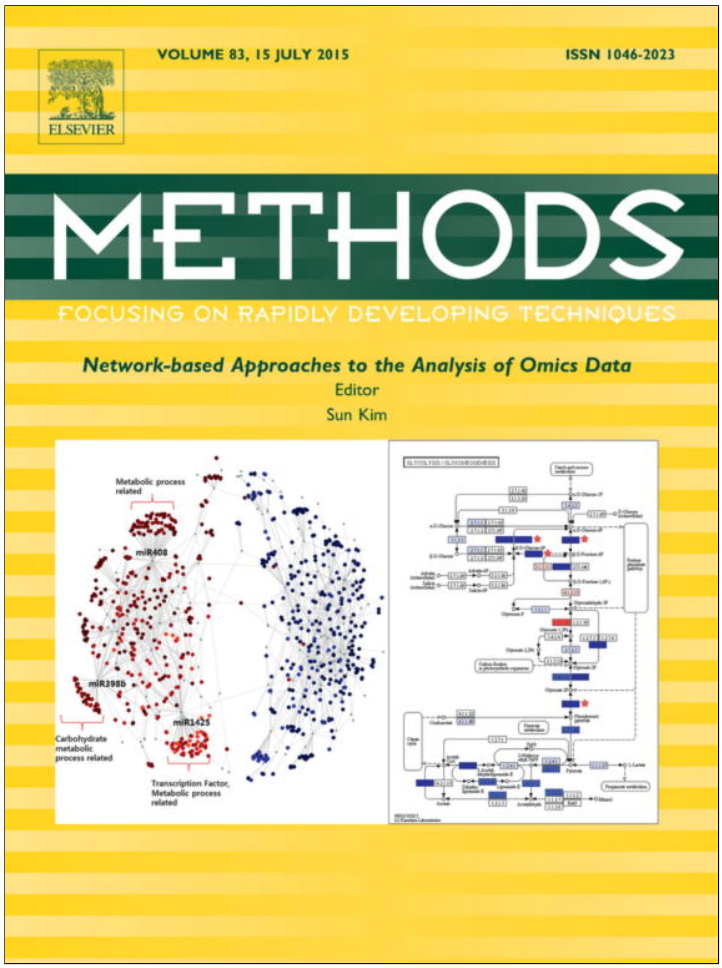

Methods 83 (2015) 36–43

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Cerebrospinal fluid protein dynamic driver network: At the crossroads of brain tumorigenesis

Zhou Tan [a,b,1], Rui Liu [b,c,1], Le Zheng [b,d,1], Shiying Hao [b,1], Changlin Fu [b,e], Zhen Li [b], Xiaohong Deng [b], Taichang Jang [b], Milton Merchant [b], John C. Whitin [b], Minyi Guo [e], Harvey J. Cohen [b,2], Lawrence Recht [b,2], Xuefeng B. Ling [b,2,*]

[a] Hangzhou Normal University, Zhejiang 311121, China
[b] Stanford University, Stanford, CA 94305, USA
[c] South China University of Technology, Guangzhou 510640, China
[d] Tsinghua University, Beijing 100084, China
[e] Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form 2 May 2015
Accepted 5 May 2015
Available online 14 May 2015

Keywords:
Transition state
Dynamical driver network (DDN)
Critical transition
Tumorigenesis progressing
Network entropy

ABSTRACT

To get a better understanding of the ongoing *in situ* environmental changes preceding the brain tumorigenesis, we assessed cerebrospinal fluid (CSF) proteome profile changes in a glioma rat model in which brain tumor invariably developed after a single *in utero* exposure to the neurocarcinogen ethylnitrosourea (ENU). Computationally, the CSF proteome profile dynamics during the tumorigenesis can be modeled as non-smooth or even abrupt state changes. Such brain tumor environment transition analysis, correlating the CSF composition changes with the development of early cellular hyperplasia, can reveal the pathogenesis process at network level during a time before the image detection of the tumors. In our controlled rat model study, matched ENU- and saline-exposed rats' CSF proteomics changes were quantified at approximately 30, 60, 90, 120, 150 days of age (P30, P60, P90, P120, P150). We applied our transition-based network entropy (TNE) method to compute the CSF proteome changes in the ENU rat model and test the hypothesis of the critical transition state prior to impending hyperplasia. Our analysis identified a dynamic driver network (DDN) of CSF proteins related with the emerging tumorigenesis progressing from the non-hyperplasia state. The DDN associated leading network CSF proteins can allow the early detection of such dynamics before the catastrophic shift to the clear clinical landmarks in gliomas. Future characterization of the critical transition state (P60) during the brain tumor progression may reveal the underlying pathophysiology to device novel therapeutics preventing tumor formation. More detailed method and information are accessible through our website at http://translationalmedicine.stanford.edu.

1. Introduction

The influence of the local environment in cancer development, clearly established in several systemic neoplasms including colon, breast and prostate cancers [1–3], remains unexplored in gliomas. An ideal approach to study the early cancer development preceding the clinical landmark of brain tumor is to analyze abnormalities in distinct time-series prior to the detection of the apparent malignancy. However, brain tumor develops with abnormal cells forming inside the brain, which significantly limits the study of its origin due to the limited access to the tissue.

Approximately 10–30% of all cerebrospinal fluid (CSF) is extrachoroidal in origin and is represented by bulk flow of the interstitial fluid from brain parenchyma into the ventricles and subarachnoid space [4–6]. With this readily accessible sample source, we previously profiled CSF proteome to survey brain environment alterations prior to impending hyperplasia by surface-enhanced laser desorption/ionization TOF mass spectrometry (SELDI-TOF-MS). SELDI-TOF-MS has been used successfully to identify biomarkers in blood from various malignancies using comparative proteomic strategies [6–8].

