Few inputs can reprogram biological networks

Liu, Slotine and Barabasi[1] identify subsets U of nodes in complex networks, which are required to exert full control of these networks. Control in this context means that for each possible state S of the network there exist inputs for all nodes in U, which are sufficient to force the network to state S[1]. Application of the methodology to gene regulatory networks suggests that roughly 80% of all nodes must be controlled to drive such a network. This seems to contradict recent empirical findings[2–6] in the cellular reprogramming field.

Mammalian cells have been shown to be efficiently reprogrammable towards other cellular phenotypes[2–6]. Controllability of such complex transitions in transcriptional networks underlying cellular phenotypes appears to be an intrinsic biological property and is being used for the development of novel disease models and cellular therapeutics in regenerative medicine. However, in contrast to technical or social networks, biological networks show a high degree of intrinsic co-regulation. Co-regulation appears to quench the admissible space of states of gene expression networks to a combinatorial expression space with relatively low dimensionality. Recent publications suggest that this subspace is spanned by only a few (five to fifteen) genome-wide differential expression patterns, which are—surprisingly—sufficient to characterize most observable biological phenotypes[7–9]. Each of these patterns is given by weighted sums over the expression of relatively extensive gene sets contributing coherently to the respective pattern[7–9]. Although these empirical studies suggest linear structures underlying the respective subspaces, nonlinear or star-shaped topologies cannot be excluded.

Moreover, genome-wide co-regulation may result in dynamic response features, as observed in chemical networks. Hence, owing to their low dimensionality, full control and reprogramming of biological systems may be achieved by only a few key control factors, seemingly contradicting the more general concept put forward by ref. 1. Recent experimental studies show proof of concept for a wide range of reprogramming of biological systems using overexpression of only one to five transcription factors[2–6,10], which effectively regulate only a subset of downstream genes. We note that the number of nodes (five or fewer genes out of about 30,000) needed to fully control a biological system[2–6,10] is much less than the estimate of 80% of all nodes[1].

Thus, for the special case of biological systems, it may be sufficient to weaken the standard definition of controllability of networks—control of large sets of nodes as suggested by ref. 1—to mean the controllability of restricted, biologically admissible network states.

Franz-Josef Müller[1] & Andreas Schuppert[2,3]
[1]Zentrum für Integrative Psychiatrie, Niemannsweg 147, 24105 Kiel, Germany.
e-mail: fj.mueller@zip-kiel.de
[2]Graduiertenschule Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, Schinkelstrasse 2, 52062 Aachen, Germany.
[3]Bayer Technology Services GmbH, Process Technology, Bdg 9115, 51368 Leverkusen, Germany.

1. Liu, Y., Slotine, J. & Barabási, A. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
3. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
4. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
5. Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
6. Huang, P. *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386–389 (2011).
7. Lukk, M. *et al.* A global map of human gene expression. *Nature Biotechnol.* **28**, 322–324 (2010).
8. Müller, F. *et al.* A bioinformatic assay for pluripotency in human cells. *Nature Methods* **8**, 315–317 (2011).
9. Dudley, J. T., Tibshirani, R., Deshpande, T. & Butte, A. J. Disease signatures are robust across tissues and experiments. *Mol. Syst. Biol.* **5**, 307, doi:10.1038/msb.2009.66 (2009).
10. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**, 411–419 (2009).

Liu *et al.* reply

Müller and Schuppert[1] describe an exception to our finding[2] that roughly 80% of the nodes must be controlled to gain full control over gene regulatory networks. Yet our result hides subtleties that reveal as much about controllability as about the limits of our current understanding of biological networks.

Reprogramming cells from one phenotype towards another is not the same as full control of the regulatory process. Controllability, as explored by us, concerns the ability to drive the system from any point to any other point in the state space in a finite time. Stable phenotypes are best seen as attractors of the system's regulatory dynamics[3], separated by potential barriers that are difficult to cross under normal circumstances[4,5]. Reprogramming therefore concerns our ability to drive the system from one attractor to another—that is, controllability within a restricted subspace of the original state space. Consequently, such a procedure may require fewer driver nodes.

