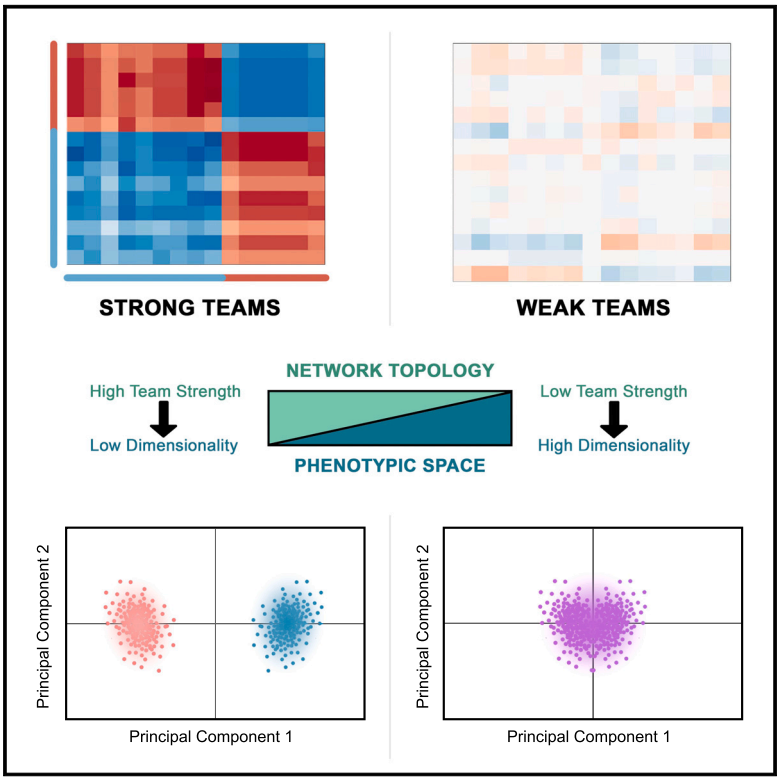
iScience

Article

Low dimensionality of phenotypic space as an emergent property of coordinated teams in biological regulatory networks

Graphical abstract



Authors

Kishore Hari, Pradyumna Harlapur, Aashna Saxena, ..., Tanisha Malpani, Herbert Levine, Mohit Kumar Jolly

Correspondence

a.hari@northeastern.edu (K.H.), h.levine@northeastern.edu (H.L.), mkjolly@iisc.ac.in (M.K.J.)

In brief

Systems biology

Highlights

- Transcriptomes of binary cell-fate decision systems are robustly low-dimensional
- Mutually inhibiting teams in GRNs enable stable low-dimensional phenotypic space
- Teams-based topology in GRNs is sufficient, but not necessary, for low dimensionality
- Teams are necessary for the robustness of low dimensionality in phenotypic space

Hari et al., 2025, iScience *28*, 111730
February 21, 2025
https://doi.org/10.1016/j.isci.2024.111730

CellPress

iScience

CellPress
OPEN ACCESS

Article

Low dimensionality of phenotypic space as an emergent property of coordinated teams in biological regulatory networks

Kishore Hari,[1,3,4,5,*] Pradyumna Harlapur,[1,5] Aashna Saxena,[1,5] Kushal Haldar,[1,2] Aishwarya Girish,[1] Tanisha Malpani,[1] Herbert Levine,[3,4,*] and Mohit Kumar Jolly[1,6,*]

[1]Department of Bioengineering, Indian Institute of Science, Bengaluru, Karnataka 560012, India
[2]Indian Institute of Science Education and Research Kolkata, Kolkata, West Bengal 741246, India
[3]Center for Theoretical Biological Physics, Northeastern University, Boston, MA 02115, USA
[4]Department of Physics, Northeastern University, Boston, MA 02115, USA
[5]These authors contributed equally
[6]Lead contact
*Correspondence: a.hari@northeastern.edu (K.H.), h.levine@northeastern.edu (H.L.), mkjolly@iisc.ac.in (M.K.J.)
https://doi.org/10.1016/j.isci.2024.111730

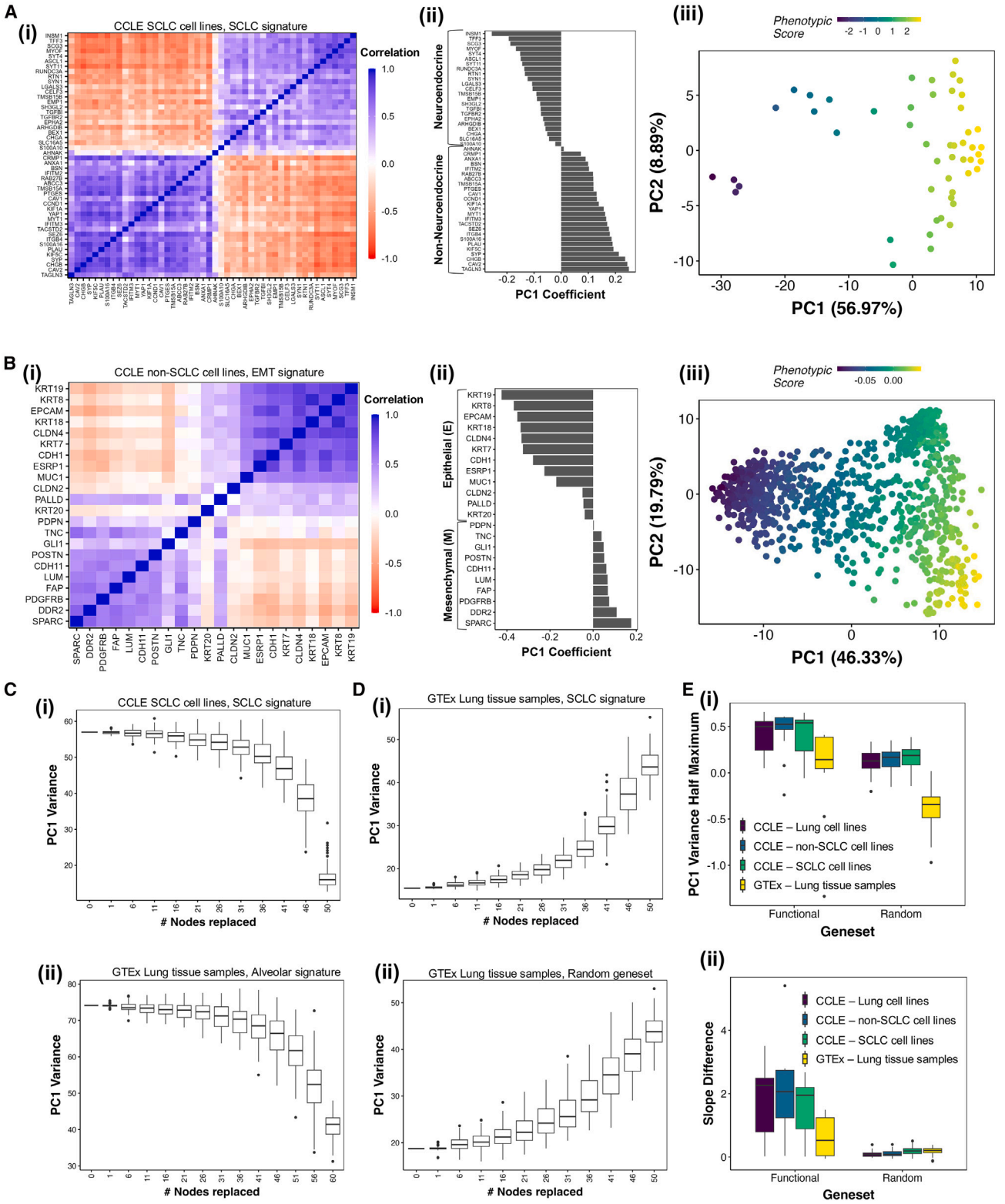
SUMMARY

Cell-fate decisions involve coordinated genome-wide expression changes, typically leading to a limited number of phenotypes. Although often modeled as simple toggle switches, these rather simplistic representations often disregard the complexity of regulatory networks governing these changes. Here, we unravel design principles underlying complex cell decision-making networks in multiple contexts. We show that the emergent dynamics of these networks and corresponding transcriptomic data are consistently low-dimensional, as quantified by the variance explained by principal component 1 (PC1). This low dimensionality in phenotypic space arises from extensive feedback loops in these networks arranged to effectively enable the formation of two teams of mutually inhibiting nodes. We use team strength as a metric to quantify these feedback interactions and show its strong correlation with PC1 variance. Using artificial networks of varied topologies, we also establish the conditions for generating canalized cell-fate landscapes, offering insights into diverse binary cellular decision-making networks.

INTRODUCTION

Cellular decision-making is driven by the complex non-linear dynamics of underlying gene regulatory networks (GRNs). Despite involving a large number of interacting components, these networks often enable only a limited number of phenotypes/cell types.[1,2] For example, in differentiation events, the cells move from a high-entropy progenitor state to a low-entropy differentiated phenotype.[3] Such changes in entropy suggest a large-scale modification of gene expression patterns during differentiation. Another well-studied example of cellular decision-making is epithelial-mesenchymal plasticity (EMP), a crucial process implicated in development, regeneration, and cancer metastasis.[4] Populations capable of undergoing EMP exhibit two major phenotypes: epithelial (E) and mesenchymal (M). The E and M phenotypes have mutually exclusive biochemical signatures, and their master regulators inhibit each other,[5–7] reminiscent of the presence of a "toggle switch" between master regulators at many developmental cell-fate bifurcation events.[8] A minor fraction of EMP populations also exhibit hybrid E/M phenotypes[9–11]; these states have been implicated as major drivers of the metastatic process.[12] The hybrid E/M phenotypes have complex non-uniform biochemical signatures as compared to the E and M phenotypes, making them hard to identify and therapeutically target.[13] Specifically, both progenitor phenotypes and the hybrid E/M phenotypes lack the coherent expression profiles of differentiated cell types and the E and M phenotypes, respectively. In other words, these phenotypes exist in a higher dimensional space than the E and M states, as the variation between E and M states can be confined to the composite axis composed of canonical epithelial and mesenchymal genes.[14,15]

A recent analysis of EMP models and other related GRNs suggested that most of the variance in phenotypic space could be determined by the first principal component (PC1) axis,[16] suggesting that many cell-fate decision systems, including those of EMP, operate in a low-dimensional space. However, what underlying network features allow for an abnormally high PC1 variance has been unclear. In our previous work,[5] we demonstrated a seemingly different phenomenon in diverse networks of EMP. Despite their complexity, their repertoire of emergent steady states is limited; thus, most of the phenotypic space converges to epithelial or mesenchymal states. We demonstrated that this convergence was a direct consequence of underlying network topology. These networks comprised well-coordinated "teams" of nodes, such that the nodes belonging to the same team effectively activate each other, while those across teams inhibited one



(legend on next page)

another. How connected (or not) the concepts of ''teams'' and PC1 dominance remains elusive.

Here, we reconcile the observations related to PC1 variance and the presence of ''teams'' of nodes as a design principle underlying robust phenotypic landscapes. First, we demonstrate this low dimensionality trait in RNA-sequencing data and show that this behavior is specific to relevant genes forming ''teams'' but not for housekeeping genes. Next, we investigate GRNs across several biological contexts, including EMP, small-cell lung cancer (SCLC), and gonadal cell fate determination. Here, we show that networks forming strong teams show a high PC1 variance, a property lost upon the weakening of teams by deleting specific edges and/or swapping edges within the network. The weaker the team strength, the more the number of principal components are needed to explain the underlying variance and, thus, the higher the dimensionality of phenotypic space. To generalize our results further, we generated artificial team-based networks of varying densities to highlight how underlying network topology governs the team strength and, consequently, the PC1 variance. We additionally establish the necessity of teams' structure to achieve PC1 stability, a feature of robustness observed in transcriptomic data for functional genesets. Overall, our results elucidate how the presence of well-coordinated ''teams'' of nodes in underlying networks leads to low dimensionality of phenotypic space and concomitant cell-fate canalization during decision-making.

