npj | systems biology and applications

Article

Published in partnership with the Systems Biology Institute

https://doi.org/10.1038/s41540-025-00504-2

Gene regulatory network inference during cell fate decisions by perturbation strategies

Qing Hu[1], Xiaoqi Lu[1], Zhuozhen Xue[1] & Ruiqi Wang[1,2] ✉

With rapid advances in biological technology and computational approaches, inferring specific gene regulatory networks from data alone during cell fate decisions, including determining direct regulations and their intensities between biomolecules, remains one of the most significant challenges. In this study, we propose a general computational approach based on systematic perturbation, statistical, and differential analyses to infer network topologies and identify network differences during cell fate decisions. For each cell fate state, we first theoretically show how to calculate local response matrices based on perturbation data under systematic perturbation analysis, and we also derive the wild-type (WT) local response matrix for specific ordinary differential equations. To make the inferred network more accurate and eliminate the impact of perturbation degrees, the confidence interval (CI) of local response matrices under multiple perturbations is applied, and the redefined local response matrix is proposed in statistical analysis to determine network topologies across all cell fates. Then in differential analysis, we introduce the concept of relative local response matrix, which enables us to identify critical regulations governing each cell state and dominant cell states associated with specific regulations. The epithelial to mesenchymal transition (EMT) network is chosen as an illustrative example to verify the feasibility of the approach. Largely consistent with experimental observations, the differences of inferred networks at the three cell states can be quantitatively identified. The approach presented here can be also applied to infer other regulatory networks related to cell fate decisions.

Cell fate decision, a complex dynamical process involved in the whole life, is intricately regulated by various molecules[1]. The transitions between distinct cell fates, such as differentiation, reprogramming, or transdifferentiation, occur during developmental processes[2,3]. The dynamical variations of various cell fates can be quantitatively captured through gene regulatory networks[4–7]. However, despite these advancements, accurately identifying the regulations and their intensities involved in these networks using solely in vivo experimental data is still difficult[8–10]. Consequently, inferring gene regulatory networks based on data sources remains a fundamental yet unresolved problem in systems biology.

In the realm of biological systems involving cell fate decisions, several methods have been developed to characterize cell fates using single-cell RNA sequencing (scRNA-seq) data, e.g., TSCAN[11], Monocle 2[12], SPRING[13], and Slingshot[14]. In addition, the distance covariance entropy (DICE) approach has the ability to predict regulatory factors underlying various phenotypes[15]. However, these methods often encounter limitations in inferring topological structures of gene regulatory networks associated with identified cell fates, i.e., even with the ability to determine cell states, network topologies of the system remain unknown.

On the other hand, there are algorithms for gene regulatory network inference using scRNA-seq data based on mutual information, random forest, correlation, regression, ordinary differential equations (ODEs), or Boolean models. Examples of such algorithms include PIDC[16], GENIE3[17], PPCOR[18], SCODE[19], and SCNS[20]. However, each of these network inference algorithms possesses its specific limitations. For example, PIDC and PPCOR fail to capture the directionality of regulations, while GENIE3 is limited to identifying the existence of regulations without quantifying their strength or weight. For SCODE, regulatory networks can only be inferred when specific linear ODEs with known parameters are employed, and these parameters can be further refined through linear regression and error iteration. In

[1]Department of Mathematics, Shanghai University, Shanghai, China. [2]Newtouch Center for Mathematics of Shanghai University, Shanghai, China.
✉e-mail: rqwang@shu.edu.cn

contrast, for the SCNS algorithm, the network model based on Boolean local rules can only be reconstructed if the three parameters corresponding to each gene are predetermined. Unfortunately, none of these methods adequately account for the dynamical network differences during cell fate decisions.

Modular response analysis (MRA) is an effective approach for inferring network connections through the analysis of experimental data obtained from various perturbations. Within this framework, the local response matrix provides a quantitative representation of both the direction and intensity of the interconnected edges within the network[21]. This indicates that the local response matrix has the potential to provide significant insights into biological network topologies. However, MRA failed to account for network topological variations across different cell states and the network inferred using local response matrix dose not satisfy the sparsity requirement of regulatory networks. In addition, there is still a gap in understanding the effects of different perturbation degrees on local response matrices[22,23].

Here, we introduce a computational method aimed at inferring network topologies and identifying differences among networks during cell fate decisions. This method is achieved through statistical and differential analyses of expression data subjected to systematic perturbations. Our method is broadly applicable and theoretically suitable for a variety of expression data types, including single-cell gene expression data, experimentally obtained measurements, or stable cellular data derived from software simulations. In addition, there are no restrictions on the scale of the inferred networks. Initially, local response matrix at each cell state is numerically calculated from perturbation data. The elements within the matrix correspond to regulations between molecules, meaning that local response matrix reveals network topologies associated with cell fate decisions. Furthermore, based on statistical analysis, the confidence intervals (CIs) of local response matrix are employed to identify the sparsity of regulatory networks[24,25], as well as the influence of regulation degrees. Subsequently, we construct the redefined local response matrix to reflect the network topologies during all cell states, and the accuracy of the inferred network topologies can be verified by two types of errors. Furthermore, differential analysis is performed to determine the relative local response matrix, thereby quantifying critical regulations within each cell fate and identifying primary cell states associated with specific regulations.

To demonstrate the feasibility of our approach, we choose the epithelial to mesenchymal transition (EMT) network as an illustrative example[26]. This example demonstrates how perturbation and statistical analyses are applied to infer regulatory networks, how differential analysis is performed to identify critical regulations in three distinct cell states: epithelial (E), mesenchymal (M), and hybrid (H) state with both E and M features, and how specific cell states related to each regulation are determined. For the analysis of EMT network, we utilize mathematical model (ODEs) to simulate systematic perturbations and obtain the expression data of all molecules. However, it is worth noting that our approach is theoretically applicable to measured experimental data as well. Additionally, the calculation of local response matrices, which is based on data, is model-independent. We demonstrate the feasibility of our method by comparing it with priori knowledge[27–31], indicating its capability in inferring network topologies and identifying network differences.

Results

Fundamentals of the network inference

For a gene regulatory network comprised of n molecules, i.e., nodes, and multiple molecular regulations, i.e., links or edges, the dynamics of the network can be generally described by a set of ODEs as follows

$$\frac{d\boldsymbol{x}}{dt} = \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x};\boldsymbol{p}), \tag{1}$$

where $\boldsymbol{x} = (x_1, \cdots, x_n)^T$ is a time-dependent vector representing concentrations of the molecules, such as genes or proteins, $\boldsymbol{f} = (f_1, \cdots, f_n)^T$ is a continuously differentiable function vector related to the production and consumption of the molecules as well as the regulations between them, and $\boldsymbol{p} = (p_1,..., p_u, p_{u+1},..., p_m)$, $m > u \geq n$, is a parameter vector including rate constants, Michaelis constants, etc.

To infer network topologies and investigate network differences across different cell fates based on perturbation analysis, we divide $\boldsymbol{p}$ into two parts, the sensitive parameter set $(p_1,... p_u)$ and the constant parameter set $(p_{u\_+1},...,p_m)$. The sensitive parameter set can be perturbed and is associated with nodes, each of which contains at least one sensitive parameter, such as the basic generation rate, degradation rate, signals pertinent to the specific node, or activators and inhibitors. The perturbation applied to the sensitive parameter associated with node i meets the criterion that the expression of node i is initially and directly influenced by this perturbation, with subsequent indirect effects on the expressions of other nodes due to the regulations among them. Here, we restrict our analysis to each node contains only a single sensitive parameter, i.e., $u = n$, $\partial f_i/\partial p_i \neq 0$, and $\partial f_i/\partial p_k = 0$, $1 \leq i \neq k \leq n$.

We assume the system (1) reaches a locally stable steady state $\bar{\boldsymbol{x}} = (\bar{x}_1, \cdots, \bar{x}_n)$ at the basal unperturbed sensitive parameter set $\boldsymbol{p_b} = (p_{b,1},..., p_{b,n})$ and the constant parameter set $\boldsymbol{p_c} = (p_{c,n+1},..., p_{c,m})$, i.e.,

$$\left.\frac{d\boldsymbol{x}}{dt}\right|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})} = \boldsymbol{f}(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c}) = \boldsymbol{0},\ \forall\, t > t_0.$$

In other words, $\bar{\boldsymbol{x}}$ represents an unperturbed stable steady state at the unperturbed sensitive parameter set $\boldsymbol{p_b}$ and the constant parameter set $\boldsymbol{p_c}$. Denote the perturbed stable steady state by $\bar{\boldsymbol{x}}_k = (\bar{x}_{k,1}, \cdots, \bar{x}_{k,n})$ after sensitive parameter p_k is slightly perturbed from $p_{b,k}$ to $p_{s,k}$. Define $\boldsymbol{p_s} = (p_{s,1},..., p_{s,n})$ as the perturbed sensitive parameter set. Here, $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}_k$ must be the same type of cells, and $\bar{x}_i \neq 0$, $\bar{x}_{k,i} \neq 0$, $i, k = 1,..., n$.

