsf12a	0.000185	Fzd2	0.000451
Nploc4	0.000235	Sod3	7.40E-05
Ereg	0.000456	H2afv	0.000397
Fgfr4	3.20E-05	Fosl1	0.00037
Lox	9.10E-05	Carkd	4.70E-05
Rbp1	3.20E-05	Klf6	2.50E-05
Cmpk1	0.000439	Hspb8	7.60E-05
Pvr	3.00E-06	Adprh	7.00E-05
Samd8	0.000261	Josd2	0.000246
Hcfc1r1	0.000167	Arap1	2.00E-06
Traf1	0.000409	Psmd6	6.00E-05
Ufd1l	0.000138	Zfp36	0.000347
Sdc4	3.40E-05	Ccnl1	0.00038
Ace2	0.000165	Vapa	2.90E-05
Suds3	0.000247	St6gal1	4.10E-05
Plaur	1.40E-05	Ssbp3	4.60E-05
Dbp	0.000194	Cd48	0.000276
Psmb5	0.000169	Slc20a1	0.000126
Aldh3a1	0.000307	Pcyox1	0.000238
Gna14	0.000243	Hes6	0.000154
Mt2	4.00E-06	Taldo1	2.30E-05
Loc100043998	5.10E-05	Fgfbp1	2.00E-06
Psmd3	0.000113	Spns2	5.00E-06
Aox1	0.000367	Bcl6	4.10E-05
Loc100046560	6.20E-05	Gfpt2	1.00E-04
Eg383901	4.50E-05	H3f3a	0.00032
Larp5	0.000365	Kif5b	0.000164
Angptl2	3.00E-06	Zc3h15	0.000139
Ppl	1.00E-06	Dedd2	1.90E-05
Asns	0.000156	Esd	0
Sars	1.60E-05	Adss	0.000203
Hspa1b	2.00E-06	F3	0.000157
Phldb1	0.000204	Ifrd1	1.50E-05
Loc100047868	0.000155	Ubxn4	4.40E-05
Stat3	5.00E-06	Eif3c	0.000399
Ncald	0.000304	Myc	0.000136
Adamts10	0.000224	Prss22	2.40E-05
Marchs7	0.000302	Ptpn14	1.90E-05
Cd82	8.50E-05	Ensmusg0(	3.30E-05
Camk1	0.000329	D430019h	2.70E-05
Klf5	0.00016	Ncl	5.30E-05
Nfix	2.60E-05	C920025e(	8.00E-05
Mest	0.000207	Tinagl1	2.00E-06
Notum	0.000286	Pdk2	1.00E-05
Areg	0.000168	Eg667723	0.000165
D10ertd641e	0.000199	Rnf5	0.000224

Isoc1	0.000105	Maff	0.000159
Txnrd1	0	Pdk4	0.000105
Srxn1	0	Fkbp5	0.000233
Tspyl4	7.00E-06	Myct1	6.40E-05
Cxcr4	6.70E-05	Errfi1	9.50E-05
Lbh	3.60E-05	Isg20	0.000219
Kcnb1	0.000197	Clca1	0.000136
Ptgs1	1.10E-05	Psmd14	0.000131
Tiparp	0.000443	Hspb1	7.00E-06
Gsta4	0.00013	Ottmusg0(	0.000242
Mrpl17	0.000428	Creg1	0
Mknk1	0.000268	Slc40a1	0.000121
Pcp4l1	8.20E-05	Gimap4	0.000478
Tmem206	0.000107	Col6a3	1.00E-06
Cdkn1a	3.50E-05	Slc2a1	7.00E-06
Lrp2	0.000258	Gnl3	0.000162
Ppap2b	8.00E-05	Meis1	0.000324
Mfap2	0.000323	Cstb	5.10E-05
Sox18	0.000139	Nfkbia	0.000456
Krt8	2.50E-05	Smo	0.000109
Dnajc5	0.000457	Pim1	9.70E-05
Thbs1	0.000388	Angptl4	4.60E-05
Mafk	0.000138	Kctd9	0.000171
Basp1	4.80E-05	Npnt	0.000253
Psme4	0.00015	Rhou	0.000421
Fabp4	1.70E-05	Scn3b	3.20E-05
Afp	2.90E-05	Tuba4a	0
Lama3	1.30E-05	Apln	4.60E-05
Loc677317	4.00E-05	Rgs3	0.000104
Tie1	0.000386	Nqo1	1.00E-05
Kazald1	0.000334	Psmd1	0.000164
Atf4	2.90E-05	Tnfrsf19	2.40E-05
Anxa6	4.60E-05	Zwint	3.40E-05
Gclc	0	Psmd12	0.000346
Loc100047619	3.00E-06	Pde4b	0.000162
Ceacam1	5.80E-05	Eif5	1.60E-05
Pgd	1.00E-06	Irs1	4.00E-06
Pfn2	2.10E-05	Ppp1r14a	0.000449
Tspan13	9.20E-05	Lgals3	4.80E-05
Gp49a	0.000283	Loc10004'	2.80E-05
Aars	4.20E-05	Purg	0.000245
Hnrpdl	0	Hmga1	0.000135
Ype13	5.70E-05	Oaz2	0.000314
Smad1	0.00021	Gas5	0.000124
Hsp90aa1	8.00E-06	Tnfsf12	1.60E-05
Myd116	5.90E-05	Epha2	0.000317