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.
controls the activity of its regulators, the type of regulatory relationship (activation or inhibition), and the initial state of the node (see the "Methods" section).

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% accuracy. Alternatively, SFA was applied to an aging-related gene regulatory network (GRN) to identify potential aging reversion targets. 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.

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. 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. 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. We 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

Associated to an observable cell fate. 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.
development by setting the initial activities of two genes necessary for normal embryonic development, Gata.a and Zic-r, 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 induce the embryonic tissue fates were estimated using SFA. To evaluate the perturbations, we analyzed the attractors generated from the initial state values of the EpiSC cells and ESCs. The second cluster included the attractors generated from the initial state values of the EpiSC cells and ESCs. The third cluster included the attractors generated from the initial state values of the EpiSC cells and ESCs.

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 inputs. The system was simulated under a stochastic asynchronous update 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 100,006 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 method 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...
simulated the 729 combinations of perturbations on the FVS control nodes. Of the 729 perturbations, 375 passed the machine learning classification 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 extra-cellular signaling information to positively regulate downstream Nanog transcription, and overexpression of Nanog was found to rescue pluripotency in the case of Klf4 knockdown16. 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...
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).

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.

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 gene 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
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 cells65,66. Finally, MAPK8 is a member of...
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 COPASI to simulate two differential equation models of cell reprogramming and add different noise levels into some initial states. We used 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 genes and a stochastic DE (SDE) model of cell fate differentiation in pancreatic cells, 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-cells. 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).

DISCUSSION

With the rise in availability of multi-omics datasets and tools for constructing gene regulatory and intracellular signaling networks from these data, 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 dynamics. It is important to note that FVS control

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

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.
Kobayashi and colleagues showed experimentally if a cell et al.55, where FVS genes were permanently overexpressed or Cho74, where edge modifications. Unlike the SFA control method introduced by Lee and we observe when the system has shifted toward the attractor the original SFA algorithm47 to apply FVS control-based perturbations associating phenotypes to the attractor landscape via nodes required for cell fate reprogramming by estimating system fi revision should be performed. nodes are denoted in the second column. state cannot be reached by perturbations to the FVS control nodes, then the FVS was not correctly identi, 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 algorithm57 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 Database.75 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 development.56 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. 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 (x0) 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 experimentally54,55, 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.
the specific timings of perturbations for successful cellular reprogramming.\textsuperscript{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 activity. 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 + TSC1 thr could further increase sensitivity to treatment and the rate of apoptosis by MAPK and mTOR signaling inhibition. Alternatively, BRAFi + SRCi + GRB2i can increase sensitivity to MAPKi therapy and prevent the metastatic spread of CRC tumors. Finally, BRAFi + SRCi + MAPKi may also increase sensitivity to MAPKi therapy and drive cells toward apoptosis.

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.

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 simulations. 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 colleagues. 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 efficiency, 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%.

### 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 SFA. 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.

**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 network. This step can be performed exhaustively to identify all FVSes in a network. Then, virtual screenings of combinatorial 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. 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 Choo. 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 incoming edge. The state of the network node at the previous time step, the effect that computes the activity of a network node at a given time in terms of the incoming edge, at time $t$. $W_{ij}$ 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:

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

where $\text{sign}(ij)$ is the value of the edge between $i$ and $j$ (1 for activating edges, -1 for inhibiting edges), $D_{\text{out}}$ is the out-degree of $j$, and $D_{\text{in}}$ is the in-degree of $i$. Finally, $b$ is the logarithm of the initial state of node $i$ and $s$ is a hyperparameter used to weight the influence of the network structure and the initial state on the Signal Flow. By default, in our pipeline, the hyperparameter $s$ 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 SFA.

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 ($\log_2(FC)$) in differential gene expression analysis. 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 variation. 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 variation.