While there have been several clinical studies that attempted to identify biomarkers of brain tumor using comparative proteomic techniques [9–11], failure in controlling variables such as age,

* Corresponding author.
E-mail address: bxling@stanford.edu (X.B. Ling).
[1] Authors contributed equally as first authors.
[2] Authors contributed equally as last authors.

http://dx.doi.org/10.1016/j.ymeth.2015.05.004
1046-2023/

space occupying volume and tissue permeability prevented these studies from recognizing whether a changed protein expression pattern accurately represented an effect of the neoplastic process. To control these variables, we assessed changes in CSF proteome at days P30, P60, P90, P120 and P150 in a rat model, of which gliomas invariably developed after a single *in utero* exposure to the neurocarcinogen ethylnitrosourea (ENU).

Given that the rat gliomas are not generally detectable pathologically until approximately 90 days of age (P90), we hypothesized that brain tumor progression can be modeled into three states: (1) a pre-hyperplasia state with high resilience and robustness to perturbations; (2) a critical state defined as the prelude to catastrophic shift into the hyperplasia state, occurring before the imminent phase transition point is reached, and with low resilience and robustness due to its dynamical structure; (3) a hyperplasia state representing a seriously deteriorated stage possibly with high resilience and robustness, when the system usually finds it difficult to recover or return to the normal state even after intervention. This hypothesis was supported by the observations that there was usually sudden health catastrophic shift during the gradual progression of many chronic diseases [12–17]. The drastic or a qualitative transition in the focal system or network, from a normal state to a disease state, corresponds to a so-called bifurcation point in dynamical systems theory [18,19]. Various critical transition phenomena have been reported in climate and ecosystems [20]. When the system is near the critical point, there exists a dominant group which we defined as dynamic driver network (DDN) of features satisfying the following three conditions: (1) the correlation between any pair of members in DDN becomes very strong; (2) the correlation between one member of DDN and any other molecule of non-DDN becomes very weak; (3) any member of DDN becomes highly fluctuating during transition [21–23]. We previously employed transition-based network entropy (TNE) to effectively identify the DDN as well as the transition state [21]. The TNE was actually an improved Shannon entropy [24] that was conditional on the previous state of a local dynamical network in a Markov process, which was also the entropy rate of the state change in a feature space network, where each node represented a feature and each edge represented a regulatory relation between two features, with the assumption that a Markov process governed the dynamics of each node. Given a high dimensional feature network, we found that the TNE was drastically increasing when the system approached the transition state, whereas there were no significant TNE fluctuations at either normal or disease states.

In this study, we set to assess the CSF proteome profile dynamics and test our hypothesis of non-smooth or even abrupt state changes during the glioma tumorigenesis. Such brain tumor environment transition analysis, correlating the CSF composition changes with the development of early cellular hyperplasia, can reveal the pathogenesis process at network level during a time before the imaging detection of the tumors.

2. Materials and methods

In this section, we describe the experimental procedure and the theoretical basis, i.e., the TNE score, and the mathematical basis of DDN method (Fig. 1); some details are given in Supplementary information.

2.1. Data acquirement and ethics

Case (ENU) and control rat handling was in accordance with guidelines for animal safety and welfare. Rat CSF proteomics experiment was approved by the Stanford IUCAC (Protocol #11936).

2.2. ENU administration, rat CSF collection, histological analysis, and CSF proteomics

ENU rat glioma model, ENU administration, rat CSF collection and subsequent histological analysis were as previously described [6]. CSF proteomics profiling and subsequent data analysis were as previously described (Table 1) [6,25,26].

2.3. Markov process of the network evolution

The dynamics for the progression of complex diseases are very complicated either before or after sudden deterioration, and therefore the state equations are generally constructed in a high-dimensional space with a large number of variables and parameters. However, when the system is driven by a group of parameters to approach to a critical point, theoretically the system can be expressed in a very simple form, generally by one- or two-variable dynamical equations in an abstract phase space around a codimension-one bifurcation point. This is generally guaranteed by both the bifurcation theory and center manifold theory [23]. Based on this special feature during this special phase, we derived the dynamical characteristics of the network at this stage to detect the critical transition.

Specifically, we first defined the network state (or original variables) and transition state of a dynamical network in a Markov process. For an n-node network, let

$$Z(t) = (z_1(t), \ldots, z_n(t))$$

represent the network state at the sampling point t, where $z_i(t)$ denotes the expression value of node (i.e., feature i). Then, $x_i(t) \in \{0, 1\}$ is defined to measure whether or not node i has a large change at t, that is, if $|z_i(t) - z_i(t-1)|$ is sufficiently large ($\geqslant d_i$), $x_i(t) = 1$, otherwise $x_i(t) = 0$, where d_i is a positive constant threshold or the threshold. Thus, $X(t) = (x_1(t), \ldots, x_n(t))$ represents the transition state for the network at t.

Next, a local network structure centered on each node is defined, which is the basis to construct the conditional network entropy. Assume that node i has m linked first-order neighbor nodes $i_1, i_2, \ldots, i_m$, which composes a local network centered on node i with local transition state $X^i(t) = (x_i(t), x_{i_1}(t), \ldots, x_{i_m}(t))$ at t. Clearly, from the current state $X^i(t)$ at time t, there are totally 2^{m+1} possible state transitions (or possible transition states), which is denoted as $\{A_u\}_{u=1,2,\ldots,2^{m+1}}$ for this local network at the time point $t+1$ (see Fig. 2A). To simplify notation, $X^i(t)$ is denoted as $X(t)$, and transition state is denoted as state.

From the network structure, the Markov matrix $P = (p_{u,v})$ can be derived, where $p_{u,v}(t)$ describes the transition rate from state u to state v with

$$p_{u,v}(t) = \Pr(X(t+1) = A_v | X(t) = A_u), \tag{1}$$

where $u, v \in \{1, 2, \ldots, 2^{m+1}\}$ and $\sum_v p_{u,v}(t) = 1$. Thus, we have the following the stochastic Markov process for $X(t)$

$$\{X(t+i)\}_{i=0,1,\ldots} = \{X(t), X(t+1), \ldots, X(t+i), \ldots\} \tag{2}$$

with $X(t+i) = A_u, u \in \{1, 2, \ldots, 2^{m+1}\}$.