RESULTS

Transcriptomic signatures of cell-fate decision systems are stably low-dimensional

We started by analyzing bulk RNA-seq data from the Cancer Cell Line Encyclopedia (CCLE)[17] and the Adult Genotype-Tissue Expression (GTEx) portal.[18] While the GTEx data serves as a representative of normal adult tissue, CCLE data represents diseased tissue. In each case, we sourced genesets that capture phenotypic heterogeneity resulting from a cell-fate decision system. With the CCLE dataset, we used the small cell lung cancer (SCLC) genesets capturing the neuroendocrine (NE)/non-neuroendocrine (non-NE) heterogeneity[19,20] in conjunction with SCLC cell line samples, and epithelial-mesenchymal plasticity (EMP) geneset[21] in conjunction with Breast cancer and non-SCLC cell lines. With the GTEx dataset, we used the hallmark C8 genesets from MSigDB that distinguish different cell types.[22,23] Specifically, we used the genesets for alveolar fibroblast and epithelial signatures. Additionally, we utilized a geneset generated using a machine learning tool called FUGUE which uses transcriptional and network features to generate tissue-relevant genesets.[24]

The pairwise correlation matrix (correlation between gene expression profiles across cell lines) showed a clear separation of the gene sets into two ''teams'' in all expression levels for the correspondingly appropriate gene lists - SCLC gene sets for the SCLC cell lines (Figure 1A, i) and EMT gene sets for non-SCLC cell lines (Figure 1B, i). Each ''team'' here is defined as a set of genes that show a positive correlation within themselves and a negative correlation with genes that do not belong to the team. Such a correlation pattern is characteristic of a dominantly binary cell-fate decision system such as EMT, where epithelial and mesenchymal phenotypes are much more abundant than ''hybrid'' phenotypes. We then performed principal component analysis (PCA) on these expression matrices to see the extent to which it is possible to reduce the dimensionality of the data. We provide a brief contextual explanation of PCA to better understand the results and their significance. The expression matrix should be considered an n-dimensional dataset, where n is the number of genes in the corresponding gene set. In other words, each sample/cell line has n coordinates that uniquely define it. PCA transforms the data dimensions from n gene expression vectors to n mutually perpendicular vectors or axes. Each new axis is a unique linear combination of the gene expression vectors. Importantly, the first axis (PC1) is calculated to capture the maximum variance in the data. PC1 also identifies the ideal classification of the genes into two groups or teams, such that the genes of the same group have similar expression patterns (e.g., positive correlation, co-expression). In SCLC cell line data (Figure 1A, i), based on the clustering of the genes in the correlation matrix, it is clear that this classification should correspond to NE and non-NE genes. The coefficients of the genes in the PC1 axis reflect the team classification (Figure 1A, ii). As expected, PC1 explained a significant part of the variance in the expression data (Figure 1A, iii). Similarly, we see high variance explained by PC1 in an EMT geneset in CCLE breast cancer cell lines (Figure 1B), the alveolar fibrotic-epithelial geneset (Figure S1A), and the FUGUE geneset (Figure S1B) in GTEx. For both datasets, random genesets comprise protein-coding genes had a much lower PC1 variance (Figure S1C), highlighting that high PC1 variance is a property unique to genesets corresponding to active cell-fate decision systems in the dataset. Similarly, for the whole genome, PC1 explains a higher fraction of variance by its nature, but the difference in PC1 and PC2 variance is not significant, as is the case for the functional genesets (Figures S2A and S2B). Furthermore, the SCLC

Figure 1. Low-dimensionality in bulk transcriptomic data

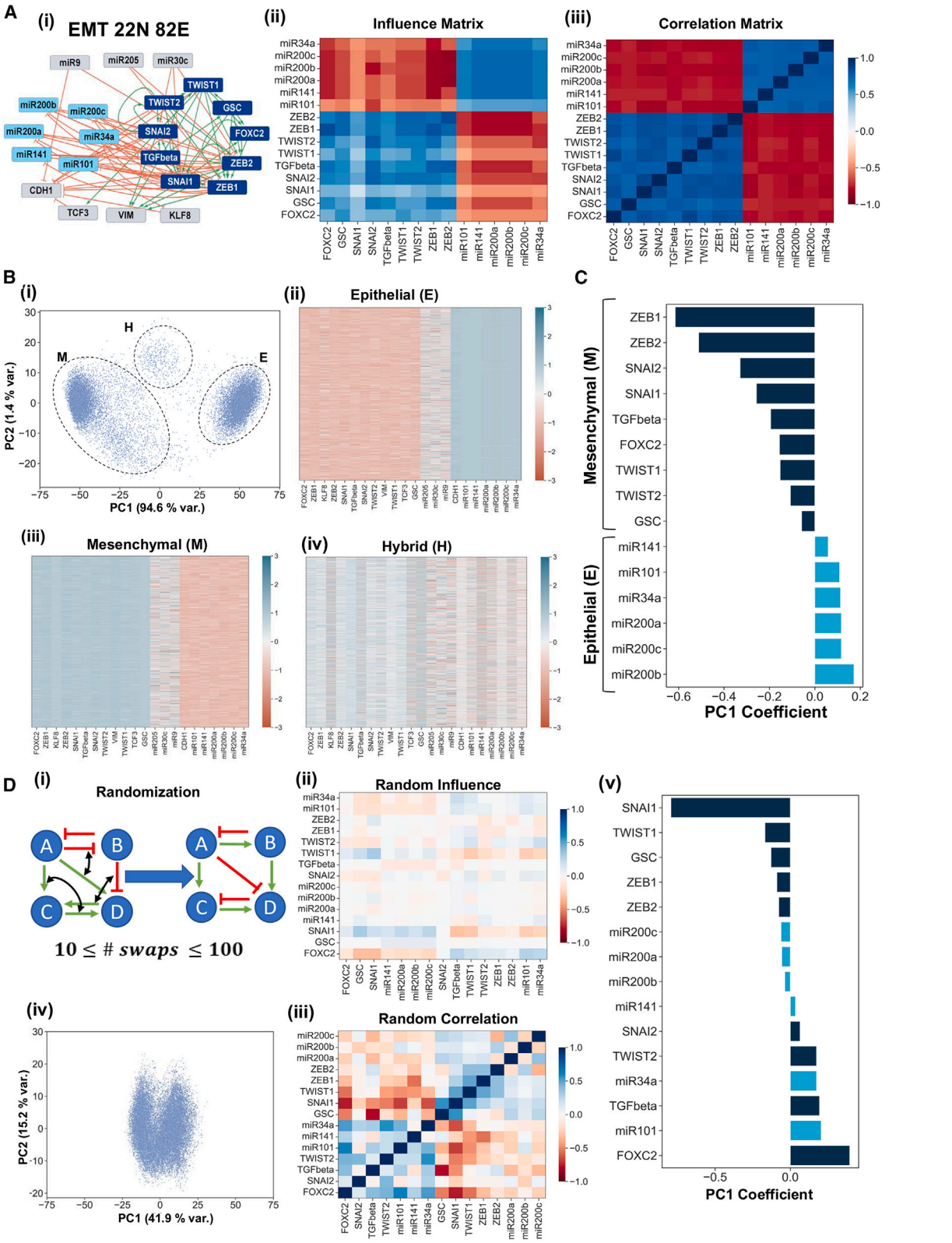
(A) (i) Pairwise correlation of expression for the 32 genes of SCLC network in the CCLE SCLC cell-line data. (ii). Coefficients of the first principal component axis for the geneset-cell line combination in (i). (iii) Scatterplot of CCLE expression data on the PC 1 and 2 axes. The color bar represents the phenotypic score of each sample.

(B) Same as A, but for CCLE non-SCLC cell lines for EMT signature.

(C) PC1 Variance against the number of genes replaced with housekeeping genes for (i) CCLE SCLC cell lines with SCLC gene signature, (ii) GTEx Lung tissue samples with Alveolar fibrotic-epithelial signature obtained from MSigDB.

(D) Same as C, but for GTEx samples with (i) 50 gene SCLC signature and (ii) a random gene signature.

(E) Boxplots depicting the comparison of (i) the number of genes replaced corresponding to half maximum PC1 and (ii) the difference in slopes before and after half maximum of mean PC1 vs. number of genes replaced between the pooled biological genesets and housekeeping genes.



(legend on next page)

signature capturing the neuroendocrine - non-neuroendocrine cell fate switching is uniquely active in cancer tissue and correspondingly has a low PC1 variance and low PC1/PC2 variance ratio (Figures S2B).

We then studied the robustness of these PC1 results to errors in the corresponding gene set. To do this, we asked how the dimensionality of these datasets would change when geneset components were increasingly replaced by irrelevant genes, thereby forming a somewhat erroneous geneset. For a geneset of length N, we replaced n genes at a time with a set of housekeeping genes[25] and calculated the principal components for the new set of N genes. We repeat this exercise 100 times for each value of n. First, note that PC1 variance decreases with an increasing number of genes replaced with housekeeping genes (Figure 1C), while the number of PC axes required to explain 90% variance increases (Figure S1D). Both the results suggest that the dimensionality of these datasets increases with increased error in the geneset composition. Interestingly, the increase in dimensionality is minimal until nearly 80% of the genes are replaced with housekeeping genes. We refer to this phenomenon as PC1 stability. We found PC1 stability for all the relevant geneset-dataset combinations, as shown in Figure 1C. However, as a control, we studied the SCLC geneset in normal tissue and found a very low PC1 variance and lack of PC1 stability in GTEx Lung tissue data (Figures 1D, i and S1E, i). This behavior was recapitulated by choosing a random set of protein-coding genes as well (Figures 1D, ii and S1E, ii), suggesting that PC1 stability could be a unique feature of transcription factors that are ''functional'' in the context of the dataset studied.

We quantified PC1 stability using two metrics, half maximum (the number of genes needed to be swapped to induce half of the maximum change in mean PC1 variance) normalized against the maximum relative change in PC1 variance, and the Slope difference (difference in the rate of change of mean PC1 variance with the number of genes replaced, see STAR Methods for details of calculation). The half maximum and slope difference for the functional genesets are much higher than that of random genesets (Figures 1E and S1F). Overall, the PC1 stability suggests the presence of strong control structures unique to functional sets of transcription factors. We had previously identified the presence of strong control structures in GRNs underlying cell-fate decision networks, which were called ''teams'' of nodes.[5] Boolean simulations of these mutually inhibiting teams resulted in a bimodal phenotypic landscape resilient to dynamical and structural perturbations. These networks also satisfy the property of low frustration that was found to be underlying high PC1 variance by Tripathi et al.[16] Hence, we wanted to understand the effect of teams on the dimensionality of the continuous state space.

Teams of nodes in EMP network lead to low dimensional steady state space

We started by analyzing a 22 node 82 edge (22N 82E) GRN underlying EMP[26] (Figure 2A, i). Each node in this network is either a transcription factor or a micro-RNA. Each edge represents the regulation (either transcriptional or post-transcriptional) of the target node (gene) by the source node. These regulations can be activating or inhibiting. As demonstrated in our previous work, the influence matrix generated from this network (which takes a weighted sum of direct and indirect interactions between each node pair) can be effectively considered a higher-order toggle switch consisting of two ''teams'' of nodes (Figure 2A, ii).[5] Nodes belonging to the same team have a positive influence within themselves and a negative influence on the nodes of the other team. One team consists of the epithelial factors in the network: miR101, miR141, miR200a, miR200b, miR200c, and miR34a. The second team comprises mesenchymal transcription factors: ZEB1/2, TWIST1/2, TGFβ, SNAI1/2, GSC, and FOXC2. The emergent phenotypes from this network exhibit high expression from all the nodes of either one of the two teams, while the nodes of the other team are strongly suppressed.[5] Additionally, there are peripheral nodes (grey-colored nodes in Figure 2A, i) in the network: signal nodes (miR9, miR205, and miR30c) and output nodes (CDH1, TCF3, VIM, and KLF8). We have previously shown that these peripheral nodes have no major consequence on the phenotypic landscape resulting from this network.

