PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY
SYSTEMS BIOLOGY

Cell reprogramming design by transfer learning of functional transcriptional networks

Thomas P. Wytock[a,b] and Adilson E. Motter[a,b,c,d,e,1]

Edited by Yamini Dalal, National Cancer Institute, Bethesda, MD; received July 28, 2023; accepted January 26, 2024 by Editorial Board Member Tadatsugu Taniguchi

Recent developments in synthetic biology, next-generation sequencing, and machine learning provide an unprecedented opportunity to rationally design new disease treatments based on measured responses to gene perturbations and drugs to reprogram cells. The main challenges to seizing this opportunity are the incomplete knowledge of the cellular network and the combinatorial explosion of possible interventions, both of which are insurmountable by experiments. To address these challenges, we develop a transfer learning approach to control cell behavior that is pre-trained on transcriptomic data associated with human cell fates, thereby generating a model of the network dynamics that can be transferred to specific reprogramming goals. The approach combines transcriptional responses to gene perturbations to minimize the difference between a given pair of initial and target transcriptional states. We demonstrate our approach's versatility by applying it to a microarray dataset comprising >9,000 microarrays across 54 cell types and 227 unique perturbations, and an RNASeq dataset consisting of >10,000 sequencing runs across 36 cell types and 138 perturbations. Our approach reproduces known reprogramming protocols with an AUROC of 0.91 while innovating over existing methods by pre-training an adaptable model that can be tailored to specific reprogramming transitions. We show that the number of gene perturbations required to steer from one fate to another increases with decreasing developmental relatedness and that fewer genes are needed to progress along developmental paths than to regress. These findings establish a proof-of-concept for our approach to computationally design control strategies and provide insights into how gene regulatory networks govern phenotype.

biological networks | data-driven control | nonlinear dynamics | cell reprogramming

Significance

The lack of genome-wide mathematical models for the gene regulatory network complicates the application of control theory to manipulate cell behavior in humans. We address this challenge by developing a transfer learning approach that leverages genome-wide transcriptomic profiles to characterize cell type attractors and perturbation responses. These responses are used to predict a combinatorial perturbation that minimizes the transcriptional difference between an initial and target cell type, bringing the regulatory network to the target cell type basin of attraction. We anticipate that this approach will enable the rapid identification of potential targets for treatment of complex diseases, while also providing insight into how the dynamics of gene regulatory networks affect phenotype.

The major bottleneck in designing protocols to control cell behavior no longer lies with the availability of experimental tools to manipulate cellular dynamics, microenvironment, or genetics, but in the ability to triage the combinatorial explosion of possible interventions to rationally direct experimental efforts. Advances in synthetic biology are steadily increasing the breadth and precision of possible intervention tools, whether they be nanoparticles (1, 2) and minicells (3) for targeted drug delivery, CRISPR, and its variants for targeted perturbation of the genetic code (4) and cellular dynamics (5, 6), or immunotherapy-based approaches for cancer treatment (7–9).

The large corpus of potential interventions and combinations thereof make brute-force trial and error approaches too expensive and time-consuming to be feasible. Unlike engineered systems in which control theory provides an equation-based framework to design interventions (10, 11), biological systems are only beginning to attain genome-scale mathematical descriptions (12), while technical limitations constrain the number of actuable degrees of freedom (genes) to be much smaller than number of components to be controlled. These features present a challenge given that underactuated control is onerous even in physical systems that admit a closed-form mathematical description (13). As a result, the biological control problem is often relaxed to steering between natively stable states (14–17), rather than stabilizing natively unstable ones. Since transcriptomic measurements are the most frequently employed technique for querying the cell state, the formulation of data-driven control presented here requires only publicly available data and is robust against the high dimensionality, multi-cell averaging, and low temporal resolution typical of these data. Our goal is to design a general data-driven control approach tailored to these aspects of the data. Our approach contrasts with related control-theoretic formulations in the literature, which cannot be easily deployed as they usually require targeted experiments to design the controller (18), temporally rich data (19), individually resolved system trajectories (20, 21), and/or the availability of microscopic agent-based models (22). It also contrasts

Author contributions: T.P.W. and A.E.M. designed research; T.P.W. performed research; T.P.W. analyzed data; and T.P.W. and A.E.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. Y.D. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: motter@northwestern.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2312942121/-/DCSupplemental.

Published March 4, 2024.

with existing heuristic approaches to manipulate cell behavior, which can be categorized as network-based or annotation-based.

The network-based approaches require an explicit reconstruction of network interactions (15–17, 23, 24) and may additionally rely upon a description of the network dynamics (25, 26). On the other hand, the annotation-based approaches focus on whether specific transcription factors are highly expressed in the target state (27–29), without further considering network interactions. While each type of approach successfully addresses the problems for which they were conceived, they possess specific attributes that preclude their direct application to the present problem. Network-based approaches may be structural or dynamical. The structural approaches assume comprehensive knowledge of the network structure and are valid under a restricted set of dynamical relationships, whereas the dynamical approaches require laborious experimental validation, so they offer high reliability within a limited scope. In contrast, the annotation-based approaches downplay the role of gene-gene interactions and make qualitative predictions.

The control approach introduced here employs transfer learning on transcriptomic data to retain the strengths of both the network-based and the annotation-based approaches while addressing their limitations. Transfer learning entails pre-training on a broad-based dataset followed by the incorporation of application-specific data (30). In our case, we use broad-based gene expression and bulk RNA-seq datasets consisting of observations across a range of unperturbed cell types to pre-train a machine learning model that maps transcriptional states to cell type. Pre-training consists of calculating the gene-gene correlation matrix, decomposing the matrix into eigengenes—combinations of genes that vary approximately independently of one another (31, 32)—and selecting the eigengenes that best distinguish cell types. The eigengene selection is implemented by iteratively selecting the eigengene that minimizes the cross-validation error of a distance weighted k-nearest neighbors (KNN) model mapping gene expression to cell type, until the error stops decreasing. The KNN model plays the role of an objective function in our control approach, and the selected eigengenes capture the functional network of regulatory interactions between genes without the need to explicitly reconstruct the underlying network of biochemical interactions. The functional network of regulatory interactions stabilizes cell types, which can be identified as attractors of the regulatory dynamics because cells of a given type exhibit stable phenotypes with distinct expression profiles (33). Naturally, the empirical evidence for the existence of attractors does not require the specification of a dynamical model, which is consistent with our data-driven approach (11, 34, 35). Because the KNN model maps out regions of the transcriptional state space associated with each cell type, it can be interpreted as estimating the regions of this space in the neighborhood of the corresponding cell type attractors. These regions can extend beyond the attractors themselves due to stochasticity while nevertheless remaining in each case within the attraction basin of the cell type, which is the set of transcriptional states that would deterministically converge to the attractor.

Equipped with the pre-trained KNN model, we incorporate transcriptomic data associated with gene perturbations, which constitutes our application-specific data. Each gene perturbation is either a knockdown or an overexpression (typically of a single gene), and the transcriptomic data include associated mock-treated experiments serving as negative controls. Fig. 1*A* illustrates the mock-treated (filled symbols) and perturbed states (open symbols) for an overexpression (green) and a knockdown (blue). Arrows indicate the transcriptional response to the perturbation, defined as the mean difference in expression between the perturbed and mock-treated experiments. To identify which perturbations can predictively alter the transcriptional state from one cell type to another, we start from unperturbed states of an initial cell type and add the corresponding transcriptional perturbation responses until reaching the basin of attraction of the target cell type as inferred by the KNN model (Fig. 1*B*). In these predictions, the selected eigengenes are the same for all reprogramming tasks, and the selection of perturbations is made application-specific by scaling the projections of the initial-target state distance in the eigengene basis according to the transcriptional variance of the target cell type.

Results

Data Description. We apply our approach to human cells using a gene expression microarray dataset ("GeneExp") and an RNA-sequencing dataset ("RNASeq"), which are described in Table 1. Each dataset has a fixed set of measured genes $\mathcal{G}_D$ and a fixed set of selected eigengenes $\mathcal{F}_D^*$, where D labels the dataset. These datasets are partitioned into unperturbed cell states used for training the KNN model, perturbed states used for defining the transcriptional response matrix, and (in the case of the GeneExp dataset) reprogrammed states used for validating the predictions. The summary statistics for the partitions include the number of experiments N_D, the number of experimental series E_D, the set of cell types $\mathcal{C}_D$, and the set of perturbations $\mathcal{P}_D$.

Each partition serves a distinct role in our approach. In both datasets, the unperturbed partition consists of cells free from exogenous stimulation, drug treatment, and genetic knockdown or overexpression. These data are used for training our recently developed machine learning method, previously used to distinguish cell type (36), which selects the optimal set of eigengenes $\mathcal{F}_D^*$. This method produces a distance-weighted KNN model that maps the latent space to a vector indicating the probability of belonging to each of the $\mathcal{C}_D$ cell types. Here, cell type refers to the phenotypic characterization assigned to the cell sample based on histological and morphological characteristics. Thus,

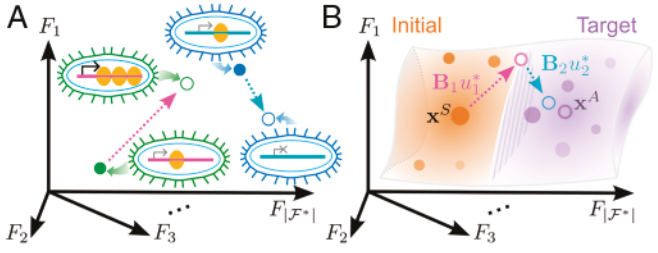

Fig. 1. Schematic overview of the data-driven control approach. (*A*) Construction of the library of transcriptional responses to gene perturbations in the latent space, which is defined as the subspace of selected eigengenes $\mathcal{F}^*$. The pink and teal arrows indicate the experimentally measured shift in transcription from a mock-treated state to a perturbed state (filled and empty circles, respectively) in different cell types (green and blue colors). (*B*) Perturbation optimization algorithm, where the goal is to drive the initial state $\mathbf{x}^S$ (orange filled circle, "S" for starting) to the target state $\mathbf{x}^A$ (open purple circle, "A" for attractor), which is the average of the individual states of the target cell type (filled purple circles). This is achieved by linearly combining the transcriptional responses to steer the system to a state (open teal circle) that minimizes the distance to the target. Within the algorithm, perturbation responses are added incrementally until the state is predicted to cross the cell type boundary (marked by the patterned surface) as determined by the KNN model. The order in which the incremental perturbations are selected within the algorithm does not imply a temporal ordering in the implementation of the perturbations.

