Computational and Structural Biotechnology Journal 20 (2022) 1189–1197

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Mini review

# Development of a dynamic network biomarkers method and its application for detecting the tipping point of prior disease development

Chongyin Han [a], Jiayuan Zhong [b], Qinqin Zhang [a], Jiaqi Hu [a], Rui Liu [a], Huisheng Liu [a], Zongchao Mo [a], Pei Chen [b,*], Fei Ling [a,*]

[a] Guangdong Key Laboratory of Fermentation and Enzyme Engineering, School of Biology and Biological Engineering, South China University of Technology, Guangzhou, Guangdong, China
[b] School of Mathematics, South China University of Technology, Guangzhou, Guangdong, China

ARTICLE INFO

*Article history:*
Received 14 October 2021
Received in revised form 21 February 2022
Accepted 21 February 2022
Available online 24 February 2022

*Keywords:*
Dynamic network biomarker
Tipping point
Cancer
Complex disease

ABSTRACT

The dynamic network biomarker (DNB) method has advanced since it was first proposed. This review discusses advances in the DNB method that can identify the dynamic change in the expression signature related to the critical time point of disease progression by utilizing different kinds of transcriptome data. The DNB method is good at identifying potential biomarkers for cancer and other disease development processes that are represented by a limited molecular profile change between the normal and critical stages. We highlight that the cancer tipping point or premalignant state has been widely discovered for different types of cancer by using the DNB method that utilizes bulk or single-cell RNA sequencing data. This method could also be applied to other dynamic research studies and help identify early warning signals, such as the prediction of a pre-outbreak of COVID-19. We also discuss how the identification of reliable biomarkers of cancer and the development of new methods can be utilized for early detection and intervention and provide insights into emerging paths of the widespread biomarker candidate pool for further validation and disease/health management.



**Contents**



* Corresponding authors.
*E-mail addresses:* fling@scut.edu.cn (P. Chen), chenpei@scut.edu.cn (F. Ling).

https://doi.org/10.1016/j.csbj.2022.02.019

## 1. Introduction

In disease development, the disease slowly progresses from a normal state, and after a critical state, it rapidly develops into an irreversible disease state [1]. There are numerous differences between the normal state and the disease state, and many biomarkers have been developed based on this case-control method. However, an increasing amount of evidence shows that these biomarkers might have very limited clinical effects, which indicates that the case-control research method might be ineffective and suggests that new theories and methods could serve as alternatives. If the critical state signals can be captured for early warning and intervention, it would be possible to prevent the disease from entering an irreversible disease state. Because the difference between the normal state and the critical state is small and thus becomes challenging to detect, here, we mainly review the DNB method that can solve and address this dilemma by quantifying the early warning signals of the critical state using network theory. We summarize the recent application of DNBs in cancer and other diseases for tipping point identification. By capturing early warning signals and performing an early intervention for critical health conditions, better management might be achieved based on different perspectives to help prevent major disease progression and improve quality of life.

## 2. Cancer biomarkers, cancer tipping points, and dynamic network markers

### 2.1. Cancer biomarkers

Molecular markers are molecules or molecular groups that have an indicative effect on the process, type, or therapeutic effect of the disease [2]. Generally, such molecules are divided into three relatively important types.

(1). Diagnostic molecular biomarkers: These markers indicate whether the patient is suffering from the disease. Such molecular markers are usually obtained from the analysis of disease-normal sample pairs. These markers may be mutation sites, mutant genes, or genes with significantly high/low expression or significant changes in characteristic metabolites [3]. Some markers can define the type of disease and are of great value to the choice of treatment methods for patients. Diagnostic markers usually can only provide clinical information that has already appeared but cannot be used as a predictor of disease.
(2). Therapeutic molecular biomarkers: Markers directly related to the treatment process of disease patients are called therapeutic markers and are the most important type of markers for patients with diseases. They are usually used to judge or predict the response of a patient to a specific type of treatment. For example, in colorectal cancer, mutations in the *KRAS* gene and *BRAF* gene are used to predict the sensitivity of the patient to chemotherapy [4].
(3). Molecular biomarkers of patient prognosis: All molecules that are significantly related to patient survival are called prognostic markers, and there are generally two categories, namely, those related to the high survival rate of the patient or related to the low survival rate of the patient. Clinically, the most direct indicator to measure whether the treatment is effective is the patient's survival rate. Moreover, the overall survival (OS) rate is the most objective indicator to measure the treatment effect [5].

For the discovery of molecular markers of diseases, DNA-level information, such as point mutations, structural variations, hotspot gene mutations, and genome stability as markers, might be appropriate. Proteins can also act as biomarkers that reflect the clinical status of the patient [6]. Among them, the selection of RNA as a biomarker is a very good choice. RNA levels can reflect the cell's response to the environment and important gene expression regulation information at a considerable level.

### 2.2. Cancer tipping points

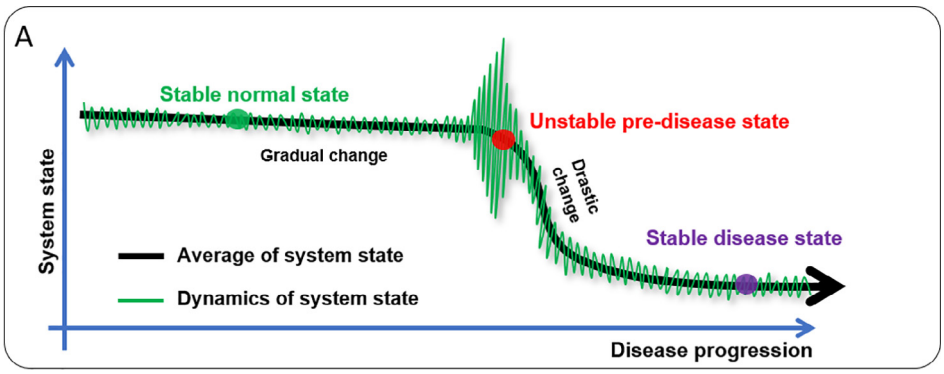
Critical transitions or critical points widely exist in many natural and social systems, including biological systems. Before the onset of disease or the appearance of symptoms, it is very important to determine the critical point or critical state of a complex disease for disease prevention and early treatment. Generally, regardless of the specific biological and pathological differences in the progression of the disease, the progression of complex diseases can be roughly divided into three stages [7–8]: (1) Normal state, which is a stable state and represents a relatively healthy and high resilience condition; (2) Critical state, a kind of state that is not normal but exists before the disease state, represented by low resilience and high susceptibility; and (3) Illness or diseased state, a stage in which quality is decreased but there is still high resilience.

