SYSTEMS BIOLOGY

Attractors and Democratic Dynamics

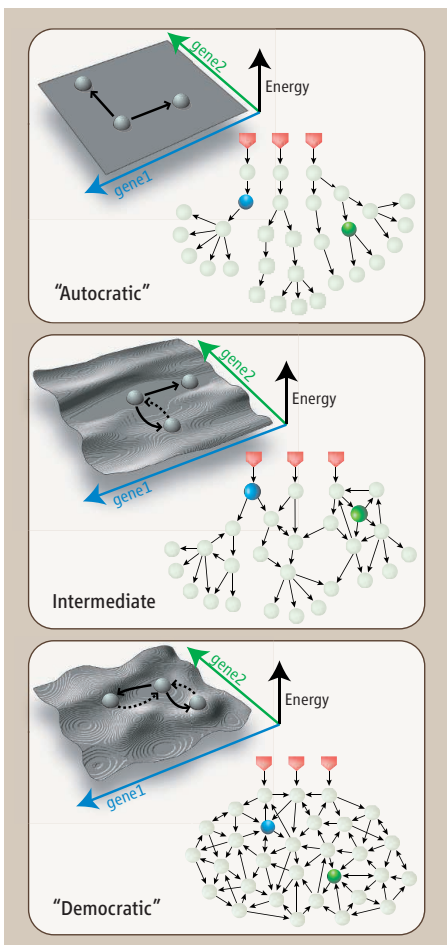
Cellular transcription networks are conceptualized as distributed control systems that regulate gene expression.

Yaneer Bar-Yam,[1] Dion Harmon,[1] Benjamin de Bivort[1,2]

The functional identity of a cell is largely determined by the regulated expression (transcription) of thousands of genes, so how it maintains a particular transcriptional state is of critical importance. Developmental biologists study how embryonic cells navigate a series of intermediate transcriptional states before settling into a final adult state; microbiologists identify the mechanisms by which transcription is altered by environmental perturbation; and oncologists seek to identify how cells switch from benign to cancerous. Consider two concepts of transcriptional regulation. In a "molecular autocracy," master genes respond to environmental or developmental stimuli by regulating thousands of genes, either directly or through other transcription factors. In a "molecular democracy," all genes exert a regulatory influence on all other genes, and phenotypic change (altered cell behavior) is brought about through the concerted action of thousands of genes. These scenarios are extreme and cells operate under a condition that is somewhere intermediate (see the figure) (*1*). But the choice of concept affects how regulation is studied.

The autocratic framework can be directly investigated by studies of individual molecular mechanisms and has been the starting point for discussions of biological processes. But a broader understanding of regulatory mechanisms is needed that incorporates essential features of both extreme views. The democratic framework relies on mutual regulation, which tends toward a self-consistent gene expression state that is stable in the face of fluctuations. In other words, this view has its roots in the conceptual understanding of stability and homeostasis of cell types (*2*, *3*). The democratic view has only recently gained empirical support, perhaps because its characterization involves studies of genome-wide dynamical processes.

A dynamic system with extensive mutual regulation tends to transition toward particular states, known as attractors, over time (often envisioned as valleys in a landscape). Background "noise" causes deviation in one cell over time, and among cells at one instant, but they recover. That there is an attractor state in the space of transcriptional states (*4*, *5*) supports the prediction of a democratic system. Chang *et al*. (*6*) recently identified transcriptional variability in clonally related mouse hematopoietic precursor cells, and separated the cells into several groups with expression differences in thousands (but still a minority) of genes. Over days, these cell group lineages converged to the same transcriptional state distribution. That is, the cell groups became indistinguishable, having the same average gene expression, as well as noise-induced variation, among individual cells. Such convergence is the signature of an attractor, in which many individual differences in transcription are insufficient to change the overall cell phenotype (a "controlling" majority of transcribed genes does not change) and mutual interactions among the genes cause trends toward specific mutually reinforcing states.



Transcription regulatory architecture. In autocratic regulatory networks (**top**), individual master regulator genes (pointed squares) are stimulated by external signals and control many other genes (circles). As shown by the energy landscape, the transcriptional states (spheres) may have no preferences (black arrows represent changes in expression of genes 1 and 2). In democratic networks (**bottom**), all genes act as mutual regulators. A few specific gene expression patterns become stable, shown as basins of attraction (cell types) in the landscape. Once a cell reaches one of these states, changing the expression of one gene is unlikely to switch the cell type (black arrows). Intermediate networks (**middle**) have mutual regulation, but certain genes (blue circle) are major controllers.