The direct regulation from node j to node i can be quantified by the relative change of $\Delta x_i/\bar{x}_i$ with respect to $\Delta x_j/\bar{x}_j$ under the mild perturbation of sensitive parameters. It is defined as local response coefficient r_{ij}, i.e.,

$$\begin{aligned} r_{ij} &= \lim_{\Delta x_j \to 0}\left(\frac{\Delta x_i/\bar{x}_i}{\Delta x_j/\bar{x}_j}\right) = \left.\frac{\partial \ln x_i}{\partial \ln x_j}\right|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})} \\ &= \left.\frac{\bar{x}_j}{\bar{x}_i}\cdot\frac{\partial x_i}{\partial x_j}\right|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})}, i \neq j, \end{aligned}$$

where Δx_i represents the change of $\bar{x}_i$ under the perturbation to one sensitive parameter, and we define $r_{ii} = -1$. The sign of r_{ij} reflects the type of regulation from node j to node i, the absolute value of r_{ij} indicates the strength of the regulation, where $r_{ij} > 0$ (or $r_{ij} < 0$) represents activated (or inhibitory) regulation, and $r_{ij} = 0$ indicates no direct regulation from node j to node i, $i \neq j$. Meanwhile, the relative change of node i to the change of an infinitesimal perturbation for a sensitive parameter p_k defines the global response coefficient R_{ik}, i.e.,

$$R_{ik} = \left.\frac{\partial \ln x_i}{\partial \ln p_k}\right|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})} = \left.\frac{p_{b,k}}{\bar{x}_i}\cdot\frac{\partial x_i}{\partial p_k}\right|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})},\ 1 \leq i, k \leq n.$$

We use $\boldsymbol{r} = (r_{ij})$ and $\boldsymbol{R} = (R_{ik})$ to represent the local response matrix and the global response matrix[21], respectively. Therefore, the local response matrix $\boldsymbol{r}$ can be used to characterize the network topology of the system.

The relationship between local response matrix and global response matrix

Usually, a slight perturbation to p_k initially affects node k, and subsequently propagates to node i via both direct regulation from k to i and indirect path from k to i, which includes intermediate nodes. Therefore, it is impractical to determine the precise proportion of direct and indirect regulations that contribute to the ultimate change in node i under the perturbation p_k. In other words, the local response matrix $\boldsymbol{r}$ can not be measured directly based on the experimental data before and after perturbations. However, the global response matrix $\boldsymbol{R}$ can be calculated directly.

In experimental measurements, based on the work of literature[21], the global response coefficient R_{ik} can be quantitatively expressed as $\Delta \ln x_i(k)/\Delta \ln p_k$, where $\Delta \ln x_i(k) = \ln \bar{x}_{k,i} - \ln \bar{x}_i$ and $\Delta \ln p_k = \ln p_{s,k} - \ln p_{b,k}$. Using central difference approximations and ignoring the higher order nonlinear terms for the logarithmic function, we have

$$R_{ik} = \frac{\Delta \ln x_i(k)}{\Delta \ln p_k} = \frac{\bar{x}_{k,i} - \bar{x}_i}{\bar{x}_{k,i} + \bar{x}_i} \cdot \frac{p_{s,k} + p_{b,k}}{p_{s,k} - p_{b,k}}. \quad (2)$$

Based on the multivariate implicit function theorem and the multivariate chain rule, the local response matrix $\boldsymbol{r}$ is related to the global response matrix $\boldsymbol{R}$ (see Method or equation [5] in ref. 21), i.e.,

$$\boldsymbol{r} = -[\mathrm{diag}(\boldsymbol{R}^{-1})]^{-1}\boldsymbol{R}^{-1}. \quad (3)$$

Consequently, utilizing only the data obtained before and after perturbations to sensitive parameters, without the need for specific ODEs and regardless of the network size, the local response matrix $\boldsymbol{r}$ can be directly inferred from equations (2) and (3). We refer to the analysis involving perturbations to sensitive parameters as systematic perturbation analysis. Therefore, network topologies and regulation strength at each cell state can be inferred. It is important to note that the calculation of $\boldsymbol{r}$ is only valid when the data obtained after perturbations are in close proximity to the unperturbed stable steady states, i.e., there is no qualitative differences in these data (no transition from one cell state to another).

Additionally, when the ODEs and corresponding parameter values of system (1) are known, we can get

$$\boldsymbol{r} = -\boldsymbol{X}^{-1}\left[\mathrm{diag}(\boldsymbol{J})\right]^{-1}\boldsymbol{J}\boldsymbol{X}, \quad (4)$$

where the Jacobian matrix $\boldsymbol{J}$ at the stable steady state $\bar{\boldsymbol{x}}$ takes the form

$$\boldsymbol{J} = \left(\frac{\partial f_i}{\partial x_j}\right)\Bigg|_{(\bar{\boldsymbol{x}};\boldsymbol{p}_b;\boldsymbol{p}_c)}, \ 1 \le i,j \le n,$$

and the nonsingular diagonal matrix $\boldsymbol{X} = \mathrm{diag}(\bar{\boldsymbol{x}})$ (see Method). Therefore, the local response matrix $\boldsymbol{r}$ at $\bar{\boldsymbol{x}}$ can also be calculated when the Jacobian matrix $\boldsymbol{J}$ (or specific ODEs) of system (1) at $\bar{\boldsymbol{x}}$ is known. For simplicity and to distinguish $\boldsymbol{r}$ computed by Eqs. (2) and (3), we call the local response matrix obtained by Eq. (4) as the wild-type (WT) local response matrix, denoted as $\boldsymbol{r}^{\boldsymbol{0}}$.

However, obtaining ODEs and their corresponding parameter values which accurately represent a gene regulatory network is often challenging in practice, making the calculation of $\boldsymbol{r}^{\boldsymbol{0}}$ impractical. Fortunately, experimental measurements of the data before and after systematic perturbations are feasible, allowing for the calculation of $\boldsymbol{r}$ using Eqs. (2) and (3).

The inference of network topologies and identification of network differences

There are numerical inaccuracies in deriving $\boldsymbol{r}$ through the utilization of the cell data before and after sensitive parameter perturbations, mainly due to fluctuations in actual measurements caused by internal and external environmental factors and the difference in perturbation intensity of sensitive parameters. Additionally, the omission of higher-order nonlinear terms (central difference approximation) further contributes to the inaccuracies. Furthermore, the majority of elements in $\boldsymbol{r}$ calculated using Eqs. (2) and (3) are nonzero (see ref. 21), which is contradict with the sparsity of regulatory networks. Consequently, it is imperative to diminish the significance of most predicted regulations to minimize the disparity between the inferred and the true networks, thus reducing the inaccuracy.

Based on perturbation and statistical analyses, we construct the redefined local response matrix $\hat{\boldsymbol{r}} = (\hat{r}_{ij})$ to reflect the network topologies more accurately (see Methods). For a gene regulatory network with s different cell fates, there are s redefined local response matrices, i.e., one $\hat{\boldsymbol{r}}(v)$ for the v-th state, $v = 1,..., s$. And there exist quantitative variations in the network topologies for these s cell fates. Further, based on differential analysis, we define the relative local response matrix $\tilde{\boldsymbol{r}}(v) = (\tilde{r}_{ij}(v))$ to quantify the differences of the regulations across all cell fates (see Methods).

In summary, to infer network topologies and identify network differences across all cell fates, the specific steps are as follows.

