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

Kishore Hari[1†*], Pradyumna Harlapur[1 †], Aashna Saxena[1‡], Aishwarya Girish[1‡], Herbert Levine [2,3], Mohit Kumar Jolly[1*]

[1] Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore 560012, India

[2] Center for Theoretical Biological Physics, Northeastern University, Boston, MA 02115

[3]Department of Physics, Northeastern University, Boston, MA 02115

[*] Corresponding author: mkjolly@iisc.ac.in, kishorehari@iisc.ac.in

[†] Equal contribution (first authors)

[‡] Equal contribution (second authors)

Abstract

Cell-fate decisions are driven by complex regulatory networks. Despite their complexity, these networks often exhibit low-dimensional dynamics and allow only a limited number of phenotypes. What design principles in network topology allow for these salient features remains unclear. Previously, we demonstrated that networks driving epithelial-mesenchymal transition (EMT) were comprised of two well-coordinated teams of nodes engaged in mutual antagonism; one team drove an epithelial phenotype, while the one reinforced a mesenchymal one, thus canalizing cell-fate decisions. However, it remains elusive whether teams can drive the low-dimensional dynamics of such networks. Here, we investigate networks across diverse biological scenarios (EMT, small cell lung cancer, pluripotency, gonadal cell fate) and show that they all comprise two teams of nodes that can be identified without simulations and drive two mutually exclusive phenotypes. Moreover, a stronger team strength is associated with low dimensionality in the emergent phenotypic space. Our analysis elucidates how the presence of teams may be a fundamental design principle in networks driving cellular decision-making.

Keywords: Phenotypic space, Gene Regulatory networks, Design Principles, Teams of nodes

1. Introduction

Cellular decision-making is driven largely by complex and non-linear dynamics of underlying gene regulatory networks. While the regulatory networks involve a large number of interconnected molecules, the observed phenotypic landscape is often very simple. In other words, the functional classification of the cells driven by these networks has a limited number of strata. Consequently, despite a large number of interacting components, only a limited number of phenotypes/cell types are reported [1, 2]. In differentiation events, for example, the cells move from a high-entropy progenitor phenotype to a low-entropy differentiated phenotype [3]. Such changes in entropy suggest a large-scale modification of gene expression during differentiation. Another well-studied example of cellular decision-making is epithelial-mesenchymal plasticity, a crucial process for the progression of cancer metastasis [4]. Populations capable of undergoing EMP show majorly two phenotypes: Epithelial (E) and Mesenchymal (M). The biochemical signatures of these phenotypes are mutually exclusive, and their master regulators are known to be antagonistic to one another [5, 6, 7]. A minor fraction of such populations also exhibits hybrid E/M phenotypes with a mixed biochemical signature of varying extents [8, 9]. The hybrid E/M phenotypes have complex non-uniform biochemical compositions compared to the E and M phenotypes, making them hard to identify and therapeutically target [10].

Thus, both progenitor phenotypes and the hybrid E/M phenotypes lack the cohesive expression profiles of the differentiated phenotypes and the E and M phenotypes, respectively. In other words, these phenotypes exist in a higher dimensional space than E and M states, respectively; as the E and M states can be reduced to the composite axis composed of canonical Epithelial and Mesenchymal genes; but hybrid cells occupy various places in that two-dimensional projection [11, 12]). Recently, Tripathi *et al.* showed that most of the variance in the phenotypic space emergent from the regulatory networks is explained by the first principal component of the steady state space (PC1) [13]. This observation indicates that a single axis of phenotypic switching is often what regulatory networks underlying binary cell-fate decision systems, including those of Epithelial-Mesenchymal Plasticity, operate along. In our previous work, we demonstrate a similar phenomenon in diverse networks of Epithelial-Mesenchymal Plasticity, and despite their complexity, emergent steady states are limited in number, and most of the state space converges to the Epithelial or the Mesenchymal phenotypes. We further found these networks to have well-coordinated ”teams” of nodes, such that the nodes belonging to the same team have a positive/activatory influence on each other, while those in

different teams have a negative/inhibitory influence. This structure of teams leads to the convergence of a majority of the state space to epithelial and mesenchymal phenotypes [5]. Here, we reconcile the observations related to PC1 of the steady state space and that of the teams of nodes as a design principle underlying robust phenotypic landscapes. We show that networks with strong teams show PC1 dominance, a property that is lost upon weakening the teams. We further demonstrate an increase in the dimensionality of the steady state space upon weakening the teams structure, a property conducive to the stabilization of the hybrid phenotypes in EMP. We further show that the relationship between team strength and low dimensionality holds in cell fate decisions of different biological contexts: epithelial-mesenchymal plasticity, stem cell differentiation, neuroendocrinal/non-neuroendocrinal, and gonadal fate determination. Thus, our results elucidate that low-dimensionality of phenotypic space is an emergent property of coordinated teams in biological regulatory networks.

2. Results

2.1. Teams of nodes in EMP network and the dimensionality of the steady state space

We start by analyzing a 22 node, 82 edge EMP network [14]. Each node in this network is either a transcription factor or a micro-RNA. Each edge represents the transcriptional regulation 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 (that takes a weighted sum of direct and indirect interactions between each node pair) contains a higher-order toggle switch. We refer to each "supernode" (collection of similarly behaving nodes) in this toggle switch as a "team" consisting of a set of nodes positively influencing each other and negatively influencing the nodes of the other team (Figure 1A). We simulate this network (Figure 1B) over many kinetic parameter sets using the software tool RACIPE [14]. Briefly, RACIPE first generates a system of coupled Ordinary Differential Equations (ODEs). It then randomly samples each parameter of the ODE system from a predefined range, generating a large set of parameter sets. RACIPE then simulates the ODEs for each parameter set 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 was reflected in the pairwise correlation matrix as well, where the epithelial nodes (such as miR200c, miR200b) correlated strongly positively with the epithelial nodes and negatively with the mesenchymal nodes (such as ZEB1, SNAIL) (Figure 1B).

