*Bioinformatics*, 36(5), 2020, 1522–1532
doi: 10.1093/bioinformatics/btz758
Advance Access Publication Date: 9 October 2019
Original Paper



Systems biology

# Single-sample landscape entropy reveals the imminent phase transition during disease progression

**Rui Liu[1], Pei Chen[1,*] and Luonan Chen[2,3,4,*]**

[1]School of Mathematics, South China University of Technology, Guangzhou 510640, China, [2]Key Laboratory of Systems Biology, Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, [3]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China and [4]Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai 201210, China

*To whom correspondence should be addressed.

Associate Editor: Lenore Cowen

Received on January 10, 2019; revised on July 5, 2019; editorial decision on September 30, 2019; accepted on October 5, 2019

## Abstract

**Motivation:** The time evolution or dynamic change of many biological systems during disease progression is not always smooth but occasionally abrupt, that is, there is a tipping point during such a process at which the system state shifts from the normal state to a disease state. It is challenging to predict such disease state with the measured omics data, in particular when only a single sample is available.

**Results:** In this study, we developed a novel approach, i.e. single-sample landscape entropy (SLE) method, to identify the tipping point during disease progression with only one sample data. Specifically, by evaluating the disorder of a network projected from a single-sample data, SLE effectively characterizes the criticality of this single sample network in terms of network entropy, thereby capturing not only the signals of the impending transition but also its leading network, i.e. dynamic network biomarkers. Using this method, we can characterize sample-specific state during disease progression and thus achieve the disease prediction of each individual by only one sample. Our method was validated by successfully identifying the tipping points just before the serious disease symptoms from four real datasets of individuals or subjects, including influenza virus infection, lung cancer metastasis, prostate cancer and acute lung injury.

**Availability and implementation:** https://github.com/rabbitpei/SLE.

**Contact:** chenpei@scut.edu.cn or lnchen@sibs.ac.cn

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

## 1 Introduction

The time evolution of many biological systems is not always smooth but occasionally abrupt, that is, the persistent effects or perturbations of various internal or external factors sometimes result in drastic or qualitative changes of system states at a tipping point. For example, in some chronic diseases such as cancer (He *et al.*, 2012; Scheffer *et al.*, 2009; Yang *et al.*, 2018), irreversible deterioration may occur suddenly within a short period of progression, while the disease (e.g. chronic inflammation) may progress gradually and steadily for years or even decades of a long incubation period before such a catastrophic transition. The disease progression can be viewed as the evolution of a nonlinear dynamical system with a tipping point occurring after progressive changes in a certain organ or the whole organism. Regardless of specific differences in either biological processes or observed symptoms among diseases, the progression of illness can be generally divided into three stages or states (Supplementary Fig. S1 and Fig. 1B), i.e. a normal state, a pre-disease state and a disease state, where the pre-disease state is a critical state or tipping point just before disease appearance (Achiron *et al.*, 2010; Chen *et al.*, 2012; Liu *et al.*, 2014a). In other words, during the course of illness there is a drastic transition from a relatively healthy/normal stage via a critical/pre-disease stage to an irreversible disease stage (Litt *et al.*, 2001; Liu *et al.*, 2002; McSharry *et al.*, 2003; Paek *et al.*, 2005; Roberto *et al.*, 2003; Venegas *et al.*, 2005). Therefore, to prevent from or at least get ready for the catastrophic deterioration of a disease, it is crucial to detect its early-warning signals by quantifying its critical state or tipping point (Chen *et al.*, 2012).

Although hunting for disease tipping point or pre-disease state is of great importance, it is still a challenging task to identify the pre-disease state or predict the disease state by finding robust biomarkers specific to respective diseases. Traditional biomarkers, including molecular- or module-biomarkers, are employed to distinguish disease samples from normal samples or identify the disease state by exploiting the information of the differential expressions of genes/proteins between the normal and disease states, rather than signaling the pre-

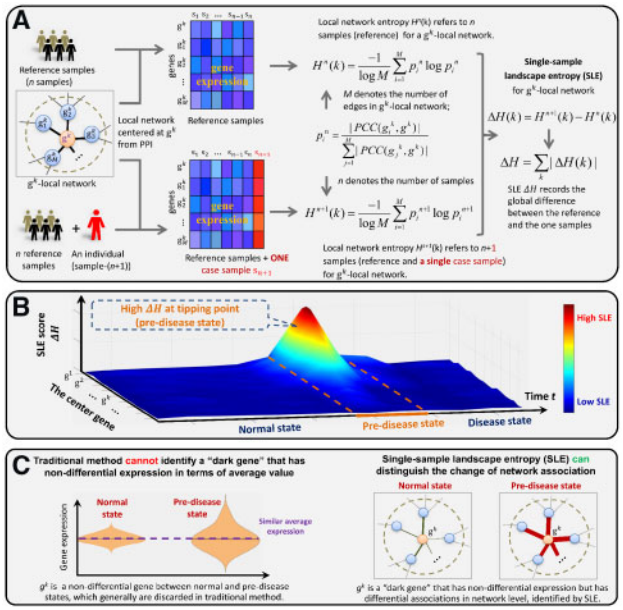


**Fig. 1.** The schematic illustration of single-sample landscape entropy (SLE). (**A**) Given a number of reference samples which can be derived from normal cohort, the SLE is calculated based on a single-sample from any individual. Specifically, both the reference samples and the to-be-determined single-sample are mapped to the existing PPI network or other reference network, which can be partitioned into local networks. For each local network centered on gene $k$, the local SLE score $\Delta H(k)$ is calculated. (**B**) Give a series of samples or time-course samples from an individual, the tipping point or the pre-disease state can be identified through the significant increase of SLE, i.e. the SLE changes gradually when the system is in the normal state, while it increases abruptly when the system approaches the tipping point, due to the dynamic nature of SLE. (**C**) Different from the traditional biomarkers based on differential-expression genes, SLE actually identified the pre-disease sample by exploring the dynamic and network features that result in the high SLE even with non-differential genes, that is, our SLE method can uncover the 'dark genes' (in the sense of differential associations), which plays an important role when the system approaches the critical point

disease state which generally are similar to the normal state in terms of phenotypes and gene expressions (Supplementary Fig. S1). This is also the reason why the clinical judgments through traditional biomarkers may fail to identify a pre-disease state. By exploring the information of differential associations of the observed molecules between the normal and pre-disease states, we proposed a new type of dynamic network biomarker (DNB) (Chen *et al.*, 2012; Liu *et al.*, 2012a) with three statistical conditions to detect the tipping point at the network level. Specifically, it was theoretically proved that when a biological system from a normal state approaches the critical state/tipping point, a DNB or a group of molecules (or variables) appear and satisfy the three statistic conditions (Chen *et al.*, 2012): i.e. correlations between the variables among this group rapidly increase, correlations between this group and other variables rapidly decrease and standard deviations of the variables among this group drastically increase. Based on these three necessary conditions, DNB score is constructed to serve as an indicator of the imminent state transition for predicting the disease state. In contrast to the information of differential expressions widely used in traditional molecular biomarkers to 'diagnose disease', DNB is a type of network biomarker based on the information of differential associations, thereby being capable to 'predict disease' or 'diagnose pre-disease' state. Having been applied to real biological and clinical data by many groups, the DNB method and its follow-up modifications had identified the pre-disease states of several diseases (Chen *et al.*, 2016, 2017, 2019; Liu *et al.*, 2014b; Liu *et al.*, 2015), detected the tipping points of cell fate decision (Mojtahedi *et al.*, 2016) as well as cellular differentiation (Richard *et al.*, 2016) and further investigated the immune checkpoint blockade (Lesterhuis *et al.*, 2017) and so on. However, to quantify DNB, its three statistical conditions such as standard deviation and correlation, require multiple samples from an individual, which are generally unavailable for many biomedical studies, thus significantly restricting the application of DNB method in most real cases. The single-sample problem is demanded from many biomedical fields, particularly in clinical practice, that is, there is usually only one single case sample available. The lack of samples for one individual results in the failure of traditional statistic method to analyze the biomedical problem at a network level.