1. Determine nodes and their respective sensitive parameters. Additionally, identify the number of cell fates present within the network. Assume the network has n nodes and s different cell fates, with each node having only one sensitive parameter. For each cell fate, measure unperturbed steady state data $\bar{\boldsymbol{x}}(v)$ under unperturbed sensitive parameter set $\boldsymbol{p_b}(v)$, $v = 1, ..., s$.
2. Randomly select values for one perturbed sensitive parameter set $\boldsymbol{p_s}(v)$ from distributions such as the uniform or normal distribution under systematic perturbation analysis. Calculate the global response matrix $\boldsymbol{R}(v)$ by Eq. (2) and using the measured stable steady state data before and after perturbations. Then, infer the local response matrix $\boldsymbol{r}(v)$ by Eq. (3).
3. Repeat step 2 multiple times to obtain the distribution of $\boldsymbol{r}_{ij}(v)$ and then verify whether the 95% confidence interval (CI) of the distribution contains zero for all $1 \le i \ne j \le n$ in statistical analysis. Further, infer the redefined local response matrix $\hat{\boldsymbol{r}}(v)$ which reveals network topologies across all s cell fates.
4. Calculate the relative local response matrix $\tilde{\boldsymbol{r}}(v)$ based on $\hat{\boldsymbol{r}}(v)$ in differential analysis, which serves as a quantitative tool to identify and compare network differences across all cell fates.

Application to the EMT network

To demonstrate feasibility of the approach, we apply it to the EMT network[26]. The approach presented here can be applied to diverse biological systems, enabling the inference of network topologies and the identification of crucial network differences across various cell fates.

EMT plays crucial roles in both physiological and pathological processes, such as tissue repair, normal embryonic development, wound healing, and cancer metastasis[32–34]. During EMT, epithelial cells undergo a morphological transformation characterized by the loss of cell-cell polarity and adhesion, ultimately transforming into mesenchymal cells with enhanced motility and invasive capabilities. A hybrid phenotype exists during this transition process[35–37]. Consequently, the EMT process involves three stable steady states: the epithelial (E) state, the hybrid (H) state, and the mesenchymal (M) state[38,39].

The core EMT network is composed of nine molecules, including one endogenous transforming growth factor-β (TGF-β), two transcription factors, SNAIL1 and ZEB, two mRNAs, snail1 and zeb, two microRNAs, miR-34 and miR-200, and two markers, E-cadherin and N-cadherin, as shown in Fig. 1. TGF-β promotes the transcription of snail1 mRNA, which is translated to SNAIL1[40]. SNAIL1 promotes the expression of zeb mRNA,

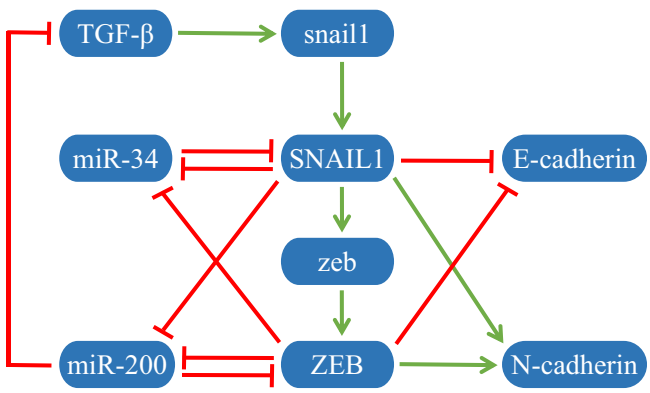

Fig. 1 | The core EMT regulatory network. Green arrows and red bars denote activation and inhibition, respectively.

which is translated to ZEB[41]. Both SNAIL1 and ZEB repress the transcription of miR-34 and miR-200. In turn, miR-34 and miR-200 inhibit the expression of SNAIL1 and ZEB, respectively[42,43]. In addition, miR-200 also suppresses the expression of TGF-β[44]. SNAIL1 and ZEB inhibit the expression of E-cadherin and promote the expression of N-cadherin.

Detailed ODEs of the dynamical model are given in Supplementary Note 1. Since E-cadherin and N-cadherin do not affect the expression of other molecules, namely, there is no regulation from them to other molecules, we exclude them in the following analysis. That is, only 7 nodes and 11 regulations are left. The details on all regulations are given in Table 1. By using the parameters in Supplementary Note 2 (Supplementary Table 1) to solve the ODEs under different initial conditions, the expression data of the three stable phenotypes, E, H, and M, can be calculated concretely. The diagonal matrix $\boldsymbol{X}(v)$ constituted by expressions of the seven molecules at each state and the corresponding Jacobian matrix $\boldsymbol{J}(v)$ can be obtained, where v = 1, 2, 3 represents E, H, and M states, respectively.

Table 1 | Regulations of EMT network

Source nodes	Target nodes	Types	Regulations
miR-200	TGF-β	-1	L_{17}
TGF-β	snail1	1	L_{21}
snail1	SNAIL1	1	L_{32}
miR-34	SNAIL1	-1	L_{34}
SNAIL1	miR-34	-1	L_{43}
ZEB	miR-34	-1	L_{46}
SNAIL1	zeb	1	L_{53}
zeb	ZEB	1	L_{65}
miR-200	ZEB	-1	L_{67}
SNAIL1	miR-200	-1	L_{73}
ZEB	miR-200	-1	L_{76}

L_{ij} represents the regulation from source node j to target node i, i, j = 1,..., 7 represent TGF-β, snail1, SNAIL1, miR-34, zeb, ZEB, and miR-200, respectively. And 1 and − 1 represent activation and inhibition, respectively.

Table 2 | Four perturbation types applied to the sensitive parameter kd_i, i = 1,..., 7

Type	Distribution	Perturbation
I	uniform	$kd_{s,i} \in U(0.8kd_{b,i}, 1.2kd_{b,i})$
II	uniform	$kd_{s,i} \in U(0.5kd_{b,i}, 1.5kd_{b,i})$
III	normal	$kd_{s,i} \in N(kd_{b,i}, 0.1kd_{b,i})$
IV	normal	$kd_{s,i} \in N(kd_{b,i}, 0.2kd_{b,i})$

$kd_{b,i}$, the basal value of kd_i, is the unperturbed parameter value, and $kd_{s,i}$ is the perturbed parameter value.

Inferring the EMT network

To infer network topology across all cell fates based on perturbation and statistical analyses, i.e., to derive the redefined local response matrix $\hat{\boldsymbol{r}}(v)$ using the stable steady state data before and after perturbations, instead of relying on specific ODEs and parameter values of the system, we initially determine the set of sensitive parameters (kd_1,..., kd_7), which comprised by the degradation rates of all nodes. This set satisfies that a perturbation to kd_i directly affects only the expression of node i, and subsequently, indirectly alters the expression of other nodes, i = 1,..., 7 represent TGF-β, snail1, SNAIL1, miR-34, zeb, ZEB, and miR-200, respectively. Once the perturbed sensitive parameter set is determined, the local response matrix can be calculated by using only the measured stable data before and after perturbations to sensitive parameters.

Since the local response matrix calculated from one perturbed sensitive parameter set may have randomness and contingency, we adopt a resampling approach to mitigate these effects. By repeatedly perturbing the sensitive parameter set 2000 times and calculating the corresponding local response coefficients for each set, i.e., N = 2000, we obtain these resampling local response matrices. For each perturbed sensitive parameter set, we assume it follows either a uniform or normal distribution. Recognizing that the perturbation levels applied to sensitive parameter set can lead to differences in measured data, which may further cause differences in calculated local response matrices, we select two distinct perturbation levels for each distribution: 20% or 50% for the uniform distribution (Type I or II), and 10% or 20% for the normal distribution (Type III or IV), the details of each perturbation are outlined in Table 2.

By applying statistical analysis to these resampling local response matrices, we can derive the distribution of each local response coefficient and its corresponding 95% CI. Among them, only certain elements of $\boldsymbol{r}(v)$ satisfy the criterion: 0 does not fall within the 95% CI. The diagonal elements of $\boldsymbol{r}(v)$, − 1, are excluded in the following analysis. For these four perturbations, the probability distribution of calculated local response coefficients satisfying the criterion are shown in Fig. 2. Obviously, the total count and the intensities of identified regulations differ in E, H, and M states. For each state, the disparity of the total count is mainly caused by elements approaching zero. The closer the calculated local response coefficient r_{ij} approaches zero, the more significant the differences become, e.g., r_{76} in E state and r_{73} in H state. For the distributions of r_{ij} satisfying $0 \notin$ the 9 5% CI across all the four perturbation types, there exist minor disparities near the left and right endpoints of these distributions at each state, and the distribution associated with each perturbation type exhibit distinct patterns, satisfying the principle that as the perturbation degree increases, the distribution tends to become more gradual.