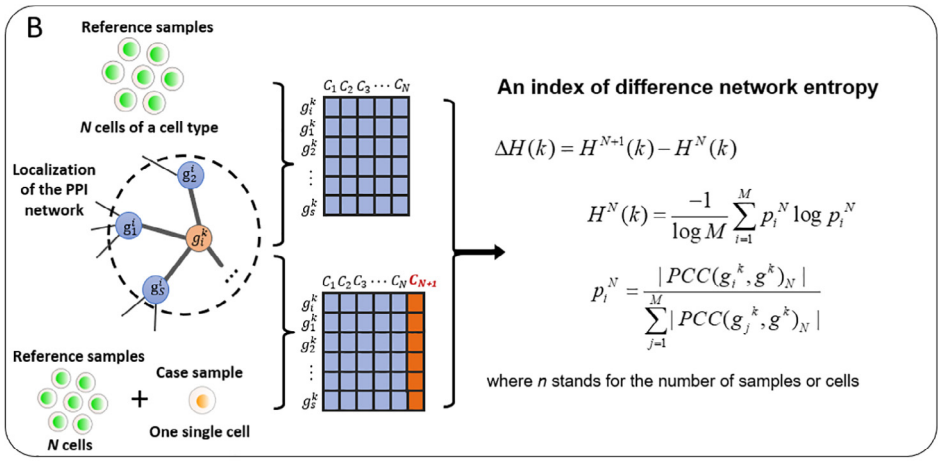
It is important to predict the critical state to prevent or at least prepare for the coming deterioration by making appropriate interventions. However, in contrast to disease states, determining the critical state is a difficult task because it is similar to the normal state in terms of phenotype and molecular expression, which renders traditional static biomarkers ineffective. An increasing number of studies have found that the occurrence of diseases is due not only to single factors but also to more complex and interactive network structures. In the context of the dynamic system theory of bifurcation and critical slowdown [9–10], a conceptual biomarker called a dynamic network was developed to quantitatively describe the critical state in the process of disease progression [9].

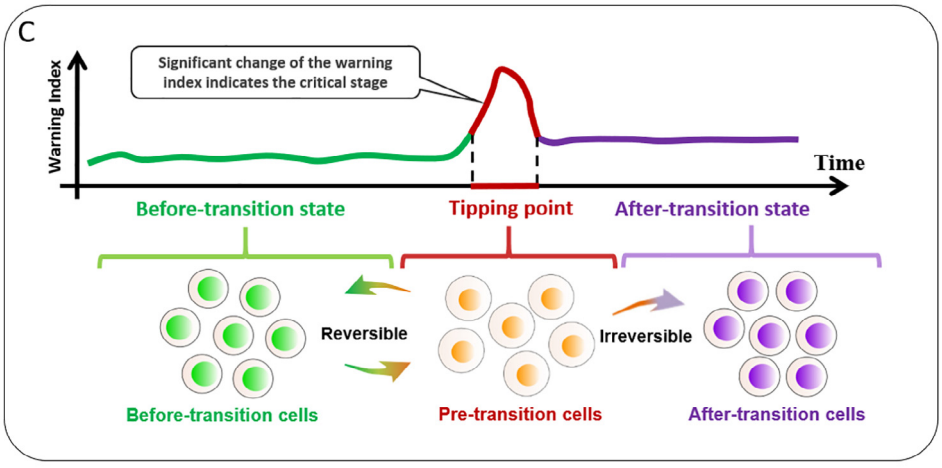
### 2.3. DNB algorithm and its development

Specifically, the dynamic development process of the disease can be considered a time-dependent nonlinear dynamic system. The dynamic network biomarker (DNB) concept provides a statistical method for measuring critical state-related variables; that is, a small group of closely related variables communicates early warning signals of the upcoming critical transition through drastic statistical indicators [7–8]. DNBs are a set of molecules (genes or proteins) or molecular modules that can signal a critical point or critical state before the rapid deterioration associated with a complex disease. When the system reaches a critical state, DNB molecules have the following properties [11]:

(1). In the critical period, the contact between molecules within the group is rapidly strengthened, that is, the correlation rapidly increases.



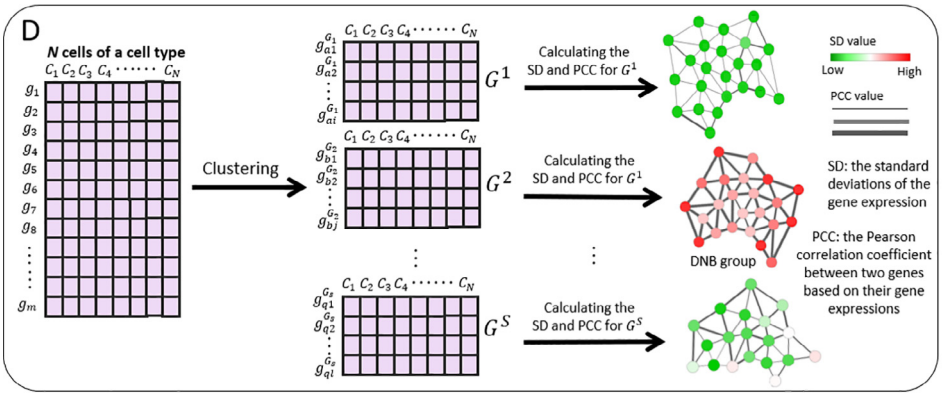




**Fig. 1.** Introduction of dynamic network biomarkers. (A) Key theory of dynamic network biomarkers. (B) Algorithm and basic principle of dynamic network biomarkers. (C) Dynamic network biomarkers have an early warning function for the critical stage before the occurrence of disease. (D) The schematic illustration for identifying the dynamical network biomarkers (DNBs).

