Research
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Identification of Critical States in Complex Biological Systems Using Cell-Specific Causal Network Entropy

Jiayuan Zhong[1†], Ziyi Huang[2†], Jianqiang Qiu[1], Fei Ling[2*], Pei Chen[3*], and Rui Liu[3*]

[1]School of Mathematics, Foshan University, Foshan 528000, China. [2]School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510640, China. [3]School of Mathematics, South China University of Technology, Guangzhou 510640, China.

*Address correspondence to: fling@scut.edu.cn (F.L.); chenpei@scut.edu.cn (P.C.), scliurui@scut.edu.cn (R.L.)
†These authors contributed equally to this work.

Citation: Zhong J, Huang Z, Qiu J, Ling F, Chen P, Liu R. Identification of Critical States in Complex Biological Systems Using Cell-Specific Causal Network Entropy. *Research* 2025;8:Article 0852. https://doi.org/10.34133/research.0852

Submitted 9 June 2025
Revised 29 July 2025
Accepted 30 July 2025
Published 26 August 2025

Copyright © 2025 Jiayuan Zhong et al. Exclusive licensee Science and Technology Review Publishing House. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Abrupt shifts, referred to as critical transitions, are frequently observed in complex biological systems, characterized by marked qualitative changes occurring from one stable state to another through a pre-transitional/critical state. Pinpointing such critical states, along with the signaling molecules, can provide valuable insights into the fundamental mechanisms of intricate biological processes. However, the identification and early warning of the critical state remains a challenge, particularly in model-free cases with high-dimensional single-cell data, where traditional statistical methods often prove inadequate due to the inherent sparsity, noise, and heterogeneity of the data. In this study, we propose a novel quantitative method, cell-specific causal network entropy (CCNE), to infer the specific causal network for each cell and quantify dynamic causal changes, thereby enabling the identification of critical states in complex biological processes at the single-cell level. We validated the accuracy and effectiveness of the proposed approach through numerical simulations and 5 distinct real-world single-cell datasets. Compared to existing methods for detecting critical states, the proposed CCNE exhibits enhanced effectiveness in identifying critical transition signals. Moreover, CCNE score is a computational tool for distinguishing temporal changes in cellular heterogeneity and demonstrates satisfactory performance in clustering cells over time. In addition, the reliability of CCNE is further emphasized through the functional enrichment and pathway analysis of signaling molecules.

Introduction

The complex biological processes, such as pericyte-to-neuron transition [1], pluripotency-to-hepatocyte transition [2], and epithelial-to-mesenchymal transition [3], involve pre-transition or critical states where marked and qualitative shifts occur. From a dynamics viewpoint, complex biological processes are typically seen as a time-dependent nonlinear dynamical system, characterized by 3 main phases: a before-transition state with stability and resilience, a pre-transition state marked by high sensitivity and instability where cell state transition occurs, and a subsequent stable after-transition state (Fig. 1A) [4]. Identifying the critical state and associated key molecules is pivotal for many biological phenomena, such as cell differentiation, where it is crucial for cellular reprogramming and holds great significance for progress in regenerative medicine [5]. However, the high dimensionality, non-linearity, and dynamic complexity of biological systems make it difficult to detect precursory signs of the pre-transition state, especially when working with inherently sparse, noisy, and heterogeneous single-cell datasets. Although various methods have achieved effective outcomes in single-cell data analysis, particularly in cell clustering, trajectory inference, and cellular heterogeneity [6–8], the task of analyzing molecular causal regulatory relationships and predicting pre-transition states continues to present a great challenge [9,10].

Our recently proposed concept, the dynamic network biomarker (DNB), rooted in nonlinear dynamical systems theory, has been developed to identify critical states and predict key molecules in complex biological processes by focusing on groups of collectively fluctuating genes. The DNB method and its extended versions have been successfully applied to identify pre-disease states for complex diseases [11–13], uncover cell-fate transitions during embryonic development [14], and explore the mechanism of immune checkpoint blockades [15]. However, these approaches mainly aim to detect early warning signals of critical transitions by utilizing the dynamical characteristics of critical states derived from bulk omics data and still continue to face robustness issues in high-noise data, particularly evident in single-cell transcriptomic analysis. Among omics data, single-cell data stand out for their ability to uncover gene associations and offer a wealth of unprecedented information crucial for

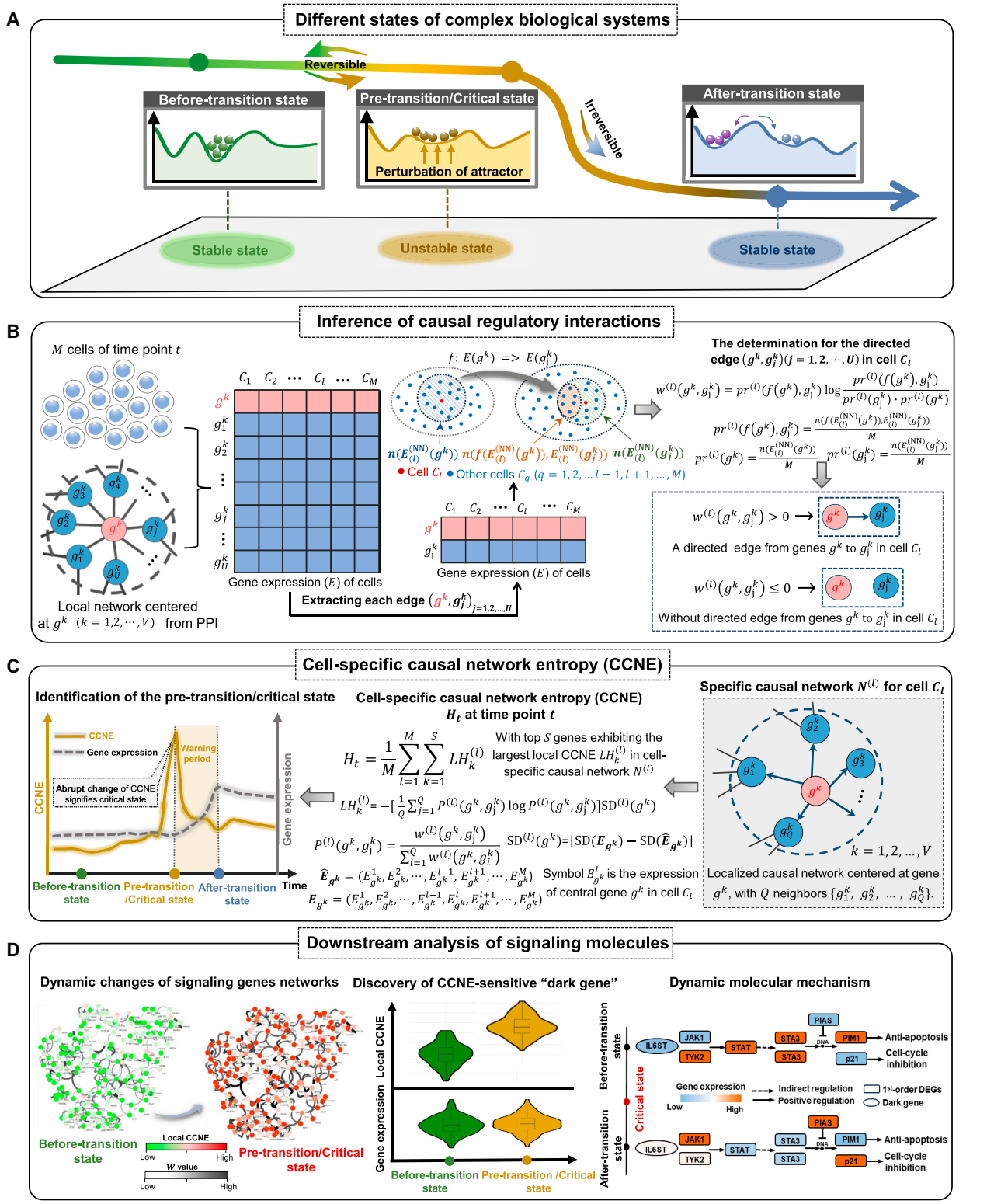


Fig. 1. A schematic depiction of the proposed CCNE for identifying pre-transition states in complex biological systems. (A) The complex biological process can be generally classified into 3 distinct states: a stable before-transition state, a highly sensitive and unstable pre-transition state where cell state transition occurs, and another stable after-transition state. (B) The causal relationship and its strength between 2 molecules can be inferred using the continuity scaling law, allowing the construction of cell-specific causal networks for each cell. (C) The local CCNE value is computed for each localized causal network, while the CCNE at each time point is utilized to assess the criticality of the complex biological system process, with a marked increase indicating an approaching pre-transition or critical state. (D) Three main analyses are performed using CCNE: the dynamic changes in signaling gene networks, identification of CCNE-sensitive "dark genes", and exploration of functional pathways implicated in signaling gene.

detecting critical states [16,17]. Causality represents a directional relationship between 2 variables and is a fundamental interaction in complex systems, providing valuable insights to advance our understanding of system dynamics. For instance, causality metrics derived from Takens' embedding theorem effectively capture the relationship between non-linearly coupled nodes and tackle a key causal inference issue related to non-separability [18]. Combining causality measures with the early-warning signal framework has the potential to enhance the robustness and informativeness of pinpointing critical transitions or system bifurcations [18,19]. Therefore, it is necessary to develop innovative causality network-based methods tailored to single-cell expression data, enabling the identification of pre-transition states in complex biological processes and the discovery of the associated key molecules.

