Biological Networks Regulating Cell Fate Choice Are Minimally Frustrated

Shubham Tripathi,[1,2,3] David A. Kessler,[4,*] and Herbert Levine[2,3,†]
[1]PhD Program in Systems, Synthetic, and Physical Biology, Rice University, Houston, Texas 77005, USA
[2]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, USA
[3]Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA
[4]Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

(Received 22 November 2019; accepted 14 July 2020; published 17 August 2020)

Characterization of the differences between biological and random networks can reveal the design principles that enable the robust realization of crucial biological functions including the establishment of different cell types. Previous studies, focusing on identifying topological features that are present in biological networks but not in random networks, have, however, provided few functional insights. We use a Boolean modeling framework and ideas from the spin glass literature to identify functional differences between five real biological networks and random networks with similar topological features. We show that minimal frustration is a fundamental property that allows biological networks to robustly establish cell types and regulate cell fate choice, and that this property can emerge in complex networks via Darwinian evolution. The study also provides clues regarding how the regulation of cell fate choice can go awry in a disease like cancer and lead to the emergence of aberrant cell types.

DOI: 10.1103/PhysRevLett.125.088101

Biological regulatory networks establish cell-type-specific gene expression patterns [1] and regulate cell fate choice in response to various signals. These networks present a contradiction analogous to the famed Levinthal paradox in protein folding [2]. Networks as large and complex as those regulating cell fate typically exhibit a huge number of stable states [3]. Each such stable state, or collection of stable states with a reasonably shared pattern of gene expression, represents a cell type [4,5]. This relationship, however, predicts a number of cell types much larger than that seen in multicellular organisms. A smaller number of cell types can be attained via the evolutionary fine-tuning of network parameters [6] or by putting cells through a precise sequence of events during development [7]. In both scenarios, cell fate will be highly sensitive to intra- and extracellular perturbations, an undesirable property.

Features that distinguish biological regulatory networks from random networks may provide a clue regarding how these networks can robustly establish the smaller than expected number of cell types. Biological networks have been shown to often exhibit a scale-free degree distribution [8] which might allow these networks to define topologically stable cell types [9]. Regulatory networks in *Escherichia coli* and *Saccharomyces cerevisiae* have been shown to be hierarchically organized [10]. Certain network patterns, called motifs, are known to recur far more frequently in biological networks than in random networks [11], and often mediate cell fate choice [12]. However, these investigations of topological differences between biological and random networks have provided few insights into how the functional characteristics of biological networks differ from those of random networks. In this Letter, we compare the dynamical behavior of biological regulatory networks with that of random networks which have similar topological features and observe some remarkable differences. These could hold the key to elucidating the design principles that allow biological regulatory networks to carry out biological functions.

Boolean modeling of biological networks.—A Boolean modeling framework [13] has proven useful for characterizing the behavior of large networks in cases where the use of ordinary differential equations-based modeling frameworks becomes challenging due to the numerous and hard to estimate kinetic parameters involved. In this framework, the only knowledge required is whether each regulatory relationship between network nodes is activating or inhibitory and how the inputs to a given node combine to regulate node activity. The state of an N-node network in such a framework may be specified via a sequence $\{s_i\}$ of N binary variables: $s_i = \pm 1$. When modeling a biological regulatory network, each network node represents a molecular species such as a transcription factor or micro-RNA. When species i (the molecular species represented by node i) is highly expressed, $s_i = +1$, otherwise $s_i = -1$. If the inputs to network nodes combine additively to activate or inhibit node activity, regulatory relationships between molecular species can be specified by an $N \times N$ matrix J where $J_{ij} = +1$ if species j promotes the expression of species i and $J_{ij} = -1$ if species j inhibits the expression of species i. The absence of any regulatory relationship from species i to species j is indicated by

$J_{ij} = 0$. The discrete-time network dynamics can then be simulated using [14]

$$s_i(t+1) = \begin{cases} +1 & \\ -1 & \text{if} \\ s_i(t) & \end{cases} \begin{array}{l} \sum_j J_{ij} s_j > 0 \\ \sum_j J_{ij} s_j < 0 \\ \sum_j J_{ij} s_j = 0. \end{array} \quad (1)$$

The network state is updated asynchronously, i.e., at each discrete time point, a network node is chosen at random, and its state updated using Eq. (1). Clearly, a state $\{s_i\}$ is a stable state of the network if s_i is a fixed point of Eq. (1) for all i.