In this paper, solely based on one sample, we proposed a sample-specific method, called single-sample landscape entropy (SLE), to identify the critical state just before the disease state or detect the early-warning signals of the critical transition (Fig. 1). Specifically, by exploring the dynamical difference between the normal and pre-disease states, a local network-based entropy, i.e. SLE score, is constructed to characterize the statistical perturbation brought by each individual sample against a group of given control/reference samples (Fig. 1A). The normal/reference samples refer to samples collected either previously from an individual, or from a number of healthy/relatively healthy individuals, which are regarded as the background or the normal samples in the SLE algorithm. With solid theoretical background of DNB theory, the proposed SLE score and its corresponding algorithm have the following four advantages. (i) Different from traditional differential-gene or other static methods that focus on the disease state, SLE method aims at capturing dynamical information and identifying the pre-disease state during the progression of a biological system (Fig. 1B). (ii) It detects individual-specific entropy biomarkers, depending on the temporal and spatial information from both individual (one-sample) information and the reference data. (iii) It is capable of analyzing a single-sample and thus applicable to most biomedical cases. (iv) It helps to reveal 'dark genes', which are non-differential but sensitive to network entropy scores and perform well in prognosis (Fig. 1C). This new computational method was validated by both simulation and real datasets. Applying the method of SLE to an individual-sample dataset of influenza virus infection, the significant increase of SLE scores signaled the respective (upcoming) influenza symptoms for all symptomatic individuals, while the smooth SLE scores showed no false positive signals for the asymptomatic subjects. SLE method was then applied to the TCGA dataset of lung adenocarcinoma (LUAD). Before a critical transition into cancer distant metastasis in Stage IV, the significant increase of SLE score during stages IIIA–IIIB provided early-warning signals of the disease deteriorations (Chiang and Massagué, 2008; Klein, 2008). Functional enrichment analysis for the genes in the top SLE module showed that the functions of SLE-sensitive genes are consistent with the phenotype of viral infection for influenza virus infection and cancer processes of metastasis. Moreover, SLE score identified the pre-disease states for acute lung injury (Supplementary Information SC) and prostate cancer (Supplementary Information SD). For all these complex diseases, the critical states were identified successfully before the appearance of severe disease deterioration. These applications are all coincident with the clinical or experimental observation. The analyses of real data also provided biological insights into the molecular mechanisms of the critical transitions from the perspectives of both molecules and networks for these complex diseases.

In summary, SLE provides an effective network approach for studying complex diseases based on a novel entropy-induced data transformation. It opens a new way to identify the pre-disease state and its leading network even with 'dark genes' during disease progression using only a single-sample of each individual. Therefore, SLE is of great potential in clinic application, based on which it is possible to characterize individual-specific state through only a single-sample and thus achieve the personalized disease prediction.

## 2 Materials and methods

### 2.1 Theoretical background

The disease progression can be modeled by three states or stages (Supplementary Fig. S1 and Fig. 1B) (Chen *et al.*, 2012; Scheffer *et al.*, 2009; Li *et al.*, 2017; Liu *et al.*, 2019): (i) normal state, which

is a stable state with high resilience and robustness to perturbations; (ii) pre-disease state, which is the tipping point just before the catastrophic shift into the irreversible disease state and is thus characterized by low resilience and robustness due to its critical dynamics, but is still reversible to the normal state with appropriate treatments; and (iii) disease state, which is another stable state generally with high resilience and robustness and is thus usually very difficult to return to the normal state even with intensive medical treatment.

The SLE method aims at identifying the pre-disease state before an irreversible critical transition into the disease state. The theoretical background of SLE approach is our recently proposed DNB theory, which was developed to quantitatively identify the critical state or tipping point during the progression of a complex system based on multiple samples. Theoretically, when a complex system is near the critical point, among all observed variables there exists a dominant group defined as the DNB variables/biomolecules, which satisfy the following three conditions based on the observed data (Chen *et al.*, 2012):

- The correlation ($PCC_{in}$) between any pair of members in the DNB group rapidly increases;
- The correlation ($PCC_{out}$) between one member of the DNB group and any other non-DNB member rapidly decreases;
- The standard deviation ($SD_{in}$) or coefficient of variation for any member in the DNB group drastically increases.

In other words, the above conditions are necessary conditions of the phase transition, and can also be approximately stated as: the appearance of a strongly fluctuating and highly correlated group of features/variables implies the imminent transition into the disease state. Then, these three conditions are adopted to quantify the tipping point as the early-warning signals of diseases, and further, the identified dominant group of biomolecules consists of DNB members. The DNB theory has been applied to a number of analyses of disease progression and biological processes to predict the critical states as well as their driven factors (Liu *et al.*, 2016, 2018; Yang *et al.*, 2018; Yu *et al.*, 2017).

## 2.2 Algorithm to identify the tipping point based on SLE

Given a number of reference samples (samples from normal cohort which are used as the background that represents the healthy or relatively healthy individuals), we carry out the following algorithm to identify the tipping point by using only one sample (see Fig. 1A).

[Step 1] Map the genes to protein-protein interaction (PPI) network (or other template network), forming a global network $N^G$. In this work, we use the PPI network downloaded from STRING (https://string-db.org) (Szklarczyk *et al.*, 2015), which incorporates the interactions of the selected genes with a confidence level 0.800. All the isolated nodes, without any links to other nodes, were discarded. Clearly, for all individual samples, the PPI network $N^G$ is identical as a template network.

[Step 2] Extract each local network from the global network $N^G$, such that each local network $N^k$ ($k = 1, 2, \ldots, Q$) is centered at a gene $g^k$, whose 1st-order neighbors $\{g_1^k, g_2^k, \ldots, g_M^k\}$ are the edges (Fig. 1A). There are totally $Q$ local networks if there are $Q$ genes in $N^G$.

[Step 3] For each local network $N^k$ ($k = 1, 2, \ldots, Q$) at a time point $t$, calculate the local entropy $H^n(k,t)$ based on $n$ reference samples $\{s_1(t), s_2(t), \ldots, s_n(t)\}$ (Fig. 1A), i.e.

$$H^n(k,t) = -\frac{1}{\log M}\sum_{i=1}^{M} p_i{}^n(t)\log p_i{}^n(t), \tag{1}$$

with

$$p_i{}^n(t) = \frac{|\mathrm{PCC}^n(g_i{}^k(t), g^k(t))|}{\sum_{j=1}^{M}|\mathrm{PCC}^n(g_j{}^k(t), g^k(t))|}, \tag{2}$$

where $\mathrm{PCC}^n(g_i{}^k(t), g^k(t))$ represents the Pearson Correlation Coefficient between the center gene $g^k$ and a neighbor $g_i^k$ based on $n$ reference samples. In Eq. (1), the superscript $k$ records that the local network is centered at $g^k$, the subscript $n$ denotes the number of samples and constant $M$ is the number of edges/neighbors in the local network $N^k$. In Eq. (2), symbols $g^k(t)$ and $g_i^k(t)$ respectively denote the expressions of genes $g^k$ and $g_i^k$ at time point $t$.

[Step 4] For a single sample $s_{case}(t)$ of an individual, mix it with $n$ reference samples. Calculate the entropy $H^{n+1}(k,t)$ based on $n+1$ mixed samples $\{s_1(t), s_2(t), \ldots, s_n(t), s_{case}(t)\}$ (Fig. 1A)

$$H^{n+1}(k,t) = -\frac{1}{\log M}\sum_{i=1}^{M} p_i{}^{n+1}(t)\log p_i{}^{n+1}(t), \tag{3}$$

In Eq. (3), the definition of $p_i{}^{n+1}$ is similar to that in Eq. (2), but the correlation $\mathrm{PCC}^{n+1}(g_i{}^k(t), g^k(t))$ is based on $n+1$ mixed samples $\{s_1(t), s_2(t), \ldots, s_n(t), s_{case}(t)\}$.

[Step 5] Calculate the differential entropy $\Delta H(k,t)$ between $H^n(k,t)$ and $H^{n+1}(k,t)$, i.e.

$$\Delta H(k,t) = \Delta \mathrm{SD}(k,t)|H^{n+1}(k,t) - H^n(k,t)|, \tag{4}$$

with

$$\Delta \mathrm{SD}(k,t) = |\mathrm{SD}^{n+1}(k,t) - \mathrm{SD}^n(k,t)|, \tag{5}$$

where $\mathrm{SD}^n(k,t)$ and $\mathrm{SD}^{n+1}(k,t)$ are the standard deviations of the gene expression for center gene $g^k$ respectively based on $n$ reference samples $\{s_1(t), s_2(t), \ldots, s_n(t)\}$ and $n+1$ mixed samples $\{s_1(t), s_2(t), \ldots, s_n(t), s_{case}(t)\}$. The differential entropy $|H^{n+1}(k,t) - H^n(k,t)|$ between $H^n(k,t)$ and $H^{n+1}(k,t)$ characterizes the differences caused by the single case sample $s_{case}(t)$. In other words, comparing with the local entropy $H^n(k,t)$ based on $n$ reference samples $\{s_1(t), s_2(t), \ldots, s_n(t)\}$, $H^{n+1}(k,t)$ based on $n+1$ mixed samples $\{s_1(t), s_2(t), \ldots, s_n(t), s_{case}(t)\}$ records the perturbation brought by the single sample $s_{case}(t)$ for local network $N^k$. Besides, to bring the fluctuation of genes into consideration, the differential standard deviation $\Delta \mathrm{SD}(k,t)$ is regarded as the weight coefficient.

