MIWE: detecting the critical states of complex biological systems by the mutual information weighted entropy

Yuke Xie[1,2], Xueqing Peng[1] and Peiluan Li[1*]

*Correspondence:
15038522015@163.com

[1] School of Mathematics
and Statistics, Henan University
of Science and Technology,
Luoyang 471000, China
[2] Key Laboratory of Systems
Health Science of Zhejiang
Province, School of Life
Science, Hangzhou Institute
for Advanced Study, University
of Chinese Academy of Sciences,
Hangzhou 310024, China

Abstract

Complex biological systems often undergo sudden qualitative changes during their dynamic evolution. These critical transitions are typically characterized by a catastrophic progression of the system. Identifying the critical point is critical to uncovering the underlying mechanisms of complex biological systems. However, the system may exhibit minimal changes in its state until the critical point is reached, and in the face of high throughput and strong noise data, traditional biomarkers may not be effective in distinguishing the critical state. In this study, we propose a novel approach, mutual information weighted entropy (MIWE), which uses mutual information between genes to build networks and identifies critical states by quantifying molecular dynamic differences at each stage through weighted differential entropy. The method is applied to one numerical simulation dataset and four real datasets, including bulk and single-cell expression datasets. The critical states of the system can be recognized and the robustness of MIWE method is verified by numerical simulation under the influence of different noises. Moreover, we identify two key transcription factors (TFs), CREB1 and CREB3, that regulate downstream signaling genes to coordinate cell fate commitment. The dark genes in the single-cell expression datasets are mined to reveal the potential pathway regulation mechanism.

Keywords: Dynamic network biomarker (DNB), Mutual information, Critical state, Differential entropy

Background

The development of complex disease systems can be categorized into three stages [1]: normal state, critical state and disease state. The human system has high elasticity and strong robustness in normal state and disease state. In the critical state, the human system is unstable and reversible, with low rebound and weak robustness. If the system is disturbed at this time, it may transition to the subsequent stable state or revert to the preceding stable state. Most diseases are discovered at this stage of the onset of symptoms. Despite receiving appropriate treatment, returning to a normal state remains challenging [2]. Being able to identify critical states of complex diseases at an early stage and identify tipping points before serious complications occur allows for more precise

personalized treatment. In experiments conducted at the single-cell level, cell fate commitment marks a pivotal transition, and the essential endeavor of understanding and foreseeing this shift is crucial for tailoring disease models and performing personalized assessments of therapeutic efficacy in individual patients [3]. Therefore, it holds significant biomedical importance to describe the dynamic features of biological systems and accurately detect the critical stages.

In the study of complex biological systems, researchers had made great achievements in the detection of preliminary alerts of complex systems by using dynamic network markers, differential network and network entropy. The recently proposed DNB concept theoretically derived a DNB-based indicator that acts as a basis for detecting the approach of critical state [2]. Single-cell graph entropy quantified the robustness and pivotal nature within gene regulatory networks between cellular communities and could be used to provide key signals of cell fate determination [4]. At the small-sample level, evaluating the critical state can also be achieved by calculating the network entropy difference generated by perturbation using a single perturbed sample [5].

Although many studies had contributed to the development of areas related to warning signs of qualitative changes in detection systems, a large amount of research was currently conducted on bulk datasets. Compared with traditional bulk omics information, single-cell analysis is impacted by high dimensionality, noise, sparsity, and heterogeneity in samples. Characterizing the dynamics of biological systems from single-cell datasets and accurately detecting critical state is a complex task.

In this research, we suggest a differential entropy method utilizing mutual information network, i.e., mutual information weighted entropy (MIWE), which uses the differential entropy information of each stage to detect the critical state. The gene expression is transformed into probability distribution and the mutual information network is constructed at each stage. Then, according to the weight between genes in each stage network, the weighted differential entropy of each local network is calculated to quantitatively describe the fluctuations of the system at each stage, thus identifying the critical state. The MIWE method is utilized on a numerical simulation dataset and four real biological datasets, encompassing bulk sequencing and single-cell RNA sequencing (scRNA-seq) data. We effectively identify critical states of colon adenocarcinoma (COAD) and thyroid carcinoma (THCA). In addition, signals related to cell fate commitment are detected in datasets related to cell differentiation, encompassing mouse embryonic fibroblast (MEF) to neuron and mouse embryonic stem cell (mESC) to mesoderm progenitor (MP). The predicted results align with the original experimental results, which support the validity and stability of the MIWE method.

The MIWE method offers a reliable way for identifying critical states in the evolution of the complex biological systems. This approach possesses the following four benefits: (1) From the perspective of continuous variables, MIWE method can more accurately describe the mutual influence between genes than discrete variables, and can capture small changes and trends when dealing with complex data structures and nonlinear relationships, with strong robustness. (2) MIWE method is suitable for both bulk and single-cell expression data. By using edge weights to calculate phase entropy and make full use of network information, MIWE method can accurately reflect the dynamics and complexity of system changes and enhance effectiveness. (3) Based on MIWE

method, critical states can be detected before critical qualitative changes occur in complex biological systems and the signaling genes of the critical state can be detected. (4) Based on the MIWE method, key TFs related to embryonic differentiation and more potential dark genes that are not detectable by traditional biomarkers are discovered. Although these dark genes are non-differential signaling genes, they have been demonstrated to participate in embryonic differentiation processes through functional pathway mechanisms.

Methods

Data progression and functional analysis

The MIWE method has been utilized on a numerical simulation dataset and four real biological datasets, encompassing bulk sequencing data including COAD and THCA from The Cancer Genome Atlas (TCGA) database (http://cancergenome.nih.gov) and scRNA-seq data (embryonic differentiation of MEF to neurons (GEO: GSE67310) [6] and mESC to MP (GEO: GSE79578) [7]. from the NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo).

The functional annotation analysis relies on the DAVID Bioinformatics Resources (https://david.ncifcrf.gov/) and Circos (http://www.circos.ca/). Potential upstream regulators of signaling genes are identified based on ChEA3 (https://amp.pharm.mssm.edu/chea3/). Protein–Protein Interaction (PPI) networks are constructed utilizing STRING (https://string-db.org/) and the client software Cytoscape (https://cytoscape.org/).

Theoretical background

The dynamic change of complex biological system can be regarded as irregular process, which will undergo qualitative change when approaching the critical stage. DNB theory proposed that when system approaches critical point, a set of genes or protein molecules, known as the DNB group, emerges that fulfills the following conditions: the connection between any two molecules in the DNB group swiftly grows, while the correlation with any other non-DNB molecule declines. The standard deviation of any member of the DNB group grows sharply. The system state may show small significant changes before reaching the critical point, and traditional biomarkers or methods cannot successfully predict the critical state, while the DNB index acts as a basis for identifying the approach of critical state [2]. Therefore, it is the active changes in molecular binding and spatial fluctuations, instead of differences in gene expression, that lead to differences in biological systems [8].

MIWE method transforms gene expression into probability distribution and constructs mutual information network at each stage. The edge between genes in each local network is used as the weight to calculate the weighted differential entropy of each stage. The dynamic difference changes of each stage can be measured by the difference of entropy value. The global MIWE score at every stage functions as a precursor signal for identifying the critical state.

