

# CSF protein dynamic driver network: at the crossroads of brain tumorigenesis

Changlin Fu<sup>1,2,#</sup>, Zhou Tan<sup>2,#</sup>, Rui Liu<sup>2,3,#</sup>, Shiyong Hao<sup>2,#</sup>, Zhen Li<sup>2</sup>, Pei Chen<sup>2</sup>, Taichang Jang<sup>2</sup>,

Milton Merchant<sup>2</sup>, John C. Whitin<sup>2</sup>, Oxford Wang<sup>2</sup>, Minyi Guo<sup>1</sup>, Harvey J. Cohen<sup>2,\*</sup>, Lawrence Recht<sup>2,\*</sup> and Xuefeng B. Ling<sup>2,\*</sup>

1. Shanghai Jiao Tong University, Shanghai, China, 200240

2. Stanford University, Stanford, CA, USA, 94305

3. South China University of Technology, Guangzhou, China, 510640

#. Authors contributed equally as first authors.

\*. Authors contributed equally as last authors. Correspondence: bxling@stanford.edu (XBL)

**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 develop 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 this 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. An improved understanding of the critical transition state (P60) during the brain tumor progression can provide the scientific groundwork to device novel therapeutics preventing tumor formation.

**Keywords**— transition state; dynamical driver biomarker (DDN); critical transition; tumorigenesis progressing; network entropy;

## I. INTRODUCTION

The influence of the local environment, clearly established in the development of several systemic neoplasms including colon, breast and prostate cancers, remains unexplored in gliomas. An ideal approach to study the early relationships 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 form inside the brain which significantly limits

the study of its origin due to the relative inaccessibility of the tissue.

Approximately 10-30% of all CSF is extrachoroidal in origin and is represented by bulk flow of the interstitial fluid from brain parenchyma into the ventricles and subarachnoid space [1]. With this readily accessible sample source, we previously profiled cerebrospinal fluid (CSF) proteome to survey brain environment alterations prior to impending hyperplasia by surface-enhanced laser desorption/ionization TOF mass spectrometry (SELDI TOF MS). Surface-enhanced laser desorption/ionization TOF mass spectrometry (SELDI TOF MS) has been used successfully to identify biomarkers in blood from various malignancies using comparative proteomic strategies [1].

While there have been several clinical studies that attempted to identify biomarkers of brain tumor using comparative proteomic techniques, they all suffer from an inability to control such factors as age, space occupying volume and tissue permeability, thus obscuring whether a changed protein expression pattern accurately represents an effect of the neoplastic process. To control for these variables, we assessed changes in CSF proteome at days P30, P60, P90, P120 and P150 in a rat model, of which gliomas invariably develop 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, therefore, 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 treatment. This is supported by the observations that there is usually sudden health catastrophic shift during the gradual progression of many chronic diseases [2-7]. 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