[Step 6] Calculate the weighted sum of $\Delta H(k)$ for all the local networks, i.e.

$$\Delta H(t) = \frac{1}{Q}\sum_{k=1}^{Q}\Delta H(k,t), \tag{6}$$

where constant $Q$ is the number of all genes. In Eq. (6), $\Delta H(t)$ records the overall impact caused by the single sample $S_{case}(t)$, and thus is called the **global SLE score** or just **SLE score** for the global network $N^G$, while $\Delta H(k,t)$ in Eq. (4) is called the **local SLE score** for the local network $N^k$ centered at gene $g^k$.

According to the DNB theory (Chen *et al.*, 2012), when the system approaches the critical state, the DNB biomolecules exhibits significantly collective behaviors with fluctuations. Thus, in a local network within which the nodes are DNB biomolecules, the correlation coefficient $\mathrm{PCC}^{n+1}(g_i{}^k(t), g^k(t))$ in the probability $p_i{}^{n+1}$ become more similar or equalized when the system is near a critical point (see Section E in Supplementary Information for detailed derivation), leading to the increase of local SLE score $\Delta H(k)$ in Eq. (4). Moreover, in Eq. (6), the term $\Delta \mathrm{SD}(k,t)$ increases accordingly, which also contributes to the boost of SLE score $\Delta H(t)$. Thus, SLE score can provide the early-warning signals of the critical transition. From above algorithm, it is seen that the proposed method is not governed by a specific kinetic system, and thus model free.

## 2.3 Data processing and functional analysis

We applied the SLE scheme to four time-course or stage-course datasets, i.e. lung adenocarcinoma (LUAD) from TCGA database, the time-course dataset for influenza virus infection process (GSE30550), the microarray data of acute lung injury induced by phosgene inhalation (GSE2565), and prostate cancer (GSE5345) which were downloaded from the NCBI GEO database (www.ncbi.nlm.nih.gov/geo). For all these omics data, we discarded the probes without corresponding NCBI Entrez gene symbol. For each gene mapped by multiple probes, the average value was employed as the

gene expression. The procedure of building a molecular interaction network was as follows. First, the biomolecular association networks for *Homo sapiens* and *Mus musculus* were downloaded from several public databases, e.g. protein–protein interactions from STRING (http://string-db.org) (Szklarczyk *et al.*, 2015), and transcriptional regulations from TRED (rulai.cshl.edu/cgi-bin/TRED/tred.cgi? process=home). We integrated this linkage information together without redundancy into a whole molecular interaction network including 65 625 functional linkages in 11 451 molecules for *Homo sapiens*, and 37 950 linkages in 6683 molecules for *Mus musculus*. Second, the genes from each microarray dataset were mapped to the integrated network to extract the related linkages. The molecular network was used for consequent analysis. Finally, our main results were visualized by Cytoscape (www.cytoscape.org) in the post-processing step. The functional analysis including gene ontology and pathway enrichment was based on GO database (http://www.geneontology.org/page/go-database) and KEGG mapper tool (http://www.genome.jp/kegg/tool/map_pathway2.html).

Dataset GSE30550 comprises expression profiles of humans with influenza virus infection. The gene expression profiles were obtained and measured on whole peripheral blood drawn from all subjects at an interval of 8 h post-inoculation (hpi) through 108 hpi. There are 11 961 probe sets and 17 samples in the original GSE30550 dataset, and 11 619 gene symbols were mapped from the ID of the probe sets.

The LUAD dataset contained RNA-Seq data and included both tumor and tumor-adjacent samples. The tumor samples were divided into different stages based on clinical (stage) information from TCGA, and the samples without stage information were ignored.

Functional annotations were performed by searching the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene). The enrichment analyses were separately obtained using web service tools from the Gene Ontology Consortium (GOC, http://geneontology.org) and client software from Ingenuity Pathway Analysis (IPA, http://www.ingenuity.com/products/ipa).

## 3 Results

We present the definition and theoretical explanation of SLE score in Section 2. Here, we used a single-sample with high-throughput omics data, to identify the pre-disease state or early warning signals of the disease deterioration based on the SLE score. Achieving reliable identification with only one sample is of great importance in clinic application since it is usually difficult to obtain multiple samples from an individual who does not yet exhibit any disease symptoms during a short period. To illustrate how SLE works, we applied our method first to a simulated dataset, and then four real datasets, including influenza infection (GSE30550), prostate cancer (GSE5345) and acute lung injury (GSE2565) from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) and LUAD from the TCGA database (http://cancergenome.nih.gov). The applications of SLE score in influenza infection and LUAD are illustrated in the main text, while the others are provided in Supplementary Information. The successful identification of the pre-disease states in these diseases validated the effectiveness of SLE method in quantifying the tipping point just before the critical transitions into severe disease states.

### 3.1 Validation based on numerical simulation

In order to validate the proposed SLE method, we employed a theoretical network with sixteen nodes (Fig. 2A) to illustrate the identification of the early-warning signals when the system approaches a tipping point. Such a regulatory network with Michaelis-Menten or Hill form is often employed in the analysis to study gene regulatory activities such as transcription and translation (Garcia-Ojalvo *et al.*, 2004; Sherman and Cohen, 2012; Cantone *et al.*, 2009), and multistability and nonlinear biological processes (Chen *et al.*, 2015; Li *et al.*, 2006). In addition, the bifurcation in Michaelis-Menten form is often employed to model the state transition of gene regulatory networks (Gardner *et al.*, 2000; O'Brien *et al.*, 2012). Detailed description of the network characterized by a set of sixteen stochastic

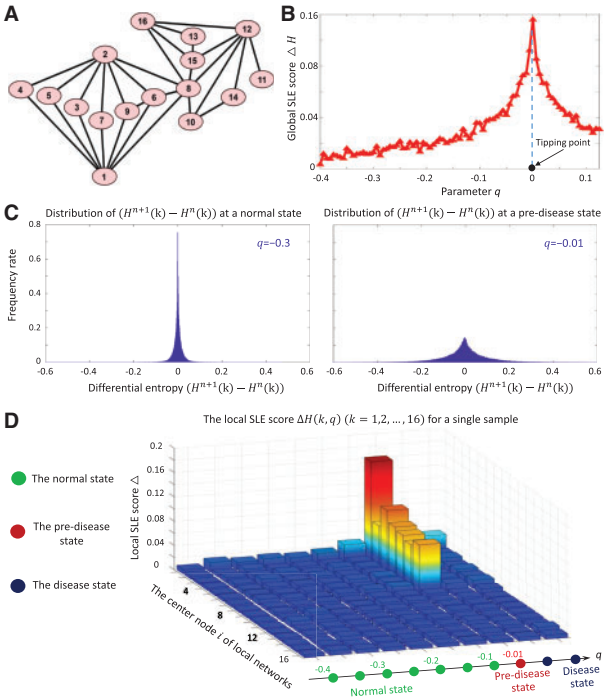


**Fig. 2.** The validation of SLE method through a numerical experiment. (**A**) A network with 16-nodes, based on which the numerical simulation is conducted. (**B**) The curve of global SLE score $\Delta H$ defined in Eq. (6). It is clear that the sudden increase of SLE score indicates the upcoming state transition at $q = 0$, which is in accordance with the bifurcation parameter value at $q = 0$ [see Eq. (S3) in Supplementary Information]. (**C**) The distributions of differential entropy $(H^{n+1}(k) - H^n(k))$ of the network respectively at $q = -0.3$ (left) and $q = -0.05$ (right). The distribution of the differential entropy follows the so-called volcano distribution (Liu *et al.*, 2016). Comparing with the distribution of SLE at the normal state, the volcano is much flatter when the system is at the pre-disease state, indicating the significant difference between the reference and single case sample in the critical stage. (**D**) The landscape of local SLE scores $\Delta H(k)$ $(k = 1, 2, \ldots, 16)$ [as defined as in Eq. (4)] respectively for 16 local networks. It clearly shows that when the system approaches the tipping point at $q = 0$, the drastic increase of SLE scores for some local networks $\Delta H(k)$ $(k = 1, 2, \ldots, 7)$ signals the tipping point of the system. Actually, those indicative local networks are all centered at the DNB variables. See the detailed description in Supplementary Information SB

differential equations Eq. (S3) in Michaelis-Menten form, was provided in Supplementary Information SB. Based on a parameter $q$ varying from $-0.4$ to 0.15 with $q = 0$ as the tipping point, a dataset was generated for numerical simulation from the network.

