

ARTICLE OPEN



NETISCE: a network-based tool for cell fate reprogramming

Lauren Marazzi¹, Milan Shah¹, Shreedula Balakrishnan¹, Ananya Patil¹ and Paola Vera-Licona^{1,2,3,4}✉

The search for effective therapeutic targets in fields like regenerative medicine and cancer research has generated interest in cell fate reprogramming. This cellular reprogramming paradigm can drive cells to a desired target state from any initial state. However, methods for identifying reprogramming targets remain limited for biological systems that lack large sets of experimental data or a dynamical characterization. We present NETISCE, a novel computational tool for identifying cell fate reprogramming targets in static networks. In combination with machine learning algorithms, NETISCE estimates the attractor landscape and predicts reprogramming targets using signal flow analysis and feedback vertex set control, respectively. Through validations in studies of cell fate reprogramming from developmental, stem cell, and cancer biology, we show that NETISCE can predict previously identified cell fate reprogramming targets and identify potentially novel combinations of targets. NETISCE extends cell fate reprogramming studies to larger-scale biological networks without the need for full model parameterization and can be implemented by experimental and computational biologists to identify parts of a biological system relevant to the desired reprogramming task.

npj Systems Biology and Applications (2022)8:21 | <https://doi.org/10.1038/s41540-022-00231-y>

INTRODUCTION

Cell reprogramming redefines a cell's identity by altering its epigenetic or transcriptional landscapes through the forced expression of transcription factors, small molecules, non-coding RNAs, or microenvironment-mediated changes. One type of cellular reprogramming, cell fate reprogramming, aims to control the internal state of a cell so that it is driven from a selected state to a target state or phenotype^{1–7}. Practical applications of cell reprogramming in stem cell engineering^{8–10} and cancer biology^{11–13} have generated a great interest in the task of cell-fate reprogramming. Identifying combinations of reprogramming targets is especially useful for treating complex diseases, regenerating tissue, or reversing acquired resistance to canonical treatment regimens, where multi-drug approaches may be more effective than single-drug therapy^{14–18}.

Genome-wide and computational systems biology approaches are being broadly adopted for cell fate reprogramming studies. The majority of methods can be divided between those that take a network-based approach^{19–24} and those that take a dynamical systems-based approach^{25–30} to cell fate reprogramming^{30–37}. However, these methods either do not capture all the essential information for cellular reprogramming³⁸, or require mechanistic details and kinetic parameters to build a mathematical model of the system (for a brief overview of these approaches, please see Supplementary Text 1 and Supplementary Fig. 1).

Although many network-based and dynamical systems-based methods take a trial-and-error approach to identify reprogramming targets, cell fate reprogramming can be viewed as a classical attractor-based control theory problem. The goal is to systematically determine how to shift the cell's system from one attractor (stable state) to another with some degree of optimality. Previous theoretical studies on the controllability of systems show that even for large, non-linear biological systems, few targets need to be controlled to guide a system toward a biologically admissible target state^{3,5}. This has been shown experimentally in cell fate reprogramming studies, such as in the transition of embryonic stem cells to somatic cells via knockdown of pluripotency-

associated transcription factors³⁹, the reversion of tumorigenesis by impairment of oncogenic signaling⁴⁰, and fibroblast cell reprogramming⁴¹. The use of control theory to identify cellular reprogramming targets has been limited and is not directly applicable to large cell signaling networks^{42,43}. Among the reasons for this limitation is the scarcity of available mechanistic details and kinetic parameters to build a mathematical model of the system, and, when the mechanistic rules are known, linear functions are used to describe them; however, it is unclear how the commonly observed switch-like behavior of biochemical processes^{44,45}, can influence the results^{1,27,46}.

A new method for identifying attractor-based reprogramming targets using only network topology extends from control theory for non-linear dynamics. The feedback vertex set (FVS) control is a structure-based, attractor-based control method suited to non-linear dynamical systems⁵. A network's minimal FVS is the minimal set of nodes that intersects all cycles in a graph. FVS control states that appropriate perturbations on an FVS, which we refer to herein as FVS control nodes, can drive the system from any arbitrary initial state to any of the attractors of that system (Fig. 1).

While the FVS control method provides a powerful cell fate reprogramming framework, it does not identify the specific perturbations needed on FVS control nodes (knockouts or overexpressions) to drive the system toward a particular set of attractors. By proceeding analogously to Boolean networks, estimating the system's attractor landscape can aid in the search for the perturbations needed to be applied on FVS control nodes. To that end, signal flow estimation algorithms aim to estimate steady states without complete dynamical information of the system. They can be helpful to evaluate the effect of node perturbations in static networks^{47,48} and have been applied in biochemical⁴⁹ and disease networks⁵⁰. The signal flow analysis (SFA) method is especially suited to estimate system dynamics for non-linear complex systems⁴⁷, and its application to biology has been recently explored^{47,51}. SFA estimates a steady-state value for each network node based on a signal propagation equation that

¹Center for Quantitative Medicine, University of Connecticut School of Medicine, Farmington, CT 06030, USA. ²Department of Cell Biology, University of Connecticut School of Medicine, Farmington, CT 06030, USA. ³Center for Cell Analysis and Modeling, University of Connecticut School of Medicine, Farmington, CT 06030, USA. ⁴Institute for Systems Genomics, University of Connecticut School of Medicine, Farmington, CT 06030, USA. ✉email: veralicona@uchc.edu

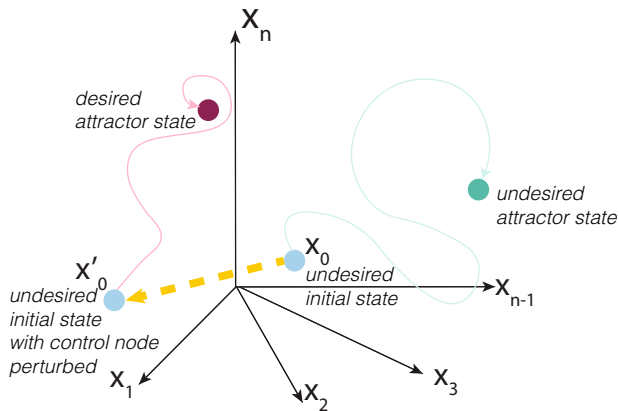


Fig. 1 Control theory view of cell fate reprogramming. The feedback vertex set control is a structure-based control method that can be applied to cell-fate reprogramming. By performing appropriate concerted perturbations to the minimal Feedback Vertex set on an initial state of the system that leads to an undesired attractor (green circle), the system can be shifted (yellow dashed lines) to a trajectory that leads to a desired attractor (magenta circle).

considers the activity of its regulators, the type of regulatory relationship (activation or inhibition), and the initial state of the node (see the “Methods” section). The signal propagation equation is solved for all network nodes synchronously until the difference between $x(t+1)$ and $x(t)$ is less than a tolerance threshold. Thus, for a sufficiently small tolerance threshold (by default, this tolerance threshold is 10^{-6}), a fixed-point attractor is generated for every initial state provided to SFA. As a result, unlike Boolean networks, where many initial states can arrive at one attractor, the attractor landscape estimated by SFA lacks large basins of attraction. In NETISCE, we use machine learning clustering algorithms to systematically partition the estimated attractors⁵² and produce the SFA-estimated attractor landscape. A phenotype is defined by the resultant set of attractors within a cluster and associated with the experimental sample whose attractor is contained in that cluster.

SFA has been used to successfully reproduce the steady-states of signaling networks derived from ODE models and the changes in expression for network elements under different perturbations from perturbation biology experiments with up to 80% accuracy⁴⁷. Latterly, SFA was applied to an aging-related gene regulatory network (GRN) to identify potential aging reversion targets⁵¹. Although Lee and colleagues did not take a control theory-based approach, aging reversion targets were predicted by evaluating SFA-simulated single node perturbations that decreased the estimated expression values of aging-related biomarker nodes compared to an unperturbed simulation.

Previously, we introduced OCSANA+ (Optimal Control and Simulation of Signaling Networks from Network Analysis). This Cytoscape application implements an FVS-finding algorithm and SFA to perform perturbation analyses on static networks⁵³. OCSANA+ can be used to observe the effect of perturbations on FVS control nodes for biological networks in a user-friendly graphical interface. However, the application cannot customize the SFA algorithm to include gene expression data or perform an attractor landscape estimation. In addition, the current software requires the user to configure every perturbation simulation manually.

In this work, we introduce NETISCE (NETwork-driven analysis of CELLular reprogramming), a novel computational method for identifying cell fate reprogramming targets. Specifically, in our context, we define cell fate reprogramming as guiding the system from an initial state to any of its attractors, which can be

associated to an observable cell fate⁴³. Our approach can be applied to any static network and only requires gene expression data from the undesired phenotype for the initial state of the system. NETISCE employs SFA and a machine learning clustering method to estimate the attractor landscape. Then SFA computes the attractor state reached by a given set of perturbations on an FVS control set. Finally, NETISCE uses machine learning classification methods to evaluate whether this newly computed attractor drives the system to the region of the attractor landscape associated with the desired cell fate.

To illustrate and validate our approach, we apply NETISCE to three different examples of cell fate reprogramming in developmental, stem cell, and cancer biology using GRNs and signaling networks. We show that NETISCE can reproduce the results of experimentally validated cell fate reprogramming studies and identify new reprogramming targets and perturbations.

We conclude that NETISCE extends the usefulness of static biological networks to analyses that currently require full parameterization. This is implemented by approaching cell fate reprogramming from the framework of control theory and dynamical systems and applying these concepts to a network-based analysis. Our method provides a practical and informative step for researchers designing experimental or mathematical modeling studies of cell fate reprogramming by identifying parts of the system relevant to the desired reprogramming task. NETISCE is a user-friendly tool implemented as a command-line Nextflow pipeline and Galaxy Project workflow that non-experts can use in modeling or computational approaches to analyze the biological systems of their interest.

RESULTS

NETISCE identifies combinations of perturbations applied on a GRN or signaling network to trigger a shift from an undesired to the desired cell fate (Fig. 2a). The core of the pipeline is (1) the application of the FVS control to identify reprogramming targets and (2) an attractor landscape estimation coupled with machine learning methods to predict the precise perturbations that drive the system from an initial state (that would lead to an attractor associated with an undesired phenotype) and toward an attractor associated with the desired phenotype.

Method's validation

We provide here three application examples of NETISCE in experimentally validated studies of cell fate reprogramming. In addition, we present a comparative study of NETISCE applied to the *Drosophila melanogaster* patterning specification system described in silico both as Boolean and Ordinary Differential Equations Models (Supplementary Text 2, Supplementary Tables 1, 2, Supplementary Fig. 2).

Reproducing experimentally validated perturbations to FVS control nodes for cell fate specification in ascidian embryos

Using a GRN of cell fate specification in ascidian embryos, Kobayashi and colleagues experimentally verified that concerted perturbations to the network's FVS could induce embryonic cells to the epithelial, mesenchymal, endodermal, notochord, brain, and pan-neural, and muscle tissue fates^{54,55}. We performed simulations of the experimentally verified perturbations on FVS control nodes for cell fate specification in ascidian embryos using SFA.