Algorithm to detect the tipping point based on MIWE

Given the chronological datasets of scRNA-seq or bulk sequencing, we design the following algorithm to detect the critical state (Fig. 1).
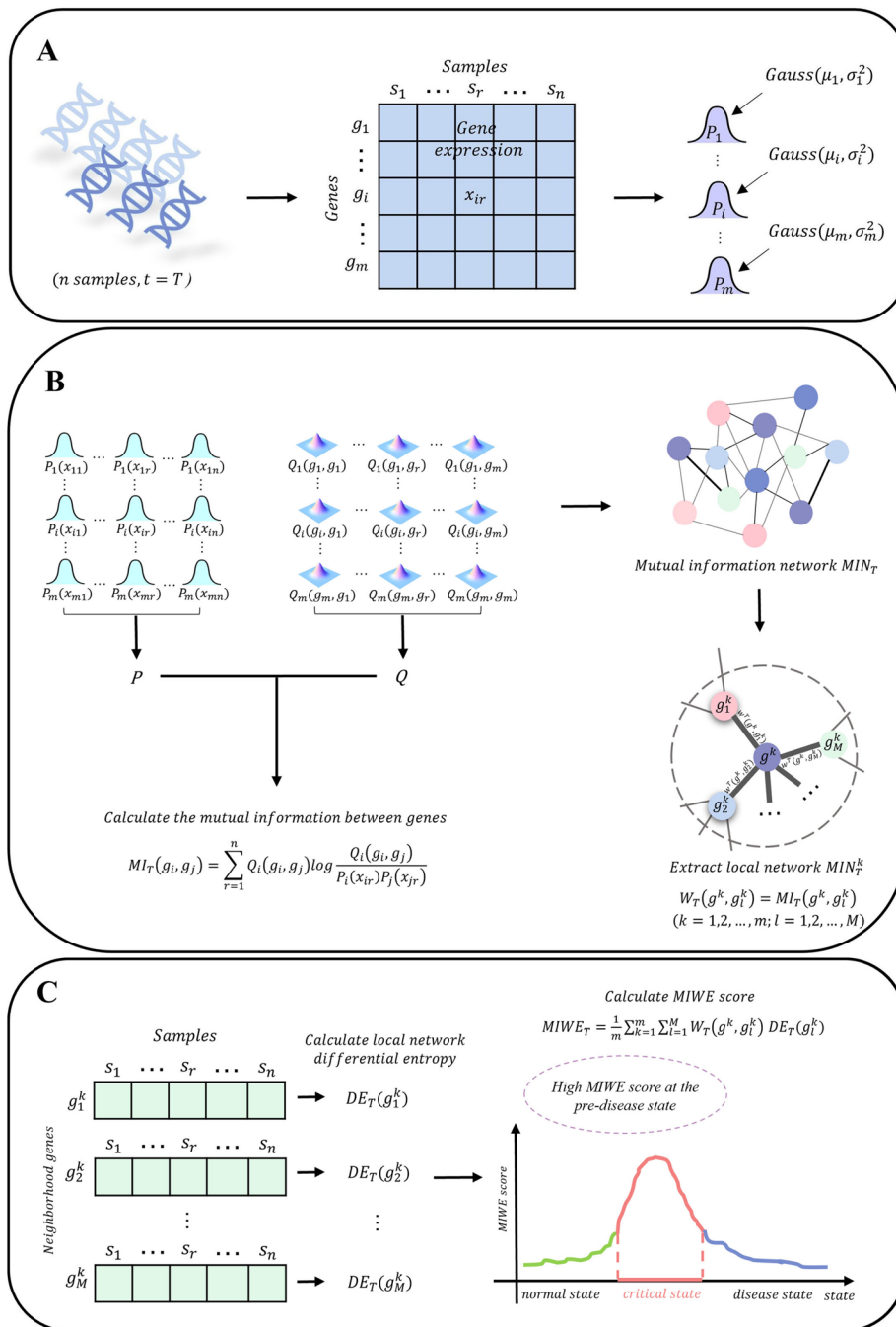
Fig.1 The schematic of the MIWE method. **A** Gaussian distribution is fitted for each gene. **B** Mutual information network is constructed by taking mutual information between genes as edge weight, and local network is extracted from global network. **C** The weighted differential entropy of the global network is calculated. When the system is in the critical state, the MIWE score is at a low level, and once it reaches the critical state, the MIWE score increases sharply

[step 1] Fit Gaussian distribution of each gene at different time T.

Based on given samples and transform gene expressions into probability distribution.

The Gaussian distribution is fitted according to the expression of g_i $(i = 1, 2, \ldots, m)$ in the n samples $\{S_1, S_2, \ldots, S_n\}$ at time T. The goodness of fit test is performed on the fitted Gaussian distribution. The gene expression values among the samples are converted to cumulative probability $P_i(x_{ir})$. If any linear combination of genes g_i and g_j obey one-dimensional normal distribution, then the joint distribution between the two genes as the bivariate normal distribution, and their joint probability is $Q_i(g_i, g_j)$.

$$P_i(x_{ir}) = \frac{1}{\sigma_i \sqrt{2\pi}} \int_0^{x_{ir}} e^{-\frac{(u-\mu_i)^2}{2\sigma_i^2}} \, du, \tag{1}$$

$$Q_i(g_i, g_j) = \frac{1}{2\pi \sigma_i \sigma_j \sqrt{1 - \rho^2}} \int_0^{\overline{g_i}} \int_0^{\overline{g_j}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(u-\mu_i)^2}{\sigma_i^2} - 2\rho \frac{(u-\mu_i)(v-\mu_j)}{\sigma_i \sigma_j} + \frac{(v-\mu_j)^2}{\sigma_j^2} \right]} \, du \, dv, \tag{2}$$

where x_{ir} is the gene expression values of gene g_i $(i = 1, 2, \ldots, m)$ in the samples r $(r = 1, 2, \ldots, n)$, $\overline{g_i}$ and $\overline{g_j}$ are the average expression values of genes g_i and g_j in n samples at time T respectively, ρ is the correlation coefficient between gene g_i and g_j at time T, μ_i, σ_i $(i = 1, 2, \ldots, m)$ are the mean expression value and standard deviation of gene g_i in n samples at time T.

[step2] Construct mutual information network MIN_T at each time T.

The edge association in the MIN_T can quantitatively characterize the correlation degree between genes, in which the edge weight between genes g_i and g_j is determined by the $MI_T(g_i, g_j)$ index.

$$MI_T(g_i, g_j) = \sum_{r=1}^n Q_i(g_i, g_j) \log \frac{Q_i(g_i, g_j)}{P_i(x_{ir}) P_j(x_{jr})}, \tag{3}$$

the degree of correlation between genes is described from the perspective of information. In the presence of a certain level of gene correlation, increased mutual information is observed when there is less randomness between genes.

[step3] Extract the local network from the global network.