Table 1. Statistics of the GeneExp and RNASeq datasets

	Genes, $\lvert\mathcal{G}_D\rvert$	Eigengenes, $\lvert\mathcal{F}_D^*\rvert$	Category	Profiles, N_D	Series, E_D	Cell types, $\lvert\mathcal{C}_D\rvert$	Perturbations, $\lvert\mathcal{P}_D\rvert$
GeneExp	17,525	4	Unperturbed	3,103	136	91	0
			Perturbed	5,735	356	368	207
			Reprogrammed	296	24	13	10
RNASeq	17,361	10	Unperturbed	9,851	1	36	0
			Perturbed	1,348	24	20	138

each transcriptomic measurement in this partition is identified with one cell type, implying that the aspects of the transcriptional state associated with this phenotype are sufficiently long-lived to be considered stable over the timescale of the experiment. Our KNN method can successfully infer cell type without explicitly reconstructing the regulatory network or the dynamical equations of the system (*SI Appendix*, Fig. S1). The ability of the KNN model to infer the behavior of high-dimensional regulatory networks using a latent space of much lower dimension without losing the relevant biological features may be interpreted as a by-product of the minimal frustration recently recognized in these networks (37).

The perturbed partition also applies to both datasets and consists of experimental series, which are sets of experiments associated with the same series-accession number in the Gene Expression Omnibus (GEO) database and are usually associated with a single study or publication. These series have transcriptional measurements of one or more gene knockdowns or overexpressions in addition to associated mock-treated experiments. The elements of the set $\mathcal{P}_D$ are metadata identifying the gene and kind of perturbation and are associated with a transcriptional response to that perturbation. The transcriptional states of genetic perturbations are regarded as steady states that generally persist only as long as the perturbation is induced, implying that the cell type remains unchanged. The final transcriptomic measurements of these states are usually taken 24 to 96 h after the initiation of the induction. The transcriptional responses derived from this partition are central to our data-driven control approach described in the next subsection.

The reprogrammed partition in the GeneExp dataset consists of experimental series associated with cell reprogramming experiments. Since reprogramming is used to refer to several processes in the literature, we clarify that in the remainder of the paper, we exclusively use the term reprogramming to refer to the process of transforming differentiated cells into a pluripotent state (i.e., embryonic stem cell-like state capable of redifferentiating into another cell lineage). When discussing our results, we use transdifferentiation to refer to changing the behavior of a differentiated cell without transitioning through a pluripotent state. Compared to perturbation experiments, reprogramming experiments involve more extensive passaging and selection to remove nonresponding cells, and the remaining cells do not generally return to their original cell type when the perturbation is removed. The reprogramming experiments in this partition serve as a validation set for our approach.

Data-Driven Control Approach. We now describe how we leverage the partitions of each dataset to arrive at our control approach. Suppressing the dataset label D, we refer to the transcriptional state in the full gene expression space and eigengene space using primed $\mathbf{x}' = (x_i') \in \mathbb{R}^{|\mathcal{G}|}$ and unprimed $\mathbf{x} = (x_i) \in \mathbb{R}^{|\mathcal{F}^*|}$ symbols, respectively. Revisiting Fig. 1*A*, each arrow corresponds to a column of the transcriptional response matrix, $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_{|\mathcal{P}|}) = (B_{ij}) \in \mathbb{R}^{|\mathcal{F}^*|\times|\mathcal{P}|}$, represented in the coordinates $\mathcal{F}^* = \{F_1, \ldots, F_{|\mathcal{F}^*|}\}$. Our approach to identify perturbations whose transcriptional responses facilitate the transitions between cell types finds the sum of an initial transcriptional state $\mathbf{x}^S$ and transcriptional responses (i.e., columns of $\mathbf{B}$ scaled by control inputs $\mathbf{u}$) that is as close as possible to the target $\mathbf{x}^A$ (Fig. 1*B*). The distance-weighted KNN model $\mathbf{K}(\mathbf{x})$ operates on transcriptional states to infer the probability of cell type membership and assigns a transcriptional state to the most probable cell type—as indicated by the orange and purple background. The control inputs $\mathbf{u} = (u_j) \in \mathbb{R}^{|\mathcal{P}|}$ scale the arrows to indicate the extent to which each perturbation is applied. Specifically, $u_j = 0$ means that the jth perturbation is inactive, while $u_j = 1$ means that this perturbation is active as in the data used to determine $\mathbf{B}_j$.

The control problem in Fig. 1*B* can be framed as an optimization problem:

$$\mathbf{u}^\dagger = \arg\min ||\mathbf{K}(\mathbf{x}^S + \mathbf{B}\mathbf{u}) - \mathbf{K}(\mathbf{x}^A)||_2, \qquad [1]$$

where $||\cdot||_2$ is the Euclidean distance between the probability vectors that are output by $\mathbf{K}$. The element u_j prescribes the extent to which the jth perturbation $\mathbf{B}_j$ is active. We consider three increasingly restrictive scenarios for the control inputs u_j: $-\infty < u_j < \infty$, $|u_j| \leq 1$, and $0 \leq u_j \leq 1$. In the first scenario, the control input can alter the initial state to any point on the line $\mathbf{x}^I + \mathbf{B}_j u_j$, while in the second and third scenarios, the range of achievable states is bounded by the magnitude and magnitude and direction of the measured transcriptional response, respectively. Moreover, solutions to Eq. **1** that require fewer perturbations are in principle easier to implement experimentally, suggesting a constraint $g = ||\mathbf{u}||_0$ (g nonzero elements in $\mathbf{u}$). These constraints make solving Eq. **1** expensive due to calculating numerical derivatives of $\mathbf{K}$.

To facilitate the identification of experimentally feasible $\mathbf{u}^*$, we approximate Eq. **1** as:

$$\mathbf{u}^* = \arg\min ||\mathbf{x}^S + \mathbf{B}\mathbf{u} - \mathbf{x}^A||_2. \qquad [2]$$

Eqs. **1** and **2** yield the same solution whenever the nearest neighbor to $\mathbf{x}^S + \mathbf{B}\mathbf{u}^*$ is substantially closer than the kth-nearest neighbor (*Materials and Methods*). Note that the approximation in Eq. **2** has the advantage of transforming a nonlinear and nonconvex optimization into one that is linear and convex. This is achieved by approximating the impact of multiple perturbations as their linear sum and by approximating differences in KNN-estimated probabilities as differences in transcriptional states. The former approximation implies that our approach does not temporally order the constituent perturbations within a combination. The latter approximation is consistent with the observed stability of cell type states, which guarantees that small

perturbations to the observed transcriptional states converge to the same attractor because otherwise the cell type would be unstable. Since measurements of a given cell type are in the neighborhood of the same attractor, the convex hull of the measurements tends to reside within the cell type basin of attraction. Eq. **2** can also be expressed as a constrained mixed-integer quadratic program, enabling us to take advantage of specialized software. Once $\mathbf{u}^*$ is obtained, $\mathbf{K}(\mathbf{x}^S + \mathbf{B}\mathbf{u}^*)$ is evaluated without approximations to determine whether the target has been reached.

Comparison with Existing Approaches. We benchmark our data-driven approach against existing approaches to identify candidate reprogramming perturbations using the $D =$ GeneExp dataset. Approaches that rely on network structure or dynamics are not applicable here due to the lack of a method to generate predicted transcriptional response from a reconstruction of the gene regulatory network. The remaining annotation-based approaches select perturbation candidates by compiling lists of genes that are significantly differentially expressed (DE) between initial and final states (27–29). These lists are used to identify statistically enriched annotations (38, 39), from which differences in pathway regulation and/or transcription factor binding are inferred. Annotation-based methods have shown promise in attributing changes to single transcription factors, but we demonstrate here that they have limited ability to predict the impact of combinations of factors.

We emulate these methods by assigning $u_j^{\mathrm{DE}} = x_j'^A - x_j'^S$ for all perturbations that are measured in the gene expression, i.e., $j \in \mathcal{P}_D \cap \mathcal{G}_D$. Using $d(\mathbf{x}^S, \mathbf{x}^A, \mathbf{u}; \mathbf{B})$ to represent the Euclidean distance on the right-hand side of Eq. **2**, we compare $\mathbf{u}^{\mathrm{DE}}$ with our method $\mathbf{u}^{\mathrm{OPT}} = \arg\min_{\mathbf{u}} d(\mathbf{x}'^S, \mathbf{x}'^A, \mathbf{u}; \mathbf{B}')$ under the three constraint scenarios described above. This is done using the coefficient of determination

$$R^2(\mathbf{u}; \mathbf{x}'^S, \mathbf{x}'^A, \mathbf{B}') = 1 - d(\mathbf{x}'^S, \mathbf{x}'^A, \mathbf{u}; \mathbf{B}')/d(\mathbf{x}'^S, \mathbf{x}'^A, \mathbf{0}; \mathbf{B}'), \quad [3]$$

where $R^2 \to 1$ as $\mathbf{x}'^S + \mathbf{B}'\mathbf{u}$ (the sum of the initial state and perturbation responses) approaches the target $\mathbf{x}'^A$ and $R^2 < 0$ if it is farther away. For each $\mathbf{x}^S$ in the unperturbed GeneExp partition, we obtain the $\mathbf{u}^{\mathrm{DE}}$ and $\mathbf{u}^{\mathrm{OPT}}$ using the mean expression of each cell type (different from that of $\mathbf{x}^S$) as $\mathbf{x}^A$. We then calculate Eq. **3** for each $\mathbf{u}^{\mathrm{DE}}$ and $\mathbf{u}^{\mathrm{OPT}}$ and take the median over all states in each initial cell type. Fig. 2 *A* and *B*, respectively, present these results as box-and-whisker plots across all target cell types for $\mathbf{u}^{\mathrm{DE}}$ and $\mathbf{u}^{\mathrm{OPT}}$. Surprised by the poor performance of the annotation-based methods given their ubiquity in the literature, we performed the same analysis for a single gene in *SI Appendix*, Fig. S2, which shows much better agreement between these methods and ours. We infer that the annotation-based method tends to move the state farther away from the target (i.e., $R^2 < 0$) because they have no way to calibrate the contributions of each individual perturbation in the sum, since off-target effects are not quantitatively accounted for. Our optimization-based approach, on the other hand, explicitly accounts for these effects, and as a result, it reduces the initial transcriptional difference by approximately 10% (when constraining $0 \le u_j \le 1$) to 25% (in the remaining cases). This recovery is 10 to 25 times larger than the fraction of perturbed genes ($|\mathcal{P}_D|/|\mathcal{G}_D| \approx 0.01$), which would be the expected recovery for perturbations that are randomly generated.

