Journal of Molecular Cell Biology (2022), 14(8), mjac052 https://doi.org/10.1093/jmcb/mjac052
Published online September 7, 2022

Article

The single-sample network module biomarkers (sNMB) method reveals the pre-deterioration stage of disease progression

Jiayuan Zhong[1,2], Huisheng Liu[4], and Pei Chen[2,3,*]

[1] School of Mathematics and Big Data, Foshan University, Foshan 528000, China
[2] School of Mathematics, South China University of Technology, Guangzhou 510640, China
[3] Pazhou Lab, Guangzhou 510330, China
[4] School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
* Correspondence to: Pei Chen, E-mail: chenpei@scut.edu.cn

Edited by Luonan Chen

The progression of complex diseases generally involves a pre-deterioration stage that occurs during the transition from a healthy state to disease deterioration, at which a drastic and qualitative shift occurs. The development of an effective approach is urgently needed to identify such a pre-deterioration stage or critical state just before disease deterioration, which allows the timely implementation of appropriate measures to prevent a catastrophic transition. However, identifying the pre-deterioration stage is a challenging task in clinical medicine, especially when only a single sample is available for most patients, which is responsible for the failure of most statistical methods. In this study, a novel computational method, called single-sample network module biomarkers (sNMB), is presented to predict the pre-deterioration stage or critical point using only a single sample. Specifically, the proposed single-sample index effectively quantifies the disturbance caused by a single sample against a group of given reference samples. Our method successfully detected the early warning signal of the critical transitions when applied to both a numerical simulation and four real datasets, including acute lung injury, stomach adenocarcinoma, esophageal carcinoma, and rectum adenocarcinoma. In addition, it provides signaling biomarkers for further practical application, which helps to discover prognostic indicators and reveal the underlying molecular mechanisms of disease progression.

Keywords: critical point, pre-deterioration stage, critical transition, dynamic network biomarker (DNB), single-sample network module biomarkers (sNMB)

Introduction

An abrupt switch to a contrasting state through a critical transition occurs in many complex systems, such as ecosystems (Beck et al., 2018; Chen et al., 2019), financial systems (Drehmann and Juselius, 2014; Huang et al., 2017), climate systems (Lenton, 2011), and infectious disease spreading (Scarpino and Petri, 2019). Similarly, the progression of many complex diseases is not always smooth but occasionally abrupt; i.e. there is a so-called critical point at which a sudden and qualitative state transition may occur (Chen et al., 2012; Liu et al., 2021). Accordingly, regardless of specific differences in clinical symptoms and biological processes, disease progression can be roughly classified into three stages, i.e. a before-deterioration stage, a pre-deterioration stage, and a deterioration stage (Figure 1A). The before-deterioration stage is viewed as a relatively healthy state with stability and high resilience. The pre-deterioration stage refers to a critical transition before switching to the onset or deterioration of symptoms and is characterized by low resilience and high susceptibility. The deterioration stage is another stable state with high resilience after the catastrophic transition to the deterioration of the disease. In contrast to the irreversible deterioration stage, the pre-deterioration stage is sensitive to perturbation and is thus usually considered to be reversible to the before-deterioration stage by appropriate intervention strategies. However, many complex diseases, such as cancers, are difficult to cure unless

Received January 27, 2022. Revised May 27, 2022. Accepted September 2, 2022.