We simulated this network over many kinetic parameter sets using the software tool RACIPE.[26] RACIPE generates a system of coupled Ordinary Differential Equations (ODEs) for a particular topology of GRN. It then randomly samples values for each parameter of the ODE system from a corresponding self-consistently determined physiological range to generate an ensemble of parameter sets (Table 1). RACIPE then simulates the ODEs for each parameter set over multiple initial conditions to identify the steady states. Despite the heterogeneity in the parameter sets, the expression levels of the nodes of the network were well correlated with each other. Notably, the team structure seen in the influence matrix (obtained directly from the network structure, without any simulations) was reflected in the pairwise correlation matrix obtained from dynamical simulations, where the epithelial nodes (such as miR200b and miR200c) are strongly positively correlated with other epithelial nodes and negatively with the mesenchymal nodes (such as ZEB1 and SNAIL) (Figure 2A, iii). Note that influence, unlike correlation, is a directional quantity, making the influence matrix asymmetric. Despite this

Figure 2. Structural similarities between teams and PC1 axis

(A) The network diagram (i) and the influence matrix (ii) for the 22N 82E EMP network. (iii) Correlation matrix depicting the pairwise correlations between the node expression levels across all parameter sets in RACIPE.

(B) (i) Scatterplot mapping the solutions generated from RACIPE on the axes of PC1 and PC2. (ii-iv) Heatmaps depicting the expression levels of the nodes of the network for each individual cluster seen in (i).

(C) The loadings of the nodes for each PC1 axis. The colors of the bars represent the team identity of the nodes.

(D) (i) Depiction of network randomization. (ii) Influence matrix of a random network. (iii) Correlation matrix generated for a random network. (iv) Scatterplot similar to C (i) but for a random network. (v) Similar to D but for a random network.

Table 1. Parameter ranges used by RACIPE

Parameters	Minimum	Maximum
Production Rate (G)	1	100
Degradation Rate (k)	0.1	1
Fold Change (Inhibition λ)	0.01	1
Fold Change (Activation λ)	1	100
Hill Coefficient (n)	1	6
Threshold	The ranges depend on the in-degree - half functional rule	

distinction, we find that the influence matrix of the 22-node EMT network strongly resembles the correlation matrix obtained from simulations across a large range of parameter sets.

We then performed a principal component analysis (PCA) over the collection of all steady states obtained for this network. When the solutions are plotted along the first and second principal component axes (PC1 and PC2, respectively), we found that the steady states separated into 3 clusters (Figure 2B, i). Analyzing the expression patterns of the steady states forming these clusters revealed that the two most prominent clusters were the epithelial and mesenchymal phenotypes. In contrast, the smaller cluster showed a mixed expression of epithelial (E) and mesenchymal (M) nodes and hence could be classified as a hybrid (E/M) cluster (Figure 2B, ii-iv). Furthermore, we observed that the PC1 axis was enough to distinguish between the epithelial and mesenchymal steady states (Figure 2B, i). Given our previous understanding of the connection between teams of nodes and the emergent phenotypes, we hypothesized that the two-team structure in the network could drive the relative weights of node expressions that make up the PC1 axis. To verify this hypothesis, we looked at the composition of the PC1 axis (using the node coefficients of PC1) and compared it against the composition of teams (Figure 2C, nodes belonging to different teams are colored differently). In line with our hypothesis, we see that the PC1 contributions from nodes belonging to different teams have opposite signs, and those belonging to the same team have the same sign. Together, these results support the hypothesis that team structure leads to PC1 dominance, which explains most of the variance in steady-state space.

An immediate implication of the above hypothesis would be that in the absence of strong teams, the PC1 axis would no longer be able to account for the majority of the variance in the steady-state space. To test the validity of this implication, we generated random networks by swapping many pairs of edges in the network (Figure 2D, i)). We have previously shown that such randomization can disrupt the team structure and thus the stability and frequency of E, M, and H phenotypes (Figure 2D, ii). We simulated each generated random network using RACIPE and performed PCA on the steady-state values given by the simulations. Unlike the wild-type (WT) EMP network, in the random networks, the PC1 and PC2 axes could not as well separate the steady states into distinct clusters (Figures 2D, iii and 2D, iv). We also found that the classification of the nodes based on their PC1 coefficients did not match their teams' classification (Figure 2E, v). Therefore, two mutually inhibiting teams in WT networks appear necessary to enable low dimensionality in the emergent phenotypic space.

We further calculated the stability of PC1 variance emergent from the EMT network by replacing individual nodes with random expression vectors. We find that PC1 variance remains high till >80% nodes (18 out of 22) are replaced with random expression vectors (Figure 3A), similar to what we found for the transcriptomic data. Therefore, two mutually inhibiting teams in WT networks appear necessary to enable low dimensionality in the emergent phenotypic space.

Next, we compared the percentage variance explained along the PC1 axis for WT networks against that of random networks. This measure was much lower in random networks than in the WT networks. In the WT network, PC1 alone explained nearly 90% of the variance, while the PC1 variance of random networks was around 0.4 (Figure 3B, i). A scatterplot between the team strength of the random and WT networks against the variance explained by PC1 (Figure 3B, ii) revealed a saturating increase of PC1 variance with the team strength of random networks, suggesting that an increase in team strength can lead to a corresponding increase in the variance explained by PC1. After a team strength of about 0.25–0.3, the variance explained by PC1 starts to saturate.

Next, we wanted to estimate the dimensionality of the steady-state space emergent from these networks. More than 90% of the variance in the phenotypic space of the EMP network is explained by the PC1 axis, making it effectively one-dimensional.[16] We then asked how many principal components are necessary for explaining 90% variance in random networks. While the WT network only requires one principal component (PC1), the random networks require a higher number of PCs, ranging from 7 to 13 out of the 22 axes (Figure 3B, iii), demonstrating the high-dimensionality of the phenotypic space of random networks as compared to the biological networks. The number of PCs also showed a strong negative correlation with team strength (Figure 3B, iv), showing that strong teams' presence reduces the steady-state space's dimensionality.

To further emphasize the role of the network topology, we compared random networks generated from the WT over a range of edge swaps. Swapping the edges changes the topological features of the biological network. The more swaps there are, the greater the changes in the topology, which should be reflected in the emergent behavior. Indeed, we find that the PC1 variance decreases with an increasing number of swaps (Figure 3C) along with the team strength (color of boxes in Figure 3C).

Strong teams underlie the low-dimensionality of phenotypic space in multiple contexts

Our observations establish that strong teams underly the PC1 dominance in the 22N 82E EMP network. We then asked if the aforementioned results hold in other biological contexts. We investigated three additional GRNs underlying cell fate decisions: SCLC subtypes, pluripotency, and gonadal fate determination (Figure S3A).[19,27,28] The SCLC network has been shown to regulate the phenotypic heterogeneity underlying small-cell lung cancer, allowing the cells to switch between neuroendocrine (NE) and non-neuroendocrine (non-NE) phenotypes. The pluripotency network describes the differentiation of

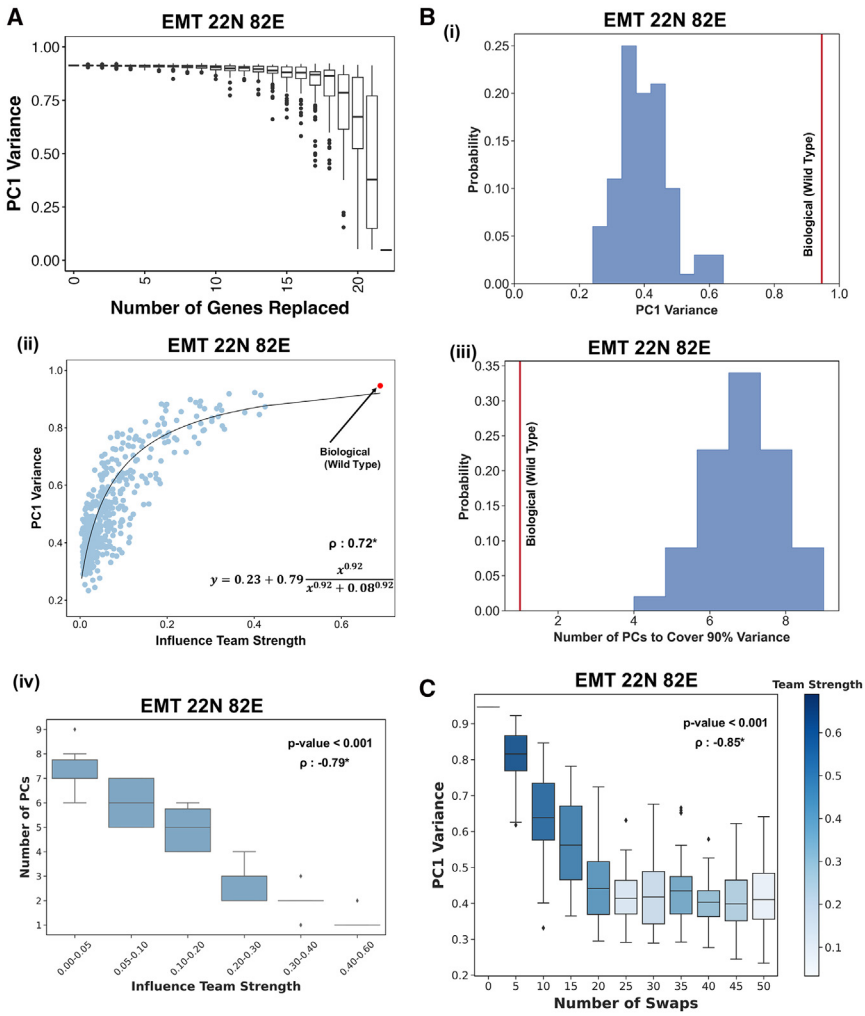


Figure 3. Strong teams lead to a low-dimensional steady-state space

(A) (i) Histogram depicting the variance explained by PC1 for random networks. A vertical red line represents the EMP network. (ii) Scatterplot depicting the dependence of variance explained by PC1 on the team strength. Each point is a random network. The biological network has been pointed out. The Spearman correlation coefficient, ρ, is reported, with * indicating a p – value < 0.05.

(B) (i) Histogram depicting the number of PCs required to explain 90% variance in random networks. A vertical red line represents the EMP network. (ii) Boxplot depicting the dependence of the number of PCs required to explain 90% of the variance in steady states on the team strength. The p-value for one-way ANOVA and the Spearman correlation coefficient (ρ) have been reported, with * indicating a p – value <0.05.

(C) PC1 variance as a function of the number of swaps used to generate random networks.

embryonic stem cells (hESCs). The network was constructed to identify pathways to achieve induced pluripotent stem cells (iPSCs). The gonadal fate determination network describes the cell fate determination between Sertoli and granulosa cell types, events in the early stages of the development of males and females, respectively. These three networks display a spectrum of characteristics. The pluripotency network has hub genes (NANOG, SOX2, OCT4) with high out-degree (i.e., these nodes regulate multiple genes). SCLC and pluripotency networks have relatively lower team strengths (0.25 and 0.2), which puts them near the border of saturation in the PC1 variance vs. team strength plot in Figure 3B. The gonadal and SCLC networks show a clear team structure in their influence matrices (Figures 4A, i and 4A, ii). Such a team structure is relatively harder to discern in the iPSC influence matrix (Figure 4A, iii). However, our clustering-based method of identifying teams allowed us to classify the nodes into two teams. This classification agreed with the biological function of these molecules and with the groups seen in the correlation matrix (Figure S3B). For all three networks, the PC1 could separate the steady states into clusters, defining clear phenotypes for all three networks (Figure S4). The nodes of different teams showed opposite signs in PC1 loading (Figure S5A). This finding is particularly noteworthy for the pluripotency network, given its weak team strength.

These three networks demonstrated PC1 stability similar to the EMP network (Figures S5B and S5C). Upon randomization, we observed a similar saturating increase in PC1 variance with team strength (Figure 4B), leading to a positive correlation between PC1 variance and team strength (Figure 4C). Correspondingly, we observed a negative correlation between the number of swaps in the network and PC1 variance (Figures 4C and S5D), further establishing that strong teams are sufficient for low-dimensional phenotypic space. The random networks also had lower PC1 variance as compared to their wild-type counterparts, suggesting low dimensionality as a significant condition imposed on the evolution of these networks.