In Figure 2B, we demonstrated the global SLE score, i.e. $\Delta H$ in Eq. (6), for the whole 16-node network. It is seen that a sharp increase of the global SLE score indicated the upcoming tipping point at a bifurcation parameter value $q = 0$. The underlying mechanism for SLE lies in the fact that the differential entropy $(H^{n+1}(k) - H^n(k))$ follows distinct distributions respectively in the normal state $(q = -0.3)$ and in the pre-disease state $(q = -0.05)$ (Fig. 2C), that is, when the system is in the normal state, most probably each differential entropy is at a low level; on the other hand, when the system approaches a tipping point or is in the pre-disease state, some of the differential entropies increase significantly. The sudden increase of a local SLE score is due to the collective behaviors within the corresponding local network, that is, the DNB critical properties include the drastic fluctuation of nodes and the strong correlation among them in terms of their expressions. The detailed theoretical explanation is provided in Section 2 and Supplementary Information SA.

To exhibit the distinct dynamics of the system between the normal state and the pre-disease state, we present the dynamical changes in local SLE scores $\Delta H(k)$ $(k = 1, 2, \ldots, 16)$ [as defined as in Eq. (4)] respectively for 16 local networks, illustrating the landscape of the network entropy in a global view (Fig. 2D). It is clear that when the system

is far away from the tipping point, all the local SLE scores are smooth and at a low level; when the system approaches the tipping point $q = 0$, the SLE scores of some local networks $\Delta H(k)$ $(k = 1, 2, \ldots, 7)$ drastically increase, which indicates the upcoming tipping point or pre-disease state. It should be noted that those local networks with sharp increasing SLE scores are actually centered in the so-called DNB members, which is regulated by the dominant eigenvalue (see Supplementary Information SA for details). Through this numerical experiment, it is clear that the SLE approach is capable of exploiting the high-dimensional information, even if there is only a single-sample. The numerical experiment validates that the SLE score is reliable and accurate in signaling the critical stage. We present the simulation and calculation details in Supplementary Information SB. Our method is a model-free method, which does not require a specific dynamical model of the system under study. The model Eq. (S3) presented in Supplementary Information is only for generating input data or samples to numerically test our method, and is irrelevant with the real-data applications.

### 3.2 Tipping points of individual influenza infection

In dataset GSE30550, the time-series data records the influenza virus infection process of 17 human adult subjects, who were inoculated with live influenza virus H3N2/Wisconsin (Huang *et al.*, 2011). Gene expressions were derived for these subjects at 16 sampling time points (−24, 0, 5, 15, 21, 29, 36, 45, 53, 60, 69, 77, 84, 93, 101 and 108 h) (Fig. 3C). Nine of the 17 adults were symptomatic subjects, who developed clinical symptoms of influenza infection. The other eight were asymptomatic subjects, who did not have any clinical symptom at all time points (Fig. 3C). For each individual, the gene expression profiles of the former 4 time points, i.e. -24, 0, 5 and 12 h, were regarded as reference samples, that is, the reference samples which reflect the relatively healthy state of a volunteer were also the individual-based samples.

Following the algorithm provided in Section 2 (also see Fig. 1), 17 SLE scores [$\Delta H$ defined in Eq. (6)] were calculated respectively based on 17 individual samples in each time point (Fig. 3A). Specifically, for each individual sample, the local SLE scores ($\Delta H(k, t)$ in Eq. (4)) were calculated for every piece of local network on the basis of an integrated *Homo sapiens* network (see Section 2 section for more details about the network). Then the global SLE score [$\Delta H(t)$ in Eq. (6)] for each individual subject was presented as the final score in Figure 3A. A drastic increase of SLE score indicates the imminent appearance of flu infection symptoms (also see Fig. 3B for each symptomatic subject).

The increase of SLE score provides an early-warning signal for the disease state, i.e. the stage that clinical symptom arises. The score of SLE in the nine symptomatic subjects drastically increases before the appearance of influenza symptom, and the score in the eight non-symptomatic subjects is stable on the whole (Fig. 3A). Hence, the early-warning signals for the influenza symptom had been detected in the 9 symptomatic subjects before the appearance of the influenza symptoms, and there was no warning signal detected for the 8 non-symptomatic subjects (Fig. 3A and C). Clearly, the new SLE approach can effectively identify the pre-disease samples and accurately detect the early-warning signals for influenza virus infection on an individual basis. Therefore, at each time point, the SLE score can be employed to identify the critical state/stage of complex diseases based on solely an individual single-sample. The landscape of local SLE score for the 9 symptomatic subjects are presented in Figure 4.

For each individual sample at the corresponding signaling time point, i.e. the time point with blue star mark in Figure 3B, the center genes of all local networks were ranked according to the local SLE scores. And the top 5% genes with the largest local SLE scores were regarded as the SLE genes that are highly related to the occurrence of symptoms. Then functional analysis was performed based on these selected SLE genes.

### 3.3 Functional analysis of SLE genes for influenza infection

For each individual subject at the respective tipping point, the top 5% genes with the largest local SLE scores were regarded as the SLE genes. To validate the effectiveness of SLE score, the functional

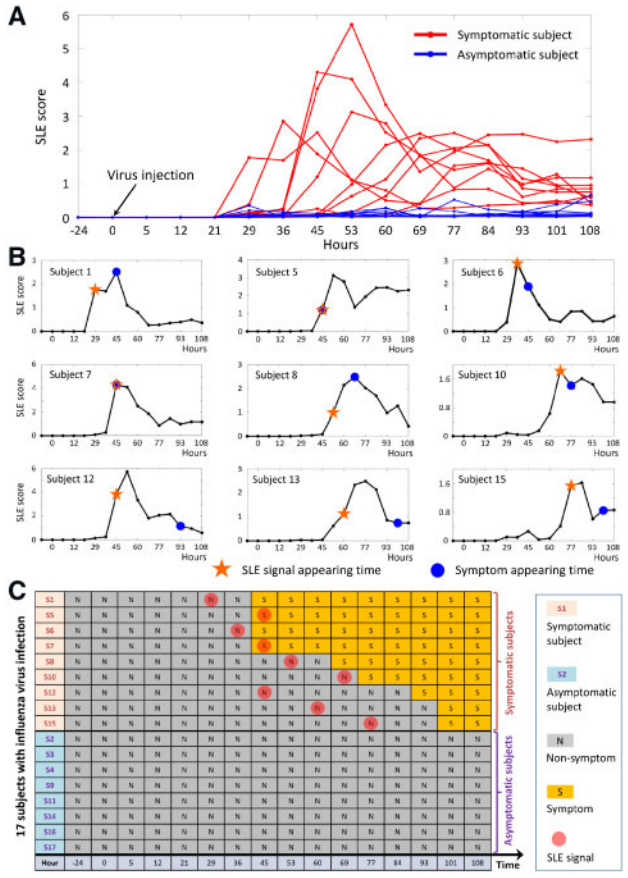


**Fig. 3.** Identification of the tipping point of H1N2 influenza infection based on a single sample. (**A**) The curves of SLE scores [$\Delta H$ defined in Eq. (6)] for 17 subjects. Each red curve records for the SLE score based on the individual data of a symptomatic subject, while each blue curve corresponds to an asymptomatic individual. (**B**) The individual SLE score curves of 9 symptomatic subjects. For each symptomatic subject, the purple circle stands for the time point at which the initial flu symptoms arises, i.e. the clinically diagnosed infection-time, and the blue star mark denotes the identified tipping point or the pre-disease state by SLE score. (**C**) The summarized prediction results. It is seen that for each symptomatic subject, the SLE score successfully indicates the upcoming infection symptoms, while for each asymptomatic subject, there is no false warning signal

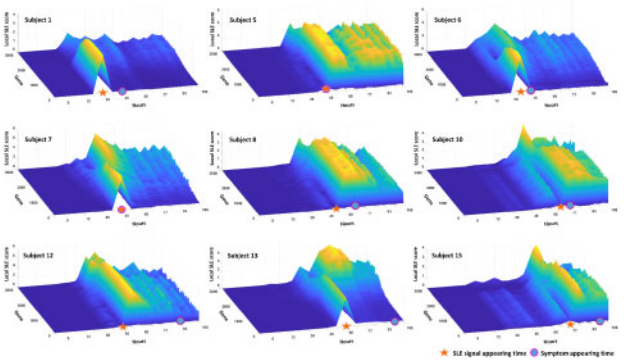

**Fig. 4.** The landscapes of the local SLE score for 9 symptomatic subjects. The dynamical change of local SLE scores demonstrates the landscape of the network entropy in a global view. For each subject, the purple circle stands for the time point at which the initial flu symptoms arises, i.e. the clinically diagnosed infection-time, and the orange star mark denotes the identified tipping point or the pre-disease state. (Color version of this figure is available at *Bioinformatics* online.)

analysis is carried out based on the common SLE genes that appear in multiple symptomatic SLE gene sets. Functional analyses of these SLE genes were performed through Ingenuity Pathway Analysis (IPA). Besides, web analysis tool Gene Ontology Consortium (Ashburner *et al.*, 2000) analysis (GO analysis, http://www.geneontology.org/) was also employed to find biological features of SLE gene set associated with influenza infection.