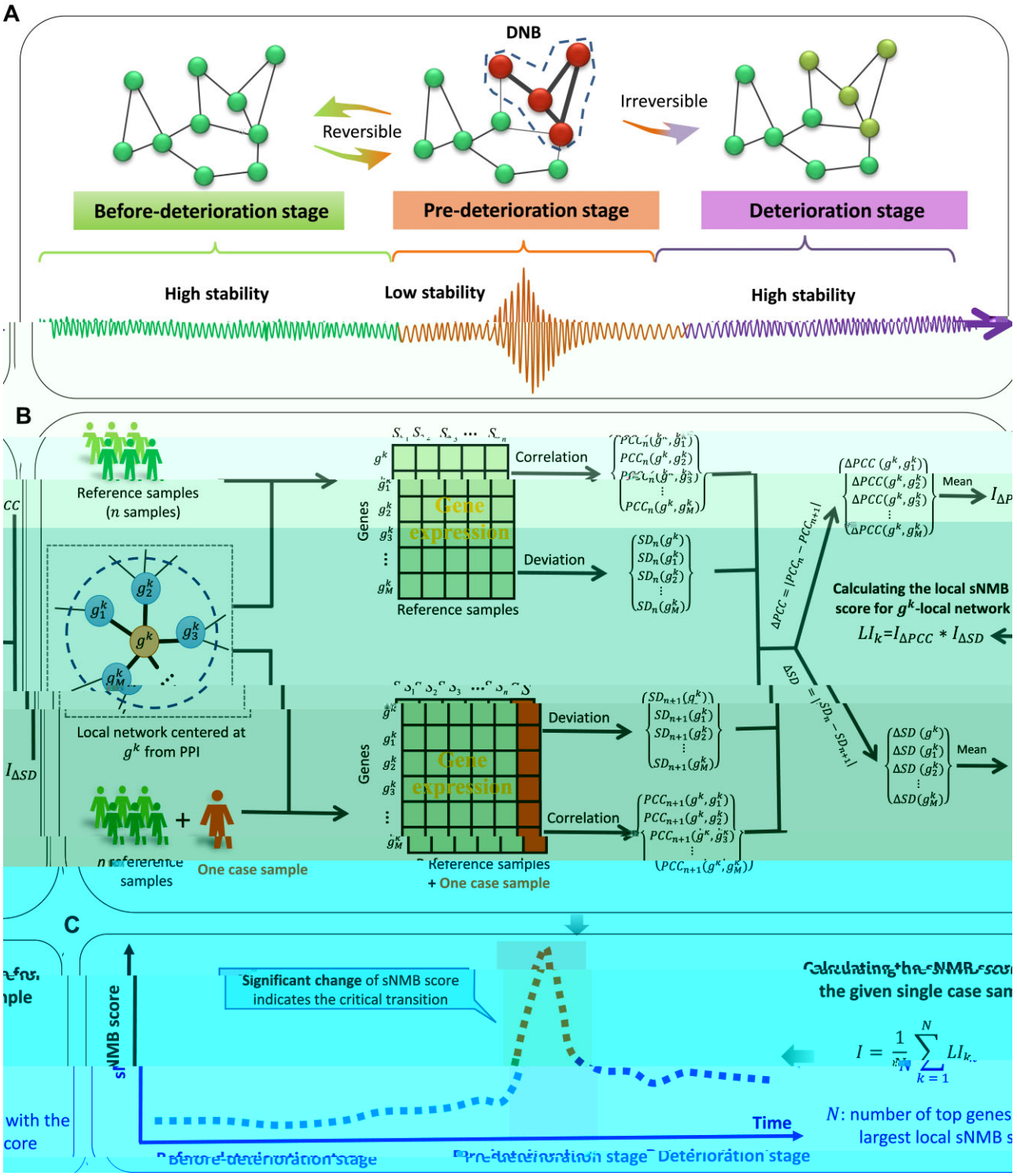
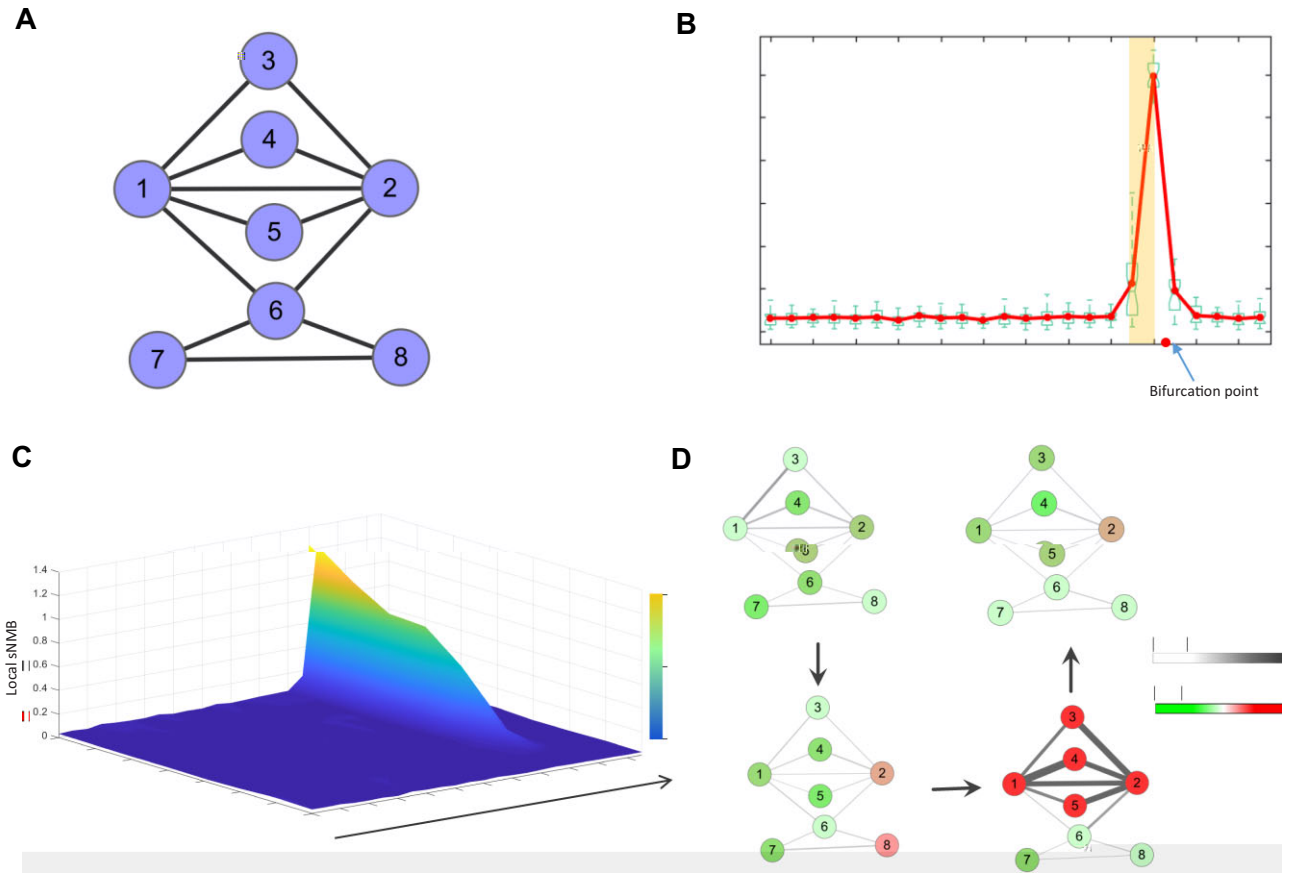
