OXFORD

Briefings in Bioinformatics, 2024, **25(5)**, bbae406
https://doi.org/10.1093/bib/bbae406
Problem Solving Protocol

Canalizing kernel for cell fate determination

Namhee Kim, Jonghoon Lee, Jongwan Kim, Yunseong Kim, Kwang-Hyun Cho*

Laboratory for Systems Biology and Bio-inspired Engineering, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

*Corresponding author. Department of Bio and Brain Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. Tel: +82-42-350-4325; Fax: +82-42-350-4310, E-mail: ckh@kaist.ac.kr

Abstract

The tendency for cell fate to be robust to most perturbations, yet sensitive to certain perturbations raises intriguing questions about the existence of a key path within the underlying molecular network that critically determines distinct cell fates. Reprogramming and trans-differentiation clearly show examples of cell fate change by regulating only a few or even a single molecular switch. However, it is still unknown how to identify such a switch, called a master regulator, and how cell fate is determined by its regulation. Here, we present CAESAR, a computational framework that can systematically identify master regulators and unravel the resulting canalizing kernel, a key substructure of interconnected feedbacks that is critical for cell fate determination. We demonstrate that CAESAR can successfully predict reprogramming factors for de-differentiation into mouse embryonic stem cells and trans-differentiation of hematopoietic stem cells, while unveiling the underlying essential mechanism through the canalizing kernel. CAESAR provides a system-level understanding of how complex molecular networks determine cell fates.

Keywords: cell fate change; canalizing kernel; master regulator; positive feedback loop; Boolean network

Introduction

Cellular systems tend to robustly maintain their committed fates, despite diverse perturbations, including external signals, internal noise, or genetic alterations [1, 2]. However, they also show flexible cell fate changes to certain perturbations. For instance, Yamanaka *et al.* found that fully differentiated fibroblasts can be reprogrammed into induced pluripotent stem cells by regulating only four transcription factors (OCT4, SOX2, KLF4, and c-MYC; OSKM), showing the potential to override inherent cellular robustness [3, 4]. This capability is further underscored by many other studies showing successful reprogramming or trans-differentiation with regulation of only a few molecular switches [5–9]. Such findings pose fundamental and longstanding questions, which have yet to be answered: which molecular regulators induce cell fate changes and what is the underlying mechanism for cell fate determination?

Previous observations on cell fate determination hint the existence of an essential regulatory path that determines distinct cellular states. This idea recalls Waddington's epigenetic landscape, wherein a developmental pathway is likened to a marble moving towards a valley, symbolizing a stable cellular phenotype [10, 11] (Fig. 1A). However, certain perturbations of cellular systems have been shown to change predefined trajectories, resulting in dramatic cell type transitions [12, 13]. This implies that the canal, which typically ensures stable cellular states, can be rewired by a small number of selective perturbations.

Since cellular state changes within a dynamical system follow changes in complex gene regulations, the underlying canal can be delineated by expounding such regulations. Molecular network models are key to understanding emergent properties of these complex regulations [14, 15]. In such a regulatory network model, an attractor represents the equilibrium state reached after repeated state transitions in a high-dimensional state space [16, 17]. Hence, cell fate change can be considered as an attractor-to-attractor transition, control of which requires identifying molecular targets that direct cell fate change. Despite previous studies [18, 19] on complex network control in this direction, difficulties remain in identifying optimal molecular targets for a desired state transition. A recently suggested one-step temporary perturbation (OT) [20] suggests a possible solution for attractor-to-attractor control, yet it requires finely modulating time intervals, which is impractical in biological experiments.

Previous studies have demonstrated the significance of positive feedback loops (FBLs), which constitute a circular chain of interactions in the network [21, 22]. These positive FBLs play a crucial role in maintaining multiple stable states, making them indispensable for understanding cellular decision-making processes such as (trans-)differentiation and reprogramming [23–26]. Following this insight, Crespo *et al.* proposed an approach to identify control targets by only investigating differentially expressed positive circuits (DEPCs) between two states [27]. However, relying solely on FBL structures without incorporating the network dynamics caused by the underlying regulation logic is overly simplified, and thus fails to fully comprehend the underlying mechanisms and identify optimal control targets.

While extensive approaches have been employed to identify master regulators capable of inducing desired state transitions,

Received: March 7, 2024. **Revised:** July 14, 2024. **Accepted:** August 1, 2024

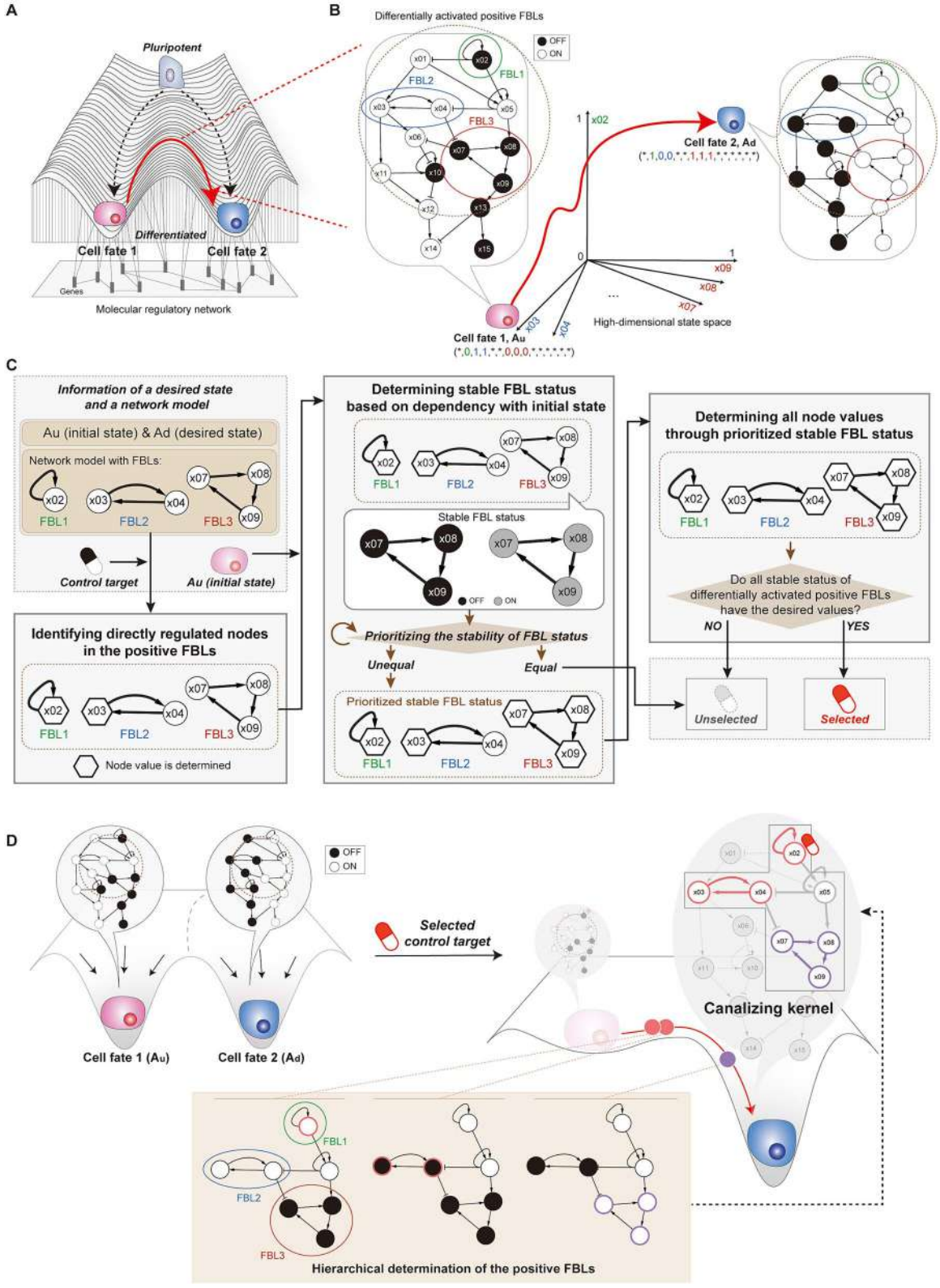


Figure 1. Workflow of CAESAR. (A) Waddington's epigenetic landscape. (B) A control framework for cell fate change. Each attractor state, corresponding to each cell fate, is represented by collection of the FBL status using the following codes: 0 for OFF, 1 for ON, and ∗ for non-FBL status. (C) A schematic view of CAESAR; first, CAESAR requires the following inputs: A_u for an initial state, A_d for a desired state, and a network model (top left); next, CAESAR identifies directly regulated nodes (bottom left) and prioritizes FBL status based on its dependency with the initial state (middle); finally, CAESAR determines all node states based on the finally prioritized stable FBL status and identifies a potential master regulator (right). (D) Canalizing kernel for cell fate determination. The selected control targets (master regulators) specify a hierarchical connection of FBLs which constitute the resulting canalizing kernel. Further details can be found in Fig. 2.

the resulting canal which includes a critical regulatory structure that facilitates such cell fate transitions remains elusive. Here, we present canalizing kernel-based systematic analysis for state transition control (CAESAR), a computational framework designed to pinpoint both the master regulator and the corresponding canalizing kernel for a desired cell fate determination. The canalizing kernel is an ensemble of interconnected FBLs that delineate the sequential regulations within the fate-determining path to the desired cell state. In particular, CAESAR can evaluate which positive FBLs predominantly influence the fate-determining path given the initial network state.