Based on GO and IPA, the enrichment analysis was proceeded for the common SLE genes that appear in multiple tipping points (multiple subjects). As shown in Table 1, these common SLE genes are enriched into the biological processes of immunity or defense against the influenza virus, e.g. 'positive regulation of response to external stimulus' (GO: 1903047), 'regulation of immune response' (GO: 0050776), 'regulation of lymphocyte mediated immunity' (GO: 0002706), 'defense response to virus' (GO: 0051607), 'positive regulation of natural killer cell differentiation' (GO: 0032825) and others in Gene ontology. Besides, SLE genes are enriched in virus-infection functions annotation including 'viral infection', 'activation of lymphocytes', 'response of macrophages', 'antimicrobial response', 'replication of virus' and others in IPA analysis. Moreover, six genes with high frequency in all SLE gene groups that respectively come from nine symptomatic subjects were listed in Table 2. These genes are involved in the biological processes related to influenza infection. The whole list of common SLE genes is provided in Supplementary Table S1. All enriched functions of common SLE genes for influenza infection is provided in Supplementary Table S2. The functional network based on common SLE genes is provided in Supplementary Figure S2.

Clearly, the functional analysis shows that the SLE genes are involved in the key biological processes or functions during influenza infection. It is thus consistent with the process of influenza virus infection, implying that the SLE genes not only identified the pre-disease state prior to the emergence of symptoms but also were involved in the key biological processes of influenza infection. These SLE genes and their mutual interactions may be considered as potential drug targets against influenza infection, which will be our future topic.

### 3.4 Critical state of lung adenocarcinoma (LUAD)

For the lung adenocarcinoma (LUAD) datasets, 459 tumor samples and 58 tumor-adjacent samples were obtained from TCGA. Based on clinical information, the samples were grouped into seven stages, i.e. Stage IA, IB, IIA, IIB, IIIA, IIIB and IV of lung cancer (Table 3). The tumor-adjacent (TA) samples were considered as reference samples or normal controls in this study.

Metastasis, the process by which cells leave a tumor and colonize distant sites, is the major cause of death for cancer patients with solid malignant tumors (Steeg *et al.*, 1988). Especially for non-small cell lung cancer, tumor metastasis is primarily responsible for the low 5-year survival rate (Marcus and Zhou, 2010). Stage IV is usually an advanced or metastatic cancer in which the tumor cells have invaded into distant tissues of other organs (Chiang and Massagué. 2008). Thus, it is crucial to identify the tipping point prior to cancer distant metastasis, so that chemotherapy and radiotherapy or other strategies that are employed in treating non-small cell lung cancer in Stage III, can be carried out timely to prevent from serious deterioration or slow down cancer progression (Bareschino *et al.*, 2011). In order to detect the early-warning signal of cancer metastasis, the SLE score was applied to the LUAD datasets. At each stage, the SLE score, i.e. the differential entropy $\Delta H(t)$ in Eq. (6), was calculated for every single sample. Different from influenza-infection dataset, there are no individual-based samples across all stages. Thus, in each stage, an average SLE score against all single-sample-based entropies was employed to quantify the risk of critical transition into cancer distant metastasis.

As shown in Figure 5A, the sharp increase of the global SLE score [$\Delta H(t)$ in Eq. (6)] was detected around Stage IIIB of LUAD, indicating the imminent critical transition into the cancer metastasis stage (Stage IV). It should be noted that there is different number of samples in each stage, while according to SLE scheme, for each single sample there is an SLE score. To quantify the tipping point, the average value was adopted and presented as a unified score for each stage. To show the local SLE scores [$\Delta H(k,t)$ in Eq. (4)] in a global view, the landscape of the entropy values for 6804 local networks was illustrated (Fig. 5B). It can be seen that around IIIB stage, there is a group of genes whose local SLE scores abruptly increase (Fig. 5B). This critical phenomenon resulted from the drastic increase of the correlations between molecules in this group when the system approaches the tipping point (see Section 2). In Figure 5C, we illustrated the evolution of the top SLE gene group/module, i.e. a mapped STRING network of top 200 genes with the largest local SLE scores. It can be seen that a significant change in the network structure occurs at Stage IIIB, signaling the critical transition into cancer distant metastasis in Stage IV. The whole dynamics of the top SLE module across all 7 stages are given in Supplementary Figure S3. We also presented the dynamical evolution of the whole gene network as in Figure 5D. At the lower left corner, a group of top 200 genes with the largest local SLE scores are intentionally grouped, whose collaborative fluctuation occurs in Stage IIIB (Fig. 5D). The whole dynamics of the network composed by all genes across the 7 stages are given in Supplementary Figure S4.

Now that Stage IIIB is a critical stage for the cancer distant metastasis, we ranked the genes by their corresponding local SLE score in Stage IIIB, that is, the entropy value for the local network centered in a gene. For each case sample, a group of top 5% genes with the largest local SLE values were then selected to be further analyzed. It was showed in the following sections that these SLE genes not only are involved in metastasis processes or metastasis-related functions, but may perform better in prognosis.

### 3.5 Functional analysis of SLE genes for LUAD metastasis

The KEGG enrichment analysis (Kanehisa and Goto, 2000) showed that the SLE genes (top 5% genes with the largest local SLE values in Stage IIIB) were highly associated with biological processes of cancer metastasis, such as, focal adhesion, chemokine signaling pathway, cell cycle, mTOR signaling pathway, Ras signaling pathway, EGFR tyrosine kinase inhibitor resistance, migration and invasion.

Through literature searching, some genes in the common SLE gene group have been shown to be associated with the process of cancer metastasis. For instance, LIMK1, whose mutation occurs commonly in lung adenocarcinomas and large cell carcinoma, functionally modulates non-small-cell lung cancer metastasis (Ji *et al.*, 2007). LIMK1 is also a cancer drug (Dabrafenib) target against

**Table 1.** The enrichment analysis for common SLE genes based on GO and IPA analysis

| Gene Ontology Consortium | | IPA | |
|---|---|---|---|
| Enriched function | *P* value | Enriched biological process | *P* value |
| Positive regulation of response to external stimulus (GO: 1903047) | 2.45E-08 | viral infection | 1.80E-15 |
| Regulation of immune response (GO: 0050776) | 5.11E-07 | activation of lymphocytes | 3.97E-14 |
| Regulation of lymphocyte mediated immunity (GO: 0002706) | 1.75E-06 | response of macrophages | 1.65E-12 |
| Defense response to virus (GO: 0051607) | 6.74E-06 | antimicrobial response | 4.27E-11 |
| Positive regulation of natural killer cell differentiation (GO: 0032825) | 2.55E-04 | replication of virus | 5.52E-09 |

**Table 2.** The genes with high frequency in nine SLE gene groups respectively from nine symptomatic individual samples

| Gene | Frequency | Location | Family[a] | Relation with cancer metastasis |
|---|---|---|---|---|
| ACADL | 9 | Cytoplasm | Enzyme | During influenza infection, multiple steps in the $\beta$-oxidation of long chain fatty acids including ACADL were disrupted at the mRNA level (Tarasenko *et al.*,2015). |
| MMP9 | 9 | Extracellular Space | Peptidase | MMP9 is closely related to influenza pathogenesis by mediating excessive neutrophil migration into lung, which is involved with viral replication (Bradley *et al.*, 2012). Mmp9 is essential for tissue and cellular repair and development in the lung during influenza infection (Jamieson *et al.*, 2013). |
| MNX1 | 9 | Nucleus | Transcription regulator | MNX1, a motor neuron marker, is enriched to inflammatory response according to IPA. |
| EWSR1 | 8 | Nucleus | Ion channel | During influenza infection, EWSR1 may disturb gene expression by mimicking or interfering with the normal function of CTD-POLII within the transcription initiation complex (Bavagnoli and Maga, 2013). |
| FLT4 | 8 | Plasma Membrane | Transmembrane receptor | FLT4, also known as VEGFR-3, disorders of NF-$\kappa$B-mediated immunity (Puel *et al.*, 2004). |
| FZD2 | 8 | Plasma Membrane | G-protein coupled receptor | Influenza virus infection downregulates the expression of proteins of Wnt/$\beta$-catenin signaling pathway like FZD2 (Almansa *et al.*, 2017). |
| HIGD1A | 8 | Cytoplasm | Other | In the complex seizures due to influenza A(H1N1) pdm09, HIGD1A which is related to cellular response to stress is significantly downregulated (Tsuge *et al.*, 2014). |
| NFATC1 | 8 | Nucleus | Transcription regulator | NFATC1 is critical for host resistance to viral infections (Maruya *et al.*, 2011). |

[a]From Ingenuity Pathway Analysis.

