communications physics

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42005-025-02160-8

Landscape control for cell fate transitions

Jinchao Lv[1,2,5], Song Zhang[1,2,5], Wenjia Zhou[3,5] & Chunhe Li[1,2,4]

Cell fate decision making is a core issue in systems biology with profound implications for cellular development and disease. Although dynamical system approaches using gene network models have advanced our knowledge of cell fate transitions, accurately and stably controlling these transitions remains a great challenge. Here, we present a landscape control (LC) approach based on energy landscape theory, which manipulates specific gene targets to direct cell fate. Through testing on a two-gene mutual inhibition and self-activation (MISA) model, an epithelial-mesenchymal transition (EMT) network, and a human embryonic stem cell (HESC) network, we demonstrate that LC significantly outperforms the previous optimal least action control (OLAC) approach in both effectiveness and computational efficiency. Moreover, LC can identify key transcription factors and integrate sparse control strategies to induce specific transitions. Overall, the LC framework provides a valuable tool for studying and engineering cell fate, with potential applications in therapeutic innovation and regenerative medicine.

The gene regulatory networks (GRNs) consist of interacting genes, and the diverse regulatory interactions control cellular processes, such as growth, division, differentiation, and development[1]. These processes have been extensively studied using GRN models, from a deterministic[2,3] or stochastic perspective[4,5]. However, comprehending the mechanism of cell fate decision making and further controlling the cell fate transition remain a great challenge[6].

A possible solution is introducing control theory to GRN models. Pioneering researchers have systematically applied the linear structural controllability theory to complex networks with directed interactions[7,8]. These studies provide a framework for studying linear dynamical systems, and focus on identifying the set of driver nodes to control system dynamics. The theoretical foundation derives from the mathematical condition for controllability, and is known as Kalman's controllability rank condition[9]. However, the real-world GRNs are governed by strong nonlinear dynamics, and control strategies for linear systems may yield unreliable predictions[10].

With regard to the nonlinear system control which is more practically significant, researchers have also conducted meaningful explorations, although the comprehensive and universally applicable framework remains elusive. For example, Wells et al. proposed the optimal least action control (OLAC) approach[11], achieving control over stable-state occupancy. Even though Wells et al. claimed that the computational cost scales linearly with the number of tunable parameters and the dimensionality of the state space[11], in practical implementation, the time required to calculate the transition action in high-dimensional GRNs substantially exceeds expectations. Moreover, to capture the optimal transition path, the selection of hyper-parameters requires careful consideration[12]. These factors significantly constrain the applicability of OLAC in real-world complex GRNs.

Undoubtedly, developing computational methods for controlling cell fate decision making is a critical issue in both theoretical and applied research. The Waddington landscape, initially proposed as a metaphor for cell differentiation[13], presents a promising approach to address this challenge. Recently, landscape quantification approaches have been proposed, providing a valuable way for the quantitative study of cell fate transition problems[5,14,15]. The landscape paradigm effectively characterizes the global stability and stochastic dynamics of diverse cell states, thereby offering an intuitive perspective for implementing control strategies. Previous study has projected trajectories onto one-dimensional landscape to simulate cellular evolution processes and control cell fate transitions[6]. However, the accuracy of representing high-dimensional system with one-dimensional landscape remains a subject of debate. Following the procedures of prior research[16,17], we assume that perturbations in the regulation strengths and protein degradation rates can be employed to achieve control over this high-dimensional, multi-parameter biophysical system. Essentially, the manipulation of regulatory parameters serves to reshape the topology of the underlying potential energy landscape. This reconfiguration can destabilize undesired stable states, which means inducing transitions toward desired stable states. Furthermore, it is crucial to investigate interventions that consider both optimality and sparsity, where sparsity is characterized by targeting only a subset of all available tunable parameters. Such is desirable, as most biologically feasible interventions are capable of modulating only a limited number of parameters.

[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [2]Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China. [3]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. [4]School of Mathematical Sciences, Shanghai Key Laboratory for Contemporary Applied Mathematics and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [5]These authors contributed equally: Jinchao Lv, Song Zhang, Wenjia Zhou. e-mail: chunheli@fudan.edu.cn

In this study, we propose a landscape control (LC) approach for controlling cell fate transitions within the framework of GRNs, which is firmly grounded in landscape theory. In this paradigm, cell fate transitions are modeled as barrier-crossing processes, with the likelihood of transition quantitatively characterized by the barrier height (BH). Specifically, lower BH corresponds to higher transition probability. BH serves as a quantitative measure of landscape topography, and naturally provides a powerful tool for manipulating cell fate decision making. By integrating non-equilibrium potential landscape and saddle point dynamics methods[18–20], we can determine locations of saddle points and achieve more accurate estimations of BH, ultimately yielding a transition probability matrix between stable states. It is noteworthy that the transition paths are irreversible, resulting in an asymmetric transition rate matrix. It differs from previous research that modeled the driving force of GRNs as the negative gradient of potential energy (equilibrium system assumption)[6].

Here, by applying the LC method to a two-gene mutual inhibition and self-activation (MISA) network, an epithelial-mesenchymal transition (EMT) network and a human embryonic stem cell (HESC) network, we demonstrated that LC not only significantly improves computational efficiency, but also achieves better control effectiveness compared to OLAC. Furthermore, through the integration of key control node identification algorithms[21], we proposed sparse control strategies that enable targeted and efficient modulation of cell fate. Applying LC to the identified nodes yields the directional and quantitative change required for each gene to drive the system towards the desired state, thereby providing verifiable predictions for experimental validation. Overall, our findings provide quantitative insights into the mechanisms underlying cell fate decision making, and present a valuable computational framework for guiding experimental studies in cellular reprogramming and regenerative medicine.

Results

Stochastic differential equations (SDEs) for gene regulatory networks (GRNs)

In this study, we will demonstrate the superiority of our LC method, integrated with the identification of key gene targets, for achieving stable-state control across multiple instances of GRNs. Initially, we illustrate the operation of LC in a synthetic two-gene circuit, featuring the MISA model. It represents a fundamental component of large-scale complex GRNs and encompasses the most essential regulatory aspects. Subsequently, we extend LC to high-dimensional GRNs exhibiting multiple stable states, specifically focusing on the EMT network[22,23] and the HESC network[14]. This extension aims to elucidate the applicability of the LC method to real-world GRNs, while simultaneously generating testable biological predictions.

In contrast to deterministic equations modeling the temporal evolution of GRNs including multiple interacting genes[2], stochastic fluctuations have been demonstrated to be crucial factors influencing gene switching behavior[24,25], and thus should not be neglected. To quantify the stochastic dynamics of gene expression in GRNs, SDEs incorporating additive models have been widely adopted[14,23]. Specifically, the SDEs governing the evolution of gene expression dynamics are represented by the following Langevin equations:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\Gamma}(t), \tag{1}$$

where the gene expression vector is denoted by $\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_N(t))$, the extrinsic noise $\boldsymbol{\Gamma}(t)$ is modeled as the Gaussian white noise[14,15,26], and the driving force $\boldsymbol{f}(\boldsymbol{x})$ is a vector composed of $f_i(\boldsymbol{x})$, $i = 1, 2, \ldots, N$. Here, $f_i(\boldsymbol{x})$ is represented by an additive model comprising Hill functions, which has been validated to effectively capture the nonlinear saturation and threshold effects in activation and inhibition. The closed-form of this model is given by:

$$f_i(\boldsymbol{x}) = \frac{dx_i}{dt} = \sum_{j=1}^{N} \frac{A_{ji} x_j^n}{S_{ji}^n + x_j^n} + \sum_{j=1}^{N} \frac{B_{ji} S_{ji}^n}{S_{ji}^n + x_j^n} - k_i x_i, \tag{2}$$

where k_i denotes the basal degradation rate, and is set to unity across all genes without loss of generality. The activation response is modeled by a sigmoidal function, where S_{ji} represents its threshold, and the Hill coefficient n determines its steepness. Conversely, the inhibitory dynamics are described by a monotonically decreasing function of gene expression. The regulatory interactions are characterized by two matrices, (A_{ji}) and (B_{ji}), which quantify the activation and inhibition strengths exerted by gene j on gene i, respectively. In cases where no regulatory relationship exists between two genes, the corresponding entry in (A_{ji}) and (B_{ji}) is set to zero. This formalism forms the basis for our subsequent modeling and systematic analysis of GRN dynamics.