The ascidian embryo GRN contained 92 nodes and 329 edges (Fig. 3a). We identified all 26 FVSes within the ascidian embryo GRN, including the set of six FVS control nodes experimentally tested by Kobayashi et al.⁵⁵: Foxa.A, Foxd, Erk Signaling, Neurog, Tbx6-r.b, and Zic-r.b (Fig. 3b). Without available normalized expression data, we simulated in silico unperturbed embryonic

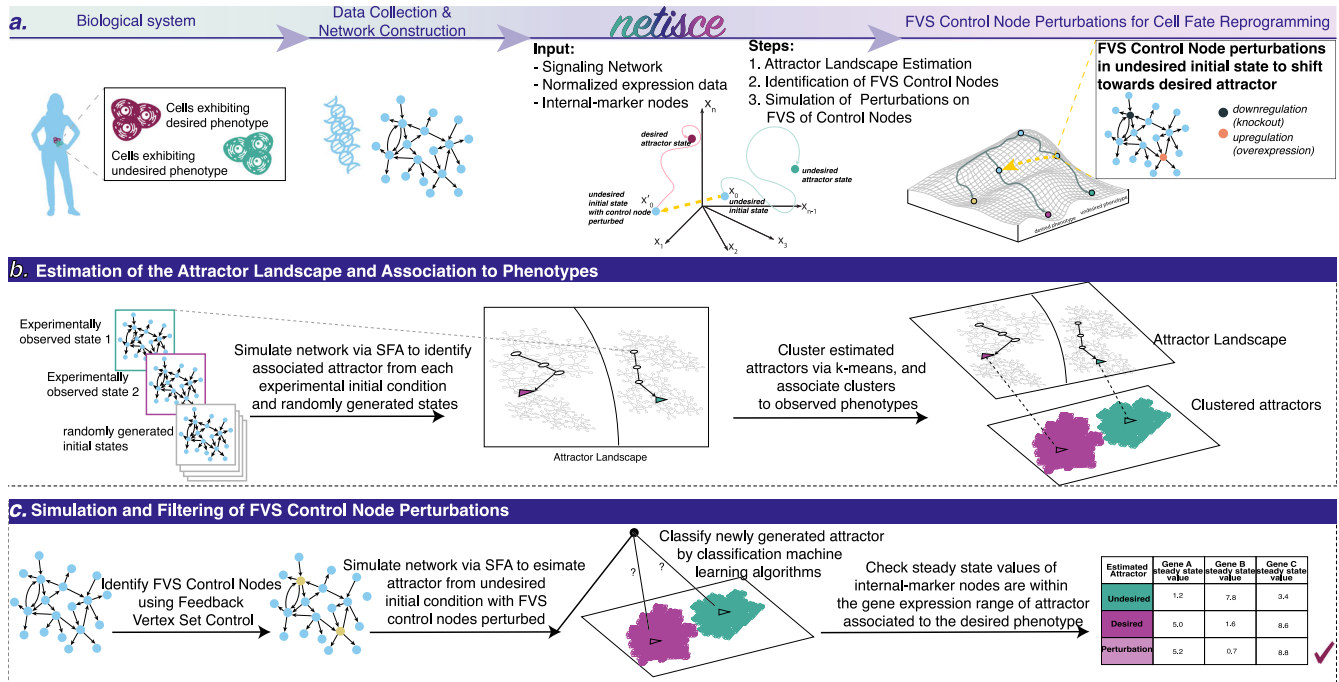


Fig. 2 NETISCE pipeline and method overview. **a** Researchers can collect data from cells exhibiting desired and undesired phenotypes and construct the signaling and regulatory networks governing cell reprogramming processes. The outputs of NETISCE are the combinations of network perturbations that shift the system from an undesired to a desired phenotype. **b** With the signal flow analysis (SFA) algorithm⁴⁷, in the first step, attractors are estimated by simulating the network with the initial states from normalized expression data and randomly generated initial states. The attractors are clustered via *k*-means, and the clusters are associated with desired (purple) and undesired (green) phenotypes. **c** In the second step, FVS control nodes are identified using an FVS-finding algorithm⁸⁴. Perturbations on FVS control nodes are performed by setting the initial states of the system to the gene expression value of the undesired phenotype and overriding the states of FVS control nodes. In the third step, the sets of perturbations on FVS control nodes that achieve the desired reprogramming are identified using two filtering criteria. In the first criterion, the attractors generated from the perturbations on FVS control nodes are filtered using machine learning classification algorithms to obtain a set of perturbations whose attractors shifted from the cluster associated with the undesired phenotype to the cluster associated with the desired phenotype. In the second criterion, for user-defined internal-marker nodes, the expression values in the attractors produced by the perturbations that passed the first filtering criterion (light purple) are evaluated to determine if their values have shifted to the expression range of the attractors associated with the desired phenotype.

development by setting the initial activities of two genes necessary for normal embryonic development, *Gata.a* and *Zic.r.a*⁵⁶, to 1 (the activated state) and all other nodes to 0 (inactivated state). The attractors of the unperturbed state and the seven experimentally validated combinations of perturbations (synchronous overexpression and knockout simulations) on FVS control nodes that induced the embryonic tissue fates were estimated using SFA. To evaluate the perturbations, we analyzed the attractors' expression values of the seven internal-marker nodes (one marker for each tissue) representing genes measured in the experimental study (Fig. 3c, Supplementary Table 3). Attractors estimated by SFA can be compared analogously to the logarithm of the fold-change (\log_2FC) in differential gene expression analysis⁴⁷, where the difference between the expression value of a node in attractors generated from different initial states or perturbations indicates that the gene expression of the node is upregulated if the difference is positive, or downregulated if the difference is negative (see "Methods" for details).

We successfully reprogrammed the unperturbed embryo to six of the seven tissue fates using the corresponding experimentally verified perturbations on FVS control nodes, giving the SFA in our pipeline an overall 85% accuracy (Fig. 4, Supplementary Table 4). We could not reproduce reprogramming to the pan-neural tissue fate. However, the internal-marker node for the pan-neural cell fate (*Celf3.a*) was upregulated when we simulated the perturbation on FVS control nodes that experimentally induced the brain +pan-neural tissue fates.

Identification of perturbations on FVS control nodes for induced pluripotent stem cell reprogramming from primed to naive pluripotency

The mechanisms that maintain stem cells' pluripotency or signal development are complex. Yachie-Kinoshita and colleagues constructed a Boolean Model to study pluripotent cell fate transitions under the constraint of signaling inputs⁵⁷. The system was simulated under a stochastic asynchronous updating scheme. They simulated perturbations to every network node and identified targets to reverse primed pluripotency to naive pluripotency. Yachie-Kinoshita et al. experimentally validated that the predicted targets reprogrammed primed pluripotency epiblast stem cells (EpiSCs) toward a naive pluripotency embryonic stem cell (ESC) state. We used NETISCE to identify combinations of perturbations on FVS control nodes that reprogram EpiSCs toward the ESC state (Fig. 5a).

The pluripotency signaling network contained 36 nodes and 143 edges (Fig. 5b). Using NETISCE, we estimated the six attractors from EpiSC and ESC gene expression data and attractors from 100,000 randomly generated initial states for a total of 100,006 attractors. On these 100,006 attractors, we performed *k*-means clustering. The optimal number of clusters identified by the elbow and silhouette metrics was $k=2$. One cluster contained the attractors generated from the initial state values of the EpiSC cells' gene expression. The second cluster included the attractors generated from the ESC cell gene expression initialization.

NETISCE identified only one FVS in the network, comprising six nodes: *Nanog*, *Oct4*, *Klf4*, *Sox2*, *Gata6*, and *Tbx3* (Fig. 5c). Then, we

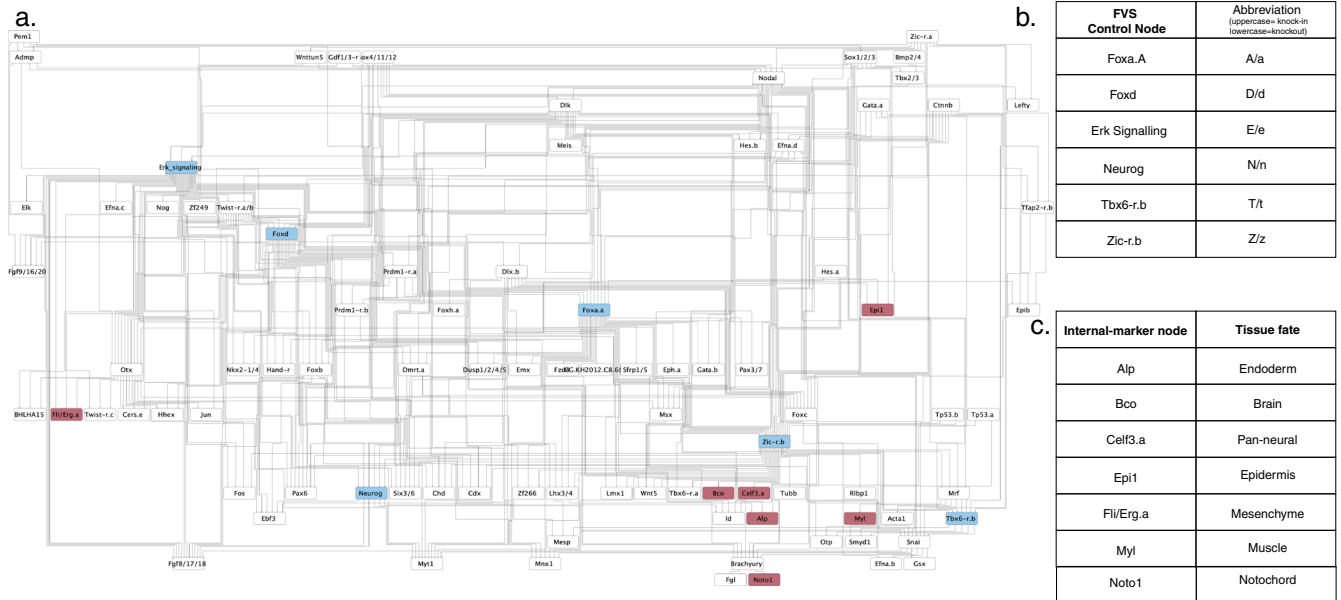


Fig. 3 Cell fate specification in ascidian embryo model. **a** Gene regulatory network of cell fate specification in the ascidian embryo *Ciona intestinalis* from Kobayashi et al.⁵⁵. The network contains 92 nodes and 329 edges. Nodes highlighted in blue are FVS control nodes. Nodes colored in magenta are the internal-marker nodes used to evaluate if the perturbations on FVS control nodes successfully specified the desired cell fate. **b** The six control nodes identified by Kobayashi et al. that comprise the network's minimal feedback vertex set (FVS). Uppercase or lowercase abbreviations for each node indicate up or downregulation of the FVS control node in a given combination of perturbations (see Fig. 4). These nodes are used as the FVS control nodes in our simulations. **c** The seven internal-marker nodes and the respective tissue fates where they are upregulated as identified by Kobayashi et al. experimental studies. These nodes were also used to identify successful perturbations in our simulations.

simulated the 729 combinations of perturbations on the FVS control nodes. Of the 729 perturbations, 375 passed the machine learning classification filtering criterion. We identified each machine learning classification algorithm's top 10 percent ranked features through feature importance analysis. Two out of the three top SVM features were FVS nodes. The top features across all three classification algorithms were mutually exclusive, alluding to their complementary nature (Supplementary Text 3, Supplementary Table 6). We used the same four internal-marker nodes as Yachie-Kinoshita et al.: Oct4, Sox2, and Nanog as markers of naive pluripotency, and EpiTFs as the marker of enriched transcription factors in EpiSCs. The expression values of the internal-marker nodes in 132 of the 375 attractors calculated from the perturbations on FVS control nodes were in the range of gene expression values of the ESC-associated attractors and thus passed criterion 2 for all replicates. Notably, one perturbation on FVS control nodes that passed both filtering criteria—overexpression of Nanog (pert_Nanog_ovr)—was also identified and experimentally validated by Yachie-Kinoshita et al. (Fig. 6). In the Boolean simulations and experimental validation by Yachie-Kinoshita et al., Klf4 overexpression induced the ESC fate. Although Klf4 was an FVS control node in the network and its overexpression passed the machine learning filtering criterion, the perturbation (pert_Klf4_ovr) did not pass the internal-marker node filtering criterion (Fig. 6). In this case, the attractor expression values of Nanog, Sox2, and Oct4 (when considered as internal-marker nodes) did not reach the gene expression levels of the ESC state (Fig. 6). Overall, we show that overexpression of the FVS control node Nanog results in cell fate reprogramming to naive pluripotency, in agreement with the results from the Boolean Model simulations and experimental validations.

We explored the ability to further filter the perturbations on FVS control nodes by increasing the number of internal-marker nodes. We identified three additional nodes from gene expression data provided by Yachie-Kinoshita et al. These included Lefty1, Pitx2 (transcription factors active in EpiSCs), and Esrrb (a transcription

factor active in ESCs). This reduced the 132 perturbations that passed filtering criterion 2 to 15 perturbations. Nanog upregulation was present in all 15 perturbations (Supplementary Fig. 3). Pert_Nanog_ovr and the combination of Nanog+Klf4 overexpression (pert_Nanog_ovr+Klf4_ovr) were two of the fifteen perturbations that passed the internal-marker node criteria (Fig. 6, Supplementary Table 5). Pert_Nanog_ovr+Klf4_ovr was not previously identified by Yachie-Kinoshita et al. However, the combination of Klf4 and Nanog overexpression may play an essential role in maintaining pluripotency, as Klf4 links extracellular signaling information to positively regulate downstream Nanog transcription, and overexpression of Nanog was found to rescue pluripotency in the case of Klf4 knockdown⁵⁸. Finally, based on the results of NETISCE, we identified a potential error in the underlying network and performed revisions to the network structure to improve simulation results (Supplementary Text 4).