We next investigate whether $\mathbf{u}^{\mathrm{DE}}$ agrees qualitatively with $\mathbf{u}^{\mathrm{OPT}}$, despite the poor performance of the former. This is quantified using the sign alignment metric

$$\sigma(\mathbf{u}^{\mathrm{DE}}, \mathbf{u}^{\mathrm{OPT}}) = \frac{1}{|\mathcal{P}_D \cap \mathcal{G}_D|} \sum_{j=1}^{|\mathcal{P}_D \cap \mathcal{G}_D|} \mathrm{sgn}\left(u_j^{\mathrm{DE}}\right) \mathrm{sgn}\left(u_j^{\mathrm{OPT}}\right), \quad [4]$$

where $\mathrm{sgn}(x)$ is 1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. Eq. **4** is directly applied to $\mathbf{u}^{\mathrm{DE}}$ and $\mathbf{u}^{\mathrm{OPT}}$ in the first two constraint scenarios (red and green in Fig. 2), but it is applied to $2\mathbf{u}^{\mathrm{DE}} - 1$ and $2\mathbf{u}^{\mathrm{OPT}} - 1$ in the $0 \le u_j \le 1$ scenario (blue in Fig. 2). This transformation enables all three scenarios to have equal ranges. Fig. 2*C* shows that the annotation-based perturbations fail to agree with the optimization-based ones qualitatively in terms of the perturbation direction (first two scenarios) or identity (third scenario). Together, these results show that our approach can identify candidate perturbation combinations when annotation-based approaches cannot.

Data-Driven Reproduction of Existing Protocols. We next validate our approach by confirming that it can reproduce existing reprogramming protocols. A protocol consists of a set of perturbations that have been experimentally observed to drive a differentiated cell type to a pluripotent cell type. For this and all subsequent analyses, we project the data onto eigengenes and estimate cell type using the KNN models. Our validation dataset contains a set of 63 successful reprogramming protocols $\mathcal{R}$ and 220 other perturbations (*SI Appendix*, Fig. S3). We order the set of all perturbations $\mathcal{Q}$ according to the minimum distance achieved by the optimal single-gene perturbation under the three constraint scenarios averaged across initial-target pairs. The initial states $\mathbf{x}^S$ are drawn from a fixed differentiated cell type and the $\mathbf{x}^A$ are drawn from a fixed pluripotent cell type in given GEO series belonging to reprogrammed partition of the GeneExp dataset. Using $\mathcal{Q}_\ell$ to denote be the first ℓ elements of the set of perturbations, the true positive rate is $|\mathcal{R} \cap \mathcal{Q}_\ell|/|\mathcal{R}|$, and the false positive rate is $|\mathcal{Q}_\ell \backslash \mathcal{R}|/|\mathcal{Q}_\ell \cup \mathcal{Q}_\ell^C|$, where $\backslash$ is the set difference operator and $\mathcal{Q}_\ell^C$ denotes the set complement. Fig. 3 plots the true positive rate as a function of the false positive rate for $\ell \in \{1, \ldots, 283\}$ (i.e., the receiver operator characteristic curves) for each constraint case. As the constraints become more restrictive, the area under the curve increases. This trend indicates that the restriction of perturbation strengths to experimentally realizable values improves the identification of viable reprogramming strategies. Unlike in Fig. 2*B*, in which the unconstrained and $|u_j| \le 1$ scenarios produced nearly identical results, the $|u_j| \le 1$ and $0 \le u_j \le 1$ scenarios produce broadly similar results in Fig. 3. We attribute the difference between the first two scenarios to our approach's ability to quantitatively discriminate between candidate reprogramming perturbations based on the magnitude of their transcriptional response (i.e., $||\mathbf{B}_j||_2$). In other words, while many single-gene perturbations have the potential to move the cell state toward the target (pluripotency in this case), they have more limited impacts on gene expression than the overexpression of the Yamanka factors (KLF4, POU5F1, MYC, SOX2) (27). The similarity of predictive performance between the $|u_j| \le 1$ and $0 \le u_j \le 1$ scenarios suggests that the $|u_j| \le 1$ case can be useful for hypothesis generation in spite of the empirical observation that the impact of knocking down a gene is not an exact additive inverse of overexpressing one (40). Specifically, genes implicated in the $|u_j| \le 1$ case correspond to portions of the gene regulatory network that may be targeted

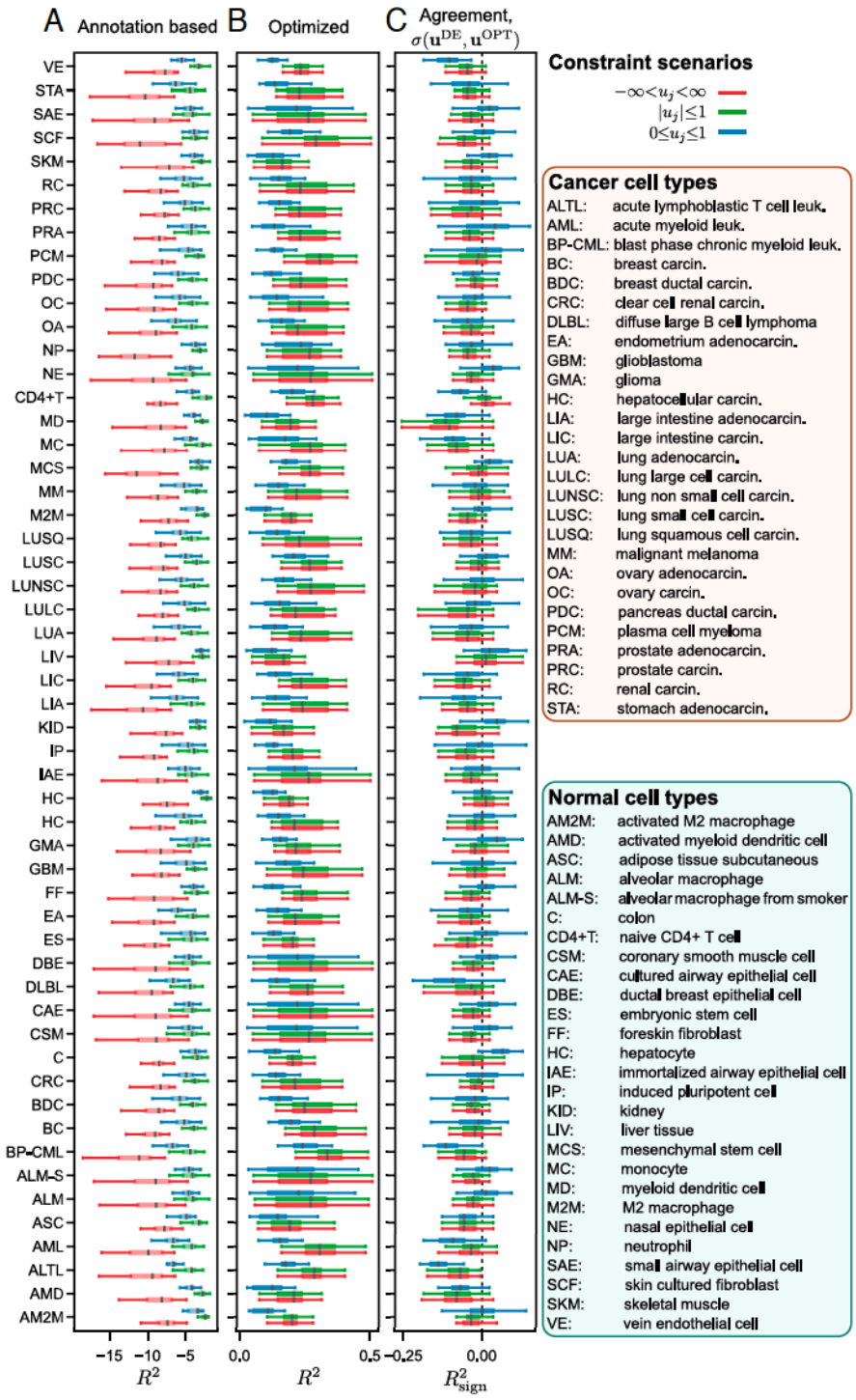

Fig. 2. Comparison of annotation-based methods, which do not account for off-target effects, with our control approach, which does. (*A*) Box-and-whisker plots of the coefficient of determination (R^2) of the perturbation predicted using annotation-based methods over all initial cell types for the unconstrained (red), size-constrained (green), and sign-constrained (blue) constraint scenarios applied to each target cell type. For each method, the left, center, and right of each box represent the 25th, 50th (median), and 75th percentiles of the distribution, respectively; the whiskers mark the minimum and maximum, excluding outliers, which are suppressed for clarity. (*B*) Results corresponding to those in *A* for our control approach. (*C*) Coefficient of determination of the sign of the optimal u_j in each method.

by new perturbation experiments to evaluate their utility for reprogramming.