Landscape control (LC) approach

Our main contribution is the development of a computational landscape control (LC) algorithm (Fig. 1), which is inspired by landscape theory and aimed at quantitatively manipulating transitions between stable states in GRNs. The methodology begins with quantifying the landscape topography of GRNs under stochastic fluctuations (see Methods in details). Then, the BH (defined as the potential energy difference between the saddle point and corresponding stable state) is calculated, which characterizes the difficulty of state transitions. Based on our computed BH, stochastic dynamics theory provides an asymptotic formula for approximating the transition rate. It enables the calculation of the limiting occupancy for different stable states (see "Methods" in details), facilitating precise control over the occupancy of the desired state and the effective identification of key gene targets.

We now detail the specific procedure for our LC approach. The stochastic dynamics of gene expression, as formulated in Eq. (1), can be equivalently represented by the corresponding Fokker-Planck equation (FPE)[27], which governs the temporal evolution of the density function of state variables. Through the implementation of the truncated moment equation approach (see "Methods" in details)[23], we derive the approximate solution of FPE, and subsequently obtain its stable-state solution as time approaches infinity. Specifically, the probability density function of a multistable system can be approximated via a Gaussian mixture model, with the mixture weights determined through stochastic initial sampling. Then the non-equilibrium potential energy is quantified through $U = -\ln p_{ss}(x)$[28,29], where $p_{ss}(x)$ represents the stable-state density function. This formulation enables the quantification of the BH, wherein the locations of saddle points in the landscape should be precisely determined. A computationally efficient saddle dynamics approach is implemented to deal with this problem[19,20,30].

Through systematic adjustment of the parameters specified in Eq. (2), the resultant modification in transition rate from state i to state j is determined via the asymptotic formula: $R_{ij}^{\epsilon}(\Omega) = \exp(-\frac{1}{\epsilon}\mathrm{BH}_{ij}(\Omega))$[29,31,32], where ϵ represents the noise intensity and Ω denotes the control parameter set. It is also important to note that the asymptotic approximation holds for small noise conditions, which are indeed characteristics of gene regulatory systems.

Given that the stable-state transitions are governed by Markov processes, the limiting occupancy is characterized by its stationary distribution, where the transition probability matrix is constructed based on our estimated transition rates (see Methods in details). Here the optimization objective is to maximize the occupancy of the desired state i: $\max_{\Omega} G(\{\mathrm{BH}_{ij}(\Omega)\}; \varepsilon) := o_i^{\epsilon}$, in which the function G can be analyzed through the asymptotic formula and balance equation. Meanwhile, the optimization process must incorporate constraints imposed by biophysical principles and experimental limitations, including the non-negativity of gene expression and the sparsity of manipulated parameters.

Our proposed LC methodology demonstrates superior computational efficiency in comparison with direct numerical simulation of the stochastic dynamics characterized by Eq. (1) and subsequent statistical analysis of rare events. Furthermore, extending beyond traditional linear control framework, LC accommodates nonlinear formulations in both the objective functions and control strategies. These characteristics not only align with the intrinsic nonlinear nature of gene regulatory processes, but also enrich the applicability to a comprehensive range of dynamic behavioral patterns. The

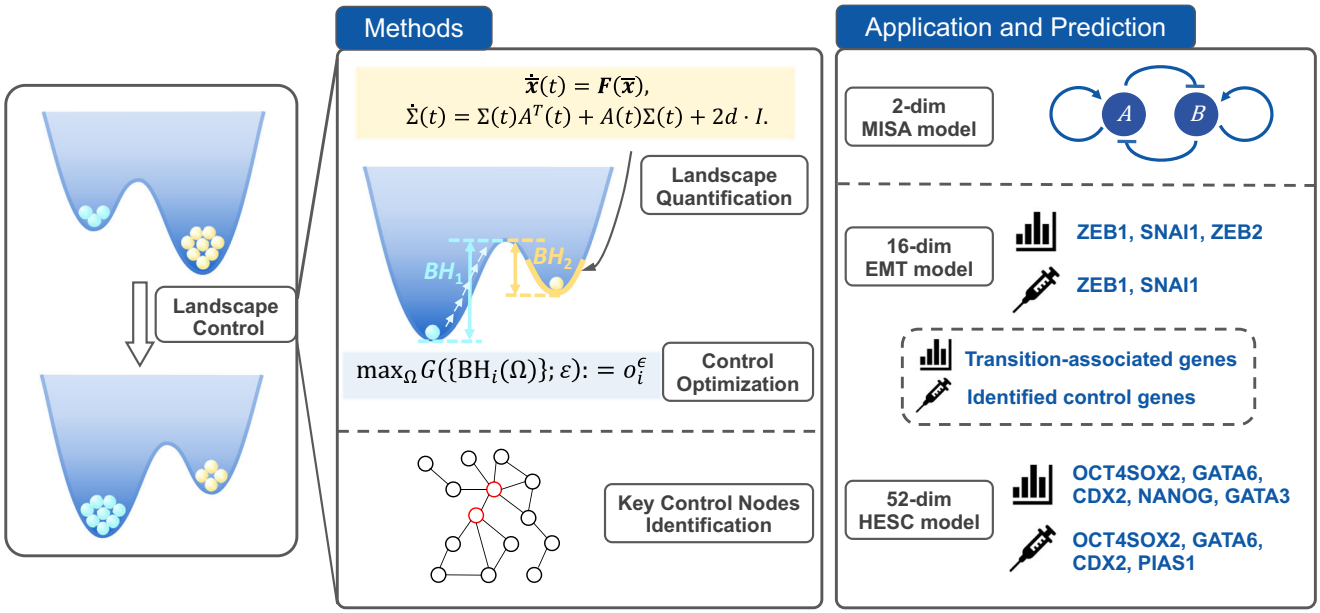

Fig. 1 | The schematic workflow of the landscape control (LC) approach. The LC methodology aims to amplify the desired state by modulating model parameters Ω without inducing phase transitions. The approach is predicated on landscape quantification, wherein the probability distribution around each stable-state is approximated as Gaussian through truncated moment equations. The mean value and covariance matrix are computed via the ordinary differential equations highlighted in the yellow box. The control optimization objective is to maximize the occupancy rate o_i^ϵ of the desired state i (ε is the noise intensity), which can be formulated as a function of the barrier height (BH). The LC framework serves dual purposes: it enables the identification of critical transition-associated gene targets, and facilitates the control of large-scale networks when integrated with control node identification algorithms. LC's efficiency and effectiveness for stable-state control are demonstrated by applying it to three representative networks with different dimensions involved in cell fate decision making.

comparative analysis also demonstrates that LC outperforms OLAC in both stable-state control and computational efficiency across multiple GRNs: a 2-dimensional MISA model, a 16-dimensional EMT network[22], and a 52-dimensional HESC network[14]. Regarding the latter two higher-dimensional GRNs, we establish systematic strategies for identifying and manipulating crucial gene targets, wherein our computational predictions can be tested by future experimental studies.

Landscape control for a two-dimensional MISA model

To illustrate the operational principles of LC, we presented its implementation in a simplified two-dimensional MISA GRNs. This model reveals a fundamental regulatory module governing cell fate decision making, as depicted in Fig. 2a. The two-gene expression dynamics, described by Eq. (1), demonstrate parameter-dependent transitions between monostable and multistable regimes.

Through our numerical analyses of GRNs characterized as high-dimensional non-gradient systems, we demonstrate a near proportional relationship between BH quantified from the potential landscape and minimum action calculated via the principle of least action (Figs. 2b and S1 and Supplementary Note 1), as well as an approximately linear relationship between BH and mean first passage time computed through Eq. (1) (Fig. S2). This observation is demonstrated in gradient systems (Supplementary Note 2), and motivates the investigation of BH regulation as a potential mechanism for directing system dynamics toward desired states.

As an illustrative example, we considered a specific set of parameters in the tristable regime: $n = 4$, $A_{ji} = 1$, $B_{ji} = 1$, $S_{ji} = 0.5$, $k_i = 1$ for all j and i (see bistability and tetrastability in Figs. S3 and S4 and Supplementary Note 3). We quantified the landscape through the application of truncated moment equations (Ω expansion). Then the direction and magnitude of control parameters were systematically optimized to maximize the occupancy rate of the intermediate state. Here, four parameters: S_{11}, S_{12}, S_{21}, and S_{22} are assumed to be adjustable, which govern the GRN by controlling the interaction thresholds of the corresponding regulatory elements. Given the initial parameter configuration (prior to control), the intermediate state maintained an occupancy rate below 30%. Under phase stability constraints, the LC algorithm determines the optimal parameters: $S_{11} = 0.08$, $S_{12} = 0.67$, $S_{21} = 0.09$, and $S_{22} = 0.68$. It led to the occupancy of the intermediate state above 99% (Fig. 2c), verifying the successful amplification efficiency. Meanwhile, the optimal amplification parameters obtained through OLAC were $S_{11} = 0.10$, $S_{12} = 0.68$, $S_{21} = 0.10$, and $S_{22} = 0.64$. Comparative analysis revealed that the parameter control strategies identified by LC and OLAC converge to comparable solutions in this MISA implementation.