Encouraged by the strong trends between team strength and PC1 variance, we wanted to ask if the presence of teams and team strength can more generally be used as a metric to predict the effective dimensionality of a network (than just the four networks studied so far). While the random networks do capture a wide range of PC1 variance and team strength values, they may have topological features inherited from their corresponding biological networks that are not captured by team strength but affect the dimensionality. Hence, we decided to analyze a larger cohort of networks from the cell collective database,[29] which contains networks of different sizes, densities, and biological functions (e.g., cell fate decision, signaling, and metabolism). We extracted the topologies of these networks, simulated them using RACIPE, and estimated the dimensionality of the emergent phenotypic space via PCA. Across 70 networks, we found (as predicted) a significant positive correlation between

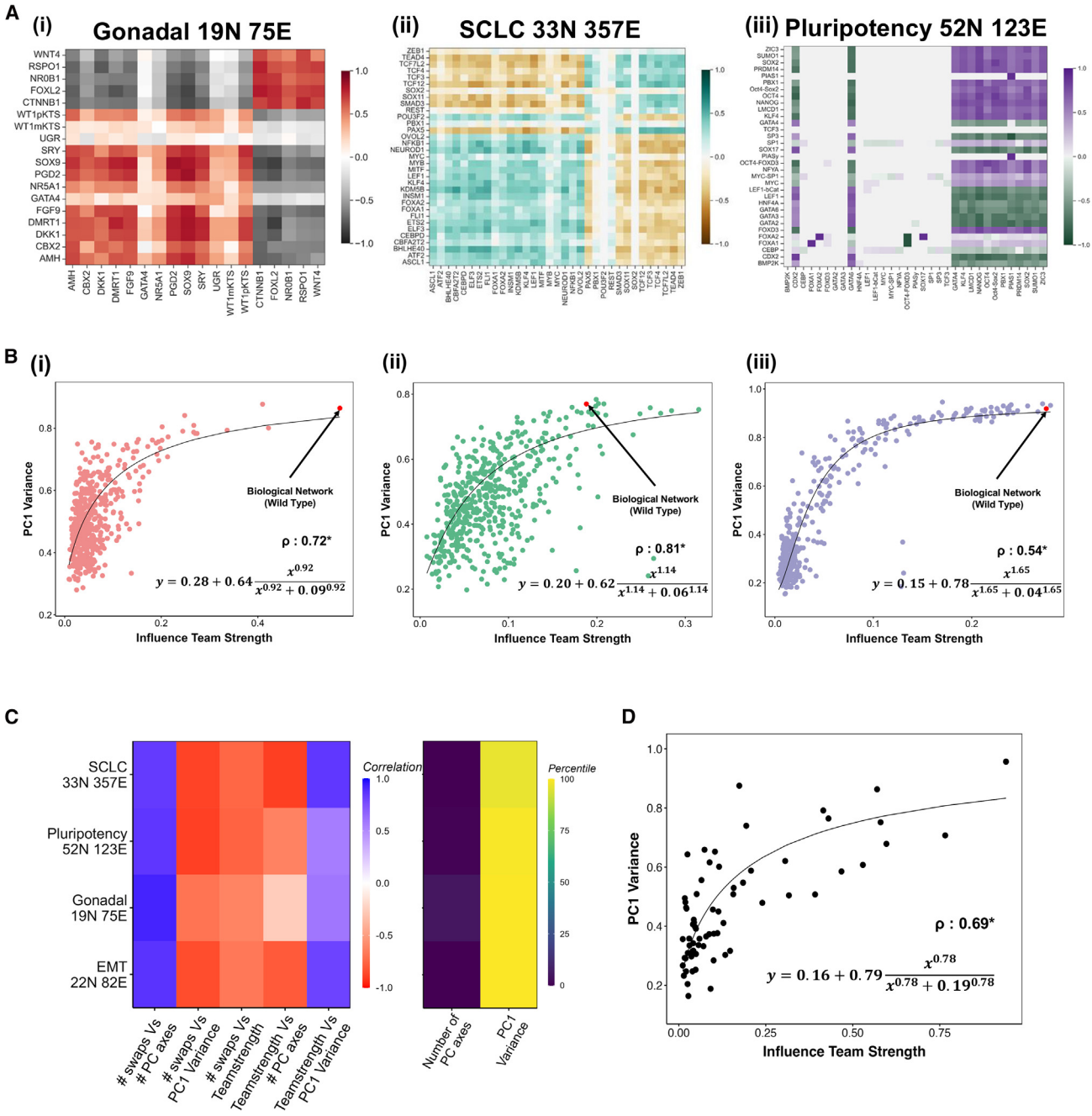


Figure 4. Team structure leads to low-dimensionality of steady-state space in SCLC, iPSC, and gonadal cell-fate decision networks

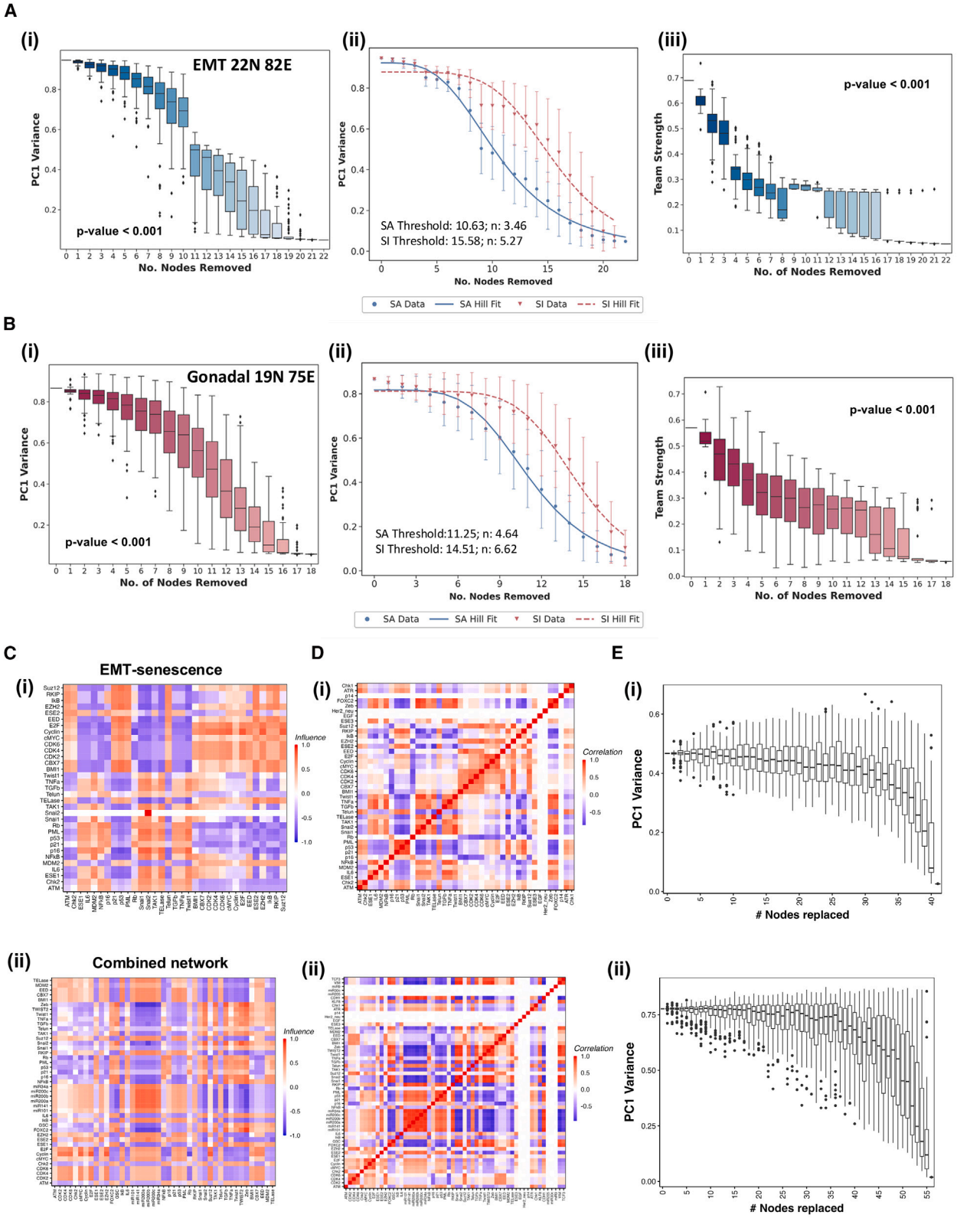
(A) Influence matrices for (i) Gonadal, (ii) SCLC, and (iii) iPSC network.

(B) Scatterplot depicting the dependence of variance explained by PC1 on the team strength for (i) Gonadal, (ii) SCLC, and (iii) iPSC network. Each point is a random network. The biological network has been pointed out. The Spearman correlation coefficient (ρ) is reported, with * indicating a p – value <0.05.

(C) Heatmap depicting the percentile of Biological networks for PC1 Variance, number of PC axes in random networks, and correlation strength between different metrics analyzed here.

(D) Scatterplot of PC1 variance against team strength for networks from cell collective database. The Spearman correlation coefficient (ρ) is reported, with * indicating a p – value <0.05.

team strength and PC1 variance (Figure 4D). Interestingly, we found that only a small percentage of networks (10%) in this cohort have a large team strength and correspondingly high PC1 Variance. Together, these results assert that strong teams can regulate the dimensionality of the emergent phenotypic space of regulatory networks in various contexts and establish



(legend on next page)

the utility of team strength to predict the dimensionality of the phenotypic space.

Strong teams provide robustness to low-dimensionality against perturbations

Our random network analysis suggests that the team structure in biological networks gives rise to low-dimensional phenotypic space due to the strong phenotypic reinforcement provided by the teams. Swapping edges reduces team strength by introducing inconsistent links such as inhibitions within teams and activations across teams, thereby proposing team strength as a metric to predict the emergent dimensionality of a network. To further establish the causal relationship between team strength and low dimensionality, we wanted to employ other ways of weakening teams that would affect dimensionality directly. Since teams are formed due to the strong interactions within the network, we chose to disrupt teams by disconnecting nodes from the network. This method maintains the consistency of the edges, unlike randomization, while reducing the density of the network ($\frac{E}{N^2}$) Biologically, disconnected nodes can imply mutations or nodes erroneously included in the network.

Briefly, we sampled n nodes at a time ($n \in [1, N]$), disconnected them by removing their incoming and outgoing edges, and adding either a self-activation (SA) or self-inhibition (SI). We repeated this exercise 100 times for each value of n. As expected, team strength decreases with an increase in the number of disconnected nodes (Figure 5A, i). For the EMT 22 node network, the PC1 variance showed a sigmoidal decline as the number of disconnected nodes increased (Figure 5A, ii). The mean values of the PC1 variance distributions of these networks showed a sigmoidal decrease with an increasing number of nodes disconnected (Figure 5A, iii). Adding SI to the disconnected nodes leads to a higher threshold value than adding SA. A possible explanation could be the noise suppression characteristics of the self-inhibition motif, in contrast to the noise-amplifying behavior of the self-activation motif. Thus, the self-activation-based curve can represent the maximum resilience offered by the team structure against erroneous inclusion of unrelated genes in genesets corresponding to cell-fate decision-making. Note that the number of nodes at which the PC1 variance starts to decrease corresponds to a mean team strength of around 0.3. A similar trend could also be seen for the gonadal network: a decrease in the team strength (Figure 5B, i), a sigmoidal decrease in the PC1 variance with an increase in the number of nodes disconnected (Figures 5B, ii and 5B, iii). Note that the PC1 variance remains unchanged for a larger percentage of disconnected nodes in the Gonadal network (6/19 = 32%) compared to the EMP network (4/22 = 19%). This contrast can be explained by the reduction rate of team strengths in both cases. EMP network sees a much sharper decrease in team strength against the number of nodes disconnected (Figure 5A, iii) compared to the Gonadal network (Figure 5B, iii). Together, these results support a causal relationship between the team strength and the low dimensionality of the phenotypic space.