At each state and for each perturbation type, the number of non-zero elements in the redefined local response matrix $\hat{\boldsymbol{r}}(v)$ equals the total number

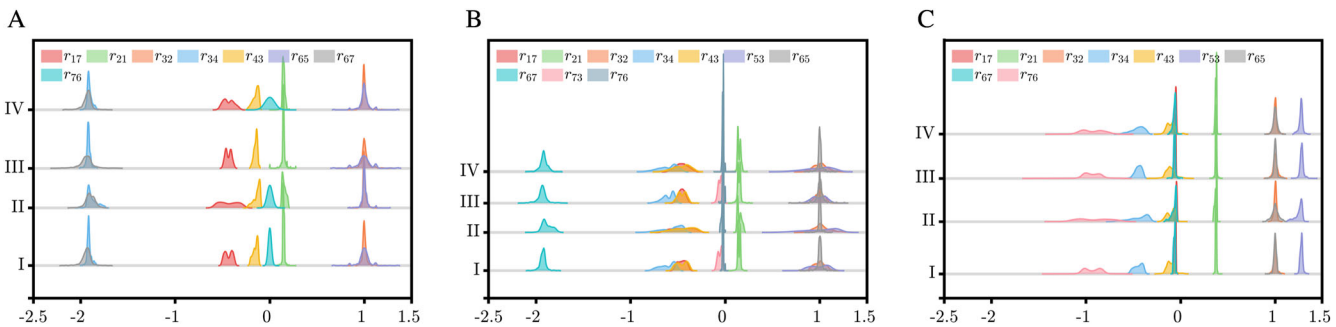

Fig. 2 | The probability distributions of the resampled local response coefficients under these four perturbations satisfying the condition that $0 \notin$ the 95% CI at E, H, and M states, corresponding to (A), (B), (C), respectively. The horizontal axis depicts the values of r_{ij}, while the vertical axis represents corresponding distribution frequency.

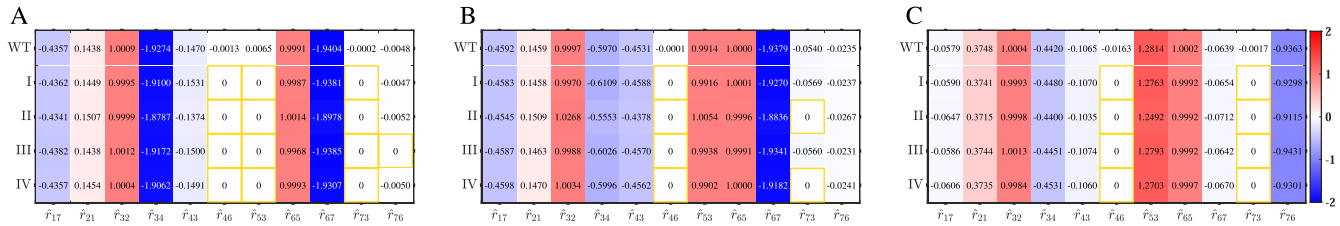

Fig. 3 | The non-zero elements of the WT and redefined local response matrices at E, H, and M states, corresponding to (A), (B), (C), respectively. Each element's value is depicted by a distinctively colored box, with the intensity of the color indicating the strength of the regulation. Specifically, a positive value denotes activation, whereas a negative value denotes inhibition. This contrast reflects the difference between true and inferred networks. The top row represents the non-zero elements of the WT local response matrices. The elements satisfying $\hat{r}_{ij}(\nu) = 0$ are marked by orange boxes.

of probability distributions identified in Fig. 2, i.e., most elements are zero and only some elements are non-zero. To assess the accuracy of $\hat{\boldsymbol{r}}(\nu)$, which is calculated using Eqs. (2), (3), (6), in reflecting the true network topologies, we calculate the wild-type (WT) local response matrix $\boldsymbol{r}^0(\nu)$ based on Eq. (4) for comparison. Only some elements of $\boldsymbol{r}^0(\nu)$, corresponding to true regulations listed in Table 1, are non-zero, and the value and sign of r^0_{ij} reflect the regulatory information on L_{ij}. To verify the effects of perturbation degrees and distributions on $\hat{\boldsymbol{r}}(\nu)$, as well as to characterize the difference between the redefined and WT local response matrices, the contrast between the non-zero elements in $\boldsymbol{r}^0(\nu)$ and corresponding elements in $\hat{\boldsymbol{r}}(\nu)$ under four perturbation types is shown in Fig. 3.

When $r^0_{ij}(\nu)$ significantly deviates from zero, corresponding $\hat{r}_{ij}(\nu) \neq 0$ and fluctuates around $r^0_{ij}(\nu)$ under four perturbation types, i.e., there is a slight difference between $\hat{r}_{ij}(\nu)$ and $r^0_{ij}(\nu)$ quantitatively. However, for $r^0_{ij}(\nu)$ close to zero, corresponding $\hat{r}_{ij}(\nu)$ may either be zero or extremely close to $r^0_{ij}(\nu)$ and be non-zero under four perturbation types, i.e., a few elements exhibit qualitative differences. Specifically, the actual regulation from ZEB to miR-34 exists in all cell phenotypes, i.e., $r^0_{46} \neq 0$, while we obtained $\hat{r}_{46} = 0$ at each perturbation analysis. The main reason for this difference we considered is that the value of r^0_{46} at each state is very close to zero, resulting in the 95% CI of r_{46} under resamplings including zero, the same for the regulation from SNAIL1 to zeb in E state (i.e., $r_{53}(\nu)$, $\nu = 1$), and almost the same for the regulation from SNAIL1 to miR-200 (i.e., r_{73}), with the only difference is that $\hat{r}_{73} \neq 0$ in H state and for the perturbation types I and type III (small degree of perturbation).

When the true element $r^0_{ij}(\nu)$ is far from zero, the perturbation degrees and distributions barely alter $\hat{r}_{ij}(\nu)$. Conversely, if $r^0_{ij}(\nu)$ is exceedingly small, the perturbation degrees may significantly affect the identification of regulations. Specifically, a smaller perturbation degree results in a more precise identification of regulations, while the distribution type of perturbation has little effect on the identification. Consequently, the network topologies inferred from perturbation data closely resemble the true network topologies obtained based on the dynamical model across all the three cell states, although the regulations inferred from four perturbation types exhibit subtle differences in each state.

To verify whether the redefined local response matrix $\hat{\boldsymbol{r}}$ can be used to reflect the WT local response matrix $\boldsymbol{r}^0$, i.e., how precisely the inferred network topologies based on stable steady-state data before and after perturbation capture the true network topologies, we compute two types of errors. One is the inference error between $\hat{\boldsymbol{r}}$ and $\boldsymbol{r}^0$, $\hat{e} = \sqrt{\sum_{i,j}(\hat{r}_{ij} - r^0_{ij})^2}$. The other is the sample error between $\boldsymbol{r}^l$ and $\boldsymbol{r}^0$, $e_l = \sqrt{\sum_{i,j}(r_{l,ij} - r^0_{ij})^2}$, $l = 1,\ldots, N$. The two types of errors for each state under the four distinct perturbation types are presented in Fig. 4. In each state, the colored dots (green, orange, blue, and yellow dots) represent the values of e_l for resamplings under four perturbation types, which form violin-shaped distributions to illustrate the sample errors across all perturbation types. The corresponding red dot marks the mean of each distribution, while the error bar indicates its standard deviation. Additionally, the red + sign positioned at each distribution represents the value of $\hat{e}(\nu)$.

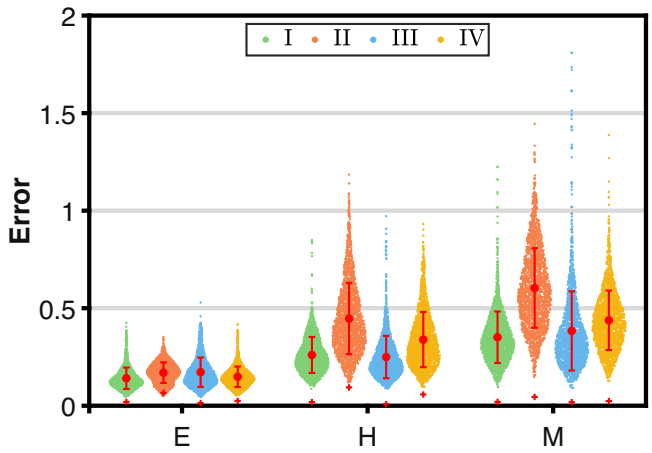

Fig. 4 | The inference and sample errors at E, H, and M states. The inference errors are represented as the red + signs. The sample errors under perturbation type I, II, III, and IV are represented as green, orange, blue, and yellow dots, respectively.