We also compare our approach against those of an existing method, Mogrify (29). Unlike our approach, Mogrify does not take the initial state into account and only provides a ranked list of transcription factors, with the assumption that all transcription factors are overexpressed. We calculate all single-gene perturbations that facilitate reprogramming between the 214 initial cell types in the GeneExp dataset to the 54 tissues considered by Mogrify. Using the set of 71 overlapping transcription factors between knockouts in the GeneExp dataset and those in Mogrify, we compare the set of transcription factors

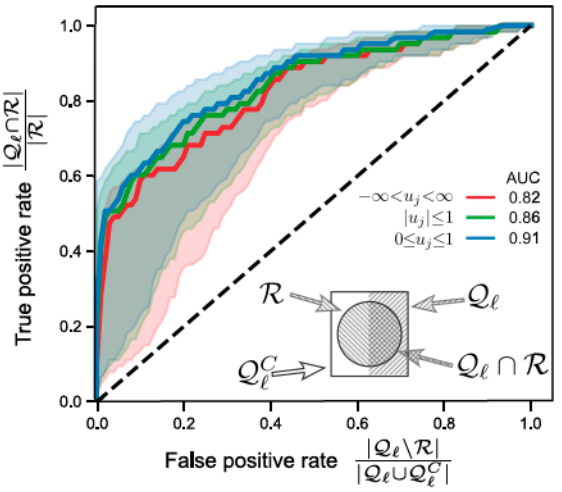

Fig. 3. Receiver operator characteristic (ROC) curves demonstrating the ability of our approach to reproduce known reprogramming protocols. The ROC curves are constructed by comparing single-perturbation strategies identified by our approach ($\mathcal{Q}$, upper diagonal hatching in the rectangle) ranked in order of their distance to the target (Eq. **2**) against 63 experimentally confirmed reprogramming protocols from the literature ($\mathcal{R}$, lower diagonal hatching in the circle). The sizes of $\mathcal{Q}$ and $\mathcal{R}$ and their overlap are characterized by the true positive rate and false positive rate as defined in the vertical and horizontal axis labels, respectively. The color-coded curves and backgrounds correspond to the median and interquartile range for the constraints indicated in the legend, including the median area under the curve (AUC).

predicted by our approach with those in the top 1% for the same target cell type in Mogrify. In all cases, we find at least one overlapping transcription factor between our predictions and those of Mogrify. However, we additionally identify other transcription factors in our predictions that reprogram a wider range of initial states, demonstrating that taking the initial state into account can generate more state-specific and broadly effective reprogramming strategies.

Analysis of Predicted Transdifferentiation Transitions. We next apply our approach to the GeneExp dataset and the RNASeq dataset and examine the transdifferentiation strategies it generates in each case. The states $\mathbf{x}^S$ from initial cell type s are drawn from the unperturbed partition in each dataset, while each $\mathbf{x}^A$ is the mean of all states in the partition belonging to a target cell type $a \neq s$. Constraining $0 \leq u_j \leq 1$, we determine the smallest number of applied perturbations g that reaches the target basin of attraction for each pair $\{\mathbf{x}^S, \mathbf{x}^A\}$. We create a transdifferentiation transition network from these results, in which the nodes are cell types and edges indicate the transdifferentiation transitions from s to a that are possible with g or fewer genes for more than a fraction f of the possible states $\mathbf{x}^S$ in the dataset. Fig. 4 shows the size of the largest strongly connected component (LSCC) of each transdifferentiation transition network as a function of g and f for the (*A*) RNASeq and (*B*) GeneExp datasets. For each value of f, there is a number of applied perturbations g for which $g \rightarrow g+1$ results in a rapid increase of the LSCC size, indicating a jump from fragmented subnetworks to a single giant component. Such a pattern is consistent with the hypothesis that few genes are necessary to reprogram cells (14) (a similar trend is observed for small increases in f for fixed g). The LSCC of the GeneExp dataset, for example, contains all cell types for only $g = 2$ and $f = 0.5$, meaning that it becomes possible to steer from one cell type to any other with $g \leq 2$ genes.

We illustrate the transdifferentiation transition network obtained at $g = 5$ and $f = 0.5$ in Fig. 5 (the circled instance in Fig. 4*A*), which is on the cusp of being fully connected. Our approach finds that most transdifferentiation transitions occur between related cell types, in accordance with observed developmental patterns. Cardiac, circulatory, fatty, skin, and to a lesser extent, neurological and digestive tissues illustrate this pattern. The main exceptions to this are reproductive or secretory tissues, which do not seem to preferentially transdifferentiate within their group.

Prominent Genes in Transdifferentiation Transitions. In addition to examining the pattern of transitions, we statistically test whether any gene perturbations are associated with reaching particular cell types. For each pair, we test whether the first-selected perturbation from a particular initial cell type to a particular target occurs more frequently than would be expected by chance. This expectation is set by the average of the observed frequencies of perturbations from the initial cell type to all other targets. Perturbations with frequencies exceeding this expectation are associated with exiting the initial cell type. Conversely, perturbations with frequencies exceeding the average frequencies observed for transitions into the target cell type from all initial cell types are associated with entering the target cell type (details in *Materials and Methods*).

We present these perturbations and associated cell types in *SI Appendix*, Fig. S4. While few perturbations appear to be associated with specific cell types, we find that digestive cell types share the long non-coding RNA (lncRNA) chromatin-associated transcript 10 (CAT10) (41), SYNCRIP, SULT2B1, and BEGAIN knockdowns between two digestive cell types. Furthermore, fatty tissues shared the double knockdown of lysine methyltransferases MLL1 (KMT2A) and MLL2 (KMT2D) and arterial tissues shared the lncRNA LINC00941. The prevalence of lncRNAs associated with transdifferentiation is consistent with recent experiments establishing the role of these factors in

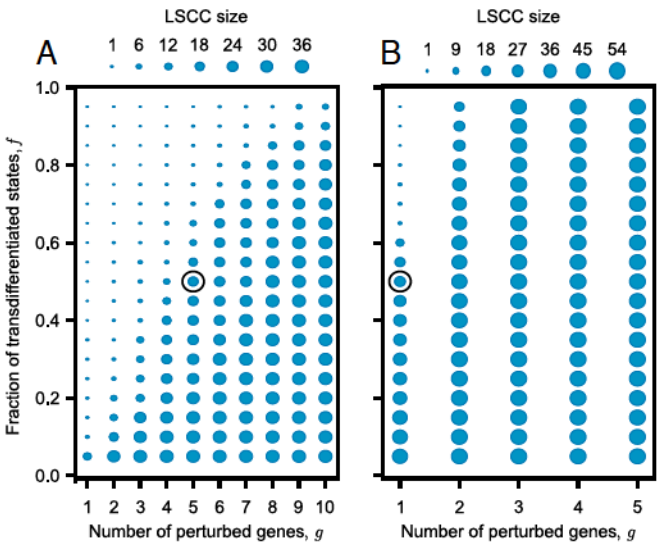

Fig. 4. Possible transdifferentiation transitions as a function of the number of genes perturbed and the fraction of successful transitions. (*A*) Largest strongly connected component sizes of the networks created when including an edge for each initial-target pair in the RNASeq dataset for which at least a fraction f of the initial states (vertical axis) are transdifferentiated using at most g perturbations (horizontal axis). (*B*) Corresponding results for the GeneExp dataset. The circled cases are considered further in subsequent figures.

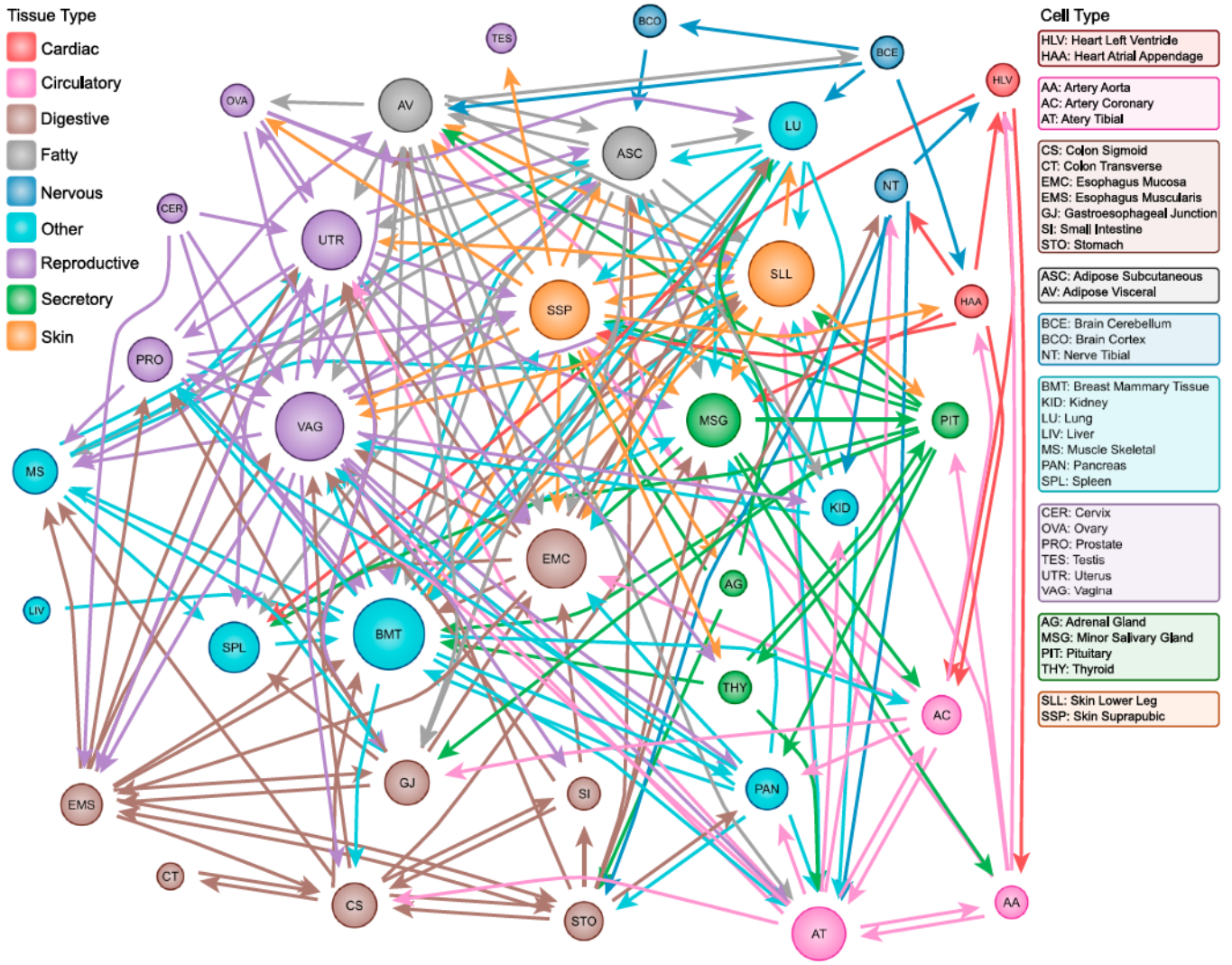

Fig. 5. Network of transitions (edges) between cell types (nodes) for the parameters indicated by the circle in Fig. 4*A*. The nodes and outgoing edges are color coded by tissue type. The node size increases with the total number of edges (i.e., the sum of incoming and outgoing edges).