We also mentioned that, due to the non-convex nature of the objective function, LC may yield multiple possible control parameter solutions; however, these solutions achieve comparably effective control outcomes (Fig. S4). Furthermore, when parameters are perturbed within a certain range, averaging the barrier heights can provide a more accurate estimate of transition rates (Fig. S5). We additionally adjusted the gene-dependent Hill coefficients to verify the robustness of LC in achieving desirable control performance (Table S1). We further applied LC to regulate transcription rates (corresponding to A_{ji} and B_{ji} in the additive model), which offers greater experimental feasibility (Fig. S6). Besides, LC exhibits the capability in addressing noise-induced transitions, through the incorporation of noise intensity as a component within control parameter sets. We have investigated the effects of varying noise intensity d on both the BH and the limiting occupancy rates (Fig. S7).

A critical advantage of LC lies in its computational efficiency. Experimental results indicate that LC not only achieves comparable or superior control performance in comparison with OLAC, but also substantially reduces the computational time. It is reasonable, since the core of OLAC lies in expressing the objective function in terms of the Freidlin-Wentzell action between stable states, and each iteration requires searching for optimal transition paths in high-dimensional space. The computation of the action employs an implementation of the adaptive minimum action method[12], and the time cost of OLAC undergoes substantial increase with the growing number of stable states (Fig. 2d).

In contrast, LC effectively addresses this limitation by quantifying the BH based on the global landscape topography. For complex multistable scenarios, the computational cost of LC does not exhibit significant changes, rendering it particularly advantageous for analyzing more complex systems.

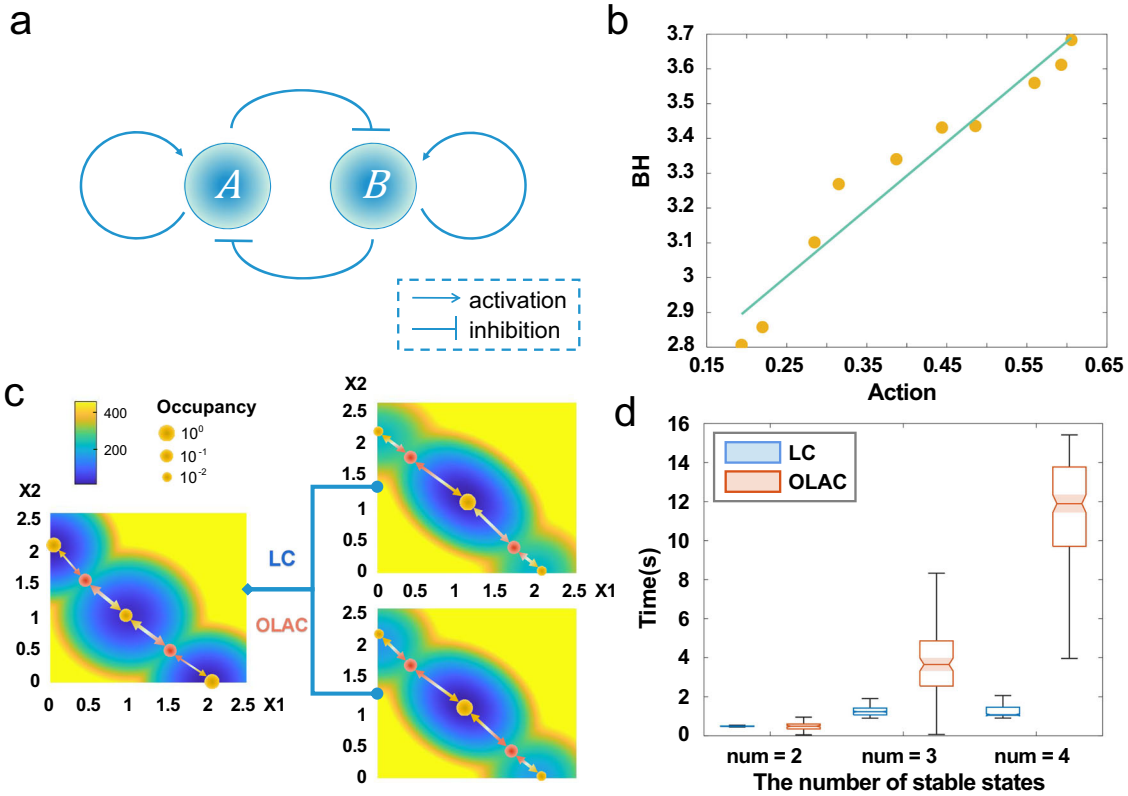

Fig. 2 | Landscape control (LC) for a two-dimensional mutual inhibition and self-activation (MISA) model. a The network structure of the MISA motif, consisting of gene A and gene B, is characterized by nonlinear self-activation loops within each gene and mutual inhibitory regulation between the two genes. **b** An approximately linear relationship was observed between the BH quantified from the landscape and the minimum action corresponding to the most probable path. The orange dots represent computational results obtained through perturbation of model parameters, while the green line indicates the least squares linear fitting. **c** Landscape modifications were presented both before and after the application of LC and OLAC approaches. Both control strategies successfully amplified the occupancy rate of the intermediate state, which increases from less than 30% to over 99%. In the landscape visualization, yellow dots represent the positions of stable states, with dot sizes proportional to their occupancy rates. Red dots locate the saddle point positions, which were determined through saddle dynamics calculations. **d** LC demonstrated superior computational efficiency compared to OLAC, particularly for systems with multiple stable states. The computational time required for single optimization iteration was evaluated across multiple parameter sets with different stable-state configurations. For each boxplot, the central line represents the median, and the box indicates the first quartile (Q1) to the third quartile (Q3). The whiskers are defined as the minimum and maximum values within 1.5 times the interquartile range (IQR) from Q1 and Q3, respectively, where the IQR is calculated as Q3 minus Q1.

In the context of this MISA model, LC and OLAC manifest comparable computational demands in the bistable regime. However, in multistable scenarios, especially in cases of tetrastability, two control approaches display distinct computational efficiency differences (Fig. 2d). Furthermore, LC reveals distinctive strengths in its inherent capability for parallel implementation. The fundamental component of LC lies in the landscape quantification, in which stable-state weight computation is accomplished through Monte Carlo simulations, consequently facilitating parallel computational acceleration. Conversely, the optimal path computation integral to OLAC exhibits an iterative nature, precluding parallel acceleration. Such aforementioned attributes endow LC with exceptional computational performance while maintaining control effectiveness.

Landscape control for a 16-dimensional EMT network

Following the successful exploration of LC in the simplified MISA model, we extended our analytical framework to investigate a 16-dimensional core GRN of EMT, with the aim of elucidating its regulatory mechanisms and transition dynamics. The ultimate goal of our comprehensive analysis is to generate experimentally verifiable predictions that can improve biological design paradigms. EMT is considered as a fundamental cell fate decision making process associated with cancer progression. Our previous work has established a core GRN of EMT[22], which consists of 12 transcription factors, 4 microRNAs and their regulatory relationships (Fig. 3a). Experimental evidence and computational analyses have revealed that the transitions between epithelial (E) and mesenchymal (M) states occur typically via intermediate states (IM), rather than through direct conversion[22,33].

For subsequent analysis, we selected a parameter set that exhibits tristable behavior (Supplementary Note 4). The three stable states are characterized as E, M, and IM states, wherein *CDH1* and *VIM* serve as key markers for discriminating between the E state and the M state. Then we utilized LC to modulate the model parameters, thereby modifying gene regulation and controlling transitions among E, IM, and M states. In accordance with our previous observations, an approximately linear relationship between BH and minimum action is maintained in this 16-dimensional system (Figs. 3b and S8), suggesting the broader applicability of LC to high-dimensional GRNs. While energy barriers in high-dimensional space cannot be visualized directly, the locations of saddle points and stable points can be determined through saddle dynamics and stable-state solutions of Eq. (2), respectively. Similarly, high-dimensional landscape also cannot be visualized clearly, so we projected it onto two-dimensional coordinates using our model-based dimension reduction approach for landscape[23].