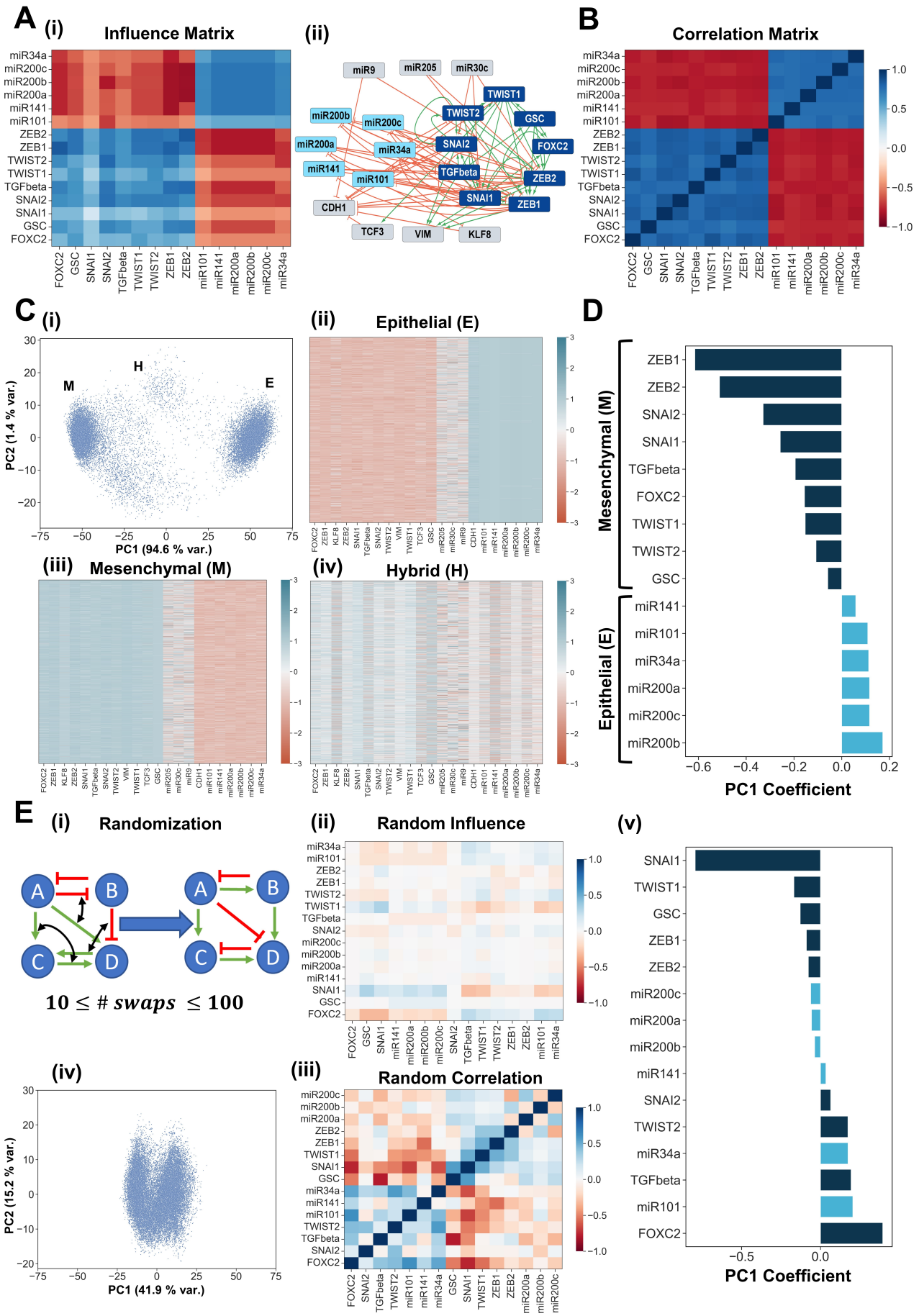

A
(i)
Influence Matrix
(ii)
miR9
miR205
miR30c
TWIST1
TWIST2
GSC
miR200b
miR200c
miR200a
miR34a
SNAI2
FOXC2
miR141
miR101
TGFbeta
ZEB2
SNAI1
ZEB1
CDH1
TCF3
VIM
KLF8
B
Correlation Matrix
C
(i)
PC2 (1.4 % var.)
PC1 (94.6 % var.)
M
H
E
(ii)
Epithelial (E)
(iii)
Mesenchymal (M)
(iv)
Hybrid (H)
D
Mesenchymal (M)
Epithelial (E)
PC1 Coefficient
E
(i)
Randomization
$10 \leq \# swaps \leq 100$
(ii)
Random Influence
(iii)
Random Correlation
(iv)
PC2 (15.2 % var.)
PC1 (41.9 % var.)
(v)
PC1 Coefficient

Figure 1: Structural similarities between teams and PC1. ***A*** *Influence matrix (i) and the network diagram (ii) for the 22N 82E EMP network.* ***B*** *Correlation matrix depicting the pairwise correlations between the node expression levels across all parameter sets in RACIPE.* ***C*** *(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).* ***D*** *The loadings of the nodes for each PC1 axis. The colors of the bars represent the team identity of the nodes.* ***E*** *(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.*

We then performed principal component analysis over the collection of all steady states obtained for this network. When the solutions are plotted along the first and second principal component axes, we found the steady states separated into 3 clusters (Figure 1C, i). Analyzing the expression patterns of the steady states forming these clusters revealed that the two 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 the hybrid

(E/M) cluster (Figure 1C ii-iv). Furthermore, we observed that the PC1 axis separated the E and M clusters, i.e., the PC1 coordinates of the epithelial steady states had opposite signs to those of the mesenchymal ones (Figure 1C, i). Given our previous understanding of the connection between teams of nodes and the emergent phenotypes, we hypothesized that the relative weights of node expressions that make up the PC1 axis could be driven by the two-team structure in the network. 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 1D, nodes belonging to different teams are colored differently). In line with our hypothesis, we see that the PC1 contribution of nodes belonging to different teams has opposite signs and that belonging to the same team has the same sign. Together these results support the hypothesis that team structure leads to PC1 explaining 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 explain 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 1E, 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 1E, ii). We

simulated each random network using RACIPE and performed PCA. Unlike the wild-type EMP network, the PC1 and PC2 axis could not separate the steady states into distinct clusters in the random networks (Figure 1E, iii, iv). We also found that the classification of the nodes based on their PC1 coefficients does not match with their teams' classification (Figure 1E, v). Therefore, the disruption of teams leads to a loss of low dimensionality in the phenotypic space.

