


TPD: a web tool for tipping-point detection based on dynamic network biomarker

Pei Chen, Jiayuan Zhong, Kun Yang, Xuhang Zhang, Yingqi Chen and Rui Liu 

Corresponding author. Rui Liu. E-mail: scliurui@scut.edu.cn

Abstract

Tipping points or critical transitions widely exist during the progression of many biological processes. It is of great importance to detect the tipping point with the measured omics data, which may be a key to achieving predictive or preventive medicine. We present the tipping point detector (TPD), a web tool for the detection of the tipping point during the dynamic process of biological systems, and further its leading molecules or network, based on the input high-dimensional time series or stage course data. With the solid theoretical background of dynamic network biomarker (DNB) and a series of computational methods for DNB detection, TPD detects the potential tipping point/critical state from the input omics data and outputs multifarious visualized results, including a suggested tipping point with a statistically significant *P* value, the identified key genes and their functional biological information, the dynamic change in the DNB/leading network that may drive the critical transition and the survival analysis based on DNB scores that may help to identify ‘dark’ genes (nondifferential in terms of expression but differential in terms of DNB scores). TPD fits all current browsers, such as Chrome, Firefox, Edge, Opera, Safari and Internet Explorer. TPD is freely accessible at <http://www.rpcomputationalbiology.cn/TPD>.

Keywords: dynamic network biomarker, critical transition, tipping-point detection, leading network, web tool

Introduction

A complex biological process is generally regarded as the progression of a high-dimensional nonlinear dynamic system. Such a continuum of progressive changes always occurs in the microenvironment and macroenvironment of a certain organ or the whole organism and is involved in complicated regulations among massive biomolecules [1–5]. The time evolution or dynamic change of many biological systems, such as disease progression, the cell fate commitment, is not always smooth but occasionally abrupt; that is, there is a tipping point or critical state during such a process at which the system state shifts irreversibly from one state (e.g. normal state) to another state (e.g. disease state) [6–10]. There are many studies investigated the microscopic mechanisms of complex diseases [11–13], and a group of researchers put an emphasis on network biomarkers in a biological process [14–18]. Regardless of differences in specific biochemical environments, the time evolution of a complex biological system is usually modeled as a time-dependent nonlinear dynamic system, in which the abrupt state transition is viewed as the phase shift at a bifurcation point [6]. The progression of a complex biological process with a tipping point can be roughly divided into three states or stages: the before-transition state when the system is far away from the bifurcation point, the critical/pre-transition state

when the system approaches the bifurcation point and the after-transition state when the system passes the bifurcation point [6, 19–21]. Recently, there are some well-designed methods proposed to capture the signal of state transitions, such as BioTIP, which is capable of inferring lineage-determining transcription factors governing critical transition from single-cell transcriptomes [22].

Based on a solid theoretical background, the dynamic network biomarker (DNB) identifies the critical state during biological processes or observed symptoms among diseases [1, 19], and the progression of a biological system is divided into three stages or states from a computational point of view i.e. a before-transition state, a critical state and an after-transition state. For a given omics time series, the design of a suitable tipping point detector (TPD) based on DNB involves the consideration of the following factors:

- 1) When a complex system is before the tipping point/critical state, the system state is stable with high resilience, and the curve of a TPD index is expected to be steady and smooth; when the system is near the tipping point, the system state is unstable with low resilience, and there is a sudden increase in the corresponding detector index; when it is after the tipping point, the system is in another stable state, and the

Pei Chen received her BS and MS degrees from Peking University, and PhD degree from South China University of Technology. Currently she is an Associate Professor at the School of Mathematics, South China University of Technology. Her research interest includes deep learning, data mining and computational biology.

Jiayuan Zhong is a Research Fellow in School of Mathematics and Big Data at Foshan University. His research interest mainly focuses on developing computational approaches to detect the tipping points of nonlinear dynamical systems.

Kun Yang received his master degrees in South China University of Technology, and is a software engineer currently.

Xuhang Zhang received his master degrees in South China University of Technology, and is a software engineer currently.

Yingqi Chen received his master degrees in South China University of Technology, and is a software engineer currently.

Rui Liu is a full Professor at the School of Mathematics, South China University of Technology, and also affiliated with Pazhou Lab, Guangzhou. He received the BS and PhD degrees in applied mathematics from Peking University. His research interest includes nonlinear dynamics, modeling and computational methods.

Received: May 22, 2022. **Revised:** August 4, 2022. **Accepted:** August 16, 2022