Here we aimed to amplify the E state (or the M state) without inducing phase transitions, and then compared the control effectiveness between LC and OLAC. Considering the practicality in experiments, where manipulating individual genes is more feasible than adjusting interaction strengths between genes, we selected the degradation rates of all genes (except for the downstream genes *CDH1* and *VIM*) as control optimization parameters.

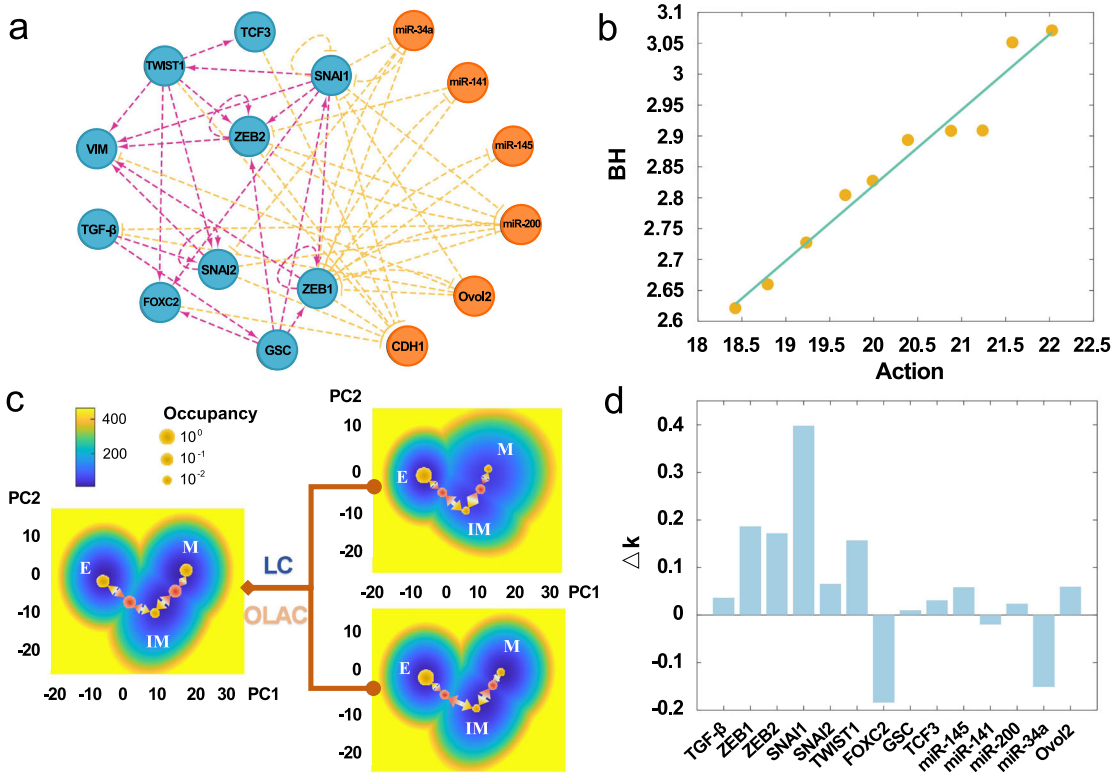

Fig. 3 | Landscape control (LC) for a 16-dimensional epithelial-mesenchymal transition (EMT) network. a The core network governing EMT consists of 12 transcription factors and 4 microRNAs, interconnected through activation (magenta) and inhibition (yellow) links. The network comprises 6 orange nodes representing epithelial (E) state markers and 10 blue nodes representing mesenchymal (M) state markers. **b** An approximately linear relationship between BH and minimum action is preserved in the high-dimensional EMT system. **c** Landscape modifications of the EMT system were presented both before and after the application of LC and OLAC approaches. While LC successfully increases the occupancy of the E state to over 99%, OLAC achieves a more modest increase to less than 91%. The legend follows that of Fig. 2c. **d** LC approach determined the directional changes and magnitudes of degradation rates for 14 genes optimized for E state amplification.

Under default parameters, the occupancy rates of E and M states were 35% and 39%, respectively. Following our LC implementation, the occupancy of the E state was adjusted to exceed 99%, whereas OLAC reached a maximum occupancy below 91% (Fig. 3c). Analogous results were observed in M state amplification, where LC implementation resulted in the occupancy of M state exceeding 81%, whereas OLAC only achieved its value below 76% (Fig. S9). The suboptimal amplification efficiency of the M state can be attributed to the presence of the intermediate state. The experimental findings demonstrate that LC shows improved control efficiency against OLAC, maintaining the performance advantage within high-dimensional cases.

We try to explain the observed differences between two control approaches. OLAC optimizes transition paths through iterative optimization and requires calculations under identical parameter configurations to ensure comparability. However, the hyper-parameters such as the time intervals corresponding to different optimal paths demonstrate substantial variability, thereby requiring precise specifications (Figs. S10–S12). This computational setting may introduce instability and bias in action calculations, which negatively impact the control outcome. In contrast, the BH metric required in LC is quantified based on a global landscape perspective, inherently maintaining analytical comparability without involving specific parameter selections necessitated by OLAC. This distinction potentially accounts for the enhanced control performance observed in LC implementation.

We also presented a set of gene degradation rate modifications under LC that facilitate E state amplification (Fig. 3d). The regulatory adjustments are mechanistically plausible: increase of the E state occupancy requires accelerated degradation rates of M state-specific marker genes and reduced degradation rates of E state-specific markers. Additional control strategies and degradation rate modulations for M state amplification are provided in Fig. S13. Comprehensive analysis reveals that M state markers, such as ZEB1, ZEB2, SNAI1, exhibited greater parameter fluctuations, while no similar significant changes were observed among E state marker genes. We discovered that these gene nodes have higher out-degrees, meaning that perturbations to them influence a larger number of other nodes in the network, potentially triggering cascading effects.

Regarding computational efficiency, LC demonstrates more pronounced advantages over OLAC in high-dimensional systems (Fig. S14). This superiority is partly due to differences in BH and transition paths calculations, and partly because OLAC requires predetermined parameter settings, whereas LC does not (Figs. S10–S12).

Landscape control of HESC network identifies optimal targets for reprogramming

We next apply LC to a 52-dimensional HESC network, a higher-dimensional GRN that governs the core regulatory process of embryonic stem cell fate decision making[14] (Fig. 4a). By modeling intricate interactions among transcription factors, this network captures the regulation mechanisms of cell differentiation potential.

Our previous studies have elucidated that the HESC system exhibits a broad bistable regime[14]. Under our uncontrolled parameters (Supplementary Note 5), the occupancy rates were 40% for the stem cell state and 60% for the differentiated state. By perturbing model parameters, we quantified the BH of the landscape and the transition action of the most probable path, and found that they maintained the significant linear relationship observed previously (Figs. 4b and S15). Subsequently, we employed LC and OLAC to amplify the stem cell state (Fig. 4c) and differentiated cell state (Fig. S16), respectively, assuming the degradation rates of all 52 genes were tunable