Using CAESAR, we were able to identify a minimal number of regulators, which can effectively alter cell states, even for large complex networks. Moreover, when applied to real biological networks, including those of mouse embryonic stem cell (mESC) and hematopoietic stem cell (HSC), CAESAR could successfully decode the fundamental principles with which cellular identity is determined under the regulation of specific molecular switches, which were previously opaque despite experimental observations. For instance, during the transition of B-cells into macrophages, we unraveled the role of the specific FBL in the canalizing kernel, comprising Cebpb and Spi1, which corroborated previous experimental data. Finally, we showed that how positive FBLs sequentially coordinate cellular states within the canalizing kernel, thereby proposing a comprehensive computational framework for uncovering master regulators and their underlying mechanisms during cell fate determination.

Materials and Methods

Workflow of CAESAR

To systematically identify master regulators and decipher the underlying mechanism of the resulting cell fate change after controlling the master regulator, we employ a mathematical model that can describe the dynamics of cellular states in terms of gene regulatory interactions. In particular, we use a Boolean network model that is the simplest mathematical representation that can still capture the essential dynamics of complex biological behavior [28–30]. Here, a network state (or state) is defined by a collection of node values. In other words, a state $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is a vector composed of node values mapped to either 1 (active) or 0 (inactive). In general, an attractor indicates a steady state in the state space of a dynamical system, representing either a single point or a limit cycle of points. For instance, differentiated cellular states are often well described by point attractors [31, 32]. Based on this concept, let us consider cellular states as point attractors and possible transitions between them caused by some external perturbations. These point attractors can be categorized into undesired ($\mathbf{A_u}$) or desired ($\mathbf{A_d}$) depending on the expression of specific phenotype markers. A given problem is to induce a transition from $\mathbf{A_u}$ to $\mathbf{A_d}$. For this attractor-to-attractor transition control, we note that interconnected positive FBLs can function as a bi-stable switch between those cell fates [23, 25, 33]. Hence, we define a master regulator as a set of nodes that is capable of changing the node values of positive FBLs such that they can significantly contribute to inducing a state transition from undesired to desired point attractors.

Figure 1 illustrates our proposed control framework, CAESAR, using an example Boolean network model with 15 nodes. In this illustration, there are undesired and desired attractors, A_u = (1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0) and A_d = (0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1), and three positive FBLs with different values between A_u and A_d. We define these FBLs as differentially activated positive FBLs (diff-FBLs) to specify the most important components for a transition from A_u to A_d (Fig. 1B; Materials and Methods). In CAESAR, A_u, A_d, and the Boolean network model with the information of diff-FBLs are used as inputs (Fig. 1C, top left). Using this information, CAESAR considers a set of nodes in diff-FBLs with node values of A_d as the primary control target candidates. For each control target candidate, CAESAR initially screens directly regulated nodes, which refer to the nodes that are uniquely determined by the regulation effect of the control target regardless of other node values. Then CAESAR checks the necessity of additional perturbation by comparing the regulation effect of the control target candidate and the given initial state ($\mathbf{A_u}$) if not all the node values are determined as shown in Fig. 1C, bottom left.

A positive FBL in diff-FBLs usually has two stable stationary states with opposite node values: one corresponds to the desired and the other to the undesired, which we denote as FBL status. The FBL status is not determined by the FBL alone, but also by the status of other multiple FBLs connected to it [34]. In CAESAR, we aim to identify the most effective targets among the control target candidate such that its control leads to the desired status of all diff-FBLs, which ultimately results in the state transition from A_u to A_d. To do this, we first introduce the concept of FBL stability, which denotes whether a specific FBL status can maintain its current status when exploring the regulation effects of other nodes. In other words, a stable FBL status can keep their node values in a specific condition. However, in complex dynamics, the stability of an FBL status can be changed due to the regulation effects from another stable FBL status or a given initial state with perturbations. Hence, we introduce the notion of prioritizing to specify which FBL status keeps its stability in a certain condition during the update of node values from the given undesired state. For instance, under a specific condition, FBL3 can maintain two possible stable FBL status with different node values due to the given initial state (Fig. 1C, middle). However, when sequentially updating the node values from the undesired state, the stability of a FBL status can be changed to its opposite status following the logical regulation dynamics after target perturbation. For this, CAESAR prioritizes the stable FBL status for each FBL, and iteratively evaluates whether the priority of each stable FBL status remains consistent when incorporating a set of prioritized stable FBL status. Finally, for FBL3 illustrated in Fig. 1C, if two FBL status have equal priorities, the control target is not selected as a master regulator because both FBL status are stable, so the desired state cannot be guaranteed. On the other hand, if one status becomes destabilized depending on another, the latter has a higher priority (unequal case) (Fig. 1C, middle). Based on this assessment, CAESAR designates the final prioritized stable FBL status (Fig. 1C, middle and top right). Consequently, all node values are determined through finally prioritized stable FBL status, and then CAESAR verifies whether all values of diff-FBLs correspond to the desired values. If all values have the desired values, the control target is ultimately identified as the master regulator for this transition (Fig. 1C, right).

By following the steps outlined in Fig. 1C, if there is any master regulator (selected control target), CAESAR can unravel a subnetwork that most critically contributes to driving the desired cellular state change, when the master regulator is controlled from A_u. This is illustrated upon the epigenetic landscape, which includes two attractors and the transition trajectory (fate-determining path toward the desired cell fate) converging to the desired attractor (Fig. 1D). Together, these results of hierarchically determined FBLs are used to identify the resulting canalizing kernel (Fig. 1D,

right). More detailed procedures of CAESAR are described in the next section.

Systematic identification of master regulators and their canalizing kernels

When a Boolean network model is given, all attractors are computed and two specific attractors are categorized into A_u and A_d (Fig. 2A, left). By comparing the node values of all positive FBLs within the network between A_u and A_d, CAESAR identifies the diff-FBLs, a set of FBLs which have different node values between A_u and A_d. In the example Boolean network model described in Fig. 1, there are two attractors, $A_u = (1,0,1,1,1,1,0,0,0,0,1,1,0,1,0)$ and $A_d = (0,1,0,0,1,0,1,1,1,0,0,0,1,0,1)$, and the diff-FBLs is identified as a set of three positive FBLs (annotated as FBL1–3, respectively; Fig. 2A, right). From this, we initially determine control target candidates as the nodes inside the diff-FBLs only with their corresponding node values of A_d. In this case, the control target candidates is C= {x_{02} (+), x_{03} (−), x_{04} (−), x_{07} (+), x_{08} (+), x_{09} (+)} (Fig. 2A, Stage 1). For each $c_i \in C$, CAESAR introduces the expanded network (also called the prime-implicant hypergraph) [35, 36] that is useful to investigate the status of all FBLs and their logical relationships using the graph theory. The expanded network represents each Boolean variable as an individual node (e.g. x_{01} for active and $\sim x_{01}$ for inactive states) and denotes each AND logic operation with a composite node (e.g. $\sim x_{01} \& \sim x_{02}$). In this expanded network representation, each positive FBL can have at least two structures with distinct node values (e.g. $\sim x_{01}$ and x_{01} for an auto-positve FBL, $x_{01} \rightarrow x_{01}$). Since it can represent FBL status, this structure is referred to as a feedback motif, similar to the stable motif described in previous studies [18, 37] (see Text S1A for the detailed execution process of identifying feedback motifs in the expanded network).

In Stage 2, in the expanded network without any perturbation, two stable states can be mapped to corresponding undesired (from A_u) and desired set of node values (from A_d) (Fig. 2B, left). In Boolean functions, canalization refers to the capability of determining the output node value through one of multiple regulatory input node values, called a canalizing input, irrespective of the other input node values [38]. CAESAR calculates the canalizing effects after controlling each control target (c_i) using the logical domain of influence (LDOI) [19] on the expanded network. The LDOI of a control target contains nodes and their corresponding values that are stabilized after the first logical update of the set of neighboring nodes. Through this, CAESAR identifies all node values that are sufficiently stabilized by fixing the control target (referred to as directly regulated nodes). For instance, the directly regulated nodes for x_{02} (+) are represented as hexagon shapes and six of those nodes (including x_{02} and x_{03}) have desired values (filled hexagon shapes) (Fig. 2B, right). After that, some nodes are determined in their values, but other nodes' values still need to be determined.

CAESAR next explores undetermined node values of FBLs in the residual expanded network, which is the expanded network remaining after perturbation of the control target. CAESAR initially postulates undetermined node values as the given initial state (A_u). Integrating this information with the concept of feedback motif, CAESAR evaluates the stability of each feedback motif by propagating the undetermined node values through their connected edges in the residual expanded network. In the expanded network, certain node values may distribute their logical influences and determine other node values, which allows CAESAR to include opposite node values under certain conditions. Thus, during the sequential update (through their connections in the expanded network) of the node values starting from a given initial state, both feedback motifs can be stable: one with an undesired value and the other with a desired value. In other words, the feedback motif can maintain the undesired or desired node value, respectively, under this condition. Based on this, CAESAR can obtain a set of feedback motifs (Algorithm 1 in Text S1B). For example, when investigating the stability of feedback motifs for FBL3, (x_{04},$\sim x_{09}$&x_{04},$\sim x_{07}$,$\sim x_{08}$,$\sim x_{09}$) is stable due to the regulation effect of the given initial state, and (x_{05},x_{05}&x_{07},x_{07},x_{08},x_{09}) is also stable due to the propagation of x_{05} node value (Fig. 2C, left; see Supplementary Materials for the detailed execution process of Algorithm 1).