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

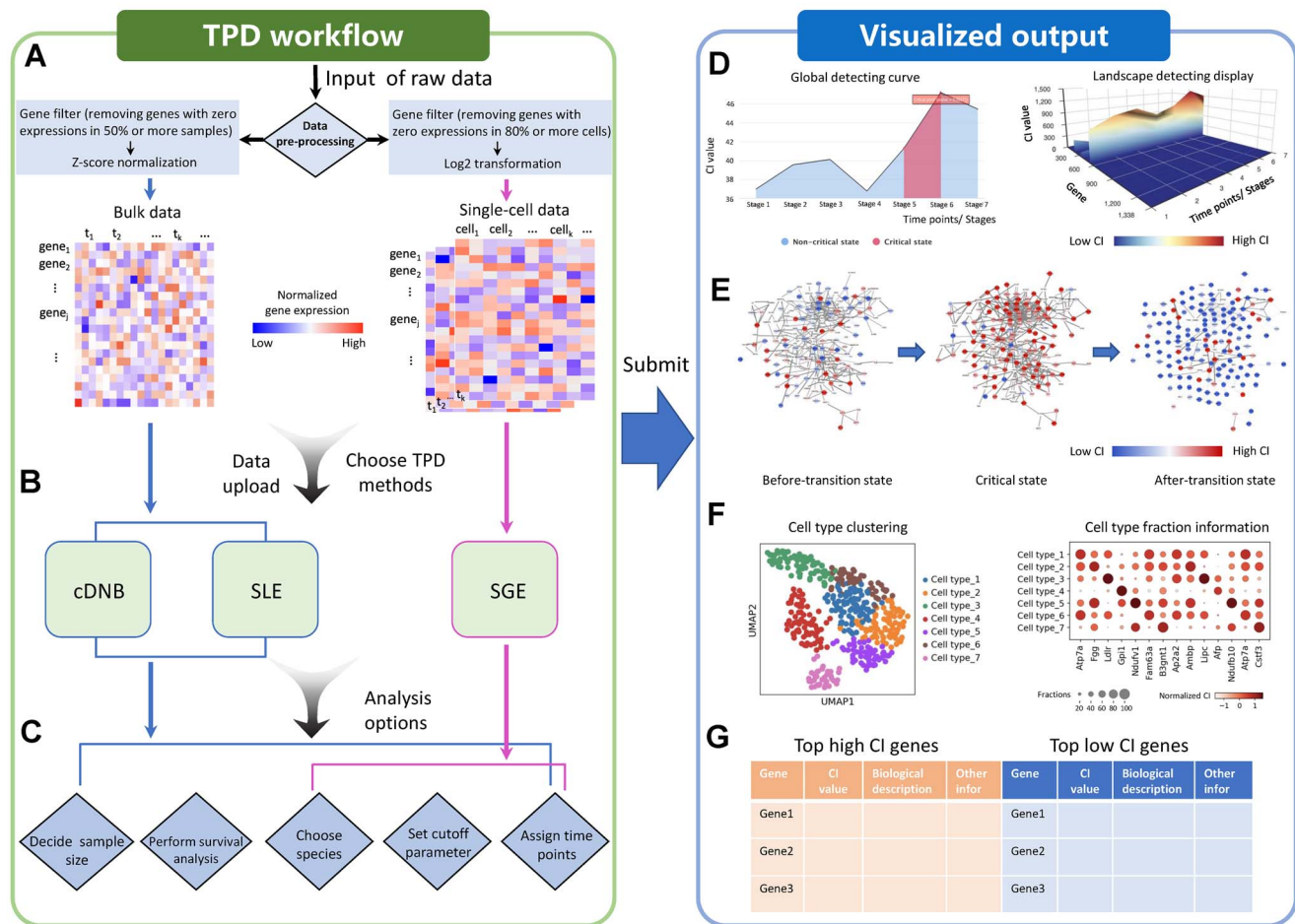


Figure 1. Schematic presentation of the TPD toolset. TPD consists of two parts: input workflow (A–C) and visualized outputs (D–G). (A) TPD accepts gene expression microarray data including bulk and single-cell data as input after pre-processing steps. (B) There are three detection methods based on the DNB concept provided by TPD for different data types (e.g. bulk or single-cell data). (C) For an input dataset, TPD provides options for users to assign samples to time points, set the cutoff parameter, choose species (*Homo sapiens* or *Mus musculus*) and decide whether to perform survival analysis, etc. (D) The tipping point is indicated by both the global detection curve and the landscape display of local genes. (E) Dynamic changes in the molecular network, which consists of DNB genes, evolving from the before-transition state and critical state to the after-transition state. (F) For the scRNA-seq data analysis, TPD outputs the analysis results, including cell type clustering and fraction information. (G) Finally, the biological description of signaling genes with the top highest and lowest CI values is shown.

corresponding curve of the detector index becomes smooth again [1, 9, 19, 23].

- When the complex system is in a critical state, among all observed genes, there exists a dominant group of biomolecules defined as the DNB, which are intuitively a group of strongly correlated molecules with simultaneously fluctuating expressions. This group of biomolecules can be identified together with the tipping point by suitable detection methods [1, 24].
- The identified key genes (i.e. signaling genes with the top highest and lowest DNBs or modified DNB scores) may play important roles in the progression of the complex biological system and are involved in the dynamic changes in the network, survival analysis, cell type clustering results, etc. [25–28].

Based on these requirements, we developed the TPD, which is the first publicly accessible web tool for the detection of the tipping point and the analysis of the identified DNB. An overview of TPD workflow is demonstrated in Figure 1.