(2). In the critical period, the connection between the molecules in the group and other molecules weakens, that is, the correlation drops sharply; and
(3). In the critical period, the standard deviation of the molecules in the group increases rapidly, that is, the standard deviation of the molecules in the group increases.

According to such mathematical assumptions, the DNB framework uses a development chain of hidden Markov model (HMM) to transform the progress of complex diseases into a static hidden Markov model (sHMM) and a dynamic hidden Markov model (dHMM) process. Overall, the DNB framework is a time series model, representing the gradual development of the entire complex process, and there are signal values that can be measured or calculated manually. The overall model is composed of three submodels, which represent the DNB hypothesis.

DNB methods and theories have been applied to the study of many biological processes, such as detecting the critical point of cell fate determination [12] and cell differentiation [13], studying immune checkpoint blockade [14], and determining the stage before the deterioration of different diseases [15–17].

However, the traditional DNB method requires multiple samples at each time point, which is usually difficult to obtain in clinical and other practice settings, which greatly limits the application of the traditional DNB method in most practical situations. The rapid development of high-throughput technology provides new insights for computational analysis. Even when there is only one sample, it is possible to quantify critical changes in the dynamic development of diseases based on high-dimensional data from a single sample.

Single-Sample Network (SSN): Single-sample DNB methods based on DNB theory, such as the single-sample hidden Markov model [18], single-sample Kullback–Leibler divergence method [19], and single-sample dynamic network biomarker method [20], have been widely developed and applied. SSN theory uses a group of individuals (N individuals) as a reference, and the method of mapping each individual can make the individual level a network dimension. To achieve this, a network of a new group (N + 1 individuals) is built and compared to the original network (N individuals) to obtain the difference between the two networks. The difference network is the network of individuals relative to the reference group (Fig. 1).

Disease diagnosis at this stage is based on the disease that has already developed, and follow-up intervention and treatment are performed. The DNB theory focuses on predicting the occurrence and development of the disease from the perspective of never occurring. The key is to study the critical point of a dynamic network marker change related to the disease. The state of the critical point is characterized by a small difference from the healthy state, but the relationship between the factors has changed. Through the three characteristics mentioned above, DNBs can be screened one by one, and the information of individual network critical points can be used to make relevant predictions. This study provides a new method and way of thinking, adding a new dimension to complex disease assessment and prediction networks. Furthermore, this network can be based on the individual level.

Chen developed a new model-free method called the landscape dynamic network biomarker (l-DNB) method [1]. This method is based on the bifurcation theory and uses only one-sample omics data to determine the critical point before the disease seriously worsens. This method evaluates local criticality or the local DNB score gene by gene and then compiles the overall local DNB score into a landscape. Then, the global critical score IDNB can be calculated from the landscape of the sample (or patient) by selecting those genes as the DNB members with the highest local DNB score.

Initially, Chen proposed a DNB method to investigate the gene module that is associated with disease development [7]. With the development of DNB methods, there have been many improvements in expanding the application of DNBs to other research. For example, Zhang proposed a new method, GNIPLR, to infer gene regulatory networks (GRNs) [21]. Alcudia developed a metaheuristic multiobjective optimization method for DNB identification using two steps [22]. First, the prefiltering procedure contained several steps to identify informatic genes. Second, the artificial bee colony based on dominance (ABCD) algorithm was developed for DNB identification [22]. To achieve better DNB performance, Guo built an assessment pipeline to evaluate the performance of a single sample method and found that different methods showed diverse performance in network building. The undirected method surpassed the directed method for sample-specific networks [23]. Improvements in network building will facilitate dynamic network marker identification.

### 2.4. DNB experimental design

Protein-based dynamic network biomarkers will provide more information to distinguish states between normal and predisease, allowing for earlier disease detection. In a study that combined proteomic expression profiling of inflammatory mediators with clinical informatics in patients with acute exacerbations, a panel of inflammatory mediators was found to change dynamically with disease progression [24]. This is a specific biomarker for patients with chronic obstructive pulmonary disease or acute exacerbation of chronic obstructive pulmonary disease [25]. The method paves the way for the development and validation of disease-specific dynamic network biomarkers.

In terms of DNB experimental design, a study modeled orthotopic liver cancer transplantation mice, carried out DNB research, and detected early warning signals of lung metastasis of liver cancer based on transcription gene expression map [26]. The following was an example of an experimental design: First, transcription data were continuously collected at different time points in the animal model, and the samples were halted when the disease phenotype was observed in animals, such as liver cancer lung metastasis. Then, the transcription data were evaluated using the DNB approach to find signals of liver cancer before lung metastasis at various time points. Finally, the experiment was designed to confirm using the method of biomarker discovery. Our research group is currently investigating the tipping point for the abnormal aggregation of α-syn, a fundamental pathology of Parkinson's disease, using methods comparable to those described above.

For another example, early diagnostic markers or intervention markers for the occurrence of colorectal cancer could collect clinical samples from four stages of normal-polyp-adenoma-stage I colorectal cancer. Transcriptase sequencing and analysis were then performed to obtain early markers, which could be used as the primary diagnosis or intervention to prevent the progression of colorectal cancer. The DNB experimental design is the endpoint of sample collection when the key biological process or clinical event under investigation happens.