Subsequently, CAESAR prioritizes feedback motifs by considering other feedback motifs (Algorithm 2 in Text S1B). For instance, (x_{05},x_{05}&x_{07},x_{07},x_{08},x_{09}) has a higher priority, because (x_{04},$\sim x_{09}$&x_{04},$\sim x_{07}$,$\sim x_{08}$,$\sim x_{09}$) can no longer be stabilized when x_{04} is turned off since $x_{05} \rightarrow \sim x_{04}$. In this step, CAESAR scans through all feedback motifs to check the priority consistency of feedback motifs. Thus, a set of prioritized feedback motifs are iteratively extended to include what are not found in Algorithm 2 (Algorithm 3 in Text S1B). When there is no need to repeat prioritizing or extending a set of prioritized feedback motifs following the criteria described in Text S1B, all node values are determined based on finally prioritized feedback motifs (Fig. 2C, right; see Supplementary Materials for the detailed execution processes of Algorithms 2 and 3).

If all feedback motifs for the diff-FBLs eventually correspond to the desired values, then this control target is selected as a master regulator for the transition. After following all steps of Stage 2 for all $c_i \in \mathbf{C}$, CAESAR suggests a set of master regulators, enabling state transitions from A_u to A_d by simultaneously and permanently fixing the value of control nodes. When there is no solution by controlling individual single node, CAESAR increases the number of nodes by one and examines all possible combinations that include the initial candidate to investigate if they induce the desired state transitions. This process continues until the desired state is achieved (see Text S1B). In this example network model, without considering combinations, CAESAR identifies x_{02} (+) as the master regulator for A_u to A_d and set its state to be active.

In Stage 3, CAESAR identifies the canalizing kernel with a critical structure that plays a significant role in understanding the fate-determining path when the identified master regulator is controlled. All nodes are characterized by their features according to the output from the Stage 2 (e.g., canalization effect region and Algorithms 1–3), and together constitute a map that represents a direct or indirect perturbation effect after the regulation of the master regulator (Fig. 2D, left). Next, CAESAR classifies the feedback motifs into three subtypes: (i) cell fate-determining master regulator, characterized by a set of master regulators, (ii) a motif connection node, which can be fixed only by a given initial state, and (iii) a canalizing feedback motif, which corresponds to a feedback motif of diff-FBL(s) with desired values. In particular, for each canalizing feedback motif, the hierarchically determined sequence (1 through 5) is calculated by averaging the perturbation effect of nodes with a specific threshold (Fig. 2D, middle). Consequently, CAESAR can identify the canalizing kernel which is the largest connected component in the subgraph composed of canalizing feedback motifs and motif connection nodes (Fig. 2D, right; Algorithm 4 in Text S1B). The canalizing kernel provides a systematic understanding on how interconnected FBLs guide the fate-determining path for desired cellular states when the master regulator is controlled. More detailed procedures of

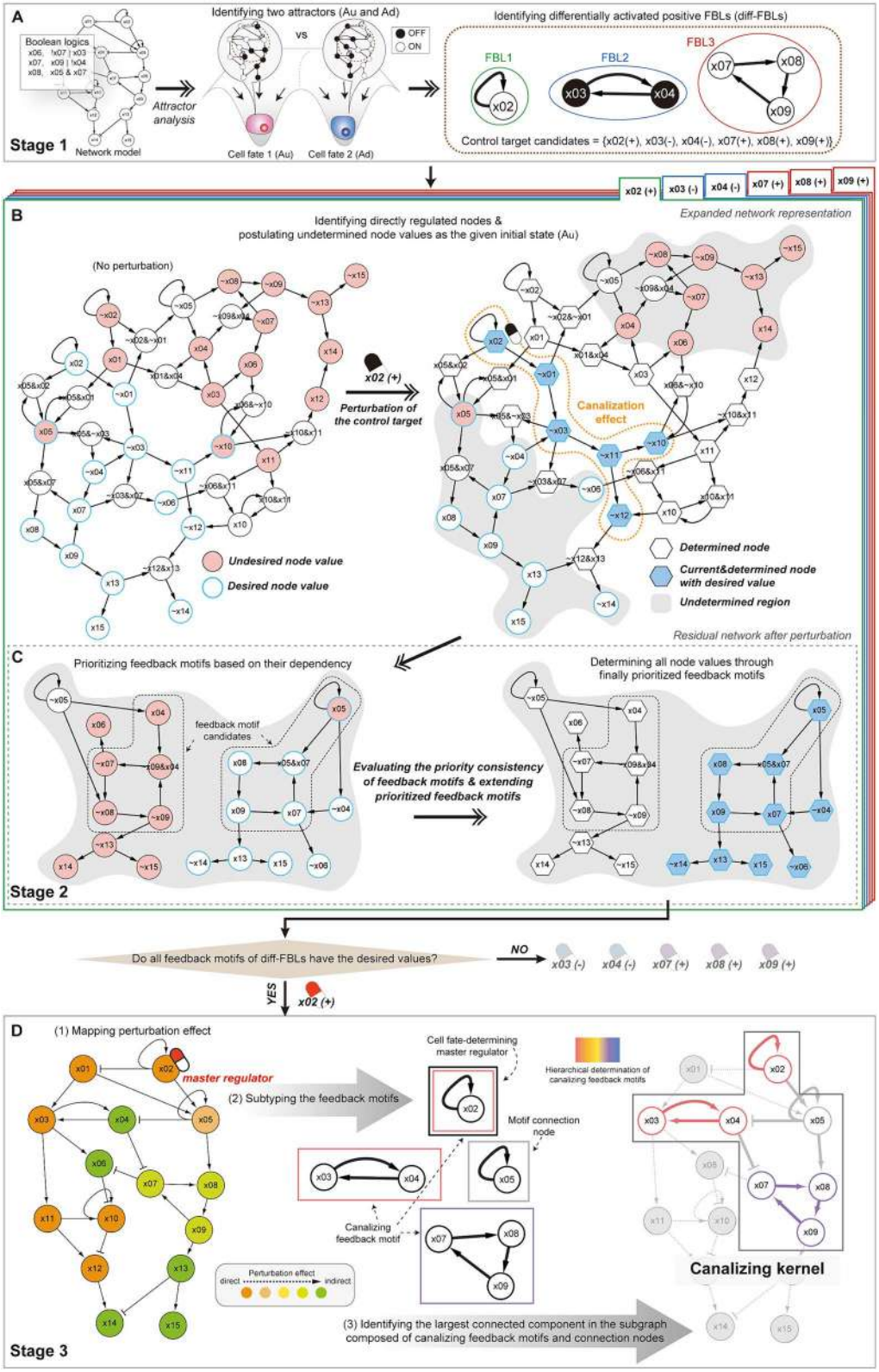


Figure 2. Systematic identification of master regulators and their canalizing kernels based on dynamic regulation of feedback motifs. (A) Identification of control target candidates; in stage 1, CAESAR designates A_u and A_d, and computes diff-FBLs. (B and C) Detailed steps for identifying master regulators; in stage 2, for each control target, CAESAR initially identifies directly regulated nodes (B), and then, from the residual network of the corresponding undetermined region, the prioritized feedback motifs are finally determined based on their dependency (C). (D) Identification of the canalizing kernel through the following steps: (1) mapping the perturbation effects from the results of Fig. 2B and C; (2) categorizing feedback motifs into three subtypes: cell fate-determining master regulator, canalizing feedback motif, and motif connection node; (3) identifying the largest connected component in the subgraph as the canalizing kernel. Nodes in each feedback motif are distinguished according to the hierarchically determined sequence by different colors and lines.

CAESAR, including Algorithms 1–4, are explained in Figs S1–6 and Text S1A-D.

Analysis of attractor transitions

An attractor is a stable state within the state space of the underlying dynamic system, representing either a single point or a limit cycle of points [28]. We only focused on the transitions between point attractors. To designate undesired and desired attractors, all point attractors were computed using compute_steady_states function from PyBoolNet [39] with Python 3.7.12. To guarantee that the simulated network converges to the predefined desired attractor after controlling the master regulator starting from the given initial (undesired) attractor, we simulated knockout (−) or overexpression (+) perturbations by fixing the values of nodes to OFF or ON, respectively, using the asynchronous update scheme. Under asynchronous update scheme, only a single node is updated per one transition, chosen randomly or according to a specific order [40]. We identified reachable attractor states from the initial attractor using find_attractor_state_by_randomwalk_and_ctl function from PyBoolNet with 100 repetitions.

Synthetic network analysis

To examine whether CAESAR can identify nearly optimal master regulators in complex networks, we constructed 100 different synthetic Boolean network models which have different network sizes ($n = 10$, $n = 20$, $n = 30$, $n = 50$, $n = 100$, with 20 models for each size) using BNGenerator function from https://github.com/choonlog/BNSimpleReduction. BNGenerator generates random directed graph, and each node has a Boolean logic from the biological networks in Cell Collective with power law in-degree distribution. To analyze these synthetic networks, we only selected one pair of attractors which have a median hamming distance in each network.