Extract the local network MIN_T^k $(k = 1, 2, \ldots, m)$ from the global network MIN_T at each time T, which contains a central gene g^k and first-order neighbors $\{g_1^k, g_2^k, \ldots, g_M^k\}$, where the edge weight $W_T(g^k, g_l^k) = MI_T(g^k, g_l^k)$ in the local network.

[step4] Calculate differential entropy of the neighborhood gene g_l^k $(l = 1, 2, \ldots, M)$ in local network MIN_T^k $(k = 1, 2, \ldots, m)$.

For each local network MIN_T^k $(k = 1, 2, \ldots, m)$ at time T, the differential entropy of neighborhood gene g_l^k $(l = 1, 2, \ldots, M)$ is denoted as:

$$DE_T\left(g_l^k\right) = -\int_0^{\overline{g_l^k}} f(x) \log f(x) \, dx, \tag{4}$$

$$f(x) = \frac{1}{\sigma_l^k \sqrt{2\pi}} e^{-\frac{\left(x - \mu_l^k\right)^2}{2\left(\sigma_l^k\right)^2}}, \tag{5}$$

where g_l^k are the average expression values of genes g_l^k in n samples at time T, μ_l^k, σ_l^k ($l = 1, 2, \ldots, M$) are the mean expression value and standard deviation of gene g_l^k in n samples at time T.

[step5] Calculate mutual information weighted entropy of the global network $MIWE_T$.

Calculate the weighted entropy value $MIWE_T^k$ ($k = 1, 2, \ldots, m$) of each local network at time T, namely,

$$MIWE_T^k = \sum_{l=1}^{M} W_T(g^k, g_l^k) DE_T\left(g_l^k\right), \tag{6}$$

then the weighted entropy score of the global network is:

$$MIWE_T = \frac{1}{m} \sum_{k=1}^{m} MIWE_T^k, \tag{7}$$

Signaling biomolecules exhibit significant collective behavior and intense fluctuations during the critical transition of a complex dynamic system. The weighted entropy of the local network containing signal biomolecules in the critical state is significantly different from that in the pretransition state. If $MIWE_T$ sharply increases, then time point T is the critical point, and the top 5% genes of $MIWE_T^k$ are signaling genes that regarded as DNBs in this work.

Results

Validation based on numerical simulation

We use a theoretical model to validate the robustness of MIWE method, and construct a 10-node monitoring network based on the Michaelis–Menten equation [9], which is mainly used to study transcription and translation processes [10], nonlinear biological processes [11, 12]. The 10-node monitoring network can generate datasets for numerical simulation, and as the parameter p varies from -0.5 to 0.25, the system experiences the critical transition when the parameter value is $p = 0$.

Figure 2A shows the gene regulatory network composed of 10 nodes with both activating and inhibitory interactions. Before the system reaches the critical point, MIWE score is at a low level. When the parameter value $p = 0$, MIWE score increases sharply, providing a precursor signal for the upcoming state change (Fig. 2B). Considering the existence of strong noise in real datasets, we verify the MIWE method under the influence of different noises, and compare it with SLE [5] and sJSD [13] methods (Fig. 2C). As the noise intensity increased, MIWE consistently offers early warning signals for impending tipping points with heightened sensitivity, indicating that the MIWE method is more robust and efficient in detecting critical points in biological processes. Additional information regarding the numerical simulation is available in the Additional file 1: Section A.
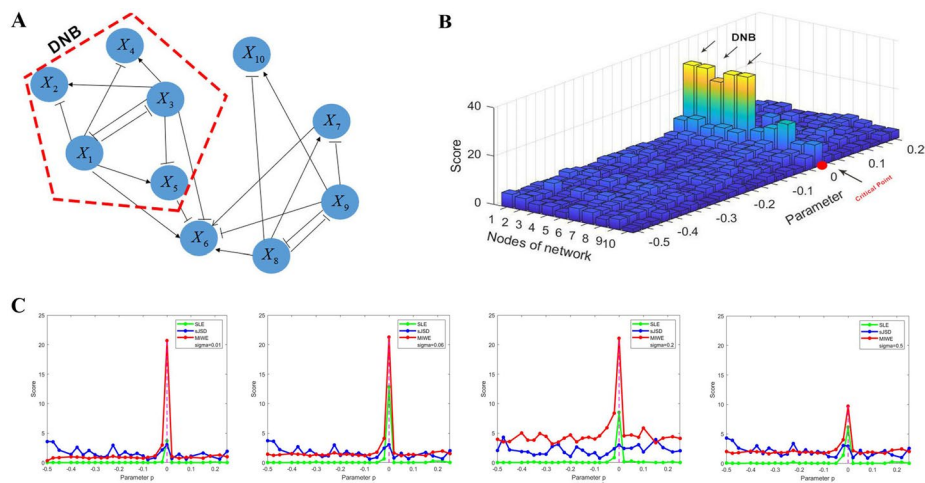
Fig.2 Numerical simulation dataset is used to verify the effectiveness of MIWE. **A** Gene regulatory network model, where the arrow represents positive regulation. **B** MIWE score for each parameter of 10 nodes. **C** Comparison of the robustness of the MIWE method with the SLE, sJSD method at various levels of noise strength

Identifying cell fate commitment during embryonic differentiation

To verify the validity of the MIWE method and detect the transformation of cell fate commitment, the method is utilized on two datasets of cell differentiation, including MEF to neurons (GSE67310) and mESC to MP (GSE79578) data. The weighted entropy of each local network is calculated according to the steps of the algorithm. Finally, the average weighted entropy (Eq. 7) is taken at each time point to quantitatively characterize the criticality of the single-cell community.

We use the MIWE score curve across time points to show the fluctuations of cell differentiation at each stage. For MEF to neurons data, MIWE scores increase significantly from day 5 to day 20 (Fig. 3A), providing a precursor signal for the imminent differentiation into neurons, indicating that cell fate commitment began on day 22. In mESC to MP data, MIWE scores at 24 h are significantly different from those at adjacent stages (Fig. 3B), indicating that transition is about to take place after 24 h, namely mouse embryonic stem cells differentiate into mesoderm. The algorithm detection results of the two datasets are consistent with the original experimental observation. Moreover, to prove the robustness of the proposed method, box graphs of weighted entropy at each stage are presented based on samples at each time point. The median value of the block diagram provides obvious signal for the critical point, indicating that the MIWE value is highly robust to the sample noise.