Our results have dealt with GRNs underlying cell-fate decision systems, networks that drive the decision between two fates. However, each of these networks was focused on a specific biological system and did not necessarily depict the behavior of the phenotypic space when these networks interact with the global regulatory network of the corresponding organism.

Based on our results from Figures 4 and 5, we hypothesized that even the networks with weak teams could exhibit a low-dimensional phenotypic landscape when interacting with an appropriate team structure in a more global context.

To this end, we identified a network that models the immortalization of epithelial cells through the inhibition of senescence.[30] This network has an edge density of 5% and has a set of negative feedback loops involving p53 and p16 (Figure S6). The low density and negative feedback loops together lead to poor team strength (0.06), and no clear team structure is visible either in the influence matrix or in the correlation matrix generated by simulating this network using RACIPE (Figures 5C, i and S6). We calculated the PC1 variance for this network to be 50%. In addition to having weak teams, this network has very limited overlap with the high-team strength EMT network (22N 82E) we analyzed in this manuscript. Based on our hypothesis, if the low-team strength EMT-senescence network is combined (using known interactions) with EMP nodes and edges from a strong-team strength network, then the PC1 variance of the combined network's phenotypic space should increase.

The combined network with an edge density 4.5% (Figure S6) shows a better resolution of teams in the influence matrix (Figure 5C, ii). We see the effect even more clearly in the correlation matrix, with an increased size of positively correlated node groups (red squares) and strengthened negative correlation between teams (Figure 5D). While the combined network still has a relatively weak team strength of 0.18, it has a PC1 variance of 75% (Figure 5E). Surprisingly, no significant differences were observed in the value of the half-max fraction of genes replaced between the original and the combined network. It is important to note that despite the low density of the network and the fact that the size of the network with weak teams (43N 91E) was larger than the 22N 82E EMT network, the combined network achieves high PC1 variance just by having a part of the network having strong teams. This analysis suggests that teams can reduce the dimensionality emergent from the networks they interact with globally. This may have significant implications for

Figure 5. Teams provide robustness to low-dimensionality against internal and external perturbations

(A) For the 22N EMT network, (i) Boxplot describing the change in PC1 Variance with the number of nodes replaced with self-activating disconnected nodes, (ii) Mean ± sd PC1 variance against the number of nodes replaced with self-activating (blue) and self-inhibiting (red) disconnected nodes, (iii) Team strength against the number of nodes replaced. Fits to the sigmoidal curve are shown using blue and red lines in (ii), and the corresponding parameters are reported in the plot.

(B) Same as A but for the Gonadal cell-fate network.

(C) Influence matrices of the EMT senescence network before (i) and after (ii) addition of the strong teamed 22N 82E EMP network.

(D) Correlation matrices of the EMT senescence network before (i) and after (ii) addition of the strong teamed 22N 82E EMP network.

(E) Boxplots depicting PC1 stability of the (i) EMT senescence and (ii) combined networks.

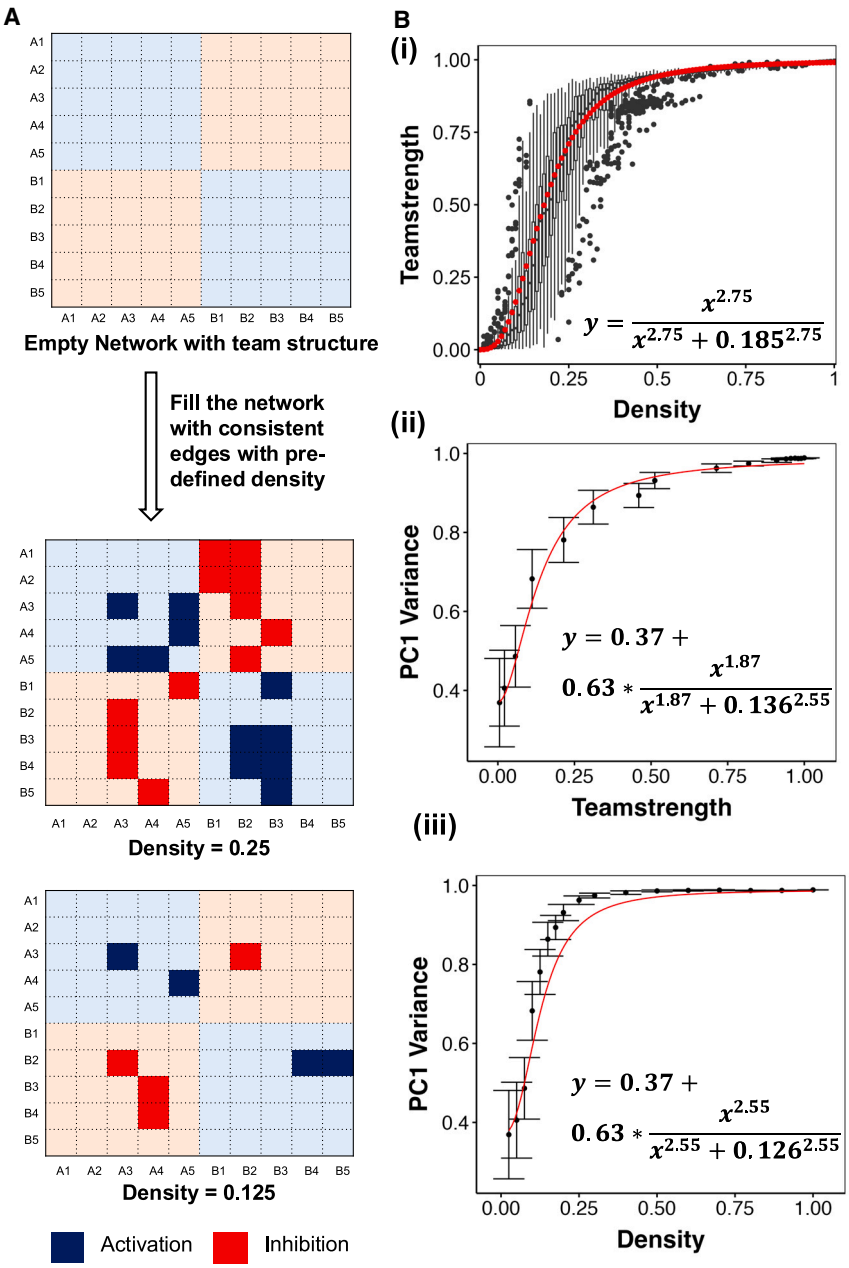


Figure 6. Artificial network analysis to validate the generalizability of team strength as a predictor of PC1 variance

(A) Schematic for generating artificial networks with teams.

(B) (i) Boxplots depicting the change in team strength as a function of the density of the artificial networks. (ii) Scatterplot depicting the change in PC1 Variance with team strength of artificial networks. (iii) Same as (ii), but with density. The error bars in (ii) and (iii) represent mean ± standard deviation.

understanding how cell fate transitions such as EMT coordinate with other physiological processes such as metabolism.[31]

Artificial networks reveal edge density as a key variable affecting team strength and dimensionality

Our results have been obtained for biological networks, and randomized networks have been generated therefrom. While these networks feature mutually inhibiting teams of nodes and low dimensionality of the emergent phenotypic space, topological structures other than team structure (such as positive and negative feedback loops or degree distributions) may have a significant role. Therefore, to further elucidate the causative connection between team structure and low dimensionality, we generated artificial networks of size 20 with pre-defined team structure and varying edge densities (Figure 6A). By predefining the team structure, we ensure that the edges between the nodes belonging to the same team are only activating, while those between nodes belonging to different teams are always inhibiting. Each edge in the network has an equal probability of being filled, and the probability value determines the network density. Since the influence between a pair of nodes is a weighted sum of the paths running between them, we hypothesized that by controlling the edge density, we could control the typical number of paths between nodes and, hence, the team strength. This hypothesis is supported by the biological networks analyzed in this manuscript. The pluripotency network with the lowest team strength has an edge density of 4.5% (123 connected node pairs out of 52*52 possible), whereas all other networks have higher densities (20–30%). This low density is visible in the influence matrix of the pluripotency network, where a fraction of the nodes have a very small fraction of nodes influencing them (see the columns in Figure 4A, iii). Note that these nodes can still influence other nodes, explaining this network's relatively dense correlation matrix (Figures S3B, iii). As expected, the mean team strength increases sigmoidally with density, with a threshold of 0.185 (18.5% edges filled, Figure 6B, i). Note that most change happens only under a density limit of 0.3 (30% edges filled), which can serve as a cutoff on the density of binary cell-fate decision networks. Indeed, the biological networks we have studied so far, with SCLC being the densest, obey this limit.

We generated 50 networks at each density and calculated their PC1 variance and the number of PC axes required to explain 90% variance. The dependence of PC1 Variance on team strength (Figure 6B, ii) had similar parameters for artificial networks as that of biological and random networks, further

establishing the usefulness of team strength as a predictor of dimensionality across contexts. The PC1 variance showed a sigmoidal increase with the density, with a similar Hill coefficient but hitting saturation earlier than that of team strength (Figure 6B, iii). The number of PC1 axes required to explain 90% variance also showed a sigmoidal decrease with increasing team strength (Figure S7). Together, these results establish that the relationship between team strength and dimensionality of the phenotypic space is generalizable to a broad class of networks, further emphasizing team strength as an important design principle that exerts strong control over the emergent phenotypic space.

Teams are necessary for stable low dimensionality

So far, we have established that mutually inhibiting teams of nodes, a topological feature observed across various contexts of cell-fate decision systems, are sufficient for the emergence of low-dimensionality in the phenotypic space. However, binary cell fate decisions can also be performed using a toggle switch. Therefore, we asked if teams are necessary for low-dimensional phenotypic space. Tripathi et al. have shown that high PC1 variance is a characteristic feature of "low-frustration" networks.[16] This study defined low frustration as a property of the steady state space derived by simulating the network using an Ising model formalism. The frustration of a steady state is calculated as the fraction of frustrated edges in the network, and a frustrated edge, in turn, is defined as an edge with the opposite sign as the product of the expression of the nodes connected by the edge. For example, an activating edge would be frustrated if the nodes had opposite expressions (−1 and 1) in a given steady state. Here, we have developed a more network-topology-based understanding of PC1 variance and low dimensionality. We have shown that the presence of mutually inhibiting teams is sufficient for robust low-dimensionality, with team strength as a satisfactory metric to qualify the dimensionality of a network structure. While networks with teams satisfy the condition of low frustration,[5] other network structures might also exhibit low frustration.

Therefore, we constructed null models that can lead to binary cell-fate decisions with varying structural features. Binary cell fate decision systems are often modeled as toggle switches (i.e., two transcription factors inhibiting the expression of each other) that can give rise to two mutually exclusive phenotypes (low-high and high-low). Furthermore, each phenotype is driven by the coexpression of multiple genes. An intuitive way to achieve this could be to have the nodes of the toggle switch act as master regulators activating all the other genes in the corresponding phenotype, creating a "hub" (Figure 7A, i). We simulated this network using RACIPE and estimated the PC1 properties of the emergent steady-state space.

We found that a hub network can indeed result in a correlation matrix showing a clear separation between the two phenotypes (Figure 7A, ii) and a high PC1 variance (72%), despite the heterogeneity in the parameter sets. We then simulated the gene replacement experiment performed on transcriptomic data by replacing the steady-state expression vectors of nodes with those of unregulated nodes of randomly sampled production and degradation rates. The hub network could not stabilize the PC1 variance, showing a steady decrease in the mean PC1 variance with the number of nodes replaced (Figure 7A, iii). We observed a steady increase in the number of PC axes required to capture 90% variance in the data (Figure S8A, i). To reinforce the phenotypes, we added activation edges within the nodes of each phenotype, in addition to the hub (Figure 7B, i). While this addition led to a small decrease in the PC1 variance, in the correlation matrix (Figure 7B, ii), we see an increase in the correlation of nodes within a phenotype but a reduction in the correlation between the nodes across the phenotypes. In addition, we found that the network now shows a stable PC1 variance (Figures 7B, iii and S8A, ii). We further updated the network by adding interactions in different directions (activations from downstream nodes to the toggle switch nodes, inhibitions from toggle switch to opposite phenotype, inhibitions between phenotypes). However, uniformly strong correlation and high and stable PC1 variance were achieved only for networks with activation links within a phenotype and inhibitions across, i.e., mutually inhibiting teams (Figure 7C), as seen in biological networks. Furthermore, we find that low dimensionality is independent of the symmetry in team sizes in the network (Figure S8B). These results thus establish that teams are necessary and sufficient for robust low-dimensionality in the phenotypic space of a cell-fate decision-making system.