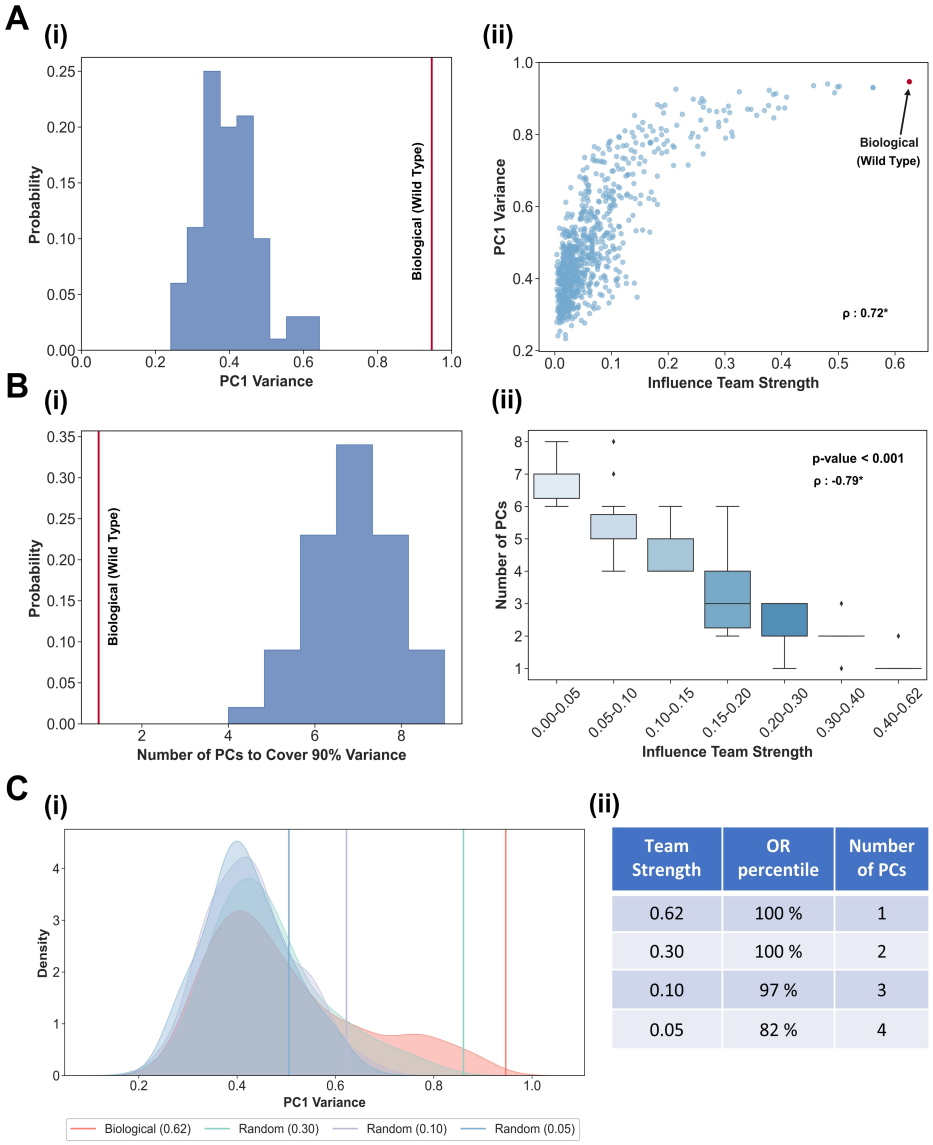
2.2. Team strength correlates with PC1 variance and dimensionality of random networks

Next, we compared the percentage variance explained along the PC1 axis for WT networks against that of random networks. We found the measure much lower in random networks than in the WT networks, where PC1 alone explained nearly 90% of the variance (Figure 2A, i). A scatter plot between the team strength of the random and WT networks against the variance explained along the PC1 axis (Figure 2A, ii) revealed a positive correlation between the two measures, suggesting that an increase in team strength can lead to a corresponding increase in the variance explained by PC1. An important observation here is the sigmoidal nature of the data. After a team strength of about 0.25-3, the variance explained by PC1 starts to saturate.

We then wanted to estimate the dimensionality of the steady-state space emergent from the random networks. More than 90% of the variance in the steady-state space of the EMP network is explained by the first principal component axis, making it majorly one-dimensional [13]. 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 2B, i), suggesting that while the steady-state space of WT network is low-dimensional, that of the random networks is high dimensional. The number of PCs also showed a strong negative correlation with team strength (Figure 2B, ii), suggesting that strong teams reduce the dimensionality of the steady-state space.

We then performed a causative analysis of the relationship between team strength and PC1 dominance. We selected three random networks, each with different team strengths. For each such random network that we refer to as the Original Random (OR) network, we generate "secondary" random (SR) networks, i.e., randomly swapping the edges in the OR network. We then simulated the OR and the corresponding SR networks and calculated the variance explained by PC1. Given the positive correlation between team strength and the variance explained by PC1 (Figure 2A, ii), we expect that the dominance of PC1 will decrease with decreasing team strength of the OR. As expected,

we find that the percentile of the OR in the distribution of PC1 variance of the corresponding SRs decreases with decreasing team strength (Figure 2C).



Team Strength	OR percentile	Number of PCs
0.62	100 %	1
0.30	100 %	2
0.10	97 %	3
0.05	82 %	4

Figure 2: Strong teams lead to a low-dimensional steady-state space. A *(i) Histogram depicting the variance explained by PC1 for random networks. EMP network is represented with a vertical red line. (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. Spearman correlation coefficient 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. EMP network is represented with a vertical red line. (ii) Boxplot depicting the dependence of the number of PCs required to explain 90% of the variance in steady states on the team strength. P-value for one-way ANOVA and the spearman correlation coefficient have been reported, with * indicating a p − value < 0.05.* ***C*** *(i) Density plots showing the PC1 variance for secondary random networks (SR). The original random networks (OR) are shown by the vertical lines. The group strengths of OR are given in the legend. (ii) Table depicting the percentile of OR in the PC1 variance distribution of SRs and the number of PCs required to explain 90% variance in the OR.*