Obviously, there is some examples with significant sample errors within each distribution, although majority clusters around the mean of distribution, suggesting that the inferred regulations under a single perturbation set may deviate substantially from the true network. Moreover, as perturbation degree increases, the sample errors tend to grow larger. Specifically, the majority of sample errors under perturbation type II, depicted by orange dots, are generally higher than those under other perturbation types. However, irrespective of the degree of the perturbation and perturbation types, the inference error $\hat{e}(\nu)$ remains remarkably close to zero. Moreover, it is much smaller than the mean of $e_l(\nu)$ across all three states, even lower than the individual sample error under each perturbation type, showing good feasibility of using $\hat{\boldsymbol{r}}(\nu)$ as a representation of $\boldsymbol{r}^0(\nu)$ even when the perturbation degree is not particularly negligible.

Regardless of whether the distribution of perturbation is normal or uniform, as the degree of perturbation decreases, the inference error converges towards zero, indicating a more accurate inference. Therefore, we use the redefined local response matrix $\hat{\boldsymbol{r}}(\nu)$ under perturbation type I to infer network topology and identify network differences across three cell fates, the corresponding elements are shown in Fig. 5. By analyzing whether redefined local response coefficients remain non-zero across all states, we can determine the presence of specific regulations within the network topology. As a result, the comprehensive EMT network can be reconstructed, as depicted in Fig. 6.

Compared to the true EMT network presented in Fig. 1, it is evident that two networks exhibit a remarkable similarity except that our approach fails to identify the regulation from ZEB to miR-34, i.e., $\hat{r}_{46}(\nu) = 0$. The reason for the difference is that the actual regulation intensity $r^0_{46}(\nu)$ is almost zero. While the values of $r_{46}(\nu)$ calculated by Eqs. (2) and (3) under all resamplings satisfy the condition 0 within the 95% CI, resulting in

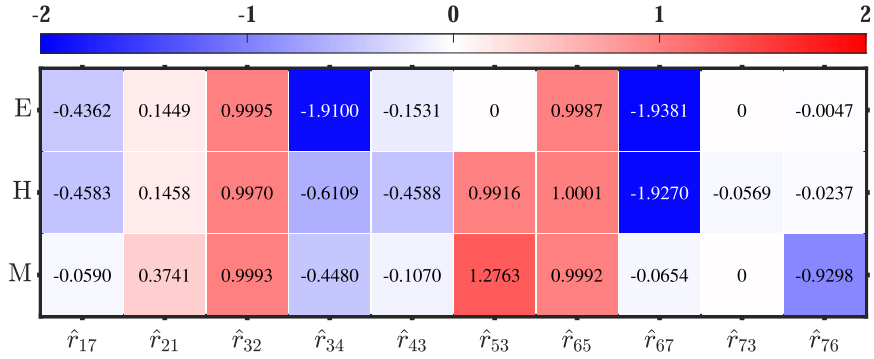

Fig. 5 | The particular elements of the redefined local response matrix $\hat{r}(v)$ under perturbation type I for E, H, and M states. The value of element under each state reflects the regulation strength and regulation type (activated or inhibitory). The network topologies can be deduced by integrating these elements in three states.

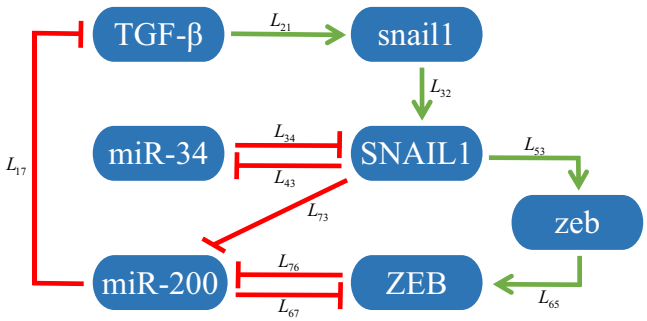

Fig. 6 | The inferred EMT network revealed by $\hat{r}(v)$ under perturbation type I. The regulation of node i by node j is denoted by L_{ij}, where i = 1,..., 7 represent TGF-β, snail1, SNAIL1, miR-34, zeb, ZEB, and miR-200, respectively.

$\hat{r}_{46}(v) = 0$. Furthermore, the network topologies inferred in the EMT network using the redefined local response matrix are largely in agreement with experimental observations. Specifically, these inferred regulations are indeed present in the EMT network, as reported in studies[27,30,31]. For instance, there are double-negative regulations between SNAIL1 and miR-34[42], as well as between ZEB1 and miR-200[43]. Additionally, miR-200 negatively regulates TGF-β[44], etc. In conclusion, it is remarkably accurate to utilize the redefined local response matrix to reveal the network topology, regardless of the perturbation degrees and perturbation distributions. This approach is especially advantageous in biology because there exist various noises and precise measurement is impossible in experiments.

Identifying critical information in EMT network

The redefined local response elements $\hat{r}_{ij}(v)$ offer qualitative insights into types and intensities of direct regulations between two nodes in each state. However, they are insufficient for quantitatively determining dominant regulations within each state or precisely identifying primary states associated with each regulation, solely based on whether $\hat{r}_{ij}(v)$ is non-zero. To overcome this limitation, we calculate the relative local response matrix $\tilde{\boldsymbol{r}}(v)$ under perturbation type I, which captures the relative importance of regulations in all states. Additionally, the relative WT local response matrix $\tilde{\boldsymbol{r}}^{\boldsymbol{0}}(v)$ is also calculated to verify the feasibility of our approach, where $\tilde{r}^0_{ij}(v) = r^0_{ij}(v)/\sum_{v=1}^{3}|r^0{}_{ij}(v)|$.

To assess the accuracy of determining dominant regulations at each state or identifying primary states associated with specific regulations, i.e., characterize the difference between $\tilde{\boldsymbol{r}}(v)$ under perturbation type I and $\tilde{\boldsymbol{r}}^{\boldsymbol{0}}(v)$, non-zero elements of $\tilde{\boldsymbol{r}}^{\boldsymbol{0}}(v)$ and their corresponding elements of $\tilde{\boldsymbol{r}}(v)$ at E, H, and M states are depicted in Fig. 7A, B and C, respectively. By comparing the absolute value of $\tilde{r}_{ij}(v)$ under perturbation type I or $\tilde{r}^0_{ij}(v)$ with the threshold, 1/3, we can quantitatively identify primary cell states associated with specific regulations (Table 3), as well as determine dominant regulations at each cell state (Fig. 8). For each regulation, the darker the color at one state (the corresponding absolute value greater than the threshold 1/3), the more dominant in this state. Obviously, there is little difference between the information of $\tilde{\boldsymbol{r}}(v)$ under perturbation type I and $\tilde{\boldsymbol{r}}^{\boldsymbol{0}}(v)$.

Specifically, to analyze $\tilde{\boldsymbol{r}}(v)$ under perturbation type I, we identify that the inhibitory regulations imposed by miR-200 on TGF-β and ZEB, i.e., L_{17} and L_{67}, are dominant in both E and H states. The activation of snail1 by TGF-β, i.e., L_{21}, and the inhibition of miR-200 by ZEB, i.e., L_{76}, are critical in M state. The translation processes from snail1 to SNAIL1 and from zeb to ZEB, i.e., L_{32} and L_{65}, are critical across all cell states because the relative regulation intensities are almost equal to the threshold 1/3. The inhibition of SNAIL1 by miR-34, i.e., L_{34}, dominates in E state. While the inhibitions of miR-34 and miR-200 by SNAIL1, i.e., L_{43} and L_{73}, are critical in H state. The activation of zeb by SNAIL1, L_{53}, is dominant in H and M states. These results are shown in the second row of Table 3. Further, we identify the dominant regulations are L_{17}, L_{32}, L_{34}, L_{65}, and L_{67} in E state, L_{17}, L_{32}, L_{43}, L_{53}, L_{65}, L_{67}, and L_{73} in H state, and L_{21}, L_{32}, L_{53}, L_{65}, and L_{76} in M state (the left panels of Fig. 8). These results are almost exactly consistent with those in[26], i.e., the critical regulations for common recognition are L_{17}, L_{34}, and L_{67} in E state, and L_{17}, L_{43}, and L_{67} in H state, and L_{21}, L_{32}, L_{53}, L_{65}, and L_{76} in M state.