**Table 3.** The number of tumor samples within each stage in the LUAD dataset from TCGA

| Stage | IA | IB | IIA | IIB | IIIA | IIIB | IV | TA[a] |
|---|---|---|---|---|---|---|---|---|
| Samples | 106 | 124 | 39 | 59 | 62 | 10 | 21 | 58 |

[a]TA refers to the tumor-adjacent samples, which are employed as the reference.

LUAD by inhibiting non-small-cell lung cancer cell growth and metastasis (Koscielny *et al.*, 2017). EGFR is considered as a key gene in the progression and treatment of lung adenocarcinomas, whose overexpression has been discovered in 40% of lung adenocarcinomas cases (Walker *et al.*, 2009). CXCL12, a prognostic marker significantly associated with poor prognosis in lung cancer patients (Wen *et al.*, 2011), may activate the canonical extracellular signal–regulated kinases (ERK) signaling pathway, which promotes cell invasion and is involved in lung cancer metastasis (Cristea and Sage, 2016). PRNP is identified as highly associated with invasion in non-small cell lung cancer (Lader *et al.*, 2004). Transcription factor SOX18, a critical switch for lymphangiogenesis, plays an important role in cancer metastasis (Tobler and Detmar, 2006). Researchers also showed that suppressing SOX18 function is sufficient to impede tumor metastasis (Duong *et al.*, 2012). TP53 codon 72 polymorphism was associated with an increased risk of lung cancer (Piao *et al.*, 2011). DLX4, a gene inhibited motility and invasion via both hematogenous and lymphogenous routes, is one of the metastasis signature genes (Tomida *et al.*, 2007). Overexpression of TREM2 may enhance tumor cell proliferation and invasion (Araki *et al.*, 2009). By mediating heterotypic and homotypic intercellular adhesion, gene IGSF1 is involved in the invasion and metastasis process (Xue *et al.*, 2005). Mutation in gene TNFRSF6B, which may block Fas ligand, has been observed to be amplified in almost half of lung tumors (Cooper, 2005). Eight overlapped genes appearing in all 10 SLE gene groups that respectively come from 10 stage-IIIB single samples were listed in Table 4. These genes have all been reported to be associated with metastasis process or metastasis related

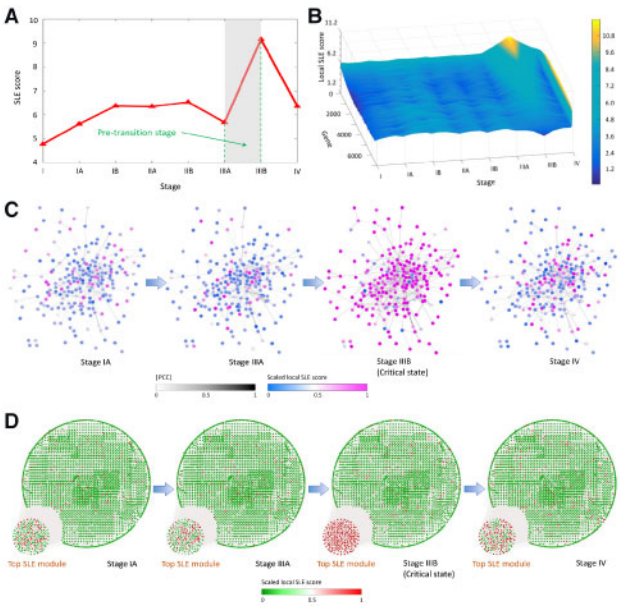


**Fig. 5.** Identification of the critical state of LUAD distant metastasis. (**A**) It exhibits SLE score curve of LUAD progression, which shows the critical state around IIIA–IIIB stages. (**B**) The dynamical change of local SLE scores demonstrates the landscape of the network entropy in a global view. (**C**) The evolution of the top SLE gene group/module, i.e. the top 200 genes with the largest local SLE scores, illustrates that a significant change in the network structure occurs at Stage IIIB. The network structure is derived by mapping genes to the STRING PPI network. All the isolated nodes, without any links to other nodes, were discarded. The color of each node represents the value of the scaled local SLE score, while the color of each edge stands for the absolute value of Pearson correlation coefficient |PCC|. The whole dynamics of the top SLE module across all 7 stages are given in Supplementary Figure S3. (**D**) We illustrate dynamical evolution of the whole gene network (the co-expressed network). All the top 200 SLE genes are intentionally put together at the lower left corner. It is seen that by using SLE approach, the early-warning signals of cancer distant metastasis from a network perspective can be detected at IIIB stage, before the critical transition into distant metastasis in Stage IV. The whole dynamics across the seven stages are given in Supplementary Figure S4

**Table 4.** The genes with high frequency in 10 SLE gene groups in the critical stage (Stage III)

| Gene | Frequency | Location | Family[a] | Relation with cancer metastasis |
|---|---|---|---|---|
| SOX18 | 10 | Nucleus | Transcription regulator | SOX18 is a critical switch for lymphangiogenesis thus relates to cancer metastasis (Tobler and Detmar, 2006). |
| TREM2 | 10 | Plasma membrane | Transmembrane receptor | TREM2 enhances tumor cell proliferation and invasion (Araki *et al.*, 2009). |
| ACYP2 | 10 | Cytoplasm | Enzyme | ACYP2 associates with colorectal cancer metastasis (Zhang *et al.*, 2016). |
| FBXO32 | 10 | Cytoplasm | Enzyme | FBXO32 is a negative regulator of epithelial to mesenchymal transition (EMT), while acquisition of EMT is important for cancer progression and metastasis (Tanaka *et al.*, 2016). |
| TTK | 10 | Nucleus | Kinase | TTK associates with metastasis via chromosomal Instability (Harima *et al.*, 2009). |
| VDAC2 | 10 | Cytoplasm | Ion channel | The up-regulated expression of VDAC2 would disturb the mitochondrial ion homeostasis in the oxidative stress (Liu *et al.*, 2012b), which drives tumor progression and metastasis (Sotgia *et al.*, 2011). |
| YWHAQ | 10 | Cytoplasm | Other | YWHAQ coordinates the regulation of proliferation, survival and metastasis (Liang *et al.*, 2016). |
| YY1 | 10 | Nucleus | Transcription regulator | YY1 positively regulates the expression of HLJ1, whose promoter contains four YY1-binding sites (Tsai *et al.*, 2014). HLJ1, a tumor suppressor, closely relates to both tumor growth and metastasis in non-small cell lung cancer (Chen *et al.*, 2008). |

[a]From Ingenuity Pathway Analysis.

**Table 5.** The functional enrichment of common SLE genes in the critical stage samples for LUAD

| Gene Ontology Consortium | | IPA | |
|---|---|---|---|
| Enriched function | *P* value | Enriched biological process | *P* value |
| Regulation of cell cycle (GO: 0051726) | 1.17E-06 | Metastatic non-small cell lung cancer | 8.16E-20 |
| Regulation of cell-cell adhesion (GO: 0022407) | 6.74E-06 | Progression of carcinoma | 4.08E-11 |
| Regulation of kinase activity (GO: 0043549) | 5.28E-05 | Locally advanced non-small cell lung carcinoma | 1.03E-08 |
| Tube morphogenesis (GO: 0035239) | 1.22E-04 | Metastatic large cell lung carcinoma | 1.08E-08 |
| Ras protein signal transduction (GO: 0007265) | 5.12E-04 | Stage IV metastatic solid tumor | 1.07E-07 |

functions. The whole list of common SLE genes is provided in Supplementary Table S3. All enriched functions of common SLE genes for influenza infection is provided in Supplementary Table S4. The functional network based on common SLE genes of LUAD was provided in Supplementary Figure S5.

Functional enrichment through GO analysis illustrated that the common SLE genes are involved in the biological processes including the regulation of cell cycle, the regulation of cell adhesion, the regulation of kinase activity and others (Table 4). All these biological processes are associated with the cancer metastasis or progression of cancer. Furthermore, in IPA (Ingenuity Pathway Analysis), these common genes were also enriched to metastasis-related function annotation, such as non-small-cell lung carcinoma and metastatic non-small-cell lung cancer at Stage IV by functional enrichment (Table 5).

### 3.6 Revealing non-differential 'dark genes' by SLE method

In most pathological and biomedical studies, differential expression genes draw much attention in finding new biomarkers, key regulators and drug targets. However, similar to non-coding RNAs (or non-coding regions of DNA) that are now considered as the 'dark matter' in sequence, the SLE-based analysis shows that some non-differential genes may play important roles in disease progression and also performs well in prognosis (Fig. 6), not at the gene level but at the network level. Thus, these non-differential genes which are sensitive to SLE score are considered as the 'dark matter' in expression.