Note that the dynamical behavior of a network in the above modeling framework is equivalent to the zero-temperature dynamics of an asymmetric spin glass on a graph. Using this equivalence, we characterize an edge $j \to i$ in state $\{s_i\}$ as frustrated [15] if $J_{ij} s_i s_j < 0$. Then, the frustration of a state can be defined as the fraction of network edges that are frustrated in that state. If a network involves regulatory relationships that conflict with one another, all regulatory relationships cannot be satisfied in any state. Hence, such a network will have stable states with nonzero frustration. Further, if $\{s_i\}$ is a stable state of the network, $\{s_i'\}$ with $s_i' = -s_i$ will also be a stable state of the network. The state $\{s_i'\}$ will have the same frustration as the state $\{s_i\}$.

Comparison of biological and random networks.—We determined the stable states of five biological networks taken from the literature [14,16–19] [see Supplemental Material (SM) [20], Sec. I A] and compared the frustration of the stable states of each of these networks with the frustration of the stable states of random networks with similar topological features (each random network had the same total number of nodes and edges, node in- and out-degree distributions, and the total number of activating and inhibitory relationships between nodes in the network as the corresponding biological network) [Figs. 1(a)–1(e); also see SM [20], Figs. S1–S4 and Secs. I B and I D]. In the case of each biological network, most stable states had frustration comparable to the frustration of the stable states of the corresponding random networks. However, each biological network had a set of stable states with frustration much lower than the frustration of the stable states of random networks. Crucially, biological networks are highly likely to end up in one of these minimally frustrated stable states when their dynamics is simulated starting from random initial conditions (Fig. 1, pink violin in each panel; also see SM [20], Fig. S5). Minimally frustrated stable states are thus likely to be biologically significant, with most cells in a population exhibiting gene expression patterns corresponding to these states.

Relation between stable states and biological phenotypic states.—We next investigated if any structural patterns underlie the organization of stable states of biological networks. The distribution $P(q_{\alpha\beta})$ of the overlap between network states $q_{\alpha\beta} = \sum_i (s_i^\alpha s_i^\beta)/N$ was found to be very

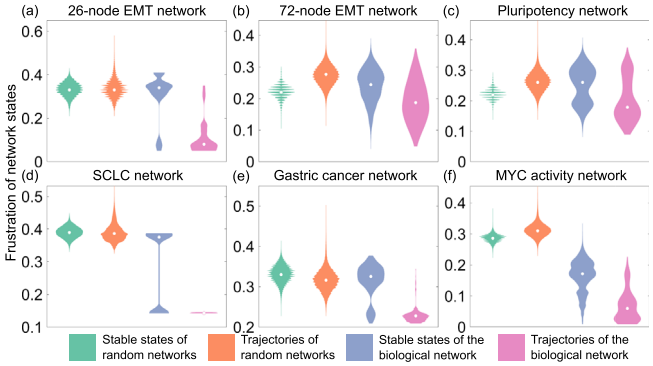

FIG. 1. Distribution of frustration of the stable states of biological networks and of random networks. 500 random networks corresponding to each biological network were considered. The orange and pink violins in each panel show the distribution of the frustration of the states one ends up in when simulating the dynamics of the network starting from 500 random initial conditions (each state is thus counted as many times as it is encountered while each stable state is counted only once in the other two violins). Biological networks in (a)–(e) were taken from the literature. The MYC activity network in (f) was inferred from gene expression data (see SM [20], Sec. I G). In (b), (c), and (f), the green violin shows the distribution when only the 5000 least frustrated stable states of each random network are considered. The white circle in each violin indicates the median. SCLC stands for small cell lung cancer.