Materials and methods

TPD provides three different methods for tipping point detection with the solid DNB theoretical background i.e. conventional dynamic network biomarker (cDNB) [1, 19] and single-sample landscape entropy (SLE) [24] for bulk datasets and single-cell graph entropy (SGE) [28] for single-cell RNA sequencing (scRNA-seq) datasets (Figure 1B), which quantitatively identify each candidate tipping point with a method-specific composite index (CI) (Figure 1D). A detailed description of these methods is provided as below.

cDNB

Given a bulk dataset, the cDNB algorithm in the proposed webtool TPD is for detecting the dynamic differences in statistical indices. First, TPD mapped the gene expression to the default *Homo sapiens* or *Mus musculus* protein–protein interaction (PPI) network. Second, the network was partitioned into many local networks. Each local network contained a center node/gene and all of its 1st-order neighbors based on the network structure. The local-network

index I -score of a center node at time point t for a local network with $M + 1$ members (i.e. one center node with M 1st-order neighboring nodes) was then calculated through the following definition:

$$I_t = \text{SD}_{\text{in}}(t) \frac{\text{PCC}_{\text{in}}(t)}{\text{PCC}_{\text{out}}(t)}, \quad (1)$$

where

$$\text{SD}_{\text{in}}(t) = \frac{\sum_{i=1}^n \text{SD}(\mathbf{g}_i(t))}{M} \quad (2)$$

is the average differential standard deviation (SD) of the nodes inside the local network. Vector $\mathbf{g}_i(t)$ stands for the gene expressions of a nodes i inside the local network;

$$\text{PCC}_{\text{in}}(t) = \frac{\sum_{i,j=1, i \neq j}^n |\text{PCC}(\mathbf{g}_i(t), \mathbf{g}_j(t))|}{M(M-1)} \quad (3)$$

is the average value of Pearson correlation coefficients (PCCs) (in absolute value) inside the local network i.e. both nodes i and j are in the same local network;

$$\text{PCC}_{\text{out}}(t) = \frac{\sum_{i,l=1, i \neq l}^n |\text{PCC}(\mathbf{g}_i(t), \mathbf{g}_l(t))|}{M(M-1)} \quad (4)$$

is the average PCC (in absolute value) between a member (node i) in the local network and that (node l) outside the local network.

Thus, the network-based index, which is also called the global composite cDNB score, I_t , can quantitatively identify the criticality of the state for each DNB node. Each node has an I_t value, and hence those I_t scores for all of nodes with the time evolution construct a landscape. When the system approaches the critical state, I_t of each DNB node increases drastically based on the three statistic conditions of DNB, whereas I_t of other non-DNB node may have no significant change.

SLE

Given a dataset with reference samples, such as the gene expressions of the healthy cohort, and one single case sample from a to-be-test individual, the SLE algorithm in the proposed webtool TPD is for determining whether the single case sample is from a critical/pre-disease state. The algorithm of SLE can be described as follows:

- 1) Map the genes to the PPI network (or other template networks), forming a global network N^G .
- 2) Extract each local network from the global network N^G , such that each local network N^k ($k = 1, 2, \dots, Q$) is centered at a gene \mathbf{g}^k , whose 1st-order neighbors $\{\mathbf{g}_1^k, \mathbf{g}_2^k, \dots, \mathbf{g}_M^k\}$ are the edges. There are totally Q local networks if there are Q genes in N^G .
- 3) For each local network N^k ($k = 1, 2, \dots, Q$) at a time point t , calculate the local entropy $H^n(k, t)$ based on n reference samples $\{\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_n(t)\}$ i.e.

$$H^n(k, t) = -\frac{1}{\log M} \sum_{i=1}^M p_i^n(t) \log p_i^n(t), \quad (5)$$

with

$$p_i^n(t) = \frac{|\text{PCC}^n(\mathbf{g}_i^k(t), \mathbf{g}^k(t))|}{\sum_{j=1}^M |\text{PCC}^n(\mathbf{g}_j^k(t), \mathbf{g}^k(t))|}, \quad (6)$$

where $\text{PCC}^n(\mathbf{g}_i^k(t), \mathbf{g}^k(t))$ represents the PCC between the expressions of center gene \mathbf{g}^k and that of a 1st-order neighbor \mathbf{g}_i^k based on n reference samples. In Equation (5), the superscript k records that the local network is centered at gene \mathbf{g}^k , the subscript n denotes the number of samples and constant M is the number of edges/neighbors in the local network N^k . In Equation (6), vectors $\mathbf{g}^k(t)$ and $\mathbf{g}_i^k(t)$ respectively denote the expressions of genes \mathbf{g}^k and \mathbf{g}_i^k at time point t .

- 4) For a single sample $\mathbf{s}_{\text{case}}(t)$ of an individual, mix it with n reference samples. Calculate the entropy $H^{n+1}(k, t)$ based on $n + 1$ mixed samples $\{\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_n(t), \mathbf{s}_{\text{case}}(t)\}$

$$H^{n+1}(k, t) = -\frac{1}{\log M} \sum_{i=1}^M p_i^{n+1}(t) \log p_i^{n+1}(t), \quad (7)$$

In Equation (7), the correlation $\text{PCC}^{n+1}(\mathbf{g}_i^k(t), \mathbf{g}^k(t))$ is based on the $n + 1$ mixed sample set.