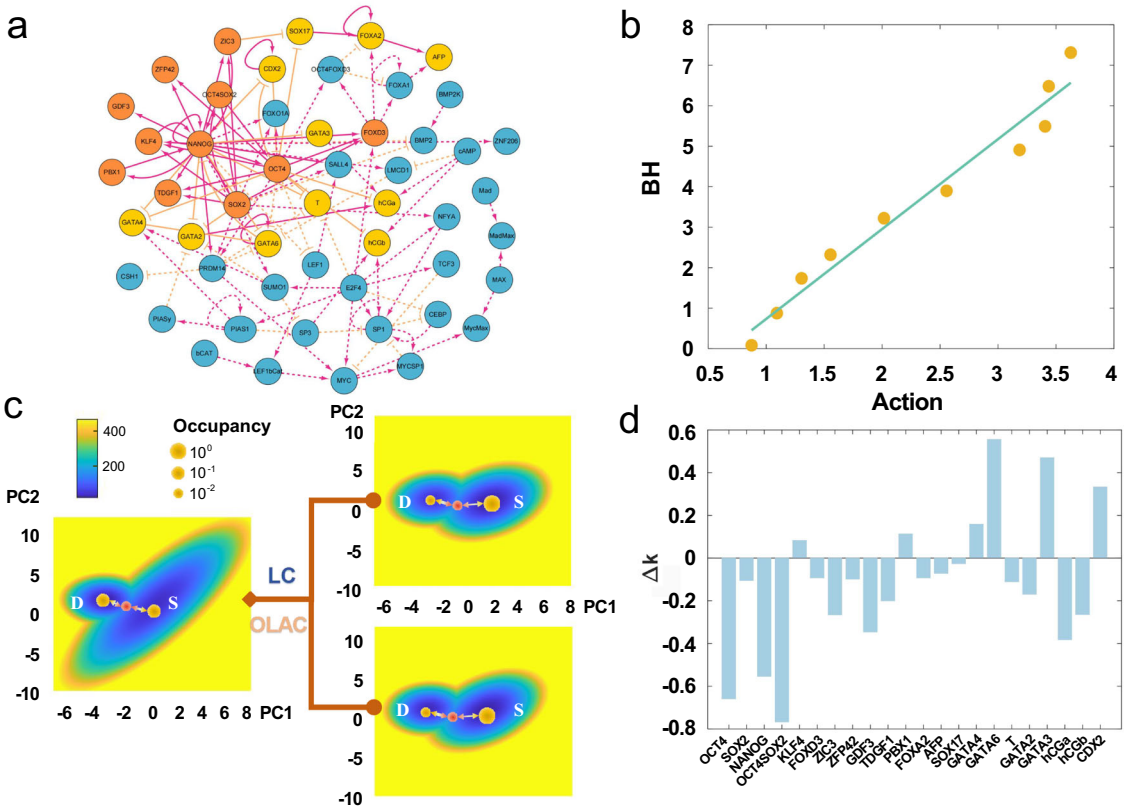

Fig. 4 | Landscape control (LC) for a 52-dimensional human embryonic stem cell (HESC) network. a The core network governing cell fate decision making in HESC comprises 52 genes and their activation (magenta) and inhibition (yellow) interactions. The network consists of 11 orange nodes representing stem cell state markers, 11 yellow nodes representing differentiated state markers, and 30 blue nodes representing other regulatory genes. Solid lines indicate connections between markers, while dashed lines represent other interactions. **b** A near-linear relationship between BH and minimum action is again observed in this 52-dimensional system. **c** Landscape modifications of the HESC system were presented both before and after the application of LC and OLAC methods. While LC successfully increases the occupancy of the stem cell state to over 81%, OLAC yields a lower efficiency with an increase to less than 73%. **d** Directional changes and magnitudes of degradation rates were displayed for 11 stem cell state markers and 11 differentiated state markers, as determined by LC to optimize the stem cell state amplification.

parameters. As with previous findings, LC achieved significantly higher maximum occupancy rates than OLAC (for stem state amplification, LC: larger than 81%, OLAC: less than 73%, Fig. 4c; for differentiated state amplification, LC: larger than 99%, OLAC: less than 98%, Fig. S16).

Considering that simultaneously manipulating 52 genes in biological experiments is challenging, and the computational simulations suggest that the proliferation of regulatory parameters frequently results in multiple control strategies, we aimed to reduce the number of regulated genes. To achieve more robust and feasible results, we assumed that only the degradation rates of 22 marker genes associated with stem and differentiated cells were tunable. Then we discovered that, amplification efficiency using 22 genes was not significantly different from that with 52 genes, emphasizing the critical regulatory roles of these 22 marker genes within the network. For stem cell state amplification, our analysis revealed that the majority of stem cell marker genes had decreased degradation rates (Fig. 4d), indicating enhanced expression of these genes, while most differentiation marker genes showed increased degradation rates. A comparison between stem cell and differentiated cell state amplification demonstrated that stem cells required more substantial modifications in gene degradation rates (Fig. S17). This is consistent with biological intuition, as reprogramming typically requires stronger regulatory interventions and stimulations than differentiation.

Besides, the identification of key gene targets in high-dimensional GRNs remains crucial and urgent. We aim to incorporate constraints governing the quantity or magnitude of control parameters in our computational model, thereby enhancing the reliability of gene target identification. Based on multiple simulations under this configuration, several genes exhibited significantly higher regulatory amplitudes relative to others, thus these genes are identified as key regulatory factors: the stem cell markers NANOG and OCT4SOX2, together with the differentiation markers GATA6, GATA3, and CDX2. These observations demonstrate concordance with prior global sensitivity analyses conducted under parametric perturbations[14], and some are corroborated by experimental validation studies[34,35].

Integrating landscape control and DCGS achieves better control efficiency

We aimed to propose a solution for applying LC to large-scale networks. By quickly identifying key nodes, we can limit the number of tunable parameters, thus achieve efficient control of desired state. Specifically, we integrated LC with the Divide and Conquer for Global Stabilization (DCGS) approach[21] (Fig. 5) and leveraged the critical nodes identified by DCGS for precision regulation. This combination accelerates the optimization process and enhances the practicality of the LC methodology for more complex networks.

We first applied the DCGS approach to identify key control genes in the GRN, enabling effective control by manipulating fewer genes. The approach first uses the structural properties of the given network to divide it into a union of strongly connected components (SCCs), and subsequently identifies all feedback vertex sets (FVSs) for each component. By integrating the dynamic information of Boolean networks, DCGS further classifies FVS into a canalizing set and canalized set (Fig. 5), where the canalizing set controls the behavior of other nodes and plays a pivotal role in the network[36]. In the following simulations, the degradation rates of nodes in the canalized set are treated as adjustable parameters, while the rates for the rest of nodes

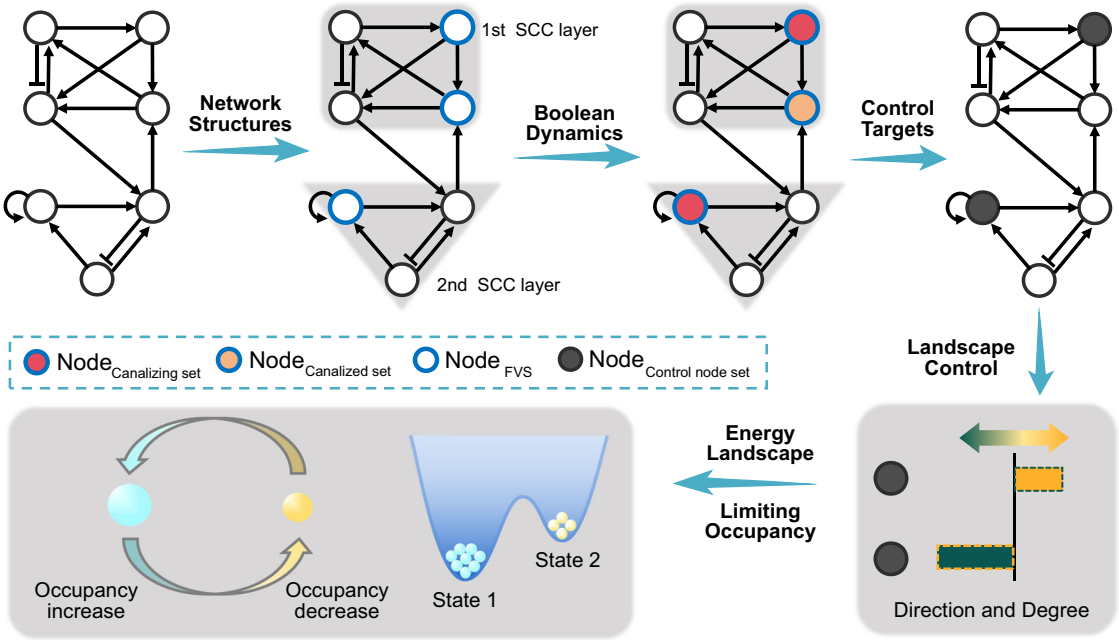

Fig. 5 | Integration of landscape control (LC) and key node identification algorithms for controlling large-scale networks. The upper panel illustrates the divide and conquer for global stabilization (DCGS) workflow, which identifies critical control nodes through the integration of structural and dynamical information to determine canalizing sets. The lower panel demonstrates the application of LC to the key nodal parameters, optimizing both direction and magnitude of regulation to maximize the limiting occupancy of the desired state. SCC strongly connected component, FVS feedback vertex set.

remain fixed. This strategy allows for efficient and effective control of the system by using LC. Next we demonstrated the application in the EMT network previously explored, as well as in the HESC network.