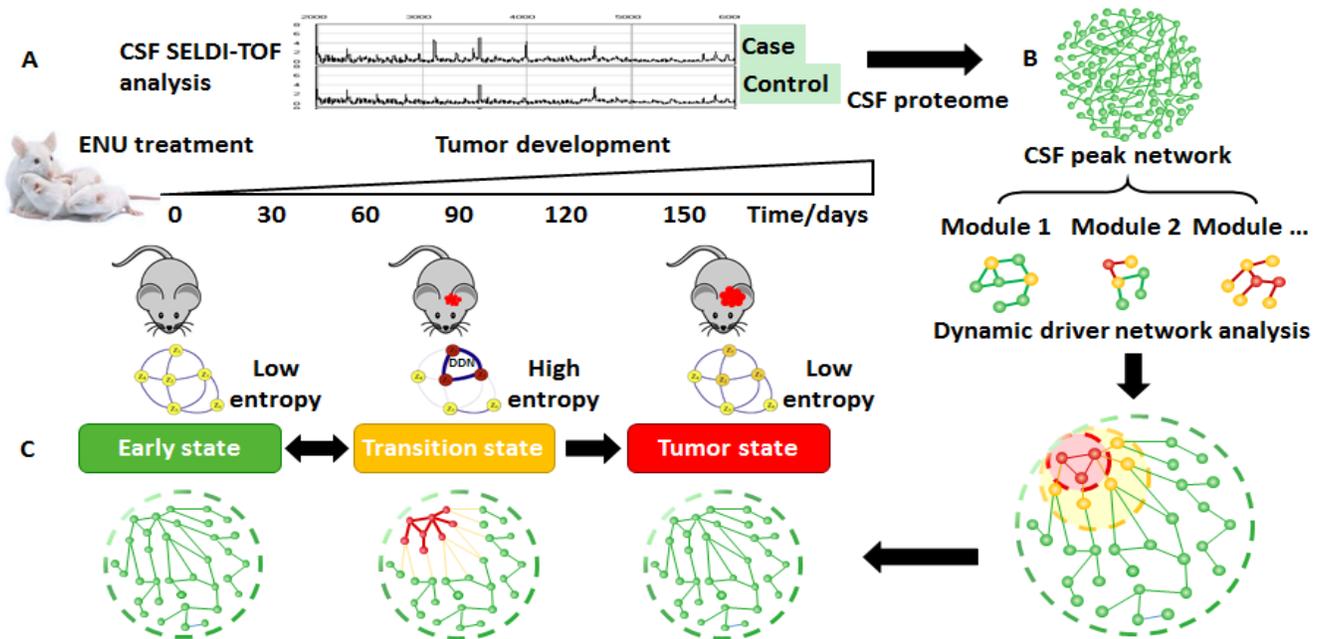


Figure 1. The sketch of study design. (A). Based on SELDI/TOF proteomics profiling, we studied the tumor development of rats with ENU treatment. The time-course data ranged over 5 sampling time points, i.e., 30, 60, 90, 120 and 150 days. The occurrence of hyperplastic micro tumors is at P90 as previously observed. (B). With the dynamic driver network (DDN) analysis, we localize the CSF proteome feature network and calculate the network entropy, through which the whole feature network can be classified into three layers. (C) Based on the DDN, which locates in the inside layer, we can identify the transition state and detect the early-warning signal of the imminent critical deterioration into hyperplasia state.

in dynamical systems theory [8-9]. 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: The correlation between any pair of members in DDN becomes very strong; The correlation between one member of DDN and any other molecule of non-DDN becomes very weak; Any member of DDN becomes highly fluctuating during transition [10-12]. We previously employed transition-based network entropy (TNE) to effectively identify the DDN as well as the transition state [11]. The TNE is actually an improved Shannon entropy [13] that is conditional on the previous state of a local dynamical network in a Markov process, which is also the entropy rate of the state change in a feature space network, where each node represents a feature and each edge represents a regulatory relation between two features, with the assumption that a Markov process governs the dynamics of each node. Given a high dimensional feature network, we find that the TNE is drastically increasing when the system approaches the transition state, whereas there are 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 image detection of the tumors.

## II. METHODS

### A. Data acquisition and Ethics

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

### B. 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 [1]. CSF proteomics profiling and subsequent data analysis were as previously described [1, 14-15].

TABLE I. SAMPLE DESCRIPTION

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

### C. Markov process of the network evolution

In this section, we study the qualitative behaviors in dynamics of the nodes to characterize the critical transition by an  $n$ -node network, which is used to describe the regulation relationship among features. Generally, the dynamics for the progression of complex diseases is 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 some parameters to approach 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 the bifurcation theory and center manifold theory. Just because of this special feature, during this special phase, we can derive the dynamical characteristics of the network at this stage to detect the critical transition.

Specifically, we first define the network state (or original variables) and transition state of a dynamical network in a Markov process.

For a  $n$ -node network, let  $Z(t) = (z_1(t), \dots, z_n(t))$  represent the network state at  $t$ , 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 sampling point  $t$ , that is, if  $|z_i(t) - z_i(t-1)|$  is a sufficiently large ( $\geq d_i$ ), then  $x_i(t) = 1$ , otherwise  $x_i(t) = 0$ , where  $d_i$  is a constant threshold. Thus,  $X(t) = (x_1(t), \dots, x_n(t))$  is the transition state for the network at  $t$ .

Next, we define a local network structure centered on each node, which is the basis to construct a conditional network entropy. Assume that node  $i$  has  $m$  linked first-order neighbor nodes  $i_1, i_2, \dots, i_m$ , which composes a local network centered on node  $i$  with local transition state  $X^i(t) = (x_i(t), x_{i_1}(t), \dots, x_{i_m}(t))$  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  $\{A_u\}_{u=1,2,\dots,2^{m+1}}$  for this local network at the next time point  $t+1$  (see Fig. 2A). To simplify notation, we hereafter drop  $i$  to denote  $X^i(t)$  as  $X(t)$ , and also denote transition state simply as state.