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

$$\Delta H(k, t) = \Delta \text{SD}(k, t) |H^{n+1}(k, t) - H^n(k, t)|, \quad (8)$$

with

$$\Delta \text{SD}(k, t) = |\text{SD}^{n+1}(k, t) - \text{SD}^n(k, t)|, \quad (9)$$

where $\text{SD}^n(k, t)$ and $\text{SD}^{n+1}(k, t)$ are the SDs of the gene expression for center gene \mathbf{g}^k based on n reference samples $\{\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_n(t)\}$ and $n + 1$ mixed samples $\{\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_n(t), \mathbf{s}_{\text{case}}(t)\}$, respectively. The differential entropy $|H^{n+1}(k, t) - H^n(k, t)|$ between $H^n(k, t)$ and $H^{n+1}(k, t)$ characterizes the differences caused by the single case sample $\mathbf{s}_{\text{case}}(t)$. In other words, comparing with the local entropy $H^n(k, t)$ based on n reference samples $\{\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_n(t)\}$, $H^{n+1}(k, t)$ based on $n + 1$ mixed samples $\{\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_n(t), \mathbf{s}_{\text{case}}(t)\}$ records the perturbation brought by the single sample $\mathbf{s}_{\text{case}}(t)$ for local network N^k . Besides, to bring the fluctuation of genes into consideration, the differential standard deviation $\Delta \text{SD}(k, t)$ is regarded as the weight coefficient.

- 6) Calculate the weighted sum of $\Delta H(t)$ for all the local networks i.e.

$$\Delta H(t) = \frac{1}{Q} \sum_{k=1}^Q \Delta H(k, t), \quad (10)$$

where constant Q is the number of all genes (i.e. the number of all local networks). In Equation (10), the CI $\Delta H(t)$ records the overall impact caused by the single sample $\mathbf{s}_{\text{case}}(t)$, and thus is called the global SLE score or just SLE score for the global network N^G , whereas $\Delta H(k, t)$ in Equation (8) is called the local SLE score for the local network N^k centered at gene \mathbf{g}^k .

SGE

Given a scRNA-seq dataset. The SGE algorithm in the proposed webtool TPD is for detecting the critical transition of cell differentiation or cell fate commitment. Suppose there are totally N cells in a time point. The algorithm of SGE can be described as follows:

- 1) Plotting a scatter diagram for each pair of genes $(\mathbf{g}_i, \mathbf{g}_j)$, in a cartesian coordinate system where the vertical- and horizontal- axes are the expression values of the two genes, respectively. In the scatter diagram of genes \mathbf{g}_i and \mathbf{g}_j , for cell C_k ($k = 1, 2, \dots, N$),

we set two boxes near gene expression values $E_i^{(k)}$ (the gene expression of g_i in cell C_k) and $E_j^{(k)}$ (the gene expression of g_j in cell C_k), representing the strip neighborhood of $E_i^{(k)}$ and $E_j^{(k)}$ respectively.

2) Calculating statistical dependency index $r_{ij}^{(k)}$ as follows

$$r_{ij}^{(k)} = \frac{n^{(k)}(E_i, E_j)}{N} - \frac{n^{(k)}(E_i)}{N} \cdot \frac{n^{(k)}(E_j)}{N}, \quad (11)$$

where $n^{(k)}(E_i)$ and $n^{(k)}(E_j)$ represent the number of the points/cells in the strip neighborhood of $E_i^{(k)}$, and that of $E_j^{(k)}$, respectively. $n^{(k)}(E_i, E_j)$ denote the number of points/cells in the box neighborhood of (E_i, E_j) .

3) Building a specific network for each cell. If the statistical dependency index $r_{ij}^{(k)}$ is larger than zero, there is an edge between g_i and g_j in the cell C_k , otherwise there is no edge. In this way, a cell-specific network $N^{(k)}$ for cell C_k is built, where weight of each edge is determined by the dependency index $r_{ij}^{(k)}$.

4) Extracting each local network/subnetwork from the cell-specific network. Specifically, for a cell C_k , there are Q local networks $LN_i^{(k)}$ ($i = 1, 2, 3, \dots, Q$) corresponding to its Q genes.

5) Calculating the gene-specific local SGE value $H_i^{(k)}$ for each local network. Given a local network $LN_i^{(k)}$ centered at a gene g_i , the corresponding local SGE value is obtained from

$$H_i^{(k)} = -\frac{1}{\log(M)} \sum_{j=1}^M p_{ij}^{(k)} \log p_{ij}^{(k)}, \quad (12)$$

with

$$p_{ij}^{(k)} = \frac{r_{ij}^{(k)} \cdot E_j^{(k)}}{\sum_{j=1}^S r_{ij}^{(k)} \cdot E_j^{(k)}}, \quad (13)$$

where the value $E_j^{(k)}$ represents the gene expression of a neighbor g_j in C_k and constant M is the number of neighbors in the local network $LN_i^{(k)}$.

6) Calculating the cell-specific SGE value $H^{(k)}$ based on a group of genes with highest local SGE values i.e.

$$H^{(k)} = \sum_{i=1}^T H_i^{(k)}, \quad (14)$$

where constant T is an adjustable parameter representing the number of top genes with the highest local SGE values. In the web tool TPD, the default parameter $T = 5\% * Q$, that is, the top 5% genes with the highest SGE values are chosen as the signaling features. $H^{(k)}$ called the composite SGE index, which is capable to detect the tipping point of the single cell data during the progression of cells, like cell differentiation.