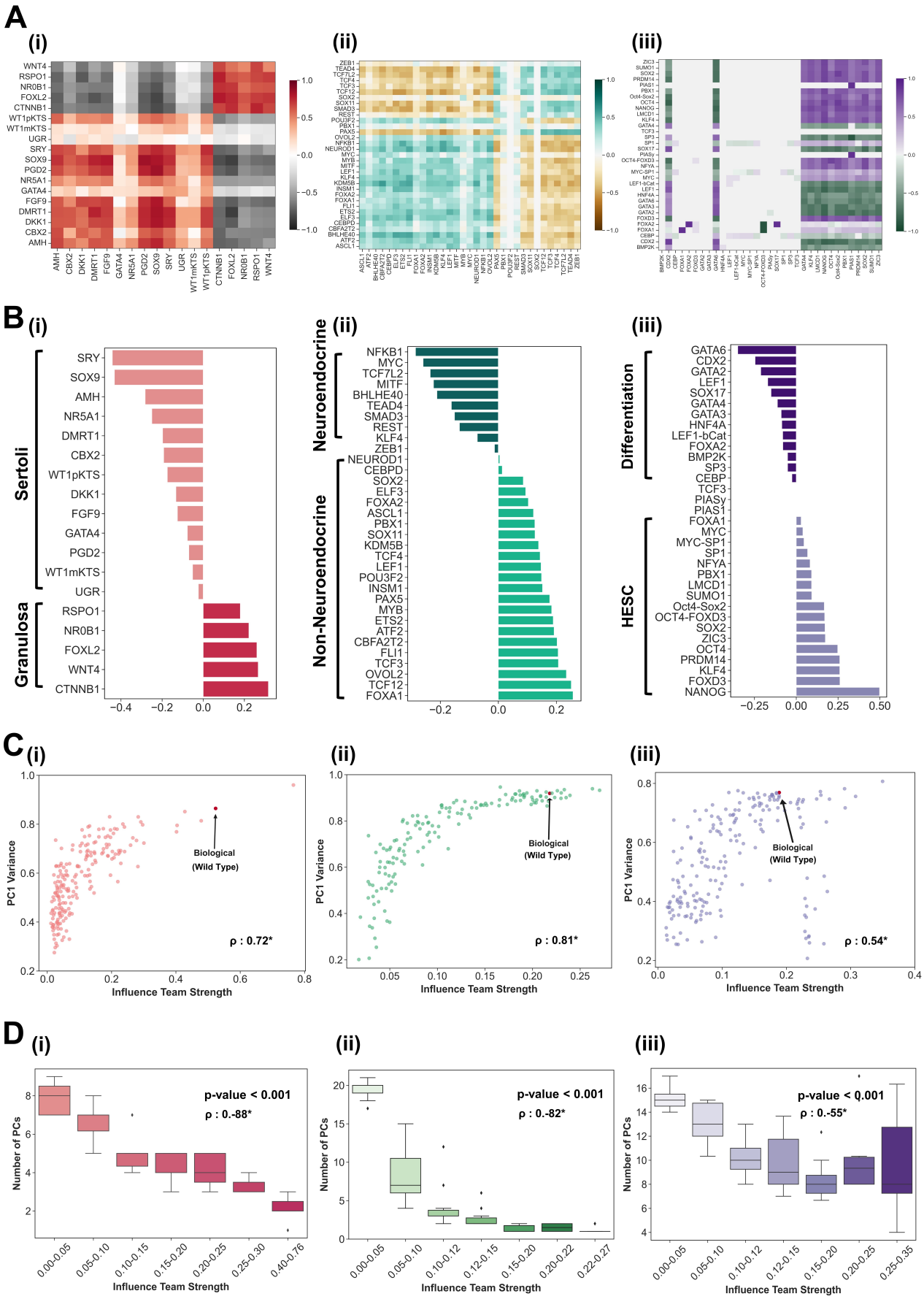

A
(i)
(ii)
(iii)
B
(i)
Sertoli
Granulosa
SRY
SOX9
AMH
NR5A1
DMRT1
CBX2
WT1pKTS
DKK1
FGF9
GATA4
PGD2
WT1mKTS
UGR
RSPO1
NR0B1
FOXL2
WNT4
CTNNB1
(ii)
Neuroendocrine
Non-Neuroendocrine
NFKB1
MYC
TCF7L2
MITF
BHLHE40
TEAD4
SMAD3
REST
KLF4
ZEB1
NEUROD1
CEBPD
SOX2
ELF3
FOXA2
ASCL1
PBX1
SOX11
KDM5B
TCF4
LEF1
POU3F2
INSM1
PAX5
MYB
ETS2
ATF2
CBFA2T2
FLI1
TCF3
OVOL2
TCF12
FOXA1
(iii)
Differentiation
HESC
GATA6
CDX2
GATA2
LEF1
SOX17
GATA4
GATA3
HNF4A
LEF1-bCat
FOXA2
BMP2K
SP3
CEBP
TCF3
PIASy
PIAS1
FOXA1
MYC
MYC-SP1
SP1
NFYA
PBX1
LMCD1
SUMO1
Oct4-Sox2
OCT4-FOXD3
SOX2
ZIC3
OCT4
PRDM14
KLF4
FOXD3
NANOG
C
(i)
(ii)
(iii)
PC1 Variance
Influence Team Strength
Biological
(Wild Type)
ρ : 0.72*
ρ : 0.81*
ρ : 0.54*
D
(i)
(ii)
(iii)
Number of PCs
Influence Team Strength
p-value < 0.001
ρ : 0.-88*
ρ : 0.-82*
ρ : 0.-55*
0.00-0.05
0.05-0.10
0.10-0.15
0.15-0.20
0.20-0.25
0.25-0.30
0.40-0.76
0.10-0.12
0.12-0.15
0.15-0.20
0.20-0.22
0.22-0.27
0.20-0.25
0.25-0.35

Figure 3: 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*** *The loadings of the nodes for each PC1 axis for (i) Gonadal, (ii) SCLC, and (iii) iPSC network. The colors of the bars represent the team identity of the nodes.* ***C*** *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. Spearman correlation coefficient is reported, with * indicating a p − value < 0.05.* ***D*** *Boxplot depicting the dependence of the number of PCs required to explain 90% of the variance in steady states on the team strength for (i) Gonadal, (ii) SCLC, and (iii) iPSC network. P-value for one-way ANOVA and the spearman correlation coefficient have been reported, with * indicating a p − value < 0.05.*