2.4. Theoretical derivation near the critical point

Consider the following discrete-time dynamical system representing dynamical evolution of a network

$$Z(t+1) = f(Z(t); P), \tag{3}$$

where $Z(t) = (z_1(t), \ldots, z_n(t))$ is an n-dimensional state vector or variable at time instant k representing feature values,

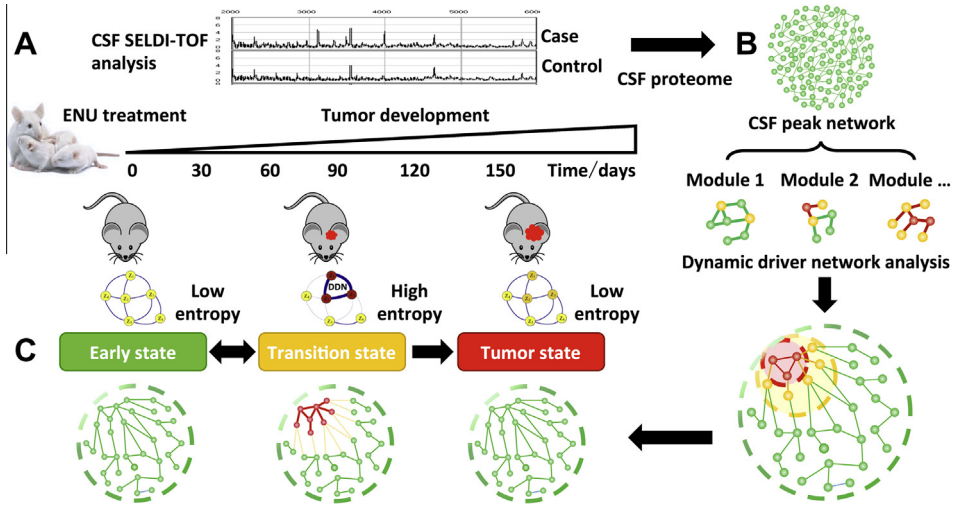

Fig. 1. Study outline. (A) Based on SELDI/TOF proteomics profiling, we studied the tumor development of rats with ENU treatment. The time-course data ranged over 5 sampling time points, i.e., 30, 60, 90, 120 and 150 days. The occurrence of hyperplastic micro tumors was at P90 as previously observed. (B) With the dynamic driver network (DDN) analysis, we localized the CSF proteome feature network and calculated the network entropy, through which the whole feature network was classified into three layers: inside, boundary and outside. (C) Based on the DDN locating in the inside layer, we identified the transition state and detected the early-warning signal of the imminent critical deterioration into hyperplasia state.

Table 1
Cohorts of ENU rats and SELDI spectra revealed CSF proteome.

Times	Sample description		
	Case (samples)	Control (samples)	Features
P30	13	11	247
P60	16	16	
P90	22	23	
P120	6	5	
P150	7	5	

$P = (p_1, \ldots, p_s)$ is a parameter vector or driving factors representing slowly changing factors. $f : R^n \times R^s \to R^n$ is a nonlinear function. Furthermore, the following conditions are assumed to be held for Eq. (3).

- $\bar{Z}$ is a fixed point of system such that $\bar{Z} = f(\bar{Z}; P)$.
- There is a value P_c such that one or a pair of the eigenvalues of the Jacobian matrix $\frac{\partial f(Z;P_c)}{\partial Z}|_{Z=\bar{Z}}$ equals to 1 in modulus.
- When $P \neq P_c$, the eigenvalues of system (3) are not always 1 in modulus.

The above three conditions with other transversal conditions imply that the system undergoes a phase change at $\bar{Z}$ or a codimension-one bifurcation when P reaches the threshold P_c. The bifurcation is generic, i.e., almost all of bifurcations for a general system satisfy these conditions.

For system (3) near $\bar{Z}$, before P reaches P_c, the system is supposed to stay at a stable fixed point $\bar{Z}$ and therefore all the eigenvalues are within $(0, 1)$ in modulus. The parameter value P_c at which the state shift of the system occurs is called a bifurcation parameter value, or a critical transition value.

Now we consider the linearized approximate equations of Eq. (3). Specifically, by introducing new variables $Y(t) = (y_1(t), \ldots, y_n(t))$ and a transformation matrix S, i.e., $Y(t) = S^{-1}(Z(t) - \bar{Z})$, we have

$$Y(t+1) = \Lambda(P)Y(t) + \zeta(t),$$

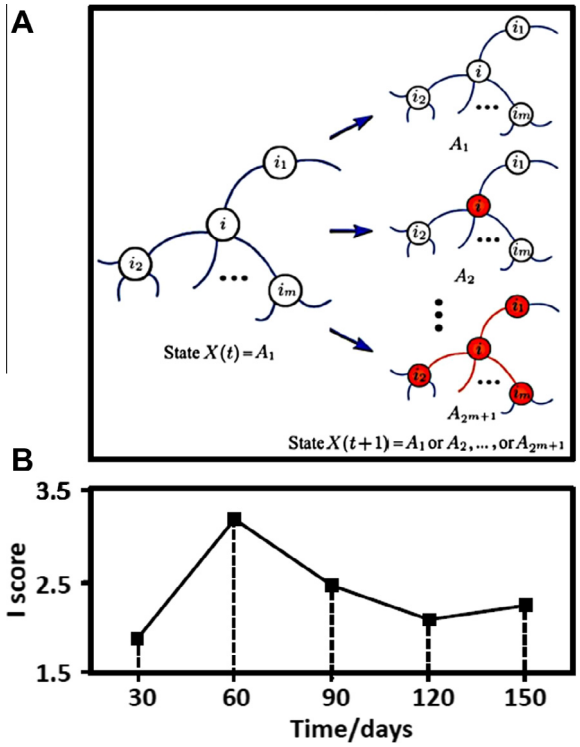

Fig. 2. (A) Local network centered on node i with m linked neighbor nodes $i_1, i_2, \ldots, i_m$. For any state at time t, there are totally 2^{m+1} possible state transitions (or possible transition states) to the state in the next time $t+1$. Such state transition process is modeled as a Markov process. (B) Based on the state transition process, we derived the transition-based network entropy (TNE). For the data of brain tumor development, the composite TNE index I increase sharply around 60 days, indicating the critical transition and reflecting the emerging hyperplasia after P60.

where $\zeta(t) = (\zeta_1(t), \ldots, \zeta_n(t))$ are small Gaussian noise with zero means. ζ_i has a small standard deviation σ_i for all k.