This can be illustrated by the example of a directed star with one central hub and $N - 1$ leaf nodes, capturing the common situation in which the expression of $N - 1$ genes is co-regulated by a single transcription factor. To gain full control of the system's dynamics we need to control $N_D = N - 1$ nodes, that is, all but a single node. (Here, N_D denotes the number of driver nodes and N is the total number of nodes in the network.) Yet, if we do not care about the relative differences in the expression pattern of the leaf nodes, then it is sufficient to control the central hub, that is, $N_D = 1$. This does not imply full
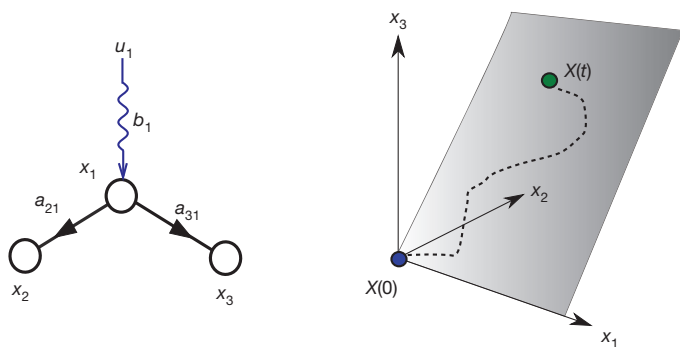
BRIEF COMMUNICATIONS ARISING

Figure 1 | Controlling the central hub of a directed star will cause the system to get stuck in a particular subspace. In the left panel, the control input u_1 (represented by the blue wavy line) is applied to the central hub x_1 of the directed star with $N = 3$ nodes. The system will be stuck in a plane described by $a_{31}x_2(t) = a_{21}x_3(t)$, as shaded in the right panel. The system is controllable within this subspace, that is, it can be driven from any initial state (blue point) to any desired final state (green point) in finite time. However, it is not controllable in the whole state space.

control, because the system is stuck in a subspace of the full state space (see Fig. 1), but such limited control may be sufficient for reprogramming.

Our estimate that approximately 80% of nodes must be controlled to fully control the gene regulatory network is based on very incomplete maps of regulatory networks. Current estimates indicate that the protein–protein interaction maps cover less than 10% of all potential protein–protein interactions[6,7]. For regulatory networks we lack such estimates, but probably the coverage is even lower. Assuming that with advances in mapping the number of regulatory links will increase tenfold, our analytical results (equations (4) and (5) in ref. 2) predict that n_D ($= N_D/N$) will drop to $(0.8)^{10} \approx 11\%$. Hence, the observed large n_D is partially driven by the incompleteness of regulatory maps. Furthermore, the presence of gene auto-regulation will create self-edges (that is, nodes pointing to themselves), making n_D drop even further (M. Pósfai, Y.-Y.L., J.-J.S. and A.-L.B., manuscript in preparation). Finally, n_D could be further reduced by nonlinearities in the dynamics, which would potentially help the

system explore its state space, enhancing its controllability and thus decreasing n_D (refs 8 and 9).

Thus, controllability offers an appropriate framework within which to explore many cellular processes, potentially including cell reprogramming. To reach its full potential, however, we need a better understanding of both the structure and the dynamics of regulatory processes.

Yang-Yu Liu[1,2], Jean-Jacques Slotine[3,4] & Albert-László Barabási[1,2,5]
[1]Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts 02115, USA.
e-mail: alb@neu.edu
[2]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.
[3]Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
[4]Department of Mechanical Engineering and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
[5]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

1. Müller, F.-J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478,** doi:10.1038/nature10543 (this issue).
2. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473,** 167–173 (2011).
3. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101,** 4781–4786 (2004).
4. Waddington, C. H. *The Strategy of the Genes* (Geo Allen & Unwin, 1957).
5. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94,** 1–4 (2005).
6. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nature Methods* **6,** 83–90 (2009).
7. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Rev. Genet.* **12,** 56–68 (2011).
8. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
9. Isidori, A. *Nonlinear Control Systems* (Springer, 1995).