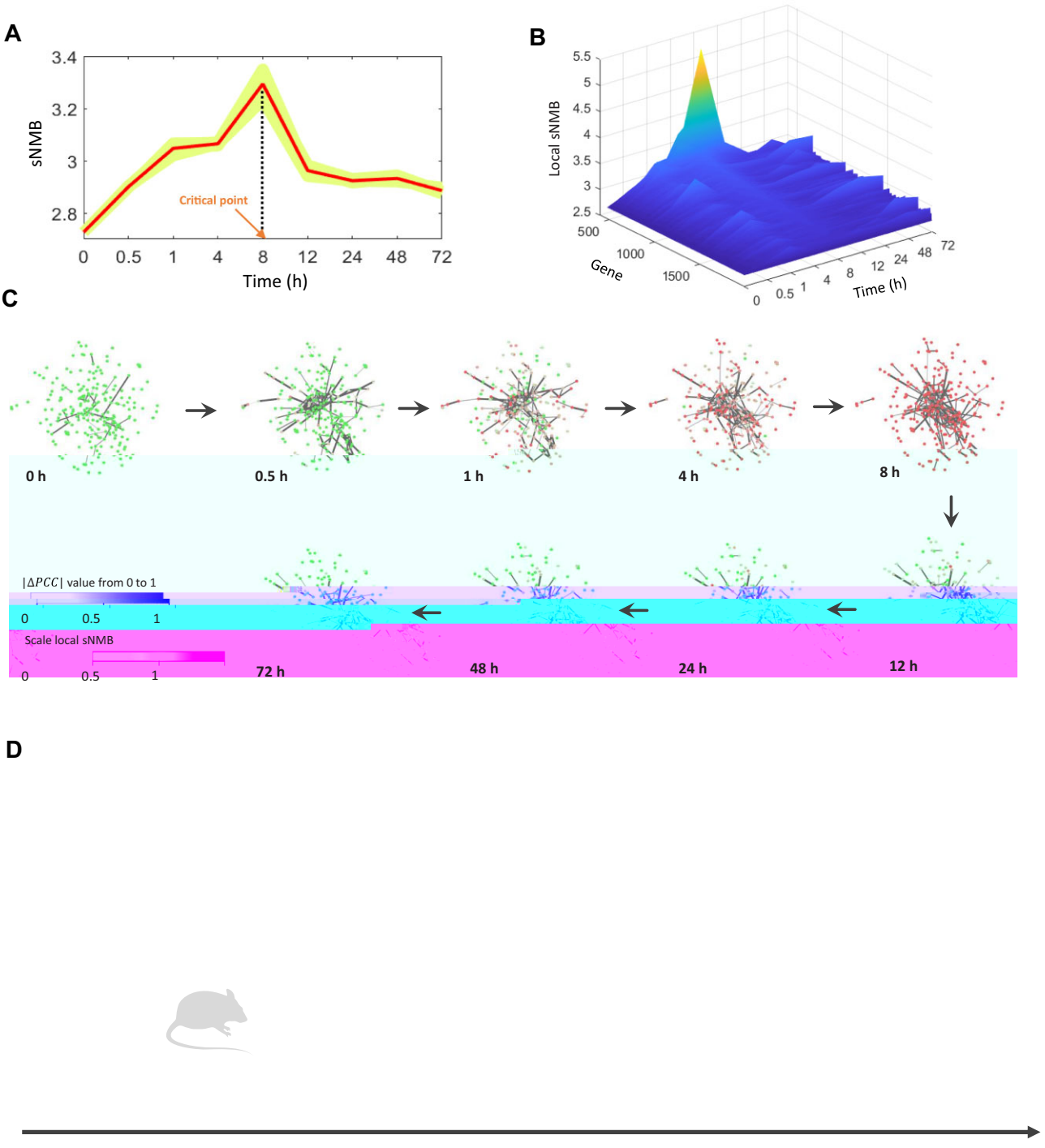