Identification of perturbations on FVS control nodes to overcome adaptive resistance to targeted MAPK inhibitor therapy in colorectal cancers

BRAF inhibition (BRAFi) therapy is a form of MAPK inhibitor (MAPKi) therapy used to treat cancer patients with mutant BRAF. BRAFi inhibits the MAPK signaling pathway, suppressing proliferation and inducing apoptosis. In colorectal cancers (CRCs), adaptive resistance emerges against BRAFi through the activation of MAPK signaling by upstream regulator EGFR. Park et al. were interested in identifying a gene that, when perturbed in combination with BRAFi, could prevent the development of adaptive resistance to BRAFi⁵⁹. However, instead of inhibiting upstream molecules like in the BRAFi + EGFRi treatment, they searched for a target within the MAPK signaling pathway that could sensitize HT29 CRC cells to MAPKi therapy. By constructing a Boolean model of signaling pathways in CRC and simulating perturbations to every node in the model under an asynchronous updating scheme, Park et al. showed that BRAFi combined with SRC inhibition (SRCi) prevented

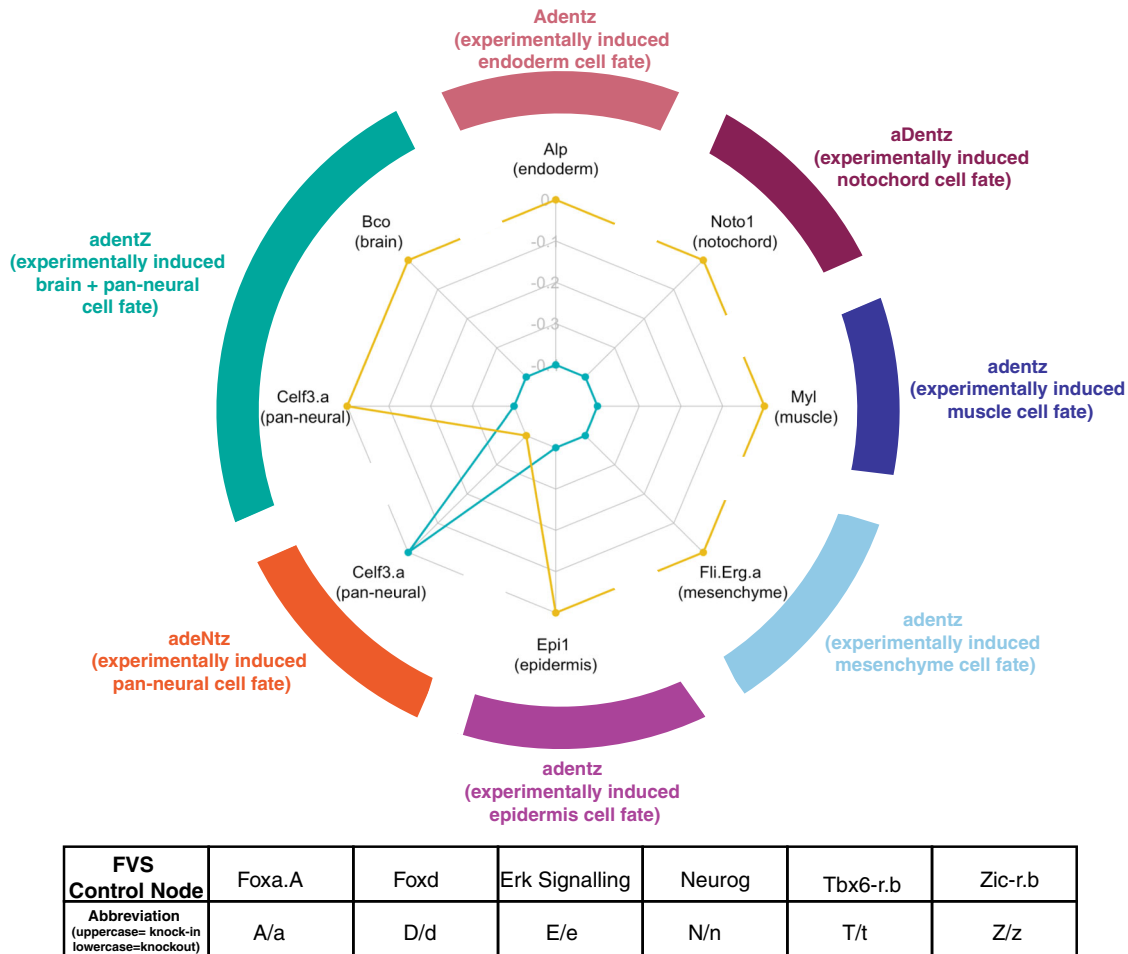


Fig. 4 SFA simulations of perturbations on FVS control nodes for cell fate specification in ascidian embryos. Kobayashi et al.⁵⁵ performed experimental combinations of perturbations on FVS control nodes to induce seven tissue fates in ascidian embryos. We aimed to reproduce these results in silico by simulating the combinations of perturbations on FVS control nodes using SFA. The results of each combination of perturbations are displayed in the radar plot. Each axis on the radar plot displays the expression value for an internal-marker node, with the tissue each internal-marker node represents in parentheses. The teal polygon shows the expression values of the internal-marker nodes in the attractor associated with the unperturbed state. The yellow polygon displays the internal-marker node attractor state expression values produced by applying one combination of perturbations on FVS control nodes to the unperturbed initial state. Note that the precise steady-state values have been scaled for plotting, and the raw data can be found in Supplementary Table 3. The outer colored ring denotes the combination of perturbations on FVS control nodes performed with the respective results displayed at that axis. The tissue fate that was induced experimentally by a perturbation is denoted in parentheses. Each letter and capitalization stand for a separate FVS control node and its perturbed state, as described in the table at the bottom. For a simulation of the perturbations on FVS control nodes to be considered successful, the expression value of the internal-marker node must be greater in the attractor produced by the perturbation on the FVS control nodes than the expression value in the attractor associated with the unperturbed state (the yellow polygon extends out past the teal polygon on the radar plot). We reproduced the cell fate specification results for 6/7 cell fates, excluding adentZ inducing the pan-neural cell fate.

the development of adaptive resistance via inhibition of ERK (MAPK1), a member of the MAPK signaling pathway. This result was validated experimentally in HT29 CRC cells. Therefore, with only the CRC signal pathway network structure, RNA-seq data from untreated HT29 cells, and functionally annotated mutational information as input to NETISCE, we sought to identify perturbations on FVS control nodes in combination with BRAFi that overcome adaptive resistance to MAPKi therapy (Fig. 7a).

Park and colleagues built a CRC signaling network containing 95 nodes and 337 edges (Fig. 7b). We adapted the PROFILE method⁶⁰ to verify that the generic CRC network and our SFA simulations preserved the phenotypic signatures of apoptosis and proliferation (see Supplementary Text 5 and Supplementary Fig. 4).

The normalized gene expression data from an untreated HT29 sample was used for the initial activities for all SFA simulations, as data for HT29 cells with BRAFi and BRAFi+EGFRi treatment is unavailable. In HT29 cells, PIK3CA and BRAF have

gain-of-function mutations, while APC, SMAD4, and TP53 have loss-of-function mutations. Therefore, states of nodes with gain-of-function/loss-of-function mutations were overridden to the appropriate overexpression or knockout state using our modified SFA equation for perturbations (see “Methods”). Next, we simulated the treatment of an untreated HT29 cell with BRAFi (HT29_BRAFi) or BRAFi+EGFRi (HT29_BRAFi+EGFRi) to obtain attractors related to the MAPKi therapy-resistant state and MAPKi therapy-sensitive state, respectively. These simulations used the normalized expression data of the untreated HT29 sample as initial state values and included the appropriate mutational overrides. To simulate BRAFi and EGFRi, the state of these nodes was overridden to a knockout state using the modified SFA equation for perturbations. The optimal *k*-means clustering of the attractors obtained from the untreated HT29, HT29_BRAFi, HT29_BRAFi+EGFRi treated initial conditions, and the attractors from 100,000 randomly generated initial states was *k*=3; as

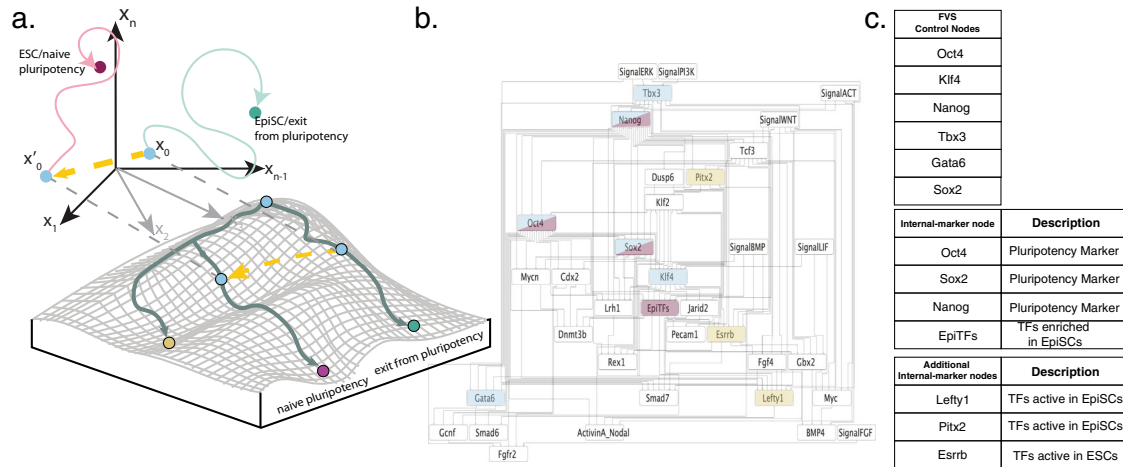


Fig. 5 Reprogramming of epiblast stem cells (EpiSCs) to embryonic stem cells (ESCs). **a** The goal of this study is for NETISCE to identify the perturbations on FVS control nodes that shift EpiSCs toward ESCs, reverting from a primed to a naive pluripotent state. To that end, we initialize the system with the gene expression data from EpiSC cells (x_0), and, with the correct combination of perturbations on FVS control nodes, shift the system to a state (yellow arrow leading to x'_0) that leads to the ESC cell fate (magenta circle). **b** Pluripotent stem cell signaling network as constructed by Yachie-Kinoshita et al.⁵⁷. The network contains 36 nodes and 143 edges. Nodes colored in blue are FVS control nodes, nodes with magenta coloring are the internal-marker nodes used by Yachie-Kinoshita et al., and gold nodes are the additional internal-marker nodes we used for further selection of perturbations on FVS control nodes. Note that the nodes Sox2, Nanog, and Oct4 are FVS control nodes and internal-marker nodes, and thus colored with both blue and magenta. **c** Key nodes for the desired cell fate reprogramming. There are six FVS control nodes within the system. Four internal-marker nodes were used by Yachie-Kinoshita et al. to identify successful reprogramming targets for the ESC state. We identified an additional three internal-marker nodes to filter perturbations on FVS control nodes. In the second column for the internal-marker nodes, we denote which behavior they represent in pluripotency signaling.

desired, the untreated-state, MAPKi therapy-resistant state and MAPKi therapy-sensitive state associated attractors were found in separate clusters.

NETISCE exhaustively identified 68 FVSes in the CRC static signaling network; the union of all FVSes contained 25 nodes, and each FVS had a combination of 14 out of the 25 nodes. All FVSes contained SRC. In addition, each FVS included TP53, a loss-of-function mutant gene in HT29 cells whose state was already overridden to a knockout state in our simulations; therefore, additional perturbations to TP53 were not performed, reducing the number of FVS control nodes that could be perturbed to 13 (Supplementary Data 1).

We present the results of using one FVS, referred to as Set 1 (Fig. 7c), to identify combinations of perturbations that shift the system from the MAPKi therapy-resistant phenotype to the MAPKi therapy-sensitive phenotype. We simulated 1,594,323 combinations of perturbations to the 13 FVS control nodes in Set 1 to identify their corresponding attractors. First, 232,114 of the 1,594,323 attractors generated from the perturbations on FVS control nodes were classified to the MAPKi therapy-sensitivity associated cluster by at least two out of three machine learning classification algorithms. FVS control nodes or internal-marker nodes were part of the top 10 percent of ranked features revealed by the feature importance analysis on the three machine learning classification algorithms (Supplementary Text 6, Supplementary Table 7). Then, the 232,114 perturbations that produced the attractors that passed criterion 1 were filtered by the three internal-marker nodes that were used as output readout nodes in Park et al. (Fig. 7c): CASP3 (apoptosis marker), Cyclin E (proliferation marker), and MAPK1/ERK (MAPK signaling activity marker). The internal-marker nodes in the attractors produced by 52,703 of the 232,114 perturbations on FVS control nodes had expression values in the range of gene expression values of the MAPKi therapy sensitivity associated attractors. Notably, the combination of BRAFi + SRCi (pert_BRAFi+SRCi) was the smallest perturbation set that passed both filtering criteria (Fig. 8).

We noted that the expression values of Cyclin E and MAPK1 in the attractors produced from the HT29_BRAFi and HT29_BRAFi +EGFRi treated initial conditions were very similar. Although the difference between these values of Cyclin E and MAPK1 had the correct relationship as expected from the literature (the attractor-state values in the HT29_BRAFi+EGFRi-associated attractor were lower than the values in the HT29_BRAFi-associated attractor), we considered an additional ten internal-marker nodes related to apoptosis, proliferation, and MAPK signaling (Fig. 7c) to filter and gain confidence on our predicted perturbations on FVS control nodes. Of the 52,703 perturbations on FVS control nodes, 2085 met the filtering threshold of 90% for the apoptosis, proliferation, and MAPK internal-marker node steady-state values. The combination of BRAFi + SRCi alone did not pass the filtering criteria, as the value of MLK3, a member of the MAPK signaling pathway whose reactivation is linked to BRAFi resistance⁶¹, did not shift into the range of MAPKi sensitive attractors' gene expression. However, SRCi was present in all the combinations of control nodes that passed two filtering criteria (Supplementary Fig. 5).