broad when α, β pairs were chosen randomly from the set of stable states of biological networks. However, when the pairs were sampled from among the minimally frustrated stable states, $P(q_{\alpha\beta})$ was bimodal with peaks near $+1$ and -1 [Figs. 2(a) and 2(d); also see SM [20], Figs. S10 (left-hand column) and S11]. Since $q_{\alpha\beta}$ is a measure of similarity between the states $\{s_i^\alpha\}$ and $\{s_i^\beta\}$, a collection of states with $q_{\alpha\beta}$ close to $+1$ for all pairs represents a collection of cells with reasonably similar gene expression profiles, to be associated with a distinct cell phenotype. A bimodal $P(q_{\alpha\beta})$ for the minimally frustrated stable states of biological networks considered here thus suggests that these states constitute two stable phenotypic states.

In the case of the network regulating epithelial-mesenchymal transition (EMT) [14], minimally frustrated stable states represent the two canonical phenotypic states, epithelial and mesenchymal [Figs. 2(b) and 2(c); also see SM [20], Fig. S10]. In the case of the network regulating pluripotency and differentiation in human embryonic stem cells [18], minimally frustrated stable states define stem and differentiated cell types [Figs. 2(e) and 2(f)]. In contrast, high frustration stable states of biological networks involve copresentation of molecular markers corresponding to conflicting biological behaviors—high frustration stable states of the network regulating EMT involve copresentation of epithelial and mesenchymal markers [Figs. 2(b) and 2(c); also see SM [20], Fig. S10] while high frustration stable states of the