Biological network analysis

To examine whether CAESAR can identify nearly optimal master regulators in complex networks, we collected 34 biological Boolean networks from Cell Collective (https://cellcollective.org/), and 8 Boolean networks from the previously published literature. All the details are in Table S1. For each network, we randomly selected attractor pairs having a wide range of hamming distances. In particular, to confirm whether CAESAR can unravel the fundamental principles of cell fate transitions, we employed five biological Boolean network models in this study. First, a mESC network reported by Yachie-Kinoshita *et al.* [41] includes core pluripotency-associated genes. This network recapitulates the behavior of mESCs and epiblast stem cells (EpiSCs) under various input conditions, like the presence of bFGF and Activin A, by defining distinct attractor states. Second, another mESC network model by Dunn *et al.* [42] explores the gene regulations responsible for maintaining and inducing naive pluripotency in mESCs. The model can predict experimental observations, focusing on transcription factors such as Oct4, Sox2, and Nanog. Third, a B-cell network by Méndez *et al.* [43] recapitulates B-cell terminal differentiation based on transcription factors such as Pax5, Bcl6, Bach2, Irf4, and Blimp1. The model simulates differentiation into specific phenotypes (Naive, GC, Mem, and PC cells) and accurately predicts the effects of genetic mutations. Fourth, a EMT regulatory network model by Kim *et al.* [25] includes distinct molecules related to key functions during EMT process. The model describes the mesenchymal and epithelial states by their corresponding node activities, and also characterizes a chemosensitive epithelial state. It incorporates computational predictions with a FBL-related method and experimental validations. Lastly, a HSC-network reported by Collombet *et al.* [44] highlights the key regulatory interactions involved in lymphoid and myeloid cell specification and trans-differentiation. This model successfully predicts the impact of various genetic perturbations on cell fate decisions, well in accord with experimental observations.

Comparison of Boolean control frameworks

We compared the performance of CAESAR with four other Boolean control frameworks using synthetic and biological Boolean networks. To find a set of master regulators using LDOI and stable motif (SM) controls, we used the following functions: drivers.minimal_drivers for brute-force (BF) search, drivers.GRASP for greedy randomized adaptive search procedure (GRASP) search, and AttractorRepertorie.from_primes and succession_diagram.reprogram_to_trap_spaces for SM control from https://github.com/jcrozum/pystablemotifs. To find a set of master regulators using OT control, we used the cabean package [45] in Python. Lastly, we found a set of master regulators for DEPC control through the following steps: (1) extracting independent strongly connected components (SCCs) by transforming the network with DEPCs, as outlined in the study [27], (2) finding node sets with the highest out-degree for each SCC. All performance tests were conducted using the same computer specification.

Analysis of gene expression profiles

To compare the regulating effects between the predicted and experimental results, we used publicly available GEO datasets to analyze the relative expression levels when cellular states represent specific phenotypes (GSE62155) [46] or cell reprogramming is induced by a specific perturbation (GSE17316) [47]. We made the data matrix for each sample from GSMs by matching the probe sets in the given platform using the GEOquery package in R. We finally acquired the log2-transformed gene expression profiles for EpiSCs ($n = 3$) versus ESCs ($n = 3$) and perturbed C10 cells (each $n = 2$ for 5 time points). For EpiSC versus ESC, we computed the changes of the indicated genes by dividing their expression levels with the average values and then performed *t*-test on the mean values of the two states using scipy.stats.ttest_ind function in Python. From the result in Table S2, we determined OFF or ON of each gene depending on the relative values of the two states. For perturbed cells with Cebpa overexpression, the varying genes were first filtered on the basis of variance and diff-value, obtained by subtracting the minimum value from the maximum value, of each gene expression (we used 0.5 and 1.5 as the threshold of variance and diff-value, respectively). In the next, we simply employed the mean value of data as the threshold of each gene. Using the threshold, each gene was assigned a binarized value (0 for OFF and 1 for ON) over time (Table S3).

Data availability

Publicly available datasets (GSE62155 and GSE17316) we used in this study, were obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62155; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17316). CAESAR is available at: https://github.com/namheee/CAESAR.

Results

Overview of CAESAR

CAESAR takes a Boolean model, and the desired and undesired attractor states of the model as inputs. CAESAR analyzes the FBLs and identifies the control nodes (referred to as master regulators) that induce the transition from the undesired to the desired attractor. In addition, CAESAR decodes the underlying molecular mechanism for cell fate change by identifying the sequential connection of key FBLs (referred to as the canalizing kernel). To do this, CAESAR integrates the following four steps: (i) Identification of control target candidates for inducing a state transition from the undesired to the desired attractor (Stage 1). CAESAR first characterizes differentially activated positive FBLs (diff-FBLs) containing any node with a different value between undesired and desired attractors. All nodes within the diff-FBLs are designated as control target candidates. (ii) Prioritization of FBL status for each control target candidate (Stage 2). CAESAR uses the expanded network to analyze regulation effects after perturbation of a control target. When the value of the control target is fixed, it can determine the status of certain diff-FBLs. In the expanded network, CAESAR calculates the canalizing effects and scans through all FBL status using the information of the given initial state. Their priority is determined by the sequence in which their status are determined through the logical relationships among the diff-FBLs. By evaluating the priority of FBL status, CAESAR determines finally prioritized stable FBL status. (iii) Identification of master regulators using the final prioritized stable FBL status (Stage 2). If all values of diff-FBLs correspond to the desired values, then CAESAR identifies that control target candidate as a master regulator for the state transition. (iv) Identification of the canalizing kernel for the master regulator (Stage 3). CAESAR calculates the perturbation effects and identifies the corresponding essential subnetwork, canalizing kernel. Please see the Materials and Methods for the detailed procedures of CAESAR.

Identifying a small number of master regulators with CAESAR

Previous studies have shown that there is no need to control many molecular targets for cell fate changes [5–8, 48]. To confirm that only a few targets suffice to control a state transition, we identified master regulators in 100 synthetic networks with CAESAR. We found that the desired attractor can be reached by controlling only a single target in 74 out of 100 synthetic networks (Fig. 3A, left; Materials and Methods). In addition, the number of master regulators was relatively small compared to the network size (the averaged fraction of master regulator nodes in synthetic networks is 0.15) (Fig. 3A, right). To examine whether CAESAR can identify nearly optimal master regulators (i.e., small number of molecular targets) in complex networks, we compared the performance of CAESAR with four other control frameworks including one-step temporary perturbation (OT) [20], differentially expressed positive circuit (DEPC) [27], logical domain of influence (LDOI) [19], and stable motif (SM) [18] (Fig. 3B; Materials and Methods). CAESAR, OT, and DEPC could successfully find a minimal number of master regulators for relatively small networks in most cases (Fig. 3B and C; Fig. S7A). However, for large biological networks, CAESAR was able to find smaller sets of master regulators compared to OT and DEPC (Fig. 3C). In addition, DEPC showed the best time complexity, but was less accurate ($0.67 \sim 0.9$) than other frameworks (Fig. S7B and S7C). Altogether, these results suggest that CAESAR is effective in identifying a minimal number of biologically feasible master regulators in complex networks.

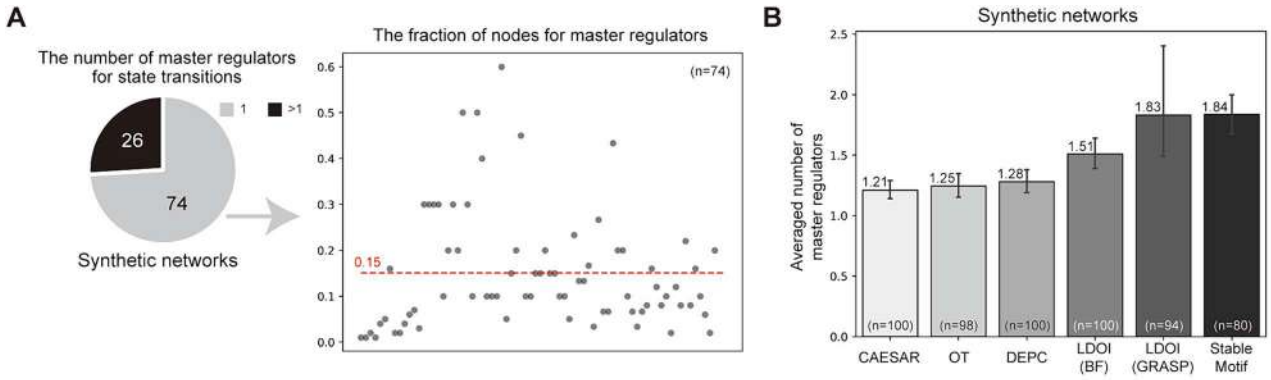
Canalizing kernel for reprogramming EpiSCs to ESCs

We applied CAESAR to real biological network models to investigate the underlying mechanism during cellular reprogramming. First, we utilized the Boolean network model reported by Yachie-Kinoshita *et al*. [41]. This network recapitulates the transition in pluripotent stem cells like mESCs and EpiSCs by using pluripotency-associated genes such as Oct4, Sox2, Nanog, Klf4, and c-Myc. We initially analyzed attractors given the specific input conditions of activated bFGF and Activin A, as reported in the study [41]. After defining EpiSC and ESC phenotypes in terms of EpiTFs (a single component composed of EpiSC-enriched factors such as Fgf5, Eomes, Otx2, and Brachyury), Pecam1, and Rex1 genes, we focused on two transitions (transition-1 and transition-2) across different EpiSC and ESC attractors (Fig. 4A).

CAESAR identified any one of Klf2, Klf4, Nanog, Oct4, or Sox2 as a single master regulator of transition-1, and any one of Klf4, Oct4, or Nanog as a single master regulator of transition-2 (Fig. 4A, right). Specifically, Oct4, Sox2, and Klf4 correspond to the Yamanaka factors (OSKM) [3], and Oct4, Sox2, and Nanog (OSN) are also well-known pioneer factors for maintenance of pluripotency [49]. In contrast, DEPC identified Nanog alone, and LDOI identified three combinations of Cdx2, Oct4, and SignalLIF or Cdx2, Klf4, and Oct4, which are ineffective sets of master regulators, compared to those identified by CAESAR (Fig. 4A, right). These results demonstrate that CAESAR can identify a minimal sufficient number of previously established master regulators in the complex biological networks, surpassing other methods in its efficacy.