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

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

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

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

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

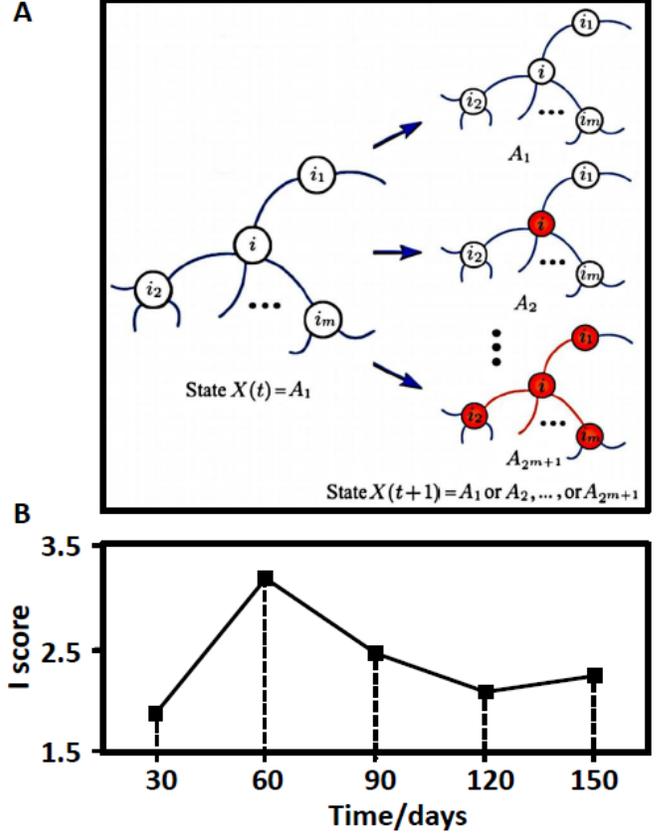


Figure 2. (A). For any state at time  $t$ , there are totally  $2^{m+1}$  possible state transitions (or possible transition states) to the state in the next time  $t+1$ . Such state transition process is modeled as a Markov process. (B) Based on the state transition process, we derived the transition-based 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.

### D. Theoretical derivation near the critical point

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

$$Z(t+1) = f(Z(t); P), \quad (3)$$

where  $Z(t) = (z_1(t), \dots, z_n(t))$  is an  $n$ -dimensional state vector or variables at time instant  $k$  representing feature values,  $P = (p_1, \dots, p_s)$  is a parameter vector or driving factors representing slowly changing factors.  $f: \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}^n$  are generally nonlinear functions. Furthermore, assume that the following conditions hold 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  $\frac{\partial f(Z; P_c)}{\partial Z} |_{Z=\bar{Z}}$  equal 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  $P_c$ , suppose the system is 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, introducing new variables  $Y(t) = (y_1(t), \dots, y_n(t))$  and a transformation matrix  $S$ , *i.e.*,  $Y(t) = S^{-1}(Z(t) - \bar{Z})$ , we have

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

where  $\Lambda(P)$  is the diagonalized matrix of  $\left. \frac{\partial f(Z; P_c)}{\partial Z} \right|_{Z=\bar{Z}}$ ,  $\zeta(t) = (\zeta_1(t), \dots, \zeta_n(t))$  are small Gaussian noises with zero means. Denote  $\sigma_i$  as the small standard deviation of  $\zeta_i$  for all  $k$ .

Without loss of generality, the diagonalized matrix  $\Lambda = (\lambda_1, \dots, \lambda_n)$  with each  $\lambda_i$  between 0 and 1.

Among eigenvalues of  $\Lambda$ , the largest one (in modulus), say  $\lambda_1$ , first approaches to 1 in modulus when parameter  $P \rightarrow P_c$ . The eigenvalue  $\lambda_1$  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  $\lambda_1 \rightarrow 1$ . Without loss of generality, assume that the first variable  $y_1$  in  $Y$  is with  $\lambda_1$ . Near a fixed point, we have proven that there exists a dominant group or a dynamical driver network (DDN), 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 [10].