determining cell fate (42). Interestingly, the knockdown of the translation initiation factor eIF4A1, which has been suggested to up-regulate expression of oncogenes (43), appears to facilitate the departure from the lower-leg skin to the suprapubic skin, which highlights the potential of using gene knockdowns to mitigate tissues that have accumulated damage.

Fig. 6 diagrams the pattern of transdifferentiation transitions in the GeneExp dataset, which contains a number of normal and cancerous tissues. Specifically, we observe that cancerous states tend to be reachable from normal states by a single gene, but not vice-versa. Of the 25 normal states, only 5 lie downstream of a cancerous cell type. In contrast, 12 of 29 cancerous cells are downstream of normal cells. These results are consistent with the observation that cancers tend to arise spontaneously but rarely resolve spontaneously. In *SI Appendix*, Fig. S5, the equivalent to *SI Appendix*, Fig. S4 for the GeneExp dataset, we find that the Yamanka factors play a central role in transdifferentiating between cell types. In addition, the prominence of multiple microRNA (MIR31$^+$, MIR34A$^-$), mechanotransduction (CDHR2$^-$, TWST1$^+$, VEGFC$^-$), and metabolic (IDH1$^-$, IDH2$^-$, ALDH1A1$^-$, ALDH3A1$^-$) gene perturbations is consistent with the observed interplay between cancer progression, mechanosensitivity (44), and metabolic reprogramming (45).

Analysis of the Required Number of Gene Perturbations. Having analyzed the networks of transdifferentiation transitions under the most restrictive case, we consider the impact of increasing g and relaxing constraints on $\mathbf{u}$. Fig. 7 *A* and *B* shows the fraction of transitions possible as a function of g for the three constraint scenarios. In the RNASeq (GeneExp) dataset, the fraction of transitions requiring only a single gene increases 3.5-fold (4.8-fold) when relaxing the sign constraint and 38-fold (9.6-fold) when relaxing all constraints. The more dramatic increase in Fig. 7*A* compared to Fig. 7*B* may reflect the greater precision of the RNASeq data, which is also reflected in the larger $|\mathcal{F}^*|$. Fig. 7*C* shows that the number of genes needed to transdifferentiate from normal to cancerous states is larger than for the reverse for the constraints $0 \leq u_j \leq 1$ in GeneExp dataset, quantifying the pattern illustrated in Fig. 6. These findings demonstrate that our control approach can not only predict candidate genes but also offer a framework for interpreting the relative stability of cell types on the basis of their gene expression. Indeed, stability can be characterized by the number genes that need to be perturbed to reach or to leave a given cell state, forging a connection with studies of cell type stability in the context of Boolean reconstructions of regulatory networks (46).

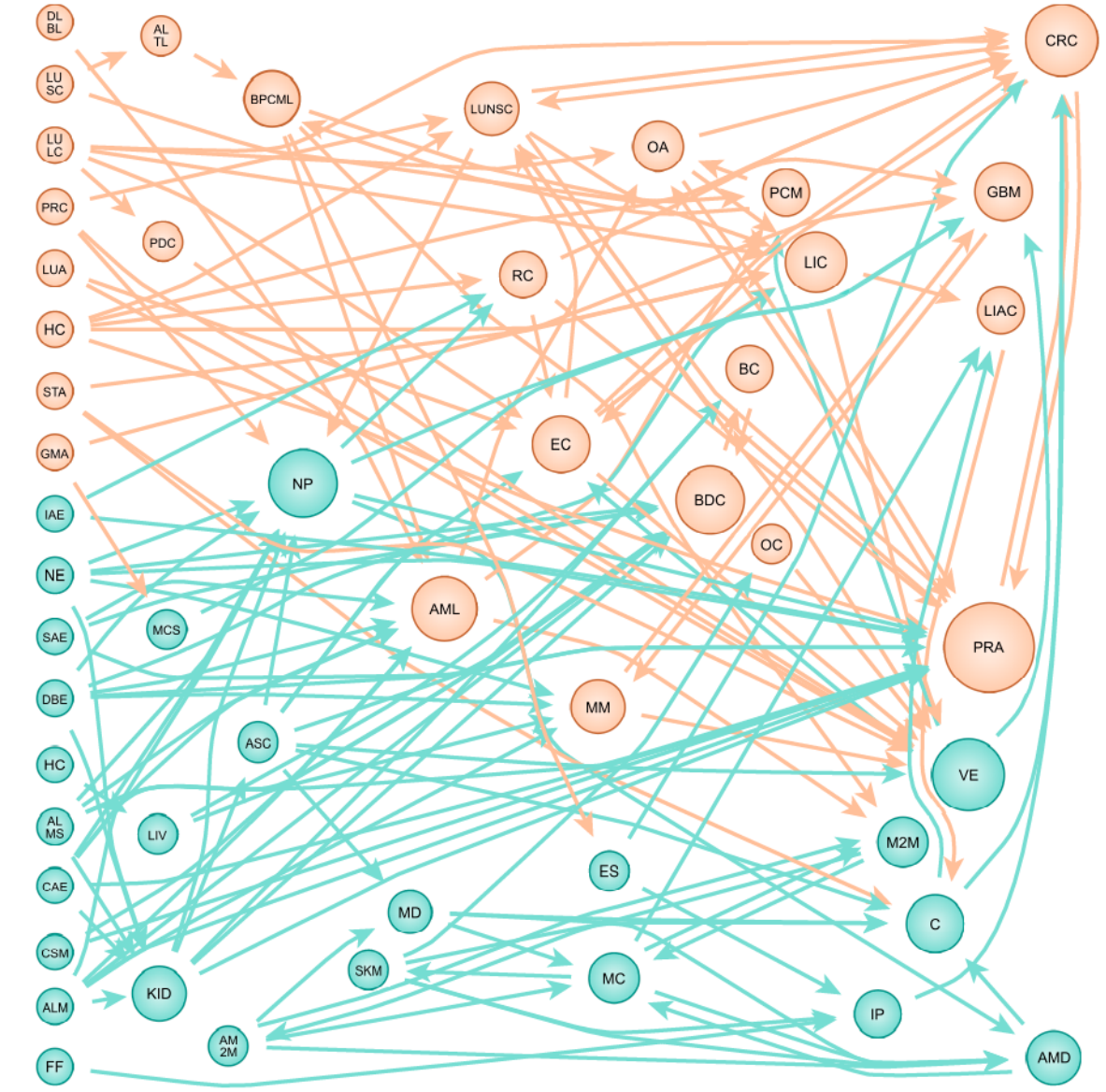

Fig. 6. Network of transitions (edges) between cell types (nodes) for the parameters indicated by the circle in Fig. 4*B*. The number of upstream cell types for a given node increases from left to right, so that nodes with zero incoming edges appear on the left. The node size increases with the number of incoming edges and the node color encodes normal (teal) and cancerous (orange) cell types.

Discussion

The results show that our control approach offers advantages over both annotation-based approaches and network-based approaches because it improves the quantitative predictive power of the former and reduces the effort required by the latter to adapt to new systems. In particular, the approach has the ability to circumvent the problem of combinatorial explosion in the number of multi-target interventions. This is achieved by using the transcriptional difference between states to computationally triage the combinations of single-target perturbations that are best suited to achieve a particular biological goal, whether that be inducing multipotency and pluripotency in differentiated tissues, driving transitions between differentiated cell types, or mitigating the progression of cancer.