2.3. Strong teams underlie the low-dimensionality of phenotypic space in multiple contexts of cell-fate decisions

Our observations establish that strong teams underly the PC1 dominance in the 22N 82E EMP network. We then asked if these above-mentioned results hold true in other biological contexts. We also looked into three more regulatory networks underlying cell-fate decisions: SCLC, iPSC, and Gonadal fate determination (Figure S1A) [15, 16, 17]. 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 iPSC network describes the differentiation of hESCs. The network was constructed to identify pathways in which iPSCs can be achieved. 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 male and female fates, respectively. These three networks display a spectrum of behavior. The iPSC network has hub genes (NANOG, SOX2, OCT4) with high out-degree (i.e., these nodes regulate multiple genes). SCLC and iPSC 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 2C. Gonadal and SCLC networks show a clear team structure in their influence matrices (Figure 3A, i, ii). Such team structure is relatively harder to discern in the iPSC influence matrix (Figure 3A, iii). However, our clustering-based method of identifying teams allowed us to classify the nodes into two teams. This classification agreed with both the biochemical orientation of these molecules as well as the groups seen in the correlation matrix (Figure S1B). For all three networks, the nodes belonging to different teams

showed opposite signs in PC1 loading (Figure 3B). This is particularly noteworthy for iPSC network, given its weak team strength.

For all three networks, the PC1 was able to separate the steady states into clusters. The shape that the steady state space takes in each network was unique, mainly in terms of the "hybrid" phenotypes. The gonadal network showed multiple clusters, albeit on the same axis as that of the most well-separated ones (Figure S2A). SCLC network, on the other hand, showed a hybrid cluster similar to the EMP network, but the clusters were more continuous than that of the EMP network (Figure S2B). iPSC network does not show any hybrid clusters at all, despite having the weakest teams of all three networks (Figure S2C). These results show the variability of the ability of these networks in being able to give rise to high-dimensional states that do not seem to follow the team structure. Furthermore, we found that the sigmoidal relationship between the team strength and PC1 variance seen for the EMP network (Figure 2A, ii) was maintained for these three networks (Figure 3C) as well. Lastly, we also confirmed that the inverse trend seen between team strength and the number of PCs needed to explain 90% variance in steady-state space was also seen for these three networks. However, the monotonicity of the relationship between team strength and the number of PCs was broken for the iPSC network due to it having the weakest team composition of the three. Together, these results suggest that strong teams universally regulate the dimensionality of the emergent phenotypic space of cell-fate decision-making networks.

3. Discussion

Cell fate decisions are ubiquitous in biological systems. Cells dynamically change phenotypes in response to varying biochemical and biomechanical signals. A well-known example is the sequence of phenotypic transitions involved in the development of an organism that starts from a pluripotent stem cell and gives rise to cells with specific functions, often through binary branching points [18, 19]. Such decisions are also shown to be extremely crucial in the pathology of cancer, where cells dynamically acquire and lose different phenotypes in order to overcome hurdles such as anoikis, inflammation, and hypoxia [20]. In developmental contexts, the property of canalization is frequently observed, where cellular phenotypes are sensitive only to a small set of specific perturbations [21]. Despite the volume and complexity of the mechanisms underlying such phenotypic changes, the

number of paths through which this can happen is limited. An implication of such canalization is that the number of phenotypes that the cells can achieve is also limited by the number of perturbations that the cells are sensitive to. Regulatory networks from various biological contexts have been shown to exhibit canalization [22]. Recent studies have shown how complex regulatory networks underlying such canalizing systems give rise to relatively simple, low-dimensional dynamics [13]. Our results suggest that this low-dimensionality can be emergent from the latent structure in these regulatory networks, called teams of nodes.

We had previously identified teams of nodes in regulatory networks driving the phenotypic decisions in Epithelial-Mesenchymal Plasticity [5], SCLC [23], and melanoma [24]. We found that teams of nodes can give rise to a variety of properties in the emergent phenotypes, including minimal frustration for a majority of them. Recently, Tripathi *et al.* showed that biological networks are minimally frustrated and can give rise to low-dimensional phenotypic space. Our results, therefore, provide a mechanistic understanding of these trends as we find that only in networks with high team strength most of the variance in the steady-state space (> 90%) is explained by the first principal component
axis.

Our current analysis is limited to one of the axes of plasticity: epithelial-mesenchymal, stemcell differentiation, gonadal cell fate determination, and non-neuroendocrine/neuroendocrine cell fate determination. Cancer cells have been shown to exhibit multiple axes of phenotypic transitions: epithelial-mesenchymal plasticity, drug resistance, stemness, and metabolic plasticity, to name a few [25, 26, 27, 28]. Studies suggest that these transitions can happen together (e.g., EMT leading to the acquisition of stemness and/or drug resistance), suggesting interconnectivity between the underlying regulatory networks [29, 30, 31, 32]. An important direction for the current study is to understand how the low-dimensional dynamics scale across these interdependent axes of phenotypic heterogeneity and how the team structure evolves under such interactions.

4. Methods

4.1. RAndom CIrcuit PErturbaiton (RACIPE)

RACIPE [14] is a tool that simulates transcriptional regulatory networks (TRNs) in a continuous manner. Given a TRN, 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. $P_i{}^0{}_T$ is the threshold value of P_i concentration at which the nonlinearity in the dynamics of T due to P_i is seen. n is termed as Hill-coefficient and represents the extent of nonlinearity in the regulation. λ represents the fold change in the target node concentration upon over-expression of 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:

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

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 predefined set of parameter ranges estimated from BioNumbers [33]. The ranges, as reported by Huang *et al.* [14] are as follows:

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 inward regulations - half functional rule	

At each parameter set, RACIPE integrates the model from multiple initial conditions 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 10000 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 method of numerical integration.

4.2. Processing RACIPE output

For a given network with i = [1,n] nodes, the steady-state expression levels of the nodes were normalized in the following way:

$$E_{in} = \frac{E_i}{f_i}$$

$$f_i = \frac{g_i}{k_i} \prod_j \lambda_{ij}$$

Where, for the ith node, E_{in} is the normalized expression level, E_i is the steady state expression level, f_i is the normalization factor, g_i, and k_i are the production and degradation of the ith node corresponding to the current steady state, and λ_{ij} are the fold change in expression of i due to node j = [1,n]. The collection of these normalized steady states was used for PCA.

4.3. 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.