By analyzing $\tilde{\boldsymbol{r}}^{\boldsymbol{0}}(v)$ (the third row of Table 3 and the right panels of Fig. 8), the inferred and true EMT networks at three cell states exhibit remarkable similarities with the exception of the identification of the regulation from ZEB to miR-34 i.e., L_{46}, which plays a dominant role in M state in the true network. Our approach fails to identify the critical states for this specific regulation because $\hat{r}_{46} = 0$ in all these three states. Therefore, the dominant regulations identified at each state by our method, which is based on the relative local response matrix calculated from the perturbation data, aligns almost perfectly with the results obtained at the E, H, and M states using the WT relative local response matrix derived from the ODEs in[26]. Evidently, the network inferences made by our method closely align with the true networks defined by ODEs, indicating good inference ability of the proposed approach.

Discussion

Cell fate decision is a dynamic process based on mutual regulations between multiple molecules, and plays crucial role in the process of cell development. These intricate regulations form complex regulatory networks, serving as a fundamental tool in computational biology for exploring cell fate decisions. However, inferring the topologies of these gene regulatory networks remains a challenge due to their inherent complexity. In cell fate decisions, regulatory networks exhibit significant variations under distinct cell fates. As a result, if we are able to reconstruct gene regulatory networks which underlie these fate decisions, we can gain profound insights into the regulatory mechanisms behind various phenotypes. Consequently, we can precisely manipulate cells, guiding them into the desired specific state through quantitative control, which offers a practical solution for the treatment of various diseases, including cancer.

In this paper, we propose a general inference approach based on the data (experimental data, single-cell data, or stable steady states data obtained

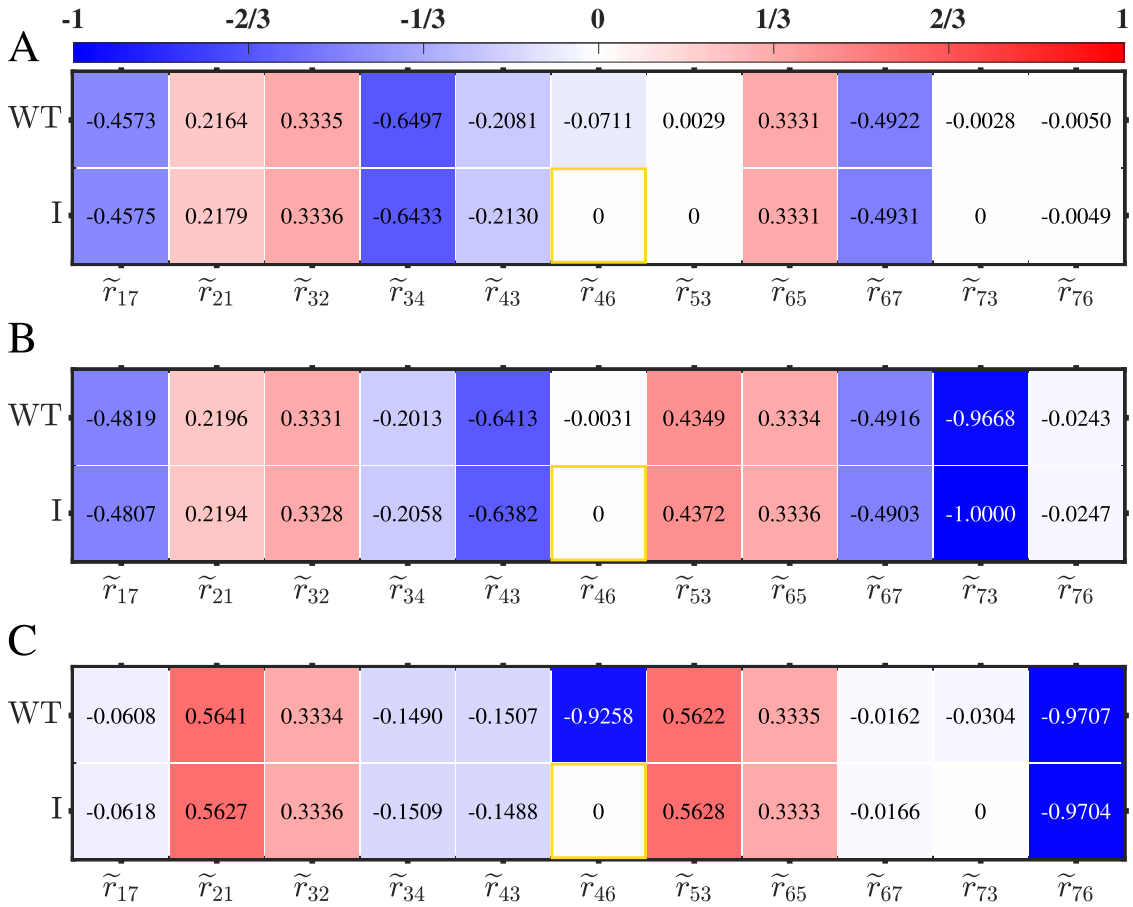

Fig. 7 | The particular elements of the relative WT local response matrix $\widetilde{r}^0(v)$ and the relative local response matrix $\widetilde{r}(v)$ under perturbation type I at E, H, and M states, corresponding to (A), (B), (C), respectively. Each element within the boxes, distinguished by different colors, represents the relative regulation strength. The darker the color, the more important the corresponding regulation. The threshold for the relative strength is set at 1/3. Elements corresponding to unidentified regulations under perturbation type I are highlighted with orange boxes.

Table 3 | Comparison between identified dominant states for specific regulations by the relative local response matrix under perturbation type I and the WT case

Types	Regulations	L_{17}	L_{21}	L_{32}	L_{34}	L_{43}	L_{46}	L_{53}	L_{65}	L_{67}	L_{73}	L_{76}
I		E, H	M	E, H, M	E	H	×	H, M	E, H, M	E, H	H	M
WT		E, H	M	E, H, M	E	H	M	H, M	E, H, M	E, H	H	M

Here, the regulation of node *i* by node *j*, i.e., L_{ij}, corresponds to $\widetilde{r}_{ij}$ and $\widetilde{r}^0_{ij}$ in Fig. 7, where × represents the unidentified regulation that is not recognized in our perturbation analysis.

in silico analysis) before and after perturbations to accurately reconstruct gene regulatory networks, and further qualitatively identify network differences during cell fate decisions. Furthermore, our approach does not impose any limitations on the size of regulatory networks. Specifically, using the steady state data before and after perturbations for each cell fate, we first calculate the local response matrix based on Eqs. (2) and (3), where the value of element reflects regulation between molecules, and the larger the value, the stronger the regulation intensity. Then, based on statistical and differential analyses, i.e., the relative local response matrix from Eqs. (6) and (7), we can quantify network differences across all cell fates, namely identify critical regulations in each cell state and dominant cell states associated with specific regulations. Additionally, we have provided a verified method by comparing our results to the relative local response matrix calculated based on the ODEs of the system, referred to as the WT local response matrix. We use the EMT core network to verify the feasibility of our approach. By the differential analysis, critical regulations in each state and the main states associated with each regulation are accurately identified, which was consistent with some experimental observations.

Based on systematic perturbation, differential, and statistical analysis, a computational approach is developed to accurately infer gene regulatory network and identify network differences during cell fate decisions. The approach provides a general framework which can also be applied to infer regulatory systems, such as embryonic stem cell differentiation[45], cancer metastasis[46], hematopoietic cell differentiation[47], and other networks related to cell fate decisions[48,49]. However, a limitation of the approach is that it cannot identify self-feedback loops, the direct regulation from one molecule to itself. Due to current limitation in experimental conditions, obtaining experimental data under perturbation conditions, such as single-cell data, is another challenge. Our future research will focus on how to infer gene regulatory networks with self-feedback loops by using experimental data under perturbations.