Figure 1 Schematic illustration for identifying the pre-deterioration stage based on the sNMB score. (**A**) The progression of complex diseases is roughly divided into three stages, including a before-deterioration stage with high stability, an unstable pre-deterioration stage, and another deterioration stage with high stability. The pre-deterioration stage is a critical state just before disease onset or deterioration and is usually considered to be reversible to the before-deterioration stage through appropriate intervention, since it is sensitive to perturbation. (**B**) The sNMB score is calculated based on a single case sample and is capable of quantifying the statistical disturbance yielded from the single case sample against a group of given reference samples collected from a relatively healthy population. (**C**) The significant change in sNMB signals the pre-deterioration stage; i.e. the sNMB score sharply increases when the system is close to the critical point.

they are diagnosed at an early stage; i.e. missing the best time for preemptive clinical interventions results in patients having no choice but to undergo high-risk therapies. Unfortunately, because many complex diseases cause few symptoms during the early stages, the disease has already reached an advanced stage when the pathological symptoms are easily diagnosed clinically (Miller et al., 2019). Therefore, identifying the pre-deterioration stage is of great significance for preventing or delaying the occurrence of catastrophic deterioration. However, it is challenging to accurately identify the tipping point or pre-deterioration stage of complex diseases, because complex biological systems exhibit relatively little state change before approaching the catastrophic transition.

By exploiting the information of differential expression between the before-deterioration and pre-deterioration stages, traditional biomarkers mainly focused on distinguishing the deterioration stage rather than identifying the pre-deterioration stage, which is similar to the before-deterioration stage in terms of phenotype and gene expression. Recently, a novel concept of the dynamic network biomarker (DNB) (Chen et al., 2012) provided




A
sNMB
Critical point
Time (h)
B
Local sNMB
Gene
Time (h)
C
0 h
0.5 h
1 h
4 h
8 h
|ΔPCC| value from 0 to 1
Scale local sNMB
72 h
48 h
24 h
12 h
D