DISCUSSION

Cell-fate decisions are ubiquitous in biological systems. Cells dynamically change phenotypes in response to varying biochemical and biomechanical signals. During the development of an organism, starting from a totipotent stem cell, a sequence of decision-making events plays out, often through binary branching points, to facilitate cellular differentiation, which is well-controlled in space and time.[32] The robustness of decision-making is frequently observed in developmental contexts, where phenotypes are sensitive only to a small set of specific perturbations, a property known as canalization.[33] Thus, it becomes imperative to understand how only a few phenotypes are generated despite the size and complexity of GRNs underlying such phenotypic changes. Recent studies have begun to elucidate the design principles of underlying regulatory networks from various biological contexts: canalization[34] and giving rise to relatively simple, low-dimensional dynamics.[16]

Our results suggest that low dimensionality in phenotypic space is an emergent consequence of a latent structure in these GRNs —the presence of "teams" of nodes where the members within a team activate each other, and those across teams inhibit one another, forming a higher-order toggle switch.

Studies dealing with cell-fate decisions often identify the presence of individual toggle switch motifs as the origin of a binary decision.[8] However, regulatory networks underlying such systems are often large and complex and consist of multiple toggle switches. This strongly suggests a role for higher-order structural features in driving decision-making, an idea not presented in those earlier works. We had previously identified mutually inhibiting teams of nodes in GRNs driving the phenotypic decisions in EMP,[5] SCLC,[35] and melanoma.[36] We had shown that mutually inhibiting "teams" allow for binary cell-fate decisions, much like a single toggle switch, while ensuring the robustness

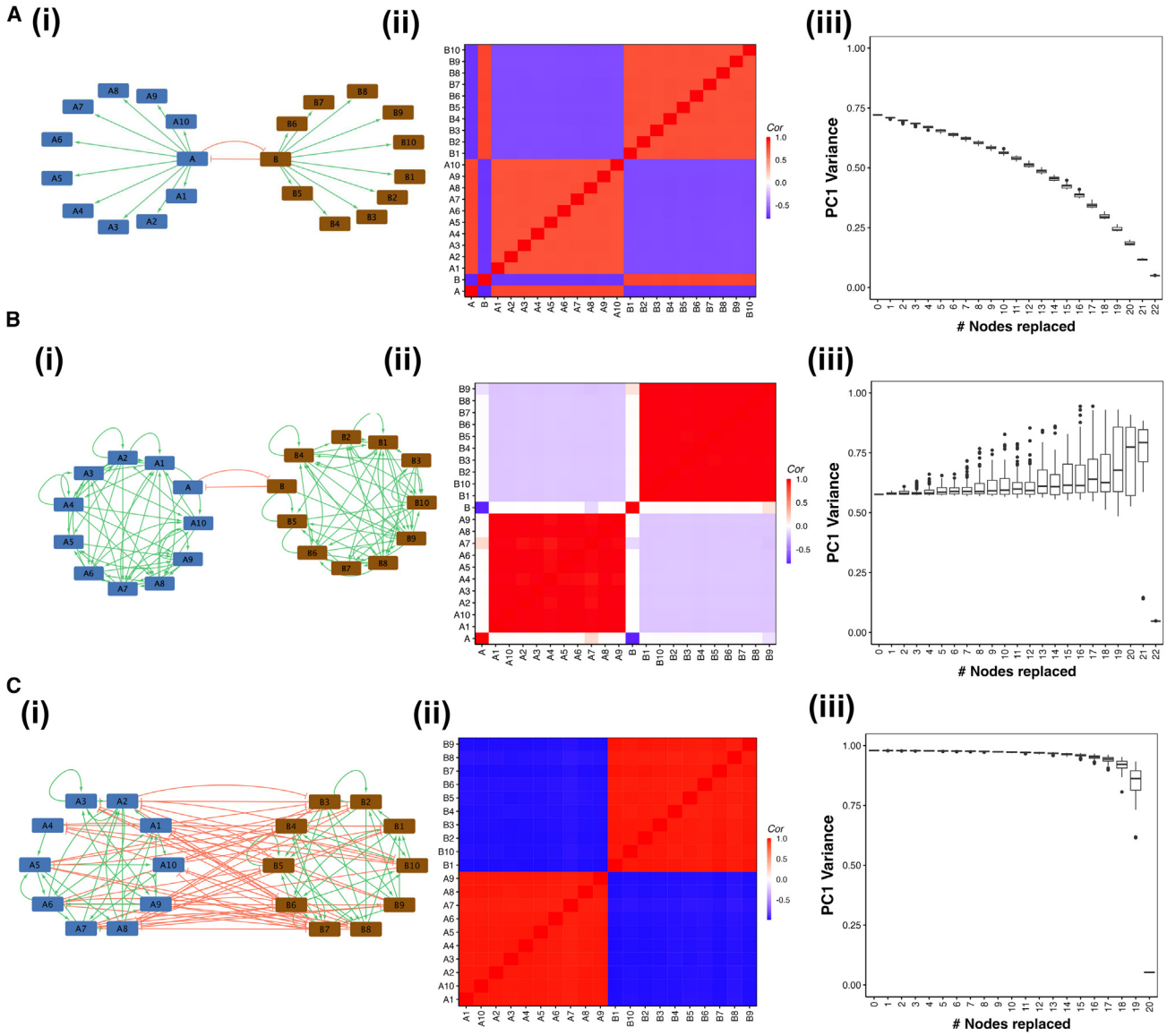


Figure 7. Teams are necessary to achieve PC1 stability

(A) Correlation and PC1 stability analysis for a toggle switch between master regulators. Each master regulator activates multiple nodes corresponding to their phenotype, forming a hub.

(B) Same as A, but for master regulators activating teams of nodes.

(C) Same as A, but for a network with mutually inhibiting teams.

of the phenotypic space to a large set of perturbations to the network structure (Figure 5).

We found that "teams" could lead to minimal frustration in biological networks. Recently, multiple biological networks were found to be minimally frustrated and capable of enabling a low-dimensional phenotypic space.[16,37] Our results provide a mechanistic understanding of these trends; we find that only in networks with very high team strength is most of the variance in the steady-state space (>90%) explained by PC1. By comparing the behavior of WT biological networks with their randomized counterparts, we establish that the stronger the "teams", the higher the variance percentage explained by PC1. Thus, networks showing weak team strength have higher dimensional dynamics in phenotypic space because it takes many more PCs together to be able to recapitulate 90% variance. It is important to note that the team strength of a network can serve as a strong predictor of the PC1 variance of the emergent phenotypic space, as evidenced by the similarity of fit parameters across random networks sourced from WT networks (Figures 3A, ii, 4B, and 5D exponents of x are close to 1).

This analysis is also strengthened by our calculations for two-team artificial networks, where we show how team strength and

PC1 variance increase sigmoidally as a function of the density of edges in the network. By approximately 25% density, team strength and PC1 variance both show largely saturating behavior, suggesting that only one-fourth of edges connected in a two-team topology may be sufficient to allow for low dimensional dynamics in phenotypic space. Intriguingly, biological networks studied here show similar density values in terms of the mean number of edges per node. Therefore, artificial networks provide an amenable tool to derive generalizable design principles. Future directions include identifying how impurities in network topology (i.e., members within a team inhibiting each other and/or members across teams activating each other) or how the presence of more than two teams (e.g., as expected to be present in $CD4^+$ T cell differentiation[38]) changes these trends observed here. Overall, we establish that mutually inhibiting two "teams" of nodes is a design principle underlying the low dimensionality of phenotypic space often seen experimentally.

Importantly, we detected low dimensionality in corresponding transcriptomic data of SCLC and EMT-associated gene sets in CCLE datasets and lung tissue genesets in the GTEx dataset. Low dimensionality is maintained despite replacing many genes in the signatures with housekeeping genes, suggesting the presence of strong network topology features such as teams of nodes underlying the control of coordinated expression levels. We further showed through the simulations of the EMT-senescence network[30] that subnetworks with strong teams can control the dimensionality of the larger networks they interact with. This observation can potentially explain the retention of low-dimensionality of transcriptomic data despite contaminating the gene signatures with potentially unrelated genes.

As two mutually inhibiting teams lead to the emergence of two robust phenotypes with mutually exclusive compositions, it can be intuitive that the teams are sufficient for high PC1 variance. But are teams necessary? Our analysis reveals that teams are not necessary for high PC1 variance and correspondingly low dimensionality. However, teams are necessary to obtain the PC1 stability feature observed in transcriptomic data. We studied other potential network architectures that can lead to stable binary cell-fate decisions and found that these networks can indeed lead to a high PC1 variance and low-dimensional phenotypic space. However, PC1 stability was not achieved until we established team-like connections within nodes of a particular phenotype. It is important to note that in the hub network in Figure 7, the network toggle switch can also represent a mutually inhibiting team structure. Thus, another implication of the necessity of teams can be that genes in transcriptomic data that show group behavior in correlation may not be connected if they do not exhibit PC1 stability. Thus, our results have the potential to contribute toward constructing GRNs from transcriptomic data corresponding to cell-fate decision systems.

Our current analysis is largely limited to one axis of plasticity in each context, e.g., epithelial-mesenchymal, Sertoli-granulosa, and non-neuroendocrine/neuroendocrine cell fate determination. Cancer cells have been shown to exhibit multiple interconnected axes of phenotypic transitions (e.g., EMP, drug resistance, stemness, and metabolic plasticity).[39–42] In most cases, these interconnections are coherent, i.e., the larger network can be resolved into two teams with each phenotype exclusively belonging to only one of the two teams.[43–46] Thus, for coherently connected axes, the emergent phenotypic space remains low-dimensional. However, while EMT networks may be enriched for higher-order feedback loops such as teams,[47] the cases of incoherent connections, such as the EMT-senescence network presented in Figure 5, can lead to interesting dynamics, including oscillations.[48,49] An important direction for the current study is understanding the low-dimensional dynamics scale across these incoherently connected axes of phenotypic heterogeneity and how the team structure evolves under such interactions. While such interactions can lead to interesting behaviors, such as the reduction of dimensionality emergent from the senescence network when coupled with an EMT network with strong teams, our analysis asserts the need for a comprehensive map of regulatory interactions for a better understanding of the dynamics of multiple axes of phenotypic plasticity.

A related direction of inquiry is the case of multi-fate systems. The design principles underlying the emergence of more than two cell fates from a single decision point are not well understood. One of the classic examples of multi-fate systems is the differentiation of pluripotent stem cells into different specialized cells, a process assumed to be driven by sequential activation of toggle switches.[8,50] Another example is the differentiation of T cells into phenotypes with mutually exclusive molecular signatures, where 3- and 4-node mutually inhibiting networks have been proposed as the possible core regulatory architecture.[38] A similar architecture of mutually inhibiting transcription factors via the formation of inactive heterodimers has been used to generate synthetic multi-fate systems in mammalian cells.[51] However, larger networks are required to explain the stability of these phenotypes (such as the two-team architecture, which allows two stable and robust phenotypes). Intuitively, such systems would give rise to multi-dimensional phenotypic spaces, requiring more than one PC axes to explain the variance. Analyzing the PC-robustness of these multiple axes could potentially provide clues to the design principles of the underlying GRNs.