The attractor paradigm has practical implications: If distinct cell types (such as a precursor cell and a fully differentiated cell) correspond to distinct attractors, then there are multiple parallel ways to shift the transcriptional state from one attractor to another. Such families of trajectories are expected to engage multiple interconnected signaling pathways whose collective behavior (and outcome) is simple. In a limited way, this has been observed in the differentiation of immune cells (*4*) and stem cells (*7*). If the attractor picture is generically valid, it should be possible to create cocktails of large numbers of gene products that switch cells between different types. Any sufficiently large subset of gene products should be sufficient to cause the switch. Consider the number of gene expression levels that are needed to robustly characterize distinct cell types. An analysis (see fig. S1) of the transcriptional profiles of 79 human tissues and tumor cell types (*8*) reveals that about 200 highly variable gene expression values are sufficient to capture the relationships among the tissues and tumor cells, whereas fewer than 80 are not. By this measure, cocktails with a couple of hundred gene products chosen to mimic the differences between two cell types should generically cause transitions between them.

Still, paradoxically, Chang *et al*. (*6*) segregated cells according to the expression of a single gene, and showed that specific genes can

[1]New England Complex Systems Institute, 24 Mt. Auburn Street, Cambridge, MA 02138, USA. [2]Roland Institute at Harvard University, 100 Edwin Land Boulevard, Cambridge, MA 02139, USA. E-mail: yaneer@necsi.edu

Published by AAAS

control the overall cell state (and cell fate). This has also been observed in developing cells that are highly sensitive to external signaling molecules (*9*), but attain a highly stable differentiated state (*4*). More generally, cells are robust to noise and small perturbations in transcription (*10*), but sensitive to small changes in specific external (*11*) and internal (*12*) cues. Indeed, whereas regulatory networks have been characterized as robust to random failure and vulnerable to targeted attack (*13*), from a regulatory perspective, generic stability with sensitivity to specific perturbations is a positive property rather than a negative one (*14*). What is missing is a framework in which individual genes and collective states can be considered together.

What framework should be used to study collective state control? The difficulty is that for individual gene effects, individual transcription levels are important. For attractors, collective dynamics of the transcriptome within a cell type, rather than specific gene expression signatures, characterize cell behavior (not just cell type differences). What is needed are control coefficients that measure change in collective states relative to archetypes (see supporting text), and in relation to individual gene transcription level changes. Relating the variation of small sets of gene expression values to deviation or conformity to archetypes can provide a framework to study the interplay of attractors and master regulators. Such observations, best taken from unaveraged data, should identify the dispersal and convergence of cells near an attractor, and the mechanisms of homeostatic control. Using multiple archetypes also should enable the study of cell fate trajectories.

References

1. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
2. C. H. Waddington, *Principles of Embryology* (Allen and Unwin, London, 1956).
3. S. A. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
4. S. Huang *et al.*, *Phys. Rev. Lett.* **94**, 128701 (2005).
5. P. Ao *et al.*, *Med. Hypotheses* **70**, 678 (2008).
6. H. H. Chang *et al.*, *Nature* **453**, 544 (2008).
7. A. G. Bang, M. K. Carpenter, *Science* **320**, 58 (2008).
8. A. I. Su *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6062 (2004).
9. K. L. Medina *et al.*, *Dev. Cell* **7**, 607 (2004).
10. J. A. de Visser *et al.*, *Evolution* **57**, 1959 (2003).
11. G. Balazsi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7841 (2005).
12. R. Losick, C. Desplan, *Science* **320**, 65 (2008).
13. R. Dobrin *et al.*, *BMC Bioinformatics* **5**, 10 (2004).
14. Y. Bar-Yam, I. R. Epstein, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4341 (2004).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5917/1016/DC1
SOM Text
Figs. S1 and S2

10.1126/science.1163225

GEOCHEMISTRY

Making a Crust

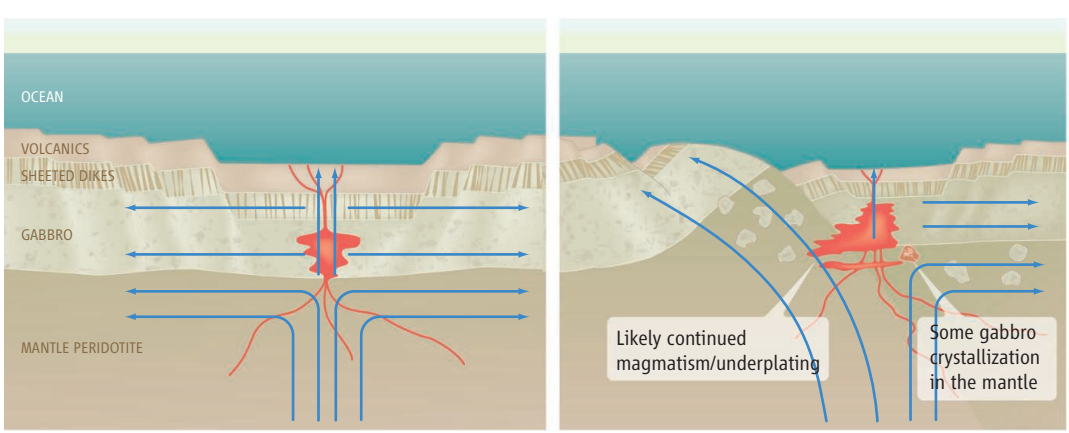
Dating zircons from the mid-Atlantic ridge provides clues about how the oceanic crust is formed.