The determination of a tipping point

To quantify how a CI captures the criticality, the one sample t-test offers a unified way to determine whether there is a significant difference between the before-transition and the critical states. To determine whether a value x is statistically different from the mean of an n -dimensional vector $\mathbf{X} = (x_1, x_2, \dots, x_n)$, the one sample t-test is defined as the following equation

$$TS = \sqrt{n} \frac{\text{mean}(\mathbf{X}) - x}{\text{SD}(\mathbf{X})} \quad (15)$$

where $\text{mean}(\mathbf{X})$ represents the mean of vector \mathbf{X} , and $\text{SD}(\mathbf{X})$ represents the SD of the vector \mathbf{X} . To estimate the statistical significance between $\text{mean}(\mathbf{X})$ and x , the P value associated with TS is obtained from the t-distribution. There is a statistical significance between $\text{mean}(\mathbf{X})$ and x if $P < 0.05$, otherwise statistically it shows little significant difference. Therefore, if $C(t)$ denotes the value of a CI at time point t , then the time point t is determined as a tipping point if the CI satisfies the following two conditions. (i) $C(t) > C(t-1)$ and (ii) $C(t)$ is significantly different ($P < 0.05$) from the prior information (i.e. the mean of vector $\{C(1), C(2), \dots, C(t-1)\}$). Here, $C(t)$ is the CI based on any of the above three detection methods i.e. the cDNB score in Equation (1), the SLE score in Equation (10) or the composite SGE index in Equation (13).

The running environment and resources

The above three detection methods embedded in TPD were written in Python.NET Core and HTML 5.0. The platform is deployed on a computing server provided by Ali Cloud with the following instance type: 2 cores, 4 GB compute-optimized type c6 generation V, system disk: ultra cloud disk/dev/xvda 40 GB. In TPD, the embedded PPI networks of *H. sapiens* and *M. musculus* are STRING networks from <https://cn.string-db.org/> [29], with default confidence level 800. The gene annotation is based on DAVID [30].

TPD web tool

Data preparation and input files

TPD provides the analysis for both bulk data and single-cell datasets (Figures 1A and 2). It allows users to upload the time series or stage course data [e.g. molecular expression matrices, in which each row represents the expressions of a gene labeled with its official gene symbol, the first column lists the gene IDs (official symbols), and other columns are data of samples]. The accepted data formats include TXT, CSV and XLS files. There is no header of the input file.

Before unloading bulk and scRNA-seq data, TPD generally requires a pre-processing step that makes the gene expressions filtered and normalized (Fig. 1A). Specifically, genes with zero expressions in 50% or more samples for the bulk data, and those with zero expressions in 80% or more cells for the single-cell data, should be removed. Then the gene expressions should be normalized through Z-score (for bulk data) or through Log2 transformation (for single-cell data). TPD allows users to assign samples to specific time points or stages, and offers several analysis options (Figure 1C). In addition, TPD provides examples of different types of inputs, including seven bulk datasets (acute lung injury dataset for *M. musculus*, 6 The Cancer Genome Atlas (TCGA) datasets for *H. sapiens*) and four scRNA-seq datasets (MEFs-to-Neurons and mouse hepatoblast cells (MHCs)-to-hepatocyte cholangiocyte cells (HCCs) datasets for *M. musculus*, neural progenitor cells (NPCs)-to-Neurons and hESCs-to-definitive endoderm cells (DECs) for *H. sapiens*), which can be downloaded from the web tool directly. For each example, there is a Readme file illustrating the details of the corresponding dataset.

Options of analysis

TPD allows users to decide (i) whether to perform survival analysis, (ii) the type of species from which the input data were derived, (iii) the specific method of tipping point detection that should be used, (iv) the cutoff of identified key genes (the DNB group) (default: top 5% genes with the highest CI value) and (v) the number of time points/stages (Figure 2). For users to easily learn

A Bulk data analysis

Data upload

Single-Sample CI Multiple-sample CI

Case data
UCEC_tumor_marker_gene_matrix.xlsx

Reference data

Survival Analysis

Do survival analysis Without survival analysis

clinical data
UCEC_Survival.csv

Species selection

Homo sapiens Mus musculus

Method selection

Single-sample Landscape Entropy (SLE)

Cutoff threshold

Top Gene: %

Data selection

The uploaded dataset contains 529 columns. Please indicate the information time points.

There are time points/ stages.

The 1st time point / stage : --

The 2nd time point / stage : --

The 3th time point / stage : --

The 4th time point / stage : --

B Single-cell data analysis

Data upload

Multiple-sample CI

Case data
MEF-to-Neurons.xlsx

Species selection

Homo sapiens Mus musculus

Method selection

Single-cell Graph Entropy (SGE)

Cutoff threshold

Top Gene: %

Data selection

The uploaded dataset contains 405 columns. Please indicate the information time points.

There are time points/ stages.