The signaling genes are identified as the top 5% of genes with the highest local MIWE scores, which may be highly correlated with cell differentiation. The landscape map shows dynamic changes in the distribution of local MIWE values of signaling genes in the global view (Fig. 3C, D), and the local MIWE values of the signaling genes in the two datasets increase sharply at day 20 and 24 h, respectively. Changes in local MIWE values of all genes are shown in Additional file 1: Fig. S2. In addition, signaling genes are mapped to PPI networks to observe the dynamic changes of networks at different stages. For both datasets, significant changes in network structure
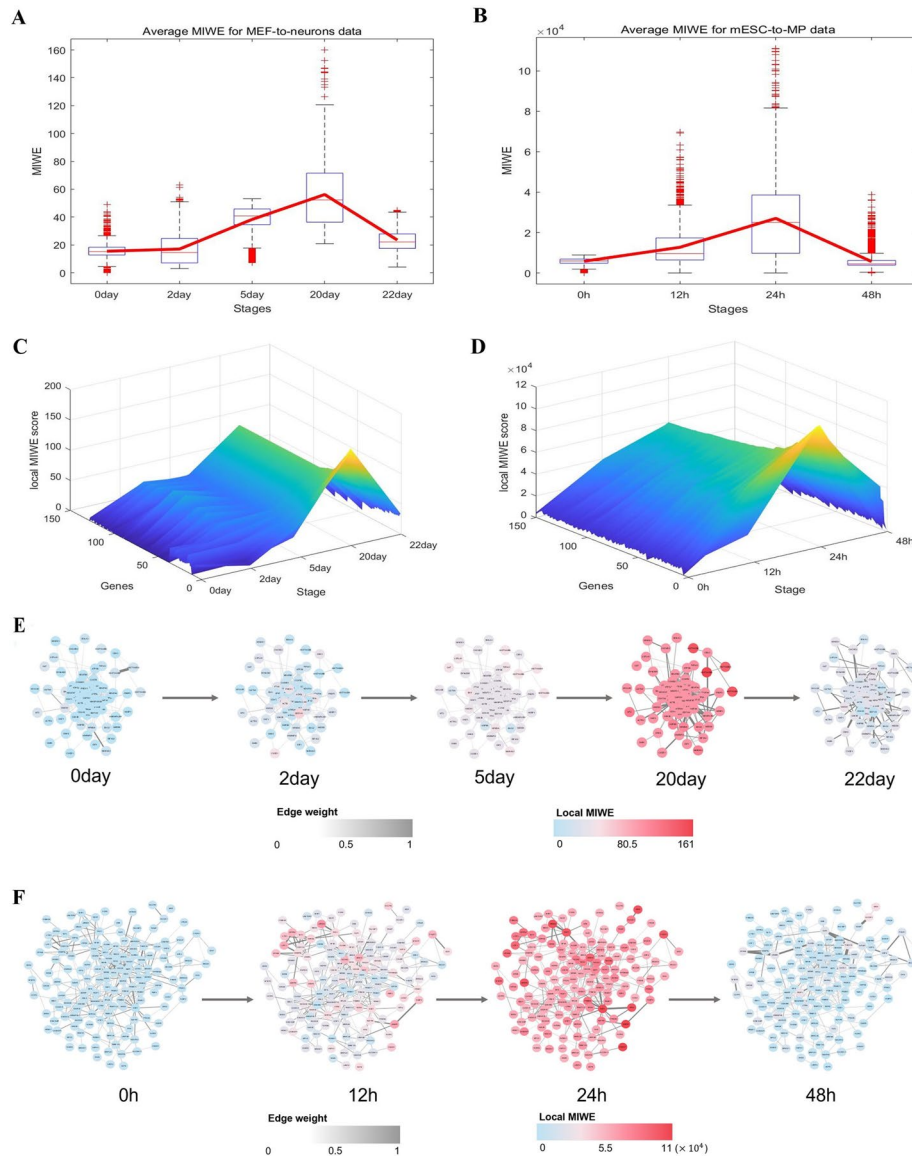
Fig.3 Detecting the signal of cell fate commitment. The MIWE value is calculated for **A** MEF to neurons and **B** mESC to MP. The landscape of local MIWE values illustrates the dynamic evolution of network entropy in a global view for **C** MEF-to-neuron, **D** mESC to MP. The dynamical evolution of gene regulatory networks for the **E** MEF-to-neuron, **F** mESC to MP

are observed at day 20 and 24 h, respectively, indicating an upcoming cell fate commitment (Fig. 3E, F).

Detecting potential upstream TFs

TFs are important molecules that control gene expression and can be considered as key players in controlling or driving cell fate commitment [14, 15]. In order to explore the involvement of the signaling genes identified in the two cell differentiation datasets in the process of cell fate commitment, we separately predict the TFs of the two groups of signaling genes on the ChEA3 website, and select the top 20 in the comprehensive average ranking as the main research content. In the GSE67310 and GSE79578 data, two sets

of TFs modulate 74% and 86% of the signaling genes at the critical point, respectively (Fig. 4A, B).

Some TFs play an important role in cell differentiation and proliferation. They are closely related to cell proliferation and self-renewal, and are crucial contributors to the early embryonic development and cell lineage specification. For GSE67310 data, the absence of CHCHD3 expression can lead to tissue undergrowth and cell proliferation defects [16], VEZF1 can regulate cell differentiation and proliferation and participate in the early vascular differentiation process [17], SP3 is required for perinatal survival in mice [18]. GTF2I indirectly contributes to the transcriptional regulation of genes controlling cell proliferation and cell cycle through encoding transcription factor TFII-I [19]. Functional annotations of TFs for GSE79578 data are in the Additional file 1: Section C.

In the analysis of TFs from GSE67310 data, we find two relatively key TFs, which can contribute to a more profound comprehension of the molecular mechanisms of embryonic development and hold significant implications for the treatment and prevention
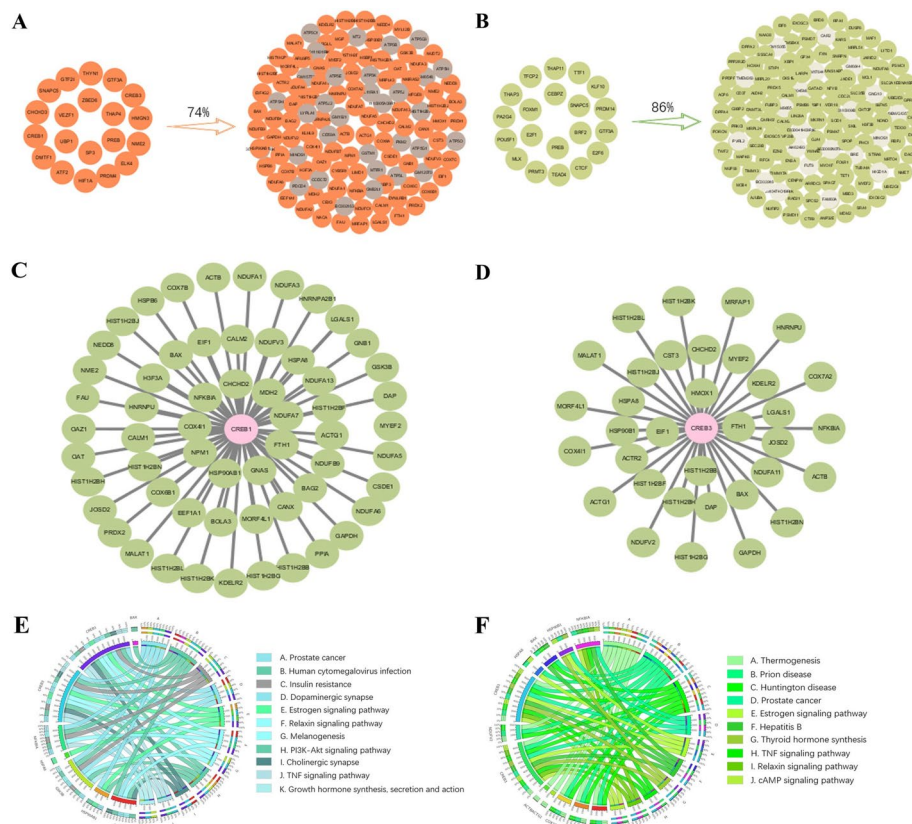