Peter J. Michael[1] and Michael J. Cheadle[2]

The discovery of seafloor spreading in the 1960s enabled the formulation of the theory of plate tectonics. Modern geology textbook wisdom provides an image of magnetic stripes being made neatly on the seafloor when new oceanic crust is produced at a very narrow mid-ocean ridge axis, and subsequently moves outward. But how is the 6-km-thick crust actually produced at the ridge axis? On page 1048 of this issue, Lissenberg *et al.* (*1*) address this and other questions of ocean crust formation by applying state-of-the-art dating techniques to date samples from the mid-Atlantic ridge. They report that the tiny zircon crystals that are relatively abundant in the oceanic crust make it easy to date the crust, thereby providing a clearer picture of the formation processes involved.

Marine geologists have a good understanding of the different rock layers that are present in ocean crust. A kilometer-thick basalt layer at the seafloor is produced by magma that is fed through sheet-like intrusions called dikes. The deepest 4 km of the crust is formed from basaltic magma that crystallizes to form intrusive gabbro rock. Numerous studies have elucidated how long it takes to form the basaltic and dike layers (*2*). Much less is known about how the gabbroic lower crust accretes, because it is less accessible, buried beneath the basalts and dikes. Does it grow from the top down, or randomly? Is the width of accretion the same as the width of the axial valley (~10 to 12 km)? Or does some of it accrete farther off axis? How deep do gabbros crystallize? Are they confined to the crust, or can they crystallize in the uppermost mantle? How quickly does the lower crust cool and become rigid after it crystallizes?

The best exposures of gabbroic crust are found at slow-spreading ridges, where faulting accounts for some of the plate spreading. Earlier work (*3*–*5*) used an ion microprobe technique to date zircons and concentrated on oceanic core complexes that expose large sections of gabbroic crust by detachment faulting (*6*). Although the zircon crystallization ages were consistent with the magnetic spreading ages, about 10% of their analyses showed anomalous old ages. These results suggested that some gabbro crystallization had occurred at depth in the upwelling mantle before being transported to shallower levels in the crust.



Crustal formation. (**Left**) Traditional model for construction of oceanic crust where there is no detachment faulting, similar to where Vema lithospheric section may have been produced as suggested by (*1*). (**Right**) Model for more complicated ridges, with some gabbro crystallization in the uppermost mantle and detachment faulting exposing crustal sections at core complexes, as advocated by (*3*).

[1]Department of Geosciences, University of Tulsa, Tulsa, OK 74104, USA. [2]Department of Geology and Geophysics, University of Wyoming, Laramie, WY 82071, USA. E-mail: pjm@utulsa.edu; cheadle@uwyo.edu

CREDIT: ADAPTED FROM T. SCHROEDER *ET AL.*, *GEOCHEM. GEOPHYS. GEOSYST.* 8, Q06015 (2007)

Supporting Online Material for

Attractors and Democratic Dynamics

Yaneer Bar-Yam,* Dion Harmon, Benjamin de Bivort

*E-mail: yaneer@necsi.edu

Published 20 February 2009, *Science* **323**, 1016 (2009)
DOI: 10.1126/science.1163225

This PDF file includes:

Global Patterns of Gene Expression

and

Master Regulators:

Supporting Online Text

for

Attractors and Democratic Dynamics

Yaneer Bar-Yam[1], Dion Harmon[1], and Benjamin de Bivort[1,2]

[1]New England Complex Systems Institute,

24 Mt. Auburn St., Cambridge, Massachusetts 02138

[2] Rowland Institute at Harvard University,

100 Edwin Land Blvd., Cambridge, MA 02139

Genome Wide Dynamics and Control: Archetypes for Collective Behavior

Differences in perspectives about regulatory structures are also related to differences in concept and representation of dynamical processes. Describing individual gene effects begins with identifying individual genes, mechanisms of gene interactions, and pathways of gene products. Describing attractors involves characterizing the convergence of transcriptome wide cell states where majorities of genes determine behavior rather than any one. The difference between low dimensional (few variable) and high dimensional (many variable) collective dynamics is central. Characterizing cellular regulatory networks more generally as distributed control systems where individual genes can exert strong influence requires bridging these two views.