In this study, from the perspective of gene-level causal relations, we introduce a novel quantitative method, cell-specific causal network entropy (CCNE), to infer a distinct causal network for each cell and identify the pre-transition state of complex biological system by analyzing the dynamic changes in molecular causal relationships. Specifically, the cell-specific causal network is constructed based on the continuity scaling law [20,21], a rigorous mathematical framework consistent with the natural interpretation of functional dependency, allowing for the quantification of "continuous causality" and its causal strength between 2 molecules through varying neighbor size (Fig. 1B). Furthermore, the local CCNE is calculated to evaluate the dynamic alterations of causal relationships among molecules within localized causal networks, and the CCNE at each time point serves to quantify the criticality of complex biological process, with a marked increase indicating an impending pre-transition or critical state (Fig. 1C). To demonstrate the reliability and effectiveness of CCNE, we conducted a verification using a numerical simulation and 5 distinct real-world single-cell datasets. Our results show that CCNE is more effective and accurate in pinpointing pre-transition states compared to existing methods, with the predicted pre-transition states aligning well with experimental observations. Furthermore, by converting the sparse single-cell expression matrix into a non-sparse entropy matrix, CCNE can effectively distinguish cellular heterogeneity over time and facilitate analysis of temporal clusters. Besides, we highlight the effectiveness of CCNE signaling biomarkers in downstream functional analyses (Fig. 1D). Therefore, we present an innovative computational method tailored for single-cell data, which provides a new framework to track the dynamics of biological systems in terms of cell-specific causal networks.

Results

Performance and resilience of CCNE approach for simulation data

To evaluate how well the CCNE approach performs and its ability to withstand noise condition, we utilized an 8-node regulatory network (depicted in Fig. 2A) determined by a set of stochastic differential equations (Eq. S1) to simulate the behavior and evolution of complex biological systems. Such a regulatory network based on Michaelis-Menten form or Hill bifurcation is commonly applied to model how genes regulate activities in biological systems, including transcription and translation [22,23]. The network undergoes a critical transition governed by the parameter s, where $s = 0$ denotes the bifurcation point. Nodes 1 to 5 serve as DNBs or signaling molecules, while the other nodes are classified as non-DNBs. We performed numerical simulations, varying parameter s from −0.5 to 0.2, to illustrate the effectiveness of our CCNE approach in identifying the critical phase of the system near the bifurcation point. More details about the dynamical system are located in Section A of the Supplementary Materials.

As illustrated in Fig. 2B, the CCNE score demonstrates a marked increase near the bifurcation or tipping point ($s = 0$), signalling an imminent critical state. Conversely, when the system is distant from this point, the CCNE score consistently remains low. Moreover, the dynamics landscape of the local CCNE score in Fig. 2C is depicted to showcase the specific behaviors of each node or molecule. Clearly, when the system is far from the critical state, the local CCNE scores of all nodes remain low and stable, while certain nodes identified as DNBs (nodes 1 to 5) show a notable rise as the system approaches the critical state. Furthermore, the dynamic evolution of the regulatory network is showcased in Fig. 2D, revealing a dramatic change in the configuration of DNB subnetworks near the critical state, which indicates an impending transition in the network state. Besides, we conducted a comparative analysis between CCNE and molecular expression using different noise-perturbed samples to highlight the resilience of the proposed method (Fig. 2E). With increasing noise strength σ, our CCNE method proved to be more robust and effective in detecting critical signal. The numerical experiment illustrates that the proposed SCNE method identifies pre-transition or critical states with reliability and accuracy.

Identification of the pre-transition state for different biological processes

To illustrate the functionality of the proposed CCNE, we implemented it on 5 distinct single-cell datasets: pericyte-to-neuron transition [24], mouse embryonic fibroblast (MEF)-to-neuron transition [25], mouse hepatoblast cell (MHC)-to-hepatocyte and cholangiocyte cell (HCC) transition [26], epithelial cell deterioration (EPCD) transition [27], and induced pluripotent stem cell (iPSC)-to-mature hepatocyte (MH) transition [28]. The CCNE score specific to each cell was determined using Eq. 5 presented in Materials and Methods. Then, the average CCNE score at a specific time point was employed to detect potential critical signals. For the above datasets, our CCNE approach effectively pinpointed the pre-transition states during biological processes, demonstrating its accuracy and effectiveness. The algorithm's source code can be freely accessed at https://github.com/zhongjiayuan-fs/CCNE_projects.

In pericyte-to-neuron data, Fig. 3A illustrates a notable rise in the average CCNE score on day 7 ($P = 7.98\text{E}{-20}$), indicating an early indication of differentiation into induced neurons [24]. For MEF-to-neuron data, the average CCNE score depicted in Fig. 3B has shown an upward trend from day 5 to day 20 ($P = 2.77\text{E}{-5}$), after which it was observed that mouse embryonic intermediate cells are differentiated into induced neurons [25]. For MHC-to-HCC data, as presented in Fig. 2C, a significant change of average CCNE score is evident at embryonic day 12.5 (E12.5) ($P = 0.012$). This result is consistent with findings from the original experiment that hepatoblasts differentiate into hepatocytes and cholangiocytes after E12.5 [26]. For the non-time-series single-cell dataset of EPCD, the progression of EPCD can be categorized into 6 distinct clusters by constructing a pseudo-temporal trajectory (see Section B of the Supplementary