To discover these 'dark genes', we first selected SLE genes (top 5% genes with the largest local SLE score) that are not differentially expressed. Different from traditional survival analysis based on gene expression, the survival analysis exhibited in Figure 6 is carried out based on the local SLE score, i.e. the weighted network entropy for a local network defined as in Eq. (4). In other words, Figure 6 shows the survival analysis results with $P$-values $< 0.05$ based on some non-differential genes in the common SLE gene group of LUAD. We divided the samples into two groups based on the median of local SLE score level. Each group respectively included 517 samples. Clearly, a lower level of SLE score in HARBI1, BPIL2, CNN3 and IL17C is significantly associated with poor survival, while a higher level of SLE score in LAMB3 and TJP1 is associated with poor survival, which validated the effectiveness on LUAD progression for those 'dark genes' in the SLE group.

These SLE genes may get involved in the key biological processes or functions that trigger the critical deterioration, but are usually ignored due to the non-differential characteristics in gene expression. Therefore, our new method may help to find those new biomarkers, drug targets and prognosis indicators from the perspective of SLE score.

## 4 Discussion

The lack of samples is a general problem in biological studies and clinical practice, which often results in model errors and bias in

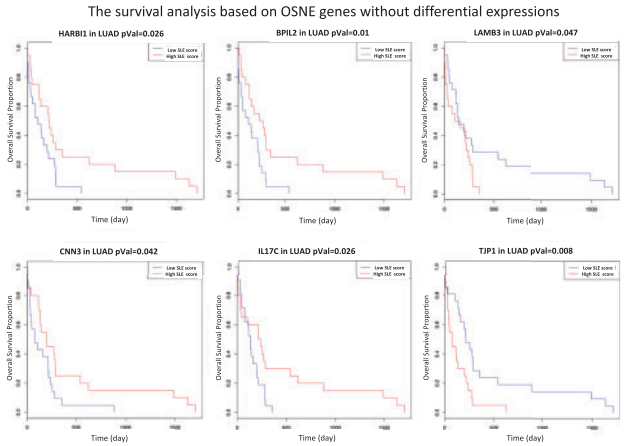


**Fig. 6.** The LUAD survival analysis based on SLE genes without differential expressions. Six genes HARBI1, BPIL2, LAMB3, CNN3, IL17C and TJP1, none of which is differentially expressed in Stage IIIB (the critical stage before distant metastasis), perform well in LUAD prognosis. Such survival analysis was based on the local SLE score, which is the SLE value for a local network centered on each above gene, not gene expression value. It is seen that the abnormally low or high local SLE scores suffice to distinguish the significantly different survival time in LUAD. The survival analysis of an aggregation of the six genes through a COX model is provided in Supplementary Figure S6

analysis. In this study, to tackle with the small-sample problem, a single-sample method was proposed to identify tipping points or critical states which appear just before the disease state. In contrast to the information of differential expressions used in traditional biomarkers to 'diagnose disease', SLE is based on the information of differential associations or differential networks among biomolecules, and thereby is capable to 'predict disease'. By exploiting the high-dimensional information of the observed data, this method was successfully applied to quantifying the tipping points from both simulation and real datasets solely based on single-sample data.

Specifically, on the basis of 17 individual time-series datasets of the influenza virus infection, the SLE score identified the pre-disease states for the 9 symptomatic subjects before the first occurrence of their respective symptoms without false SLE signal for the other 8 asymptomatic subjects. The common SLE genes were enriched to immune response, antimicrobial response and other influenza infection related functions. Both computational and functional results are consistent with the observation in the original study. The successful prediction in influenza infection validates the effectiveness of SLE score in quantifying the tipping point through only individual-based single-samples. For LUAD dataset, the significant increase of SLE scores during stages IIIA–IIIB signaled the imminent transition into distant metastasis in Stage IV. The common SLE genes of LUAD were associated with metastatic processes or metastasis-related functions. Moreover, SLE score succeeded in providing early-warning signals before critical transitions in acute lung injury and prostate cancer, which were provided in Supplementary Information SC–SD. These results also exhibited that the SLE-based analysis is consistent with the process of disease deterioration, implying that the SLE genes not only provide general early-warning signals to the disease states but are found to be involved in the key biological processes of the critical transition into a severe disease state. These SLE genes and their mutual interactions may be potential drug targets against influenza infection, which will be our future topic.

It should be noted that SLE method is a model-free approach. However, when applying SLE method, there are still a few empirical parameters to be chosen. First, in the algorithm, we use the PPI network downloaded from STRING (https://string-db.org), which incorporates the interactions of the selected genes with a confidence level 0.800. One could choose a different level or even other template network when preparing the network structure. Second, a unified threshold 5% is used in this paper, that is, the top 5% genes with the largest local SLE scores were selected as SLE genes for further biological analysis. Besides, the number of reference samples is important when applying SLE method. In Supplementary Information, we presented some discussions about the SLE performance under different reference sizes or conditions (Supplementary Information SF) and with randomly permuted orders of reference samples (Supplementary Information SE). Given reference samples, SLE score is capable of quantifying the difference between a new sample and the reference samples based on the differential associations or networks, and thus detect the imminent critical transition. On the other hand, due to the individual-based nature of SLE score, the dynamical change of SLE score and the composition of SLE genes for different individuals may differ from each other even for the same disease, but the SLE score $\Delta H$ drastically increases whenever approaching the tipping point. It should also be noted that when the reference samples and case sample are all from the same individual, the SLE could be regarded as a personalized heath index, that is, if an individual is continually monitored at a series of time points, it is possible to detect any significant dynamical changes based on SLE for this individual, by taking the healthy samples of this individual as reference.

The pre-disease state is regarded as a critical stage reversible to the normal state. Thus, appropriate medical care for patients in the pre-disease state is crucial and effective in contrast to the patients in the irreversible disease (deterioration) state. However, how to arrange the appropriate treatment is beyond the scope of this work, and will be a future topic. Moreover, theoretically, any omics data (e.g. transcriptomic data, proteomics data, or metabolomics data) which can dynamically reflect the change of the disease progression, can be used to detect the critical state or tipping point. Thus, depending on the disease type, we may choose an appropriate type of the omics data. With current high-throughput technologies, generally RNAs can be quantified in a relatively robust way in contrast to proteins and metabolites. Therefore, the transcriptomic data (e.g. RNA-Seq or microarray) are effective for SLE identification from the computational viewpoint, although metabolomics and proteomics data can also be used to identify the critical state in a similar way.

In summary, we proposed a novel computational method SLE score solely based on a single-sample of each individual. As exhibited above, SLE score is capable to identify the tipping point as well as the dynamic network biomarkers during disease progression for individuals. This method is of great potential in personalized pre-disease diagnosis and prevention medicine. The identification of SLE genes is also helpful in elucidating molecular mechanism of disease progression at the network level, discovering new network biomarkers, 'dark genes', drug targets and prognosis indicators.

## Funding

This work was supported by National Natural Science Foundation of China (Nos. 11771152, 11901203, 31930022 and 31771476), the National Key R&D Program of China (No. 2017YFA0505500), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB13040700), Guangdong Basic and Applied Basic Research Foundation (No. 2019B151502062), the Fundamental Research Funds for the Central Universities (No. 2019MS111), and Shanghai Municipal Science and Technology Major Project (No. 2017SHZDZX01).

*Conflict of Interest*: none declared.

## References

Achiron,A. *et al.* (2010) Microarray analysis identifies altered regulation of nuclear receptor family members in the pre-disease state of multiple sclerosis. *Neurobiol. Dis.*, **38**, 201–209.

Almansa,R. *et al.* (2017) Pulmonary transcriptomic responses indicate a dual role of inflammation in pneumonia development and viral clearance during 2009 pandemic influenza infection, *PeerJ*, **5**, e3915.

Araki,K. *et al.* (2009) Decorin suppresses bone metastasis in a breast cancer cell line. *Oncology*, **77**, 92–99.

Ashburner,M. *et al*. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet*., **25**, 25.

Bareschino,M.A. *et al*. (2011) Treatment of advanced non small cell lung cancer. *J. Thorac. Dis*., **3**, 122–133.

Bavagnoli,L. and Maga,G. (2013) Identification of host cell factors involved in influenza A virus infection. *Future Virol*., **8**, 195–208.

Bradley,L.M. *et al*. (2012) Matrix metalloprotease 9 mediates neutrophil migration into the airways in response to influenza virus-induced toll-like receptor signaling. *PLoS Pathogens*, **8**, e1002641.