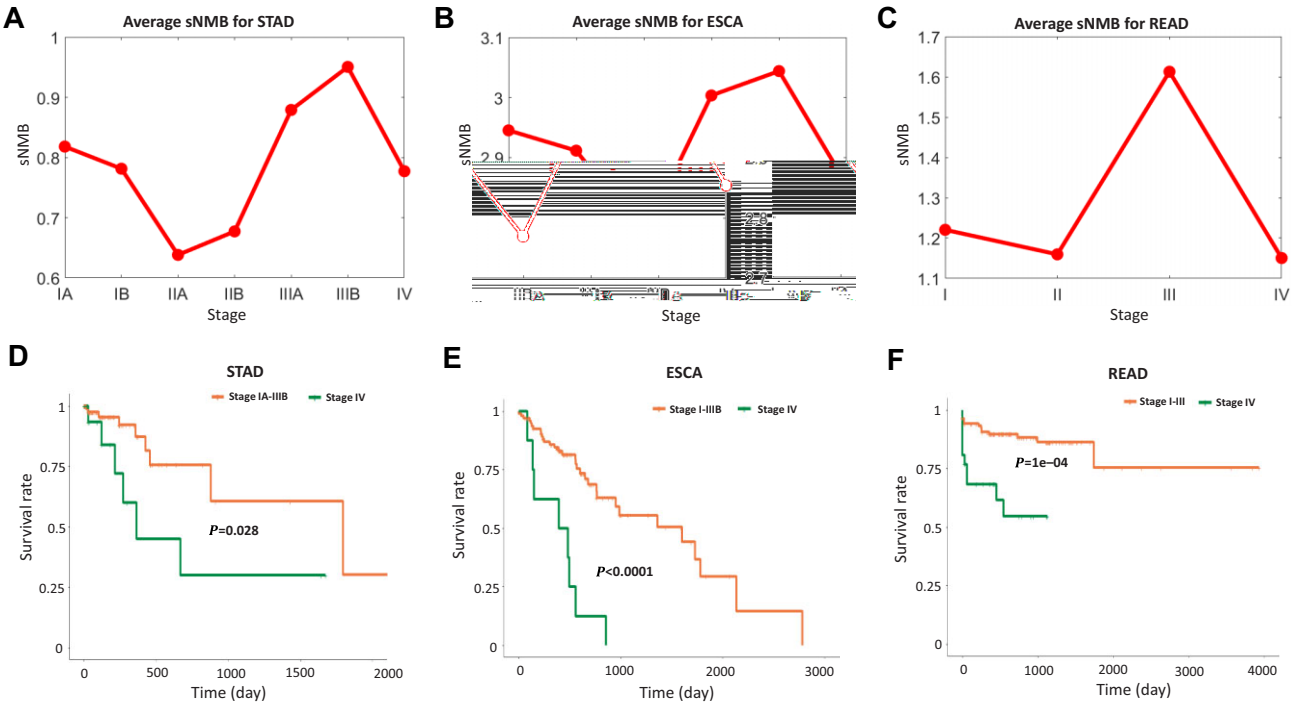


Figure 4 Identification of the critical transition of tumor distant metastasis in three cancers. (**A**–**C**) Identifying the critical transition for STAD (**A**), ESCA (**B**), and READ (**C**). (**D**–**F**) Comparing survival curves between the before-transition stage and the after-transition stage in STAD (**D**), ESCA (**E**), and READ (**F**).

Clearly, a notable change in the network structure occurs at 8 h, indicating an upcoming critical transition. These results are consistent with the observation in the original experiment (Figure 3D); i.e. the severe phosgene-induced acute lung injury occurred around 12 h, and ∼50%–60% of deaths were observed around 24 h (Sciuto et al., 2005).

Identifying the critical state for tumor diseases

To validate the effectiveness of the sNMB method in detecting the early warning signal of the pre-deterioration stage for tumor diseases, the proposed method was applied to three tumor datasets (ESCA, READ, and STAD) from TCGA. The tumor-adjacent samples that represent the relatively healthy condition were viewed as reference samples, and then the sNMB score of each tumor sample was calculated according to the algorithm described in Materials and methods. At each stage, the mean sNMB value was adopted to quantitatively measure the pre-deterioration stage of tumor diseases. By analysis with the proposed method, the pre-deterioration stage was identified in stage IIIB for ECSA, stage III for READ, and stage IIIB for STAD (Figure 4A–C). To validate the identification of the critical stage, prognostic analysis of before-transition and after-transition samples was performed and compared through Kaplan–Meier (log-rank) survival analysis (Figure 4D–F; Supplementary Figure S2). Specifically, compared with the samples from the after-transition stage, there was usually a higher life expectancy for before-transition samples.

For ESCA, as shown in Figure 4A, a sudden increase in the sNMB score was detected in stage IIIB, after which there existed a tumor invading other adjacent structures, such as the aorta and vertebral body, and distant metastasis occurred at stage IV (Stahl et al., 2010). Figure 4D shows that there is a significant difference ($P = 0.028$) between the survival curves of the before-transition samples and the after-transition samples. Clearly, samples from stages I–IIIB present significantly longer survival periods than samples from stage IV. For the samples from only two stages (stages IIIB and IV) around the critical stage, the survival time of stage IIB samples was much longer than that of stage IIIA samples ($P = 0.0215$; Supplementary Figure S3A). In addition, there was a statistically non-significant difference among the survival curves of samples from before the critical stage ($P = 0.295$; Supplementary Figure S3B). In Figure 4B, the peak of the sNMB value appears at stage IIIB, implying an imminent critical transition of STAD after stage IIIB. References show that stage IV is a severely deteriorated stage, in which the tumor has spread to nearby tissues or metastasized to other organs and ultimately causes distant metastasis (Kwon, 2011). Figure 4E shows that there is a significant difference ($P < 0.0001$) between the survival periods of the two groups of samples, i.e. samples from the before-transition stage (stages IA–IIIB) and samples from the after-transition stage (stage IV)

survival curves among samples from the before-transition period (stages IA–IIIA). For READ, as shown in Figure 4C, the drastic transitions of the sNMB score appear in stage III, after which the tumor invades other parts of the human body and distant metastasis occurs at stage IV (Jessup et al., 2011). Figure 4F shows that there was a significant difference ($P = 1e{-}04$) between the survival curves of samples from stages I–III and samples from stage IV. The before-transition samples showed a significantly longer survival time than the after-transition samples. For the samples solely from two stages (stages III and IV) around the critical state, the survival periods of stage III samples were significantly longer than those of stage IV samples ($P = 0.0167$; Supplementary Figure S3E). There was no significant difference ($P = 0.819$; Supplementary Figure S3F) in survival curves among samples before the critical stage (stages I–II). These results demonstrate that the sNMB score can detect the early warning signals of a critical transition of survival time, i.e. the critical transition associated with distant metastasis at stage IV can be identified by the sNMB score.