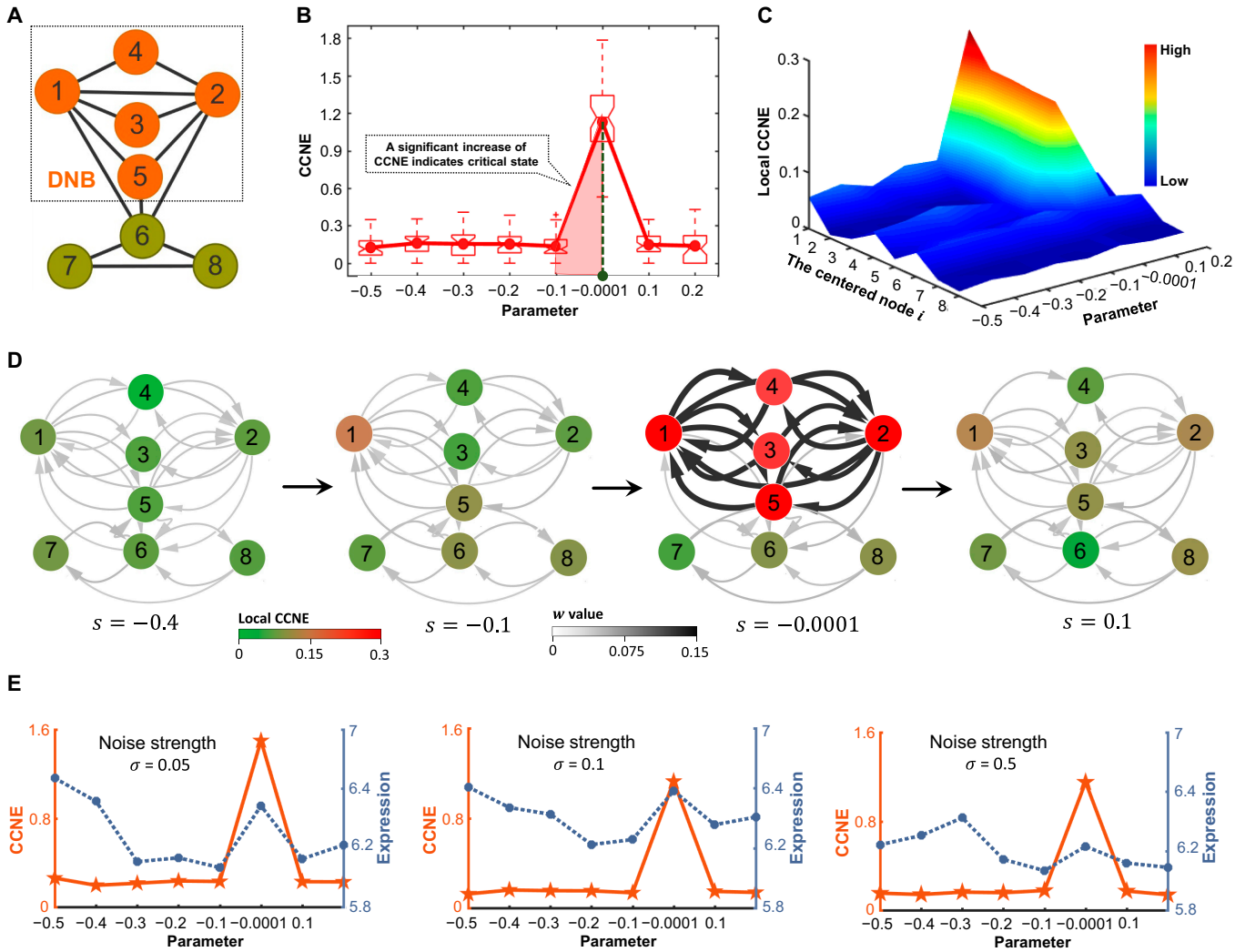
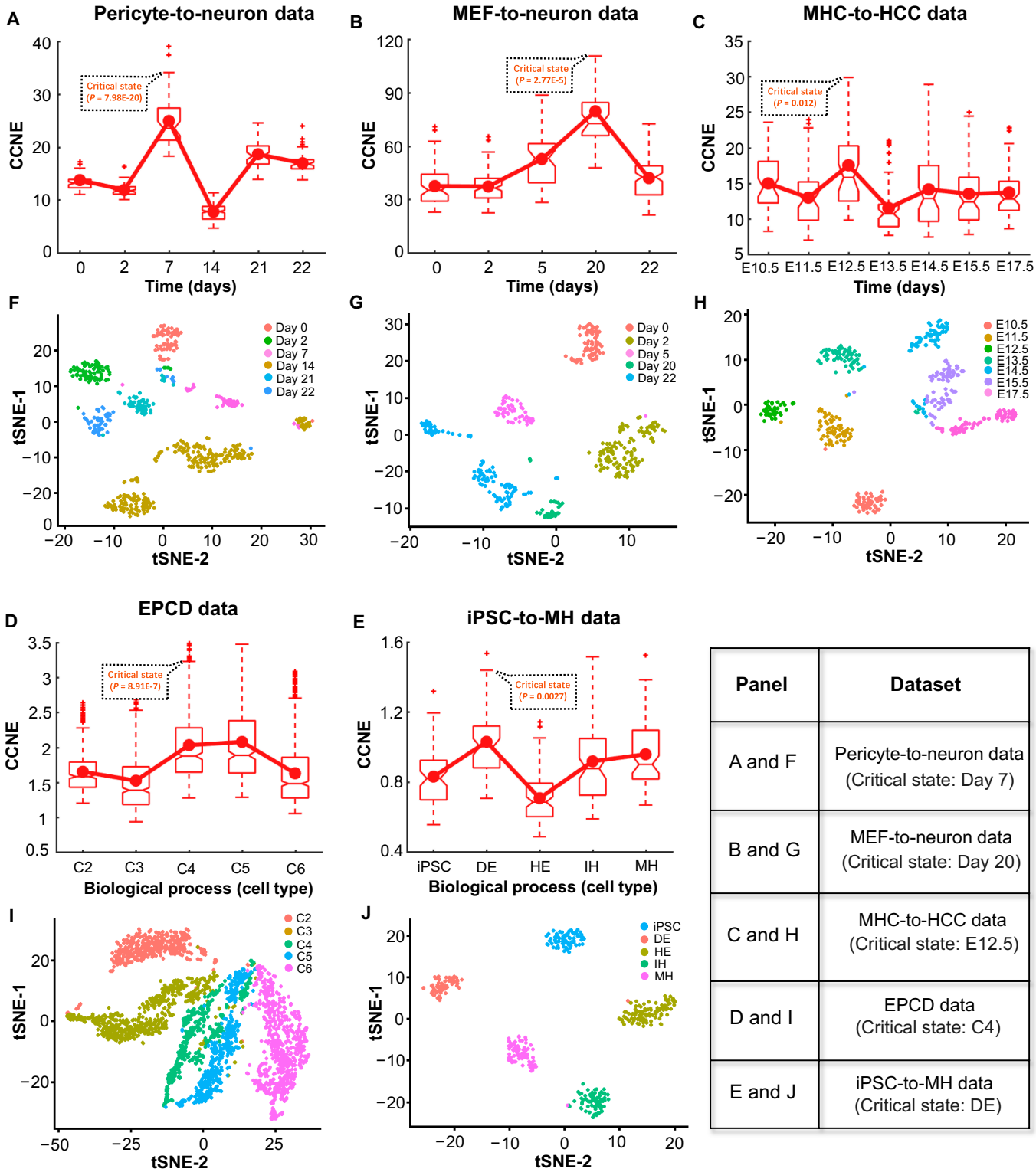


Fig. 2. Performance and resilience of CCNE approach for simulation data. (A) An 8-node regulatory network employed as the foundation for generating numerical simulations. (B) The CCNE score exhibits a sharp increase near the bifurcation or tipping point ($s = 0$). (C) The dynamics landscape of local CCNE score is illustrated across different nodes. Remarkably, certain nodes identified as DNBs (nodes 1 to 5) show a marked increase when the system approaches the critical state. (D) The dynamic evolution of the regulatory network reveals a pronounced change in the configuration of DNB subnetworks near the bifurcation point. (E) A comparison between the performance of CCNE and molecular expression reveals that the CCNE method is more robust and effective in detecting critical signals.

Materials for details). It is seen from Fig. 3D that a marked shift ($P = 8.91E{-}7$) in the average CCNE score appears at cluster 4 (C4), signaling a critical transition toward deteriorated epithelial cells and uncovering a subset of epithelial cells in the pre-deterioration stage [27]. When applied to iPSC-to-MH data, Fig. 3E shows a significant increase ($P = 0.0027$) in the average CCNE score observed in the definitive endoderm (DE) cell type, followed by the transition of DE cells into hepatic endoderm (HE) cell types [28]. In addition, the median values displayed in the red box plot of the CCNE score in Fig. 3A to E emphasize the robustness of our CCNE in pinpointing critical state. Besides, the proposed CCNE achieves better performance than existing critical state detection methods, including single-cell gene association entropy (SGAE) [29], single-sample landscape entropy (SLE) [30], BioTIP [31], and landscape DNB (LDNB) [32], for identifying pre-transition states during complex biological processes (Table 1). However, from the perspective of gene expression patterns, the dynamic changes and critical states of complex biological processes cannot be characterized as effectively by the CCNE value (Fig. S2). In addition to identification of the pre-transition state, our approach can transform gene expression data into a CCNE matrix, enabling CCNE-based cell clustering analysis. As illustrated in Fig. 3F to J, CCNE-based clustering analysis can effectively group cells from different time points or types utilizing t-distributed stochastic neighbor embedding (t-SNE) [33]. Similarly, the Uniform Manifold Approximation and Projection (UMAP) results distinguish cellular states at various time points as effectively as the t-SNE analysis, indicating that the day-to-day transitions can also be captured by CCNE scores using UMAP-based clustering (Fig. S3).