Different from the analysis on the original variables  $Z$  in [10], here we focus on the variation equation of Eq. (3) with variation variables  $\Delta Z$ .

Noting

$$z_i = s_{ij}y_1 + \dots + s_{in}y_n + \bar{z}_i, \quad (4)$$

let the variation variables

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

then from Eq. (4) we have

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

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

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

where  $\xi(t) = \zeta(t) - \zeta(t-1)$  are Gaussian noises with zero means and covariances  $\kappa_{ij} = Cov(\xi_i, \xi_j)$ . It is clear that the standard deviation of  $\xi_i(t)$  is  $\sqrt{2}\sigma_i$ , where  $\sigma_i$  is the standard deviation of  $\zeta$  for all  $t$ . Obviously, variable  $\Delta y_1$  corresponds to the dominant eigenvalue  $\lambda_1$ .

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

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

Clearly, the summation of the coefficients for the covariance matrices for  $T$  Gaussian noises, 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  $T \rightarrow +\infty$  it holds

$$\Delta Y(t+T) = \varepsilon(t) \quad (5)$$

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

Back to the original variables  $Z$ , it can be referred that

$$\Delta z_i(t+T) = s_{ij}\Delta y_1(t+T) + \dots + s_{in}\Delta y_n(t+T), \quad (6)$$

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  $T$ , which generally holds 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.

Now we discuss the case near the critical transition when the dominant eigenvalue  $\lambda_1 \rightarrow 1$  (for  $\lambda_1 \rightarrow -1$ , the derivation is similar and thus is omitted).

Notice that the variation variable  $y_1$  is related to the dominant eigenvalue  $\lambda_1$ .

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

holds for any integer  $T$ , we have

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

Therefore,

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

Hence when  $\lambda_1 \rightarrow 1$  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 a small noise. In other words, the dominant variables  $\Delta y_1(t)$  are strongly dependent of its previous state when  $P$  is near  $P_c$ . It is obviously that the same result holds when  $\lambda_1 \rightarrow -1$ .

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

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

Notice that the variable  $\Delta y_1(t)$  is related to the dominant eigenvalue  $\lambda_1$ .

There are a special group of variables  $z_j$ , whose variables  $\Delta z_j$  are related to  $\Delta y_1$ , i.e., the  $\Delta z_j$  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 [10].

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

$$\begin{aligned} \Delta z_j(t+T) &= s_{j1}\Delta y_1(t+T) + \dots + s_{jn}\Delta y_n(t+T) \\ &= \frac{s_{j1}}{s_{i1}}\Delta z_i(t) + \rho_j(t) \end{aligned}$$

where

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

is Gaussian noise, which is assumed to be small. It is clear that when  $|\lambda_1| \rightarrow 1$ , 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  $\Delta z_j(t)$ . 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  $\Delta z_k(t)$ .

### E. Dynamical increase of network entropy

For a local structure centered on node  $i$  with its  $m$  linked first-order neighbor nodes  $i_1, i_2, \dots, i_m$ , 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  $\pi = (\pi_1, \dots, \pi_{2^{m+1}})$  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  $\chi$  represents the state-transition process  $X(t), X(t+1), \dots, X(t+T), \dots$  of the local network. This

entropy is actually the conditional entropy while it also describes the average transition entropy (11), depending on the state transition, i.e.,  $H_i(t) = H(X(t) | X(t-1)) = H(X(t), X(t-1)) - H(X(t-1))$ . We also note that  $X(t)$  (or  $Z(t) - Z(t-1)$ ) are variation variables. Clearly, in an early state (or a disease state), a system recovers from a small perturbation quickly because of high resilience, i.e.,  $X(t)$  and  $X(t-1)$  are almost independent. Thus, we have  $H_i(t) \approx H(X(t))$  due to  $H(X(t), X(t-1)) \approx H(X(t)) + H(X(t-1)) > 0$ , which results in a high TNE. By contrast, the system has difficulty recovering from a small perturbation in a transition state because of low resilience, i.e.,  $X(t)$  and  $X(t-1)$  are strongly correlated, which implies that  $H_i(t)$  rapidly approaches the minimum,  $H_i(t) \approx 0$  due to  $H(X(t), X(t-1)) \approx H(X(t-1))$ .