**Identification of the minimal feedback vertex set**

Structure-based methods study the controllability of systems based solely on the structure of the network. 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. 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:

$$x_n = F_n(x_n, \ldots, x_1), \quad n = 1, 2, \ldots, N$$

L. Marazzi et al.

Published in partnership with the Systems Biology Institute

npj Systems Biology and Applications (2022) 21
where for the dynamics $x$ of node $n \in V$, $f_n$ is the set of nodes that regulate node $n$, such that self-regulatory loops ($n \in h_n$) are only positive. In addition, we assume $f_n$ satisfies decay condition:

$$\partial_t (F_n(x_n, x_i) < 0$$

for all $n$ where $\partial_t$ is the partial derivative w.r.t. the first occurrence of $x_n$ and not $x_i$.

**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 $\lim (x_j(t) - x_i(t) \rightarrow 0$ whenever $\lim (x_j(t) - x_i(t) \rightarrow 0$ for all components $h \subseteq J \subseteq V$.

In Fielder et al. and Mochizuki et al., these two definitions were proven to be equivalent for dynamics in a network. 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 $(x_i(t) - x_i(t) = 0$ 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 Žanud et al., we use a simulated annealing local search approach, SA-FVSP, originally described in Galinier et al. SA-FVSP has been shown to outperform the greedy adaptive search procedure. 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° 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:

$$p^\text{regulated}_i = \max(\text{normexp}) + 2.5(\max(\text{normexp}))$$

$$p^\text{downregulated}_i = \min(\text{normexp}) - 2.5(\min(\text{normexp}))$$

Where $\max(\text{normexp})$ and $\min(\text{normexp})$ are the maximum and minimum normalized expression values of $i$ across the experimental samples of the undesired phenotype. 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 $p^\text{regulated}_i$ 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 Naïve Bayes, Support Vector Machines, and Random Forest classifiers. Naïve Bayes and Support Vector Machines are well suited for high dimensional datasets, while Random Forest Classifiers improve predictive accuracy and reduce over-fitting. 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 algorithm.

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 lenient 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.. 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 embryonic samples, 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 state. 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 (FosxA.A, Foxd, Erk Signaling, Neuro, 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 $p^\text{regulated}_i$ or $p^\text{downregulated}_i$ equations. Since there was no gene expression data for all FVS control genes, the $p^\text{regulated}_i$ and $p^\text{downregulated}_i$ were set to 2.5 and $-2.5$. 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.\textsuperscript{57}. The normalized expression data for EpiSCs and ESCs were downloaded from the Gene Expression Omnibus (GSE88928)\textsuperscript{38}. 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 set of 68 nodes identified 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.\textsuperscript{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 Kif4, were not included as internal-marker nodes as they were FVS control nodes. Radar plots were used to visualize the perturbations of Kif4 upregulation, Nanog upregulation, and the combined Kif4+ 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.\textsuperscript{59}. The RNA-seq from untreated HT29 cells was obtained from the Cancer Cell Line Encyclopedia (CCLE)\textsuperscript{81}. 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.\textsuperscript{90}, 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 cell as there was no available gene expression data for treated HT29 cells. As annotated in Park et al.\textsuperscript{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 downregulated or upregulated value. To simulate BRAFi (HT29_BRAFi), $p_{i}$downregulated or $p_{i}$upregulated override was applied to the state of BRAF. To simulate BRAFi+EGRFi (HT29_BRAFi+EGRFi), BRAF and EGRF had the appropriate overrides applied using the $p_{i}$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 $p_{i}$upregulated value due to the loss-of-function mutation in HT29 cells (Supplementary Data 1). Based on feedback vertex set control theory\textsuperscript{47}, all FVSes of the 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 control nodes. Perturbations on FVS control nodes were filtered first by the set of 3 internal-marker nodes used by Park et al.\textsuperscript{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 based on literature evidence. This resulted in an additional ten internal-marker nodes: two genes related to proliferation (CCND1, CDC25A), and five 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 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 value 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 package\textsuperscript{85}. Scripts for analyzing the internal-marker node values are written in R. NETISCE is implemented as a Nextflow workflow\textsuperscript{92}. 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 reproduced 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 simulations\textsuperscript{93}. 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 workflows\textsuperscript{84}.

DATA AVAILABILITY Data files are available at https://github.com/VeraLiconaResearchGroup/Netisce.