Revealing the potential □

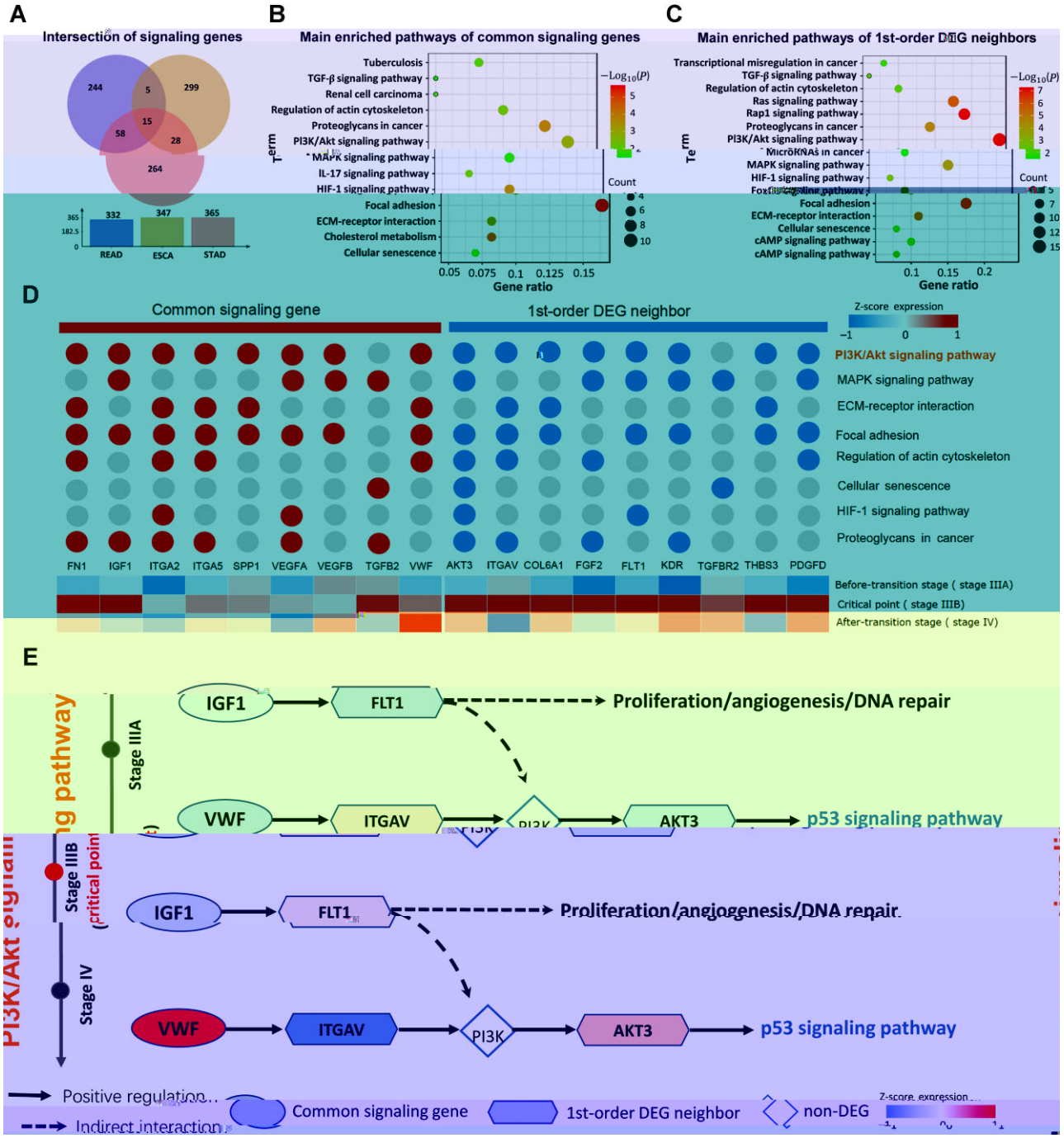


Figure 5 Functional analysis of common signaling genes in different tumor datasets. (**A**) Venn diagrams for the intersection of signaling genes in three different cancers. There were 106 common signaling genes shared among the three datasets. (**B**) Dot plot of KEGG enrichment analyses for 106 common signaling genes. (**C**) Dot plot of KEGG enrichment analyses for 1st-order DEG neighbors in the STAD dataset. (**D**) The important common pathways are shared between common signaling genes and their 1st-order DEG neighbors. The common signaling genes mapping into the pathway are depicted with red circles, while their 1st-order DEG neighbors mapping into the pathway are depicted with blue circles. (**E**) For the TCGA-STAD dataset, the underlying signaling mechanisms involve both common signaling genes and their 1st-order DEG neighbors.