Evolution of regulatory network and identification of "dark genes" for signaling genes

At the determined critical state, the top 5% of genes exhibiting the highest local CCNE values were selected as candidate signaling genes for further functional and biological process analysis. Specifically, they were mapped onto the protein–protein interaction (PPI) network, from which the most extensive



Panel	Dataset
A and F	Pericyte-to-neuron data (Critical state: Day 7)
B and G	MEF-to-neuron data (Critical state: Day 20)
C and H	MHC-to-HCC data (Critical state: E12.5)
D and I	EPCD data (Critical state: C4)
E and J	iPSC-to-MH data (Critical state: DE)

Fig. 3. Identification of the pre-transition state for complex biological process. The performance of the CCNE approach for 5 single-cell datasets related to cell-fate transition processes: (A) pericyte-to-neuron data, (B) MEF-to-neuron data, (C) MHC-to-HCC data, (D) EPCD data, and (E) iPSC-to-MH data. Temporal clustering analysis based on the CCNE matrix was conducted for 5 single-cell datasets of different biological processes: (F) pericyte-to-neuron data, (G) MEF-to-neuron data, (H) MHC-to-HCC data, (I) EPCD data, and (J) iPSC-to-MH data.

connected subgraph was extracted for exploring the dynamic changes of the signaling genes network. For pericyte-to-neuron data, a noticeable shift in the network structure on day 7 was observed (Fig. 4A), implying that cell fate begins to transition toward the differentiation into induced neurons after day 7 [24]. When applied to the MEF-to-neuron data, there is a distinct alteration in the network structure on day 20 (Fig. 4B), suggesting an impending critical transition of mouse embryonic

Table 1. Comparison of the performance among different critical state detection methods

Dataset	Pericyte to neuron	MEF to neuron	MHC to HCC	EPCD	iPSC to MH
CCNE	Day 7 ($P = 7.98E{-}20$)	Day 20 ($P = 2.77E{-}5$)	E12.5 ($P = 0.012$)	Cluster 4 ($P = 8.91E{-}7$)	DE ($P = 0.0027$)
SGAE	Day 7 ($P = 6.28E{-}4$)	Day 20 ($P = 0.0027$)	E12.5 ($P = 0.023$)	None	None
SLE	Day 14 ($P = 0.064$)	Day 22 ($P = 2.72E{-}7$)	E14.5 ($P = 0.0039$)	Cluster 6 ($P = 2.18E{-}5$)	MH ($P = 2.11E{-}6$)
BioTIP	Day 14 ($P = 7.53E{-}10$)	None	E15.5 ($P = 0.008$)	Cluster 5 ($P = 1.69E{-}21$)	DE ($P = 0.0082$)
LDNB	Day 14 ($P = 0.053$)	Day 20 ($P = 1.71E{-}7$)	E13.5 ($P = 0.013$)	Cluster 6 ($P = 0.024$)	IH ($P = 1.13E{-}6$)

MEF, mouse embryonic fibroblast; MHC, mouse hepatoblast cell; HCC, hepatocyte and cholangiocyte cell; EPCD, epithelial cell deterioration; iPSC, induced pluripotent stem cell; MH, mature hepatocyte; CCNE, cell-specific causal network entropy; SGAE, single-cell gene association entropy; SLE, single-sample landscape entropy; LDNB, landscape dynamic network biomarker

intermediate cells differentiating into induced neurons by day 22 [25]. The overall dynamic evolution of the regulatory network for these 2 datasets is illustrated in Fig. S4. Moreover, signaling genes are utilized to discover non-differential genes sensitive to the CCNE score, which may play an important role in biological processes. Despite its effectiveness in discovering new drug targets and biomarkers, differential expression analysis typically fails to consider non-differentially expressed genes (non-DEGs) that are vital to key biological functions. It is evident that such genes are involved in immune cell function [34] and are commonly associated with development pathways [35]. In our study, some genes among DNBs or signaling molecules do not exhibit differential expression at the molecular level but are highly responsive to variations in the CCNE score. Signaling genes are categorized as "dark genes" and meet 2 criteria: (a) they do not differ significantly at the level of gene expression, and (b) they demonstrate a significant distinction between before-transition and pre-transition states based on CCNE score. Specifically, focused on signaling molecules, we compared the dynamic changes of each gene at both the gene expression and CCNE score levels to discover "dark genes". As depicted in Fig. 4C and D, some "dark genes" identified in the pericyte-to-neuron and MEF-to-neuron data show no significant differences in gene expression, yet there are notable changes in the CCNE score. Additional dark genes for pericyte-to-neuron, MEF-to-neuron, and iPSC-to-MH data can be found in Section F of the Supplementary Materials. Notably, the CCNE score of "dark genes" exhibits marked upward trends, serving as an early indication of the upcoming pre-transition state. This finding also underscores the capability of the CCNE algorithm to accurately forecast impending critical states. In addition, Table 2 illustrates that certain "dark genes" have been demonstrated to play crucial functional roles in the associated biological processes. Moreover, according to Gene Ontology enrichment analysis, the "dark genes" are enriched in pathways functionally associated with cellular proliferation, tissue differentiation, and morphogenesis during embryonic development (Fig. S5). In addition, these "dark genes" are capable of regulating various non-coding RNAs (Fig. S6A and B). Specifically, *ITGA7*, *MMP2*, and *LAMTOR1* are likely to interact with specific microRNAs (miRNAs) that are mainly enriched in embryogenesis-related pathways (Fig. S6C), such as the Wnt signaling pathway [36], the JAK-STAT signaling pathway [37], and the FoxO signaling pathway [38], suggesting their potential roles as regulatory factors during embryonic development. Similarly, "dark genes" including *Ccnd2, Dctn2, Fars2, and Ripk1* may contribute to embryonic development through the regulation of miRNAs associated with developmental pathways (Fig. S6D).

Uncovering potential regulatory mechanisms in cell development and differentiation

Monocle analysis was performed on single cells from various cell types to infer the lineage relationships during iPSC development (Fig. 5A). Subsequently, by heatmap visualization of Monocle-derived gene expression, we uncovered the temporal sequence of gene expression events during differentiation (Fig. 5B). These temporally ordered changes in gene expression of signaling molecules and first-order neighbor may regulate iPSC differentiation, revealing the lineage progression from iPSC cells to DE, HE, immature hepatocyte (IH), and MH cells. In addition, Fig. 5C shows that the primary enriched pathways for dark genes or a specific group of signaling molecules are closely related to pluripotent stem cell development and differentiation. To further explore the regulatory mechanisms of iPSC differentiation, we analyzed the dynamic changes in the subnetwork formed by dark genes and their first-order DEG neighbors within the PPI network (Fig. 5D), which offers a detailed insight into the interactions and relationships influencing iPSC cell differentiation. In our research, the first-order DEG neighbors are identified based on the following 2 criteria: (a) they are directly connected to dark genes in the PPI network, and (b) they are DEGs, exhibiting statistically significant expression changes before and after the identified critical transition state. Before and after the critical state, there was a substantial shift in gene expression within the subnetwork, marked by a marked increase from low to high levels. Additionally, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed that first-order DEG neighbors are primarily enriched in pathways related to cell development and differentiation (Fig. 5E). For example, the cell cycle pathway was crucial in regulating cell proliferation and differentiation [39], while the mitogen-activated protein kinase signaling pathway influenced various cellular processes, including differentiation and development [40]. Furthermore, the JAK-STAT signaling pathway was critical in regulating iPSC reprogramming, maintaining pluripotency, and guiding differentiation [41]. Specifically, we found that the long-range signaling axis involving *IL6ST* and *JAK* may exhibit dynamic spatiotemporal changes