The approach makes two central assumptions: i) The transcriptional state is the main determinant of cell behavior, and ii)

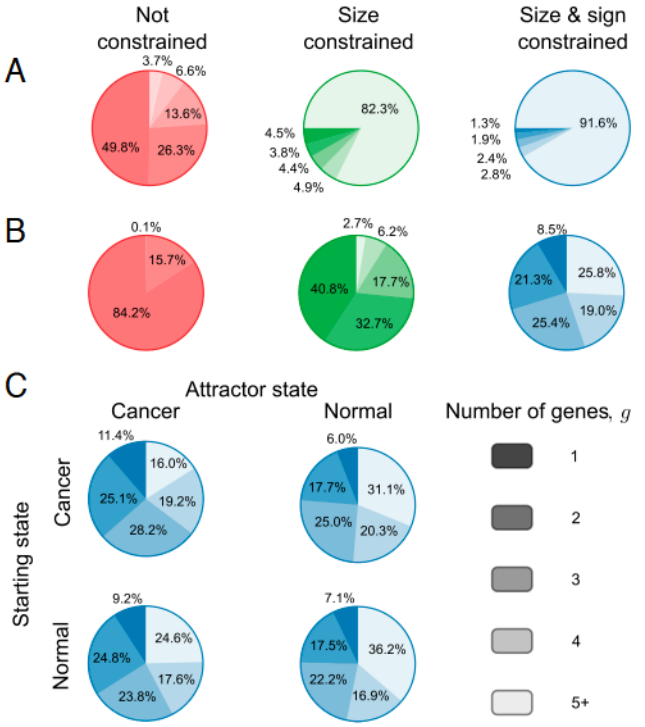

Fig. 7. Comparison of transdifferentiation transitions based on the number of genes required. (*A*) Color-coded fraction of directed cell type pairs in the RNASeq dataset that are able to be transdifferentiated as a function of the number of genes perturbed for all three constraint scenarios. (*B*) Same as (*A*) but for the GeneExp dataset. (*C*) Number of genes required for transitions as a function of the cell type class of the initial state (rows) and target state (columns) for the size-and-sign constrained case in (*B*).

the transcriptional responses to perturbations add approximately linearly to the transcriptional state. An important quality of the resulting model is that it can be trained on single-perturbation transcriptional data, which facilitates convergence due to its abundance and coverage of the space of potential perturbations. Support for (i) follows from our demonstration that it is possible to construct an accurate mapping from transcription to cell behavior (*SI Appendix*, Fig. S1). Since the approach is based on genome-wide data, it necessarily relies on destructive measurements (47). This fact imposes a tradeoff between the depth and time-resolution of the data required for any machine learning method and motivates our focus on long-term dynamics.

Concerning assumption (ii), our approach also serves as a starting point to investigate the role of nonlinearity at the scale of the whole cell, since it evaluates a linear approximation of known responses, and thus strong deviations from it would be evidence that nonlinear mechanisms are at work. Comparison of our approach with deep learning models based on variational autoencoders (VAEs) (48, 49) reveals that both offer comparable estimates of the mean expression of the final transcriptional state after application of a perturbation (*SI Appendix*, Table S1). This similarity in performance is surprising given that VAEs can in principle learn nonlinear behavior. One interpretation of this result is that cells are organized into mostly independent modules whose responses to disparate perturbations are independent and thus combine mostly linearly as proposed for *E. coli* (50). Indeed, the prevalence of pairwise nonlinear interactions as indicated by statistically significant epistasis is about 7% in *E. coli* (51) and 4% in *S. cerevisiae* (52), with most interactions organized by their genes' functional modules (53). If these trends apply to human cells, it could explain the limited amount of nonlinearity seen in previous applications of VAEs (48, 49). Moreover, nonlinearity is expected to be less pronounced in bulk averages over many cells than in the phenotype of individual cells.

Bulk transcriptomic data have the advantages that rare transcripts can be detected and that histological and morphological observations can be used to supervise the learning of cell type from transcriptional data. While this limits the ability to detect single-cell heterogeneity (47), we highlight potential applications of the approach that minimize or account for the impact of cell heterogeneity. In particular, the approach is suitable for identifying candidate gene perturbations to substitute for potentially oncogenic transgenes when creating naïve stem cells (54, 55). It can also be relevant for the management of diseases in which healthy tissues can be treated with gene and/or drug perturbations in ex vivo culture before autologous re-transplantation (7). The interventions designed by the approach need not be permanent if the target cell state is stable, and they need not be applied to all cells if the modified cells can be selected to out-compete unmodified ones (8). Moreover, the approach can be tailored to precision medicine applications by incorporating transcriptional data from individual patients' healthy and diseased tissues to identify treatments that account for differences between individuals. These examples illustrate the potential of the approach to computationally screen for effective regenerative therapies (56).

Finally, we note that the approach can readily incorporate forthcoming transcriptomic data, be applied to modalities other than transcriptional, and take advantage of rapidly advancing innovations in machine learning. The algorithms are designed to incorporate new transcriptomic data without recalculating the latent space, which enables them to capitalize on the exponentially increasing abundance of sequencing data (57, 58)—another advantage over knowledge-based control approaches that require a specific dynamical model. The versatility of the approach with respect to data modalities is important because recent research shows the effectiveness of data on complementary attributes of cell state, such as chromatin accessibility, in identifying key transcription factors for reprogramming (59). In particular, the approach is amenable to the incorporation of deep transfer learning (60), in which deep neural networks could be used to transfer knowledge across data modalities. Given its many uses and possible extensions, our approach has the potential to become a standard tool to translate bioinformatic data into biomedical applications.

Materials and Methods

Acquisition of the Training Data. The summary statistics for each dataset whose acquisition described below are provided in Fig. 1. The RNASeq dataset consists of i) unperturbed cell type data and ii) gene perturbation data. Part (i) was obtained from the GTEx consortium https://www.gtexportal.org/home/ (access date: 09/24/2019), while part (ii) was curated by searching BioProjects from the Sequencing Read Archive (SRA) (57) using the search terms "crispri" and "knockdown" and retaining the top 40 largest projects (in terms of number of sequencing runs). Specific details regarding each profile in this dataset, including the accession numbers, are provided in Dataset S1.

We constructed the GeneExp dataset using human gene expression data from the GEO repository (58), restricting to the most common platform, Affymetrix HG-U133+2 microarray (platform accession: GPL570), to facilitate the comparison of transcriptomic data from different GEO Series of Experiments (GSEs). The GeneExp dataset comprises three parts: i) unperturbed cell type data to train the KNN classifier, ii) gene perturbation data to characterize the corresponding transcriptional response, and iii) cell reprogramming data

used to validate our method. We obtained part (i) by searching GEO for the names of the NCI-60 cell lines and obtaining the relevant data from the GSEs, while supplementing this with data from the Cancer Cell Line Encyclopedia (GSE36139)(61) and the Human Body Index (GSE7307) to gain a representation of normal and cancerous cells. We curated part (ii) by querying GEO for "overexpression," "knockdown," and "RNA-interference" followed by selecting GSEs that measured gene perturbations. Finally, part (iii) was found by searching for "reprogramming," yielding 52 different protocols to de-differentiate cells toward a pluripotent state across 18 different GSEs. Specific details relating to each expression profile in this dataset, including the relevant accession numbers, are provided in Dataset S2.

Processing of the Expression and Sequencing Data. We performed batch correction on the gene expression data using covariates based on experimental series, cell line, and experimental treatment to remove systematic variation between series (62), as described in ref. 36. The data and covariates are described in Dataset S1 for the RNASeq data and Dataset S2 for the gene expression data. We used a custom Chip Definition File (CDF) that best maps the probes to the genes (63). For the RNASeq data, we used the Ensembl gene identifiers (version: GRCh38.p13) for protein-coding genes that overlapped with those in the GeneExp dataset.

Defining Perturbation Responses. To facilitate discussion of the calculation of the cellular response to perturbations, let X_{ij} be the data matrix of gene expression or transcript measurements, where $i \in \{1, \ldots, |\mathcal{G}|\}$ is an index over genes, and $j \in \{1, \ldots, N\}$ is an index over experiments. The dataset label D is suppressed to simplify notation. Let k be an index over GSEs, let m be an index over cell types and culture conditions, let p be an index over perturbation conditions, and let τ be the time the sample was collected. Then, $\mathcal{R}_{(k,m,p,\tau)} \subset \{1, \ldots, N\}$ is the set of columns that shares these experimental conditions, and we will refer to the submatrix that shares these covariates using $\mathbf{X}_{(k,m,p,\tau)}$. The data will be averaged over these experimental covariates in the following four steps:

1. Average the expression over replicates,

$$\bar{\mathbf{X}}_{(k,m,p,\tau)} = \sum_{j \in \mathcal{R}_{(k,m,p,\tau)}} \mathbf{X}_{(k,m,p,\tau)} \Big/ |\mathcal{R}_{(k,m,p,\tau)}|. \quad [5]$$

2. Average over time points,

$$\langle \bar{\mathbf{X}}_{(k,m,p)} \rangle = \sum_{\tau \in \mathcal{T}_{(k,m,p)}} \tau \bar{\mathbf{X}}_{(k,m,p,\tau)} \Big/ \sum_{\tau \in \mathcal{T}_{(k,m,p)}} \tau, \quad [6]$$

   where $\mathcal{T}_{(k,m,p)}$ is the set of time points for the experimental conditions indicated by (k, m, p).
3. Restricting to genetic perturbations $p \in \mathcal{P}$ and their controls 0, calculate differences

$$\bar{\mathbf{B}}_{(k,m,p)} = \langle \bar{\mathbf{X}}_{(k,m,p)} \rangle - \langle \bar{\mathbf{X}}_{(k,m,0)} \rangle. \quad [7]$$

4. Average over GSEs, cell types, and culture conditions,

$$\mathbf{B}_\ell = \mathbf{B}_{(p)} = \sum_{q \in \mathcal{P}} \bar{\mathbf{B}}_{(k,m,p)} \delta_{pq} \Big/ \sum_{q \in \mathcal{P}} \delta_{pq}, \quad [8]$$

   where $\ell \in \{1, \ldots, |\mathcal{P}|\}$ is an index over perturbations and δ is the Kronecker delta.

Eq. **6** weights later time-points more heavily so as to better estimate the long-term response to the gene perturbation. The responses $\mathbf{B}_\ell$ are likely to be causal, because they are the outcome of a controlled experiment, rather than merely correlative.