In particular, it was previously shown by experiments that Klf4 plays a significant role in reprogramming as it engages in direct interactions with Oct4 and Sox2 [50]. Thus, we identified the canalizing kernels for Klf4 upregulation and compared them with experimental results. For this purpose, we analyzed the regulation of Klf4 for de-differentiation from two different EpiSC states. Depending on the transition pair, each canalizing kernel included 64% of nodes (transition-1) or 43% of nodes (transition-2), respectively (Fig. 4B). Interestingly, a nine-node subnetwork exhibited shared functional structure including canalizing feedback motifs in both transitions, and was hence designated as the common canalizing kernel (Fig. 4C, left). We found that this subnetwork is also similar to the canalizing kernel when Nanog, another master regulator, is upregulated (Fig. S8). CAESAR shows the common involvement of a specific kernel that regulates the underlying mechanisms in the transitions of pluripotent stem cells.

To delineate the mechanism underlying the transition of cell states from EpiSCs to ESCs, we further analyzed the common canalizing kernel (Fig. 4C, right). Klf4 upregulation led to a change in the node values of Dnmt3b, Klf2, Esrrb, and EpiTFs, all of which share canalizing feedback motifs with Klf4. Subsequently, this change resulted in the upregulation of OSN levels, ultimately determining the ESC phenotype (Fig. 4C). To enrich our understanding of the mechanism predicted through the canalizing kernel, we further explored the gene expression profile of EpiSC and ESC states, sourced from GSE62155 [46] (Materials and Methods; see Table S2 for the analysis of regulation patterns). From this, we found that Esrrb and Klf2 are significantly activated in a pattern similar to that of Klf4. OSN levels are slightly activated but exhibit patterns similar to previously reported homogeneous expression patterns in the ESC phenotype [41] (Fig. 4D). These findings collectively indicate that our canalizing kernel can successfully capture the dynamics of co-regulating genes that are



C

The number of master regulators for state transitions (Biological networks)

	Network index	N01	N02	N03	N04	N05	N06	N07	N08	N09	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20	N21
	Network size	5	5	10	11	12	13	13	15	16	16	16	16	17	18	19	20	22	22	23	23	24
Control frameworks	CAESAR	1.67	1	1	1	2	1	1.67	1.38	1	1.1	1	3	1	2	3	1	1	1	2	2	2
	One-step source-target control (OT)	1.67	1	1	1	*NA*	1	1.67	1.38	1	1.1	1	3	1	2	2	1.25	1	1	2	2	2
	DEPC control	3	1	1	1	3	1.43	2.67	1.5	1.33	7.4	4	7.5	2.43	2	2	1	1.2	1.25	2	2	2
	LDOI target control (BF)	1.67	1	1	1	2	1.36	3	1.63	2	2.7	3.67	4	2.29	2	3	1.25	1.2	1.44	2	2	2
	LDOI target control (GRASP)	1	1	1.43	1	2	1.36	3	1.88	3	3.13	4	4	2.29	2	3	1.33	1.6	2.06	2	2	2
	Stable Motif control	1.67	1	2.71	2	4	3	4	2.63	3.67	*NA*	*NA*	*NA*	3	8	*NA*	1.5	1.8	*NA*	*NA*	5	6

	Network index	N22	N23	N24	N25	N26	N27	N28	N29	N30	N31	N32	N33	N34	N35	N36	N37	N38	N39	N40	N41	N42
	Network size	25	25	26	26	28	31	32	33	34	35	36	40	44	44	49	51	53	53	68	74	105
Control frameworks	CAESAR	2	2	1	1.33	1.05	2	1.5	1.5	2	2.33	2	1	1.38	1.33	1	1.33	2.5	2	1.25	1.14	1
	One-step source-target control (OT)	2	2	1	*NA*	1.42	*NA*	1.5	1.5	*NA*	*NA*	1	*NA*	*NA*	1.33	1.33	*NA*	2.25	*NA*	1.25	*NA*	*NA*
	DEPC control	8.25	6	1	1.33	1.21	1	1.5	2	2	3	1	1	1.38	2	1.33	3	2.25	3.5	2.25	1.14	1
	LDOI target control (BF)	4.5	6	1	1.33	1.58	2	2	1.5	2	3	2.67	1	3.5	2	1	1.33	2.5	2	3.25	1.14	1
	LDOI target control (GRASP)	4	*NA*	1	1.67	1.76	3	1	1	2	3	5	1	14.8	2	1.17	5.33	2	3.5	*NA*	1.14	1
	Stable Motif control	*NA*	*NA*	2	1.67	4.05	*NA*	4	2	2	4	3.33	3	*NA*	4.67	5.17	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

Figure 3. Comparison of CAESAR with other control frameworks. (A) The number of master regulators required for a desired state transition: a pie chart indicates the number of synthetic networks that enable changes in stable states using a single target (left); the fraction of master regulators was plotted in scatter plot (right). (B) Comparison of CAESAR with different control frameworks (CAESAR, OT, DEPC, LDOI (BF), LDOI (GRASP), and SM) using synthetic networks. Each error bar represents a 95% confidence interval, and the number of tested networks is indicated by *n*. (C) Comparison of CAESAR with different control frameworks using 42 biological networks. If any of the frameworks does not find master regulators, it is denoted as *NA*.

essential to cell fate change, enhancing our understanding of the underlying mechanisms of pluripotent stem cell transitions.

Further investigation of canalizing kernels in other networks

To further examine the underlying mechanism of the EpiSC-ESC transition, we applied CAESAR to another mESC network reported by Dunn *et al*. [42]. In this network, CAESAR identified any one of Esrrb, Klf2, Klf4, Nanog, Sall4, Sox2, or Tbx3 as a single master regulator of this transition, consistent with the reported experimental results [42, 51] (Fig. S9A). As shown in Fig. 4, we could also explore significant sequential gene regulation patterns between Klf4 and Sox2, when Esrrb (Fig. S9B) or Klf4 (Fig. S9C) was upregulated, respectively. These regulatory sequences are in accord with the regulation relationships revealed by our canalizing kernel (see Text S2A for further comparison with experimental evidence).

We applied CAESAR to the B-cell network [43] and identified either one of Blimp1 or Pax5 as a single master regulator for differentiation to plasma cells (Fig. S10A). Interestingly, CAESAR could reveal the underlying principle of the transition from germinal center B-cells to plasma cells, in accordance with previous experimental results on the differentiation mechanisms governed by Bcl6 or Irf4 [52–54] (see Fig. S10B and Text S2B for further comparison with experimental evidence).

We also applied CAESAR to the EMT network [25]. We considered the scenario where TGFβ stimulation is ON, the epithelial state (E-cad+/ZEB1-) is the desired attractor, and the mesenchymal state (E-cad-/ZEB1+) is the undesired attractor. In this case, CAESAR identified one of miR200(+)&SMAD4(−), miR200(+)&Snail(−), or p53(+)&Snail(−) as a master regulator for the trans-differentiation of mesenchymal to epithelial state (Fig. S11A). In addition, CAESAR recapitulates the mechanism of connected positive FBLs, a two-node FBL between miR200 and ZEB1 and another FBL between miR34 and Snail, as also supported by the previous experimental result [25] (Fig. S11B).

Taken together, these analyses demonstrate the capacity for the canalizing kernel to provide insight into the dynamic sequence of gene regulations during cell fate transitions.

Canalizing kernel for trans-differentiation of B-cells into macrophages

Lastly, we studied the HSC network reported by Collombet *et al*. [44]. It describes B-cells and macrophages (Macs) specification from multipotent progenitors (MPs), depending on the stimulation of input signals (Csf1 and Il7). This network includes Spi1 (PU.1) which is represented by a ternary variable (taking the values