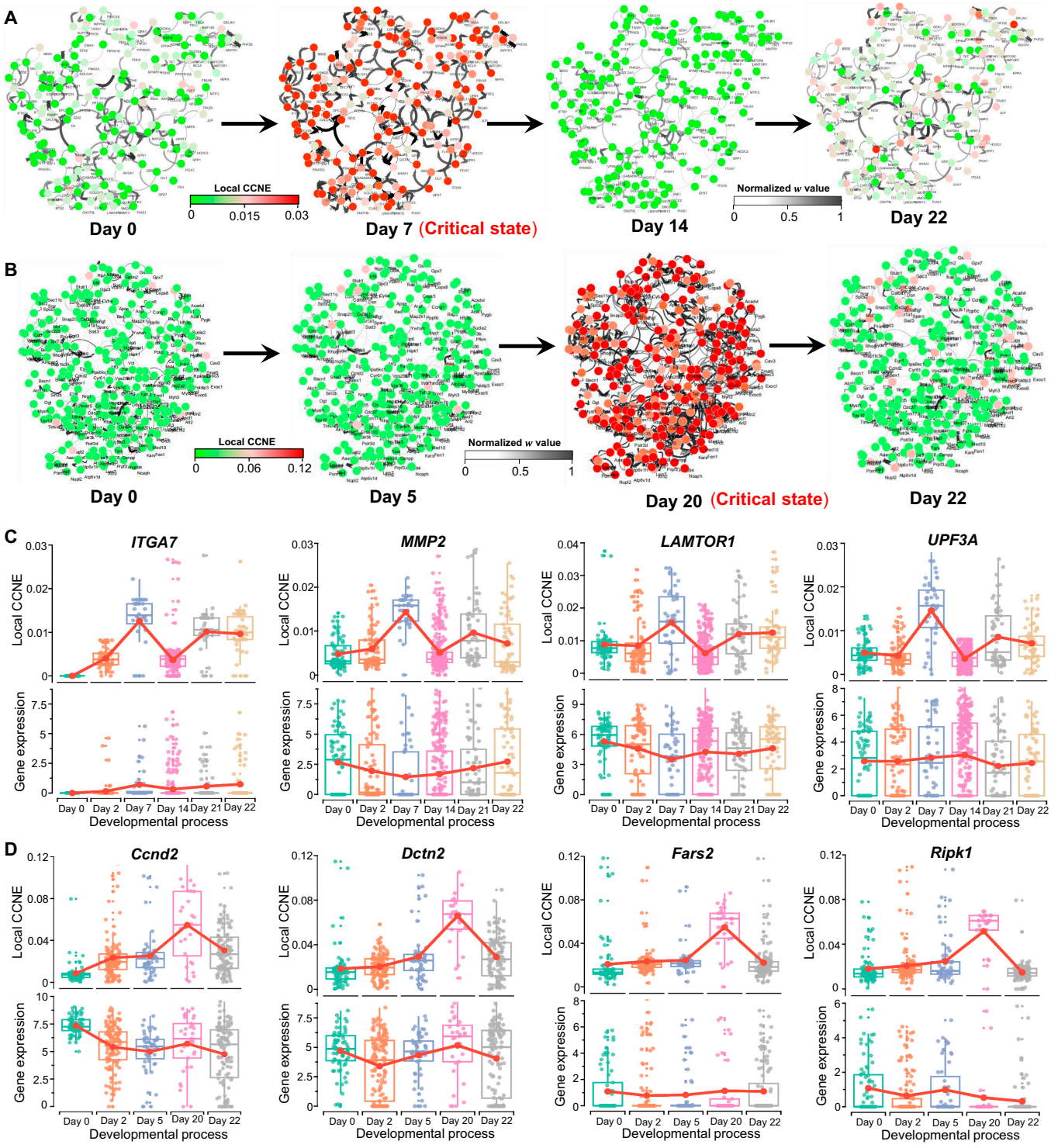


Fig. 4. Dynamic evolution of regulatory networks and identification of "dark genes" for signaling genes. Dynamic changes of signaling gene networks are shown for (A) pericyte-to-neuron data and (B) MEF-to-neuron data. The local CCNE score (top) and gene expression value (bottom) of "dark genes" are depicted for (C) pericyte-to-neuron data and (D) MEF-to-neuron data.

during iPSC lineage progression. As illustrated in Fig. 5F, *IL6ST* (a dark gene), as an upstream growth factor in the JAK-STAT pathway, activates JAK signaling after crossing the critical state, indicating that signaling begins during this period. Concurrently, the expression of downstream *JAK1* (first-order DEG neighbor) was up-regulated, enhancing the cell's response to factors like *IL-6* and promoting downstream signal transduction. Following *IL6ST* activation of various regulatory and downstream receptors, both *STAT3* activation and *PIM1* expression were inhibited, with *PIM1* change potentially facilitating the transition of iPSC cells from a proliferative stem state to mature cell differentiation [42]. Overall, the interaction between dark genes and their first-order DEG neighbors may uncover crucial signaling mechanisms in the JAK-STAT pathway. Besides, our Cellchat analysis

Table 2. The information of important "dark genes"

Gene	Location	Relation with embryonic development	PMID
ITGA7	Plasma membrane	The down-regulation of *ITGA7* could potentially hinder the process of cell differentiation during neural tube development.	33491544
MMP2	Nucleus	*MMP2* treatment of iPSCs enhanced their proliferation and differentiation into neural stem cells.	29137974
LAMTOR1	Lysosome	*LAMTOR1* regulates embryonic stem cell differentiation by GTPases Rag activation, promoting the exit from pluripotency.	28935770
UPF3A	Cytosol	*UPF3A* inhibits nonsense-mediated decay to promote the expression of genes involved in neuronal differentiation.	27837180
Ccnd2	Cytosol	*Ccnd2* can control cell fate decisions in human pluripotent stem cells by recruiting transcriptional corepressors.	26883361
Dctn2	Cytosol	*Dctn2 is* essential for normal neuronal function and may play a role in neuronal differentiation and neuroprotection.	31115483
Fars2	Mitochondrion	*Fars2* deficiency affects neuronal development and enhances neuronal apoptosis by impairing mitochondrial function.	35794642
Ripk1	Cytosol	*Ripk1* is a key effector in necroptosis and can regulate both apoptosis and necroptosis during embryonic development.	30867408

revealed that *EFNA1* and *EPHA1*, as receptor–ligand pairs specifically expressed between DE and MH, might play a key role in iPSC differentiation (Fig. 5G) [43].

Discussion

Identifying the pre-transition or critical state, along with key molecules, is essential for understanding various biological phenomena. For example, the identification of the pre-transition state or fate decision point in the cell differentiation process plays a crucial role in developing patient-specific tissue regeneration therapies and creating models for diseases related to differentiation. However, conventional methods remain limited in effectiveness and robustness when processing high-dimensional data with substantial noise, as is common in single-cell expression data. Most existing computational methods for single-cell data analysis rely on statistical metrics such as mean and variance derived from gene expression levels. In contrast to gene expression in single-cell data, gene association information remains stable across time and conditions [10], providing reliable characterization of biological processes such as critical transitions. In this study, we introduce a novel computational method at the single-cell level, termed CCNE, which aims to construct a cell-specific causal network for each cell and accurately pinpoint critical states for complex biological systems. By applying the CCNE method to both the simulated dataset and 5 single-cell datasets of different biological processes, we successfully identify their corresponding pre-transition states, validating the capability of the proposed method to detect critical signals at the single-cell level.

Our CCNE method presents the following novelty and benefits. Firstly, from the perspective of analyzing dynamic changes, it more effectively captures critical signals compared to existing critical state detection methods. Secondly, it enables the effective grouping of cells across different time points or cell types by distinguishing cellular heterogeneity based on the CCNE matrix. Thirdly, it facilitates the discovery of CCNE-sensitive "dark genes", which are typically overlooked by traditional differential expression analysis but potentially crucial in fundamental biological processes. Moreover, our CCNE strategy does not rely on models and operates without the necessity for feature selection or training of parameters. However, despite these advantages, the CCNE method has certain limitations. Specifically, it relies on the availability of a PPI network as the background network. Additionally, interpreting the biological significance of each identified critical point in multi-stage biological processes remains challenging and will be the focus of ongoing research.

Materials and Methods

Theoretical background

From the viewpoint of dynamical systems, the complex biological process is often considered a time-evolving system, wherein the pre-transition state signifies a sudden qualitative alteration observed at a bifurcation point [44]. Such a dynamical process can be divided into 3 main phases or states (Fig. 1A): a before-transition state with stability and resilience, a highly sensitive pre-transitional state where cell state transition occurs, and another stable after-transition state. Recently, the theoretical framework of DNB [45] has been applied for the quantitative detection of critical point or state transition point in complex biological processes. In particular, a cluster of DNBs with strong correlations and notable fluctuations emerges as the dynamic system approaches its critical state. Actually, the characteristics of DNB imply that dramatic variations in molecular expression and shifts in their causal regulatory strength can effectively signal the critical transition of the system [46]. Thus, by leveraging the dynamic variations in the causal relations among DNBs and their fluctuations in expression level, it is possible to predict the critical signal of a qualitative or drastic shift.