Without loss of generality, the diagonalized matrix $\Lambda = (\lambda_1, \ldots, \lambda_n)$ is assumed to has each λ_i between 0 and 1. Among the eigenvalues of Λ, the largest one (in modulus), say λ_1,

first approaches to 1 in modulus when parameter transition $P \to P_c$ occurs. The eigenvalue λ_1 characterizes the system's rate of change around the fixed point and is called the dominant eigenvalue. The early state corresponds to the period with $|\lambda_1| < 1$, whereas the transition stage corresponds to the period with $\lambda_1 \to 1$. Without the loss of generality, the first variable y_1 in Y is assumed to be associated with λ_1. We have proven that there exists a dominant group (or module) or a dynamical driver network (DDN) near a fixed point, which satisfies some generic conditions simultaneously (including high fluctuation, strong correlation within DDN, and the weak correlation between DDN-members and other nodes) when the system approaches a critical transition point [23].

Unlike the analysis on the original variables Z in [23], here we focused on the variation equation of Eq. (3) with variation variable set ΔZ.

Noting

$$z_i = s_{i1}y_1 + \cdots + s_{in}y_n + \bar{z}_i, \tag{4}$$

let the variation variable

$$\Delta Z = Z(t) - Z(t-1),$$

then from Eq. (4) we have

$$\Delta z_i = s_{ij}\Delta y_1 + \cdots + s_{in}\Delta y_n$$

where

$$\Delta Y = Y(t) - Y(t-1).$$

We call $\Delta z_i(t)$ and $\Delta y_i(t)$ as the variation variables for $z_i(t)$ and $y_i(t)$, respectively.

Obviously, it holds that

$$\Delta Y(t+1) = \Delta Y(t) + \xi(t),$$

where $\xi(t) = \zeta(t) - \zeta(t-1)$ is Gaussian noise with zero mean and covariance $\kappa_{ij} = Cov(\xi_i, \xi_j)$. It is clear that the standard deviation of $\xi_i(t)$ is $\sqrt{2}\sigma_i$, where σ_i is the standard deviation of ζ for all t. Obviously, variable Δy_1 corresponds to the dominant eigenvalue λ_1.

For any integer $T > 0$, by iteration we have

$$\begin{aligned}\Delta Y(t+T) = {} & \Lambda^T \Delta Y(t) + \Lambda^{T-1}\xi(t) + \Lambda^{T-2}\xi(t+1) + \cdots \\ & + \Lambda\xi(t+T-2) + \xi(t+T-1)\end{aligned}$$

Clearly, the summation of the coefficients for the covariance matrices for T Gaussian noise, is

$$(I - \Lambda^T)(I - \Lambda)^{-1}$$

where I is the n-dimensional identity matrix.

Note that when the system is in an early state, $\lambda_i < 1$. Hence as $T \to +\infty$ it holds

$$\Delta Y(t+T) = \varepsilon(t) \tag{5}$$

where $\varepsilon(t) = (\varepsilon_1(t), \ldots, \varepsilon_n(t))$ are small Gaussian noise with zero means. Based on the Law of Large Numbers, the deviation of is bounded when $\lambda_i < 1$.

Back to the original variable Z, it can be derived that

$$\Delta z_i(t+T) = s_{i1}\Delta y_1(t+T) + \cdots + s_{in}\Delta y_n(t+T) \tag{6}$$

Therefore, when the system is in an early state, or equivalently $|\lambda_i| < 1$, any variation variable $\Delta z_i(t+T)$ is statistically independent of its initial variable $\Delta z_i(t)$ for a sufficiently long T, because the biochemical reactions occur in a very short time interval (e.g., less than micro-seconds). In other words, any two samples can be considered to have a long T due to a large number of biochemical reactions during the intervals of their observations, and therefore, variation variables for any two samples are statistically independent of each other when the system is in the early state.

In the following section, the case near the critical transition is discussed when the dominant eigenvalue $\lambda_1 \to 1$ (for $\lambda_1 \to -1$, the derivation is similar and thus is omitted).

Notice that the variation variable y_1 is related to the dominant eigenvalue λ_1. So

$$y_1(t+T) = \lambda_1 y_1(t+T-1) + \zeta_1(t+T-1)$$

holds for any integer T. Then we have

$$\begin{aligned}&\Delta y_1(t+T) + \Delta y_1(t+T-1) + \cdots + \Delta y_1(t+1) \\ &\quad = \lambda_1(\Delta y_1(t+T-1) + \cdots + \Delta y_1(t)) + (\zeta_1(t+T-1)) - \zeta_1(t-1).\end{aligned}$$

Therefore,

$$\begin{aligned}\Delta y_1(t+T) = {} & (\lambda_1 - 1)\Delta y_1(t+T-1) + \cdots + (\lambda_1 - 1)\Delta y_1(t+1) \\ & + \lambda_1 \Delta y_1(t) + (\zeta_1(t+T-1) - \zeta_1(t-1))\end{aligned}$$

Hence when $\lambda_1 \to 1$ we have

$$\Delta y_1(t+T) = \Delta y_1(t) + (\zeta_1(t+T-1) - \zeta_1(t-1))$$

which means that $\Delta y_1(t+T)$ strongly depends on $\Delta y_1(t)$ for small noise level. In this way, the dominant variable $\Delta y_1(t)$ also strongly depends on its previous state when P is near P_c. It is obviously that the same result holds when $\lambda_1 \to -1$.