The 1st time point / stage : --

The 2nd time point / stage : --

The 3th time point / stage : --

The 4th time point / stage : --

The 5th time point / stage : --

Figure 2. The interface for the input of TPD. (A) The demo of the input for bulk data. (B) The demo of the input for single-cell data. User options of data upload, survival analysis, species selection, methods selection and sample assignment to time points, are provided.

how TPD works, the proposed web tool contains a ‘run’ button for each demo and clicking this button submits the inquiry directly.

Data upload based on samples

For the bulk dataset input, TPD works well with two situations of samples that are mostly possible to arise in real-world datasets, that is, (1) the multiple-sample situation when there are replicates of case samples at each time point, and (2) the single-sample situation when there is only one case sample at each time point. For situation (1), TPD directly calculates a series of statistical indices and the CI at each time point, thus requiring at least three case samples/replicates so that the output is statistically meaningful. For situation (2), TPD requires users to specify some samples as the reference/control samples. Then, it mixes the single case sample with the reference samples at each time point, and characterizes the statistical perturbation brought by each individual sample against the group of given control/reference samples by either SLE or cDNB. Thus, in this situation, it generally requires at least three

reference/control samples. TPD offers an option to users to select ‘Single-sample CI’ for situation (1), or ‘Multiple-sample CI’ for situation (2).

Survival analysis

If the survival analysis is selected, the related clinical information needs to be uploaded. The detailed format is shown in the embedded TCGA examples. Usually, this option is provided together with *H. sapiens* datasets.

Species selection

Users can choose the specific species for the uploaded dataset. Currently, TPD provides two options: *H. sapiens* and *M. musculus*.

Methods selection

Three classical tipping point revealing methods are provided e.g. cDNB and SLE for bulk data, and SGE for scRNA-seq data. The detailed algorithms are described in Methods section.



Figure 3. Output of TPD for bulk data based on the SLE method. **(A)** An average SLE curve for indicating the tipping point. **(B)** A landscape display for demonstrating the dynamic change of the local entropy of specific genes. **(C)** The dynamic evolution of a network composed by the signaling genes. **(D)** The lists of signaling genes with the top highest and lowest CI values. **(E)** Survival analysis based on the signaling genes.

Cutoff threshold

Users can decide the cutoff of key genes (a small group of dominated genes) with the top highest and lowest index. The default parameter is 5% i.e. the top 5% genes with the highest CI values. This gene group plays a crucial role in the biological functional analysis related to critical transition according to the DNB properties.

Analysis results

Upon receiving the input data and options, the web tool outputs multifarious visualized results, including the global curve or landscape display of the local values of the TPD index with the suggested tipping point and confidence scores (P value) (Figure 1D), the identified signaling genes with the top highest and lowest CI values in terms of the selected method (Figure 1G), the dynamic change in the DNB/leading network that may drive the critical transition (Figure 1E) and the survival analysis that may help to identify 'dark' genes (nondifferential in terms of expression levels but differential in terms of the CIs of cDNB/SLE/SGE). For scRNA-seq data, TPD also displays the clustering results of

different cell types and fractional analysis of the identified key genes (Figure 1F). While calculating, there is a progress bar of TPD indicating the running condition of this task.

For the bulk data, TPD outputs time-dependent scores of a CI according to the chosen detection method (for example, a SLE score if SLE method is selected). Thus, a CI curve is visualized which indicates the tipping point of the uploaded dataset by the significant P value i.e. the pink zone with $P < 0.05$ is critical state, and the blue zones with $P \geq 0.05$ are non-critical states (Figure 3A). By utilizing the underlying PPI network, each gene can be assigned a local CI in its local network which is constructed by this gene and its first neighbors. Thus, TPD provides the landscape display of the local CIs (Figure 3B), which can also help to identify the tipping point of the dynamic system.

Based on the DNB theory, a small group of dominant genes which play a crucial role in detecting early warning signal of the critical state are selected. TPD outputs the dynamic evolution of the networks of these genes (Figure 3C). The blue and the red nodes represent genes with low and high CI, respectively. It can be found that most of the genes in this module are red in the critical state. We demonstrated the key genes both with the top