## 3. Application of DNBs for the detection of tipping points and early warning signals in cancer development and progression

To treat the disease more effectively, it is very important to identify the critical period of disease development. The identification of molecules that play a role in the critical period is also of great help to the development of treatment strategies. There is usually no significant difference in these characteristics between the normal state and the critical state. Most traditional disease

**Table 1**
Application of the DNB method in cancer and other disease tipping points identification.

| Disease | | Data Source | Outcome | | Reference |
|---|---|---|---|---|---|
| Type | | | Tipping Point | Survival Significance | |
| Cancer | Rectum adenocarcinoma | TCGA-COAD | Stage III | log rank $p = 0.01$ | https://doi.org/10.3389/fbioe.2020.00809 |
| Cancer | Uterine Corpus Endometrial Carcinoma | TCGA-UCEC | Stage IIB | log rank $p < 0.001$ | |
| Cancer | Esophageal carcinoma | TCGA-ESCA | Stage IIIA | log rank $p < 0.001$ | |
| Cancer | Head and Neck squamous cell carcinoma | TCGA-HNSC | Stage II | log rank $p = 0.04$ | |
| Cancer | Lung adenocarcinoma | TCGA-LUAD | Stage IIB | log rank $p < 0.001$ | https://doi.org/10.1093/nsr/nwy162 |
| Cancer | Thyroid carcinoma | TCGA-THCA | Stage III | log rank $p = 5E-5$ | |
| Cancer | Kidney renal clear cell carcinoma | TCGA-KIRC | Stage II | log rank $p = 0.01$ | |
| Not | Cancer | Influenza virus infection | GSE30550 | 45 Hours | |
| Not | Applicable | | | | |
| Cancer | Hepatocellular carcinoma | GSE94016 | 3 Weeks | Not Applicable | https://doi.org/10.1038/s41467-018-03024-2 |
| Cancer | Lung squamous cell carcinoma | TCGA-LUSC | Stage IIA | log rank $p = 0.0034$ | https://doi.org/10.1186/s12864-020-6490-7 |
| Cancer | Lung adenocarcinoma | TCGA-LUAD | StageIIB | log rank $p = 3E-7$ | |
| Cancer | Stomach adenocarcinoma | TCGA-STAD | Stage IIIB | log rank $p = 0.0275$ | |
| Cancer | Thyroid carcinoma | TCGA-THCA | Stage II | log rank $p < 0.001$ | |
| Cancer | Colon adenocarcinoma | TCGA-COAD | Stage II | log rank $p < 0.001$ | |
| Not | Cancer | Lung Injury | GSE2565 | 8 Hours | Not Applicable |
| | https://doi.org/10.1038/srep00342 | Cancer | Lymphoma | GSE6136 | Level P2 |
| | | Not Applicable | | | |
| | | Cancer | HBV induced liver cancer | NA | F3 period |
| Not | Applicable | | | | |
| Cancer | Gallbladder cancer | shinyapps/human_gbc | Not Applicable | Not Applicable | https://doi.org/10.1016/j.jhep.2021.06.023 |
| Cancer | HRG-induced differentiation of cancer cells | GSE13009 | 1.5H-Agree with Experiment | Not Applicable | https://doi.org/10.3389/fgene.2015.00252 |
| Cancer | HRG-induced differentiation of MCF-7 cells | GSE6462 | 0.5H-Agree with Experiment | Not Applicable | |
| Cancer | HRG-induced differentiation of MCF-7 cells | GSE10145 | 1H-Agree with Experiment | Not Applicable | |
| Cancer | MCF-7 cells treated with tamoxifen for 12 weeks | CRA000580 | 4 Weeks | Not Applicable | https://doi.org/10.1093/jmcb/mjy059 |
| Cancer | TGF-beta treated A549 lung adenocarcinoma cells | GSE17708 | 2 Hours | Not Applicable | https://doi.org/10.1016/j.scib.2020.01.013 |
| Not | Cancer | Brain cortex development of human | GSE11512 | 0.2–0.3 Years Old | Not Applicable |
| | https://doi.org/10.1186/s12864-020-6465-8 | Not Cancer | Brain cortex development of macaque | 0.1–0.5 Years Old | Not Applicable |
| Not | Cancer | Goto–Kakizaki (GK) rats | GSE13268, GSE13269 & GSE13270 | 8 Weeks | Not Applicable |
| | https://doi.org/10.1093/bib/bbt027 | | | | |

state biomarkers are identified based on the differential expression of molecules between the disease and the normal states rather than diagnosing a critical state.

Therefore, identifying critical points or pre-disease states is an important challenge in medicine or biology. In addition to understanding the molecular mechanisms of complex diseases at the network level, it is also possible to understand the warning signs for the preventative and preemptive treatment of diseases. In particular, the DNB method has been proposed to detect critical states of many diseases using nonlinear dynamic theory. The l-DNB method described here represents a new method that can reliably and accurately identify key states and detect early warning signals of complex diseases on a single-sample basis. Therefore, the DNB is an ideal tool that can be used to detect the critical state before the onset of disease [8].

The DNB calculation model and its extended model have been applied to the analysis of real biological and clinical data in many studies, such as the detection of the critical state of a variety of complex diseases, including type I diabetes and a variety of cancers, the critical time point and core of cell differentiation gene group mining, and the identification of key nodes in the immune checkpoint blocking process [13–14,17–18,27–28]. In our review, we summarize the recent application of the DNB method in cancer tipping points and the identification of warning signals (Table 1).

### *3.1. Detection of disease critical points through landscape dynamic network biomarkers*

The L-DNB provides early warning signals of disease deterioration on a single-sample basis and detects key genes or network

biomarkers (i.e., DNB members) that promote the transition from a normal state to a disease state.

The l-DNB method was applied to the three tumor disease datasets from the TCGA and used for each patient's DNB to detect the critical stage before tumor progression [27]. Individual DNBs were further used as individual biomarkers in physiological data analysis, which led to the identification of two types of biomarkers that are very effective in predicting tumor prognosis. Biomarkers can be considered common biomarkers of cancer: one type indicates a poor prognosis, and the other indicates a good prognosis.