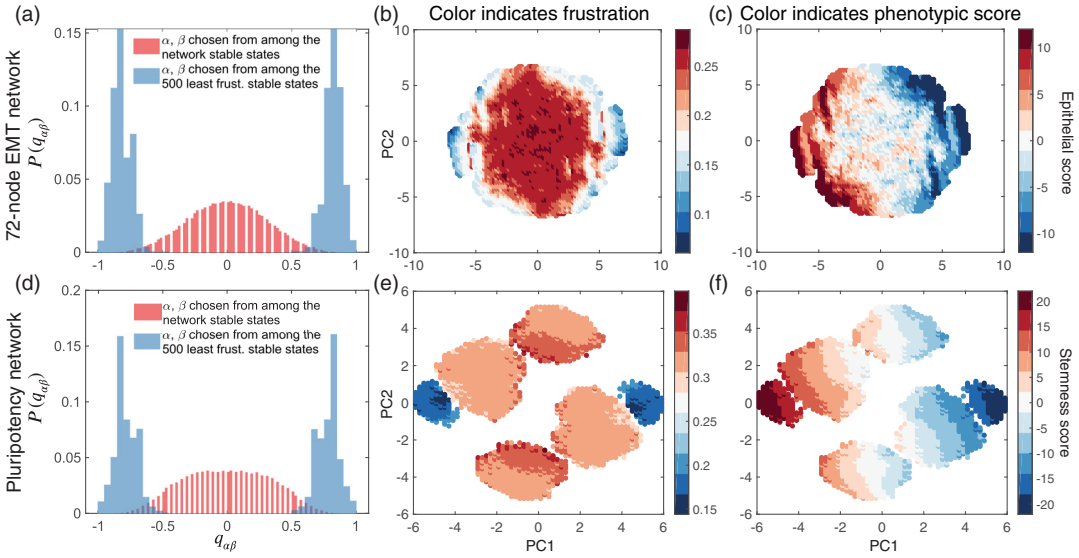

FIG. 2. Minimally frustrated stable states of biological networks define canonical cell types. (a),(d) $P(q_{\alpha\beta})$ is bimodal for the minimally frustrated stable states of both the 72-node EMT network and the pluripotency network. (b),(c),(e),(f) Principal component (PC) representation of the 200 000 observed stable states of the 72-node EMT network and of the pluripotency network. In each case, we included 100 000 least frustrated and 100 000 most frustrated stable states. In (c), a high, positive score indicates an epithelial phenotype while a low, negative score indicates a mesenchymal phenotype. In (f), a high, positive score indicates a stem cell phenotype while a low, negative score indicates a differentiated phenotype. Clusters with extreme values of the phenotypic scores represent canonical cell types. See SM [20], Sec. I E, for details of how the scores were defined.

network regulating the neuroendocrine-mesenchymal transition [16] involve copresentation of neuroendocrine and mesenchymal markers (see SM [20], Fig. S10). These stable states thus represent ambiguous cell fate choices. Such ambiguous phenotypic states have been reported in cancer cells across disease subtypes [16,34], but appear to be suppressed in healthy tissue.

Effect of noise in network dynamics.—Thus far, we have neglected stochasticity in our analyses. Noise in gene expression can have significant implications for cellular function [35]. We defined a pseudo-Hamiltonian $H = -\sum_{i,j} J_{ij} s_i s_j$ and used the finite-temperature Metropolis Monte Carlo algorithm [36,37] to probe network behavior under noisy dynamics (see SM [20], Sec. I H). As node dynamics become increasingly noisy, biological networks become more and more likely to exhibit states with high frustration [Figs. 3(a) and 3(e); also see SM [20], Figs. S13 (top row) and S14]. Functionally, this manifests as more and more cells in a population presenting with ambiguous cell fate choices [Figs. 3(c) and 3(g)]. Since cells in healthy tissue rarely exhibit ambiguous cell fate choice, we hypothesize that healthy cells must have low noise levels at which the regulatory network exhibits one of the low frustration network states.

Effect of network mutations.—Another scenario in which cells presenting ambiguous cell fate choices are frequently observed in our modeling framework is if the biological network becomes mutated [Figs. 3(b) and 3(f); also see SM [20], Figs. S13 (bottom row) and S15]. We observe here that the studied biological networks are relatively robust, and it is only after a significant number of mutations have accumulated that a significant fraction of cells in the population start exhibiting noncanonical phenotypic states.

Emergence of biological characteristics in random networks.—Under selection for networks with low frustration states, a population of randomized 26-node EMT networks can evolve to exhibit the behavior reported herein for the corresponding biological network (Fig. 4; see SM [20], Sec. I I, for details of the simulation). This includes the existence of minimally frustrated states that are frequently encountered when starting from random initial conditions [Fig. 4 (top)] and a bimodal $P(q_{\alpha\beta})$ when α, β pairs are sampled from among these minimally frustrated states [Fig. 4 (bottom)]. That such an evolutionary process is feasible lends crucial support to the hypothesis that the existence of minimally frustrated stable states is a feature that has been acquired by complex biological networks over evolutionary time. Finally, preliminary data suggest that one can relax the assumption of a Boolean modeling framework without changing the conclusion—biological regulatory networks differ from random networks in their dynamical behavior (see SM [20], Sec. I J and Figs. S15–S18).

Discussion—In the energy landscape description of protein folding [38,39], the existence of minimally frustrated structural conformations distinguishes biological proteins from random heteropolymers. Here, we have shown that the existence of minimally frustrated stable