The smallest sets of perturbations were three pairs of control nodes. The first reprogramming pair consisted of SRCi and TSC1 overexpression (pert_SRCi+TSC1ovr), the second reprogramming pair was comprised of SRCi and GRB2i (pert_SRCi+GRB2i), and the third pair contained SRCi and MAPK8i (pert_SRCi+MAPK8i) (Fig. 8). TSC1 overexpression has been studied in the context of adaptive resistance to MAPKi. TSC1 promotes cell death by inhibiting mTOR activity, and mTOR inhibition in combination with BRAFi has been shown to overcome adaptive resistance in BRAF-mutant melanoma⁶². However, a complex of TSC1 and TSC2 can be inactivated by ERK phosphorylation, leading to increased mTOR signaling⁶³. In the context of CRC, GRB2 is an important protein for transmitting oncogenic signaling and promoting tumorigenesis and metastasis⁶⁴. Interestingly, the protein Gab2, a binding partner of GRB2, was directly upregulated by BRAFi in BRAF mutated cancers⁶⁵. Impairing the interaction between GRB2 and Gab2 sensitized cells to BRAFi therapy and prevented additional oncogenic signaling and metastasis in HT29 cells^{65,66}. Finally, MAPK8 is a member of

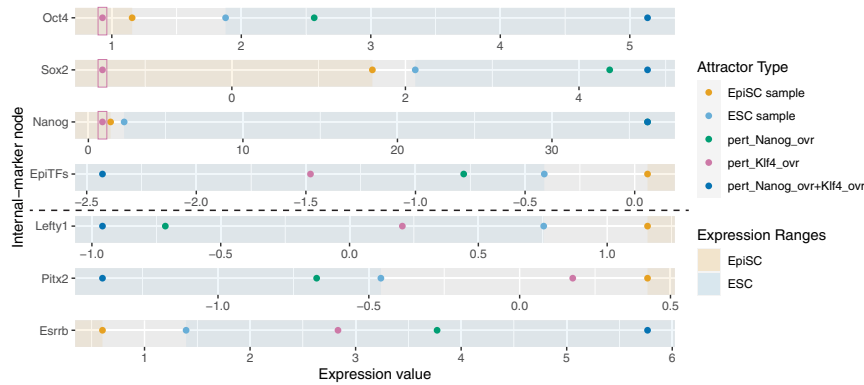


Fig. 6 Results of simulations of perturbations on FVS control nodes for EpiSC reprogramming to ESCs. Expression value plots for SFA simulations of perturbations on FVS control nodes in the pluripotent stem cell model. Each plot contains the expression values of one internal-marker node at the attractor states generated from the initial conditions: EpiSC and ESC from the available gene expression data, in addition to the specified perturbations on FVS control nodes (pert_Nanog_ovr, pert_Klf4_ovr, pert_Nanog_ovr+Klf4_ovr). Shaded regions of the plots indicate the ranges of either ESC (blue) or EpiSC (gold) expression values. A perturbation is considered successful if 90% of the internal-marker nodes have expression values within the gene expression range of the ESC-associated attractor. In other words, the colored dot for a specified perturbation should be within the blue-shaded ESC region of at least 90% of the internal-marker nodes to pass the second filtering criterion. Plots using the four internal-marker nodes used in Yachie-Kinoshita et al. are above the dotted line. Plots using the additional three internal-marker nodes identified are below the dotted line. We note that the Klf4 overexpression (pert_Klf4_ovr) suggested by Yachie-Kinoshita et al., does not pass the internal-marker node filtering criterion because the expression values of Oct4, Sox2, and Nanog have not reached the threshold of the expression value in the attractor produced from ESC gene expression data (pink boxes). However, in addition to the Nanog overexpression (pert_Nanog_ovr) perturbation identified by Yachie-Kinoshita et al., NETISCE identified the combination of Nanog and Klf4 overexpression (pert_Nanog_ovr+Klf4_ovr) to pass both machine learning and internal-marker node filtering criteria.

the MAPK signaling pathway and an important regulator of apoptosis and autophagy⁶⁷. In addition, inhibition of MAPK8 was found to increase sensitivity to BRAFi in MAPKi-resistant melanoma cell lines⁶⁸.

Evaluating the robustness to noise of NETISCE

Commonly, gene expression data can be infiltrated with extrinsic noise from experimental methods or biological variability. We evaluated the robustness to noise of NETISCE to predict perturbations on FVS control nodes with noisy experimental gene expression data. Our approach uses COPASI⁶⁹ to simulate two differential equation models of cell reprogramming and add different noise levels into some initial states. We use these generated initial states to simulate 1000 triplicates of the desired and undesired phenotype normalized gene expression data per noise level varying from 0 to 50% (see “Methods”). Overall, applying this analysis in the ODE model of *Drosophila* Segment Polarity genes⁷⁰ and a stochastic DE (SDE) model of cell fate differentiation in pancreatic cells³⁵, we found that NETISCE is strongly robust to up to 50% of noise for which more than 75% of the perturbations on FVS control nodes were still correctly identified.

The ODE model of the *Drosophila* segment polarity GRN was used in FVS control studies. The authors identified a perturbation on the FVS that shifted the system from the unpatterned to the wild-type trajectory (Supplementary Text 2). In this model, NETISCE correctly predicts the specified perturbations on FVS control nodes that shift the system from the unpatterned to the wild-type phenotype in over 92% of the simulations for the 1000 triplicates with 30% noise and 75% of simulations for the triplicates with 50% noise (Supplementary Table 8).

The Zhou et al. SDE model of pancreatic cell fate differentiation was used to simulate the reprogramming of pancreatic exocrine cells to beta-cells³⁵. We identified 31 combinations of perturbations on FVS control nodes that could reprogram exocrine cells to beta-cells (Supplementary Fig. 7, Supplementary Text 7). For more than 89% of the simulations of 1000 triplicates with up to 20% of noise, NETISCE correctly predicted the 31 combinations of perturbations on FVS control nodes (Supplementary Table 9).

At 50% noise in the initial states, 84.1% of the simulations for the 1000 triplicates correctly predict the 31 combinations of perturbations on FVS control nodes.

DISCUSSION

With the rise in availability of multi-omics datasets and tools for constructing gene regulatory and intracellular signaling networks from these data^{71–73}, there is a growing need for cell reprogramming methods that are data-driven and amenable for larger-scale biological networks where parameters to model all the system components may not be available or be challenging to estimate. We have developed NETISCE, a tool that identifies cell fate reprogramming targets using the FVS control, attractor landscape estimation, and machine learning methods. By reproducing experimental and mathematical model results, we show that NETISCE can identify cell fate reprogramming targets and their perturbations using the system’s static network, gene expression data from the undesired cell phenotype, and a set of nodes used as internal-marker nodes for the desired and undesired phenotypes.

NETISCE offers a unique approach to identifying cell fate reprogramming targets through its application of control theory and a dynamical systems-based framework. First, by employing the structure-based FVS control, we consider feedback loops that commonly regulate biological functions essential for identifying reprogramming targets. Since our target search focuses on FVS control nodes, we guarantee that the identified nodes are sufficient for cell fate reprogramming. Applying the FVS control to the problem of identifying cell fate reprogramming targets has distinct advantages. In comparison to network-based approaches, we do not need to compare the structure of multiple networks to identify cell fate reprogramming targets; in contrast to dynamical systems-based methods that do not apply control theory, full dynamical information for all network components is not required, nor is there a need to screen all network elements.

The FVS control contrasts with other control theory approaches for identifying targets like Data Guided Control, which may fail to capture cell reprogramming dynamics due to its linear assumption of regulatory dynamics⁴¹. It is important to note that FVS control

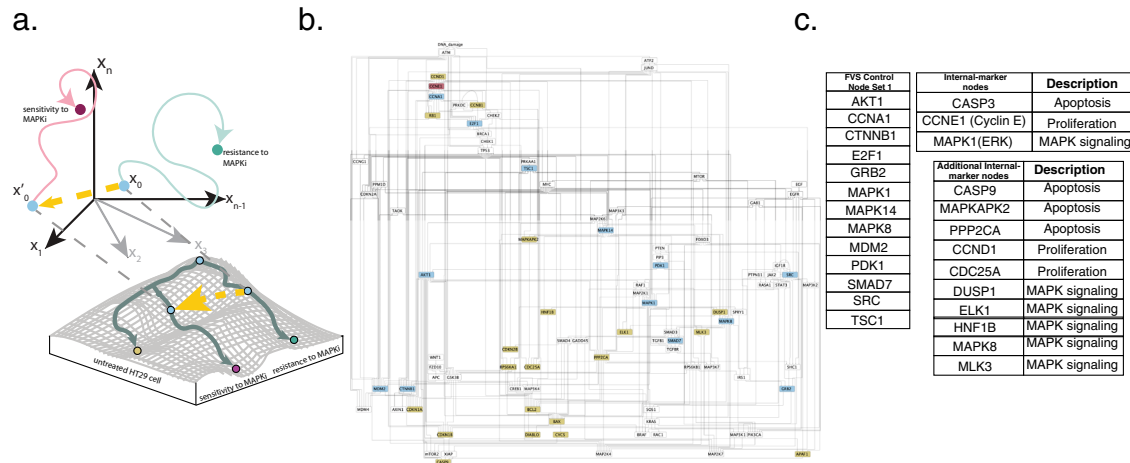


Fig. 7 Overcoming adaptive resistance in the colorectal cancer model. **a** In this study, we use NETISCE to identify perturbations on FVS control nodes that shift BRAFi-treated HT29 cells away from a state of resistance to MAPK inhibitor (MAPKi) therapy and toward a MAPKi therapy sensitivity. We initialize the system with gene expression data from untreated HT29 cells (x_0) and simulate BRAFi together with perturbations on FVS control nodes to shift the system to a state (yellow arrow leading to x'_0) that leads to sensitivity to MAPKi therapy (magenta circle). **b** Colorectal cancer signaling network introduced in Park et al.⁵⁹. Blue nodes are control nodes, magenta nodes are internal-marker nodes, and gold nodes are additional internal-marker nodes used for further filtering of perturbations of FVS control nodes. **c** Key nodes for the cell fate reprogramming. One of the 68 FVS control node sets within the network is presented here, called Set 1. There are three internal-marker nodes used by Park et al. to identify targets to reprogram the system toward the MAPKi therapy-sensitive state. The three nodes are used to measure the phenotypes of apoptosis, proliferation, and MAPK signaling activity within the perturbed system. We identified an additional 17 internal-marker nodes to filter perturbations on FVS control nodes. The phenotypes associated with the internal-marker nodes are denoted in the second column.

for identifying reprogramming targets requires a high-fidelity static network. As Mochizuki and colleagues observed⁵ and later, Kobayashi and colleagues showed experimentally^{54,55}, if a cell state cannot be reached by perturbations to the FVS control nodes, then the FVS was not correctly identified, and network revision should be performed.

Second, the SFA algorithm and machine learning methods allow us to identify the specific perturbations on FVS control nodes required for cell fate reprogramming by estimating system dynamics and the attractor landscape. In our approach, by associating phenotypes to the attractor landscape via k -means clustering and classifying the attractors produced by perturbations on FVS control nodes via machine learning classification methods, we observe when the system has shifted toward the attractor associated with the desired phenotype. Last, we have modified the original SFA algorithm⁴⁷ to apply FVS control-based perturbations. Unlike the SFA control method introduced by Lee and Cho⁷⁴, where edge modifications (removals or additions) were implemented to perform perturbations, our method of perturbations to a node's state maintains the FVS of the network. In addition, we have implemented permanent overrides on FVS control nodes rather than the original form of SFA perturbations that are transiently applied by changing only the initial states of nodes. Transient perturbations are not applicable in the context of FVS control. First, as defined by Fielder et al.⁶ and Mochizuki et al.⁵, overrides to the states of the FVS control nodes guarantee that the system will arrive at any desired attractor. Second, our implementation mimics the experimental perturbations in Kobayashi et al.⁵⁵, where FVS genes were permanently overexpressed or knocked out in the ascidian embryo. While transient perturbations could be used to simulate some types of single-drug treatments, their use in the context of the FVS control may not produce the desired reprogramming. In mathematical models, perturbations on FVS control nodes can drive the system to fixed-point attractors or cycles. However, the SFA algorithm currently identifies only fixed-point attractors due to its tolerance threshold.

A future update could modify the tolerance threshold to identify limit cycles.

NETISCE reproduced the simulated and experimentally validated results in different applications of cell fate reprogramming. We have shown that NETISCE reproduced the results of in silico simulations of the FVS control in both Boolean and ODE models and that it is highly robust to noise. Finally, although NETISCE performs synchronous, deterministic simulations, we reproduced results from asynchronous and stochastic mathematical models.