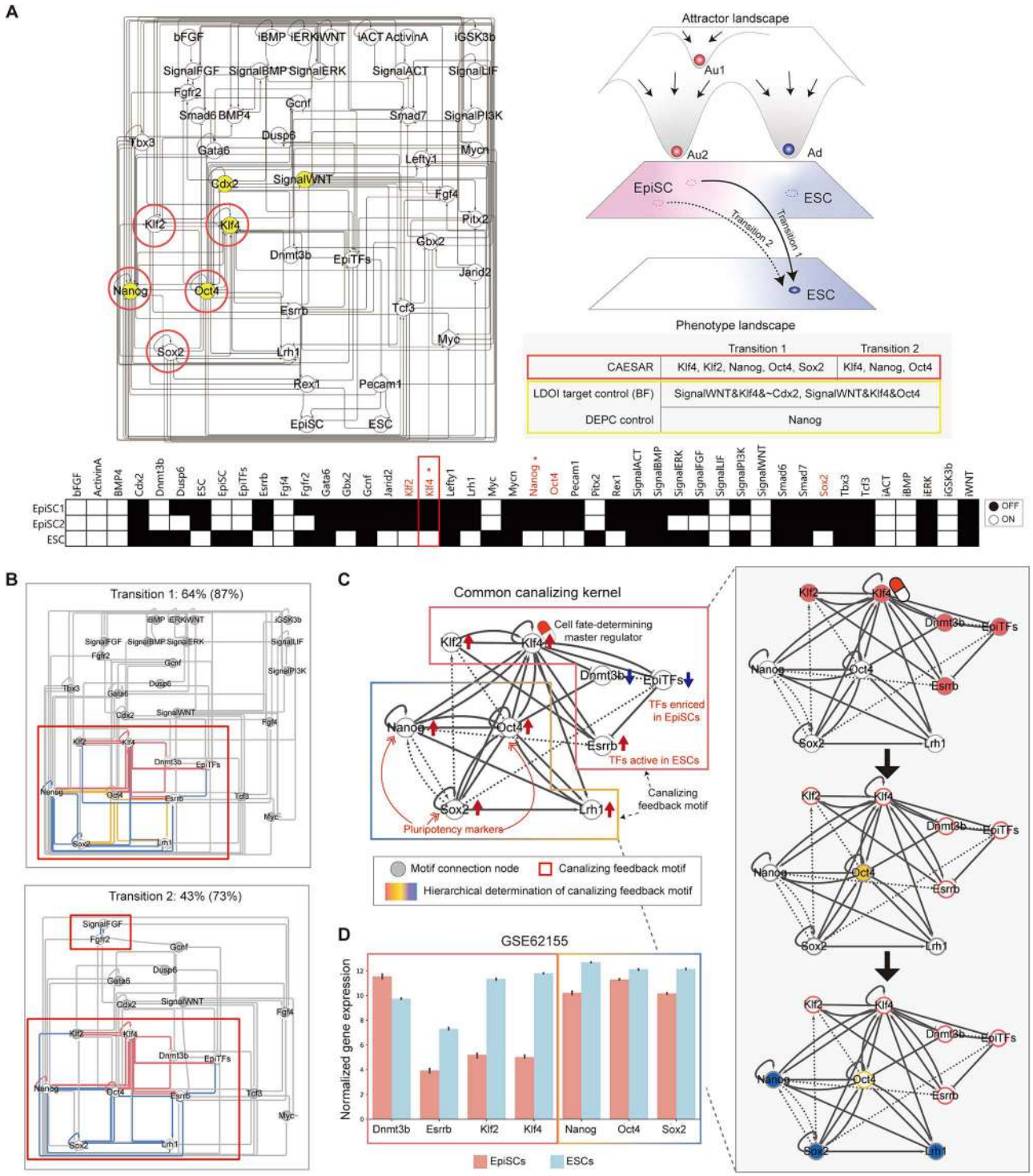


Figure 4. Canalizing kernel for cellular reprogramming in the mESC network. (A) Applying CAESAR to the mESC network; the attractor landscape illustrates two undesired attractors for EpiSC and one desired attractor for ESC. If a node was mentioned in the original study, its label is denoted with an asterisk in the bottom panel. (B) Canalizing kernel for Klf4 upregulation. With CAESAR, canalizing kernels for transition-1 (top) and transition-2 (bottom) are represented. (C) Common canalizing kernel for transitions from EpiSCs to ESCs based on the results in Fig. 4B (left) and the hierarchically determined sequence when Klf4 is upregulated (right). (D) Gene expression profiles for EpiSCs and ESCs.

0, 1, or 2). B-cells and Macs are characterized by well-known phenotype markers, Cd19 and Mac1 (also called Cd11b). With CAESAR, in the presence of Csf1 and Il7, we predicted a single regulation of any of Cebpa, Cebpb, E2a, Foxo1, or Id2, and their related nodes as a master regulator for trans-differentiation of B-cells to Macs. These predictions are consistent with previously reported experimental results [44, 55] (Fig. 5A).

Based on the above result, we focused on a well-studied reprogramming factor, Cebpa, and identified the canalizing kernel for Cebpa upregulation (Fig. 5B). Notably, with an upregulation of

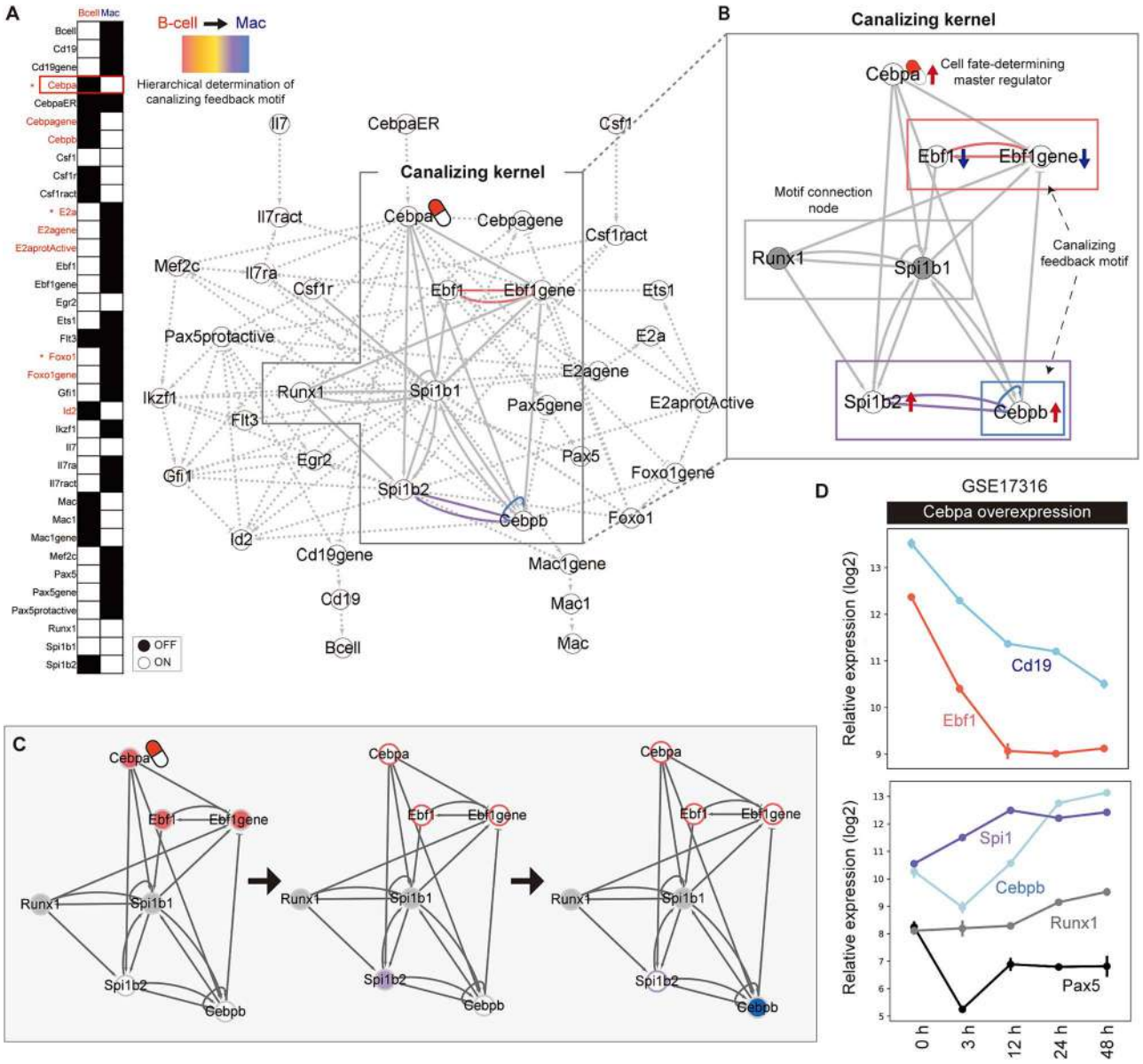


Figure 5. Canalizing kernel for trans-differentiation in the HSC network. (A) Applying CAESAR to the HSC network. All edges except the canalizing kernel are indicated by dashed lines, and edges in the canalizing feedback motifs are distinguished by different colors according to the hierarchically determined sequence. If a node was previously investigated in the original study, its label is denoted with an asterisk in the left panel. (B) Canalizing kernel for Cebpa upregulation based on the regulations outlined in Fig. 5A. (C) The hierarchically determined sequence when Cebpa is upregulated. (D) Time-course gene expression profile under Cebpa overexpression.

Cebpa, the canalizing feedback motifs were sequentially regulated by a series of Ebf1-Ebf1 gene, Spi1_b2-Cebpb, and Cebpb (Fig. 5C). To show that their regulation sequence is similar to the patterns that appear in the time-course gene expression profile with Cebpa overexpression, we analyzed the dataset obtained from GSE17316 [47] (Fig. 5D) and further compared the binarized value of each gene over time (Table S3; Materials and Methods). We found that B-cell related factors including Ebf1 and Cd19 are rapidly downregulated over time. In addition, as Runx1 was characterized as a motif connection node in the canalizing kernel, its expression level was similar to those in both B-cells and Macs, in accordance with a previous experimental study [47]. Moreover, Cebpb and Spi1 were monotonically increased or consistently sustained over time (Fig. 5D). Collectively, these findings suggest that our canalizing kernel can provide a system-level understanding of how genes are sequentially interconnected and collaboratively regulated to determine the final cellular state, rather than functioning independently.

Discussion

There still remain unsolved questions on which specific molecular regulators can induce cell fate changes and what the underlying mechanism for such cell fate determination is. While many experimental observations have implied the existence of a limited number of master regulators and their core regulation structures, systematic identification of those regulators and their underlying mechanisms remain a challenge. To answer these questions, we developed CAESAR, which can identify master regulators responsible for critical transitions between undesired

and desired cellular states. CAESAR can also unravel core FBL regulations which we named the canalizing kernel. The canalizing kernel elucidates the sequential regulations along the fate-determining path, highlighting how cellular states change when the master regulators are controlled.

In this study, a Boolean network model was employed, which can reproduce the nonlinear characteristics of cells with fewer parameters. This model is useful to capture emergent properties of living systems such as canalization, robustness, and evolvability [40, 56]. In particular, canalization is an evolutionarily established property for robustness against noise or uncertain fluctuations. A canalizing function determines an output value based on a specific canalizing variable, regardless of other values [57]. Furthermore, many previous studies have underscored the role of positive FBLs in cellular reprogramming or trans-differentiation [23–25, 27]. Based on these concepts, CAESAR identifies the canalizing kernel composed of interconnected positive FBLs, which is essential for deciphering the fate-determining path for a desired cell state.