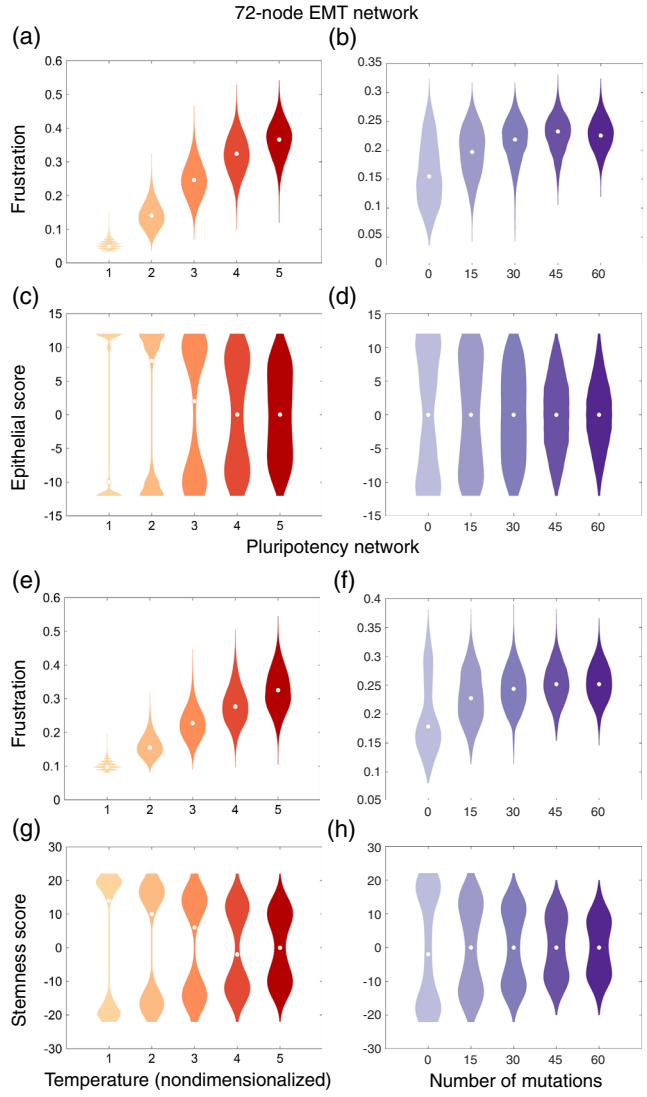

FIG. 3. High frustration network states are increasingly occupied under noisy dynamics or if the biological network accumulates mutations. (a),(e) Frustration of observed biological network states under noisy node dynamics. The dynamics become more and more noisy as the pseudotemperature is increased. (b),(f) Frustration of observed states when mutations are introduced into biological networks (without noise in the network dynamics). (c),(d) Epithelial scores of observed states under different levels of noise in the network dynamics (c) and when the network is mutated (d). (g),(h) Stemness scores of observed states under different levels of noise (g) and when the network is mutated (h). The white circle in each violin indicates the median.