Importantly, we have shown that NETISCE can be used to personalize simulations on a network when provided with expression (and, if available, mutational profiles) data for a specified sample. We use PROFILE⁶⁰ to verify the network structure and SFA simulations using patient tumor data on the CRC signaling network. PROFILE can be applied in analogous cancer studies and when the signatures of interest are specific genes related to a biological function or phenotype, such as those in Molecular Signatures Database⁷⁵. However, there is not an established model verification method for problems outside of cancer. In these problems, such as the iPSC example, NETISCE simulations and feature importance analysis can evaluate the correctness of the network and potentially correct errors if literature information or data exists (see Supplementary Text 4).

In the example of cell fate specification in ascidian embryos, our SFA simulations succeeded in reproducing 85% of the experimentally validated perturbations. This accuracy may be a limitation of the SFA estimation algorithm. The biological process of pan-neural tissue specification may explain the inability to induce the pan-neural tissue fate by Neurog overexpression. Otx—a gene downstream of Neurog involved in pan-neural cell fate specification—is inactive in the ascidian embryo until the 32-cell stage of development⁵⁶. Since overexpression of Neurog was performed at the start of the simulation, we may not be accurately simulating the timing of its pan-neural inducing effect. A future update to the SFA algorithm could perform asynchronous stochastic simulations, allowing for time-delayed perturbations and potentially producing information regarding

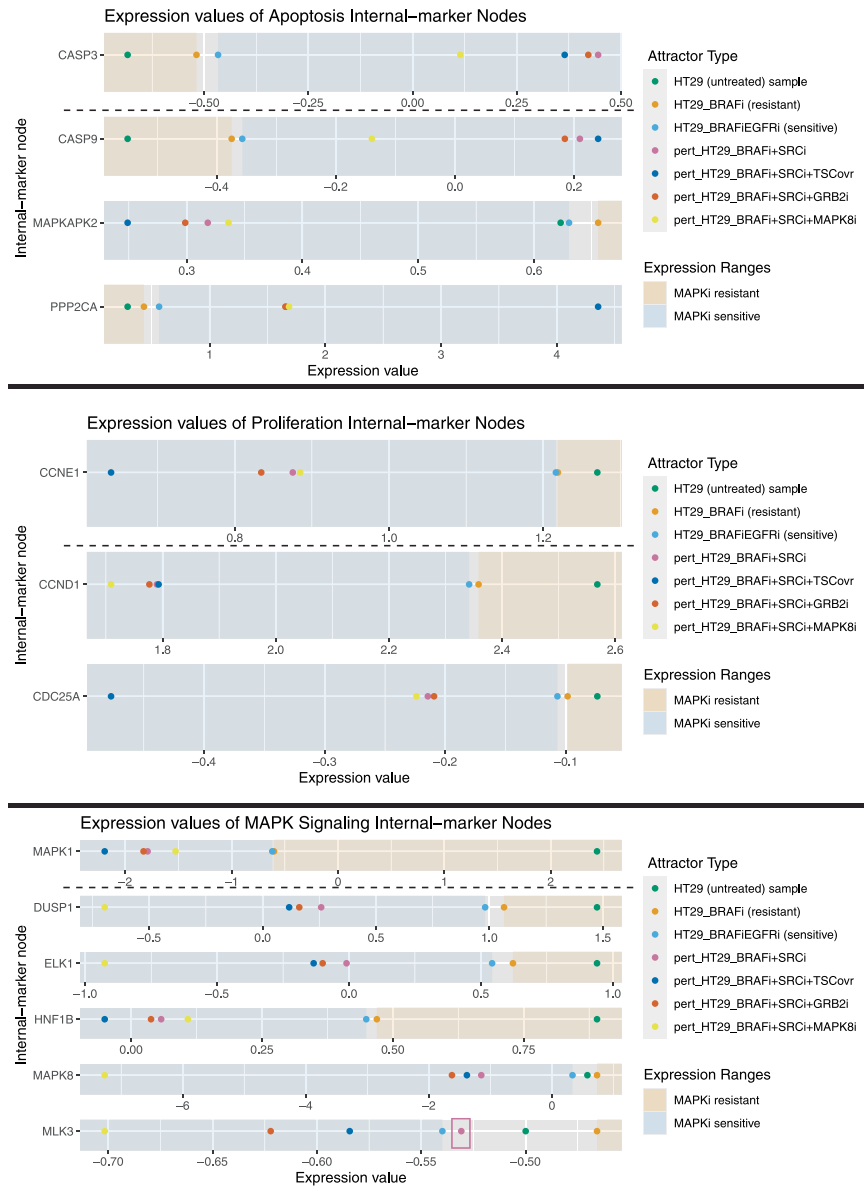


Fig. 8 Results of simulations of perturbations on FVS control nodes for overcoming adaptive resistance in colorectal cancer. Plots of internal-marker nodes' expression values at the attractors reached from the perturbations on FVS control nodes in HT29 cells in the colorectal cancer signaling network. The thirteen internal-marker nodes are divided by their associated phenotype—apoptosis (top), proliferation (middle), and MAPK signaling (bottom). Each plot contains, for one internal-marker node, the expression values at the attractors reached from the HT29_untreated, the HT29_BRAFi (MAPKi therapy-resistant), and the HT29_BRAFi+EGFRi (MAPKi therapy-sensitive), and the specified perturbations on FVS control nodes (HT29_BRAFi+SRCi, HT29_BRAFi+SRCi_TSC1ovr, HT29_BRAFi+SRCi+GRB2i, and HT29_BRAFi+SRCi+MAPK8i) initial conditions. Shaded regions of the plots indicate the MAPKi therapy-sensitive (blue) or MAPKi therapy-resistant (gold) samples expression ranges. In each panel, plots above the dotted lines are the original internal-marker nodes used by Park et al. Plots below the dotted lines are the additional internal-marker nodes. A perturbation is considered successful if 90% of the internal-marker nodes' expression values are within the expression ranges of the HT29_BRAFi+EGFRi-associated attractor (within the blue-shaded regions on the plots). We note that while the attractor state expression values of CCNE1 and MAPK1 are very similar, the difference between the values in the HT29_BRAFi+EGFRi-associated and HT29_BRAFi-associated attractors are in agreement with prior literature findings. The four attractors generated by the specified combinations of perturbations on FVS control nodes pass the second filtering criterion when using only three internal-marker nodes from Park et al., CASP3, CCNE1, and MAPK1. However, the attractor generated by Park and colleagues's perturbation of BRAFi+SRCi does not pass the second filtering criterion when considering the expanded set of internal-marker nodes, as the value of MLK3 is not within the MAPKi therapy-sensitive expression range (bottom panel, pink box).

the specific timings of perturbations for successful cellular reprogramming^{41,76}.

Our FVS control-guided method reduces the number of simulations needed to be performed to identify targets for cell fate reprogramming. In the reprogramming tasks for the pluripotent stem cells and CRC problems, some of the perturbations to the FVS control nodes that NETISCE identified as

successful were a subset of the perturbations found in Boolean models, where reprogramming targets were identified by simulating perturbations to every node in the system. In addition, NETISCE revealed combinatorial strategies for cell fate reprogramming in both models. In the case of the pluripotent stem cell model, Nanog is essential to maintain pluripotency, and additional overexpression of Klf4 could make the reprogrammed cell

unreceptive to extracellular signaling that may signal to exit from pluripotency, preserving high levels of Nanog activity⁵⁸. However, unlike in the Boolean model simulations, we could not identify that Klf4 overexpression alone could reprogram cells to the ESC cell fate. This result may be because SFA does not capture cooperative activity in the same manner as Boolean logic; as in SFA, the initial activity of a node can influence the state of a node at each time step.

In the model of adaptive resistance to MAPKi therapy in CRC, the combination of BRAFi+SRCi+TSC1ovr could further increase sensitivity to treatment and the rate of apoptosis by MAPK and mTOR signaling inhibition⁶³. Alternatively, BRAFi+SRCi+GRB2i can increase sensitivity to MAPKi therapy and prevent the metastatic spread of CRC tumors⁶⁵. Finally, BRAFi+SRCi+MAPK8i may also increase sensitivity to MAPKi therapy⁶⁸ and drive cells toward apoptosis⁶⁷.

With the potential to identify hundreds of thousands of perturbations that satisfy NETISCE's filtering criteria, a method for prioritizing perturbations on FVS control nodes is essential. A simple method employed in NETISCE is generating a secondary set of internal-marker nodes to filter perturbations. The number of necessary internal-marker nodes is dependent on the user's goals and the specific application. For example, a minimal set of marker nodes may generate thousands of successful perturbations on FVS control nodes, allowing the user to understand patterns of perturbations and potentially focus on specific network modules. However, a larger set of internal-marker nodes may allow users to gain confidence in the predicted sets and reduce the number of successful perturbations on FVS control nodes, which could help prioritize the predicted targets to be verified experimentally. This is especially true in the model of adaptive resistance to MAPKi therapy in CRC, where gene expression data for the MAPKi-resistant and MAPKi-sensitive cells was unavailable. In addition, the expanded internal-marker nodes revealed that the combination of BRAFi+SRCi may not provide the most satisfactory reprogramming, as the expression value of MAPK signaling pathway member MLK3 did not shift into the expression range of the MAPKi sensitive attractor.

Depending on the system, it may also be beneficial to prioritize the smallest combinations of perturbations on FVS control nodes to ease experiments or prevent off-target effects. Another prioritization approach could score perturbations based on the strength of their effect on the target phenotypes while minimizing side effects, similar to the method implemented by Park et al. in the CRC Boolean model simulations⁵⁹. A node received a high score if, when the probability of its activity in the results of the asynchronous stochastic simulations indicated inhibition, it would prevent adaptive resistance via ERK reactivation, promote the therapeutic side-effect of increased apoptosis, and prevent the adverse side-effect of increased proliferation. This scoring could be modified for NETISCE. Majorly, NETISCE would need to perform stochastic asynchronous simulations like the Boolean Model simulation framework used by Park and colleagues⁵⁹. The scoring also needs to consider overexpression, knockout, and combinations of perturbations on FVS control nodes. Next, the processing of internal-marker nodes could be modified to consider nodes related to side effects. Finally, an algorithm for path-finding and determining the effect of a perturbation on FVS control nodes on off-target nodes would need to be implemented.

In stem-cell reprogramming, where changes to the epigenetic profile are a significant factor in reprogramming efficiency⁷⁷, implementing a scoring of combinations of perturbations on FVS control nodes that considers epigenetic information could be highly effective to rank reprogramming targets. This information, which is incorporated into the tool, IRENE for both GRN construction and scoring of potential transcription factor-reprogramming targets, increased reprogramming protocol efficiencies in some cases by more than 900%³⁷. Similar to the

method of Park et al., this would also require implementing a stochastic simulation framework and could only be considered if the epigenetic information was available for both the undesired and desired states.

In addition, if applicable, such as in a disease model, information like druggability or drug synergies could be incorporated into prioritizing combinations of perturbations on FVS control nodes^{78,79}. For example, PHAROS⁷⁹, a meta-database of drug-target information, can be used to assess the druggability of SRC, GRB2, TSC1, and MAPK8. Currently, six drugs are approved SRC inhibitors. However, there are no approved drugs for GRB2 or MAPK8 inhibition nor known drugs or molecules that bind to TSC1 and promote its overexpression. Therefore, SRCi would be a likely first candidate for preventing adaptive resistance in combination with BRAFi.

Our dynamical systems-based analysis using static biological networks and experimental data to estimate the attractor landscape and perform combinations of perturbations to control nodes provides a valuable tool for intracellular signaling analysis. Because NETISCE can be applied to biological networks of a larger scale that are not fully parameterized, we envision it as a primary tool for cell fate reprogramming studies. Experimentalists can use the results generated from NETISCE to prioritize wet-lab perturbation experiments. At the same time, mathematical modelers can focus model construction toward regions that appear to be more relevant to the desired reprogramming task and fine-tune reprogramming target predictions by exactly solving for attractors rather than estimating steady states. Finally, our method produces useful and potentially novel combinations of perturbations for cell fate reprogramming that could eventually be applied for treatments in disease models to recover healthy cell phenotypes in biological systems.

METHODS

NETISCE overview

Input. There are three required inputs for NETISCE: (1) a static network representing a biological system, (2) a set of normalized gene expression data from cells with an undesired phenotype, and (3) a set of internal-marker nodes—user-defined nodes within the network that can be used as a point of reference to verify that their gene expression levels match the expected values in desired and undesired phenotypes. Normalized gene expression data for cells with the desired phenotype can also be provided (see the Pluripotent stem cell example) but is not required for all use cases, such as simulating adaptive resistance to treatment (see the Colorectal cancer example). Optionally, the input can include mutational data to specify the rules for the simulations involving the network's mutated genes.