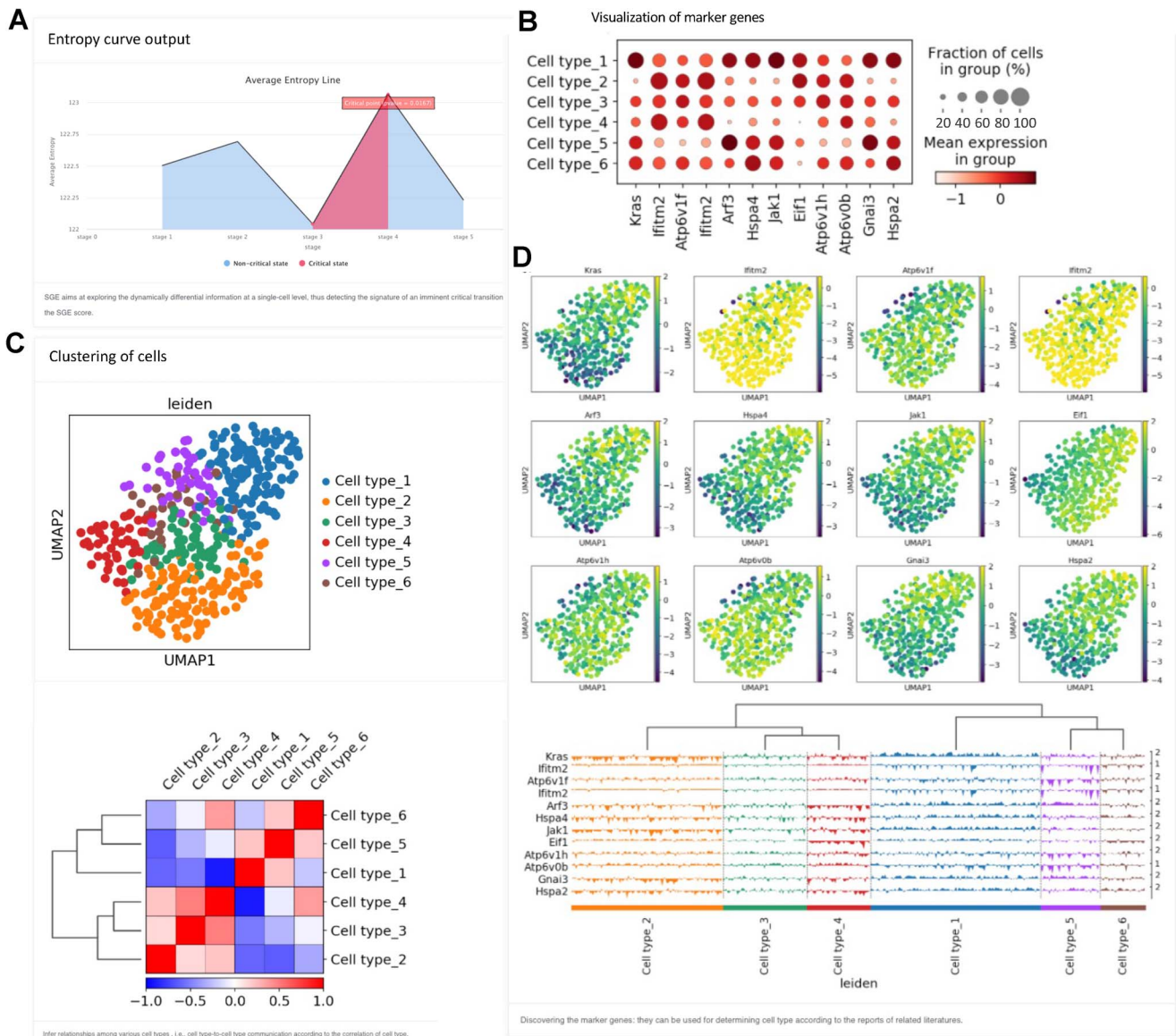


Figure 4. Output of TPD for single-cell data. (A) An average SGE curve for indicating the tipping point during the process of cell differentiation. (B) The fraction of cells based on marker genes. (C) The clustering of cells based on all signaling genes. (D) The UMAP clustering of cells based on each specific signaling gene.

highest and lowest CI values in Figure 3D, from which users can check the CI values and related biological description of these genes. If clinical information is uploaded, TPD outputs the results of survival analysis for key genes (Figure 3E). Clearly, there is a significant difference between the survival span with low CI values and that with high CI values.

For the scRNA-seq data, TPD outputs an average entropy curve which can indicate the tipping point of this dataset and show the corresponding *P* value (the pink zone is critical state, and the blue zone is non-critical state) (Figure 4A). Similarly, the genes with the top highest local SGE values are selected as the signaling genes. TPD illustrates the fraction of cells based on marker genes (Figure 4B), the UMAP and hierarchical clustering of cells based on all signaling genes or each specific gene (Figure 4C and D).

Conclusion

Generally, identifying the tipping point of a biological system is important but also difficult and troublesome due to the high

complexity and nonlinearity of the system as well as the tedious data analysis process. Supported by three TPD methods, which are all theoretically based on the DNB concept, the proposed web tool TPD provides a one-stop solution for users to achieve rapid detection of the tipping point and accurate identification of the signaling/key molecules and their dynamic networks from the input time series data. Currently, embedded with PPI networks of *H. sapiens* and *M. musculus*, TPD can analyze the high-dimensional time series of these two species, and output multifarious visualized results. We demonstrate the versatility of TPD in the use-case examples. These proposed web tools are user friendly and do not require programming knowledge or installation.

In view of the effectiveness of tipping point identification, TPD is expected to support users from academic and clinical fields. Together with the dynamic prediction method [31], TPD may not only detect the early warning signals of critical states based on omics data, but reveal the dynamically differential information that provides new insights of critical properties of a biological system in the vicinity of its tipping point.

Key points

- We present the tipping point detector (TPD) at <http://www.rpcomputationalbiology.cn/TPD>, a web tool for the detection of the tipping point during the dynamic process of biological systems.
- TPD detects the potential tipping point/critical state from the input omics data, and outputs multifarious visualized results.
- TPD provides a suggested tipping point with a statistically significant *P* value, the identified key genes and their functional biological information, the dynamic change in the DNB/leading network and the survival analysis based on DNB scores.

Data availability

All data are available at <http://www.rpcomputationalbiology.cn/TPD>.