CODE AVAILABILITY The NETISCE\textsuperscript{'}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

1. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. Nature 473, 167–173 (2011).
2. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. Proc. Natl Acad. Sci. USA 110, 2460–2465 (2013).
3. 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., Kurasawa, G. & Saito, D. Dynamics and control at feedback vertex sets. It is 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., Kurasawa, G. & Saito, D. Dynamics and control at feedback vertex sets. J. Informative and determining nodes in regulatory networks. J. Dyn. Differ. Equ. 25, 563–604 (2013).
7. Cowan, N. J., Chastain, E. J., Vlihena, 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 reprograming: a promising therapy converting malignancy to benignity. Cancer Commun. 39, 48 (2019).
12. Xiaoning, S., Feng, Y. & Cheng, L. Cellular reprogramming as a therapeutic target in cancer. Trends Cell Biol. 29, 623–634 (2019).
13. Zimmermannova, O., Caiaido, 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., Scaltitni, M. & Albert, R. A network modeling approach to elucidate drug resistance mechanisms and predict combinational drug treatments in breast cancer. Cancer Cong. 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 combinational 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, 283ra151 (2015).
19. Koh, H. W. L. et al. OmiCpass: network-based integration of multimodal data for predictive subnetwork discovery. npj Syst. Biol. Appl. 5, 22 (2019).
20. Zhang, M. et al. GeoNeCK: a web server for gene network construction and visualization. BMC Bioinform. 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. OmicNet: 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. 47, 7959–7965 (2019).
25. Jia, D., Jolly, M. K., Kulkarni, P. & Levine, H. Phenoypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. Cancers 9, 70 (2017).
26. Greenee, J. M., Gevertz, J. L. & Sontag, E. D. Mathematical approach to differentiate of complex networks. Proc. Natl Acad. Sci. USA 115, 869–876 (2018).
27. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. PLoS Comput. Biol. 6, e1000641 (2010).
28. Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A. & Satou, Y. Controlling cell fate specification system by key genes determined from network structure. Sci. Rep. 8, 15120 (2018).
29. Marazzi, L., Gainer-Dewar, A. & Vera-Licona, P. OCSANA+: optimal control and simulation of signaling networks from network analysis. Bioinformatics 36, 4960–4962 (2020).
30. Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A. & Satou, Y. Using linkage logic theory to control dynamics of a genetic regulatory network of a chordate embryo. Sci. Rep. 11, 40001 (2021).
31. Satou, Y. & Imai, K. S. Gene regulatory systems that control gene expression in the Ciona embryo. Proc. Jpn. Acad. Ser. B 91, 31–51 (2015).
32. Yachie-Kinoshita, A. et al. Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. Mol. Syst. Biol. 14, 1–18 (2018).
33. Mara, H. R., Sachs, B. D., Caron, M. G. & Lefkowitz, R. J. Pharmacological blockade of a B2AR-arrestin-1 signaling cascade prevents the accumulation of DNA damage in a behavioral stress model. Cell Cycle 12, 219–224 (2013).
34. Park, S.-M. M., Hwang, C. Y., Choi, J., Joung, C. Y. & Cho, K.-H. Feedback analysis identifies a combination target for overcoming adaptive resistance to targeted cancer therapy. Oncogene 39, 3803–3820 (2020).
35. Beal, J., Montagud, A., Traynard, P., Barillot, E. & Calzone, L. Personalization of biological models with multi-omics data allows clinical stratification of patients. Front. Physiol. 10, 1965 (2019).
36. Schroyer, A. L., Stines, 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).
37. Paton, E. L., Turner, J. A. & Schlaper, I. R. Overcoming resistance to therapies targeting the MAPK pathway in RAF-mutated tumours. J. Oncol. 2020, 1079827 (2020).
38. Corcoran, R. B. et al. EGF-mediated reactivation of MAPK signaling contributes to insensitivity of RAF-mutant colorectal cancers to RAF inhibition with vemurafenib. Cancer Disco. 2, 227–235 (2012).
39. 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 cells. 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 the molecular signatures database hallmark gene set collection. Nucleic Acids Res. 48, W395–W402 (2020).

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., Zafrudio, 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. DGlibi 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. Zafrudio, 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 naive 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-Licosa.

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.