Step 1. Estimation of the attractor landscape. The goal of the first step of NETISCE is to estimate the region of the attractor landscape containing steady-states associated with the desired and undesired phenotype (Fig. 2b). The network is simulated using an adapted version of SFA⁴⁶. SFA estimates signal flow, the information conveyed by a series of reactions as represented in a signaling network or GRN, based only on topological information in the network and an initial state of the network nodes. The output of SFA is the logarithm of the attractor's expression value for each network node. The initial states of the network nodes are based on the normalized gene expression levels. We simulate the system using SFA for each provided experimental sample until reaching the attractor.

We generate randomly sampled initial states whose values fall in the ranges of the normalized expression values for each of the supplied phenotypes in the experimental data and apply SFA to compute a sufficiently large number of attractor states³⁴. All the computed attractors are then clustered via *k*-means clustering. The elbow and silhouette metrics are calculated to determine an optimal *k*^{51,52}. The clusters are also evaluated using the internal-marker node values.

Step 2. Virtual screenings on FVS control nodes. In this step, NETISCE identifies FVS control nodes and simulates combinations of perturbations to their activity (Fig. 2c). First, the FVS control nodes are identified via a

simulated annealing algorithm to determine the FVS of the network⁵³. This step can be performed exhaustively to identify all FVSes in a network. Then, virtual screenings of combinations of perturbations on the FVS control nodes are performed using the SFA algorithm. In these simulations, the initial states of the network nodes are set to the normalized expression values of the cells with an undesired phenotype. We have modified the SFA pipeline to implement overrides to control nodes to simulate overexpression, knockout, or no change to a node's activity.

Step 3. Filtering sets of perturbations on the FVS control nodes. The final step of the pipeline aims to identify the combinations of perturbations on the FVS control nodes that result in the desired cell fate reprogramming (Fig. 2c). We employ two filtering criteria to evaluate the combinations of perturbations on FVS control nodes. The first criterion uses Random Forest, Support Vector Machine, and Naive Bayes machine learning classification algorithms to classify the attractors generated by perturbations on FVS control nodes. In this classification step, using the previously clustered attractors by the k -means analysis, the attractors generated by perturbations on FVS control nodes are classified either in the cluster(s) associated with the undesired or desired phenotype^{54–56}. To pass this filtering criterion, an attractor generated from the perturbation on FVS control nodes must be classified to the cluster associated with the desired phenotype by at least 2 out of 3 classification algorithms. After passing the first criterion, perturbations on FVS control nodes are evaluated by the second filtering criterion, which focuses on the expression values of the internal-marker nodes in the attractors. Perturbations on FVS control nodes where at least 90% of the internal-marker node expression values are within the expression ranges of the attractors associated with the desired phenotype pass this filtering step. These criteria produce a final set of perturbations on FVS control nodes that are considered capable of reprogramming from an undesired cell fate toward the desired cell fate.

Estimation of steady-states using signal flow analysis

Signal flow analysis (SFA) is based on the Signal Propagation algorithm developed by Lee and Cho⁴⁷. The algorithm is a linear difference equation that computes the activity of a network node at a given time in terms of the state of the network node at the previous time step, the effect (activating or inhibiting), and the weight of the influence of its m incoming edges, and the initial state of the node. Precisely, the logarithm (\log_2) of the steady-state activity $x_i(t+1)$ of a node i at time $t+1$ is estimated by the initial state of the node and the activities of its regulators at time t using the following equation:

$$x_i(t+1) = \alpha \sum_{j=1}^m (W_{ij}x_j(t)) + (1-\alpha)b_i \quad (1)$$

where $x_j(t)$ is the logarithm of the activity of j , a node connected to i by an incoming edge, at time t . The W_{ij} is the weight of the edge between node j and node i , which represents how much influence node j exerts on node i through the edge. W_{ij} is defined as:

$$W_{ij} = \frac{\text{sign}(ij)}{\sqrt{(D_{\text{out}})_j(D_{\text{in}})_i}} \quad (2)$$

where $\text{sign}(ij)$ is the value of the edge between i and j (1 for activating edges, -1 for inhibiting edges), D_{out} is the out-degree of j , and D_{in} is the in-degree of i . Finally, b is the logarithm of the initial state of node i and α is a hyperparameter used to weigh the influence of the network structure and initial node state on the Signal Flow. By default, in our pipeline, the hyperparameter α is set to 0.9 to provide greater weight to the network topology rather than the initial activity based on the parameter settings used in previous control studies using SFA⁷⁴.

To identify an attractor of the system, the signal propagation equation is solved for all network nodes synchronously until the difference between $x(t+1)$ and $x(t)$ is less than a tolerance threshold (by default, this tolerance threshold is 10^{-6}).

The network nodes' expression values of two attractors under different simulation inputs can be compared by computing the difference between two logarithms of the expression values produced from SFA, analogously to a logarithm of the fold-change ($\log_2\text{FC}$) in differential gene expression analysis⁴⁷. Although the actual numerical value cannot be used to measure the magnitude of the change in expression, positive difference values indicate that the specified perturbation led to a shift in signal flow that increases the gene's activity at the steady state. In contrast, negative values predict decreased activity at the steady state due to the perturbation.

When estimating the attractor landscape, we begin by solving the signal propagation equation for the system using the experimental data for each sample; the initial activity of network nodes is set to the normalized expression values. In most cases, the number of experimental samples is insufficient for landscape estimation. Therefore, NETISCE can generate randomly sampled initial states (100,000 by default), whose initial state values are calculated from the ranges of the normalized expression values for each of the supplied phenotypes in the experimental data.

Association of the attractor landscape clusters to experimental phenotypes

We employ k -means clustering to partition the attractors estimated from the normalized expression data and the randomly generated initial states. We confirm that the attractors computed from the undesired and desired experimental samples are different. We use two metrics to determine the optimal number of k clusters. The first is the elbow metric, which determines the optimal k by finding the minimal intra-cluster variation⁸⁰. The second is the silhouette metric, which aims to identify the optimal k as the number of clusters with minimal intra-cluster variation and maximal inter-cluster variations⁸¹. When the two metrics disagree on the optimal k , the smallest of the potential optimal k -values is chosen where the attractors estimated from the undesired and desired phenotype experimental samples do not appear within the same cluster(s).

Finally, we use the internal-marker nodes to confirm that the expression values of these nodes agree with experimental data or literature. NETISCE checks that the expression value of each internal-marker node in the attractors associated with the undesired phenotype and desired phenotype matches the expected differential gene expression patterns. In the scenario where only the experimental data for the undesired phenotype was provided for the initial states, NETISCE verifies that the cluster(s) containing the attractors generated from the experimental data only have attractors where the internal-marker node expression values are within the range of expression values of the undesired phenotype. If experimental data for both the undesired and desired phenotype is supplied for initial states, then NETISCE confirms that the attractors generated from the two phenotypes do not appear in the same cluster and that their internal-marker node expression values are within the appropriate expression value ranges. For example, consider a gene with a higher expression in cells in the desired phenotype than the undesired phenotype. NETISCE will verify that the expression value in the attractor associated with the desired phenotype is greater than the value in the attractor associated with the undesired phenotype (i.e., that the difference between the expression value for the internal-marker node in the attractor associated with the desired phenotype and the undesired phenotype is positive). Also, when multiple samples are given for each phenotype, NETISCE verifies that the expression values of the internal-marker nodes in the attractors of the undesired phenotype do not overlap with the values in the attractors of the desired phenotypes. If an overlap occurs, the internal-marker node is unreliable for analysis to separate the attractors in the different phenotypes. Thus, if the values of the internal-marker nodes do not match the literature or do not separate well between the attractors of the undesired and desired phenotype clusters, the user may elect to revise network structure, remove specific internal-marker nodes, or adjust simulation settings.

Identification of the minimal feedback vertex set

Structure-based methods study the controllability of systems based solely on the structure of the network^{5,82,83}. In recent years, structure-based control methods for systems with non-linear dynamics have been proposed. One such structure-based control method for non-linear dynamics is the feedback vertex set control introduced by Mochizuki et al.^{5,6}. Feedback vertex set control is a structure-based control method focused on the controllability of the system by restricting the target states to attractors. Mochizuki et al. mathematically proved that for a network governed by non-linear dynamics like cell signaling, the control action of overriding the state variables of the FVS into a targeted desired trajectory ensures that the system will asymptotically approach the desired trajectory. Consider a directed graph $G=(V, E)$ comprised of node set V and edge set E . The node states of G are described by the ODE

$$\dot{x}_n = F_n(x_n, x_n), \quad n = 1, 2, \dots, N \quad (3)$$

where for the dynamics x of node $n \in V$, I_n is the set of nodes that regulate node n , such that self-regulatory loops ($n \in I_n$) are only positive. In addition, we assume F_n satisfies decay condition:

$$\partial_1(F_n(x_n, x_{I_n})) < 0 \quad (4)$$

for all n where ∂_1 is the partial derivative w.r.t. the first occurrence of x_n and not x_{I_n} .

Definition 1.1: In G , a subset $I \subseteq V$ of nodes is feedback vertex set (FVS) if and only if removal of set $G \setminus I$ leaves a graph without directed cycles. An FVS is minimal if it does not contain a proper subset that is an FVS itself. For simplicity, in this paper, we will consider all the FVSes to be minimal.

Definition 1.2: In a dynamic system, a subset $J \subseteq V$ of nodes is a set of determining nodes if and only if two solutions satisfy $\lim \tilde{x}_j(t) - x_j(t) \rightarrow 0$ whenever $\lim \tilde{x}_n(t) - x_n(t) \rightarrow 0$ for all components $\tilde{h} \in J \subseteq V$.

In Fielder et al. and Mochizuki et al., these two definitions were proven to be equivalent for dynamics in a network^{5,6}. Therefore, observation of the long-term dynamics of the FVS is sufficient to identify all possible attractors of an entire system. Controlling the dynamics of the FVS ($x_j^*(t) - x_j(t) \rightarrow 0$) is sufficient to drive the dynamics $x(t)$ of a whole system to converge on one of any attractors $x^*(t)$.

The minimal feedback vertex set problem is a well-known NP-hard problem. Many algorithms have been developed to find the near-minimum FVS. Based on the implementation in Zaňudo et al.⁸³, we use a simulated annealing local search approach, SA-FVSP, originally described in Galinier et al.⁸⁴. SA-FVSP has been shown to outperform the greedy adaptive search procedure⁸⁵. A network may have multiple FVSes depending on the size and structure, but each FVS has the same capabilities for controlling cell fates.

Simulating perturbations on FVS control nodes

After identifying the FVS control nodes for virtual screenings, combinations of perturbations (overexpression/upregulation, knockouts/downregulation, or no change) to an FVS node's activity are generated. NETISCE generates 3^n combinations of control nodes perturbations, where n is the number of FVS control nodes.

The initial state of a node not contained in the FVS or an FVS node whose perturbation is "no change" is set to the normalized expression value of the experimental sample(s) for the selected undesired phenotype.

To simulate the perturbations on FVS control nodes, we modified the SFA algorithm to override the activity of perturbed control nodes. Specifically, the values of the perturbed FVS control nodes are fixed and unaffected by the incoming signal flow. The fixed state p of an upregulated (downregulated) FVS control nodes i is defined as:

$$p_i^{\text{upregulated}} = \max(\text{normexp}_i) + 2.5(\max(\text{normexp}_i)) \quad (5)$$

$$p_i^{\text{downregulated}} = \min(\text{normexp}_i) - 2.5(\min(\text{normexp}_i)) \quad (6)$$

Where $\max(\text{normexp}_i)$ and $\min(\text{normexp}_i)$ are the maximum and minimum normalized expression values of i across the experimental samples of the undesired phenotype, respectively. These equations are also used when gain-of or loss-of-function information from mutational data is supplied to NETISCE. For example, the value of a node representing a gene with a gain-of-function mutation is fixed to the corresponding $p_i^{\text{upregulated}}$ based on the normalized gene expression data.

Classification of perturbations on FVS control nodes

To systematically identify which perturbations on the FVS control nodes shifted the system away from the attractor associated with the undesired phenotype and toward the desired phenotype, we filter the resultant attractors with two criteria. Criterion 1 considers a combination of perturbations on FVS control nodes successful if the perturbation's corresponding attractor is classified to the cluster associated with the desired phenotype by at least two of three machine learning classification algorithms. The three classification algorithms considered in NETISCE are Naive Bayes, Support Vector Machines, and Random Forest classifiers. Naive Bayes and Support Vector Machines are well suited for high dimensional datasets^{86,87}, while Random Forest Classifiers improve predictive accuracy and reduce over-fitting⁸⁸. For each algorithm, the training data is the set of attractors generated from the provided normalized expression data and the randomly generated initial states and their associated clusters identified in the k -means clustering step. For each attractor, the entire vector of all network nodes is supplied. Last, we perform a Feature Importance Analysis. We examine the important features (network nodes and their associated

steady-state values) used by each algorithm to classify the combinations of perturbations on FVS control nodes. We extract the top 10% of ranked features for each algorithm⁸⁹.