Fig.4 TFs regulation and related enrichment analysis. **A** TFs modulated 74% of signaling genes identified by GSE67310 critical point. **B** TFs modulated 86% of signaling genes identified by GSE79578 critical point. Regulatory network of **C** CREB1, **D** CREB3. **E** CREB1 and its regulated signaling genes participate in significant biological processes and KEGG pathways. The outer ring's left side signifies the signaling genes identified by MIWE, while the right side represents the diverse biological processes associated with these genes. The inner ring depicts various enrichment pathways, with connection color and width indicating different levels of gene function significance. **F** CREB3 and its regulated signaling genes participate in significant biological processes and KEGG pathways

of related diseases, namely CREB1 and CREB3. CREB1 plays a role in cell proliferation, myogenic differentiation and other related pathways [20]. CREB3 is involved in embryonic development and the differentiation of other tissues and organs, such as osteoblast differentiation [21]. In order to visualize the downstream signaling genes regulated by these two TFs, we present the regulatory network centered on TFs (Fig. 4C, D). Combined with the TFs and their regulated signaling genes, we find that they are involved in some signaling pathways related to embryonic differentiation (Fig. 4E, F). The TNF signaling pathway is central to a range of physiological and pathological processes, influencing cell proliferation, differentiation, apoptosis, immune response regulation, and inflammation induction. Activation of TNF signaling pathway can trigger activation of PI3K-Akt signaling pathway. The interaction between CREB1 and NF-κB can modulate the transcription of downstream genes and thus contribute to the control of apoptosis and other processes. The mechanism of CREB1 in the PI3K-Akt signaling pathway is shown in Fig. 5E. The cAMP signaling pathway governs various intracellular processes, such as the modulation of cell proliferation, differentiation, and apoptosis via the activation of cAMP-dependent protein kinase (PKA) [22]. Phosphorylated PKA can then further phosphorylate CREB3 and activate its transcriptional activity. By binding to CBP, CREB3 regulates the transcription of specific genes and thus contributes to the control of various cellular physiological responses. In this way, CREB3 is crucial for cell growth and development, metabolic regulation, and stress response.

The underlying signaling mechanisms revealed by dark genes based on scRNA-seq data

Differential expression not only helps to reveal the secret of biological process, but also provides important theoretical basis for gene diagnosis and therapy. In many medical experiments and molecular studies, differentially expressed genes (DEGs) serve as markers or drug therapeutic targets, while some non-differentially expressed genes (non-DEGs) are often ignored, which will also have a significant role in biological processes and may be potential therapeutic biomarkers. In this study, genes with no differential expression but sensitive to the MIWE score are defined as dark genes, and differential MIWE analysis is performed on the two embryonic differentiation datasets to show the differences in MIWE values and gene expression of dark genes in the two datasets (Fig. 5A, B). There is a clear observation that gene expression remains relatively constant at each stage, while there are significant differences in MIWE values.

For mESC to MP data, it has been confirmed that some dark genes are closely related to embryonic differentiation, which are mainly involved in the regulation of chemical reactions in cells or organisms, macromolecular metabolism, and the frequency, rate or degree of gene expression and other biological processes. Extracellular STIP1 engages with diverse receptors to boost induced differentiation, cell proliferation, and protein synthesis [23]. Low expression of Receptor coactivator 3 (NCOA3) may lead to decreased differentiation potential of embryonic stem cells in vitro and in vivo [24]. CKS1B regulates cell cycle processes by engaging with cyclin-dependent kinase (CDK) and SCF complex to affect cell proliferation [25]. MDM2, an E3 ubiquitin ligase, plays a crucial role in the differentiation of various cell types, including osteoblasts and myoblasts [26].
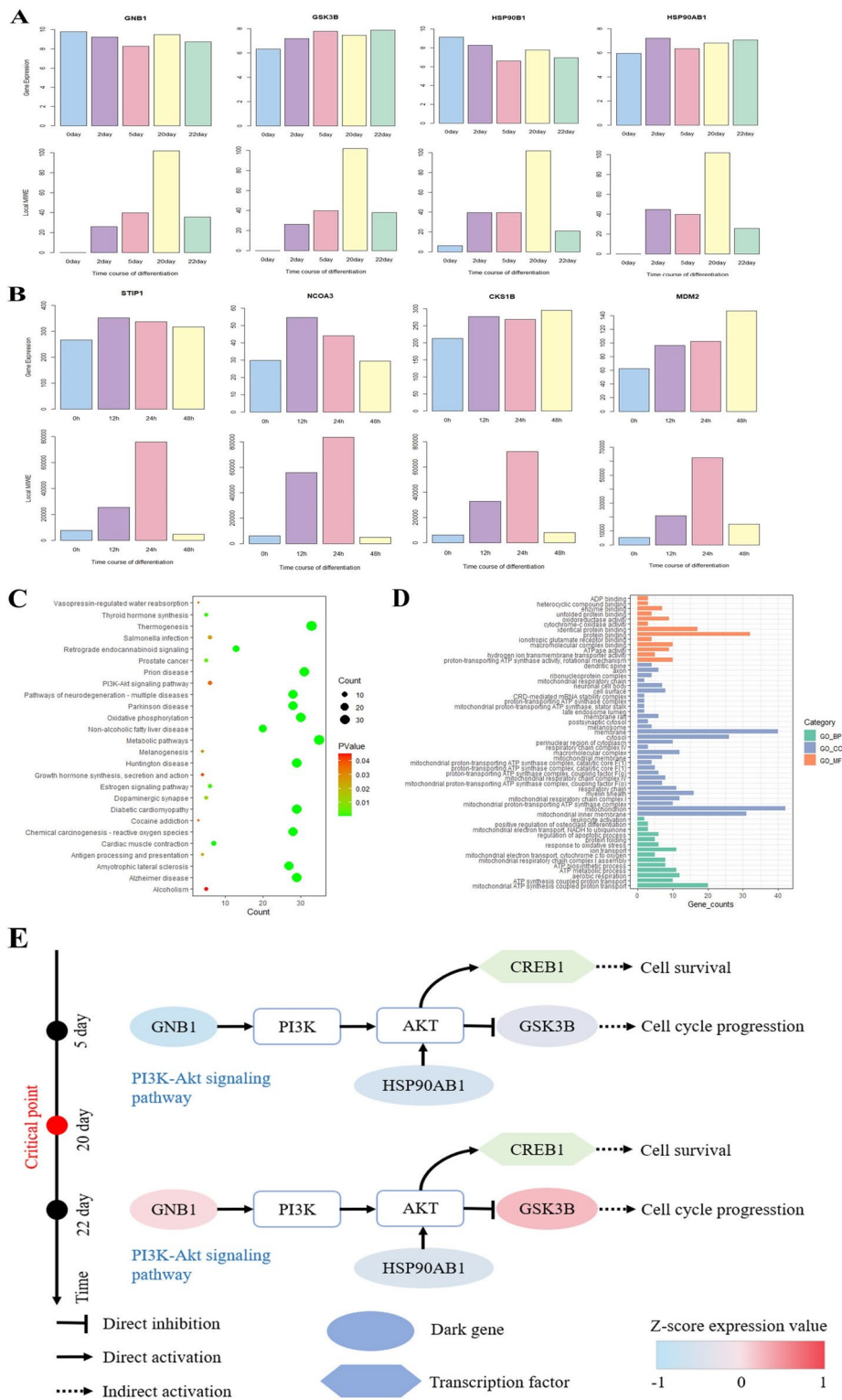
Fig.5 Potential regulatory mechanisms related to embryonic differentiation revealed by dark genes. Dynamic changes of gene expression and entropy of dark genes for **A** MEF to neurons, **B** mESC to MP. **C** Pathways enriched of MEF to neurons. **D** GO analysis of MEF to neurons. **E** The enrichment and regulation of related dark genes of MEF to neurons