On the other hand, because $|\lambda_i| < |\lambda_1|$, $i = 2,3,\ldots,n$, other variables $\Delta y_i(t+T)$ satisfy Eq. (5), that is,

$$\Delta y_i(t+T) = \varepsilon_i(t), \quad i = 2, 3, \ldots, n.$$

Notice that the variable $\Delta y_1(t)$ is related to the dominant eigenvalue λ_1.

There is a special group of variables z_j, whose variables Δz_j are related to Δy_1, i.e., the Δz_j in Eq. (6) with $s_{j1} \neq 0$, called a dominant group. Such variables z_j are called the dominant group members, or dynamical driver network (DDN) members [22].

For any two DDN members z_j and z_i with $s_{j1} \neq 0$ and $s_{i1} \neq 0$ in Eq. (6), when $|\lambda_1| \to 1$, we have

$$\Delta z_j(t+T) = s_{j1}\Delta y_1(t+T) + \cdots + s_{jn}\Delta y_n(t+T) = \frac{s_{j1}}{s_{i1}}\Delta z_i(t) + \rho_j(t)$$

where

$$\begin{aligned}\rho_j(t) = s_{j1}\Bigg(&\zeta_1(t+T-1) - \zeta_1(t-1) + \frac{s_{j1}}{s_{i1}}(s_{j2} - s_{i2})\varepsilon_2(t) \\ &+ \cdots + \frac{s_{j1}}{s_{i1}}(s_{jn} - s_{in})\varepsilon_n(t)\Bigg)\end{aligned}$$

represents Gaussian noise, which is assumed to be small. It is clear that when $|\lambda_1| \to 1$, for any two DDN members, the variable $\Delta z_j(t+T)$ is correlated to $\Delta z_i(t)$. It also holds for $i = j$, i.e., for any DDN member, the variable $\Delta z_j(t+T)$ is correlated to its previous state $\Delta z_j(t)$. On the other hand, as indicated by Eq. (3), for any non-DDN member z_k, $\Delta z_k(t+T)$ is statistically independent of $\Delta z_k(t)$.

2.5. 2.5.Dynamical increase of network entropy

For a local structure centered on node i with its m linked first-order neighbor nodes $i_1, i_2, \ldots, i_m$, we already know that its state transition process is a stochastic Markov process given as in (2). Within a period or phase, assume that there is no change on the transition matrix, i.e., the transition probabilities $p_{u,v}(t)$ in (1) between any two possible states A_u and A_v are invariant. Thus, the process $\{X(t)\}_{t\in[t_1,t_2]}$ is a stationary stochastic Markov process during a specific period, e.g., the early stage or the transition stage.

Hence, there is a stationary distribution $\pi = (\pi_1, \ldots, \pi_{2^{m+1}})$ satisfying $\sum_v \pi_v p_{u,v} = \pi_u$. Then, we can define a transition-based network entropy (TNE) as

$$H_i(t) = H(\chi) = -\sum_{u,v} \pi_v p_{u,v} \log p_{u,v} \quad (7)$$

where the subscript index i indicates the center node i of this local network, and χ represents the state-transition process $X(t), X(t+1), \ldots, X(t+T), \ldots$ of the local network.

We combine the TNEs for all nodes and define the average network entropy for the whole network with n nodes as the average TNE as follows:

$$H(t) = \frac{1}{n}\sum_{i=1}^{n} H_i(t) \quad (8)$$

Suppose that there are control samples and case samples, then we define the comparative entropy as:

$$I(t) = \frac{H_{\text{control}}(t)}{H_{\text{case}}(t)} \quad (9)$$

where $H_{\text{control}}(t)$ is the TNE based on control samples in the form of Eq. (8), and $H_{\text{case}}(t)$ is the TNE based on case samples in the form of Eq. (8).

Note that we defined the dominant group, or the DDN, as a group of nodes that collectively makes the first move toward the disease state, thereby indicating a sudden deterioration. Then, the nodes in the network can be categorized into three groups according to the local structure of the DDN or the leading network:

- *Type 1:* DDN feature is a DDN node, i.e., if node i belongs to DDN, then i is a Type 1 node.
- *Type 2:* A 1st-downstream feature is a node that is linked with at least one DDN node, i.e., if node i is a non-DDN node and some of its linked neighbors are DDN nodes, then i is a Type 2 node.
- *Type 3:* A 2nd-downstream feature is a non-DDN node that has no links with DDN nodes, i.e., if node i is a non-DDN node, and its linked neighbors $i_1, i_2, \ldots, i_m$ are all non-DDN members, then i is a Type 3 node.

Next, Table 2 shows that the comparative TNE in Eq. (9) based on case samples has the following generic properties in terms of its dynamics, which correspond to the three types of nodes described above when the system is near a critical transition:

Therefore, based on these three cases, we concluded that the average TNE (Eq. (8)) increases drastically as the system approaches a critical transition point. This critical phenomenon coincides with the physical meaning of the conditional entropy in terms of the low resilience or robustness in a transition state, i.e., the next state transition depends largely on its current state, which implies a low conditional entropy due to this dependence near the transition state. We adopted an efficient strategy for estimating the comparative entropy (Eq. (9)) using only those nodes with increasing TNEs, rather than all of the nodes in the network. Using such a scheme, the comparative entropy (Eq. (9)) is more sensitive to a transition state.

The derivation of the TNE is based on the properties of the DDN, while the TNE provides more information for studying the DDN. There are several advantages in using the TNE to detect a critical transition. First, it is easier to detect DDN nodes directly, rather than a dominant group in the overall network, which is generally a difficult task because of the scale of datasets. Second, we only need to focus on the local structure of a network node by node, which significantly reduces the computational complexity. Therefore, predictions based on the TNE method are DDN-free.