Approximating K with Transcriptional Distance. Our goal is to find the optimal perturbations $\mathbf{u}$ that steer from the initial state $\mathbf{x}^S$ belonging to cell type s to the target state $\mathbf{x}^A$ belonging to cell type a, as stated in Eq. **1**. We recall that $\mathbf{K}$ is the KNN mapping from transcriptional states to cell types, and we have suppressed the dataset labels to simplify notation. Direct solution of Eq. **1** is challenging because $\mathbf{K}$ is poorly behaved far from the data used to train it (as discussed below), making methods based on numerical derivatives too slow to employ due to the computational expense of evaluating $\mathbf{K}$. Here, we show that Eq. **2** is an appropriate approximation of Eq. **1** under the condition that $d_k \gg d_1$, where d_k and d_1 are the distances to the kth-nearest and nearest neighbor in the training data to a test point, respectively. Let $\mathcal{B}_\varepsilon(\mathbf{x}^A)$ be a ball of radius $\varepsilon > 0$ centered at $\mathbf{x}^A$ and define $P(\mathcal{B}_\varepsilon(\mathbf{x}^A))$ to be the probability that $\arg\max \mathbf{K}(\check{\mathbf{x}}^A) = \arg\max \mathbf{K}(\mathbf{x}^A)$ over all $\check{\mathbf{x}}^A \in \mathcal{B}_\varepsilon(\mathbf{x}^A)$. Then, $\lim_{\varepsilon \to 0} P(\mathcal{B}_\varepsilon(\mathbf{x}^A)) = 1$ because in this neighborhood, the magnitude of the possible discontinuity in $\mathbf{K}$ (caused by the change in the kth neighbor) is $\varepsilon d_k^{-1} / \left(1 + \varepsilon \sum_{i=2}^{k} d_i^{-1}\right)$, which vanishes in this limit. As a result, both Eqs. **1** and **2** provide the same answer at infinitesimal distances. To extend this approximation to finite distances, we note that (i) the method used to select eigengenes ensures that points within the convex hull of the $\mathcal{H} = \{\mathbf{x}^A \mid \mathbf{K}(\mathbf{x}^A) = \delta_{ja}\}$ belong to target cell type a with a high probability (36) and (ii) the target states we consider are averages of all expression profiles of target cell type a, $\bar{\mathbf{x}}^A$, which is contained within the convex hull. Thus, there is a finite distance $d_{\mathcal{H}}$ to the nearest boundary of the hull for which $P(\mathcal{B}_{d_{\mathcal{H}}}(\bar{\mathbf{x}}^A)) \approx 1$.

Calculating Transdifferentiation Transitions. The solution of Eq. **2** is underdetermined if the number of control inputs $|\mathcal{B}_D|$ is less than the number of features $|\mathcal{F}_D|$. In the underdetermined case, if the control variables are unconstrained, Eq. **2** is solvable by using the Moore–Penrose pseudoinverse of $\mathbf{B}$. If $|\mathcal{B}_D| > |\mathcal{F}_D|$, we impose that the ℓ_2 norm of the solution $||\mathbf{u}||_2$ is minimized to obtain a unique solution. The constrained problems reduce to the following program

$$\begin{aligned} \arg\min_{\mathbf{u}} \quad & d(\mathbf{x}^S, \mathbf{x}^A, \mathbf{u}; \mathbf{B}), \\ \text{s.t.} \quad & L \le u_\ell \le 1, \end{aligned} \quad [9]$$

where $L = -1$ for the size-constrained case and $L = 0$ for the size-and-sign-constrained case, which is solved using IBM ILOG CPLEX (12.10.0.0).

Limiting the Number of Genes Perturbed. Since it is experimentally infeasible to target more than a few genes simultaneously, we employ a forward selection approach to construct the perturbations. Let $\mathcal{V}_1 = \{\{j\}, j \in \{1, \ldots, |\mathcal{B}_D|\}\}$ be the set of all single-gene perturbations. Using $u^*_{v_1}$ to denote the input that minimizes d for the transcriptional response of column matrix $\mathbf{B}_{v_1}$, we compute Eq. **2** by evaluating $d(\mathbf{x}^S, \mathbf{x}^A, u^*_{v_1}; \mathbf{B}_{v_1})$ for each $v_1 \in \mathcal{V}_1$, and identify the column for which d is minimized, v_1^*. We proceed iteratively by constructing the set of all g-gene perturbations that include the best g–1-gene perturbation, denoted $\mathcal{V}_g = \{v^*_{g-1} \cup \{j\}, j \in \{1, \ldots, |\mathcal{B}_D|\} \backslash v^*_{g-1}\}$. We again evaluate $d(\mathbf{x}^S, \mathbf{x}^A, \mathbf{u}^*_{v_g}; \mathbf{B}_{v_g})$ for all elements $v_g \in \mathcal{V}_g$ and identify the element v_g^* that minimizes d. We note that $\mathbf{u}^*_{v_g}$ is now a vector (as indicated by the bold typeface) of control elements associated with v_g, which corresponds to the input values that minimize d. Accordingly, $\mathbf{B}_{v_g}$ is a matrix given by the g columns of the transcriptional response matrix associated with the elements of v_g. We continue this process, incrementing g until the target cell type is reached, i.e., $\arg\max \mathbf{K}(\mathbf{x}^S + \mathbf{B}_{v_g^*} \mathbf{u}^*_{v_g^*}) = \arg\max \mathbf{K}(\mathbf{x}^A)$.

Identifying Significant Genes. We identify significantly overrepresented genes by comparing the frequency of a gene's participation in a specific transdifferentiation transition with its frequency across all transitions for a given initial cell type s or target cell type a. For the RNASeq dataset, let $\widetilde{N}$ and $|\widetilde{\mathcal{C}}|$ be the numbers of states and cell types in the unperturbed partition, consider set of optimal control inputs $\mathbf{u}^{(j)}$, where $j \in \{1, \ldots, \widetilde{N}(|\widetilde{\mathcal{C}}| - 1)\}$, that are obtained for all pairs of initial states to all target cell types in this partition. The functions $I(j)$ and $T(j)$ map the index j to the initial cell type and target cell type, respectively. Furthermore, the number of solutions $\mathbf{u}$ associated with each

initial cell type is $H^{(s)} = \sum_{j=1}^{\widetilde{N}(|\widetilde{\mathcal{C}}|-1)} \delta_{I(j),s}$, the number associated with each target cell type is $H^{(a)} = \sum_{j=1}^{\widetilde{N}(|\widetilde{\mathcal{C}}|-1)} \delta_{T(j),a}$, and the number associated with each pair of cell types is $H^{(s,a)} = \sum_{j=1}^{\widetilde{N}(|\widetilde{\mathcal{C}}|-1)} \delta_{I(j),s}\delta_{T(j),a}$.

Then, the average inputs are

$$\bar{\mathbf{u}}^{(s,a)} = \frac{1}{H^{(s,a)}} \sum_{j=1}^{\widetilde{N}(|\widetilde{\mathcal{C}}|-1)} \mathbf{u}^{(j)} \delta_{I(j),s}\delta_{T(j),a} \qquad [10]$$

for each pair of initial cell types s and target cell types a. The gene most strongly associated with each transition is $v^{(s,a)} = \arg\max_i \bar{u}_i^{(s,a)}$, where $i \in \{1, \ldots, |\mathcal{P}|\}$. Defining $\mathbf{e}(i) = \left(e_{i'}(i)\right) = \left(\delta_{i,i'} \,|\, i' \in \{1, \ldots, |\mathcal{P}|\}\right)$ to be the unit (i.e., one-hot) vector associated with the ith perturbation, we determine the genes most strongly associated with each initial cell type and target cell type as $\mathbf{v}^{(s)} = \sum_{a' \neq s} \mathbf{e}(v^{(s,a')})/(|\widetilde{\mathcal{C}}| - 1)$ and $\mathbf{v}^{(a)} = \sum_{s' \neq a} \mathbf{e}(v^{(s',a)})/(|\widetilde{\mathcal{C}}| - 1)$, respectively. From $\mathbf{v}^{(s)}$, we determine the probability that the observed number of occurrences $h^{(a)}$ of each perturbation within $H^{(a)}$ states among a target cell type a exceeds the multinomial distribution null hypothesis

$$P(X \geq h^{(a)}) = \sum_{n=h^{(a)}}^{H^{(a)}} \sum_{\{n_\ell\}}^{|\mathcal{P}|\ \widetilde{\mathcal{C}}} \prod_{s \neq a} \binom{H^{(s,a)}}{n_\ell} \left(v_\ell^{(s)}\right)^{n_\ell}, \qquad [11]$$

where the middle sum is taken over all combinations $\{n_\ell\}$ such that $\sum_{\ell=1}^{|\mathcal{P}|} n_\ell = n$. Exchanging $s \leftrightarrow a$ in Eq. **11** yields the probability that the number of occurrences of a perturbation among an initial cell type is explained by the frequencies among the target cell types. We apply the two-stage Benjamini–Hochberg multiple hypothesis correction to the P-values obtained from Eq. **11** at an false discovery rate of 1% to obtain the significant genes associated with transdifferentiation transitions out of and into each cell type that are represented in *SI Appendix*, Figs. S4 and S5.

Comparison with Recent VAE Methods. Recent methods use VAEs to reconstruct the transcriptional states of perturbations applied to cell types when the post-perturbation transcriptional state is absent from the training data (48, 49). We compare the performance of these methods in *SI Appendix*, Table S1 using a single-cell RNASeq dataset of interferon-β stimulated and unstimulated peripheral blood mononuclear cells (64). This dataset was previously used to demonstrate the efficacy of the VAE approach for the purpose of reconstructing transcriptional states in ref. 48. We acquired the data and trained the VAE according to the documentation at https://scgen.readthedocs.io/en/stable/installation.html. We obtained the R^2 VAE estimates from the notebook file "scgen_perturbation_prediction.ipynb," available at https://scgen.readthedocs.io/en/stable/tutorials/scgen_perturbation_prediction.html.

The R^2 estimates for our method were computed using steps 1 to 4 in the subsection "Defining perturbation responses" above, with the following specifications: i) $k = 1$ since all data are from the same series of experiments and p = interferon-β, ii) all single-cell measurements of a given cell type were taken as replicates for the purposes of calculating $\langle \bar{\mathbf{X}}_{(m,p)} \rangle$ and $\langle \bar{\mathbf{X}}_{(m,0)} \rangle$, iii) $\mathbf{B}_{(p)}$ is an average over the training cell types only. Using m' to denote the test cell type, we then obtained R^2 between the actual state $\langle \bar{\mathbf{X}}_{(m',p)} \rangle$ and the predicted state $\langle \bar{\mathbf{X}}_{(m',0)} \rangle + \mathbf{B}_{(p)}$. In this case, the actual state is the transcriptional state of the stimulated test cell type and the predicted state is the sum of the transcriptional state of the unstimulated test cell type and the average transcriptional response in the training cell types.