4.4. 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:

$$Infl_{l_{max}} = \frac{\sum_{l=1}^{l_{max}} \frac{Adj^l}{Adj^l_{l_{max}}}}{l_{max}}$$

Where Adj_{max} is derived by setting all non-zero entries of the adjacency matrix to one, 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

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

4.5. Statistical analysis

Spearman correlation coefficients were calculated using the *cor.test* function from R 4.1. One-way ANOVA test was performed using the function *f oneway* from the *scipy* package in Python 3.8.

5. Author Contributions

MKJ and KH designed the research; KH and PH analyzed the data; AS, AG, PH, and KH carried out the simulations; all authors discussed the results and participated in the preparation of the manuscript.

6. Conflict of Interest

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

7. Data Availability

The raw data generated for this study and derived datasets supporting the current findings are available from the corresponding author (MKJ, KH) upon reasonable request.

8. Funding Information

MKJ was supported by Ramanujan Fellowship (SB/S2/RJN-049/2018) by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India. KH is supported by the Prime Minister's Research Fellowship (PMRF).

References

[1] Vickaryous MK, Hall BK (2006) Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biological Reviews* 81(03):425.

[2] Jones RC, et al. (2022) The tabula sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376(6594):eabl4896.

[3] Gulati GS, et al. (2020) Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367(6476):405–411.

[4] Jolly MK, Ware KE, Gilja S, Somarelli JA, Levine H (2017) Emt and met: necessary or permissive for metastasis? *Molecular Oncology* 11(7):755–769.

[5] Hari K, Ullanat V, Balasubramanian A, Gopalan A, Jolly MK (2022) Landscape of epithelial–mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. *eLife* 11:e76535.

[6] Hong T, et al. (2015) An ovol2-zeb1 mutual inhibitory circuit governs bidirectional and multistep transition between epithelial and mesenchymal states. *PLOS Computational Biology* 11(11):e1004569.

[7] Roca H, et al. (2013) Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PLoS ONE* 8(10):e76773.

[8] Pastushenko I, et al. (2018) Identification of the tumour transition states occurring during EMT. *Nature* 556(7702):463–468.

[9] Jolly MK, et al. (2019) Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. *Pharmacology & Therapeutics* 194:161–184.

[10] Subbalakshmi AR, Ashraf B, Jolly MK (2022) Biophysical and biochemical attributes of hybrid epithelial/mesenchymal phenotypes. *Physical Biology* 19(2):025001.

[11] Font-Clos F, Zapperi S, Porta CAML (2018) Topography of epithelial–mesenchymal plasticity. *Proceedings of the National Academy of Sciences* 115(23):5902–5907.

[12] Steinway SN, et al. (2015) Combinatorial interventions inhibit TGFβ-driven epithelial-tomesenchymal transition and support hybrid cellular phenotypes. *npj Systems Biology and Applications* 1:15014.

[13] Tripathi S, Kessler DA, Levine H (2023) Minimal frustration underlies the usefulness of incomplete regulatory network models in biology. *Proceedings of the National Academy of Sciences* 120(1):e2216109120.

[14] Huang B, et al. (2017) Interrogating the topological robustness of gene regulatory circuits by randomization. *PLOS Computational Biology* 13(3):e1005456.

[15] Chang R, Shoemaker R, Wang W (2011) Systematic search for recipes to generate induced pluripotent stem cells. *PLoS Comput Biol* 7(12):e1002300.

[16] Udyavar AR, et al. (2017) Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity. *Cancer Res* 77(5):1063–1074.

[17] R´ıos O, et al. (2015) A boolean network model of human gonadal sex determination. *Theor Biol Med Model* 12(1):26.

[18] Zhou JX, Huang S (2011) Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends in Genetics* 27(2):55–62.

[19] Graf T, Enver T (2009) Forcing cells to change lineages. *Nature* 462(7273):587–594.

[20] Agozzino L, Bal´azsi G, Wang J, Dill KA (2020) How do cells adapt? stories told in landscapes. *Annual Review of Chemical and Biomolecular Engineering* 11(1):155–182.

[21] Du Z, et al. (2015) The regulatory landscape of lineage differentiation in a metazoan embryo. *Developmental Cell* 34(5):592–607.

[22] Kadelka C, Butrie TM, Hilton E, Kinseth J, Serdarevic H (2020) A meta-analysis of boolean network models reveals design principles of gene regulatory networks. arXiv:2009.01216.

[23] Chauhan L, Ram U, Hari K, Jolly MK (2021) Topological signatures in regulatory network enable phenotypic heterogeneity in small cell lung cancer. *eLife* 10:e64522.

[24] Pillai M, Jolly MK (2021) Systems-level network modeling deciphers the master regulators of phenotypic plasticity and heterogeneity in melanoma. *iScience* 24(10):103111.

[25] Thankamony AP, Saxena K, Murali R, Jolly MK, Nair R (2020) Cancer stem cell plasticity – a deadly deal. *Frontiers in Molecular Biosciences* 7:79.

[26] Jia D, et al. (2019) Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. *Proceedings of the National Academy of Sciences* 116(9):3909–3918.

[27] Vipparthi K, et al. (2022) Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer. *iScience* 25(5):104317.

[28] Bhatia, et al. (2019) Interrogation of phenotypic plasticity between epithelial and mesenchymal states in breast cancer. *Journal of Clinical Medicine* 8(6):893.

[29] Sahoo S, et al. (2021) A mechanistic model captures the emergence and implications of nongenetic heterogeneity and reversible drug resistance in ER breast cancer cells. *NAR Cancer* 3(3):zcab027.

[30] Galbraith M, Levine H, Onuchic JN, Jia D (2023) Decoding the coupled decision-making of the epithelial-mesenchymal transition and metabolic reprogramming in cancer. *iScience* 26(1):105719.

[31] Pasani S, Sahoo S, Jolly MK (2020) Hybrid e/m phenotype(s) and stemness: A mechanistic connection embedded in network topology. *Journal of Clinical Medicine* 10(1):60.

[32] Biddle A, Gammon L, Liang X, Costea DE, Mackenzie IC (2016) Phenotypic plasticity determines cancer stem cell therapeutic resistance in oral squamous cell carcinoma. *EBioMedicine* 4:138–145.

[33] Milo R, Jorgensen P, Moran U, Weber G, Springer M (2009) BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research* 38(suppl 1):D750–D753.