It is reasonable to ignore any nodes with increasing TNEs because this may be caused by noise or data errors. In a transition state, the remaining nodes with decreasing TNEs form a sub-network, which is used as the index or criterion for detecting a critical transition and the leading network in the transition to a disease state. Our analysis is based on the local structure, i.e., the targeted node and its linked neighbors, and the final sub-network encapsulates the most critical information allowing the determination of the dynamics of the overall network. Therefore, our method simplifies the computation and avoids the effect of noise, providing a more accurate and reliable early warning signal. To summarize the above theoretical methods, we have the following statement: (1) Decrease of the average TNE is the early warning signal for detecting a critical transition. (2) Criterion for DDN can further identify features of the leading network.

Table 2
Node type and descriptions.

Type	Node	TNE for a local network
1	DDN feature	Increase drastically
2	1st-downstream feature	Increase
3	2nd-downstream feature	No significant change

2.6. A computational pipeline for CSF protein DDN analysis

In order to make our DDN analysis more clearly, our computational process was summarized as follows.

(1) CSF proteomics data analysis and subsequent data cleaning were as previously described [6,25,26]. All relevant raw data can be found on our website: http://translationalmedicine.stanford.edu.
(2) The DDN features were selected according to the theoretical methods described above and in Refs. [22,23].
(3) Determine the critical transition time point by using TNE and criterion described above and in Refs. [21,22]. These networks were then visualized by using Cytoscape software version 3.2 (http://www.cytoscape.org).
(4) Functional analysis of the DDN features was conducted by searching PubMed http://pubmed.com.

3. Results and discussion

3.1. Identify the transition state

Based on comparative TNE, we selected 35 features out of 247 mass spectrometric spectral features to identify the transition state of brain tumor. Specifically, the sharp increase of the TNE index (I score) (Fig. 2B) can be used as an early-warning signal for the imminent critical transition, which commits the brain tumor development. The selected 35 features are listed in Table 3.

Unlike the traditional molecular biomarkers whose differential expressions reflect the presence or severity of the disease state, the DDN is a strongly correlated feature network where the values of features dynamically change in the transition state as shown in Fig. 1. The system tends to present increasingly instability approaching the transition state, where the DDN features collectively fluctuate. These DDN characteristics at the transition state can be exploited to detect the early warning signal of a complex disease in the early stage, which however is not possible using traditional biomarkers or their associated methods. Hence, the detection of the DDN implies that a particular subject system is in the transition state with high-level system entropy.

3.2. Development of brain tumors in progeny of ENU-exposed rats

ENU exposed rats ($n = 64$) (13 from P30, 16 from P60, 22 from P90, 6 from P120 and 7 from P150) were examined histologically

Table 3
Selected DDN mass spectrometric spectral features ($n = 35$) and descriptions.

m/z	Laser energy	Note
13913	Med	Glutathionylated transthyretin
3487	Low	α1-Macroglobulin fragments (1212–1243)
3544	Low	Neuroendocrine protein 7B2 c-terminal peptide
14120	Med	Sinapinic acid adduct of glutathionylated transthyretin
22893	High	Prostaglandin D2 synthase
6953	Low	$Z = 2$ for glutathionylated transthyretin
5704	Med	
66110	High	Albumin
5789	Med	
14285	Med	
4196	Low	
6795	Med	$Z = 2$ for transthyretin
5788	Low	
15847	Med	Hemoglobin subunit beta
3641	Low	
7515	Med	
5375	Med	
6909	Med	$Z = 2$ for glu-cys-transthyretin
3992	Low	
11848	Med	
4362	Low	
8913	Med	
11861	Med	
3941	Low	
3700	Low	
4891	Low	
7061	Low	
3504	Low	
4161	Low	
5818	Low	
8569	Med	
7444	Med	
12772	Med	
22034	High	$Z = 3$ for albumin (66110)
4734	Low	

for the presence of nestin+ and OPN+ precursor lesions (nests) as well as the appearance of tumors, which was detected by MRI after day 90. Consistent with previous reports [27,28], single or multiple nestin+ precursor cell cyst were noted in all rats by P30 (100%). In contrast, microtumors were not noted in any rats sacrificed at P30, only 4 rats (25%) at P60, while ~60% of rats at P90, and 100% of rats at P120–150 (Fig. 3B, left). The rat death was firstly found after day 120, which is showed by the surviving curve in Fig. 3B, right. No macroscopic tumors were found in any animals at the time points examined.

3.3. Application of transition-based network entropy (TNE) method to identify dynamic drive network and critical transition state before hyperplasia

CSF was collected from a total of 64 ENU and 60 saline exposed rats and mass spectra of CSF applied to CM10 ProteinChip arrays were collected for the five postpartum ages (P30, P60, P90, P120 and P150) as described in Section 2 (Table 1). The relative intensities of peaks were different in the CSF of rats obtained at these five ages. For this reason we grouped the spectra by postpartum age for baseline correction, noise reduction and intensity normalization. The spectra for all five ages was then grouped together for peak findings, and then separated again for peak analysis at each age. Our DDN method was applied to analyze the case and control mass spectrometry profiles, which allowed the identification of early-warning CSF proteome DDN components. The DDN's transition-based network entropy was proposed as a general early-warning indicator for the transition to hyperplasia, which appeared to be related to the tumor initiation related CSF proteome changes and progression. It can provide better insight into the pathophysiology and give clues to the tumor environment impact. Based on the state transition process, we derived the transition-based network entropy (TNE) analyzing the CSF proteomes. As shown in Fig. 2B, the composite TNE index I increased sharply around 60 days, indicating a critical transition into hyperplasia during glioma development after P60.

3.4. DDN biomarkers vs. traditional biomarkers

Traditional disease diagnostic biomarkers differentiate disease and normal states, whose expressions reflect the presence or severity of the disease state and are required to have statistically distinct quantifications between case and control subjects. Our novel disease transition associated DDN network feature markers are different from these diagnostic biomarkers. Specifically, we first defined the TNE, and proved that TNE can serve as a general early-warning indicator of any imminent transitions. The DDN features collectively form a strongly correlated feature network where the values of features fluctuate in a coordinated manner. When the system approaches the transition state, the DDN features reveal a unique kinetic pattern of pronounced fluctuation. Furthermore, the identification of a dynamic driver network in this study validates our initial hypothesis of the disease transition state when the system switches from normal to disease state.