To investigate the potential signaling mechanisms indicated by mouse dark genes and their domain genes, we conduct a series of functional analyses of dark genes from MEF to neurons (Fig. 5C, D). HSP90B1 participates in the Thyroid hormone synthesis pathway, in which synthetic thyroid hormones bind to nuclear receptors and control the expression of numerous genes associated with cell cycle regulation and differentiation [27]. In Prostate cancer pathway, HSP90AB1 and HSP90B1 can indirectly affect cell proliferation and survival by activating Ar and thus binding to DNA sites. GSK3B phosphorylates β-catenin to further activate Cyclin D1, an important regulatory factor of cell cycle [28], it can also lead to cell proliferation. The PI3K-Akt signaling pathway serves as a crucial hub governing cell growth, proliferation and metabolism in mammalian cells [29]. Figure 5E shows the potential mechanism of dark genes in MEF to neurons data and their domain genes in pathways. During embryonic differentiation, the high expression of GNB1 activates PI3K, which is then combined with HSP90 to activate the downstream target AKT of PI3K, HSP90 regulates various biological processes, such as cell growth, differentiation, and survival [30], AKT kinase translates diverse signals into intracellular cues governing cell survival, proliferation, metabolism, and differentiation [31] and transmits them to downstream genes, affecting cell proliferation and differentiation. The gene expression of the dark genes changes significantly between day 5 and day 22, and the recognized critical point could serve as a crucial time point to guide the differentiation of MEF to neurons.

Identifying the critical state during cancer progression

In addition to identifying the critical transition of embryonic differentiation, we also apply MIWE algorithm to two cancer datasets, COAD and THCA, and take healthy samples as the reference group to participate in the entropy calculation at each stage. In the second phase, local MIWE values in the COAD and THCA data increased significantly (Fig. 6A, B), which could be identified as a critical state of disease progression. The landscape map shows the dynamic changes of local MIWE values of signaling genes (Fig. 6C, D), which also indicated the abnormal system in the second stage. In addition, genes with the top 5% maximum local MIWE value at the critical stage are used as signaling genes, Changes in local MIWE values of all genes and dynamic changes of signaling genes in PPI network are shown in the Additional file 1: Fig. S3. Detection of critical points before disease progression or metastasis is conducive to timely clinical intervention for subsequent treatment. MIWE method can provide early warning signals in the course of disease development, which is helpful for disease treatment.

We use the Kaplan–Meier method for prognostic survival analysis of clinical samples from two cancers. By comparing the survival rate of each sample and its standard error, it can be observed that the prognosis of patients diagnosed before the critical state is significantly different from that of patients diagnosed after the critical stage, with P values less than 0.05 (Fig. 6E, F). Patients treated before deterioration have higher survival rate and longer survival time. More details of survival analysis are shown in the Additional file 1: Section E.
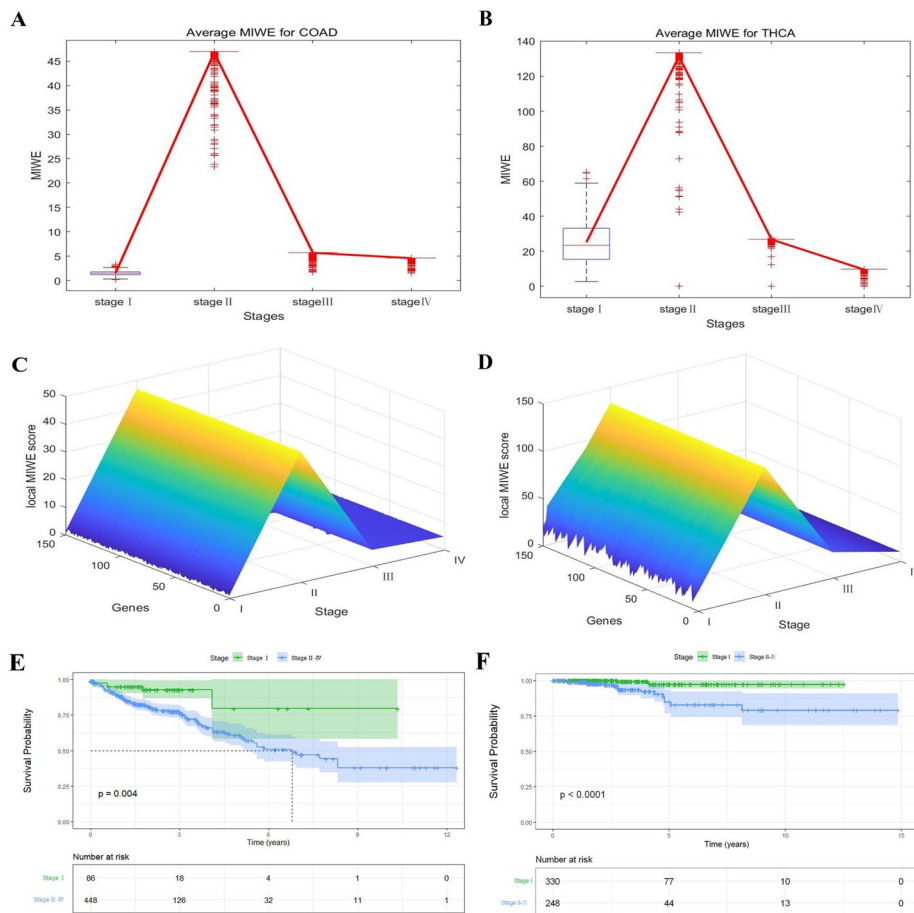
Fig.6 Detection of the critical point of cancer progression. The MIWE score for **A** COAD, **B** THCA. Landscapes of the local MIWE score for **C** COAD, **D** THCA. Survival analysis before and after the identified critical states for **E** COAD, **F** THCA

Functional analysis of the common MIWE signaling genes among two cancers

To comprehend the mechanism of signaling genes involved in disease development, we perform functional enrichment analysis of the common signaling genes of two cancers. The GO analysis results show that the signaling genes are mainly involved in the chemical reaction of protein formation in the cytoplasm, the macromolecular modification process of synthesis or assembly of ribonucleoprotein complexes, and the regulation of the rate of ubiquitin groups added to proteins (Fig. 7A). The lack of numerous ribosomal proteins can directly impact the overall translation process and the global expression of proteins, contributing to the onset of various diseases, including cancer [32]. Figure 7B shows the association between genes and biological processes. Elevated in numerous solid tumors, HSP90AB1 is believed to stimulate angiogenesis and facilitate cancer metastasis [33]. Heat shock protein family A (HSPA5) as a diagnostic and prognostic biomarker for various malignancies [34]. P4HB can influence tumor formation in a collagen-dependent or collagen-independent manner [35].