Methods

At the unperturbed stable steady state $\bar{x}$, the Jacobian matrix J is nonsingular and $J_{ii} \neq 0$ (related to degradation and self-regulation). Based on the multivariate implicit function theorem for a small open neighborhood of the

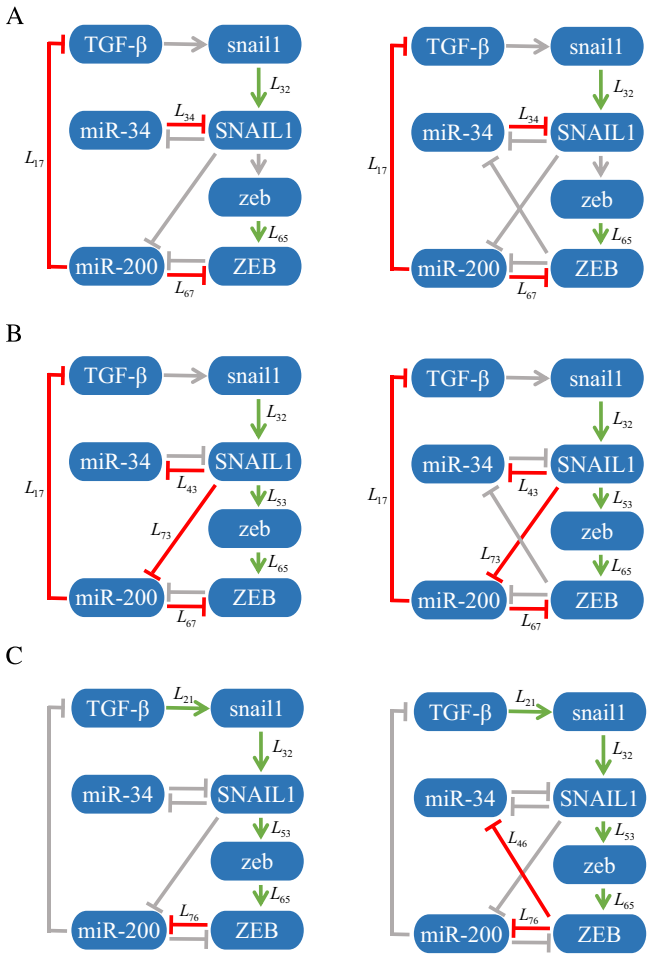

Fig. 8 | The identified (left) and the true (right) critical regulations in E, H and M states, corresponding to (A), (B), (C), respectively. Using the relative local response matrix derived from perturbation type I and the WT relative local response matrix, as depicted in Fig. 7, the dominant regulations at each state have been identified and are indicated by green arrows and red bars. The non-critical regulations are represented as grey in each state.

unperturbed stable steady state $\bar{\boldsymbol{x}}$, we have

$$x_i = x_i\Big(x_1(p_{b,k}), \cdots, x_{i-1}(p_{b,k}), x_{i+1}(p_{b,k}), \cdots, x_n(p_{b,k}), p_{b,i}, p_{c,n+1}, \cdots, p_{c,m}\Big),\ k = 1, \cdots, n,$$

and

$$\frac{\partial x_i}{\partial x_j} = -\frac{\partial f_i/\partial x_j}{\partial f_i/\partial x_i},\ j \neq i,$$

which means that the local response coefficient r_{ij} is related to the elements of Jacobian matrix at the unperturbed stable steady state. Consequently, we obtain

$$\boldsymbol{r} = -\boldsymbol{X}^{-1}\left[\mathrm{diag}(\boldsymbol{J})\right]^{-1}\boldsymbol{J}\boldsymbol{X}.$$

where $\boldsymbol{X} = \mathrm{diag}(\bar{\boldsymbol{x}})$ and $\mathrm{diag}(\boldsymbol{J})$ are nonsingular diagonal matrices composed by the unperturbed stable steady state vector and the diagonal elements of $\boldsymbol{J}$, respectively.

Perturbation analysis of sensitive parameters

Generally, under a slight perturbation to sensitive parameter p_k, the stable steady state remains confined within a small open neighborhood of $\bar{\boldsymbol{x}}$. Considering the derivative of f_i with respect to p_k, and using the multivariable chain rule, we have

$$\left(\frac{\partial f_i}{\partial p_k} + \sum_{j=1}^{n}\frac{\partial f_i}{\partial x_j}\frac{\partial x_j}{\partial p_k}\right)\Bigg|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})} = \frac{\mathrm{d}f_i}{\mathrm{d}p_k}\Big|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})} = 0.$$

Therefore,

$$\boldsymbol{R} = -\boldsymbol{X}^{-1}\boldsymbol{J}^{-1}\boldsymbol{S}\boldsymbol{P}, \tag{5}$$

where $\boldsymbol{S} = (\partial f_i/\partial p_k)|_{(\bar{\boldsymbol{x}};\boldsymbol{p_b};\boldsymbol{p_c})}$, and $\boldsymbol{P} = \mathrm{diag}(\boldsymbol{p_b})$ is a nonsingular diagonal matrix. According to $\mathrm{diag}(\boldsymbol{R}^{-1}) = -\mathrm{diag}(\boldsymbol{P}^{-1}\boldsymbol{S}^{-1}\boldsymbol{J}\boldsymbol{X}) = -\boldsymbol{P}^{-1}\boldsymbol{S}^{-1}\mathrm{diag}(\boldsymbol{J})\boldsymbol{X}$, we find $\boldsymbol{r}\boldsymbol{R} = -[\mathrm{diag}(\boldsymbol{R}^{-1})]^{-1}$. Consequently, the local response matrix $\boldsymbol{r}$ can be solved only from the global response matrix $\boldsymbol{R}$, i.e.,

$$\boldsymbol{r} = -[\mathrm{diag}(\boldsymbol{R}^{-1})]^{-1}\boldsymbol{R}^{-1},$$

which means that the local response matrix $\boldsymbol{r}$ can be directly inferred solely from $\boldsymbol{R}$[21]. The calculation of $\boldsymbol{R}$ can be derived from the cell data gathered before and after the perturbations is applied to the sensitive parameter set, namely, Eq. (2). Therefore, we are able to compute $\boldsymbol{r}$ without needing detailed ODEs or exact parameter values.

Statistical analysis

To enhance the accuracy of inferring network topology, N distinct perturbation sets are utilized and thus N local response matrices can be calculated by Eqs. (2) and (3), i.e., $\boldsymbol{r_l} = (r_{l,ij})$, $l = 1,...,N$. Specifically, a perturbed sensitive parameter set $\boldsymbol{p_s}(l) = \left(p_{s,1}(l), \cdots, p_{s,n}(l)\right)$, $p_{s,k}(l)$ is randomly chosen from a uniform or normal distribution of p_k, $k = 1,...,n$. Consequently, we can derive the distribution of each local response coefficient and its corresponding 95% CI, which is delineated by the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the distribution. This approach allows us to exclude insignificant regulations by evaluating whether their 95% CIs encompass zero, ultimately achieving sparse network. And for the nonzero regulations, we employ the mean of r_{ij} across the N sets to reflect the regulation strength from node j to node i. Then, the redefined local response matrix $\hat{\boldsymbol{r}} = (\hat{r}_{ij})$ is used to represent the inferred network topology, where

$$\hat{r}_{ij} = \begin{cases} \frac{\sum_{l=1}^{N} r_{l,ij}}{N},\ 0 \notin \text{the 95\% CI of } r_{ij}, \\ 0,\ 0 \in \text{the 95\% CI of } r_{ij}. \end{cases} \tag{6}$$

Further, an error analysis is performed to determine the degree of accuracy between inferred network, $\hat{\boldsymbol{r}}$, and true network, $\boldsymbol{r^0}$.

Differential analysis

Combining the concept of relative contributions of driving factors in ecosystem services[50], we employ the relative local response matrix, $\tilde{\boldsymbol{r}}(v)$, to quantitatively characterize the strength of a specific regulation across all cell fates, or to compare all regulations within a specific cell state. This approach enables us to identify the states where a particular regulation dominates, as well as the critical regulations which characterize each state. Specifically, the relative local response matrix $\tilde{\boldsymbol{r}}(v) = (\tilde{r}_{ij}(v))$ is defined as follows

$$\tilde{r}_{ij}(v) = \begin{cases} \frac{\hat{r}_{ij}(v)}{\sum_{v=1}^{s}|\hat{r}_{ij}(v)|},\ \hat{r}_{ij}(v) \neq 0, \\ 0,\ \hat{r}_{ij}(v) = 0. \end{cases} \tag{7}$$

Next, we compare the absolute value of $\tilde{r}_{ij}(v)$ with the threshold, $1/s$, the regulation of node i by node j is significant in the v-th state when $|\tilde{r}_{ij}(v)| - 1/s > 0$. Furthermore, the larger $|\tilde{r}_{ij}(v)|$ is, the more significant the impact of the regulation on the v-th state becomes. Therefore, we can

quantitatively represent the differences in the strength of each regulation across all cell fates, as well as evaluate the disparities of all regulations within a specific cell state.