For the EMT network, we identified an FVS comprising four genes: ZEB1, ZEB2, SNAI1, and SNAI2. Further analysis identified ZEB1 and SNAI1 as members of the canalizing set. We set the degradation rates of the two genes as adjustable parameters and aimed to control the amplification of either the E or M state. Our results showed that there was no significant difference in amplification effectiveness between controlling two parameters and controlling fourteen parameters. Specifically, for amplifying the E state (Fig. 6a), increasing the degradation rates of ZEB1 and SNAI1 raised the E state occupancy to 99.58%. For amplifying the M state (Fig. 6b), reducing the degradation rates of these two genes increased M state occupancy to 78.56%. Additionally, we compared single-gene control outcomes and found that ZEB1 plays a more critical role in controlling the EMT system than SNAI1. This can be attributed to the feedback mechanisms in the network. Although SNAI1 activates multiple M state genes and inhibits half of the E state genes, its self-inhibition limits its expression and regulatory capacity. In contrast, ZEB1 not only inhibits all E state genes but also self-activates, thereby enhancing its regulatory influence on the network.

For the HESC network, the FVS comprises six genes: NANOG, OCT4SOX2, FOXA2, GATA6, CDX2, and PIAS1, while the canalizing set consists of OCT4SOX2, GATA6, CDX2, and PIAS1. Similarly, we set the degradation rates of these four genes as adjustable parameters and aimed to achieve the amplification of the stem cell state (Fig. 6c) and the differentiated state (Fig. 6d). The control strategy utilizing four key parameters yielded performance comparable to the full-parameter adjustments, enhancing stem cell state occupancy rates to 79.13%, and differentiated state occupancy rates to 98.71%. Single-gene control analysis identified OCT4SOX2, and CDX2 as the most critical gene targets for regulating system states. This aligns with findings from network structural analysis, where OCT4SOX2 directly influences the three stem cell marker genes OCT4, SOX2, and NANOG, whereas CDX2 directly inhibits OCT4 and exhibits self-activation.

Discussion

The Waddington landscape has emerged as a powerful metaphor for cell differentiation, providing an intuitive and effective framework for quantitative analysis of transition dynamics in cell fate decision making. However, there remains great challenges in quantifying landscapes of high-dimensional systems and measuring the relative stability between attractors. Additionally, the development of universal frameworks for nonlinear system control remains elusive, particularly when addressing computational complexities in high-dimensional contexts. Our proposed LC methodology, inspired by these challenges, works by directly manipulating the underlying landscape topography. This differs from controlling individual trajectories[37], as our approach implements global and intrinsic control. LC modifies the potential energy landscape through BH modulation, effectively guiding the system toward desirable states. LC adopts the constraint from the previous OLAC approach to avoid inducing phase transitions[11], aiming to explore parameter adjustments that are efficient yet confined within a limited range. Benefiting from the global characterization provided by the landscape, LC is particularly well-suited for systems exhibiting multistability, and its computational complexity increases linearly with the system dimension N. Through computational experiments conducted on multiple GRNs, especially in high-dimensional cases, we demonstrated that LC achieves both substantially improved computational efficiency and markedly superior control outcomes compared to previous OLAC approach[11].

LC can be widely applied to the regulation of diverse parameters. In the two-dimensional model, we modulated the threshold S, which is related to the binding affinity of transcription factors and sensitivity of signaling pathways. While S characterizes gene response, precise experimental control is challenging. In the high-dimensional model, we adjusted degradation rates k, which can be experimentally modulated using RNA interference or protein degradation techniques. Additionally, transcription rates (corresponding to the activation strength A_{ji} and inhibition strength B_{ji} in the additive model) can also be adjusted, which is more achievable in experimental contexts.

Moreover, our LC framework enables systematic downstream analysis for identifying crucial gene targets (Fig. 1). To elucidate genes intimately

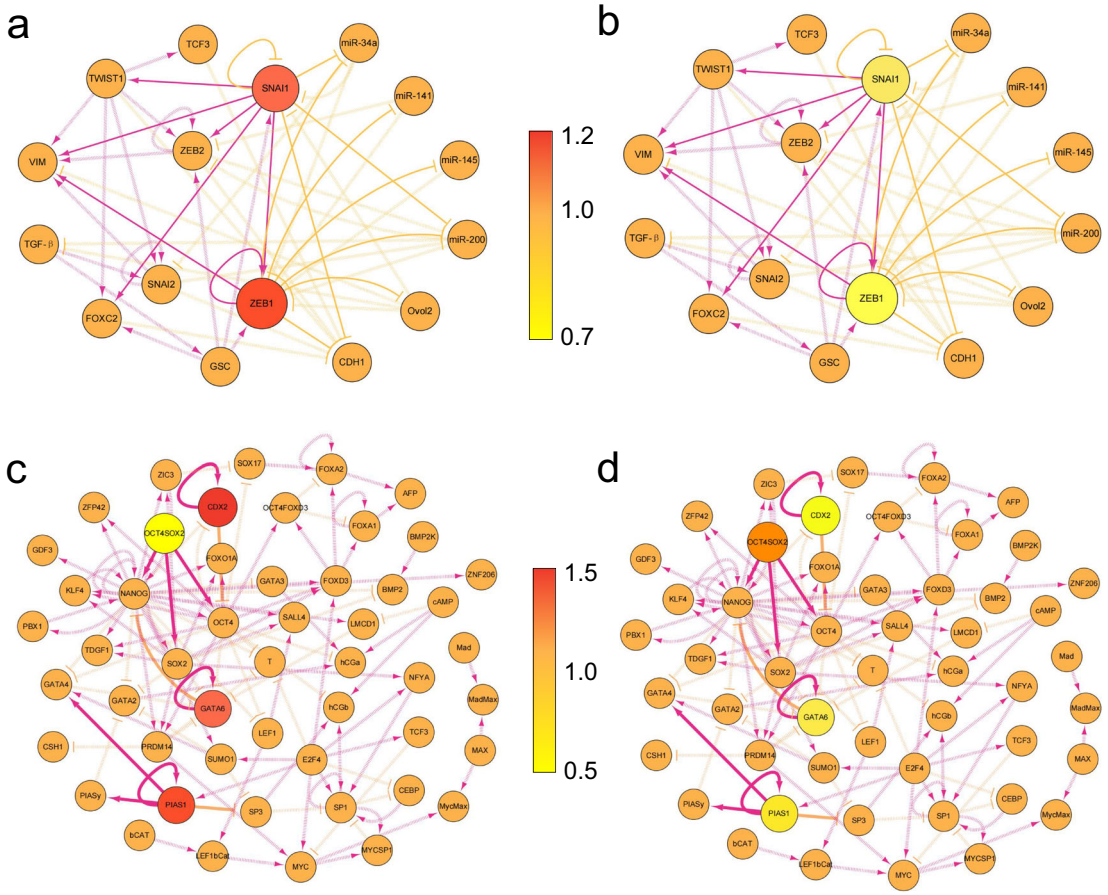

Fig. 6 | Application of landscape control (LC) to the identified critical control nodes demonstrates effective stable-state regulation. a, **b** The DCGS analysis of the EMT network identified ZEB1 and SNAI1 as critical control nodes. LC optimization of degradation rates for these two nodes achieved effective control for stable-state occupancy, with objectives set to amplify either the E state (**a**) or the M state (**b**). The node colors represent the optimized degradation rates after LC implementation. **c**, **d** For the HESC network, DCGS approach identified OCT4SOX2, GATA6, CDX2, and PIAS1 as critical control nodes. LC approach optimized the degradation rates for these four nodes yielded comparable stable-state occupancy control to that achieved by regulating all parameters, with control objectives targeting either stem cell state amplification (**c**) or differentiated state amplification (**d**). Similarly, magenta and orange arrows represent activation and inhibition, respectively, while the outgoing edges from the identified control nodes are denoted by bold solid lines.

associated with cell fate transitions, we employed two complementary analytical approaches. First, we assessed the magnitude of changes in gene degradation rates during stable-state control processes. The second approach leveraged the DCGS algorithm to prescreen a sparse set of control nodes, demonstrating that targeted manipulation of these nodes yields comparable outcomes to regulating all parameters. Significantly, the gene targets identified through both methods showed substantial overlap, generating model predictions that warrant further experimental validation. In the HESC system, OCT4SOX2 emerged as a particularly compelling target, as it both shows the largest parameter variation during transition and belongs to the control node set. OCT4SOX2 is a protein complex formed by OCT4 and SOX2 that co-bind to enhancer sequences to regulate gene expression. Indeed, OCT4 and SOX2 have been established as two of the four most crucial factors in cellular reprogramming[34,38]. Recent experiments have revealed their synergistic interaction, wherein the OCT4SOX2 complex demonstrates a 650-fold enhancement in binding affinity compared to OCT4 alone[39]. This evidence indirectly validates the effectiveness of LC predicted targets.