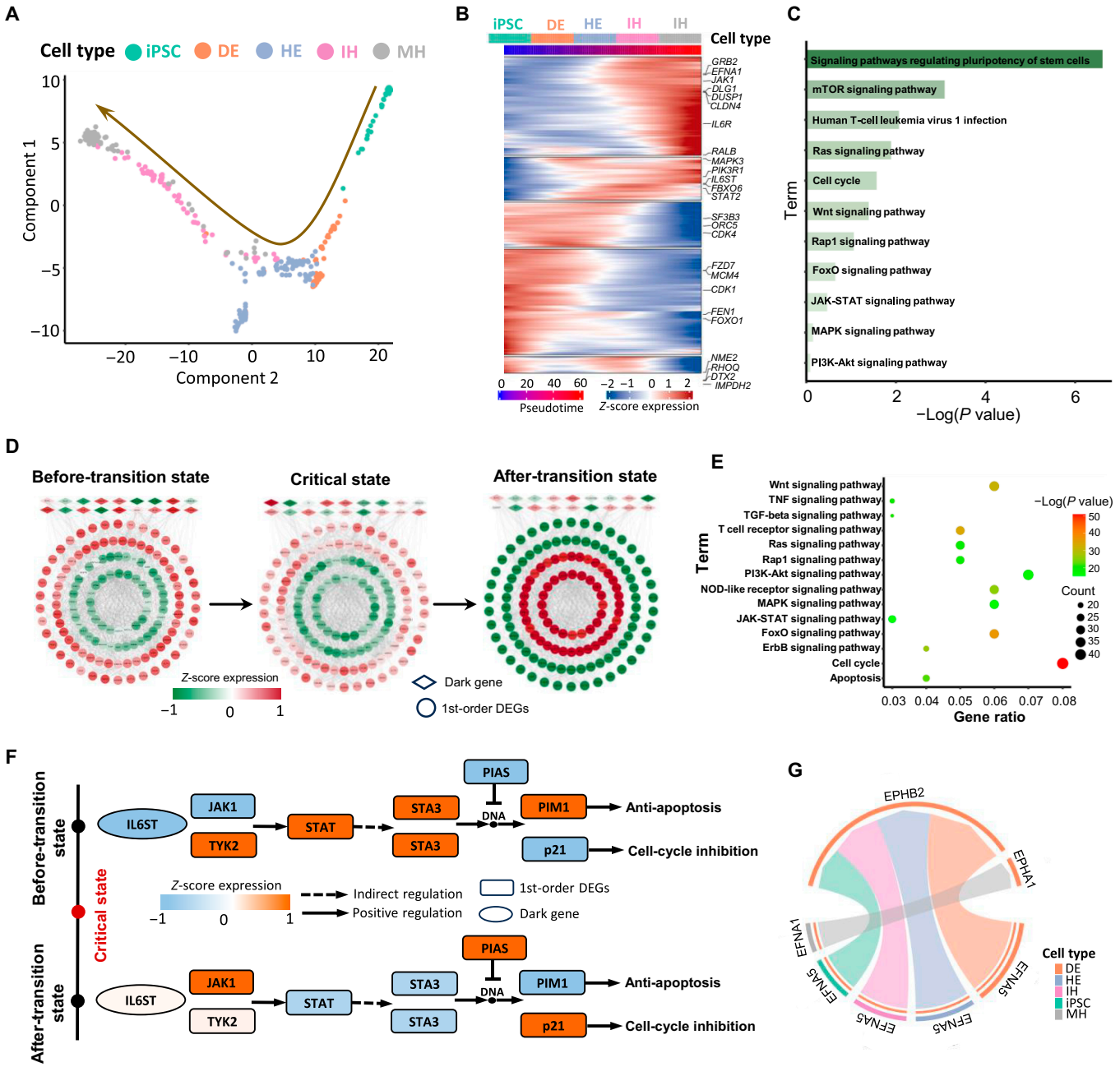


Fig. 5. The potential regulatory mechanisms during iPSC cell development and differentiation. (A) Inferred potential differentiation trajectory from iPSC cells to DE, HE, IH, and MH cells. (B) The heatmap for the temporal sequence of gene expression events during iPSC cell development and differentiation. (C) The primary enriched KEGG pathways for dark genes. (D) Dynamic evolution of the PPI subnetwork of dark genes and their first-order DEG neighbors for the iPSC differentiation process. (E) The KEGG pathway enrichment analysis for the first-order DEG neighbors. (F) The potential regulatory mechanism is uncovered by the functional analysis of dark genes and their first-order DEG neighbors. (G) The chord plot shows the interactions among cell subpopulations along the iPSC differentiation trajectory via the EPHA signaling pathway.

In this study, to determine continuous causality or functional dependency from a causal gene to an effect gene, we define continuous causality through functional continuity between variables based on a scaling law [20,21], which can be employed to infer direct gene regulatory network from single-cell data. Specifically, consider 2 genes g_x and g_y with M data points or cells, represented as $E(g_x) = \{E_1(g_x), E_2(g_x), \cdots, E_M(g_x)\}$ and $E(g_y) = \{E_1(g_y), E_2(g_y), \cdots, E_M(g_y)\}$. To quantify the causality or continuity of $g_y = f(g_x)$ from g_x to g_y in cell l, a new causal criterion or cross-map mutual information is designed and formulated as follows:

$$w^{(l)}(g_x, g_y) g_y = pr^{(l)}(f(g_x), g_y) \log \frac{pr^{(l)}(f(g_x), g_y)}{pr^{(l)}(g_x) \bullet pr^{(l)}(g_y)}, \quad (1)$$

$$pr^{(l)}\left(f\left(g_x\right),g_y\right)=\frac{n\left(f\left(E_{(l)}^{(\mathrm{NN})}\left(g_x\right)\right),E_{(l)}^{(\mathrm{NN})}\left(g_y\right)\right)}{M},$$
$$pr^{(l)}\left(g_x\right)=\frac{n\left(E_{(l)}^{(\mathrm{NN})}\left(g_x\right)\right)}{M},pr^{(l)}\left(g_y\right)=\frac{n\left(E_{(l)}^{(\mathrm{NN})}\left(g_y\right)\right)}{M}. \quad (2)$$

Here, $n\left(E_{(l)}^{(\mathrm{NN})}\left(g_x\right)\right)$ and $n\left(E_{(l)}^{(\mathrm{NN})}\left(g_y\right)\right)$ are the counts of the nearest neighbors/cells of $E_l\left(g_x\right)$ and $E_l\left(g_y\right)$ in cell l, respectively, among all M cells. The term $f\left(E_{(l)}^{(\mathrm{NN})}\left(g_x\right)\right)$ represents the cross-mapped values of $E_{(l)}^{(\mathrm{NN})}\left(g_x\right)$ to $E\left(g_y\right)$, which are the corresponding points of $E_{(l)}^{(\mathrm{NN})}\left(g_y\right)$ in $E\left(g_y\right)$. The number of points in the intersection of $n(f\left(E_{(l)}^{(\mathrm{NN})}\left(g_x\right)\right))$ and $n\left(E_{(l)}^{(\mathrm{NN})}\left(g_y\right)\right)$ is denoted as $n\left(f\left(E_{(l)}^{(\mathrm{NN})}\left(g_x\right)\right),E_{(l)}^{(\mathrm{NN})}\left(g_y\right)\right)$. A positive statistical dependency value signifies the presence of g_x's influence on g_y or functional dependency from g_x to effect g_y, implying the existence of a directed edge between them. Therefore, Eq. 1 can be utilized to infer a causal relationship from g_x to g_y and construct the cell-specific causal networks at the single-cell level.

CCNE tailored to identify the pre-transition state of complex biological process

Based on time-series single-cell data, the quantitative CCNE method is utilized to uncover cell state transitions or the pre-transition state during complex biological process, with the following specific implementation steps outlined.

Step 1: Constructing a cell-specific causal network for each cell. Based on the PPI network, the cell-specific causal network $N^{(l)}$ for cell C_l can be constructed using the causal criterion $w^{(l)}$ as defined in Eq. 1. Specifically, for a local network centered on a gene g^k within the PPI network, whose first-order neighbors are g_j^k $(j=1,2,\cdots,U)$, if $w^{(l)}\left(g^k,g_j^k\right)$ is positive, it indicates a directed edge $\left(g^k,g_j^k\right)$ from gene g^k to g_j^k; otherwise, no such directed edge exists (Fig. 1B). By evaluating each neighboring gene g_j^k in the local network of g^k based on the causal criterion $w^{(l)}\left(g^k,g_j^k\right)$, we can derive the causal subnetwork centered on g^k. Therefore, this process enables us to infer the cell-specific causal network $N^{(l)}$ for cell C_l.

Step 2: Isolating each localized causal network from the cell-specific causal network. Specifically, for a given cell C_l, its cell-specific causal network $N^{(l)}$ consists of V localized causal network $LN_k^{(l)}$ $(k=1,2,3,\ldots,V)$, each of which is uniquely identified by a specific gene and connected to its immediate outgoing neighbors (i.e., one gene for one localized network). Thus, each localized causal $LN_k^{(l)}$ is represented as a directed subnetwork centered on the specific gene g^k, where the central node g^k is connected to Q first-order outgoing neighbors or targets $\left\{g_1^k,g_2^k,\ldots,g_Q^k\right\}$ via Q out-degree edges.