In addition, common signaling genes are involved in several pathways associated with cancer progression (Fig. 7C). MHC Class I and Class II antigen processing and presentation pathways present peptides to circulating CD8+cytotoxic T cells and CD4+helper
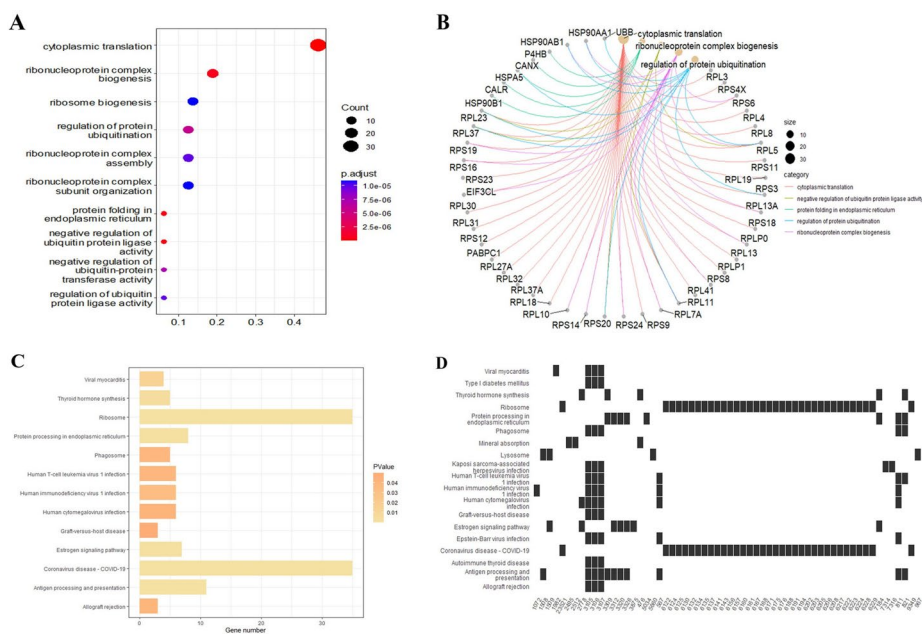
Fig.7 Functional analysis of common signaling genes in two cancers. **A** Common signaling genes involve in major biological processes. **B** The association of genes with biological processes. **C** Common signaling genes involve in cancer related pathways. **D** The association between genes and pathways, where the number represents the ENTREZ ID of the gene

T cells, respectively, to recognize pathogens and transform cells. Immune surveillance of transformed cells/tumor cells induces alterations in antigen processing and presentation pathways to evade immune response, which is an important process in tumor development [36]. Figure 7D shows the related pathways involved in each gene. β2-microglobulin (B2M) plays a physiological and pathological role in tumor cells [37]. In Antigen processing and presentation, the complex of B2M and HLA-B/C activates down-stream signals, upregulates and enhances T cell immunity, and plays an important role in controlling colon/rectal cancer growth [38]. Studies have shown that B2M is a potential tumor suppressor gene in COAD and has been identified as a potential biomarker for THCA [39]. Processing, modification, and folding of proteins in the endoplasmic reticulum (ER) are highly regulated procedures that dictate cell function, fate, and survival. Abnormal activation of the downstream signaling pathway of ER has been proven to be a key regulatory factor for tumor growth and metastasis [40]. Estrogen can affect tumor progression by regulating tumor microenvironment and plays a pivotal role in the occurrence and development of THCA [41]. GNAS is considered to be an oncogene that can be constitutionally activated by a specific point mutation of Guanine nucleotide binding protein alpha subunit (Gsα) in the Estrogen signaling pathway, thus activating multiple cancer-related pathways [42].

Discussion

Identifying critical states in complex biological systems is essential, such as critical stages of disease progression and cell fate commitments during embryonic development, early warning signs of disease progression that can prepare for treatment, and

understanding cell fate commitment that can build individual specific disease models. However, identifying critical transitions in complex biological systems is often challenging, and real biological datasets have strong noise and cannot characterize the dynamics of biological processes. In this study, we propose MIWE method for identifying cell fate transitions and complex disease critical states. The MIWE score quantifies the dynamic differences of mutual information networks at each stage based on weighted differential entropy at each time point, and converts gene expression values into probabilities to minimize the influence of strong noise. To verify the validity of the MIWE algorithm, the method is utilized on one simulated dataset and four real datasets, encompassing two scRNA-seq datasets and two bulk sequencing datasets.

Based on the MIWE method, we successfully detect the critical states the dynamic processes of complex biological systems. The function analysis of signaling genes in critical stage reveals the important role of signaling genes in embryonic differentiation or cancer development. In addition, we focus on exploring the potential signaling mechanisms of some non-differential signaling genes in embryonic differentiation pathways. Although they are not DEGs, the pathways involve are highly related to cell differentiation.

MIWE method is model-free and suitable for both bulk and single-cell expression data. However, MIWE also has limitations, as undirected networks are used in the construction of networks, which ignore causal relationships between nodes compared with directed networks. In addition, the joint distribution of two genes is binary normal distribution if and only if any linear combination of them follows a normal distribution. In general, the MIWE method helps to identify and detect critical states in complex biological systems, providing a theoretical basis for timely clinical intervention and disease modeling.

Conclusions

In this study, we propose a new method, mutual information weighted entropy (MIWE), which identifies critical states by quantifying the molecular dynamic differences at each stage by calculating the weighted differential entropy of each stage of the global network. The robustness of the proposed method under the influence of different noises is verified by numerical simulation. In addition, we identify two key transcription factors (TFs), CREB1 and CREB3, which are involved in cell proliferation and differentiation by regulating downstream signaling genes. The dark genes in the single-cell expression dataset are mined to reveal the potential pathway regulation mechanisms involved.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-024-05667-z.

> **Additional file 1.** Supplementary materials, figures, tables.

Author contributions
YX and XP designed the research; YX and XP performed the research; YX and XP analyzed and interpreted the data; YX wrote the manuscript; PL supervised and reviewed the manuscript. PL supported the funding. All authors have approved the manuscript.