In certain simplified scenarios, combining LC with bifurcation diagrams can provide a more comprehensive perspective on understanding the system dynamic behaviors. The versatility establishes LC as a promising theoretical framework for developing generalized control strategies across diverse complex biological systems. Its utility extends far beyond our presented cases of cell fate decision making, encompassing critical areas such as cancer systems biology. Integrated with synthetic biology approaches that facilitate the rational design of genetic circuits[40], LC shows therapeutic potential for cancer treatment. Moreover, in the context of epigenetic disorders, LC facilitates the identification of conditions that enhance systemic resilience against pathological perturbations, elucidates mechanistic relationships between environmental factors and disease manifestation, and generates valuable insights for developing preventive intervention strategies.

Overall, the LC methodology establishes a scalable framework for controlling complex networks with multiple variables, parameters, and attractors. It is worth noting that, due to the non-convex nature of this optimization problem, LC does not yield a unique optimal parameter set, a challenge also encountered in OLAC. However, our numerical experiments revealed that the optimized parameter sets maintain consistent trends and achieve similar control objectives. Besides, LC requires only system dynamics equations, while accommodating diverse control constraints. Future investigations will extend the LC framework in the context of more complex uncertainty models, and empirical biological data acquired through experimental observations (in the absence of explicit kinetic formulations). This will be complemented by direct validation of gene target identification using in vivo and in vitro experiments.

Methods

Quantifying the landscape through truncated moment equations

Here, we assume that the external noise $\boldsymbol{\Gamma}(t)$ is homogeneous, and independent at different time. We denote d as the diffusion coefficient, also

known as the noise intensity. Then the temporal evolution of the probability density function $p(\boldsymbol{x}, t)$ is governed by the FPE, which is equivalent to Eq. (1):

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}[f_i(\boldsymbol{x})p(\boldsymbol{x}, t)] + d\sum_i\sum_j \frac{\partial^2}{\partial x_i \partial x_j} p(\boldsymbol{x}, t). \quad (3)$$

However, solving Eq. (3) for high-dimensional systems poses significant challenges. To approximate the solution of Eq. (3), we are inspired by the theoretical framework of Ω expansion, employing a Gaussian distribution along the deterministic trajectory for each state. Following this assumption, our focus narrows down to computing its mean $\bar{\boldsymbol{x}}(t)$ and covariance matrix $\boldsymbol{\Sigma}(t)$. If d is assumed to be relatively small, the truncated moment equations conform to the following ordinary differential equations (ODEs)[28,29]:

$$\begin{aligned}\dot{\bar{\boldsymbol{x}}}(t) &= \boldsymbol{f}(\bar{\boldsymbol{x}}(t)),\\ \dot{\boldsymbol{\Sigma}}(t) &= \boldsymbol{\Sigma}(t)\boldsymbol{A}^T(t) + \boldsymbol{A}(t)\boldsymbol{\Sigma}(t) + 2d\cdot\boldsymbol{I},\end{aligned}$$

where I is the identity matrix, and $\boldsymbol{A}(t)$ is the Jacobian matrix of $\boldsymbol{f}(\boldsymbol{x})$ evaluated at $\boldsymbol{x} = \bar{\boldsymbol{x}}(t)$, i.e., $A_{ij}(t) = \frac{\partial f_i(\boldsymbol{x})}{\partial x_j}\big|_{\boldsymbol{x}=\bar{\boldsymbol{x}}(t)}$. Then the evolution of the density function for each state $p(\boldsymbol{x}, t)$ can be approximated by:

$$p(\boldsymbol{x}, t) = \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{\Sigma}(t)|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}}(t))^T\boldsymbol{\Sigma}^{-1}(t)(\boldsymbol{x}-\bar{\boldsymbol{x}}(t))\right\}.$$

When t goes to infinity, we consider $p(\boldsymbol{x}, t)$ to have converged to the stable-state function $p_{ss}(\boldsymbol{x})$. For the multistability analysis, we use weighted sum of Gaussians to estimate the density function of the system: $p_{ss}(\boldsymbol{x}) = \sum_{j=1}^{M}\phi^j p_{ss}^j(\boldsymbol{x})$, where M is the number of stable states, $p_{ss}^j(\boldsymbol{x})$ is the stable-state density of the j^{th} state, and $\phi^j (j = 1, \ldots, M)$ denotes its corresponding weight, representing the relative size of different basins of attraction. As $\sum_{j=1}^{M}\phi^j = 1$, $p_{ss}(x)$ is also a probability density function. Ultimately, the potential energy landscape is quantified through $U = -\ln p_{ss}(\boldsymbol{x})$[28,29].

Calculation of BH through saddle point dynamics

To calculate the BH from the landscape, the positions of saddle points need to be precisely determined. Numerous previous studies have focused on employing the gentlest ascent dynamics to locate the index-1 saddle points[18,41–43]. The stable fixed points of the following dynamics correspond to the (index-1) saddle points of the system $\boldsymbol{x}$, and two directional vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ that are required to track both the right and left eigenvectors of the Jacobian $\nabla\boldsymbol{f}(\boldsymbol{x})$:

$$\begin{aligned}\dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}) - 2\boldsymbol{w}^T\boldsymbol{f}(\boldsymbol{x})\boldsymbol{v},\\ \dot{\boldsymbol{v}} &= \nabla\boldsymbol{f}(\boldsymbol{x})\boldsymbol{v} - \boldsymbol{v}^T[\nabla\boldsymbol{f}(\boldsymbol{x})]^T\boldsymbol{v}\cdot\boldsymbol{v},\\ \dot{\boldsymbol{w}} &= [\nabla\boldsymbol{f}(\boldsymbol{x})]^T\boldsymbol{w} - \left\{2\boldsymbol{v}^T[\nabla\boldsymbol{f}(\boldsymbol{x})]^T\boldsymbol{w} - \boldsymbol{v}^T[\nabla\boldsymbol{f}(\boldsymbol{x})]^T\boldsymbol{v}\right\}\boldsymbol{w},\end{aligned}$$

where we take the regularization by setting $\boldsymbol{v}^T\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{w} = 1$. To approximate $\nabla\boldsymbol{f}(\boldsymbol{x})$ (corresponding to the negative Hessian matrix of potential in the equilibrium system) at each iteration step, the "dimer" method was proposed[44]. It essentially implements matrix multiplication through central difference schemes. Specifically, $\nabla\boldsymbol{f}(\boldsymbol{x})\boldsymbol{v}$ is approximated by $\frac{\mathrm{d}}{\mathrm{d}l}\boldsymbol{f}(\boldsymbol{x}+l\boldsymbol{v})|_{l=0} \approx \frac{1}{2l}(\boldsymbol{f}(\boldsymbol{x}+l\boldsymbol{v}) - \boldsymbol{f}(\boldsymbol{x}-l\boldsymbol{v}))$, with the dimer length $2l$ for some $l > 0$.

After obtaining the saddle point position through the saddle dynamics, we substituted it into the stable-state density $p_{ss}(\boldsymbol{x})$, and then achieved more accurate estimate of the potential energy at the saddle point. Thus, we calculated the energy BH that is used to measure the relative stability: $\mathrm{BH} = U_{\mathrm{saddle}} - U_{\mathrm{stable}}$, where U_{saddle} and U_{stable} represent the energy at the saddle point and the stable-state, respectively.

Formulation and implementation of LC optimization objectives

Based on the calculated BH and the asymptotic formula, we can estimate the transition rate R_{ij}^{ϵ}, which signifies the transition probability per unit of time from state i to state j. For this non-equilibrium gene regulatory process, the transition paths and rates both exhibit asymmetry, i.e., $R_{ij}^{\epsilon} \neq R_{ji}^{\epsilon}$. Then we can define the transition probability matrix P satisfying $P_{ij} = R_{ij}^{\epsilon}$ for $i \neq j$, where P_{ij} denotes the transition probability from state i to state j, and $P_{ij} = 1 - \sum_j R_{ij}^{\epsilon}$ for $i = j$, which ensures that each row of P sums equal to one.