With the CSF proteomics survey of the ENU model rats, we constructed CSF protein networks to gauge the physiological and pathological status of the cerebral compartment at different days of age occurring with the gradual appearance of cellular hyperplasia. We employed our previously developed transition-based network entropy (TNE) [23] and identified the drastic or a qualitative transition at P60 in the CSF proteome network before hyperplasia, which corresponds to a so-called bifurcation point in dynamical systems theory [29]. When the ENU rats were at P60 and CSF proteome network was near the critical point, we found that a dominant group of 35 CSF proteins which we defined as the dynamic driver network (DDN) of CNS protein features collectively increased the TNE that is conditional on the previous state of a local dynamical network in a Markov process, whereas there were no significant TNE fluctuations before and after P60.

3.5. Identification of DDN biomarkers

Our current effort in up-to-date proteomics identified 35 of the DDN CSF proteins, with 5 transthyretin species of different post-translational modifications. Consistent with our previous observation and other reports, CSF transthyretin protein species were shown to differentially express in our ENU rat model [6] and human brain tumor [7,8,30]. In this regard, CSF transthyretin is a biomarker, not only differentiates between case and control, but also functions as a DDN component with sharp TNE increase at rat age P60. Our previous results indicated that, between case and control groups, total transthyretin levels did not differ while there were significant differences of posttranslational modifications. It is possible that variation of different translational modifications may disrupt transthyretin's normal functions in the transport of both thyroxine and reinol, which may drive the critical transition of cellular hyperplasia after P60 during tumorigenesis. It seems unlikely that the fluctuation in CSF transthyretin levels before hyperplasia in this study represented release from these small nests and microtumors as variation diminished prior to and post the critical transition at P60.

Another two CSF DDN proteins identified in our study were albumin and prostaglandin D2 synthase (PGD2S). Both proteins are abundant in the CSF [31–34] and considered as indicative of

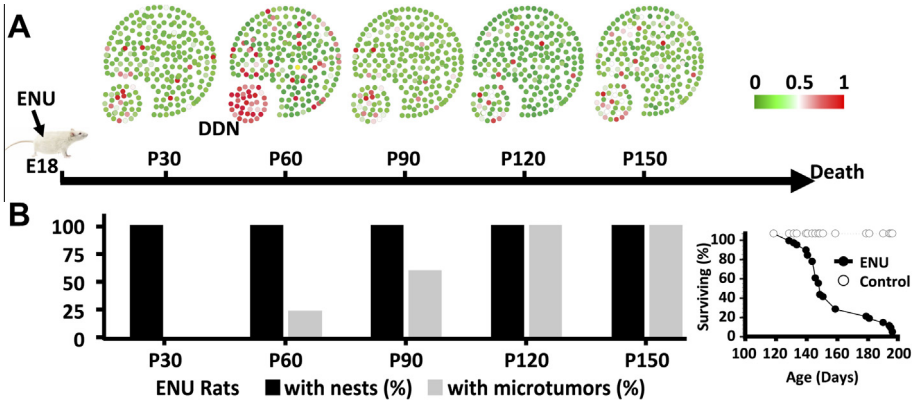

Fig. 3. Time course analysis of brain tumor development revealing the critical transition state. (A) The dynamical change is illustrated for the whole feature network, with DDN located in the lower left, where the color of the nodes represents the fluctuation strength of DDN features. DDN analysis showed a significant change at P60, which illustrates the imminent critical transition. The early-warning signal detected by DDN preceded the detection of hyperplastic microtumors observed in 90 days. (B) Left, histogram of percentage of ENU-exposed rats with nests or microtumors. Right, surviving curve of ENU-exposed rats after day 120.

absence of hyperplasia before P60, therefore, the differential expression between case and control or variations observed before P60 reflected either albumin release from tumor cells or the impact of a space occupying lesion before apparent imaging changes. Our previous hypothesis [6] of these proteins' differential CSF abundance was the disruption of the blood brain barrier during tumorigenesis.

From the network level of an ensemble of 35 biomarkers, in the critical transition state of brain tumor, the DDN nodes represent distinct biomarkers with computationally derived relationships. Our DDN-driven analysis facilitates data integration across multiple levels of biological complexity and may define contribution of specific biomarkers to systems-wide properties of brain tumorigenesis.

4. Conclusions

Our DDN discovery at P60 findings and the DDN CSF protein identification results were consistent with the hypothesis that a CSF environmental change was initiated before the hyperplasia / micro tumor stage (before P60), similar to what has been reported in systemic cancers including breast and prostate lesions [3,35]. 11/35 DDN CSF proteins were identified and 24 CSF protein identities remained to be determined. Upon the completion of all DDN CSF protein identifications, we will be in a much better position to explore CSF environmental changes committing the hyperplasia development path. Nevertheless, our dynamic network analysis suggests, in regard to tumorigenesis, to focus at P60 of the rat glioma model to probe the *in situ* environment changes preceding the development of hyperplasia abnormalities. This may lead to not only insights of host tumor environment interactions, but also the development of an effective time window for novel therapeutic strategies in primary brain tumor.

Acknowledgments

This work was supported by Stanford University Spark Program (2013–2014) to X.B.L. and H.J.C., and American Heart Association (14GRNT20510026) to X.B.L. and H.J.C. This work was supported by the National Natural Science Foundation of China (NSFC) to Z.T. (2012–2015, 31201697) and R.L. (9143920024 and 11401222) by Fundamental Research Funds for the Central Universities (Grant Number 2014ZZ0064).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2015.05.004.