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:

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

where  $H_{control}(t)$  is the TNE based on control samples in the form of Eq. (8), and  $H_{case}(t)$  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 make 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  $i_1, i_2, \dots, i_m$  are all non-DDN members, then  $i$  is a Type 3 node.

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

Type	Node	TNE for local network
1	DDN feature	Increases drastically
2	1st-downstream feature	Increases

3	2nd-downstream feature	No significant change
---	------------------------	-----------------------

We have a detailed theoretical proof for the conclusions in [11].

### III. RESULTS

#### A. Identify the transition state

Based on comparative TNE, we selected 35 features out of 247 features and thus identified the transition state of brain tumor. Specifically, the sharp increase of the TNE index provides the early-warning signal for the imminent critical transition, that is, the commitment of brain tumor. The selected 35 features are listed in Table II.

We point out that, different from the traditional molecular biomarkers used in medicine, whose expressions reflect the presence or severity of the disease state and are required to have consistent (or constant) values that are distinct in the respective tumor and early states, the DDN is a strongly correlated feature network where the values of features, however, dynamically change without keeping constant values in the transition state as shown in Fig.1. In other words,

the system tends to present increasingly instability when the system approaches to the transition state, that is, the DDN features of DDN show increasingly fluctuation while they behave dynamically in a strongly collective manner, which is a key feature of the DDN. This is why it can be used to detect the early signal of a complex disease in the early stage, which is not otherwise possible using traditional biomarkers or methods. Hence, the existence of the DDN implies that the system is in the transition state for an individual among certain high-risk cohort, whose health condition is in a highly unstable state and thus results in high-level entropy in the driver feature network.

#### B. Development of Brain Tumors in Progeny of ENU-exposed rats

ENU exposed rats (n= 72) (13 from P30, 21 from P60, 26 from P90, 6 from P120 and 6 from P150) were examined histologically for the presence of nestin+ and OPN+ precursor lesions (nests) as well as appearance of tumors, which can be detected by MRI after day 90 (Fig. 3A middle and right). Consistent with previous reports, single or