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

Chen,H.W. *et al*. (2008) Curcumin inhibits lung cancer cell invasion and metastasis through the tumor suppressor HLJ1. *Cancer Res*., **68**, 7428–7438.

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

Chen,P. *et al*. (2019) Detecting early-warning signals of influenza outbreak based on dynamic network marker. *J. Cell. Mol. Med*., **23**, 395.

Chen,P. *et al*. (2017) Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. *J. Transl. Med*., **15**, 217.

Chen,P. *et al*. (2016) Detecting critical state before phase transition of complex biological systems by hidden Markov model. *Bioinformatics*, **32**, 2143–2150.

Chen,Y. *et al*. (2015) Emergent genetic oscillations in a synthetic microbial consortium. *Science*, **349**, 986–989.

Chiang,A.C. and Massagué,J. (2008) Molecular basis of metastasis. *N. Engl. J. Med. 2008*, **359**, 2814–2823.

Cooper,D.N. (2005) Somatic mutation in lung cancer. In: *The Molecular Genetics of Lung Cancer*. Springer, Berlin, Heidelberg.

Cristea,S. and Sage,J. (2016) Is the canonical RAF/MEK/ERK signaling pathway a therapeutic target in SCLC? *J. Thorac. Oncol*., **11**, 1233–1241.

Duong,T. *et al*. (2012) Genetic ablation of SOX18 function suppresses tumor lymphangiogenesis and metastasis of melanoma in mice. *Cancer Res*., **72**, 3105.

Garcia-Ojalvo,J. *et al*. (2004) Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proc. Natl. Acad. Sci. USA*, **101**, 10955–10960.

Gardner,T.S. *et al*. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, **403**, 339.

Harima,Y. *et al*. (2009) Identification of genes associated with progression and metastasis of advanced cervical cancers after radiotherapy by cDNA microarray analysis. *Int. J. Radiation Oncol. Biol. Phys*., **75**, 1232–1239.

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

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

Jamieson,A.M. *et al*. (2013) Role of tissue protection in lethal respiratory viral-bacterial coinfection. *Science*, **340**, 1230–1234.

Ji,H. *et al*. (2007) LKB1 modulates lung cancer differentiation and metastasis. *Nature*, **448**, 807–810.

Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*., **28**, 27–30.

Klein,C.A. (2008) Cancer. The metastasis cascade. *Science*, **321**, 1785–1787.

Koscielny,G. *et al*. (2017) Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res*., **45**, D985–94.

Lader,A.S. *et al*. (2004) Identification of a transcriptional profile associated with in vitro invasion in non-small cell lung cancer cell lines. *Cancer Biol. Therapy*, **3**, 624–631.

Lesterhuis,W.J. *et al*. (2017) Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. *Nat. Rev. Drug Discov*., **16**, 264.

Li,C. *et al*. (2006) Stability of genetic networks with SUM regulatory logic: lur'e system and LMI approach. *IEEE Trans. Circuits Syst. I Regular Papers*, **53**, 2451–2458.

Li,M. *et al*. (2017) Dysfunction of PLA2G6 and CYP2C44 associated network signals imminent carcinogenesis from chronic inflammation to hepatocellular carcinoma. *J. Mol. Cell Biol*., **9**, 489–503.

Liang,F. *et al*. (2016) TSC22D2 interacts with PKM2 and inhibits cell growth in colorectal cancer. *Int. J. Oncol*., **49**, 1046–1056.

Litt,B. *et al*. (2001) Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron*, **30**, 51–64.

Liu,J. *et al*. (2012) Mitochondrial proteomics of nasopharyngeal carcinoma metastasis. *BMC Medical Genomics*, **5**, 62.

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

Liu,J.K. *et al*. (2002) Pituitary apoplexy. *Semin. Neurosurg*., **12**, 315–320.

Liu,R. *et al*. (2015) Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. *Sci. Rep*., **5**, 1–13.

Liu,R. *et al*. (2018) Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. *J. Mol. Cell Biol*., doi: 10.1093/jmcb/mjy059.

Liu,R. *et al*. (2014) Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Med. Res. Rev*., **34**, 455–478.

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

Liu,X. *et al*. (2016) Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res*. **44**, e164–e164.

Liu,X. *et al*. (2019) Detection for disease tipping points by landscape dynamic network biomarkers. *Natl. Sci. Rev*. **6**, 775–785.

Marcus,A.I. and Zhou,W. (2010) LKB1 regulated pathways in lung cancer invasion and metastasis. *J. Thorac. Oncol*., **5**, 1883–1886.

Maruya,M. *et al*. (2011) Vitamin A-dependent transcriptional activation of the nuclear factor of activated T cells c1 (NFATc1) is critical for the development and survival of B1 cells. *Proc. Natl. Acad. Sci. USA*, **108**, 722–727.

McSharry,P.E. *et al*. (2003) Prediction of epileptic seizures: are nonlinear methods relevant? *Nat. Med*., **9**, 241–242.

Mojtahedi,M. *et al*. (2016) Cell fate decision as high-dimensional critical state transition. *PLoS Biol*., **14**, e2000640.

O'Brien,E.L. *et al*. (2012) Modeling synthetic gene oscillators. *Math. Biosci*., **236**, 1–15.

Paek,S.H. *et al*. (2005) Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma. *Cancer*, **104**, 580–590.

Piao,J.M. *et al*. (2011) p53 codon 72 polymorphism and the risk of lung cancer in a Korean population. *Lung Cancer*, **73**, 264–267.

Puel,A. *et al*. (2004) Inherited disorders of NF-κB-mediated immunity in man. *Curr. Opin. Immunol*., **16**, 34–41.

Richard,A. *et al*. (2016) Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol*., **14**, e1002585.

Roberto,P.B. *et al*. (2003) Transition models for change-point estimation in logistic regression. *Stat. Med*., **22**, 1141–1162.

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

Sherman,M.S. and Cohen,B.A. (2012) Thermodynamic state ensemble models of cis-regulation. *PLoS Comput. Biol*., **8**, e1002407.

Sotgia,F. *et al*. (2011) Mitochondrial oxidative stress drives tumor progression and metastasis: should we use antioxidants as a key component of cancer treatment and prevention? *BMC Med*., **9**, 62.

Steeg,P.S. *et al*. (1988) Evidence for a novel gene associated with low tumor metastatic potential. *JNCI J. Natl. Cancer Inst*., **80**, 200–204.

Szklarczyk,D. *et al*. (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*., **43**, D447–52.

Tanaka,N. *et al*. (2016) Acquired platinum resistance involves epithelial to mesenchymal transition through ubiquitin ligase FBXO32 dysregulation. *JCI Insight*, **1**, 1–15.

Tarasenko,T.N. *et al*. (2015) Kupffer cells modulate hepatic fatty acid oxidation during infection with PR8 influenza. *Biochim. Biophys. Acta (BBA) Mol. Basis Dis*., **1852**, 2391–2401.

Tobler,N.E. and Detmar,M. (2006) Tumor and lymph node lymphangiogenesis—impact on cancer metastasis. *J. Leukocyte Biol*., **80**, 691–696.

Tomida,S. *et al*. (2007) Identification of a metastasis signature and the DLX4 homeobox protein as a regulator of metastasis by combined transcriptome approach. *Oncogene*, **26**, 4600.

Tsai,M.F. *et al*. (2014) Tumour suppressor HLJ1: a potential diagnostic, preventive and therapeutic target in non-small cell lung cancer. *World J. Clin. Oncol*., **5**, 865.

Tsuge,M. *et al*. (2014) Gene expression analysis in children with complex seizures due to influenza A (H1N1) pdm09 or rotavirus gastroenteritis. *J. Neurovirol*., **20**, 73–84.

Venegas,J.G. *et al*. (2005) Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature*, **434**, 777–782.

Walker,F. *et al*. (2009) Growth factor receptor expression in anal squamous lesions: modifications associated with oncogenic human papilloma virus and human immunodeficiency virus. *Hum. Pathol*., **40**, 1517–1527.

Wen,J. *et al*. (2011) Genetic and epigenetic changes in lung carcinoma and their clinical implications. *Modern Pathol*., **24**, 932.

Xue,F. *et al*. (2005) Expression of IgSF in salivary adenoid cystic carcinoma and its relationship with invasion and metastasis. *J. Oral. Pathol. Med*., **34**, 295–297.

Yang,B. *et al*. (2018) Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nat. Commun*., **9**, 678.

Yu,X. *et al*. (2017) Individual-specific edge-network analysis for disease prediction. *Nucleic Acids Res*., **45**, e170.

Zhang,F. *et al*. (2016) Genetic variants in the acylphosphatase 2 gene and the risk of breast cancer in a Han Chinese population. *Oncotarget*, **7**, 86704.