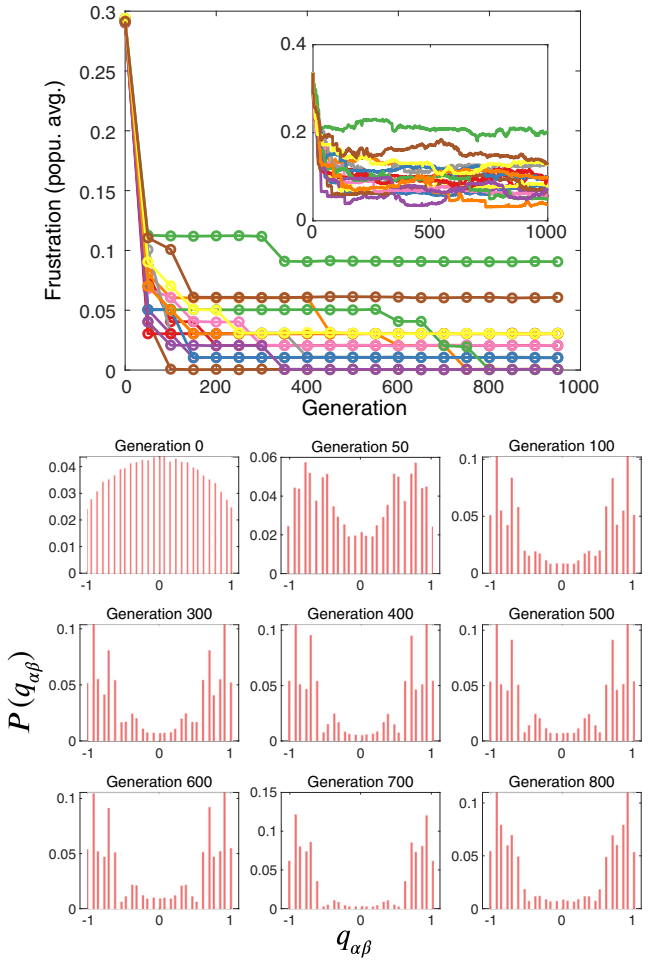

FIG. 4. Evolution of biological behavior by a population of random networks under selection for networks with low frustration states. Top: Frustration of the least frustrated observed state averaged over the networks in a population of 500 networks. Different curves indicate independent simulation runs. Inset: State frustration averaged over the end states of simulations starting from 50 random initial conditions for each network followed by averaging over the networks in the population. The initial population of 500 random networks was generated from the 26-node EMT network. Bottom: $P(q_{\alpha\beta})$ at different time points during the evolution simulation, shown for one of the simulation runs. To construct one of the histograms shown, we calculated $q_{\alpha\beta}$ for each network using the end states of simulations starting from 50 random initial conditions. This led to a total of $\binom{50}{2} = 1225$ values for each network. Each histogram shows the $q_{\alpha\beta}$ values for all the networks (a total of 500) in the population at a given time [total of $500 \times \binom{50}{k} = 612\,500$ values].

states similarly distinguishes biological regulatory networks from random networks. These minimally frustrated stable states represent canonical cell types, and because most random initial conditions dynamically evolve to one of the minimally frustrated stable states, biological networks can robustly establish cell types and regulate cell fate choice between these types. The number of commonly observed cell fates is thus limited to the number of expression patterns in these minimally frustrated stable states. The minimal frustration property distinguishes stable states corresponding to canonical cell types from other possible stable states of the biological network. In contrast, while a random network may have a collection of stable states with an expression pattern similar to that of a canonical cell type, these stable states will in no way be

special as compared to the numerous other stable states the random network can exhibit. Note that in the present study, we have only analyzed biological networks that regulate binary cell fate choice, which we argue are representative of cell fate decision making in biology (see SM [20], Sec. I A). The results presented here may not generalize to other biological networks such as signaling cascades or metabolic networks. Further, in this study, the dynamics of networks do not involve complex logic-based rules which often govern how different inputs to a network node combine to regulate node behavior. However, preliminary data involving the analysis of another network (shown in SM [20], Figs. S7 and S8) suggest that the ideas presented here can easily be generalized to networks whose dynamics are governed by more complex rules. Finally, a regulatory network inferred from gene expression data using an unsupervised, data-driven approach exhibited the same behavior as the regulatory networks taken from the literature [Fig. 1(f); also see SM [20], Sec. I G and Fig. S9]. This suggests that the results presented here are not simply an outcome of any network curation that may have been carried out in the studies from which the networks have been taken.