The process of cancer usually has three stages of development, which are clear in the clinical diagnosis: the early stage of cancer, rapidly advanced stage, and late stage of cancer. This means that the tipping point may appear between the transition from a healthy state to the early stage of cancer or between the early stage of cancer and the rapidly advanced stage of cancer. Considering the actual clinical hazards and the extreme difficulty of obtaining tissue transcriptome data from healthy people, clinically, there is usually more attention to the process of cancer patients transitioning from the early stage to the rapid progression stage. Once cancer passes this critical period, a variety of clinical symptoms will appear, such as obvious primary cancer lesions, infiltrating tissues, metastases, and obvious metabolic abnormalities. At this time, the difficulty of conventional treatment has been greatly increased. Therefore, the identification of the critical state before the advanced stage is critical for timely medical intervention in cancer.

### 3.2. Dynamic network markers applied to colorectal cancer

The latest study used the DNB method in hepatocellular carcinoma and colorectal cancer research to identify markers of key states [29,31]. In our research, we developed an algorithm called single-sample node entropy, which uses only a single sample to measure cell signal activity in key stages of cancer. It is a model-free method and does not require any model training process. We confirmed the stability and sensitivity of the simulation data and the TCGA cohort data analysis in our study. Using this model, we determined the critical points of cancer progression and found that they are as important as expected for patient survival. We also found that some pivotal genes are highly related to some important biological functions, such as cancer cell proliferation and invasion. The results of the differential expression analysis revealed that the hub genes mainly act as gatekeepers or core relays in multiple important pathways, such as FZD8/9, which controls the entry of the canonical Wnt pathway, or RRAS2, which dominates the Ras signaling pathway and has been proven to interact with the breast cancer drug tamoxifen [32]. In summary, we successfully developed a dynamic node entropy model based on single-case data to identify the critical point and possible mechanisms of cancer progression. These findings may provide new target genes for therapeutic intervention strategies.

Therefore, by appropriately adjusting the expression levels of these hub genes, we can easily affect their downstream genes, such as the *MYC* or *MAPK* gene families, in some important signaling pathways, thereby further affecting the growth of cancer. We noticed that the node entropy of some genes was significantly correlated with the survival of patients, but their expression levels were not significantly correlated, which indicates that we can use our model to obtain additional prognostic biomarkers. Therefore, if we obtain enough clinical patient samples, we can accurately measure the ability of node entropy as a prognostic factor. Finally, we identified a series of SNE core genes related to DNA repair. Among them, the *MUTYH* gene and *PARP2* gene are closely related to the two approved anticancer drug mechanisms, cisplatin and olaparib [33–34], indicating that these nonspecific genes may also become drug targets.

In addition, the results of the differential analysis showed that most of the dynamic network biomarkers identified by node entropy, such as NKD2 or DAAM1, are located upstream of many important cancer-related signaling pathways, regulating the inter-genes in the signaling pathways. We also used node entropy instead of the expression level to identify some new prognostic biomarkers, such as *PER2*, *TNFSF4*, *MMP13*, and *ENO4*. More importantly, we found that the conversion of nonspecific pathways related to DNA damage repair is the main driving force for cancer progression.

Based on DNB and single-cell data, Hu identified a subgroup of pre-exhausted CD8+ T cells that contributes to T cell exhaustion in CRC [35]. The hub genes CCT6A and TUBA1B were identified as the core contributors to T cell exhaustion. Single-cell analysis of colorectal cancer adjacent tissue B cells revealed that stage II is a critical period before lymph node metastasis, and the DHX9 gene was identified to be involved in dynamic network changes during CRC progression [36].

### 3.3. Dynamic network markers applied to hepatocellular carcinoma and gallbladder cancer

HCC is one of the most fatal cancers. In the application of l-DNB to hepatocellular carcinoma (HCC), Sun et al. mined bulk RNA-seq data to identify early warning signals from a state of cirrhosis to a very advanced HCC state in individual patients and found that both low and high dysplastic states are critical for HCC [29] based on the l-DNB method. Gao identified a group of genes capable of distinguishing the disease group from the healthy group, suggesting that this gene module might facilitate disease diagnosis [30]. Yang identified a DNB network biomarker for an indication for pulmonary metastasis in hepatocellular carcinoma [26]. In recent years, the recognition of this periodic molecular dynamic state of cancer has received considerable attention. For example, early warning signals of lung metastasis have been found in liver cancer, and with the help of time-series gene expression data in spontaneous lung metastasis mouse models, it was determined that the calcium ion conductive protein gene (Calmodulin-like-protein 3, CALML3) is significantly related to the initiation of metastasis, and it was confirmed that knocking down CALML3 can inhibit metastasis [37]. Zhang *et al.* identified a premalignant state of gallbladder cancer using the DNB method by integrating single-cell information [38]. The combination of DNBs with pseudotime analysis enables a more precise determination of cell subtypes where some subtypes are mixed.

### 3.4. Dynamic network markers applied to lung adenocarcinoma

For lung cancer research, a dataset was also included for investigation [20]**.** The l-DNB method was further applied to the analysis of three different tumors, including lung adenocarcinoma (LUAD). In particular, the criticality of LUAD was determined in the IIB stage. In addition, we found that DNB members could effectively predict prognosis when the DNB that is common in population samples was further applied to physiological data as a common biomarker. In addition, the analysis revealed that DNB members could be divided into two types of molecules to predict the prognosis of three tumors: one for samples with a poor prognosis (i.e., negative biomarkers) and the other for samples with a good prognosis (i.e., positive biomarkers). In summary, these results indicate that l-DNB can reliably identify the criticality of the disease on a single-sample basis using the DNB module. Importantly, this method quantifies early warning signs before the disease worsens and provides real network biomarkers for each person's disease prediction. Based on the DNB method, survival analyses revealed

that *SMAD7* and SERPINE1 are DNB genes and further act as prognostic biomarkers for lung adenocarcinoma [39].