Step 3: Calculating the local CCNE for each localized causal network. In particular, for the localized causal network $LN_k^{(l)}$ centered on the gene g^k with Q first-order outgoing neighbors, its local CCNE $LH_k^{(l)}$ is conducted as follows:

$$LH_k^{(l)}=-\left[\frac{1}{Q}\sum_{j=1}^{Q}P^{(l)}\left(g^k,g_j^k\right)logP^{(l)}\left(g^k,g_j^k\right)\right]\mathrm{SD}^{(l)}\left(g^k\right), \quad (3)$$

with

$$P^{(l)}\left(g^k,g_j^k\right)=\frac{w^{(l)}\left(g^k,g_j^k\right)}{\sum_{i=1}^{Q}w^{(l)}\left(g^k,g_i^k\right)},$$
$$\mathrm{SD}^{(l)}\left(g^k\right)=|\,\mathrm{SD}\left(\boldsymbol{E}_{\boldsymbol{g^k}}\right)-\mathrm{SD}\left(\widehat{\boldsymbol{E}}_{\boldsymbol{g^k}}\right)|, \quad (4)$$

where $w^{(l)}\left(g^k,g_j^k\right)$, as defined in Eq. 1, signifies the weight between the central gene g_k and its jth first-order outgoing neighbor g_j^k. The vectors $\boldsymbol{E}_{\boldsymbol{g^k}}$ and $\widehat{\boldsymbol{E}}_{\boldsymbol{g^k}}$ are represented as follows: $\boldsymbol{E}_{\boldsymbol{g^k}}=\left(E_{g^k}^1,E_{g^k}^2,\cdots,E_{g^k}^{l-1},E_{g^k}^l,E_{g^k}^{l+1},\cdots,E_{g^k}^M\right)$ and $\widehat{\boldsymbol{E}}_{\boldsymbol{g^k}}=\left(E_{g^k}^1,E_{g^k}^2,\cdots,E_{g^k}^{l-1},E_{g^k}^{l+1},\cdots,E_{g^k}^M\right)$, where the symbol $E_{g^k}^l$ denotes the expression of the gene g^k in cell C_l. $\mathrm{SD}^{(l)}$ can be viewed as evaluating the expression variation of the gene g^k in cell C_l against other cells.

Step 4: Calculating the cell-specific CCNE for each cell. Specifically, for cell C_l, the cell-specific CCNE $H^{(l)}$ is calculated by a cluster of genes with the highest local CCNE, according to the following formula:

$$H^{(l)}=\sum_{k=1}^{S}LH_k^{(l)}, \quad (5)$$

where the constant S corresponds to the count of genes ranked in the top 5% for the highest local CCNE. Our analysis indicates that parameter S within a certain range does not affect the overall trend of the signal curve (Fig. S7). Furthermore, the average cell-specific CCNE H_t at specific time point t, as described in Eq. 6, serves to capture the warning signal of cell state transition.

$$H_t=\frac{1}{M}\sum_{l=1}^{M}H^{(l)}. \quad (6)$$

Close to the pre-transition state or critical state, signaling molecules known as DNBs show the collective behaviors of fluctuations, characterized by a notable increase in molecule expression variation and a rapid strengthening effect of their causal strength. Analyzing the dynamic change of this group of DNB variables at a network level enables the prediction of qualitative state transitions. Consequently, the CCNE H_t exhibits a marked increase as the system nears the pre-transition state.

Step 5: Identifying the pre-transition state through the one-sample t test. To evaluate the effectiveness of our proposed CCNE H_t in detecting critical signals, we employ a one-sample t test [47] to ascertain the statistical significance of differences between the pre-transition state and the before-transition state. The formula for the one-sample t test statistic SC can be found below as Eq. 7 and is utilized to determine the significance of the deviation of the constant x from the average of the m-dimensional vector $\boldsymbol{X}=\left(x_1,x_2,\cdots,x_m\right)$.

$$SC = \sqrt{m}\frac{\text{mean}(\boldsymbol{X}) - x}{\text{SD}(\boldsymbol{X})} \quad (7)$$

where the terms mean($\boldsymbol{X}$) and SD($\boldsymbol{X}$) refer to the average and standard deviation of vector $\boldsymbol{X}$, respectively. The SC statistic yields a P value that evaluates the significance of statistics between mean($\boldsymbol{X}$) and constant x. If the P value is less than 0.05 (P value < 0.05), it denotes a statistically significant difference; otherwise, it suggests an absence of significant difference. Therefore, a pre-transition state is reached when the CCNE H_t meets both of the following criteria: (a) $H_t$is larger than H_{t-1}, and (b) H_t varies from prior values in a statistically significant manner (P value < 0.05), as detailed in Section H of the Supplementary Materials.

Sources of data and functional analysis

To highlight the capabilities of the CCNE method, it has been utilized in both a numerical simulation and the analysis of 5 real-world single-cell datasets: pericyte-to-neuron transition (ID: GSE113036), MEF-to-neuron transition (ID: GSE67310), MHC-to-HCC transition (ID: GSE90047), non-time-series EPCD transition (ID: GSE161277), and iPSC-to-MH transition (ID: GSE81252). Comprehensive descriptions of these datasets are available in Section I of the Supplementary Materials. An analysis of functional enrichment was conducted using the Metascape online tool [48] and the ClusterProfiler package [49]. Pathway-related information is accessible through KEGG.

Acknowledgments

Funding: This work was funded by the National Natural Science Foundation of China (nos. T2341022, 42450084, 12401630 and 12322119), the Educational Commission of Guangdong Province (2023KQNCX073), the Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems (2024B1212010004), the Natural Science Foundation of Guangdong Province (2023A1515110558 and 2024A1515011797), and the National Training Program of Innovation and Entrepreneurship for Undergraduates (202410561139).
Author contributions: R.L., P.C., and F.L. conceived the research. J.Z., Z.H., and J.Q. conducted the real data analysis. All authors contributed to writing the paper and approved the final manuscript.
Competing interests: The authors declare that they have no competing interests.

Data Availability

The data for pericyte-to-neuron transition (ID: GSE113036), MEF-to-neuron transition (ID: GSE67310), MHC-to-HCC transition (ID: GSE90047), EPCD transition (ID: GSE161277), and iPSC-to-MH transition (ID: GSE81252) were obtained from the NCBI Gene Expression Omnibus (GEO) database at http://www.ncbi.nlm.nih.gov/geo. The source code of the algorithm and related data can be found at https://github.com/zhongjiayuan-fs/CCNE_projects.