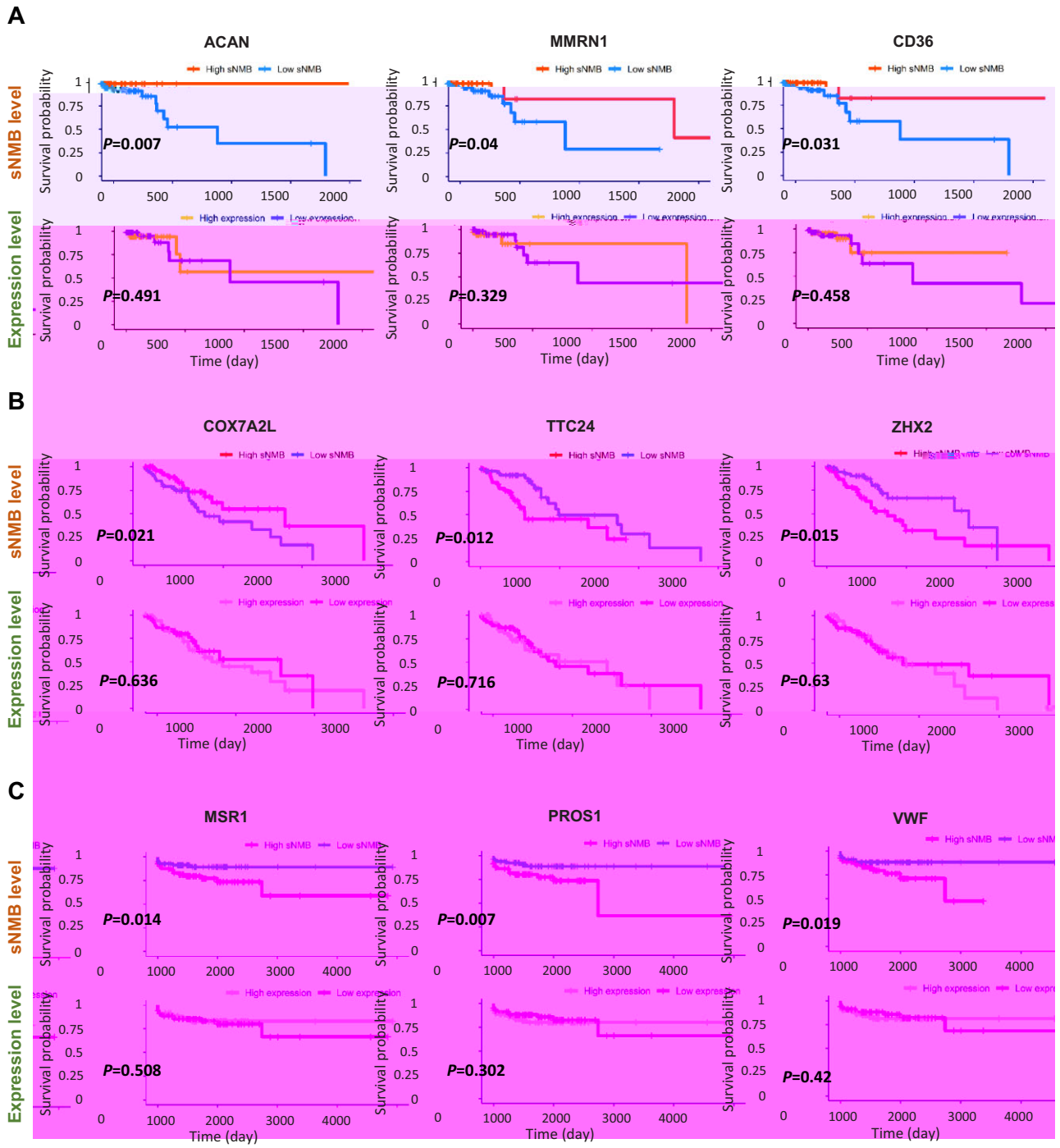


Figure 6 'Dark genes' are sensitive to the sNMB score. (**A**–**C**) Kaplan–Meier (log-rank) survival analysis of local sNMB values and gene expression for STAD (**A**), ESCA (**B**), and READ (**C**). The non-differential genes sensitive to the sNMB score are considered 'dark genes'. The dark genes perform well in prognosis and are strongly related to patient survival at the sNMB level but not at the gene expression level.

ESCA, and READ), the drastic transitions of the sNMB score signal the pre-disease stage before distant metastasis in stage IV. The successful detection of critical points for these biological datasets validates the effectiveness of the sNMB approach in identifying the criticality of complex diseases solely based on a single sample.

There are a few advantages to the sNMB method. First, compared with traditional biomarkers that aim to diagnose the deterioration stage based on the differential information of gene expression, the proposed method is capable of predicting the pre-deterioration stage based on the information of differential networks among biomolecules. Second, against a group of

give[redacted]erence samples, t[redacted]MB m[redacted]d can identify the pre-dete[redacted]tion stage or crit[redacted]oint us[redacted]only a single sample, whil[redacted] conventional D[redacted]ethod [redacted]ires multiple samples at ea[redacted]ime point to eva[redacted]e its th[redacted]statistical indices. In addi[redacted] the sNMB metho[redacted] only c[redacted]ts general early warning s[redacted]ls of critical tran[redacted] into t[redacted]eterioration stage but also [redacted]s that are involved in key [redacted]y combining with dynam[redacted]2020), it may help to identify the future critica[redacted]es based on omics data. Third, the sNMB method helps[redacted]veal the dark genes, which are [redacted]sion but sensitive to sNMB and [redacted]ly, it should be noted that sNMB [redacted] implies that the sNMB strategy does not involve feature selection or model/parameter trTc7Tc7]TJ7/GSgs7hTf7...(ba)(sed)]TJ7/GS

[redacted]

[Step 5] Calculating the sNMB score for the single sample at time point t. The sNMB score for the single sample is calculated based on a group of genes with the largest local sNMB score, i.e.

$$I_t = \frac{1}{N}\sum_{k=1}^{N} LI_k \quad (4),$$

where the constant N is an adjustable parameter representing the number of the top 5% of genes with the largest local sNMB scores. I_t is applied to quantify the overall perturbation caused by a single case sample.