Data availability

All data used and analyzed during the current study are calculated in MATLAB R2023b and can be accessed via the code at https://github.com/Qing-Hu-gif/network_inference.git. Any other relevant data can be obtained from the authors.

Code availability

The codes needed to reproduce the data and figures are deposited in GitHub at https://github.com/Qing-Hu-gif/network_inference.git.

Received: 23 September 2024; Accepted: 21 February 2025;
Published online: 04 March 2025

References

1. Sáez, M., Briscoe, J. & Rand, D. A. Dynamical landscapes of cell fate decisions. *Interface Focus* **12**, 20220002 (2002).
2. Xu, L., Zhang, K. & Wang, J. Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation. *Plos One* **9**, e105216 (2014).
3. Huang, D. S. & Wang, R. Q. Exploring the mechanisms of cell reprogramming and transdifferentiation via intercellular communication. *Phys. Rev. E* **102**, 012406 (2020).
4. Huang, B. et al. Modeling the transitions between collective and solitary migration phenotypes in cancer metastasis. *Sci. Rep.* **5**, 17379 (2015).
5. Yu, L. L. et al. Modeling the genetic regulation of cancer metabolism: interplay between glycolysis and oxidative phosphorylation. *Cancer Res.* **77**, 1564–1574 (2017).
6. Jiao, J. F., Luo, M. & Wang, R. Q. Feedback regulation in a stem cell model with acute myeloid leukaemia. *BMC Syst. Biol.* **12**, 43 (2018).
7. Akhtar, J. et al. Bistable insulin response: the win-win solution for glycemic control. *iScience* **25**, 105561 (2022).
8. Huang, B., Lu, M. Y., Jia, D. Y., Ben-Jacob, E. & Onuchic, J. Interrogating the topological robustness of gene regulatory circuits by randomization. *Plos Comput. Biol.* **13**, e1005456 (2017).
9. Li, C. H. Landscape of gene networks for random parameter perturbation. *Integr. Biol.* **10**, 92–99 (2018).
10. Huang, B. et al. RACIPE: a computational tool for modeling gene regulatory circuits using randomization. *BMC Syst. Biol.* **12**, 74 (2018).
11. Ji, Z. C. & Ji, H. K. TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* **44**, e117–e117 (2016).
12. Qiu, X. J. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
13. Weinreb, C., Wolock, S. & Klein, A. M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018).
14. Street, K., Risso, D., Fletcher, R. B., Das, D. & Dudoit, S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
15. Luo, Q., Maity, A. K. & Teschendorff, A. E. Distance covariance entropy reveals primed states and bifurcation dynamics in single-cell RNA-Seq data. *iScience* **25**, 105709 (2022).
16. Chan, T. E., Stumpf, M. P. H. & Babtie, A. C. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* **5**, 251-267.e3 (2017).
17. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, e12776 (2010).
18. Kim, S. ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665–674 (2015).
19. Matsumoto, H. et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**, 2314–2321 (2017).
20. Woodhouse, S., Piterman, N., Wintersteiger, C. M., Gttgens, B. & Fisher, J. SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst. Biol.* **12**, 59 (2018).
21. Kholodenko, B. N. et al. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc. Natl Acad. Sci. USA* **99**, 12841–12846 (2002).
22. Andrec, M., Kholodenko, B. N., Levy, R. M. & Sontag, E. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *J. Theor. Biol.* **232**, 427–441 (2005).
23. Rukhlenko, O. S. et al. Control of cell state transitions. *Nature* **609**, 975–985 (2022).
24. Kang, T., Moore, R., Li, Y., Sontag, E. & Bleris, L. Discriminating direct and indirect connectivities in biological networks. *Proc. Natl Acad. Sci. USA* **112**, 12893–12898 (2015).
25. Jimenez-Dominguez, G., Ravel, P., Stéphan, J., Vincent, C. & Colinge, J. An R package for generic modular response analysis and its application to estrogen and retinoic acid receptor crosstalk. *Sci. Rep.* **11**, 7272 (2021).
26. Tian, X. J., Zhang, H. & Xing, J. H. Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. *Biophys. J.* **105**, 1079–1089 (2013).
27. Silveira, D. A. & Mombach, J. C. M. Dynamics of the feedback loops required for the phenotypic stabilization in the epithelial-mesenchymal transition. *FEBS J.* **287**, 578–588 (2020).
28. Das, S., Becker, B. N., Hoffmann, M. F. & Mertz, J. E. Complete reversal of epithelial to mesenchymal transition requires inhibition of both ZEB expression and the Rho pathway. *BMC Cell Biol.* **10**, 1–8 (2009).
29. Drake, J. M., Strohbehn, G., Bair, T. B., Moreland, J. G. & Henry, M. D. ZEB1 enhances transendothelial migration and represses the epithelial phenotype of prostate cancer cells. *Mol. Biol. Cell* **20**, 2207–2217 (2009).
30. Nieto, M. A. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science* **342**, 1234850 (2013).
31. Zhang, J. Y. et al. TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci. Signal.* **7**, ra91 (2014).
32. Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithlial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
33. Ravikrishnan, A. et al. Regulation of epithelial-to-mesenchymal transition using biominmetic fibrous scaffolds. *ACS Appl. Mater. Interface* **8**, 17915–17926 (2016).
34. Pei, D. Q., Shu, X. D., Gassama-Diagne, A. & Thiery, J. P. Mesenchymal-Epithelial transition in development and reprogramming. *Nat. Cell Biol.* **21**, 44–53 (2019).
35. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**, 1420–1428 (2009).
36. Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. MircoRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc. Natl Acad. Sci. USA* **110**, 18144–18149 (2013).
37. Wang, H. Y., Zhang, X. P. & Wang, W. Regulation of epithelial-to-mesenchymal transition in hypoxia by the HIF-1α network. *FEBS Lett.* **596**, 338–349 (2022).
38. Jolly, M. K. et al. Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. Oncol.* **5**, 155 (2015).
39. Steinway, S. N. et al. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ Syst. Biol. Appl.* **1**, 15014 (2015).

40. Peinado, H., Quintanilla, M. & Cano, A. Transforming growth factor β-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions. *J. Biol. Chem.* **278**, 21113–21123 (2003).
41. Dave, N. et al. Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. *J. Biol. Chem.* **286**, 12024–12032 (2011).
42. Siemens, H. et al. miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions. *Cell Cycle* **10**, 4256–4271 (2011).
43. Brabletz, S. & Brabletz, T. The ZEB/miR-200 feedbackloop—A motor of cellular plasticity in development and cancer? *EMBO Rep.* **11**, 670–677 (2010).
44. Gregory, P. A. et al. An autocrine TGF-β/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol. Biol. Cell* **22**, 1686–1698 (2011).
45. Chickarmane, V. & Peterson, C. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS One* **3**, e3478 (2008).
46. Li, C. H. & Wang, J. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Res.* **75**, 2607–2618 (2015).
47. Cheng, H., Zheng, Z. F. & Cheng, T. New paradigms on hematopoietic stem cell differentiation. *Protein Cell* **11**, 34–44 (2019).
48. Huang, B. et al. Decoding the mechanisms underlying cell-fate decision-making during stem cell differentiation by random circuit perturbation. *J. R. Soc. Interface* **17**, 20200500 (2020).
49. Ye, Y. J. et al. An enriched network motif family regulates multistep cell fate transitions with restricted reversibility. *PLoS Comput. Biol.* **15**, e1006855 (2019).
50. Ma, X. N. et al. Assessing the relative contributions, combined effects and multiscale uncertainty of future land use and climate change on water-related ecosystem services in Southwest China using a novel integrated modelling framework. *Sustain. Cities Soc.* **106**, 105400 (2024).

Acknowledgements

This study was funded by the National Natural Science Foundation of China [Grants No. 12371497].

Author contributions

Q.H. performed the computational modeling, data analysis, and draft and revised the manuscript. X.L. and Z.X. performed formal analysis and validation. R.W. conceived the study and draft and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41540-025-00504-2.

Correspondence and requests for materials should be addressed to Ruiqi Wang.