Criterion 2 focuses on the expression values of the internal-marker nodes. In this second criterion, the attractor obtained after simulating the system under the studied combination of perturbations on FVS control nodes must have at least 90% of the expression values of its internal-marker nodes within the expected expression value ranges of the desired attractors. This ensures that beyond the machine learning classification based on the entire attractor, the known biological internal-marker nodes have the expected values of the desired phenotype. By default, NETISCE is set to strict filtering criteria, where the expression values of the attractor produced by the perturbations on FVS control nodes must shift into the appropriate range of the desired phenotype expression values. For example, consider an internal-marker node whose expression value in the cluster of attractors associated with the desired phenotype is 2.0. The expression values in the attractors associated with the undesired phenotype are 1.0. The difference between the internal-marker node's expression values in the desired phenotype versus the undesired phenotype would be positive, indicating that the gene's activity is higher in the desired phenotype. For a perturbation on the FVS control nodes to pass the filtering criterion, the expression value of the internal-marker node must be >2.0 . Therefore, the difference between the internal-marker node's expression values in the attractor produced by the perturbation on FVS control nodes would be positive when compared to either the desired or undesired phenotype. Alternatively, the user can select a more relaxed filtering threshold. In this case, for the example described above, a perturbation on the FVS control nodes would pass the filtering criterion if the attractor expression value of the internal-marker node is >1.0 . All the perturbations to the FVS control nodes that pass both filtering criteria are considered to successfully shift the initial state to an attractor associated with the desired phenotype. If NETISCE is run with replicates for the undesired phenotype, then a perturbation on FVS nodes must pass the first filtering criteria on all replicates. All replicates are individually analyzed in the second filtering criterion, and NETISCE produces a separate list of perturbations that pass the criterion for each replicate. In our pluripotent stem cell example that contained three replicates, the perturbations on FVS control nodes that passed both criteria in the replicates were identical. However, users familiar with their data may be interested in perturbations that only work for a subset of replicates.

Data for developmental, stem cell, and cancer biology validations

Cell fate specification in the ascidian embryo. The network structure was obtained from Kobayashi et al.⁵⁵. Since the focus of this example was to reproduce the experimental results of embryonic cell fate specification using feedback vertex set control and SFA, we performed computations separate from the NETISCE pipeline but using the essential scripts (see GitHub repository and tutorial). Without available normalized gene expression data for the unperturbed embryo, we performed in silico simulations to reproduce the cell fate specification results with SFA. The attractor for an unperturbed embryo was estimated by setting the initial activities of two genes necessary for normal embryonic development, Gata.a and Zic-r.a, to 1, representing an activated state⁵⁶. All other nodes were initialized to 0, representing an initial inactive activity. The attractor estimation function simulated the seven perturbations to the FVS control nodes that induced the seven tissue fates experimentally: (Foxa.A, Foxd, Erk Signaling, Neurog, Tbx6-r.b, and Zic-r.b.). Specifically, in these simulations, Gata.a and Zic-r.a had initial activities set to 1 and all other nodes set to 0. Then, the values of the FVS control nodes were overridden using the $p_i^{\text{upregulated}}$ or $p_i^{\text{downregulated}}$ equations. Since there was no gene expression data, for all FVS control genes the $p_i^{\text{upregulated}} = 2.5$ and $p_i^{\text{downregulated}} = -2.5$. The internal-marker node expression values for the unperturbed and perturbations on FVS node simulation results can be found in Supplementary Table 1. A perturbation was considered successful in replicating the experimental results if the difference in the expression values between the specified internal-marker node for the relevant tissue in the attractor generated from a perturbation on FVS control nodes and the attractor generated from the unperturbed initial state was positive (Supplementary Table 4). These values were additionally graphed using radar plots to visualize the respective upregulations for each perturbation (Fig. 4).

Induced pluripotent stem cell reprogramming from primed to naive pluripotency. The intracellular signaling network for induced pluripotent

stem cell signaling was obtained from Yachie-Kinoshita et al.⁵⁷. The normalized expression data for EpiSCs and ESCs were downloaded from the Gene Expression Omnibus (GSE88928)⁹⁰. There were three replicates for each experimental sample. Each replicate was used separately as initial state values to simulate the network, compute their associated steady-states, and perform perturbations on the FVS control nodes. Initially, we selected as internal-marker nodes the four output nodes used in the Boolean Model of Yachie-Kinoshita et al.: Oct4, Sox2, Nanog, and EpiTFs. Although these three nodes were also FVS control nodes, they were used as internal-marker nodes to be consistent with the output nodes in the Boolean simulations. To further filter our perturbations, we selected additional internal-marker nodes from gene expression data provided by Yachie-Kinoshita et al.⁵⁷. Based on the gene expression data for the network nodes, there were six genes whose values differed significantly between the ESC and EpiSC states. These included Lefty1, Pitx2, and Esrrb. The three other genes, Tbx3, Gata6, and Klf4, were not included as internal-marker nodes as they were FVS control nodes. Radar plots were used to visualize the perturbations of Klf4 upregulation, Nanog upregulation, and the combined Klf4+Nanog upregulation (Fig. 6).

Overcoming adaptive resistance to MAPK inhibitory therapy in colorectal cancers. The colorectal cancer (CRC) tumorigenesis signaling network and annotated HT29 mutational profile for network nodes were provided by Park et al.⁵⁹. The RNA-seq from untreated HT29 cells was obtained from the Cancer Cell Line Encyclopedia (CCLE)⁹¹.

In this study of adaptive resistance to MAPKi therapy in CRC, the ultimate therapeutic goal is to decrease proliferation and increase apoptosis in tumor cells. In a method adapted from Beal et al.⁶⁰, we verify that CRC tumors' proliferation and apoptosis signatures are preserved under the SFA simulation of the generic CRC network for patient tumors (Supplementary Text 5).

The network was simulated using as initial conditions the normalized expression and mutational profile of an untreated HT29 as there was no available gene expression data for treated HT29 cells. As annotated in Park et al.⁵⁹, PIK3CA and BRAF have gain-of-function mutations, while APC, SMAD4, and TP53 have loss-of-function mutations in HT29 cells. Therefore, the values of these nodes were fixed to the appropriate $p_i^{\text{upregulated}}$ or $p_i^{\text{downregulated}}$ value. To simulate BRAFi (HT29_BRAFi), $p_i^{\text{downregulated}}$ override was applied to the state of BRAF. To simulate BRAFi+EGFRi (HT29_BRAFi+EGFRi), BRAF and EGFR had the appropriate overrides applied using the $p_i^{\text{downregulated}}$ equation. We used the FVS-finding algorithm to search for a sufficiently large number of FVS in the CRC network, which identified 68 FVSes of size 14. TP53 was removed from the FVSes since it was already fixed to its $p_i^{\text{downregulated}}$ value due to the loss-of-function mutation in HT29 cells (Supplementary Data 1). Based on feedback vertex set control theory⁵, all FVSes of a system can guide the system to any of its natural attractors; therefore, we randomly selected the first FVS identified by the algorithm to perform virtual screenings. For simulating perturbations on FVS control nodes, the system was initialized with the same parameters as the (HT29_BRAFi) simulation, with the additional perturbations to the FVS control nodes.

Perturbations on FVS control nodes were filtered first by the set of 3 internal-marker nodes used by Park et al.⁵⁹: CASP3, a marker of apoptosis, CCNE1 (also known as Cyclin E), a proliferation marker, and ERK (also known as MAPK1), a downstream molecule of the MAPK signaling pathway whose activity after BRAFi treatment indicates adaptive resistance. Potential additional internal-marker nodes were selected from downstream signaling elements of MAPK signaling, apoptosis-related, and proliferation-related pathways⁵⁹. These sets were filtered using the internal-marker node checking step of NETISCE to ensure that their expression values in the attractors associated with MAPK inhibitor therapy and MAPK inhibitor resistance were appropriate based on literature evidence. This resulted in an additional ten internal-marker nodes: three genes from apoptosis-related pathways (CASP9, MAPKAPK2, PPP2CA), two genes related to proliferation (CCND1, CDC25A), and five genes from MAPK signaling pathways (DUSP1, ELK1, HNF1B, MAPK8, MLK3). In order for a perturbation to pass the filtering criteria using the expanded set of internal-marker nodes, it must meet the 90% threshold for each phenotype's set (apoptosis, proliferation, and MAPK signaling) independently.

Evaluating the robustness to noise of NETISCE

In COPASI, we simulated the differential equation models using the Time Course function for the undesired and desired phenotypes. We additionally

simulate the time course for the undesired initial condition when a perturbation on FVS nodes is applied to ensure the system still arrives at the desired attractor. We injected seven levels of noise (1%, 5%, 10%, 20%, 30%, 40%, 50%) in the undesired and desired initial conditions using the Random Distribution item in the COPASI's Parameter Scan function. For each node with a nonzero initial concentration, the noisy initial condition was generated using a normal distribution, where the mean was the initial state of the node, and the standard deviation was 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, or 0.50, to simulate 1%, 5%, 10%, 20%, 30%, 40%, or 50% noise, respectively. We generated 1000 initial states for each noise level for the desired initial and undesired initial states.

In COPASI, von Dassow's Drosophila Segment Polarity Gene model simulations were computed using the deterministic LSODA Solver for 500 s when a steady-state was reached. The values of the FVS control nodes were set to the values determined by Zanudo et al. (Supplementary Table 1). The SDE model was extracted from Zhou et al. and simulated using the SDE solver. To implement the time-delay perturbations of MafA, Pdx1, Ngn3, Pax4 overexpression, and Ptf1a knockout in exocrine cells, we used the Event function to increase the production or degradation rates as performed in Zhou et al.

After completing the simulations in COPASI and generating sets of noisy initial conditions, we tested the ability of NETISCE to correctly identify that the specified perturbations on FVS control nodes can shift the undesired initial state away from the undesired attractor and toward the desired attractor. It is unlikely that an experiment contains 1000 biological replicates in real circumstances. Therefore, we generated 1000 subsets of three wild-type and three unpatterned initial states at each noise level, analogous to an experiment with three biological replicates of the wild-type phenotype and three biological replicates of the unpatterned phenotype. For each set of 1000 initial states at each noise level, we run NETISCE with default settings.

NETISCE implementation

The main computational scripts of our pipeline are written in Python, utilizing the extensively optimized machine learning algorithms of the Scikit Learn package⁸⁹. Scripts for analyzing the internal-marker node values are written in R. NETISCE is implemented as a Nextflow workflow⁹². Nextflow is a state-of-the-art workflow manager tool that is language agnostic and designed for parallel processing as a dataflow manager. Checkpoints are implemented for the user to investigate any possible errors or make changes to the run configuration. The code can easily be resumed without having to re-run all computations. We provide Nextflow pipelines for local machine use and high-performance cluster implementations. We also provide NETISCE within a Docker container to further enhance the reproducibility of NETISCE simulations⁹³. In addition to the command-line tool, our pipeline is available through the Galaxy Project. This cloud-based open-source tool requires little to no programming experience for biological analysis and workflows⁹⁴.

DATA AVAILABILITY

Data files are available at <https://github.com/VeraLiconaResearchGroup/Netisce>.

CODE AVAILABILITY

The NETISCE's Nextflow pipeline version and the Docker image documentation are available at <https://github.com/VeraLiconaResearchGroup/Netisce>. The installation, tutorials, information for installing the Galaxy Project version of NETISCE, and walkthroughs for reproducing the above results are found at <http://veraliconalab.org/Netisce/>.

Received: 30 December 2021; Accepted: 31 May 2022;

Published online: 20 June 2022

REFERENCES

- Liu, Y. Y., Slotine, J. J. & Barabási, A. L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
- Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Natl Acad. Sci. USA* **110**, 2460–2465 (2013).
- Müller, F.-J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478**, E4–E4 (2011).