References

[1] A. Joshi, D. Cao, Front. Bio. 15 (2010) 180–194.
[2] A.Y. Liu, L.E. Pascal, R.Z. Vencio, E.F. Vencio, Cancer Biomarkers: Section A Dis. Markers 9 (2010) 141–155.
[3] L.C. van Kempen, K.E. de Visser, L.M. Coussens, Eur. J. Cancer 42 (2006) 728–734.
[4] H.F. Cserr, Ann. N. Y. Acad. Sci. 529 (1988) 9–20.
[5] J. Skipor, J.C. Thiery, Acta Neurobiol. Exp. 68 (2008) 414–428.
[6] J.C. Whitin, T. Jang, M. Merchant, T.T. Yu, K. Lau, B. Recht, H.J. Cohen, L. Recht, PLoS One 7 (2012) e49724.
[7] E.T. Fung, T.T. Yip, L. Lomas, Z. Wang, C. Yip, X.Y. Meng, S. Lin, F. Zhang, Z. Zhang, D.W. Chan, S.R. Weinberger, J. Int. Cancer 115 (2005) 783–789.
[8] Z. Zhang, R.C. Bast Jr., Y. Yu, J. Li, L.J. Sokoll, A.J. Rai, J.M. Rosenzweig, B. Cameron, Y.Y. Wang, X.Y. Meng, A. Berchuck, C. Van Haaften-Day, N.F. Hacker, H.W. de Bruijn, A.G. van der Zee, I.J. Jacobs, E.T. Fung, D.W. Chan, Cancer Res. 64 (2004) 5882–5890.
[9] A. Elstner, F. Stockhammer, T.N. Nguyen-Dobinsky, Q.L. Nguyen, I. Pilgermann, A. Gill, A. Guhr, T. Zhang, K. von Eckardstein, T. Picht, J. Veelken, R.L. Martuza, A. von Deimling, A. Kurtz, J. Neurooncol. 102 (2011) 71–80.
[10] A.M. Saratsis, S. Yadavilli, S. Magge, B.R. Rood, J. Perez, D.A. Hill, E. Hwang, L. Kilburn, R.J. Packer, J. Nazarian, Neuro-oncology 14 (2012) 547–560.
[11] H.F. Juan, J.H. Chen, W.T. Hsu, S.C. Huang, S.T. Chen, J. Yi-Chung Lin, Y.W. Chang, C.Y. Chiang, L.L. Wen, D.C. Chan, Y.C. Liu, Y.J. Chen, Proteomics 4 (2004) 2766–2775.
[12] J.G. Venegas, T. Winkler, G. Musch, M.F. Vidal, Nature 434 (2005) 777–782.
[13] S.H. Paek, H.T. Chung, S.S. Jeong, C.K. Park, C.Y. Kim, J.E. Kim, D.G. Kim, H.W. Jung, Cancer 104 (2005) 580–590.
[14] P.E. McSharry, L.A. Smith, L. Tarassenko, Nat. Med. 9 (2003) 241–242 (author reply 242).
[15] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter, G. Vachtsevanos, Neuron 30 (2001) 51–64.
[16] Y. Hirata, N. Bruchovsky, K. Aihara, J. Theor. Biol. 264 (2010) 517–527.
[17] D. He, Z.P. Liu, M. Honda, S. Kaneko, L. Chen, J. Mol. Cell Biol. 4 (2012) 140–152.
[18] R. Gilmore, Catastrophe theory for scientists and engineers, Dover Publication, New York, 1993.
[19] J.D. Murray, Mathematical biology, 3rd ed., Springer, New York, 2002.
[20] M. Scheffer, J. Bascompte, W.A. Brock, V. Brovkin, S.R. Carpenter, V. Dakos, H. Held, E.H. van Nes, M. Rietkerk, G. Sugihara, Nature 461 (2009) 53–59.
[21] R. Liu, M. Li, Z.P. Liu, J. Wu, L. Chen, K. Aihara, Sci. Rep. 2 (2012) 813.
[22] R. Liu, X. Wang, K. Aihara, L. Chen, Med. Res. Rev. 34 (2014) 455–478.
[23] L. Chen, R. Liu, Z.P. Liu, M. Li, K. Aihara, Sci. Rep. 2 (2012) 342.
[24] C.E. Shannon, Bell Syst. Techn. J. 27 (1948) 379–423.
[25] J. Ji, J. Ling, H. Jiang, Q. Wen, J.C. Whitin, L. Tian, H.J. Cohen, X.B. Ling, BMC Res. Notes 6 (2013) 109.
[26] S.M. Carlson, A. Najmi, J.C. Whitin, H.J. Cohen, Proteomics 5 (2005) 2778–2788.
[27] T. Jang, N.S. Litofsky, T.W. Smith, A.H. Ross, L.D. Recht, Neurobiol. Dis. 15 (2004) 544–552.

[28] T. Jang, B. Sathy, Y.H. Hsu, M. Merchant, B. Recht, C. Chang, L. Recht, J. Neurosurg. 108 (2008) 782–790.
[29] S.J. Lade, T. Gross, PLoS Comput. Biol. 8 (2012) e1002360.
[30] M.U. Schuhmann, H.D. Zucht, R. Nassimi, G. Heine, C.G. Schneekloth, H.J. Stuerenburg, H. Selle, Surg. Oncol. 36 (2010) 201–207.
[31] C. Thomsen, F. Stahlberg, M. Stubgaard, B. Nordell, Radiology 177 (1990) 659–665.
[32] J.S. Smith, T.E. Angel, C. Chavkin, D.J. Orton, R.J. Moore, R.D. Smith, Proteomics 14 (2014) 1102–1106.
[33] N. Samuel, M. Remke, J.T. Rutka, B. Raught, D. Malkin, J. Neurooncol. 118 (2014) 225–238.
[34] D.N. Melegos, M.S. Freedman, E.P. Diamandis, Prostaglandins 54 (1997) 463–474.
[35] M. Mego, M. Karaba, G. Minarik, J. Benca, T. Sedlackova, L. Tothova, B. Vlkova, Z. Cierna, P. Janega, J. Luha, P. Gronesova, D. Pindak, I. Fridrichova, P. Celec, J.M. Reuben, M. Cristofanilli, J. Mardiak, Breast J. (2015).