Supplementary Materials

Supplementary Text
Figs. S1 to S7
Table S1

References

1. Loan A, Awaja N, Lui M, Syal C, Sun Y, Sarma SN, Chona R, Johnston WB, Cordova A, Saraf D, et al. Single-cell profiling of brain pericyte heterogeneity following ischemic stroke unveils distinct pericyte subtype-targeted neural reprogramming potential and its underlying mechanisms. *Theranostics*. 2024;14(16):6110–6137.
2. Ayabe H, Anada T, Kamoya T, Sato T, Kimura M, Yoshizawa E, Kikuchi S, Ueno Y, Sekine K, Camp JG, et al. Optimal hypoxia regulates human iPSC-derived liver bud differentiation through intercellular TGFB signaling. *Stem Cell Rep*. 2018;11(2):306–316.
3. Jiang Z, Lu L, Liu Y, Zhang S, Li S, Wang G, Wang P, Chen L. SMAD7 and SERPINE1 as novel dynamic network biomarkers detect and regulate the tipping point of TGF-beta induced EMT. *Sci Bull*. 2020;65(10):842–853.
4. Zhong J, Han C, Wang Y, Chen P, Liu R. Identifying the critical state of complex biological systems by the directed-network rank score method. *Bioinformatics*. 2022;38(24):5398–5405.
5. Merrell AJ, Stanger BZ. Adult cell plasticity in vivo: De-differentiation and transdifferentiation are back in style. *Nat Rev Mol Cell Biol*. 2016;17(7):413–425.
6. Li Y, Lu Y, Kang C, Li P, Chen L. Revealing tissue heterogeneity and spatial dark genes from spatially resolved transcriptomics by multiview graph networks. *Research*. 2023;6:0228.
7. Tran TN, Bader GD. Tempora: Cell trajectory inference using time-series single-cell RNA sequencing data. *PLOS Comput Biol*. 2020;16(9):Article e1008205.
8. Nguyen QH, Lukowski SW, Chiu HS, Senabouth A, Bruxner TJC, Christ AN, Palpant NJ, Powell JE. Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. *Genome Res*. 2018;28(7):1053–1066.
9. Jiao L, Wang Y, Liu X, Li L, Liu F, Ma W, Guo Y, Chen P, Yang S, Hou B. Causal inference meets deep learning: A comprehensive survey. *Research*. 2024;7:0467.
10. Dai H, Li L, Zeng T, Chen L. Cell-specific network constructed by single-cell RNA sequencing data. *Nucleic Acids Res*. 2019;47(11):Article e62.
11. Gao R, Yan J, Li P, Chen L. Detecting the critical states during disease development based on temporal network flow entropy. *Brief Bioinform*. 2022;23(5):Article bbac164.
12. Liu X, Chang X, Liu R, Yu X, Chen L, Aihara K. Quantifying critical states of complex diseases using single-sample dynamic network biomarkers. *PLOS Comput Biol*. 2017;13(7):Article e1005633.
13. Liu R, Yu X, Liu X, Xu D, Aihara K, Chen L. Identifying critical transitions of complex diseases based on a single sample. *Bioinformatics*. 2014;30(11):1579–1586.
14. Li L, Xu Y, Yan L, Li X, Li F, Liu Z, Zhang C, Lou Y, Gao D, Cheng X, et al. Dynamic network biomarker factors orchestrate cell-fate determination at tipping points during hESC differentiation. *Innovations*. 2022;4(1):Article 100364.
15. Lesterhuis WJ, Bosco A, Millward MJ, Small M, Nowak AK, Lake RA. Dynamic versus static biomarkers in cancer immune checkpoint blockade: Unravelling complexity. *Nat Rev Drug Discov*. 2017;16(4):264–272.
16. Dai H, Jin QQ, Li L, Chen LN. Reconstructing gene regulatory networks in single-cell transcriptomic data analysis. *Zool Res*. 2020;41(6):599–604.

17. Wen L, Li G, Huang T, Geng W, Pei H, Yang J, Zhu M, Zhang P, Hou R, Tian G, et al. Single-cell technologies: From research to application. *Innovation*. 2022;3(6):Article 100342.
18. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S. Detecting causality in complex ecosystems. *Science*. 2012;338(6106):496–500.
19. Oravecz Z, Vandekerckhove J. Quantifying evidence for-and against-granger causality with bayes factors. *Multivariate Behav Res*. 2024;59(6):1148–1158.
20. Pearl J. Statistics and causality: Separated to reunite-commentary on Bryan Dowd's "separated at birth". *Health Serv Res*. 2011;46(2):421–429.
21. Ying X, Leng SY, Ma HF, Nie Q, Lai YC, Lin W. Continuity scaling: A rigorous framework for detecting and quantifying causality accurately. *Research*. 2022;2022:9870149.
22. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*. 2000;403(6767):339–342.
23. O'Brien EL, Itallie EV, Bennett MR. Modeling synthetic gene oscillators. *Math Biosci*. 2012;236(1):1–15.
24. Karow M, Camp JG, Falk S, Gerber T, Pataskar A, Gac-Santel M, Kageyama J, Brazovskaja A, Garding A, Fan W, et al. Direct pericyte-to-neuron reprogramming via unfolding of a neural stem cell-like program. *Nat Neurosci*. 2018;21(7):932–940.
25. Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SA, Sim S, Neff NF, Skotheim JM, Wernig M, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature*. 2016;534(7607):391–395.
26. Yang L, Wang WH, Qiu WL, Guo Z, Bi E, Xu CR. A single-cell transcriptomic analysis reveals precise pathways and regulatory mechanisms underlying hepatoblast differentiation. *Hepatology*. 2017;66(5):1387–1401.
27. Huang X, Han C, Zhong J, Hu J, Jin Y, Zhang Q, Luo W, Liu R, Ling F. Low expression of the dynamic network markers FOS/JUN in pre-deteriorated epithelial cells is associated with the progression of colorectal adenoma to carcinoma. *J Transl Med*. 2023;21(1):45.
28. Camp JG, Sekine K, Gerber T, Loeffler-Wirth H, Binder H, Gac M, Kanton S, Kageyama J, Damm G, Seehofer D, et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature*. 2017;546(7659):533–538.
29. Zhong J, Han C, Chen P, Liu R. SGAE: Single-cell gene association entropy for revealing critical states of cell transitions during embryonic development. *Brief Bioinform*. 2023;24(6):Article bbad366.
30. Liu R, Chen P, Chen L. Single-sample landscape entropy reveals the imminent phase transition during disease progression. *Bioinformatics*. 2020;36(5):1522–1532.
31. Yang XH, Goldstein A, Sun Y, Wang Z, Wei M, Moskowitz IP, Cunningham JM. Detecting critical transition signals from single-cell transcriptomes to infer lineage-determining transcription factors. *Nucleic Acids Res*. 2022;50(16):Article e91.
32. Liu X, Chang X, Leng S, Tang H, Aihara K, Chen L. Detection for disease tipping points by landscape dynamic network biomarkers. *Natl Sci Rev*. 2019;6(4):775–785.
33. Zhou B, Jin W. Visualization of single cell RNA-Seq data using t-SNE in R. *Methods Mol Biol*. 2020;2117:159–167.
34. Jin Q, Zuo C, Cui H, Li L, Yang Y, Dai H, Chen L. Single-cell entropy network detects the activity of immune cells based on ribosomal protein genes. *Comput Struct Biotechnol J*. 2022;20:3556–3566.
35. Zhong J, Han C, Zhang X, Chen P, Liu R. scGET: Predicting cell fate transition during early embryonic development by single-cell graph entropy. *Genom Proteom Bioinf*. 2021;19(3):461–474.
36. Sengupta S, Nie J, Wagner RJ, Yang C, Stewart R, Thomson JA. MicroRNA 92b controls the G1/S checkpoint gene p57 in human embryonic stem cells. *Stem Cells*. 2009;27(7):1524–1528.
37. Omeljaniuk WJ, Laudański P, Miltyk W. The role of miRNA molecules in the miscarriage process. *Biol Reprod*. 2023;109(1):29–44.
38. Xi B, An X, Yue Y, Shen H, Han G, Yang Y, Zhao S. Identification and profiling of microRNAs during sheep's testicular development. *Front Vet Sci*. 2025;12:1538990.
39. Liu L, Michowski W, Kolodziejczyk A, Sicinski P. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nat Cell Biol*. 2019;21(9):1060–1067.
40. Li L, Sun L, Gao F, Jiang J, Yang Y, Li C, Gu J, Wei Z, Yang A, Lu R, et al. Stk40 links the pluripotency factor Oct4 to the Erk/MAPK pathway and controls extraembryonic endoderm differentiation. *Proc Natl Acad Sci USA*. 2010;107(4):1402–1407.
41. Wu R, Liu Y, Zhao Y, Bi Z, Yao Y, Liu Q, Wang F, Wang Y, Wang X. m6A methylation controls pluripotency of porcine induced pluripotent stem cells by targeting SOCS3/JAK2/STAT3 pathway in a YTHDF1/YTHDF2-orchestrated manner. *Cell Death Dis*. 2019;10(3):171.
42. Cao L, Wang F, Li S, Wang X, Huang D, Jiang R. PIM1 kinase promotes cell proliferation, metastasis and tumor growth of lung adenocarcinoma by potentiating the c-MET signaling pathway. *Cancer Lett*. 2019;444:116–126.
43. Schmucker D, Zipursky SL. Signaling downstream of Eph receptors and ephrin ligands. *Cell*. 2001;105:701–704.
44. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B. Catastrophic shifts in ecosystems. *Nature*. 2001;413(6856):591–596.
45. Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep*. 2012;2:342.
46. Wen Z, Liu ZP, Liu Z, Zhang Y, Chen L. An integrated approach to identify causal network modules of complex diseases with application to colorectal cancer. *J Am Med Inform Assn*. 2013;20(4):659–667.
47. Rochon J, Kieser M. A closer look at the effect of preliminary goodness-of-fit testing for normality for the one-sample t-test. *Br J Math Stat Psychol*. 2011;64(3):410–426.
48. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523.
49. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287.