Supplementary figures

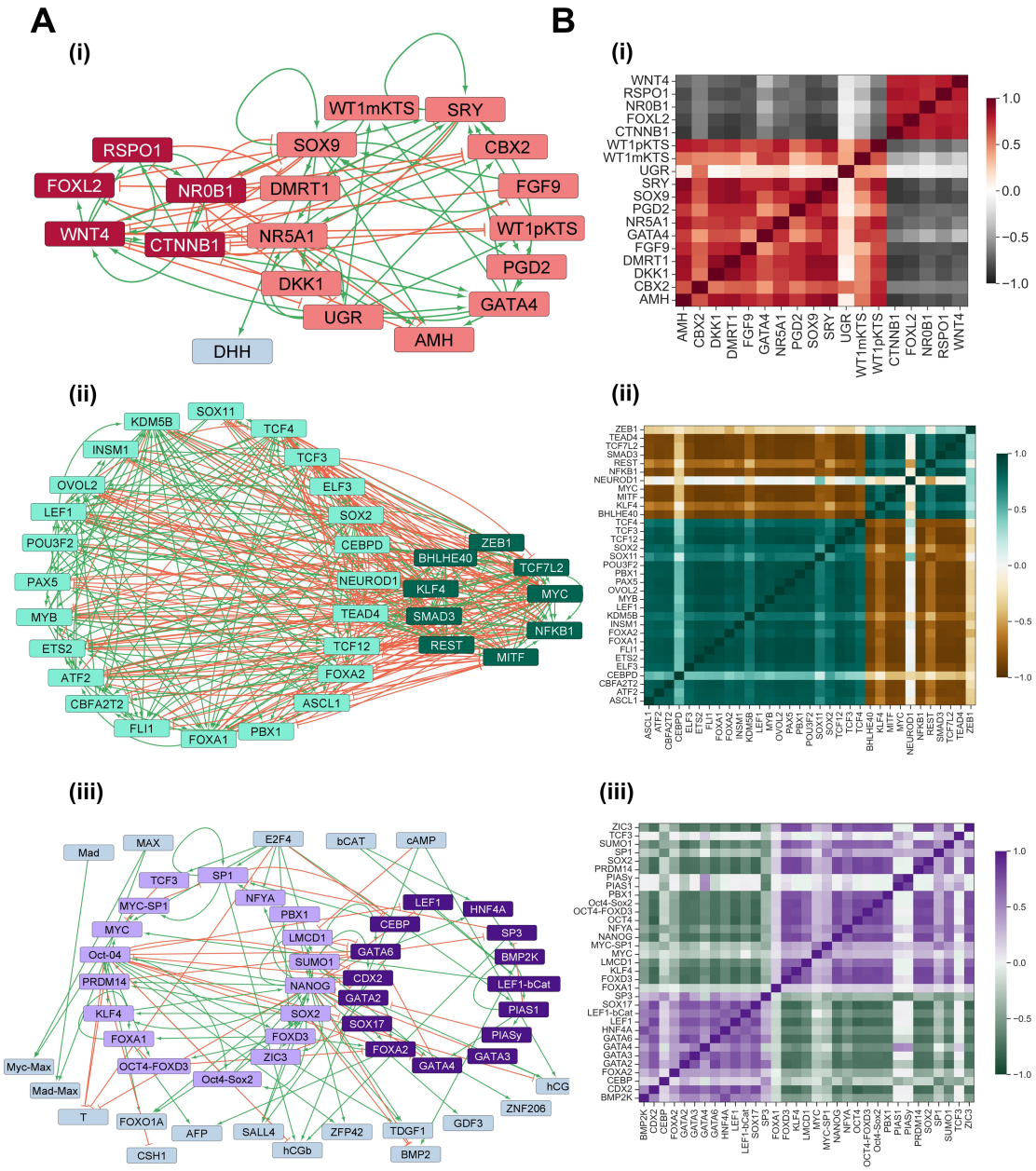


Figure S1: Structural similarities between teams and PC1. ***A*** *Network diagrams for (i) Gonadal, (ii) SCLC, and (iii) iPSC network.* ***B*** *Correlation matrix depicting the pairwise correlations between the node expression levels across all parameter sets in RACIPE for (i) Gonadal, (ii) SCLC, and (iii) iPSC network.*