Xie *et al. BMC Bioinformatics*   (2024) 25:44

Page 16 of 17

Availability of data and materials
To ensure reproducible results, all data can be found here: http://www.ncbi.nlm.nih.gov/geo/ and http://cancergenome.nih.gov, and the original code are available at https://github.com/xykxingchen/MIWE.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Liu R, Wang XD, Chen LN, et al. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev. 2014;34:455–78.
2. Chen LN, 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.
3. Bargaje R, Trachana K, Shelton MN, et al. Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. Proc Natl Acad Sci U S A. 2017;114:2271–6.
4. Zhong JY, Han CY, Zhang XH, et al. scGET: predicting cell fate transition during early embryonic development by single-cell graph entropy. Genom Proteom Bioinform. 2021;19:461–74.
5. Liu R, Chen P, Chen LN. Single-sample landscape entropy reveals the imminent phase transition during disease progression. Bioinformatics. 2020;36:1522–32.
6. Treutlein B, Lee QY, Camp JG, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature. 2016;534:391–5.
7. Semrau S, Goldmann JE, Soumillon M, et al. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat Commun. 2017;8:1096.
8. Peng H, Zhong JY, Chen P, et al. Identifying the critical states of complex dis-eases by the dynamic change of multivariate distribution. Brief Bioinform. 2022;23:bbac177.
9. Deichmann U, Schuster S, Mazat JP, et al. Commemorating the 1913 Michaelis-Menten paper Die Kinetik der Invertinwirkung: three perspectives. FEBS J. 2014;281:435–63.
10. Metzler CM, Tong DDM. Computational problems of compartment models with Michaelis-Menten-type elimination. Pharmaceutic Sci. 1981;7013:733–7.
11. Saganuwan SA. Application of modified Michaelis - Menten equations for determination of enzyme inducing and inhibiting drugs. BMC Pharmacol Toxicol. 2021;22:57.
12. Srinivasan B. A guide to the Michaelis-Menten equation: steady state and beyond. FEBS J. 2021;289:6086–98.
13. Yan JL, Li PL, Gao R, Li Y, Chen LN. Identifying critical states of complex diseases by single-sample Jensen-Shannon divergence. Front Oncol. 2021;11:684781.
14. Nakajima H. Role of transcription factors in differentiation and reprogramming of hematopoietic cells. Keio J Med. 2011;60:47–55.
15. Evans CM, Jenner RG. Transcription factor interplay in T helper cell differen-tiation. Brief Funct Genomics. 2013;12:499–511.
16. Deng QN, Guo T, Zhou X, et al. Cross-talk between mitochondrial fusion and the hippo pathway in controlling cell proliferation during drosophila development. Genetics. 2016;4:1777–88.
17. Zou Z, Ocaya PA, Sun H, et al. Targeted Vezf1-null mutation impairs vascular structure formation during embryonic stem cell differentiation. Arterioscler Thromb Vasc Biol. 2010;30:1378–88.
18. Alyssa M, Omar L, Sayane S, et al. Sp3 is essential for normal lung morpho-genesis and cell cycle progression during mouse embryonic development. Development. 2023;150:dev200839.
19. Roy AL. Pathophysiology of TFII-I: old guard wearing new hats. Trends Mol Med. 2017;23:501–11.
20. Feng YR, Raza SHA, Liang CC, et al. CREB1 promotes proliferation and differentiation by mediating the transcription of CCNA2 and MYOG in bovine myoblasts. Int J Biol Macromol. 2022;216:32–41.
21. Sampieri L, Funes Chabán M, Di Giusto P, et al. CREB3L2 modulates nerve growth factor-induced cell differentiation. Front Mol Neurosci. 2021;14:650338.

22. Yan K, Gao LN, Cui YL, et al. The cyclic AMP signaling pathway: exploring targets for successful drug discovery (Review). Mol Med Rep. 2016;13:3715–23.

23. Tan JSY, Lee B, Lim J, et al. Parkinson's disease-specific autoantibodies against the neuroprotective Co-Chaperone STIP1. Cells. 2022;11:1649.

24. Wu ZT, Yang M, Liu HJ, et al. Role of nuclear receptor coactivator 3 (Ncoa3) in pluripotency maintenance. J Biol Chem. 2012;287:38295–304.

25. Liu XT, Zhao DF. CKS1B promotes the progression of hepatocellular carcinoma by activating JAK/STAT3 signal pathway. Anim Cells Syst. 2021;25:227–34.

26. Zheng H, Yang G, Fu J, et al. Mdm2 promotes odontoblast-like differentiation by ubiquitinating Dlx3 and p53. J Dent Res. 2020;99:320–8.

27. Pascual A, Aranda A. Thyroid hormone receptors, cell growth and differentiation. Biochim Biophys Acta. 2013;1830:3908–16.

28. Montalto FI, De AF. Cyclin D1 in cancer: a molecular connection for cell cycle control, adhesion and invasion in tumor and stroma. Cells. 2020;9:2648.

29. Pompura SL, Dominguez VM. The PI3K/AKT signaling pathway in regulatory T-cell development, stability, and function. J Leukoc Biol. 2018;103:1065–76.

30. Abdullah H, Marwan ES, Hassan N. The HSP90 family: structure, regulation, function, and implications in health and disease. Int J Mol Sci. 2018;19:2560.

31. Abdullah L, Hills LB, Winter EB, et al. Diverse roles of Akt in T cells. Immunometabolism. 2021;3:e210007.

32. Reza AMMT, Yuan YG. microRNAs mediated regulation of the ribosomal proteins and its consequences on the global translation of proteins. Cells. 2021;10:110.

33. Wang HN, Deng GX, Ai M, et al. Hsp90ab1 stabilizes LRP5 to promote epithelial–mesenchymal transition via activating of AKT and Wnt/β-catenin signaling pathways in gastric cancer progression. Oncogene. 2019;38:1489–507.

34. Dong WG, Du DW, Huang H. HSPA5 is a prognostic biomarker correlated with immune infiltrates in thyroid carcinoma. Endokrynol Pol. 2022;73:680–9.

35. Shi R, Gao SS, Zhang J, et al. Collagen prolyl 4-hydroxylases modify tumor progression. Acta Biochim Biophys Sin. 2021;53:805–14.

36. Reeves E, James E. Antigen processing and immune regulation in the response to tumours. Immunology. 2016;150:16–24.

37. Lin HL, Wang KL, Zou KB, et al. Analysis of the B2M expression in colon ade-nocarcinoma and its correlation with patient prognosis. Evid Based Complement Alternat Med. 2022;2022:7264503.

38. Michelakos T, Kontos F, Kurokawa T, et al. Differential role of HLA-A and HLA-B, C expression levels as prognostic markers in colon and rectal cancer. J Immunother Cancer. 2022;10:e004115.

39. Jasim A, Mohammed A, Ibrahim A. Beta-2-microglobulin as a marker in patients with thyroid cancer. Iraqi Postgrad Med J. 2019;18:6.

40. Chen X, Cubillos-Ruiz JR. Endoplasmic reticulum stress signals in the tumour and its microenvironment. Nat Rev Cancer. 2021;21:71–88.

41. Liu J, Xu TM, Ma L, et al. Signal pathway of Estrogen and estrogen receptor in the development of thyroid cancer. Front Oncol. 2021;11:593479.

42. Wang Z, Jing CW, Cao HX, et al. Rare and novel GNAS gene mutations in Chinese patients with thyroid cancer. Precis Med Sci. 2021;10:83–5.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.