Our analysis is limited to dimensions constructed as linear combinations of genes involved in the GRNs. However, it must be noted that biological interactions and emergent behavior are non-linear. Our simulations account for the non-linearity of the interactions by using sigmoidal functions to represent the regulatory terms in the ODEs.[26] We have shown here that the effect of nonlinearity in regulation is masked in the emergent phenotypic space in the presence of strong, mutually inhibiting teams. However, assessing the dimensionality of networks with weak teams, especially the random networks generated from the WT networks and the cell-collective networks, perhaps requires non-linear methods. Accounting for non-linear aspects could allow us to identify additional design principles that can govern the dimensionality of phenotypic space without strong teams. Such an extension would also allow us to understand incoherently interacting switches leading to high-dimensional or cyclic phenotypic space[48,49] and multi-fate decision systems.

Limitations of the study

Our current analysis focuses only on GRNs governing binary cell fate determination events and, as a result, is concerned with only

one axis of plasticity. However, it has been observed that these decisions are often influenced and coupled with other programs.[39–42] Understanding how these programs coordinate and are coupled with respect to the teams framework needs to be studied in detail. Through the developmental timescale, cells have been shown to make a series of decisions that eventually lead them to differentiate into a terminally differentiated phenotype.[8,50] Many of these events are binary, but there are also cases where cells can potentially differentiate into more than two different cell types; in these cases, the current analysis of two teams would not hold.[38,51] Our analysis depends on PCA to capture the variance and, hence, the dimensionality of the data, which works very well for binary cell fate networks with two distinct phenotypes. But, if the interactions between genes are highly nonlinear and incoherent the study of steady states would not be suitable as these types of interactions vary temporally (as opposed to invariant steady-state dynamics), capturing such behaviors for studying coordination between genes in a GRN would need more sophisticated methods.[48,49]

RESOURCE AVAILABILITY

Lead contact

For additional resources or inquiries, please contact Prof. Mohit Kumar Jolly mkjolly@iisc.ac.in.

Materials availability

This paper does not include newly generated reagents.

Data and code availability

- The data used in this paper are publicly available.
- All processed data and scripts used in the manuscript are uploaded and available at the GitHub repository: https://github.com/aashnasaxena/Teams-PC1.
- The raw data generated for this study and the derived datasets supporting the current findings are available from the lead author upon reasonable request.

ACKNOWLEDGMENTS

K.H. was supported by the Prime Minister's Research Fellowship (PMRF) awarded by the Government of India and the Center for Theoretical Biological Physics, NSF PHY-2019745. P.H. was supported by the PMRF, awarded by the Government of India. M.K.J. was supported by the Ramanujan Fellowship (SB/S2/RJN-049/2018) awarded by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India. M.K.J. was also supported by Param Hansa Philanthropies. H.L. was supported in part by the Center for Theoretical Biological Physics, NSF PHY-2019745.

AUTHOR CONTRIBUTIONS

M.K.J. and K.H. designed the research; K.H., P.H., and T.M. analyzed the data; A.S., A.G., P.H., and K.H. carried out simulations; all authors discussed the results and participated in the preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing financial or non-financial interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111730.

Received: July 4, 2024
Revised: September 14, 2024
Accepted: December 30, 2024
Published: January 2, 2025

REFERENCES

1. Vickaryous, M.K., and Hall, B.K. (2006). Human cell type diversity, evolution, develop- ment, and classification with special reference to cells derived from the neural crest. Bio- logical Reviews *81*, 425–455. https://doi.org/10.1017/s1464793106007068.
2. Tabula Sapiens Consortium, Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple- organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.eprint.
3. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405–411. https://doi.org/10.1126/science.aax0249.
4. Jolly, M.K., Ware, K.E., Gilja, S., Somarelli, J.A., and Levine, H. (2017). EMT and MET: necessary or permissive for metastasis? Mol. Oncol. *11*, 755–769. https://doi.org/10.1002/1878-0261.12083.
5. Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A., and Jolly, M.K. (2022). Landscape of epithelial–mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. eLife *11*, e76535. https://doi.org/10.7554/eLife.76535.
6. Hong, T., Watanabe, K., Ha Ta, C., Villarreal-Ponce, Q., Nie, Q., and Dai, X. (2015). An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. PLoS Comp. Biol. *11*, e1004569. https://doi.org/10.1371/journal.pcbi.1004569.
7. Roca, H., Hernandez, S., Weidner, R.C., McEachin, D., Fuller, S., McEachin, R.C., Fuller, D., Sud, S., Schumann, T., Wilkinson, J.E., et al. (2013). Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. PLoS ONE *8*, e76773. https://doi.org/10.1371/journal.pone.0076773.
8. Zhou, J.X., and Huang, S. (2011). Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends Genet. *27*, 55–62. https://doi.org/10.1016/j.tig.2010.11.002.
9. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468. https://doi.org/10.1038/s41586-018-0040-3.
10. Lu, M., Jolly, M.K., Levine, H., Onuchic, J.N., and Ben-Jacob, E. (2013). MicroRNA- based regulation of epithelial–hybrid–mesenchymal fate determination. Proc. Natl. Acad. Sci. USA *110*, 18144–18149. https://doi.org/10.1073/pnas.1318192110.
11. Tian, X.-J., Zhang, H., and Xing, J. (2013). Coupled Reversible and Irreversible Bistable Switches Underlying TGFβ-induced Epithelial to

Mesenchymal Transition. Biophys. J. *105*, 1079–1089. https://doi.org/10.1016/j.bpj.2013.07.011.

12. Jolly, M.K., Somarelli, J.A., Sheth, M., Biddle, A., Tripathi, S.C., Armstrong, A.J., Hanash, S.M., Bapat, S.A., Rangarajan, A., and Levine, H. (2019). Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. Pharmacol- ogy & Therapeutics *194*, 161–184. https://doi.org/10.1016/j.pharmthera.2018.09.007.
13. Subbalakshmi, A.R., Ashraf, B., and Jolly, M.K. (2022). Biophysical and biochemical at- tributes of hybrid epithelial/mesenchymal phenotypes. Phys. Biol. *19*, 025001. https://doi.org/10.1088/1478-3975/ac482c.
14. Font-Clos, F., Zapperi, S., and La Porta, C.A.M. (2018). Topography of epithelial–mesenchymal plasticity. Proc. Natl. Acad. Sci. USA *115*, 5902–5907. https://doi.org/10.1073/pnas.1722609115.
15. Steinway, S.N., Zañudo, J.G.T., Michel, P.J., Feith, D.J., Loughran, T.P., and Albert, R. (2015). Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transi- tion and support hybrid cellular phenotypes. NPJ Syst. Biol. Appl. *1*, 15014. https://doi.org/10.1038/npjsba.2015.14.
16. Tripathi, S., Kessler, D.A., and Levine, H. (2023). Minimal frustration underlies the use- fulness of incomplete regulatory network models in biology. Proc. Natl. Acad. Sci. USA *120*, e2216109120. https://doi.org/10.1073/pnas.2216109120.eprint.
17. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wil- son, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line En- cyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607. https://doi.org/10.1038/nature11003.
18. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.
19. Udyavar, A.R., Wooten, D.J., Hoeksema, M., Bansal, M., Califano, A., Estrada, L., Schnell, S., Irish, J.M., Massion, P.P., and Quaranta, V. (2017). Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity. Cancer Res. *77*, 1063–1074.
20. Lissa, D., Takahashi, N., Desai, P., Manukyan, I., Schultz, C.W., Rajapakse, V., Velez, M.J., Mulford, D., Roper, N., Nichols, S., et al. (2022). Heterogeneity of neuroendocrine transcriptional states in metastatic small cell lung cancers and patient-derived models. Nat. Commun. *13*, 2023. https://doi.org/10.1038/s41467-022-29517-9.
21. Aiello, N.M., Maddipati, R., Norgard, R.J., Balli, D., Li, J., Yuan, S., Yamazoe, T., Black, T., Sahmoud, A., Furth, E.E., et al. (2018). EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. Dev. Cell *45*, 681–695.e4. https://doi.org/10.1016/j.devcel.2018.05.027.
22. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.
23. Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J., et al. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature *587*, 619–625. https://doi.org/10.1038/s41586-020-2922-4.
24. Somepalli, G., Sahoo, S., Singh, A., and Hannenhalli, S. (2021). Prioritizing and characterizing functionally relevant genes across human tissues. PLoS Comp. Biol. *17*, e1009194. https://doi.org/10.1371/journal.pcbi.1009194.
25. Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet. *29*, 569–574. https://doi.org/10.1016/j.tig.2013.05.010.
26. Huang, B., Lu, M., Jia, D., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2017). Interrogat- ing the topological robustness of gene regulatory circuits by randomization. PLoS Comput. Biol. *13*, e1005456. https://doi.org/10.1371/JOURNAL.PCBI.1005456.
27. Chang, R., Shoemaker, R., and Wang, W. (2011). Systematic search for recipes to gen- erate induced pluripotent stem cells. PLoS Comput. Biol. *7*, e1002300.
28. Ríos, O., Frias, S., Rodríguez, A., Kofman, S., Merchant, H., Torres, L., and Mendoza, L. (2015). A Boolean network model of human gonadal sex determination. Theor. Biol. Med. Model. *12*, 26. https://doi.org/10.1186/s12976-015-0023-0.
29. Helikar, T., Kowal, B., McClenathan, S., Bruckner, M., Rowley, T., Madrahimov, A., Wicks, B., Shrestha, M., Limbu, K., and Rogers, J.A. (2012). The Cell Collective: Toward an open and collaborative approach to systems biology. BMC Syst. Biol. *6*, 96. https://doi.org/10.1186/1752-0509-6-96.
30. Méndez-López, L.F., Davila-Velderrain, J., Domínguez-Hüttinger, E., Enríquez-Olguín, C., Martínez-García, J.C., and Alvarez-Buylla, E.R. (2017). Gene regulatory network underlying the immortalization of epithelial cells. BMC Syst. Biol. *11*, 24. https://doi.org/10.1186/s12918-017-0393-5.
31. Jia, D., Park, J.H., Kaur, H., Jung, K.H., Yang, S., Tripathi, S., Galbraith, M., Deng, Y., Jolly, M.K., Kaipparettu, B.A., et al. (2021). Towards decoding the coupled decision- making of metabolism and epithelial-to-mesenchymal transition in cancer. Br. J. Cancer *124*, 1902–1911.
32. Agozzino, L., Balázsi, G., Wang, J., and Dill, K.A. (2020). How Do Cells Adapt? Stories Told in Landscapes. Annu. Rev. Chem. Biomol. Eng. *11*, 155–182. https://doi.org/10.1146/annurev-chembioeng-011720-103410.
33. Du, Z., Santella, A., He, F., Shah, P.K., Kamikawa, Y., and Bao, Z. (2015). The Regulatory Landscape of Lineage Differentiation in a Metazoan Embryo. Dev. Cell *34*, 592–607. https://doi.org/10.1016/j.devcel.2015.07.014.
34. Kadelka, C., Butrie, T.-M., Hilton, E., Kinseth, J., and Serdarevic, H. (2020). A meta- analysis of Boolean network models reveals design principles of gene regulatory net- works. arXiv. https://doi.org/10.48550/ARXIV.2009.01216. https://www.science.org/doi/10.1126/sciadv.adj0822.
35. Chauhan, L., Ram, U., Hari, K., and Jolly, M.K. (2021). Topological signatures in regula- tory network enable phenotypic heterogeneity in small cell lung cancer. Elife *10*, e64522. https://doi.org/10.7554/elife.64522.
36. Pillai, M., and Jolly, M.K. (2021). Systems-level network modeling deciphers the master regulators of phenotypic plasticity and heterogeneity in melanoma. iScience *24*, 103111. https://doi.org/10.1016/j.isci.2021.103111.
37. Tripathi, S., Kessler, D.A., and Levine, H. (2020). Biological Networks Regulating Cell Fate Choice are Minimally Frustrated. Phys. Rev. Lett. *125*, 088101. https://doi.org/10.1103/physrevlett.125.088101.
38. Duddu, A.S., Sahoo, S., Hati, S., Jhunjhunwala, S., and Jolly, M.K. (2020). Multi-stability in cellular differentiation enabled by a network of three mutually repressing master regu- lators. J. R. Soc. Interface *17*, 20200631. https://doi.org/10.1098/rsif.2020.0631.
39. Thankamony, A.P., Saxena, K., Murali, R., Jolly, M.K., and Nair, R. (2020). Cancer Stem Cell Plasticity – A Deadly Deal. Front. Mol. Biosci. *7*, 79. https://doi.org/10.3389/fmolb.2020.00079.
40. Jia, D., Lu, M., Jung, K.H., Park, J.H., Yu, L., Onuchic, J.N., Kaipparettu, B.A., and Levine, H. (2019). Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. Proc. Natl. Acad. Sci. USA *116*, 3909–3918. https://doi.org/10.1073/pnas.1816391116.
41. Vipparthi, K., Hari, K., Chakraborty, P., Ghosh, S., Patel, A.K., Ghosh, A., Biswas, N.K., Sharan, R., Arun, P., Jolly, M.K., and Singh, S. (2022). Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer. iScience *25*, 104317. https://doi.org/10.1016/j.isci.2022.104317.
42. Bhatia, S., Monkman, J., Blick, T., Pinto, C., Waltham, M., Nagaraj, S.H., and Thompson, E.W. (2019). Interrogation of Phenotypic Plasticity between Epithelial and Mesenchymal States in Breast Cancer. J. Clin. Med. *8*, 893. https://doi.org/10.3390/jcm8060893.