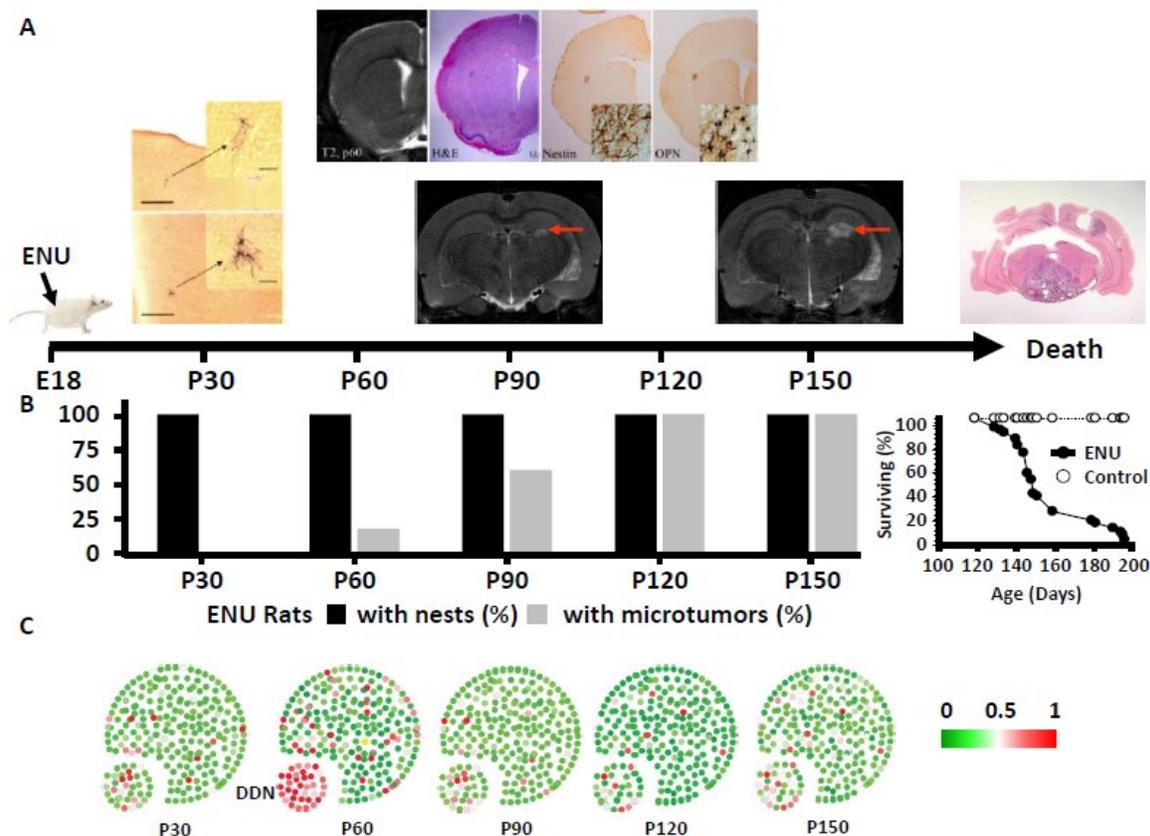


Figure 3. Development of tumor at five time points consistent with transition state. (A). Left, single and multi-cell cysts existed in all five ages of ENU-exposed rats. Middle, nestin+ and OPN+ precursor lesions (nests) presented in ENU-exposed rats by day 90, which were rare noted with MRI in any rats before P60. Right, the final histological staining by rat death. (B). Left, histogram of percentage of ENU-exposed rats with nests or microtumors. Right, surviving curve of ENU-exposed rats after day120. (C) The dynamical change is illustrated for the whole feature network, with DDN located in the lower left. It can be seen that the DDN presents a significant change at P60, which illustrates the imminent critical transition. The early-warning signal detected by DDN is in accordance with the observation in the experiment, that is, the occurrence of hyperplastic microtumors is observed in 90 days.

multiple nestin+ precursor cell cyst were noted in all rats by P30 (100%) (Fig. 3B left). In contrast, microtumors were not noted in any rats sacrificed at P30, only 4 rats (18%) at P60, while ~60% of rats at P90, and 100% of rats at P120-150 (Fig. 3B left). While the rat death was firstly found at day150, which were showed by surviving curve in Figure 3B, right. No macroscopic tumors were found in any animals at the time points examined.

*C. 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 72 ENU and 75 saline exposed rats over three independent experiments. At P30 (13 ENU- and 11 saline-treated), P60 (21 ENU- and 21 saline-treated) and P90 (26 ENU- and 29 saline-treated), 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 Methods (Table I). 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 were then grouped together for the purpose of finding peaks, and then separated again by age for further analysis of the peaks 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 and may 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), and we identified 247 DDN CSF proteins. As shown in Figure 2B, the composite TNE index I increase sharply around 60 days, indicating the critical transition into hyperplasia during glioma development after P60.

TABLE II. PART OF SELECTED FEATURES

M/Z	Functional analysis	
	Laser energy	Annotation
13913	med	glutathionylated transthyretin
14120	med	Sinapinic acid adduct of glutathionylated transthyretin
22893	high	prostaglandin D2 synthase
66110	high	albumin (z=1)
6795	med	transthyretin (z=2)
6909	med	glu-cys-transthyretin (z=2)
Features without annotation		
3487, 3544, 6953, 5704, 5789, 14285, 4196, 5788, 15847, 3641, 7515 5375, 3992, 11848, 4362, 8913, 11861, 3941, 3700, 4891, 7061, 3504 4161, 5818, 8569, 7444, 12772, 22034, 4734		

IV. DISCUSSION

Characterization of the early relationship between brain tumor cells and their environment is pivotal to the understanding of the brain tumorigenesis. Given that the rat gliomas are not generally detectable pathologically until approximately 90 days of age (P90), we tested the hypothesis 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 transition state, defined as the prelude to catastrophic shift into the hyperplasia state, occurring before the imminent phase transition point is reached, therefore, 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.

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 and nervous tissues at different days of age occurring with the gradual appearance of cellular hyperplasia. We employed our previously developed transition-based network entropy (TNE) [11] 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 [8-9]. When the ENU rats were at P60 and CSF proteome network is near the critical point, we found a dominant group of 35 CSF proteins which we defined as 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 are no significant TNE fluctuations before and after P60.

Our current up to date proteomics effort identified 6/35 of the DDN CSF proteins, and 4/6 are transthyretin species of different posttranslational modifications. Consistent with our previous observation and other reports, CSF transthyretin protein species were shown to differentially express in our ENU rat model [1] and human brain tumor [16-18]. In this regard, CSF transthyretin is a biomarker, not only differentiate between case and control, but also function as 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 represent release from these small nests and microtumors as variation diminished prior and post the critical transition at P60.