Cancer cells exhibit very noisy gene expression, which can be driven by multiple factors [40–44]. Our results suggest that given the high gene expression noise, cancer cells must frequently exhibit ambiguous cell fate choices. Such behavior has been widely reported, and noncanonical phenotypic states in cancer cells have been shown to be associated with disease aggressiveness. For example, hybrid epithelial-mesenchymal cells have been implicated in the metastatic aggressiveness of solid tumors [34]. Populations of small cell lung cancer cells treated with anticancer drugs have been shown to enrich for hybrid neuroendocrine-mesenchymal cells [16]. Lowering of network frustration upon the deletion from the EMT network of factors known to stabilize hybrid epithelial-mesenchymal cells (shown in SM [20], Fig. S12) further bolsters the evidence for a connection between noncanonical phenotypic states and high frustration in biological networks. Our model thus provides a new perspective on how noise in the dynamics of regulatory networks in cancer cells can contribute toward the failure of anticancer therapies—noise can facilitate the emergence of cancer cells that exhibit noncanonical phenotypic states. Additionally, accumulation of mutations in biological networks, another characteristic associated with cancer progression, will also promote aberrant cell fate choice. Estimation of network frustration from cancer cell gene expression data will be a direct test of the role of cell fates associated with high frustration states in cancer progression.

This work was supported by the National Science Foundation Grants No. PHY-1427654 and No. PHY-1935762. D. A. K. acknowledges support from the U.S.-Israel Binational Science Foundation Grant No. 2015/619.

*kessler@dave.ph.biu.ac.il
†h.levine@northeastern.edu

[1] O. Hobert, Science **319**, 1785 (2008).
[2] C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting held at Allerton House, Monticello, Illinois*, edited by P. Debrunner, J. Tsibris, and E. Munck (University of Illinois Press, Urbana, IL, 1969), pp. 22–24.
[3] R. Palmer, in *Lectures in the Sciences of Complexity*, 1st ed., edited by D. L. Stein (Addison-Wesley, Reading, MA, 1989), pp. 275–300.
[4] S. Huang, in *Gene Regulation and Metabolism: Post-Genomic Computational Approach*, edited by R. Hofestdt and J. Collado-Vides (MIT Press, Cambridge, MA, 2002), pp. 181–220..
[5] S. Huang, G. Eichler, B.-Y. Yaneer, and D. E. Ingber, Phys. Rev. Lett. **94**, 128701 (2005).
[6] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 1993).
[7] S. F. Gilbert, *Developmental Biology*, 3rd ed. (Sinauer Associates, Sunderland, MA, 1991), pp. 376–406; *Developmental Biology*, 3rd ed. (Sinauer Associates, Sunderland, MA, 1991), pp. 407–457.
[8] R. Albert, J. Cell Sci. **118**, 4947 (2005).
[9] M. Aldana and P. Cluzel, Proc. Natl. Acad. Sci. U.S.A. **100**, 8710 (2003).
[10] H. Yu and M. Gerstein, Proc. Natl. Acad. Sci. U.S.A. **103**, 14724 (2006).
[11] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[12] J. X. Zhou and S. Huang, Trends Genet. **27**, 55 (2011).
[13] R. Albert and J. Thakar, WIREs Syst. Biol. Med. **6**, 353 (2014).
[14] F. Font-Clos, S. Zapperi, and C. A. M. La Porta, Proc. Natl. Acad. Sci. U.S.A. **115**, 5902 (2018).
[15] P. W. Anderson, J. Less Common Met. **62**, 291 (1978).
[16] A. R. Udyavar, D. J. Wooten, M. Hoeksema, M. Bansal, A. Califano, L. Estrada, S. Schnell, J. M. Irish, P. P. Massion, and V. Quaranta, Cancer Res. **77**, 1063 (2017).
[17] D. Jia, J. T. George, S. C. Tripathi, D. L. Kundnani, M. Lu, S. M. Hanash, J. N. Onuchic, M. K. Jolly, and H. Levine, Phys. Biol. **16**, 025002 (2019).
[18] R. Chang, R. Shoemaker, and W. Wang, PLoS Comput. Biol. **7**, e1002300 (2011).
[19] S. Li, X. Zhu, B. Liu, G. Wang, and P. Ao, Oncotarget **6**, 13607 (2015).
[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.125.088101 for methods and for Figs. S1–S18, which includes Refs. [21–33].
[21] E. U. Azeloglu and R. Iyengar, Cold Spring Harbor Perspect. Biol. **7**, a005934 (2015).
[22] E. Watson, L. S. Yilmaz, and A. J. Walhout, Annu. Rev. Genet. **49**, 553 (2015).