43. Sahoo, S., Mishra, A., Kaur, H., Hari, K., Muralidharan, S., Mandal, S., and Jolly, M.K. (2021). A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer *3*, zcab027. https://doi.org/10.1093/narcan/zcab027.

44. Galbraith, M., Levine, H., Onuchic, J.N., and Jia, D. (2023). Decoding the coupled decision- making of the epithelial-mesenchymal transition and metabolic reprogramming in cancer. iScience *26*, 105719. https://doi.org/10.1016/j.isci.2022.105719.

45. Pasani, S., Sahoo, S., and Jolly, M.K. (2020). Hybrid E/M Phenotype(s) and Stemness: A Mechanistic Connection Embedded in Network Topology. J. Clin. Med. *10*, 60. https://doi.org/10.3390/jcm10010060.

46. Biddle, A., Gammon, L., Liang, X., Costea, D.E., and Mackenzie, I.C. (2016). Phenotypic Plasticity Determines Cancer Stem Cell Therapeutic Resistance in Oral Squamous Cell Carcinoma. EBioMedicine *4*, 138–145. https://doi.org/10.1016/j.ebiom.2016.01.007.

47. Nordick, B., and Hong, T. (2021). Identification, visualization, statistical analysis and mathematical modeling of high-feedback loops in gene regulatory networks. BMC Bioinfo *22*, 481. https://doi.org/10.1186/s12589-021-04405-z.

48. Deritei, D., Aird, W.C., Ercsey-Ravasz, M., and Regan, E.R. (2016). Principles of dy- namical modularity in biological regulatory networks. Sci. Rep. *6*, 21957. https://doi.org/10.1038/srep21957.

49. Deritei, D., Rozum, J., Ravasz Regan, E., and Albert, R. (2019). A feedback loop of con- ditionally stable circuits drives the cell cycle from checkpoint to checkpoint. Sci. Rep. *9*, 16430. https://doi.org/10.1038/s41598-019-52725-1.

50. Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451–467. https://doi.org/10.1016/j.cell.2016.06.011.

51. Zhu, R., Rio-Salgado, J. M. del, Garcia-Ojalvo, J., and Elowitz, M.B. (2022). Synthetic multistability in mammalian cells. Science *375*, eabg9765. https://doi.org/10.1126/science.abg9765.

52. Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Res. *38*, D750–D753. https://doi.org/10.1093/nar/gkp889.

53. R Core Team (2023). _R: A Language and Environment for Statistical Computing_ (Vienna, Austria: R Foundation for Statistical Computing).

54. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, W., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Analyzed Dataset	https://storage.googleapis.com/adult-gtex/bulk-gex/v8/rna-seq/GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz	Adult GTEx bulk RNAseq TPM dataset
Analyzed Dataset	https://storage.googleapis.com/adult-gtex/bulk-gex/v8/rna-seq/GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz	CCLE Gene expression TPM dataset
Epithelial Mesenchymal Transition geneset	Aiello et al.[21]	N/A
Small-cell lung cancer genesets	Udayavar et al.,[19] Lissa et al.[20]	N/A
Lung Alveolar genesets	Travaglini et al.[23]	N/A
FUGUE lung tissue geneset	Somepalli et al.[24]	N/A
Software and algorithms		
RACIPE	https://github.com/csbBSSE/Gene_Network_Modelling/releases/tag/v2.28	Modified from Huang et al.[28]
Analysis codes	https://github.com/aashnasaxena/Teams-PC1	N/A

METHOD DETAILS

Transcriptomic data analysis

We used the cancer cell-line expression data[17] and normal tissue bulk RNA-seq data from GTEx[18] to understand the dimensionality of the transcriptomic data. For CCLE dataset, we obtained genesets that capture the phenotypic heterogeneity in cancerous samples: SCLC[20,35] and EMP.[21] For GTEx dataset, we used genesets that capture phenotypic heterogeneity of the Fibrotic-Epithelial axis in lung (alveolar) tissue obtained from MsigDB C8 collection[22,23] and a functionally relevant geneset for lung tissue.[24] We first took the appropriate subsets of the cell lines/samples (49 SCLC cell lines, non-SCLC (848) and breast cancer (62) cell lines from CCLE data; Lung tissue samples from GTEx data). We divided each gene set into two parts corresponding to the two phenotypes of their corresponding axes (Epithelial- Mesenchymal for EMP, NE-nonNE for SCLC, and Fibrotic and Epithelial for Alveolar gene signatures). For the signatures obtained from MSigDB, we picked the top 100 genes based on absolute PC1 loading for further analysis. We took further subsets of the data corresponding to the gene sets. We then performed pca analysis on these subset data.

For each subset of data, we calculated a phenotypic score using the formula:

$$score = \frac{\sum_{i \in P1} e_i}{\#P1} - \frac{\sum_{i \in P2} e_i}{\#P2}$$

Where P1 and P2 are sets of genes corresponding to the two phenotypes in each axis, e_i is the expression of gene i in the cell line of interest, and #P1 is the number of genes in the set P1.

To perform the PC1 stability analysis, we replaced the expression of n ($0 \leq n \leq N$) randomly sampled genes for each geneset with n randomly selected housekeeping genes[25] (where N is the number of genes in the geneset), generating a new expression matrix. We generate 100 such matrices for each value of n, calculate the PC1 variance, and report the results in a boxplot.

RAndom CIrcuit PErturbaiton (RACIPE)

RACIPE[26] is a tool that simulates GRNs in a continuous state space. Given a GRN, it constructs a system of Ordinary Differential Equations representing the network. For a given node T and a set of input nodes P_i and N_j that activate and inhibit T respectively, the corresponding differential equation takes the following form:

$$\frac{dT}{dt} = G_T * \prod_i \frac{H^{S+}(P_i, P_{iT}^0, n_{P_i,T}, \lambda_{P_i,T})}{\lambda_{P_i,T}} * \prod_j H^{S-}(N_j, N_{iT}^0, n_{N_j,T}, \lambda_{N_j,T}) - k_T * T$$

Here, T, P_i, and N_j represent the concentrations of the species. G_T and k_T denote the production and degradation rates, respectively. P_i^0 is the threshold value of P_i concentration at which the non-linearity in the dynamics of T due to P_i is seen. n is termed as Hill-coefficient and represents the extent of non-linearity in the regulation. λ represents the fold change in the target node concentration

upon over-expression of the regulating node. Finally, the functions H^{S+} and H^{S-} are known as shifted hill functions and represent the regulation of the target node by the regulatory node. The hill shift function takes the following form:

$$H^{S+/-}\left(B, B_A^0, n_{B,A}, \lambda_{B,A}\right) = \frac{{B_A^0}^{n_{B,A}}}{{B_A^0}^{n_{B,A}} + B^{n_{B,A}}} + \lambda * \frac{B^{n_{B,A}}}{{B_A^0}^{n_{B,A}} + B^{n_{B,A}}}$$

Note that, for high values of the regulatory node concentration, $H^{S+/-}$ approaches λ.

For the model generated in this way, RACIPE randomly samples parameter sets from a pre-defined set of parameter ranges estimated from BioNumbers.[52] The ranges as reported by Huang et al.[26] are given in Table 1.

RACIPE integrates the model from multiple initial conditions at each parameter set and obtains steady states in the initial condition space. The output, hence, comprises the collection of parameter sets and corresponding steady states obtained from the model. For the current analysis, we used a sample size of 10,000 for parameter sets and 100 for initial conditions. The parameters were sampled via a uniform distribution and the ODE integration was carried out using Euler's numerical integration method.

Processing RACIPE output

For a given network with n nodes, the steady-state expression levels of the nodes obtained from RACIPE were normalized in the following way:

$$E_{in} = log_2\left(\frac{E_i}{f_i}\right)$$

$$f_i = \frac{g_i}{k_i}$$

Where, for the i^{th} node, E_{in} is the normalized expression level, E_i is the steady state expression level, f_i is the normalization factor, and g_i and k_i are production and degradation of the i^{th} node corresponding to the current steady state. The collection of these normalized steady states was used for PCA.

Network randomization

Network randomization is carried out by randomly swapping the edges in the network. Each iteration involves randomly selecting a pair of edges and swapping their edge type (activation/inhibition). For each random network, we carry out 10 to 100 such iterations. For each biological network, we generate 100 random networks.

Calculation of team strength

For each network, we first generate the adjacency matrix $A = \{-1, 0, 1\}^{NXN}$, where N is the number of nodes in a network. Each cell of the adjacency matrix represents an edge, with –1 for inhibition, 1 for activation, and 0 for no interaction. We then generate the influence matrix using the following formula:

$$Inf\, I_{I_{max}} = \frac{\sum_{I=1}^{I_{max}} \frac{Adj^I}{Adj^I_{I_{max}}}}{I_{max}}$$

where Adj_{max} is derived by setting all non-zero entries of the adjacency matrix to 1, and the division is element-wise. We then take the sign of the influence matrix, augment the transpose of the influence matrix with itself, and apply hierarchical clustering to identify the two teams. We then calculate the team strength as follows:

$$T_S = \frac{\sum_{k,l \in \{1,2\}} |T_{kl}|}{4}$$

Where,

$$T_{kl} = \frac{\sum_{i \in T_k, j \in T_l} Inf\, I_{ij}}{n_{kl}}, k, l \in \{1, 2\}$$

Disconnecting nodes in biological networks

We generated networks with disconnected nodes by randomly selecting n nodes ($0 \leq n \leq N$) and removing all incoming and outgoing edges connected to these nodes. Since RACIPE cannot simulate disconnected nodes, we added self-activation and self-inhibition to these nodes.

For each n we generate 50 networks, representing 50 random choices of n nodes to be disconnected.

We then simulate each network in RACIPE and estimate the PC1 variance and number of PC axes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Spearman correlation coefficients were calculated using the *cor.test* function from R 4.1.[53] The significance of correlation coefficients is reported within the figure panels as c* where c represents the correlation coefficient and * represents p-value < 0.05. One-way ANOVA test was performed using the function *f_oneway* from the *scipy*[54] package in Python 3.8. The corresponding p-value is reported within the figure. PCA analysis was performed using the *prcomp* function from R 4.1. The sigmoidal fits were performed using the *lm* function from R 4.1. Linearized forms of the Hill function used are given below:

For increasing sigmoidal function:

$$ln\left(\frac{1}{y+\epsilon 1}-1+\epsilon 2\right) = n*ln(K) - n*ln(x+\epsilon 3)$$

For decreasing sigmoidal function:

$$ln\left(\frac{1}{y+\epsilon 1}-1+\epsilon 2\right) = n*ln(x+\epsilon 3) - n*ln(K)$$

where, ln is the natural logarithm, n is the Hill coefficient, and K is the half-max constant. ϵ values are minor error terms of the order 10^{-5} added to avoid division by zero, and non-positive numbers inside logarithm functions.