Another two CSF DDN proteins identified were albumin and prostaglandin D2 synthase (PGD2S). Both proteins are abundant in the CSF [19-20] and considering absence of hyperplasia before P60, therefore, the differential expression between case and control or variations observed before P60

reflects either albumin release from tumor cells or the impact of a space occupying lesion before imaging changes are apparent. Our previous hypothesis [1] of these proteins' differential CSF abundance was the disruption of the blood brain barrier during tumorigenesis.

Our DDN discovery at P60 findings and the DDN CSF protein identification results are consistent with the hypothesis that a CSF environmental change that is initiated before the hyperplasia/ micro tumor stage (before P60), similar to what has been reported to occur early in systemic cancers such as those involving the breast and prostate [21]. 6/35 DDN CSF proteins were identified and 29 CSF protein identities remain to be determined. Once completing all DDN CSF protein identifications, we will be at 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 an effective time window for novel therapeutic strategies in primary brain tumor.

#### REFERENCES

- [1] Whitin, J. C., et al. (2012). "Alterations in cerebrospinal fluid proteins in a presymptomatic primary glioma model." *PLoS One* 7(11): e49724.
- [2] He D, Liu ZP, Honda M, Kaneko S, Chen L. Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. *Journal of molecular cell biology*. 2012 Jun;4(3):140-52.
- [3] Hirata Y, Bruchofsky N, Aihara K. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. *Journal of theoretical biology*. 2010 May 21;264(2):517-27.
- [4] Litt B, Esteller R, Echazu J, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron*. 2001 Apr;30(1):51-64.
- [5] McSharry PE, Smith LA, Tarassenko L. Prediction of epileptic seizures: are nonlinear methods relevant? *Nature medicine*. 2003 Mar;9(3):241-2; author reply 2.
- [6] Paek SH, Chung HT, Jeong SS, et al. Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma. *Cancer*. 2005 Aug 1;104(3):580-90.
- [7] Venegas JG, Winkler T, Musch G, et al. Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature*. 2005 Apr 7;434(7034):777-82.
- [8] Gilmore R. *Catastrophe theory for scientists and engineers*. New York: Dover Publications; 1993.
- [9] Murray JD. *Mathematical biology*. 3rd ed. New York: Springer; 2002.
- [10] Chen, L. et al. (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers, *Scientific Reports*, 2(342), 1-8.
- [11] Liu, R. et al. (2012) Identifying critical transitions and their leading networks for complex diseases, *Scientific Reports*, 2(813), 1-9.
- [12] Liu, R. et al. (2013) Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers, *Medicinal Research Reviews*, DOI 10.1002/med.212932013.
- [13] Shannon CE. *Mathematical theory of communication*. The Bell System Technical Journal. 1948;27:379-423.
- [14] Ji, J., et al. (2013). "Cloud-based solution to identify statistically significant MS peaks differentiating sample categories." *BMC Res Notes* 6: 109.
- [15] Carlson, S. M., et al. (2005). "Improving feature detection and analysis of surface-enhanced laser desorption/ionization-time of flight mass spectra." *Proteomics* 5(11): 2778-2788.
- [16] Fung, E. T., et al. (2005). "Classification of cancer types by measuring variants of host response proteins using SELDI serum assays." *Int J Cancer* 115(5): 783-789.
- [17] Schuhmann, M. U., et al. (2010). "Peptide screening of cerebrospinal fluid in patients with glioblastoma multiforme." *Eur J Surg Oncol* 36(2): 201-207.
- [18] Zhang, Z., et al. (2004). "Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer." *Cancer Res* 64(16): 5882-5890.
- [19] Melegos, D. N., et al. (1997). "Prostaglandin D synthase concentration in cerebrospinal fluid and serum of patients with neurological disorders." *Prostaglandins* 54(1): 463-474.
- [20] Thomsen, C., et al. (1990). "Fourier analysis of cerebrospinal fluid flow velocities: MR imaging study. The Scandinavian Flow Group." *Radiology* 177(3): 659-665.
- [21] van Kempen, L. C., et al. (2006). "Inflammation, proteases and cancer." *Eur J Cancer* 42(6): 728-734.