According to the DNB theory, when the system approaches the critical stage, the sub-network (local network) composed of DNB molecules exhibits significant changes in terms of variance and correlation, thus resulting in significantly differential information between the pre-deterioration stage and the before-deterioration stage. Similar to previous studies (Zeng et al., 2014a,b), such a sub-network is viewed as a module, which can be utilized to detect the early warning signal of critical transition. Therefore, the composite indicator I_t would abruptly increase when the system is near the tipping point, signaling an upcoming critical transition.

Data processing and functional analysis

The proposed sNMB approach was applied to a numerical simulation and four real datasets, i.e. STAD, ESCA, and READ datasets from TCGA database (http://cancergenome.nih.gov) and an acute lung injury (GSE2565) dataset from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). The tumor datasets are composed of both tumor and tumor-adjacent samples. The tumor samples are grouped into different stages according to the stage information of TCGA, and the samples lacking corresponding information are ignored. The cancer samples were grouped into seven stages (stages IA, IB, IC, IIA, IIB, IIIA, IIIB, and IV) for STAD, six stages (stages I, IB, IC, IIA, IIB, IIIA, IIIB, and IV) for ESCA, and four stages (stages I, II, III, and IV) for READ. The details of the sampling conditions are given in Supplementary Table S1. The tumor-adjacent samples that represent the relatively healthy condition are viewed as reference samples. For all these datasets, we discarded the probes without corresponding NCBI Entrez gene symbols. For each gene mapped by multiple probes, the average value was taken as its gene expression.

The analysis of pathways was performed with the KEGG database (https://www.kegg.jp). The enrichment analysis was performed by Metascape (Zhou et al., 2019) and the ClusterProfiler package (Yu et al., 2012). The functional results were based on web service tools from the Gene Ontology Consortium (http://geneontology.org) and client software from Ingenuity Pathway Analysis (IPA, http://www.ingenuity.com/products/ipa). The networks were visualized using Cytoscape.

Availability of data and materials

STAD, ESCA, and READ datasets are available from TCGA database (http://cancergenome.nih.gov). Acute lung injury dataset (GSE2565) is available from NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo). The source code of the algorithm is provided at https://github.com/zhongjiayuan/sNMB_project.

Supplementary material

Supplementary material is available at *Journal of Molecular Cell Biology* online.

Funding

This work

Lenton, T.M. (2011). Early warning of climate tipping points. Nature Clim. Change *1*, 201–209.

Lesterhuis, W.J., Bosco, A., Millward, M.J., et al. (2017). Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. Nat. Rev. Drug Discov. *16*, 264–272.

Liu, R., Chen, P., and Chen, L. (2020). Single-sample landscape entropy reveals the imminent phase transition during disease progression. Bioinformatics *36*, 1522–1532.

Liu, R., Zhong, J., Hong, R., et al. (2021). Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy. Sci. Bull. *66*, 2265–2270.

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

Liu, X., Chang, X., Liu, R., et al. (2017). Quantifying critical states of complex diseases using single-sample dynamic network biomarkers. PLoS Comput. Biol. *13*, e1005633.

Miller, K.D., Nogueira, L., Mariotto, A.B., et al. (2019). Cancer treatment and survivorship statistics, 2019. CA Cancer J. Clin. *69*, 363–385.

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

Roberts, P.J., and Der, C.J. (2007). Targeting the Raf–MEK–ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene *26*, 3291–3310.

Scarpino, S.V., and Petri, G. (2019). On the predictability of infectious disease outbreaks. Nat. Commun. *10*, 898.

Scheffer, M., Carpenter, S., Foley, J.A., et al. (2001). Catastrophic shifts in ecosystems. Nature *413*, 591–596.

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

Stahl, M., Budach, W., Meyer, H.J., et al. (2010). Esophageal cancer: clinical practice guidelines for diagnosis, treatment, and follow-up. Ann. Oncol. *21 Suppl 5*, v46–v49.

Wang, Y., Shi, M., Yang, N., et al. (2020). GPR115 contributes to laryngeal squamous cell carcinoma metastasis associated with LAMC2 and predicts a poor prognosis. Front. Oncol. *10*, 2414.

Yu, G., Wang, L.G., Han, Y., et al. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Yu, X., Zhang, J., Sun, S., et al. (2017). Individual-specific edge-network analysis for