4. Cornelius, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nat. Commun.* **4**, 1942 (2013).
5. Mochizuki, A., Fiedler, B., Kurosawa, G. & Saito, D. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. *J. Theor. Biol.* **335**, 130–146 (2013).
6. Fiedler, B., Mochizuki, A., Kurosawa, G. & Saito, D. Dynamics and control at feedback vertex sets. I: Informative and determining nodes in regulatory networks. *J. Dyn. Differ. Equ.* **25**, 563–604 (2013).
7. Cowan, N. J., Chastain, E. J., Vilhena, D. A., Freudenberg, J. S. & Bergstrom, C. T. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLoS ONE* **7**, e38398 (2012).
8. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
9. Pera, M. F. & Tam, P. P. L. Extrinsic regulation of pluripotent stem cells. *Nature* **465**, 713–720 (2010).
10. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
11. Gong, L. et al. Cancer cell reprogramming: a promising therapy converting malignancy to benignity. *Cancer Commun.* **39**, 48 (2019).
12. Xiong, S., Feng, Y. & Cheng, L. Cellular reprogramming as a therapeutic target in cancer. *Trends Cell Biol.* **29**, 623–634 (2019).
13. Zimmermannova, O., Caiado, I., Ferreira, A. G. & Pereira, C.-F. Cell fate reprogramming in the era of cancer immunotherapy. *Front. Immunol.* **12**, 2934 (2021).
14. Gómez Tejeda Zañudo, J., Scaltriti, M. & Albert, R. A network modeling approach to elucidate drug resistance mechanisms and predict combinatorial drug treatments in breast cancer. *Cancer Conver.* **1**, 5 (2017).
15. Gómez Tejeda Zañudo, J. et al. Towards control of cellular decision-making networks in the epithelial-to-mesenchymal transition. *Phys. Biol.* **16**, 031002 (2019).
16. Liu, H. et al. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res.* **48**, D871–D881 (2020).
17. Mandai, M. et al. Autologous induced stem-cell-derived retinal cells for macular degeneration. *N. Engl. J. Med.* **376**, 1038–1046 (2017).
18. Bosch, A. et al. PI3K inhibition results in enhanced estrogen receptor function and dependence in hormone receptor-positive breast cancer. *Sci. Transl. Med.* **7**, 283ra51 (2015).
19. Koh, H. W. L. et al. iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. *npj Syst. Biol. Appl.* **5**, 22 (2019).
20. Zhang, M. et al. GeNeCK: a web server for gene network construction and visualization. *BMC Bioinforma.* **20**, 12 (2019).
21. Shu, L. et al. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics* **17**, 874 (2016).
22. Zhou, G. & Xia, J. OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Res.* **46**, W514–W522 (2018).
23. Cahan, P. et al. CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).
24. Xu, Q. et al. ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *Nucleic Acids Res.* **49**, 7966–7985 (2021).
25. Jia, D., Jolly, M. K., Kulkarni, P. & Levine, H. Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. *Cancers* **9**, 70 (2017).
26. Greene, J. M., Gevertz, J. L. & Sontag, E. D. Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. *JCO Clin. Cancer Inform.* **3**, 1–20 (2019).
27. Tyson, J. J., Chen, K. C. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231 (2003).
28. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**, 437–467 (1969).
29. Ferrell, J. E. Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* **22**, R458–R466 (2012).
30. Choi, S. R., Hwang, C. Y., Lee, J. & Cho, K. H. Network analysis identifies regulators of basal-like breast cancer reprogramming and endocrine therapy vulnerability. *Cancer Res.* **82**, 320–333 (2022).
31. Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* **20**, 869–876 (2009).
32. Cho, S. H., Park, S. M., Lee, H. Y. H. S. & Cho, K. H. Attractor landscape analysis of colorectal tumorigenesis and its reversion. *BMC Syst. Biol.* **10**, 96 (2016).
33. Kim, Y., Choi, S., Shin, D. & Cho, K. H. Quantitative evaluation and reversion analysis of the attractor landscapes of an intracellular regulatory network for colorectal cancer. *BMC Syst. Biol.* **11**, 45 (2017).
34. Cho, K. et al. The reverse control of irreversible biological processes. *WIREs Syst. Biol. Med.* **8**, 366–377 (2016).
35. Zhou, J. X., Bruschi, L. & Huang, S. Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model. *PLoS One* **6**, e14752 (2011).
36. Folguera-Blasco, N., Cuyàs, E., Menéndez, J. A. & Alarcón, T. Epigenetic regulation of cell fate reprogramming in aging and disease: a predictive computational model. *PLoS Comput. Biol.* **14**, e1006052 (2018).
37. Jung, S., Appleton, E., Ali, M., Church, G. M. & del Sol, A. A computer-guided design tool to increase the efficiency of cellular conversions. *Nat. Commun.* **12**, 1659 (2021).
38. Rackham, O. et al. Challenges for computational stem cell biology: a discussion for the field. *Stem Cell Rep.* **16**, 3–9 (2021).
39. Melton, C., Judson, R. L. & Blöchl, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621–626 (2010).
40. Felsher, D. W. Reversibility of oncogene-induced cancer. *Curr. Opin. Genet. Dev.* **14**, 37–42 (2004).
41. Ronquist, S. et al. Algorithm for cellular reprogramming. *Proc. Natl Acad. Sci. USA* **114**, 11832–11837 (2017).
42. Jiang, J. & Lai, Y. C. Irrelevance of linear controllability to nonlinear dynamical networks. *Nat. Commun.* **10**, 3961 (2019).
43. Zañudo, J. G. T. & Albert, R. Cell fate reprogramming by control of intracellular network dynamics. *PLoS Comput. Biol.* **11**, e1004193 (2015).
44. Ferrell, J. E. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14**, 140–148 (2002).
45. Pomeroy, J. R. Uncovering mechanisms of bistability in biological systems. *Curr. Opin. Biotechnol.* **19**, 381–388 (2008).
46. Qi, K. et al. Expression and significance of IKBKB in pulmonary adenocarcinoma A549 cells and its Cisplatin-resistant variant A549/DDP. *Chin. J. Lung Cancer* **17**, 363–368 (2014).
47. Lee, D. & Cho, K. H. Topological estimation of signal flow in complex signaling networks. *Sci. Rep.* **8**, 5262 (2018).
48. Arakelyan, A., Nersisyan, L., Johnson, G., Riel-Mehan, M. & Pico, A. PSFC: a pathway signal flow calculator app for cytoscape. *F1000Research* **4**, 480 (2017).
49. Santolini, M. & Barabási, A. L. Predicting perturbation patterns from the topology of biological networks. *Proc. Natl Acad. Sci. USA* **115**, E6375–E6383 (2018).
50. Vanunu, G., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* **6**, e1000641 (2010).
51. Lee, H.-Y. et al. Identifying molecular targets for reverse aging using integrated network analysis of transcriptomic and epigenomic changes during aging. *Sci. Rep.* **11**, 12317 (2021).
52. Shah, O. S. et al. ATLANTIS—attractor landscape analysis toolbox for cell fate discovery and reprogramming. *Sci. Rep.* **8**, 3554 (2018).
53. Marazzi, L., Gainer-Dewar, A. & Vera-Licona, P. OCSANA+: optimal control and simulation of signaling networks from network analysis. *Bioinformatics* **36**, 4960–4962 (2020).
54. Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A. & Satou, Y. Controlling cell fate specification system by key genes determined from network structure. *iScience* **4**, 281–293 (2018).
55. Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A. & Satou, Y. Using linkage logic theory to control dynamics of a gene regulatory network of a chordate embryo. *Sci. Rep.* **11**, 4001 (2021).
56. Satou, Y. & Imai, K. S. Gene regulatory systems that control gene expression in the Ciona embryo. *Proc. Jpn. Acad. Ser. B* **91**, 33–51 (2015).
57. Yachie-Kinoshita, A. et al. Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. *Mol. Syst. Biol.* **14**, 1–16 (2018).
58. Hara, M. R., Sachs, B. D., Caron, M. G. & Lefkowitz, R. J. Pharmacological blockade of a β 2AR- β -arrestin-1 signaling cascade prevents the accumulation of DNA damage in a behavioral stress model. *Cell Cycle* **12**, 219–224 (2013).
59. Park, S.-M. M., Hwang, C. Y., Choi, J., Joung, C. Y. & Cho, K.-H. H. Feedback analysis identifies a combination target for overcoming adaptive resistance to targeted cancer therapy. *Oncogene* **39**, 3803–3820 (2020).
60. Beal, J., Montagud, A., Traynard, P., Barillot, E. & Calzone, L. Personalization of logical models with multi-omics data allows clinical stratification of patients. *Front. Physiol.* **10**, 1965 (2019).
61. Schroyer, A. L., Stimes, N. W., Abi Saab, W. F. & Chadee, D. N. MLK3 phosphorylation by ERK1/2 is required for oxidative stress-induced invasion of colorectal cancer cells. *Oncogene* **37**, 1031–1040 (2018).
62. Paton, E. L., Turner, J. A. & Schlaepfer, I. R. Overcoming resistance to therapies targeting the MAPK pathway in BRAF-mutated tumours. *J. Oncol.* **2020**, 1079827 (2020).
63. Corcoran, R. B. et al. EGFR-mediated reactivation of MAPK signaling contributes to insensitivity of BRAF-mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* **2**, 227–235 (2012).
64. Ding, C. B., Yu, W. N., Feng, J. H. & Luo, J. M. Structure and function of Gab2 and its role in cancer. *Mol. Med. Rep.* **12**, 4007–4014 (2015).

65. Herr, R. et al. BRAF inhibition upregulates a variety of receptor tyrosine kinases and their downstream effector Gab2 in colorectal cancer cell lines. *Oncogene* **37**, 1576–1593 (2018).
66. Bier, D. et al. Small-molecule stabilization of the 14-3-3/Gab2 protein-protein interaction (PPI) interface. *ChemMedChem* **11**, 911–918 (2016).
67. Vasilevskaya, I. A., Selvakumaran, M., Roberts, D. & O'Dwyer, P. J. JNK1 inhibition attenuates hypoxia-induced autophagy and sensitizes to chemotherapy. *Mol. Cancer Res.* **14**, 753–763 (2016).
68. Du, L. et al. JNK2 is required for the tumorigenic properties of melanoma cells. *ACS Chem. Biol.* **14**, 1426–1435 (2019).
69. Hoops, S. et al. COPASI—a complex pathway simulator. *Bioinformatics* **22**, 3067–3074 (2006).
70. von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. The segment polarity network is a robust developmental module. *Nature* **406**, 188–192 (2000).
71. Erbe, R., Gore, J., Gemmill, K., Gaykalova, D. A. & Fertig, E. J. The use of machine learning to discover regulatory networks controlling biological systems. *Mol. Cell* **82**, 260–273 (2022).
72. Dasgupta, B., Vera-Licona, P. & Sontag, E. in *Algorithms in Computational Molecular Biology* 941–953 (John Wiley & Sons, Inc., 2010).
73. Nguyen, H., Tran, D., Tran, B., Pehlivan, B. & Nguyen, T. A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. *Brief. Bioinform.* **22**, bbaa190 (2021).
74. Lee, D. & Cho, K.-H. Signal flow control of complex signaling networks. *Sci. Rep.* **9**, 14289 (2019).
75. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
76. Yang, G., Zaňudo, J. G. T. & Albert, R. Target control in logical models using the domain of influence of nodes. *Front. Physiol.* **9**, 454 (2018).
77. Papp, B. & Plath, K. Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape. *Cell Res.* **21**, 486–501 (2011).
78. Cotto, K. C. et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* **46**, D1068–D1073 (2018).
79. Sheils, T. K. et al. TCRD and Pharos 2021: mining the human proteome for disease biology. *Nucleic Acids Res.* **49**, D1334–D1346 (2021).
80. Satopää, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a 'kneedle' in a haystack: detecting knee points in system behavior. in *Proceedings - International Conference on Distributed Computing Systems* 166–171 (IEEE, 2011).
81. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
82. Ching-Tai Lin. Structural controllability. *IEEE Trans. Autom. Contr.* **19**, 201–208 (1974).
83. Zaňudo, J. G. T., Yang, G. & Albert, R. Structure-based control of complex networks with nonlinear dynamics. *Proc. Natl Acad. Sci. USA* **114**, 7234–7239 (2017).
84. Galinier, P., Lemamou, E. & Bouzidi, M. W. Applying local search to the feedback vertex set problem. *J. Heuristics* **19**, 797–818 (2013).
85. Pardalos, P. M., Qian, T. & Resende, M. G. C. A greedy randomized adaptive search procedure for the feedback vertex set problem. *J. Comb. Optim.* **2**, 399–412 (1998).
86. Zhang, H. & Zhang, H. The optimality of naïve bayes. in *FLAIRS Conference* (2004).
87. Platt, J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classif.* **10**, 61–74 (1999).
88. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
89. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
90. Kurek, D. et al. Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Rep.* **4**, 114–128 (2015).
91. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
92. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
93. Merkel, D. Docker: lightweight linux containers for consistent development and deployment. *Linux J.* **2014**, 2 (2014).
94. Jalili, V. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* **48**, W395–W402 (2021).

ACKNOWLEDGEMENTS

We thank Pedro Mendes and Ion Moraru for their insightful discussions and feedback in the methodology and validation of NETISCE. We additionally thank Pedro Mendes for his help with the COPASI simulations of the differential equations models. S.B. was supported by the UConn Health Research Program; A.P. and M.S. worked on this project under the UConn High School Mentored Research Program.

AUTHOR CONTRIBUTIONS

P.V.-L. and L.M. worked on the conceptualization and planning of this work. L.M., M.S., and A.P. performed and analyzed the validation simulations. L.M., S.B., and M.S. participated in the coding and implementation of NETISCE in Nextflow and Galaxy Project. L.M. and P.V.-L. drafted the manuscript. All authors read and approved 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/s41540-022-00231-y>.

Correspondence and requests for materials should be addressed to Paola Vera-Licona.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022