### 3.5. Dynamic network markers applied to breast cancer

Based on the RNA expression data of breast cancer MCF-7 cells, Chen identified the differentiation state of these cells using the DNB method. The network module identified could not only serve as a cancer biomarker but also be used to identify drug targets [15]. Specifically, many genes lying within the DNB are associated with the cancer process, which highlights the reliability of the DNB method in identifying novel breast cancer biomarkers. In another study concerning breast cancer, Liu identified the tipping point of the endocrine resistance process using the DNB method [40].

### 3.6. Dynamic network markers applied for cancer prognosis

The application of a single DNB to clinical data shows that DNB members are effective for prognostic analysis, which can be demonstrated by identifying positive and negative biomarkers for the disease status of LUAD, THCA, and KIRC (three diseases). Therefore, if the patient's DNB includes negative biomarkers, the patient's survival time may be shortened. Four genes (*PSG3*, *AFP*, and *ADH4* in LUAD, *SPANXN3* in KIRC) were identified as negative biomarkers, but they were not differentially expressed between identified and unidentified samples. Therefore, depending on differential expression patterns, traditional methods cannot detect them. This result means that the l-DNB method can reveal "dark matter" genes (genes that are not differentially expressed) that are usually ignored by traditional analyses.

## 4. Application of the DNB method to other complex diseases and health

The DNB method could also be applied to many other research fields whose data are organized in a time-series format. Epithelial-mesenchymal transition (EMT) is a complex biological process that plays a significant role in many basic biological processes, such as embryogenesis, wound healing, tissue regeneration, and cancer metastasis. EMT is one of the key changes in cancer development. However, it is still challenging to identify these states. According to Wang's review [41], the DNB method could be applied to investigate the phase change during the EMT procedure. Jiang identified a DNB group with 37 genes that can provide early-warning signals of EMT: *SMAD7* and SERPINE1 promote EMT by switching their regulatory network [39]. In a case study, l-DNB was used to predict severe flu symptoms before actual symptoms of influenza virus infection appeared. The l-DNB method was applied to the dataset of influenza virus infection as a case study [37]. As individual biomarkers, l-DNBs can reliably detect early warning signs of a disease state transition and accurately predict how one individual biomarker is better than traditional methods at least 8 h in advance. The tipping point for infant brain development in both humans and chimpanzees was also identified using the DNB method [42]. Tipping points were found at approximately 1 month and 3 months for humans and chimpanzees and opened a new way for omics-level investigations of primate brain development. Liu found that DNBs cannot only signal the emergence of critical transitions for the early diagnosis of diseases but also provide a causal network of transitions for revealing the molecular mechanisms of disease initiation and progression at a network level [43]. A similar method was also applied to predict the outbreak of COVID-19 using geographic information and daily new case data [44]. Such prediction of the preoutbreak stage might help to better monitor public health policy.

## 5. DNBs application to biomarkers research

Dynamic network biomarker (DNB) method is a network-level model of omics data with time series based on robust hidden Markov mathematical model. It is primarily used to determine the crucial state of a biological system, making it ideal for biological process research. DNB can also be utilized for risk assessment, early diagnosis, disease monitoring, disease categorization, staging and grading, efficacy prediction, and other biomarkers.

As a risk assessment biomarker, it provides a quantitative approach to determine when an individual is vulnerable to a specific form of cancer for forecasting and assessing risk. It is frequently related to genetic mutations or epigenetic modifications. For example, mutations in the liver cancer gene *EGFR* and pancreatic cancer gene *ERBB2*, are potential predictive cancer markers. Prognostic biomarkers can indicate disease prognostic features, the probability of recurrence, or disease progression. Hormone receptors in breast cancer and prostate-specific antigen in prostate cancer are common prognostic indicators. Predictive biomarkers are used to predict a patient's response to a treatment or intervention. Such as *HER2* expression and anti-HER2 therapy in breast and gastric cancers [43–44]. As a dynamic monitoring biomarker, it is possible to monitor gene expression in patients as the disease progresses. For example, carcinoembryonic antigen (CEA) is used to detect disease recurrence in colorectal cancer [45]. CA15-3 and CEA can track the response and progression of breast cancer [46].

The DNB gene can be implemented as a drug efficacy marker if it is a pharmacological target. Xu *et al.* identified biomarkers related to the development and efficacy of CML based on DNB theory from the perspective of the CML development process and treatment process [47]. In this study, the gene expression data of CML patients treated with imatinib at various time points were included. According to the DNB theory, the efficacy criteria and index were established. By observing the changing trend of the index, the time point before the condition was stable that can be detected, and the markers related to the efficacy were identified. At the same time, from the perspective of ceRNA, Xu *et al.* constructed the ceRNA network related to lncRNA in CML patients treated with imatinib for one month, and identified the ceRNA network module related to lncRNA based on the efficacy criteria and obtained the lncRNA and mRNA markers related to the drug efficacy. Therefore, the DNB gene can be used to identify the response markers of chemotherapeutic drugs.

Some biomarkers can be used as a predictive or detection marker, as well as a therapeutic or pharmacodynamic marker. For example, the *DHX9* gene can be utilized as a prognostic marker because of its low expression in mature B cells, which is linked to a bad prognosis. The high expression of *DHX9* in stage II patients may serve as a stage marker. And its expression difference in different stages of CRC can be employed as a monitoring biomarker. If studies show that *DHX9* is a targeted biomarker, it will be used as a pharmacodynamic marker in the future.

## 6. Concluding remarks

DNBs have more advantages than traditional network biomarkers. This is mainly reflected in the fact that the DNB method is a purely data-driven method that does not rely on prior knowledge and does not need predefined gene sets related to certain biological functions, ensuring that the result is unbiased. Furthermore, the data used for this method do not require labels, and the appearance of the critical point has nothing to do with the artificially set data classification but only with the time point and the actual signal value. The hidden Markov model adopted by the DNB method also ensures that the method can tolerate a considerable

degree of data noise, which is very beneficial for cancer research because cancer transcriptome data contain a considerable degree of background noise, such as noncarcinogenic factors in the high expression of multiple ribosome-related genes. Therefore, the DNB computing framework has considerable robustness, strong generalization ability, and good processing capabilities for high-dimensional complex data.