By identifying canalizing kernels, we could unveil the underlying mechanisms in cellular reprogramming and trans-differentiation of previous experimental discoveries. For instance, in ESC networks, CAESAR identified any of Klf4, Nanog, or Esrrb as a single master regulator, all of which were previously found as primary regulators of pluripotency through experiments [49]. Moreover, our canalizing kernel highlighted the FBL composed of Klf2 and Klf4 as a crucial player in determining the ESC phenotype. Along with other Yamanaka factors, Klf4 shares extensive gene co-occupancy with Oct4 and Sox2 [50]. In addition, Klf4 is an upstream regulator of a broad feed-forward loop, encompassing Oct4, Sox2, c-Myc, and other pluripotency factors like Nanog [58]. These previous experimental results further corroborate our findings on the system-level influence of Klf4 in cellular reprogramming.

In the HSC network, we identified Cebpa as a master regulator, which can induce macrophage trans-differentiation. In the previous study, Zhang *et al.* [59] mapped mathematically optimal state transition trajectories from energy landscapes and suggested related molecular interactions. On the other hand, we elucidated the underlying mechanisms which dominantly enable transitions by analyzing cooperative state stabilization dynamics with the key positive FBLs and the corresponding master regulators. Through the canalizing kernel for Cebpa upregulation, we found that molecules such as Cebpb and Spi1 do not independently respond to stimuli. Instead, they are interconnected and collaboratively drive trans-differentiation. In spite of ongoing debates regarding the deterministic versus stochastic processes of pluripotency acquisition or reprogramming [42], our study indicates that key components of cells cooperatively determine a dominant path for the desired phenotype.

Although CAESAR provides detailed insights into molecular mechanisms for cell fate determination, a method centered on attractor landscape topography [59–61] can offer valuable complements. This approach used the minimum action path theory to find the most probable transition paths between the attractors capturing a global picture of the continuous dynamics. This landscape topography-focused approach could explain brain state transition process in working memory [61] and cell state transitions in the EMT-metastasis system [60]. Integrating our approach with analysis of attractor landscape topography could be a promising direction for future research.

CAESAR addresses the problem of an attractor-to-attractor transition rather than global stabilization [18, 19, 62] and shows better performance compared to other global stabilization approaches (Fig. 3). However, CAESAR has also some limitations, particularly concerning its high time complexity and difficulty in distinguishing stable feedback motifs that arise from oscillatory logical relationships or differences in update order (see Text S1D for details). Nonetheless, CAESAR can successfully identify a minimal number of master regulators by excluding any unnecessary control targets. Moreover, CAESAR can reveal the canalizing kernels that govern state transitions after controlling the master regulators, a capability not demonstrated by any other methods, even for large networks.

Owing to the rapid explosion of high-throughput omics data, especially single-cell omics data, the scope and detail of our understanding of biological processes are being greatly expanded. In line with this advance, many useful methods for regulatory network inference, such as Single-Cell Network Synthesis [63], CellOracle [64], scTenifoldKnk [65], and SCENIC+ [66], have been suggested. We anticipate that by employing these methods, the applicability of CAESAR to predict master regulators and delineate fate-determining paths at a system-level will be significantly enhanced, even for previously uncharacterized cellular processes such as tumorigenesis and aging [67, 68].

In summary, CAESAR provides a general framework for identifying both the master regulator and its canalizing kernel, broadening our understanding of how intricate molecular networks determine cell fates upon regulation of specific targets.

Key Points

- The unresolved problem lies in identifying master regulators and their key pathways for inducing cell fate changes despite the epigenetic barrier within Waddington's epigenetic landscape.
- The study develops CAESAR, a computational control framework, which can systematically identify master regulators and their resulting fate-determining path (referred to as "canalizing kernel") governing cell fate changes when the master regulators are controlled.
- By applying CAESAR, the study successfully identifies the master regulators and their canalizing kernels, aligning well with previous experimental studies of cellular reprogramming or trans-differentiation.

Acknowledgements

The authors thank Corbin Hopper for his critical reading and comments.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Science and ICT (2023R1A2C3002619 and RS-2024-00405360).

Conflict of interest: None declared.

Data availability

Publicly available datasets (GSE62155 and GSE17316) we used in this study, were obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62155; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17316). CAESAR is available at: https://github.com/namheee/CAESAR.

Author contributions

K.-H.C. conceived the idea. N.K. and K.-H.C. designed computational works. N.K. conduced computational simulations and developed the software tool. J.L., J.K., and Y.K. provided analytic and technical supports. K.-H.C. and N.K. wrote the manuscript. K.-H.C. designed the project and supervised the research.

Declaration of interests statement

The authors declare no competing financial interests.

References

1. Alon U, Surette MG, Barkai N. *et al.* Robustness in bacterial chemotaxis. *Nature* 1999;**397**:168–71. https://doi.org/10.1038/16483.
2. Lauffenburger DA. Cell signaling pathways as control modules: complexity for simplicity? *Proc Natl Acad Sci U S A* 2000;**97**:5031–3. https://doi.org/10.1073/pnas.97.10.5031.
3. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006;**126**:663–76. https://doi.org/10.1016/j.cell.2006.07.024.
4. Takahashi K, Tanabe K, Ohnuki M. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007;**131**:861–72. https://doi.org/10.1016/j.cell.2007.11.019.
5. Dow LE, O'Rourke KP, Simon J. *et al.* Apc restoration promotes cellular differentiation and Reestablishes crypt homeostasis in colorectal cancer. *Cell* 2015;**161**:1539–52. https://doi.org/10.1016/j.cell.2015.05.033.
6. Panciera T, Azzolin L, Fujimura A. *et al.* Induction of expandable tissue-specific stem/progenitor cells through transient expression of YAP/TAZ. *Cell Stem Cell* 2016;**19**:725–37. https://doi.org/10.1016/j.stem.2016.08.009.
7. An S, Cho SY, Kang J. *et al.* Inhibition of 3-phosphoinositide-dependent protein kinase 1 (PDK1) can revert cellular senescence in human dermal fibroblasts. *Proc Natl Acad Sci U S A* 2020;**117**:31535–46. https://doi.org/10.1073/pnas.1920338117.
8. Lee S, Lee C, Hwang CY. *et al.* Network inference analysis identifies SETDB1 as a key regulator for reverting colorectal cancer cells into differentiated normal-like cells. *Mol Cancer Res* 2020;**18**:118–29. https://doi.org/10.1158/1541-7786.MCR-19-0450.
9. Choi SR, Hwang CY, Lee J. *et al.* Network analysis identifies regulators of basal-like breast cancer reprogramming and endocrine therapy vulnerability. *Cancer Res* 2022;**82**:320–33. https://doi.org/10.1158/0008-5472.CAN-21-0621.
10. Waddington CH. Canalization of development and genetic assimilation of acquired characters. *Nature* 1959;**183**:1654–5. https://doi.org/10.1038/1831654a0.
11. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell* 2007;**128**:635–8. https://doi.org/10.1016/j.cell.2007.02.006.
12. Bhattacharya S, Zhang Q, Andersen ME. A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst Biol* 2011;**5**:85. https://doi.org/10.1186/1752-0509-5-85.
13. Enver T, Pera M, Peterson C. *et al.* Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* 2009;**4**:387–97. https://doi.org/10.1016/j.stem.2009.04.011.
14. Park SG, Lee T, Kang HY. *et al.* The influence of the signal dynamics of activated form of IKK on NF-kappaB and anti-apoptotic gene expressions: a systems biology approach. *FEBS Lett* 2006;**580**:822–30. https://doi.org/10.1016/j.febslet.2006.01.004.
15. Sreenath SN, Cho KH, Wellstead P. Modelling the dynamics of signalling pathways. *Essays Biochem* 2008;**45**:1–28. https://doi.org/10.1042/BSE0450001.
16. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 1969;**22**:437–67. https://doi.org/10.1016/0022-5193(69)90015-0.
17. Kauffman S. Homeostasis and differentiation in random genetic control networks. *Nature* 1969;**224**:177–8. https://doi.org/10.1038/224177a0.
18. Zanudo JG, Albert R. Cell fate reprogramming by control of intracellular network dynamics. *PLoS Comput Biol* 2015;**11**:e1004193. https://doi.org/10.1371/journal.pcbi.1004193.
19. Yang G, Zanudo JGT, Albert R. Target control in logical models using the domain of influence of nodes. *Front Physiol* 2018;**9**:454. https://doi.org/10.3389/fphys.2018.00454.
20. Paul S, Su C, Pang J. *et al.* An efficient approach towards the source-target control of Boolean networks. *IEEE/ACM Trans Comput Biol Bioinform* 2020;**17**:1932–45. https://doi.org/10.1109/TCBB.2019.2915081.
21. Thomas R, Thieffry D, Kaufman M. Dynamical behaviour of biological regulatory networks—I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol* 1995;**57**:247–76. https://doi.org/10.1007/BF02460618.
22. Kwon YK, Cho KH. Analysis of feedback loops and robustness in network evolution based on Boolean models. *BMC Bioinformatics* 2007;**8**:430. https://doi.org/10.1186/1471-2105-8-430.
23. Seo CH, Kim JR, Kim MS. *et al.* Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. *Bioinformatics* 2009;**25**:1898–904. https://doi.org/10.1093/bioinformatics/btp316.
24. Yeo SY, Lee KW, Shin D. *et al.* A positive feedback loop bistably activates fibroblasts. *Nat Commun* 2018;**9**:3016. https://doi.org/10.1038/s41467-018-05274-6.
25. Kim N, Hwang CY, Kim T. *et al.* A cell-fate reprogramming strategy reverses epithelial-to-mesenchymal transition of lung cancer cells while avoiding hybrid states. *Cancer Res* 2023;**83**:956–70. https://doi.org/10.1158/0008-5472.CAN-22-1559.
26. Lee HS, Hwang CY, Shin SY. *et al.* MLK3 is part of a feedback mechanism that regulates different cellular responses to reactive oxygen species. *Sci Signal* 2014;**7**:ra52. https://doi.org/10.1126/scisignal.2005260.
27. Crespo I, Del Sol A. A general strategy for cellular reprogramming: the importance of transcription factor cross-repression. *Stem Cells* 2013;**31**:2127–35. https://doi.org/10.1002/stem.1473.
28. Wang RS, Saadatpour A, Albert R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol* 2012;**9**:055001. https://doi.org/10.1088/1478-3975/9/5/055001.
29. Choi M, Shi J, Jung SH. *et al.* Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. *Sci Signal* 2012;**5**:ra83. https://doi.org/10.1126/scisignal.2003363.
30. Choi M, Shi J, Zhu Y. *et al.* Network dynamics-based cancer panel stratification for systemic prediction of anticancer drug response. *Nat Commun* 2017;**8**:1940. https://doi.org/10.1038/s41467-017-02160-5.

31. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 1982;**79**:2554–8. https://doi.org/10.1073/pnas.79.8.2554.
32. Inagaki HK, Fontolan L, Romani S. *et al.* Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 2019;**566**:212–7. https://doi.org/10.1038/s41586-019-0919-7.
33. Cinquin O, Demongeot J. Roles of positive and negative feedback in biological systems. *C R Biol* 2002;**325**:1085–95. https://doi.org/10.1016/S1631-0691(02)01533-0.
34. Azpeitia E, Muñoz S, González-Tokman D. *et al.* The combination of the functionalities of feedback circuits is determinant for the attractors' number and size in pathway-like Boolean networks. *Sci Rep* 2017;**7**:42023. https://doi.org/10.1038/srep42023.
35. Zanudo JG, Albert R. An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos* 2013;**23**:025111. https://doi.org/10.1063/1.4809777.
36. Klarner H, Bockmayr A, Siebert H. Computing maximal and minimal trap spaces of Boolean networks. *Natural Computing* 2015;**14**:535–44. https://doi.org/10.1007/s11047-015-9520-7.
37. Deritei D, Rozum J, Ravasz Regan E. *et al.* A feedback loop of conditionally stable circuits drives the cell cycle from checkpoint to checkpoint. *Sci Rep* 2019;**9**:16430. https://doi.org/10.1038/s41598-019-52725-1.
38. Kauffman S. The large scale structure and dynamics of gene control circuits: an ensemble approach. *J Theor Biol* 1974;**44**:167–90. https://doi.org/10.1016/S0022-5193(74)80037-8.
39. Klarner H, Streck A, Siebert H. PyBoolNet: a python package for the generation, analysis and visualization of boolean networks. *Bioinformatics* 2017;**33**:770–2. https://doi.org/10.1093/bioinformatics/btw682.
40. Schwab JD, Kühlwein SD, Ikonomi N. *et al.* Concepts in Boolean network modeling: what do they all mean? *Comput Struct Biotechnol J* 2020;**18**:571–82. https://doi.org/10.1016/j.csbj.2020.03.001.
41. Yachie-Kinoshita A, Onishi K, Ostblom J. *et al.* Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. *Mol Syst Biol* 2018;**14**:e7952. https://doi.org/10.15252/msb.20177952.
42. Dunn SJ, Li MA, Carbognin E. *et al.* A common molecular logic determines embryonic stem cell self-renewal and reprogramming. *EMBO J* 2019;**38**:e100003. https://doi.org/10.15252/embj.2018100003.
43. Mendez A, Mendoza L. A network model to describe the terminal differentiation of B cells. *PLoS Comput Biol* 2016;**12**:e1004696. https://doi.org/10.1371/journal.pcbi.1004696.
44. Collombet S, van Oevelen C, Sardina Ortega JL. *et al.* Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *Proc Natl Acad Sci U S A* 2017;**114**:5792–9. https://doi.org/10.1073/pnas.1610622114.
45. Su, C. and J. Pang. *Cabean 2.0: Efficient and Efficacious Control of Asynchronous Boolean Networks*. in *Formal Methods: 24th International Symposium, FM 2021, Virtual Event, November 20–26, 2021, Proceedings* 24. 2021. Springer International Publishing:581–598.
46. Kurek D, Neagu A, Tastemel M. *et al.* Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports* 2015;**4**:114–28. https://doi.org/10.1016/j.stemcr.2014.11.007.
47. Bussmann LH, Schubert A, Vu Manh TP. *et al.* A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* 2009;**5**:554–66. https://doi.org/10.1016/j.stem.2009.10.004.
48. Kim J, Park SM, Cho KH. Discovery of a kernel for controlling biomolecular regulatory networks. *Sci Rep* 2013;**3**:2223. https://doi.org/10.1038/srep02223.
49. Li M, Belmonte JC. Ground rules of the pluripotency gene regulatory network. *Nat Rev Genet* 2017;**18**:180–91. https://doi.org/10.1038/nrg.2016.156.
50. Wei Z, Yang Y, Zhang P. *et al.* Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells* 2009;**27**:2969–78. https://doi.org/10.1002/stem.231.
51. Lee KC, Wong WK, Feng B. Decoding the pluripotency network: the emergence of new transcription factors. *Biomedicine* 2013;**1**:49–78. https://doi.org/10.3390/biomedicines1010049.
52. Yasuda T, Hayakawa F, Kurahashi S. *et al.* B cell receptor-ERK1/2 signal cancels PAX5-dependent repression of BLIMP1 through PAX5 phosphorylation: a mechanism of antigen-triggering plasma cell differentiation. *J Immunol* 2012;**188**:6127–34. https://doi.org/10.4049/jimmunol.1103039.
53. Muto A, Ochiai K, Kimura Y. *et al.* Bach2 represses plasma cell gene regulatory network in B cells to promote antibody class switch. *EMBO J* 2010;**29**:4048–61. https://doi.org/10.1038/emboj.2010.257.
54. Nera KP, Kohonen P, Narvi E. *et al.* Loss of Pax5 promotes plasma cell differentiation. *Immunity* 2006;**24**:283–93. https://doi.org/10.1016/j.immuni.2006.02.003.
55. Xie H, Ye M, Feng R. *et al.* Stepwise reprogramming of B cells into macrophages. *Cell* 2004;**117**:663–76. https://doi.org/10.1016/S0092-8674(04)00419-2.
56. Joo JI, Zhou JX, Huang S. *et al.* Determining relative dynamic stability of cell states using Boolean network model. *Sci Rep* 2018;**8**:12077. https://doi.org/10.1038/s41598-018-30544-0.
57. Paul E, Pogudin G, Qin W. *et al.* The dynamics of canalizing Boolean networks. *Complexity* 2020;**2020**:1–14. https://doi.org/10.1155/2020/3687961.
58. Nandan MO, Yang VW. The role of Kruppel-like factors in the reprogramming of somatic cells to induced pluripotent stem cells. *Histol Histopathol* 2009;**24**:1343–55. https://doi.org/10.14670/HH-24.1343.
59. Zhang C, Li C. Revealing the mechanism of lymphoid and myeloid cell differentiation and transdifferentiation through landscape quantification. *Physical Review Research* 2021;**3**:013186. https://doi.org/10.1103/PhysRevResearch.3.013186.
60. Li C, Balazsi G. A landscape view on the interplay between EMT and cancer metastasis. *NPJ Syst Biol Appl* 2018;**4**:34. https://doi.org/10.1038/s41540-018-0068-x.
61. Ye L, Feng J, Li C. Controlling brain dynamics: landscape and transition path for working memory. *PLoS Comput Biol* 2023;**19**:e1011446. https://doi.org/10.1371/journal.pcbi.1011446.
62. An S, Jang SY, Park SM. *et al.* Global stabilizing control of large-scale biomolecular regulatory networks. *Bioinformatics* 2023;**39**:btad045. https://doi.org/10.1093/bioinformatics/btad045.
63. Woodhouse S, Piterman N, Wintersteiger CM. *et al.* SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst Biol* 2018;**12**:59. https://doi.org/10.1186/s12918-018-0581-y.
64. Kamimoto K, Stringa B, Hoffmann CM. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* 2023;**614**:742–51. https://doi.org/10.1038/s41586-022-05688-9.
65. Osorio D, Zhong Y, Li G. *et al.* scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene

regulatory network perturbation. *Patterns (N Y)* 2022;**3**:100434. https://doi.org/10.1016/j.patter.2022.100434.

66. Bravo Gonzalez-Blas C, Matetovici I, Hillen H. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* 2023;**20**:1355–67. https://doi.org/10.1038/s41592-023-01938-4.
67. Shin D, Cho KH. Critical transition and reversion of tumorigenesis. *Exp Mol Med* 2023;**55**:692–705. https://doi.org/10.1038/s12276-023-00969-3.
68. Cho K-H, Lee S, Kim D. *et al.* Cancer reversion, a renewed challenge in systems biology. *Current Opinion in Systems Biology* 2017;**2**:49–58. https://doi.org/10.1016/j.coisb.2017.01.005.