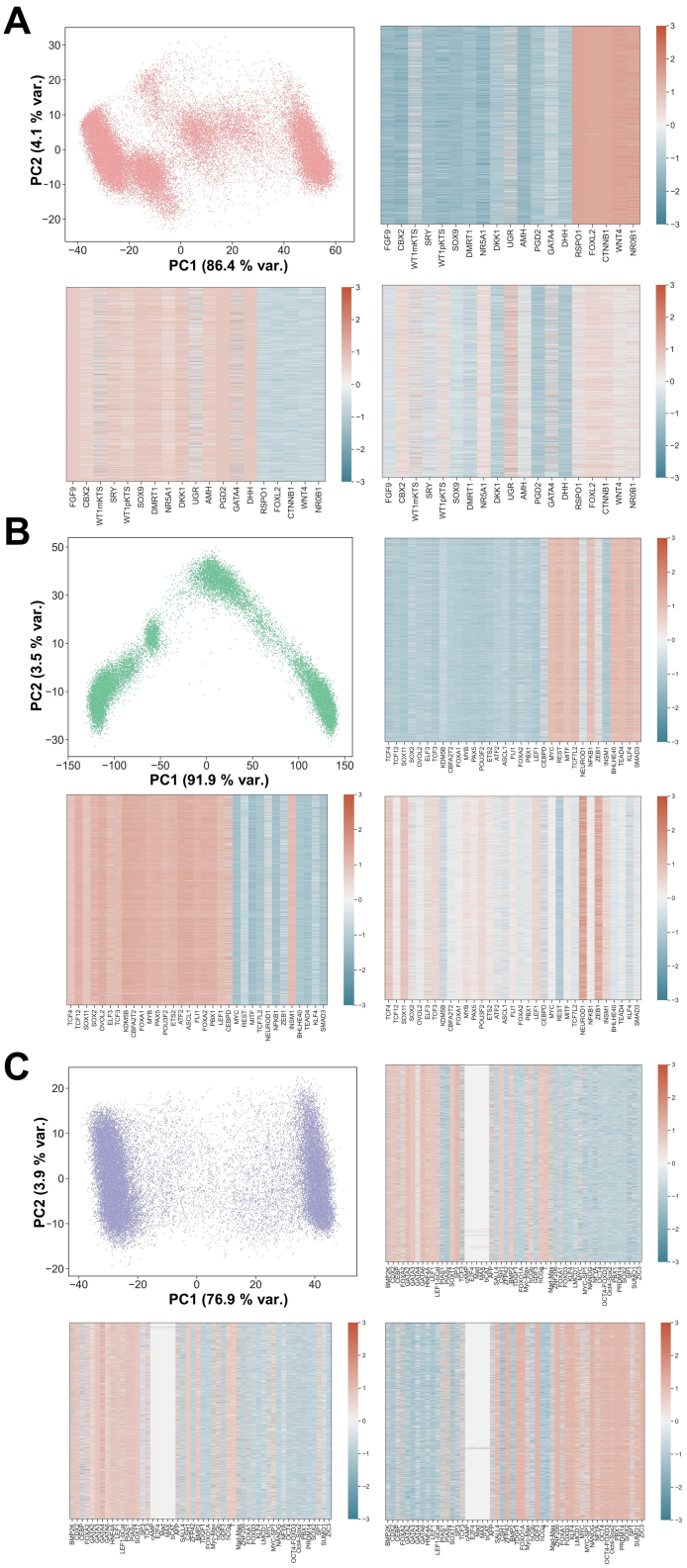


Figure S2: Structural similarities between teams and PC1. ***A*** *Scatterplot mapping the solutions generated from RACIPE on the axes of PC1 and PC2. Heatmaps depicting the expression levels of the nodes of the network for each individual cluster Gonadal fate decision network.* ***B*** *Same as* ***A*** *but for SCLC network.* ***C*** *Same as* ***A*** *but for iPSC network.*

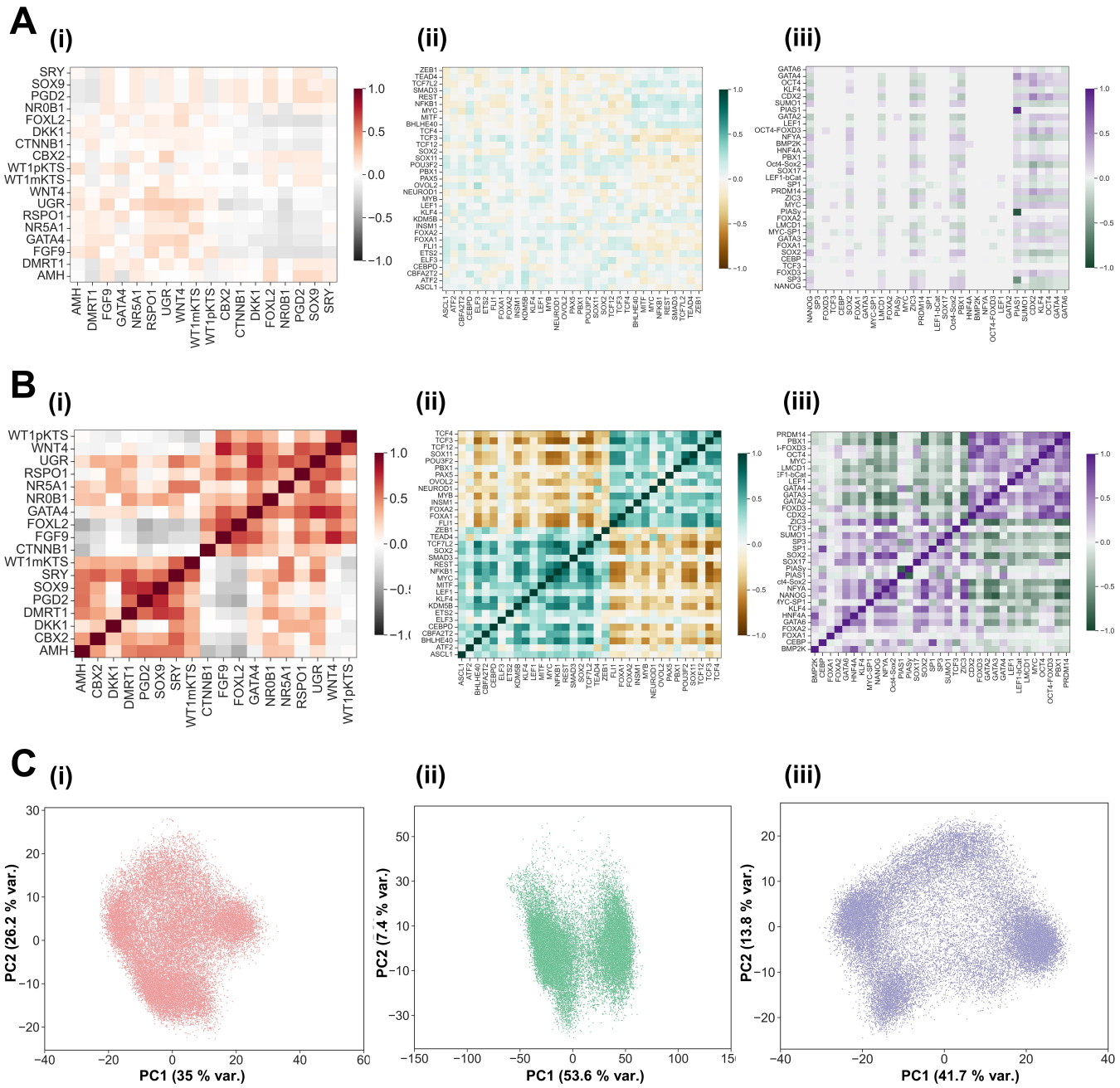


Figure S3: Structural similarities between teams and PC1. ***A*** *Influence matrices of representative random network generated from (i) Gonadal, (ii) SCLC and (iii) iPSC network.* ***B*** *Correlation matrix depicting the pairwise correlations between the node expression levels across all parameter sets in RACIPE for a representative random network generated from (i) Gonadal, (ii) SCLC and (iii) iPSC network.* ***C*** *Scatterplot mapping the solutions generated from RACIPE on the axes of PC1 and PC2 for a representative random network generated from (i) Gonadal, (ii) SCLC and (iii) iPSC network.*