Data, Materials, and Software Availability. Raw sequencing data is available through SRA (57) and GEO (58). Relevant accession numbers are included in Supporting Information. The software and processed data for employing the method are available from the "Cell reprogramming by transfer learning" repository on GitHub (65).

ACKNOWLEDGMENTS. This work was supported by ARO grant No. W911NF-19-1-0383 and NIH/NCI grants No. P50-CA221747 (through the Malnati Brain Tumor Institute) and No. U54-CA193419 (through the Chicago Region Physical Sciences Oncology Center). T.P.W. also acknowledges support from NSF GRFP grant No. DGE-0824162 and NIH/NIGMS grant No. 5T32-GM008382. The research benefitted from resources from NSF grant No. MCB-2206974 and the use of Quest High Performance Computing Facility at Northwestern University.

Author affiliations: [a]Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208; [b]Center for Network Dynamics, Northwestern University, Evanston, IL 60208; [c]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208; [d]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208; and [e]National Institute for Theory and Mathematics in Biology, Evanston, IL 60208

1. S. W. Morton *et al.*, A nanoparticle-based combination chemotherapy delivery system for enhanced tumor killing by dynamic rewiring of signaling pathways. *Sci. Signal.* **7**, ra44 (2014).
2. C. G. Da Silva, G. J. Peters, F. Ossendorp, L. J. Cruz, The potential of multi-compound nanoparticles to bypass drug resistance in cancer. *Cancer Chemother. Pharmacol.* **80**, 881–894 (2017).
3. J. A. MacDiarmid *et al.*, Sequential treatment of drug-resistant tumors with targeted minicells containing siRNA or a cytotoxic drug. *Nat. Biotechnol.* **27**, 643–651 (2009).
4. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
5. L. S. Qi *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
6. Y. Liu *et al.*, CRISPR activation screens systematically identify factors that drive neuronal fate and reprogramming. *Cell Stem Cell* **23**, 758–771 (2018).
7. H. Ludwig *et al.*, European perspective on multiple myeloma treatment strategies in 2014. *Oncologist* **19**, 829–844 (2014).
8. D. H. McDermott *et al.*, Chromothriptic cure of WHIM syndrome. *Cell* **160**, 686–699 (2015).
9. C. H. June, R. S. O'Connor, O. U. Kawalekar, S. Ghassemi, M. C. Milone, CAR T cell immunotherapy for human cancer. *Science* **359**, 1361–1365 (2018).
10. S. P. Cornelius, W. L. Kath, A. E. Motter, Realistic control of network dynamics. *Nat. Commun.* **4**, 1942 (2013).
11. D. K. Wells, W. L. Kath, A. E. Motter, Control of stochastic and induced switching in biophysical networks. *Phys. Rev. X* **5**, 031036 (2015).
12. B. Szigeti *et al.*, A blueprint for human whole-cell modeling. *Curr. Opin. Syst. Biol.* **7**, 8–15 (2018).
13. H. Yu, Y. Liu, A survey of underactuated mechanical systems. *IET Control Theory Appl.* **7**, 921–935 (2013).
14. F. J. Müller, A. Schuppert, Few inputs can reprogram biological networks. *Nature* **478**, E4 (2011).
15. G. Yang, C. Campbell, R. Albert, Compensatory interactions to stabilize multiple steady states or mitigate the effects of multiple deregulations in biological networks. *Phys. Rev. E* **94**, 062316 (2016).
16. J. G. Tejeda Zañudo, G. Yang, R. Albert, Structure-based control of complex networks with nonlinear dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7234–7239 (2017).
17. E. Newby, J. G. Tejeda Zañudo, R. Albert, Structure-based approach to identifying small sets of driver nodes in biological networks. *Chaos Interdiscip. J. Nonlinear Sci.* **32** (2022).
18. G. Baggio, D. S. Bassett, F. Pasqualetti, Data-driven control of complex networks. *Nat. Commun.* **12**, 1429 (2021).
19. J. L. Proctor, S. L. Brunton, J. N. Kutz, Generalizing Koopman theory to allow for inputs and control. *SIAM J. Appl. Dyn. Syst.* **17**, 909–930 (2018).
20. D. Canaday, A. Pomerance, D. J. Gauthier, Model-free control of dynamical systems with deep reservoir computing. *J. Phys. Complex.* **2**, 035025 (2021).
21. J. Z. Kim, Z. Lu, E. Nozari, G. J. Pappas, D. S. Bassett, Teaching recurrent neural networks to infer global temporal structure from local examples. *Nat. Mach. Intell.* **3**, 316–323 (2021).
22. D. G. Patsatzis, L. Russo, I. G. Kevrekidis, C. Siettos, Data-driven control of agent-based models: An Equation/Variable-free machine learning approach. *J. Comput. Phys.* **478**, 111953 (2023).
23. C. Campbell, J. Ruths, D. Ruths, K. Shea, R. Albert, Topological constraints on network control profiles. *Sci. Rep.* **5**, 18693 (2016).
24. L. Marazzi, M. Shah, S. Balakrishnan, A. Patil, P. Vera-Licona, NETISCE: A network-based tool for cell fate reprogramming. *NPJ Syst. Biol. Appl.* **8**, 21 (2022).
25. S. N. Steinway *et al.*, Network modelling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint Sonic Hedgehog and Wnt pathway activation. *Cancer Res.* **74**, 5963–5977 (2014).
26. J. G. Tejeda Zañudo, R. Albert, Cell fate reprogramming by control of intracellular network dynamics. *PLoS Comput. Biol.* **11**, e1004193 (2015).
27. K. Takahashi *et al.*, Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
28. A. C. D'Alessio *et al.*, A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Rep.* **5**, 763–775 (2015).

29. O. J. L. Rackham *et al.*, A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* **48**, 331–335 (2016).
30. C. V. Theodoris *et al.*, Transfer learning enables predictions in network biology. *Nature* (2023).
31. O. Alter, P. O. Brown, D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10101–10106 (2000).
32. T. P. Wytock, A. E. Motter, Predicting growth rate from gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 367–372 (2019).
33. S. Huang, G. Eichler, Y. Bar-Yam, D. E. Ingber, Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701 (2005).
34. C. H. Waddington, *Principles of Embryology* (Allen & Unwin, London, 1956).
35. S. A. Kauffman, Homeostasis and differentiation in random genetic control networks. *Nature* **224**, 177–178 (1969).
36. T. P. Wytock, A. E. Motter, Distinguishing cell phenotype using cell epigenotype. *Sci. Adv.* **6**, eaax7798 (2020).
37. S. Tripathi, D. A. Kessler, H. Levine, Minimal frustration underlies the usefulness of incomplete regulatory network models in biology. *Proc. Natl. Acad. Sci. U.S.A.* **120** (2023).
38. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
39. K. Glass, M. Girvan, Annotation enrichment analysis: An alternative method for evaluating the functional properties of gene sets. *Sci. Rep.* **4**, 4191.1–4191.9 (2014).
40. R. Sopko *et al.*, Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
41. M. K. Ray *et al.*, CAT7 and cat7l long non-coding RNAs tune polycomb repressive complex 1 function during human and zebrafish development. *J. Biol. Chem.* **291**, 19558–19572 (2016).
42. I. Sarropoulos, R. Marin, M. Cardoso-Moreira, H. Kaessmann, Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**, 510–514 (2019).
43. A. Modelska *et al.*, The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. *Cell Death Dis.* **6**, e1603 (2015).
44. E. Sullivan *et al.*, Boolean modeling of mechanosensitive epithelial to mesenchymal transition and its reversal. *iScience* **26**, 106321 (2023).
45. M. Galbraith, H. Levine, J. N. Onuchic, D. Jia, Decoding the coupled decision-making of the epithelial-mesenchymal transition and metabolic reprogramming in cancer. *iScience* **26**, 105719 (2023).
46. J. I. Joo, J. X. Zhou, S. Huang, K. H. Cho, Determining relative dynamic stability of cell states using Boolean network model. *Sci. Rep.* **8**, 12077 (2018).
47. O. Stegle, S. A. Teichmann, J. C. Marioni, Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
48. M. Lotfollahi, F. A. Wolf, F. J. Theis, scGen predicts single-cell perturbation responses. *Nat. Methods* **16**, 715–721 (2019).
49. M. Lotfollahi *et al.*, Predicting cellular responses to complex perturbations in high-throughput screens. *Mol. Syst. Biol.* **19**, e11517 (2023).
50. A. V. Sastry *et al.*, The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **10**, 5536 (2019).
51. M. Babu *et al.*, Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genet.* **10**, e1004120 (2014).
52. M. Costanzo *et al.*, A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).
53. D. Segrè, A. DeLuna, G. M. Church, R. Kishony, Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2005).
54. P. Hou *et al.*, Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* **341**, 651–654 (2013).
55. L. E. Bates, J. C. Silva, Reprogramming human cells to naïve pluripotency: How close are we? *Curr. Opin. Genet. Dev.* **46**, 58–65 (2017).
56. C. Jopling, S. Boue, J. C. I. Belmonte, Dedifferentiation, transdifferentiation and reprogramming: Three routes to regeneration. *Nat. Rev. Mol. Cell Biol.* **12**, 79–89 (2011).
57. R. Leinonen, H. Sugawara, M. Shumway, I. N. S. D. Collaboration, The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21 (2010).
58. T. Barrett *et al.*, NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
59. J. Hammelman, T. Patel, M. Closser, H. Wichterle, D. Gifford, Ranking reprogramming factors for cell differentiation. *Nat. Methods* **19**, 812–822 (2022).
60. J. Chen *et al.*, Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. *Nat. Commun.* **13**, 6494 (2022).
61. J. Barretina *et al.*, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
62. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
63. M. Dai *et al.*, Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).
64. H. M. Kang *et al.*, Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
65. T. P. Wytock, A. E. Motter, Cell reprogramming by transfer learning. GitHub. https://github.com/twytock/cell_reprogramming_by_transfer_learning. Deposited 23 December 2023.