[23] S. N. Steinway, J. G. T. Zanudo, W. Ding, C. B. Rountree, D. J. Feith, T. P. Loughran, and R. Albert, Cancer Res. **74**, 5963 (2014).
[24] S. N. Steinway, J. G. T. Zaudo, P. J. Michel, D. J. Feith, T. P. Loughran, and R. Albert, NPJ Syst. Biol. Appl. **1**, 15014 (2015).
[25] O. Ríos, S. Frias, A. Rodríguez, S. Kofman, H. Merchant, L. Torres, and L. Mendoza, Theor. Biol. Med. Model. **12**, 26 (2015).
[26] Y. Benjamini and Y. Hochberg, J. R. Stat. Soc. Ser. B **57**, 289 (1995).
[27] A. Terunuma, N. Putluri, P. Mishra, E. A. Math, T. H. Dorsey, M. Yi, T. A. Wallace, H. J. Issaq, M. Zhou, J. K. Killian *et al.*, J. Clin. Invest. **124**, 398 (2014).
[28] S. Chandriani, E. Frengen, V. H. Cowling, S. A. Pendergrass, C. M. Perou, M. L. Whitfield, and M. D. Cole, PLoS One **4**, e6693 (2009).
[29] H. Han, J.-W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, C. Y. Kim, M. Lee, E. Kim *et al.*, Nucleic Acids Res. **46**, D380 (2018).
[30] T. Moerman, S. Aibar Santos, C. Bravo Gonzlez-Blas, J. Simm, Y. Moreau, J. Aerts, and S. Aerts, Bioinformatics **35**, 2159 (2019).
[31] B. Huang, M. Lu, D. Jia, E. Ben-Jacob, H. Levine, and J. N. Onuchic, PLoS Comput. Biol. **13**, e1005456 (2017).
[32] E. Mones, L. Vicsek, and T. Vicsek, PLoS One **7**, e33799 (2012).
[33] J. A. Hartigan and P. M. Hartigan, Ann. Stat. **13**, 70 (1985).
[34] M. K. Jolly, M. Boareto, B. Huang, D. Jia, M. Lu, E. Ben-Jacob, J. N. Onuchic, and H. Levine, Front. Oncol. **5**, 155 (2015).
[35] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).
[36] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[37] W. K. Hastings, Biometrika **57**, 97 (1970).
[38] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[39] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).
[40] K. Hinohara, H.-J. Wu, S. Vigneau, T. O. McDonald, K. J. Igarashi, K. N. Yamamoto, T. Madsen, A. Fassl, S. B. Egri, M. Papanastasiou *et al.*, Cancer Cell **34**, 939 (2018).
[41] A. F. Domingues, R. Kulkarni, G. Giotopoulos, S. Gupta, S. Tan, E. Foerner, R. R. Adao, K. Zeka, B. J. Huntly, S. Prabakaran *et al.*, bioRxiv, 446096 (2018) https://doi.org/10.1101/446096.
[42] S. Yamamoto, Z. Wu, H. Russnes, S. Takagi, G. Peluffo, C. Vaske, X. Zhao, H. MoenVollan, R. Maruyama, M. Ekram *et al.*, Cancer Cell **25**, 762 (2014).
[43] A. Pastore, F. Gaiti, S. X. Lu, R. M. Brand, S. Kulm, R. Chaligne, H. Gu, K. Y. Huang, E. K. Stamenova, W. Bguelin *et al.*, Nat. Commun. **10**, 1874 (2019).
[44] A. Carrer, S. Trefely, S. Zhao, S. L. Campbell, R. J. Norgard, K. C. Schultz, S. Sidoli, J. L. D. Parris, H. C. Affronti, S. Sivanand *et al.*, Cancer Discov. **9**, 416 (2019).