Importantly, the method proposed here is model-free and does not require a learning process to identify biomarkers. This has advantages over traditional classification or machine learning methods, which require many case/control samples for supervised or unsupervised learning to avoid the problem of overfitting. Specifically, the l-DNB method is constructed based on three model-free DNB conditions for each sample that are based on the basic dynamic characteristics of the critical state of general biological systems. Therefore, this method inherently identifies individual biomarkers rather than common biomarkers without the problem of overfitting. However, it should be noted that identifying common biomarkers for all individuals for each disease or identifying a common critical threshold may require data from the entire population.

Based on the above description, we believe that using dynamic network biomarkers for sequential biological process research has considerable advantages. If the important turning points in the timing process can be identified and the core change network and important signal nodes can be obtained, this will be particularly important for disease research. For most diseases, the earlier the discovery, the earlier the intervention, the clearer the drug target, the better the clinical efficacy, and the smaller the burden of treatment for patients. In this regard, the DNB computing framework provides a feasible quantitative research method that has obtained many unique results from research on the dynamics of the network. A further application could also help to identify key warning genes during aging and health, which can be recognized early to prevent the reaction caused by "cell death" and will provide a new perspective for aging research and health research and management.

However, there are still some shortcomings to the cancer research data in the TCGA with the application of the DNB method. The first limitation is the lack of some cancer types in the TCGA. Currently, there is no pan-cancer DNB research in the TCGA. Therefore, the pan-cancer level has not yet become a critical point of study. Additionally, most studies are directed at a specific cancer type. There are specific markers for certain types of cancer, but it is difficult to explore potential drug targets across cancer types. Second, when studying genes through pure expression levels, the value of a single gene acting alone is emphasized, but there is no method to measure the value of genes in the overall intracellular signaling network. As discussed earlier, this research cannot directly interpret the mechanism of important dynamic changes in cells from the perspective of the network, and there is a certain degree of bias. Finally, traditional research based on paired difference analysis often reflects changes that have occurred in the body and cannot predict the critical period that is too dependent on clinical diagnosis experience; thus, it has certain disadvantages for early medical intervention.

Therefore, at the network level, it is undoubtedly very important to study the core genes related to the activities of important signaling pathways in cancer. Of course, some possible challenges still need to be considered in the research process:.

(1). The size of the dataset that can be obtained. The traditional DNB calculation framework is very robust for large datasets, but for small datasets, there may be difficulties. Therefore, under the condition that the amount of data that can be obtained is limited, we need to develop specific models for small datasets or even datasets based on single samples;
(2). When designing a disease model, the characteristics of the time of the disease sample need to be carefully considered. The DNB calculation framework requires a complete time series data link to obtain accurate results, which is often very rare in actual experiments. Therefore, when reviewing the data, it is necessary to ensure that the time series information of the data can be included in the model; and
(3). Although the DNB calculation framework has a certain tolerance for data noise, it also has a certain threshold that the acquired data are uniform, stable, and biologically repeatable to ensure the accuracy of the data-driven method. In addition to ensuring that the original data conform to the statistical distribution assumptions, it is also necessary to verify the usability of the model through numerical simulation, which guarantees the basic usability of the numerical model.

Therefore, if data quality and timing issues can be solved and a model that can perform well in small datasets is developed, cancer bioinformatics mining based on cancer transcriptome data and clinical information would be improved. Future integration of multi-omics datasets would help to identify a more reliable function for biomarker identification. Additionally, the application of the DNB method to spatial transcriptomics might enable distinctive discoveries.

More importantly, although the DNB method is currently widely used to identify important gene sets or networks related to the development or progression of specific diseases, more in silico and in vivo validations are still needed to ensure that the signal is real and could be biologically reliable. There is still a long way to go in obtaining overlapping biomarkers using various computational methods [20]. Specifically, it is usually difficult to obtain the pre-transiting sample for cancer research, and a comprehensive design for sampling would enable the validation of DNB gene function.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62172164, 12026608, and 11971176).We thank Professor Mingding Li of Zhejiang University and Professor Luonan Chen of Shanghai Institutes for Biological Sciences.

## Author contributions

Fei Ling and Zongchao Mo contributed to the conception of the study. Jiayuan Zhong and Pei Chen performed the experiment. Fei Ling and Zongchao Mo contributed significantly to analysis and manuscript preparation. Fei Ling and Zongchao Mo performed the data analyses and wrote the manuscript. Jiaqi Hu, Chongyin Han, Huisheng Liu and Rui Liu helped perform the analysis with constructive discussions. Manuscript is approved by all authors for publication. The work has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.02.019.