Over an extended period, these transitions result in a stationary distribution over the occupied stable states, which is referred to as the system's limiting occupancy rate and denoted by $\boldsymbol{o}^{\epsilon}$. In fact, for any finite state Markov chain, at least one stationary distribution exists. Moreover, if the Markov chain is irreducible and aperiodic, then the stationary distribution is unique. It can be obtained by solving $\boldsymbol{o}^{\epsilon}P = \boldsymbol{o}^{\epsilon}$, where $\boldsymbol{o}^{\epsilon} = [o_1^{\epsilon}, \ldots, o_M^{\epsilon}]$ and I is the identity matrix. Our goal is to modify model parameters Ω (including the degradation rates, regulatory thresholds, and interaction strengths, and among others) to optimize either the transition rate R_{ij}^{ϵ} directly or the limiting occupancy $o_i^{\epsilon} \in \boldsymbol{o}^{\epsilon}$.

Furthermore, we have to consider biophysical and experimental constraints that limit the range of possible parameter values. These constraints include non-negativity and sparsity, which significantly influence the selection of control strategies and constrain the number of targets in the control set. We formalize this problem as an optimization problem with an objective function G over the parameters Ω:

$$\begin{aligned}&\max_{\Omega} G\left(\left\{\Delta U_{ij}^*(\Omega)\right\}; \varepsilon\right) := R_{ij}^{\epsilon} \text{ or } o_i^{\epsilon}\\ &\text{subject to } \boldsymbol{x}(\Omega) \geq 0,\end{aligned}$$

where $\Delta U_{ij}{}^*(\Omega)$ represents an approximation of $\mathrm{BH}_{ij}(\Omega)$ based on saddle point dynamics, and the constraint is formulated such that all gene expressions are non-negative. This general formulation is the central part of our implementation of LC, which is independent of the system's dimension and complexity. With this framework, we are able to identify control interventions that alter the switching behavior and stability of states in desired ways. We performed the optimizations through the interior-point and limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithms. Besides, the objective function and its associated constraints are non-linear, enabling the control methodology of diverse dynamical behaviors.

Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Code availability

The computational codes are publicly available on GitHub (https://github.com/chunhelilab/LC).

Received: 15 December 2024; Accepted: 22 May 2025;
Published online: 07 June 2025

References

1. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780 (2008).
2. Huang, B. et al. Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS Comput. Biol.* **13**, e1005456 (2017).
3. Khan, F. M. et al. Unraveling a tumor type-specific regulatory core underlying e2f1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. *Nat. Commun.* **8**, 198 (2017).
4. Cao, Z. & Grima, R. Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nat. Commun.* **9**, 3305 (2018).
5. Li, C. & Wang, J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc. Natl. Acad. Sci. USA* **111**, 14130–14135 (2014).

6. Rukhlenko, O. S. et al. Control of cell state transitions. *Nature* **609**, 975–985 (2022).
7. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
8. Liu, Y.-Y. & Barabási, A.-L. Control principles of complex systems. *Rev. Mod. Phys.* **88**, 035006 (2016).
9. Kalman, R. E. Mathematical description of linear dynamical systems. *J. Soc. Ind. Appl. Math., Ser. A: Control* **1**, 152–192 (1963).
10. Jiang, J. & Lai, Y.-C. Irrelevance of linear controllability to nonlinear dynamical networks. *Nat. Commun.* **10**, 3961 (2019).
11. Wells, D. K., Kath, W. L. & Motter, A. E. Control of stochastic and induced switching in biophysical networks. *Phys. Rev. X* **5**, 031036 (2015).
12. Zhou, X. & Ren, W. et al. Adaptive minimum action method for the study of rare events. *J. Chem. Phys.* **128**, 104111 (2008).
13. Waddington, C. H. *The Strategy of the Genes* (Routledge, 2014).
14. Li, C. & Wang, J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS Comput. Biol.* **9**, e1003165 (2013).
15. Wang, J., Xu, L. & Wang, E. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. USA* **105**, 12271–12276 (2008).
16. Li, C. & Balazsi, G. A landscape view on the interplay between EMT and cancer metastasis. *NPJ Syst. Biol. Appl.* **4**, 34 (2018).
17. Chen, Z., Lu, J., Zhao, X.-M., Yu, H. & Li, C. Energy landscape reveals the underlying mechanism of cancer-adipose conversion in gene network models. *Adv. Sci.* **11**, 2404854 (2024).
18. Zhang, J. & Du, Q. Shrinking dimer dynamics and its applications to saddle point search. *SIAM J. Numer. Anal.* **50**, 1899–1921 (2012).
19. Yin, J., Zhang, L. & Zhang, P. High-index optimization-based shrinking dimer method for finding high-index saddle points. *SIAM J. Sci. Comput.* **41**, A3576–A3595 (2019).
20. Yin, J., Wang, Y., Chen, J. Z., Zhang, P. & Zhang, L. Construction of a pathway map on a complicated energy landscape. *Phys. Rev. Lett.* **124**, 090601 (2020).
21. An, S. et al. Global stabilizing control of large-scale biomolecular regulatory networks. *Bioinformatics* **39**, btad045 (2023).
22. Lang, J., Nie, Q. & Li, C. Landscape and kinetic path quantify critical transitions in epithelial-mesenchymal transition. *Biophys. J.* **120**, 4484–4500 (2021).
23. Kang, X. & Li, C. A dimension reduction approach for energy landscape: Identifying intermediate states in metabolism-emt network. *Adv. Sci.* **8**, 2003133 (2021).
24. Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).
25. Ge, H., Qian, H. & Xie, X. S. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Phys. Rev. Lett.* **114**, 078101 (2015).
26. Lv, J., Wang, J. & Li, C. Landscape quantifies the intermediate state and transition dynamics in ecological networks. *PLoS Comput. Biol.* **20**, e1011766 (2024).
27. Chandrasekhar, S. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **15**, 1 (1943).
28. Hu, G. *Stochastic Forces and Nonlinear Systems* (Shanghai Scientific and Technological Education Publishing House, 1994).
29. Kampen, N. *Stochastic Processes in Chemistry and Physics* 2nd edn (North Holland, 1992).
30. Zhang, L., Du, Q. & Zheng, Z. Optimization-based shrinking dimer method for finding transition states. *SIAM J. Sci. Comput.* **38**, A528–A544 (2016).
31. Wentzell, A. D. *Random Perturbations of Dynamical Systems* (Springer, 1998).
32. Risken, H. & Risken, H. *Fokker-Planck Equation* (Springer, 1996).
33. Jolly, M. K. et al. Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7**, 27067 (2016).
34. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
35. Theunissen, T. W. et al. Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr. Biol.* **21**, 65–71 (2011).
36. Reichhardt, C. O. & Bassler, K. E. Canalization and symmetry in Boolean models for genetic regulatory networks. *J. Phys. A: Math. Theor.* **40**, 4339 (2007).
37. Aparicio, A., Velasco-Hernández, J. X., Moog, C. H., Liu, Y.-Y. & Angulo, M. T. Structure-based identification of sensor species for anticipating critical transitions. *Proc. Natl. Acad. Sci. USA* **118**, e2104732118 (2021).
38. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
39. Michael, A. K. et al. Mechanisms of Oct_4-Sox_2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).
40. Mukherji, S. & Van Oudenaarden, A. Synthetic biology: understanding biological design from synthetic circuits. *Nat. Rev. Genet.* **10**, 859–871 (2009).
41. Weinan, E. & Zhou, X. The gentlest ascent dynamics. *Nonlinearity* **24**, 1831 (2011).
42. Gao, W., Leng, J. & Zhou, X. An iterative minimization formulation for saddle point search. *SIAM J. Numer. Anal.* **53**, 1786–1805 (2015).
43. Crippen, G. & Scheraga, H. Minimization of polypeptide energy: Xi. The method of gentlest ascent. *Arch. Biochem. Biophys.* **144**, 462–466 (1971).
44. Henkelman, G. & Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **111**, 7010–7022 (1999).

Acknowledgements

C.L. is supported by the National Natural Science Foundation of China (Grant no. 12171102) and the National Key R&D Program of China (Grant no. 2019YFA0709502).

Author contributions

C.L. designed the project, J.L., S.Z., W.Z. conducted the experiments, analyzed the results and wrote the manuscript, and C.L. revised the 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/s42005-025-02160-8.

Correspondence and requests for materials should be addressed to Chunhe Li.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.