Funding

National Natural Science Foundation of China (Grant Nos. 62172164, 12026608 and 12131020), Guangdong Basic and Applied Basic Research Foundation, China (Grant Nos. 2019B151502062 and 2021A1515012317).

Conflict of Interest

None declared.

References

1. Liu R, Wang X, Aihara K, et al. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Med Res Rev* 2014;**34**:455–78.
2. Hu S, Aisner SC, Lubitz SE. Cinacalcet for management of tertiary hyperparathyroidism associated with chronic treatment of hypophosphatemia in an adult with tumor-induced osteomalacia. *AACE Clin Case Rep* 2015;**1**:e225–9.
3. Yang B, Li M, Tang W, et al. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nat Commun* 2018;**9**:1–14.
4. Penney ME, Parfrey PS, Savas S, et al. A genome-wide association study identifies single nucleotide polymorphisms associated with time-to-metastasis in colorectal cancer. *BMC Cancer* 2019;**19**:1–12.
5. Jayadevappa G, Ravishankar SN. Risk factors and clinical profile of ischemic stroke patients attending emergency care facility in Bangalore city. *Sch J App Med Sci* 2021;**4**:572–7.
6. Scheffer M, Bascompte J, Brock WA, et al. Early-warning signals for critical transitions. *Nature* 2009;**461**:53–9.
7. Richard A, Boullu L, Herbach U, 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**:e1002585.
8. Teschendorff AE, Enver T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun* 2017;**8**:1–15.
9. Li Y, Zhang S-W. Resilience function uncovers the critical transitions in cancer initiation. *Brief Bioinform* 2021;**22**:bbab175.
10. Chen P, Li Y, Liu X, et al. Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. *J Transl Med* 2017;**15**:1–15.
11. Chen X, Sun L-G, Zhao Y. NCMCMDA: miRNA–disease association prediction through neighborhood constraint matrix completion. *Brief Bioinform* 2021;**22**:485–96.
12. Zhao Y, Wang C-C, Chen X. Microbes and complex diseases: from experimental results to computational models. *Brief Bioinform* 2021;**22**:bbaa158.
13. Chen X, Guan N-N, Sun Y-Z, et al. MicroRNA-small molecule association identification: from experimental results to computational models. *Brief Bioinform* 2020;**21**:47–61.
14. Zhang J, Guan M, Wang Q, et al. Single-cell transcriptome-based multilayer network biomarker for predicting prognosis and therapeutic response of gliomas. *Brief Bioinform* 2020;**21**:1080–97.
15. Eschrich S, Zhang H, Zhao H, et al. Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform. *Int J Radiat Oncol Biol Phys* 2009;**75**:497–505.
16. Jin G, Zhou X, Wang H, et al. The knowledge-integrated network biomarkers discovery for major adverse cardiac events. *J Proteome Res* 2008;**7**:4013–21.
17. Wu X, Chen L, Wang X. Network biomarkers, interaction networks and dynamical network biomarkers in respiratory diseases. *Clin Transl Med* 2014;**3**:1–7.
18. Zeng T, Sun S, Wang Y, et al. Network biomarkers reveal dysfunctional gene regulations during disease progression. *FEBS J* 2013;**280**:5682–95.
19. Chen L, Liu R, Liu Z-P, et al. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep* 2012;**2**:1–8.
20. Chu L-F, Leng N, Zhang J, et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol* 2016;**17**:1–20.
21. Gao R, Yan J, Li P, et al. Detecting the critical states during disease development based on temporal network flow entropy. *Brief Bioinform* 2022. <https://doi.org/10.1093/bib/bbac164>.
22. Yang XH, Goldstein A, Sun Y, et al. Detecting critical transition signals from single-cell transcriptomes to infer lineage-determining transcription factors. *Nucleic Acids Res* 2022. <https://doi.org/10.1093/nar/gkac452>.
23. Qiu Y, Ching W-K, Zou Q. Matrix factorization-based data fusion for the prediction of RNA-binding proteins and alternative splicing event associations during epithelial–mesenchymal transition. *Brief Bioinform* 2021;**22**:bbab332.
24. Liu R, Chen P, Chen L. Single-sample landscape entropy reveals the imminent phase transition during disease progression. *Bioinformatics* 2020;**36**:1522–32.
25. Lesterhuis WJ, Bosco A, Millward MJ, et al. Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. *Nat Rev Drug Discov* 2017;**16**:264–72.
26. Liu X, Chang X, Leng S, et al. Detection for disease tipping points by landscape dynamic network biomarkers. *Natl Sci Rev* 2019;**6**:775–85.
27. Huang Y, Chang X, Zhang Y, et al. Disease characterization using a partial correlation-based sample-specific network. *Brief Bioinform* 2021;**22**:bbaa062.
28. Zhong J, Han C, Zhang X, et al. scGET: predicting cell fate transition during early embryonic development by single-cell graph entropy. *Genom Proteom Bioinformatics* 2021;**19**:461–74.
29. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;**43**:D447–52.

30. Dennis G, Sherman BT, Hosack DA, *et al.* DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;**4**:1–11.
31. Chen P, Liu R, Aihara K, *et al.* Autoreservoir computing for multi-step ahead prediction based on the spatiotemporal information transformation. *Nat Commun* 2020;**11**:1–15.