## References


[1] Liu X et al. Detection for disease tipping points by landscape dynamic network biomarkers. Natl Sci Rev 2019;6:775–85.
[2] Ballman KV. Biomarker: predictive or prognostic? J Clin Oncol 2015;33:3968–71.
[3] Song HJ et al. A method of inferring the relationship between Biomedical entities through correlation analysis on text. Biomed Eng Online 2018;17:1–15.
[4] Hutchins G et al. Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. J Clin Oncol 2011;29:1261–70.
[5] Dodd LE et al. Blinded independent central review of progression-free survival in phase III clinical trials: Important design element or unnecessary expense? J Clin Oncol 2008;26:3791–6.
[6] Ward C et al. High WBP5 expression correlates with elevation of HOX genes levels and is associated with inferior survival in patients with acute myeloid leukaemia. Sci Rep 2020;10:1–12.
[7] Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep 2012;2:18–20.
[8] Liu R et al. Identifying critical transitions and their leading biomolecular networks in complex diseases. Sci Rep 2012;2.
[9] Scheffer M, Carpenter S, Foley JA, Folke C, Walker B. Catastrophic shifts in ecosystems. Nature 2001;413:591–6.
[10] Strogatz S, Friedman M, Mallinckrodt AJ, McKay S. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Comput Phys 1994;8:532.
[11] Liu R, Wang X, Aihara K, Chen L. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev 2014;34:455–78.
[12] Mojtahedi M et al. Cell fate decision as high-dimensional critical state transition. PLoS Biol 2016;14:1–28.
[13] Richard A et al. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLoS Biol 2016;14:1–35.
[14] Joost Lesterhuis W et al. Dynamic versus static biomarkers in cancer immune checkpoint blockade: Unravelling complexity. Nat Rev Drug Discov 2017;16:264–72.
[15] Chen P, Liu R, Chen L, Aihara K. Identifying critical differentiation state of MCF-7 cells for breast cancer by dynamical network biomarkers. Front Genet 2015;6:1–7.
[16] Chen P, Liu R, Li Y, Chen L. Detecting critical state before phase transition of complex biological systems by hidden Markov model. Bioinformatics 2016;32:2143–50.
[17] Liu X, Liu R, Zhao XM, Chen L. Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers. BMC Med Genomics 2013;6:1–10.
[18] Liu R, Zhong J, Yu X, Li Y, Chen P. Identifying critical state of complex diseases by single-sample-based hidden markov model. Front Genet 2019;10:1–10.
[19] Zhong J, Liu R, Chen P. Identifying critical state of complex diseases by single-sample Kullback-Leibler divergence. BMC Genomics 2020;21:1–15.
[20] Liu X et al. Quantifying critical states of complex diseases using single-sample dynamic network biomarkers. PLoS Comput Biol 2017;13:1–21.
[21] Zhang Y, Chang X, Liu X. Inference of gene regulatory networks using pseudo-time series data. Bioinformatics 2021;37:2423–31.
[22] Coleto-Alcudia V, Vega-Rodríguez MA. A metaheuristic multi-objective optimization method for dynamical network biomarker identification as pre-disease stage signal. Appl Soft Comput 2021;109:107544.
[23] Guo WF et al. Performance assessment of sample-specific network control methods for bulk and single-cell biological data analysis. PLoS Comput Biol 2021;17.
[24] Chen H, Song Z, Qian M, Bai C, Wang X. Selection of disease-specific biomarkers by integrating inflammatory mediators with clinical informatics in AECOPD patients: A preliminary study. J Cell Mol Med 2012;16:1286–97.
[25] Chen H, Wang Y, Bai C, Wang X. Alterations of plasma inflammatory biomarkers in the healthy and chronic obstructive pulmonary disease patients with or without acute exacerbation. J Proteomics 2012;75:2835–43.
[26] Yang B et al. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat Commun 2018;9.
[27] Liu R, Chen P, Chen L. Single-sample landscape entropy reveals the imminent phase transition during disease progression. Bioinformatics 2020;36:1522–32.
[28] Koizumi K et al. Identifying pre-disease signals before metabolic syndrome in mice by dynamical network biomarkers. Sci Rep 2019;9:1–11.
[29] Sun Y et al. Identifying critical states of hepatocellular carcinoma based on landscape dynamic network biomarkers. Comput Biol Chem 2020;85:107202.
[30] Gao Y et al. Identification of HCC-related genes based on differential partial correlation network. Front Genet 2021;12:1–10.
[31] Liu L et al. Identification of early warning signals at the critical transition point of colorectal cancer based on dynamic network analysis. Front Bioeng Biotechnol 2020;8:1–13.
[32] Mendes-Pereira AM et al. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. Proc Natl Acad Sci U S A 2012;109:2730–5.
[33] Verhagen CVM et al. Extent of radiosensitization by the PARP inhibitor olaparib depends on its dose, the radiation dose and the integrity of the homologous recombination pathway of tumor cells. Radiother Oncol 2015;116:358–65.
[34] Khrunin AV, Khokhrin DV, Moisseev AA, Gorbunova VA, Limborska SA. Pharmacogenomic assessment of cisplatin-based chemotherapy outcomes in ovarian cancer. Pharmacogenomics 2014;15:329–37.
[35] Hu J et al. Dynamic network biomarker of pre-exhausted CD8+ T cells contributed to T cell exhaustion in colorectal cancer. Front Immunol 2021;12:1–13.
[36] Huisheng L et al. Single-cell transcriptome analysis reveals DHX9 in mature B cell as a dynamic network biomarker before lymph node metastasis in colorectal adjacent tissues. Mol Ther - Oncolytics 2021.
[37] Zhang Y et al. Single-cell RNA-sequencing atlas reveals an MDK-dependent immunosuppressive environment in ErbB pathway-mutated gallbladder cancer. J Hepatol 2021;1–14.
[38] Jiang Z et al. SMAD7 and SERPINE1 as novel dynamic network biomarkers detect and regulate the tipping point of TGF-beta induced EMT. Sci Bull 2020;65:842–53.
[39] Liu R et al. Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. J Mol Cell Biol 2019;11:649–64.
[40] Wang P, Chen L. Critical transitions and tipping points in EMT. Quant Biol 2020;8:195–202.
[41] Li M, Zeng T, Liu R, Chen L. Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: Study of type 2 diabetes by cross-tissue analysis. Brief Bioinform 2014;15:229–43.
[42] Liu R et al. Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy. Sci Bull 2021.
[43] Schoen-Angerer V. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-ocsophageal junction cancer (ToGA): a phase 3, open-label, randomized controlled trial. J The Lancet 2010;376:1302.
[44] Untch M et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. J New England J Med 2005;353:1659–72.
[45] Locker GY et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol 2006;24:5313–27.
[46] Harris L et al. American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287–312.
[47] Xu J et al. Identifying dysregulated lncRNA-associated ceRNA network biomarkers in CML based on dynamical network biomarkers. Biomed Res Int 2020.