We describe a framework in which individual genes and collective states can be considered together to evaluate their mutual influence. The difficulty we overcome is the contrast in the quantities needed to describe the two different pictures. What is needed are analogs of control coefficients, which have been used to study the impact of individual catalysts on system metabolic flows (*S1*).

Identifying such coefficients requires a measure of differences of collective states. While Pearsons correlation might be used (*S2*), in order to define a fundamentally justifiable measure we identify an archetype, e.g. a representation of a particular cell type, using expression values of all genes $\{e_i^\alpha\}$, where i is the gene index, and α is a cell type label. These values may be taken as a representative member, or a mean over a population of cells of the same type. When a cell has that type, individual expression values may deviate from the archetype values. However, when considered over all genes, the deviations are bounded.

We measure the deviation of a gene expression value e_i relative to an archetype, normalized by the expected deviation over a reference population of cells,

$$d_i^\alpha = \left(\log(e_i) - \log(e_i^\alpha)\right)/\sigma_i. \tag{1}$$

We use logarithms of expression values to obtain better-behaved distributions. The normalization σ_i is chosen to establish a common range of values for d_i^α and can be set to the standard deviation over the reference cell population of $\log(e_i)$, i.e. a population of cells that are of a particular phenotype (σ_i may have additional labels to identify the reference population). If the population is not large we can approximate $\sigma_i = \sigma$ by considering all

genes together in taking the standard deviation, or combining multiple phenotype populations. The proximity of an arbitrary state to the archetype, the conformity or conformance, is given by

$$m^\alpha = \frac{1}{N}\sum_i \mathrm{f}(d_i^\alpha), \tag{2}$$

where f is a function that is 1 for values close to zero, i.e. when a gene expression level is proximate to the archetypical value, and goes to zero as it deviates therefrom. N is the number of genes. The purpose of this function is to prevent individual gene deviations from determining the distance, which should instead depend on whether or not many expression levels are close to archetypical values. We use

$$m^\alpha = \frac{1}{N}\sum_i \left(1 - \tanh((d_i^\alpha)^2)\right). \tag{3}$$

Control coefficients are specified by the rate of change of the collective displacement $\hat{m}^\alpha = (1 - m^\alpha)$ with respect to the control parameter, measured logarithmically—i.e. the exponent of a power law relationship. Thus we define control or sensitivity coefficients as

$$c_i = \frac{e_i}{\hat{m}^\alpha}\frac{d\hat{m}^\alpha}{e_i} = \frac{d\log(\hat{m}^\alpha)}{d\log(e_i)}. \tag{4}$$

As an example, we evaluated this for Sca-1 in the data of Chang et al. (*S2*), after performing a number of tests (see Supplementary Figure 2), and found a control coefficient of 0.52, consistent with the coupling found between this gene and collective behavior in that paper. This is obtained despite the cells having a high concentration of Sca-1 protein on the cell surface having lower mRNA expression compared to cells with intermediate protein concentrations, presumably due to feedback. We also obtained many other control coefficients from the same data, including 2.20 for Sfpi1, and -.59 for Gata1, lineage-specific transcription factors involved in stem-cell differentiation. Given the nature of the data, this indicates correlation relative to the chosen states and not causation, and the latter can also be characterized by relevant experiments.

Supplementary Figure 1

Principle component analysis reveals transcriptome attractors across tissue types. A) The transcriptional profiles of 79 human tissue and tumor cell types (*S3*) fall into several clusters when they are plotted in the two dimensions that draw the greatest distinctions among tissue types when considering the 80 genes with most varying expression levels. The dimensions of maximum variation are obtained by principal component analysis (*S4*). Tissues are color coded by category. B) As in A for analyses done with the n most varying genes as indicated. The tissue type clusters approach their final conformation when hundreds of genes are considered.

Supplementary Figure 2

Collective dynamics of cell types. Dots represent high dimensional cell states from Chang et al. (*S2*), with two replicates each of cultures distinguished by low (red), mid (purple) and high (blue) concentrations of the surface marker Sca-1, and subsequent convergence of these cultures after 6-days. An optimized two-dimensional embedding of distances given by $\hat{m}^\alpha$ (see supporting online text) is shown, with scale bar in units of $\hat{m}^\alpha$. Control coefficients can be calculated as the ratio of the $\log(\hat{m}^\alpha)$ (grey arrows) change to the log of individual gene expression value change using as reference the Sca-1 low 0 day state.

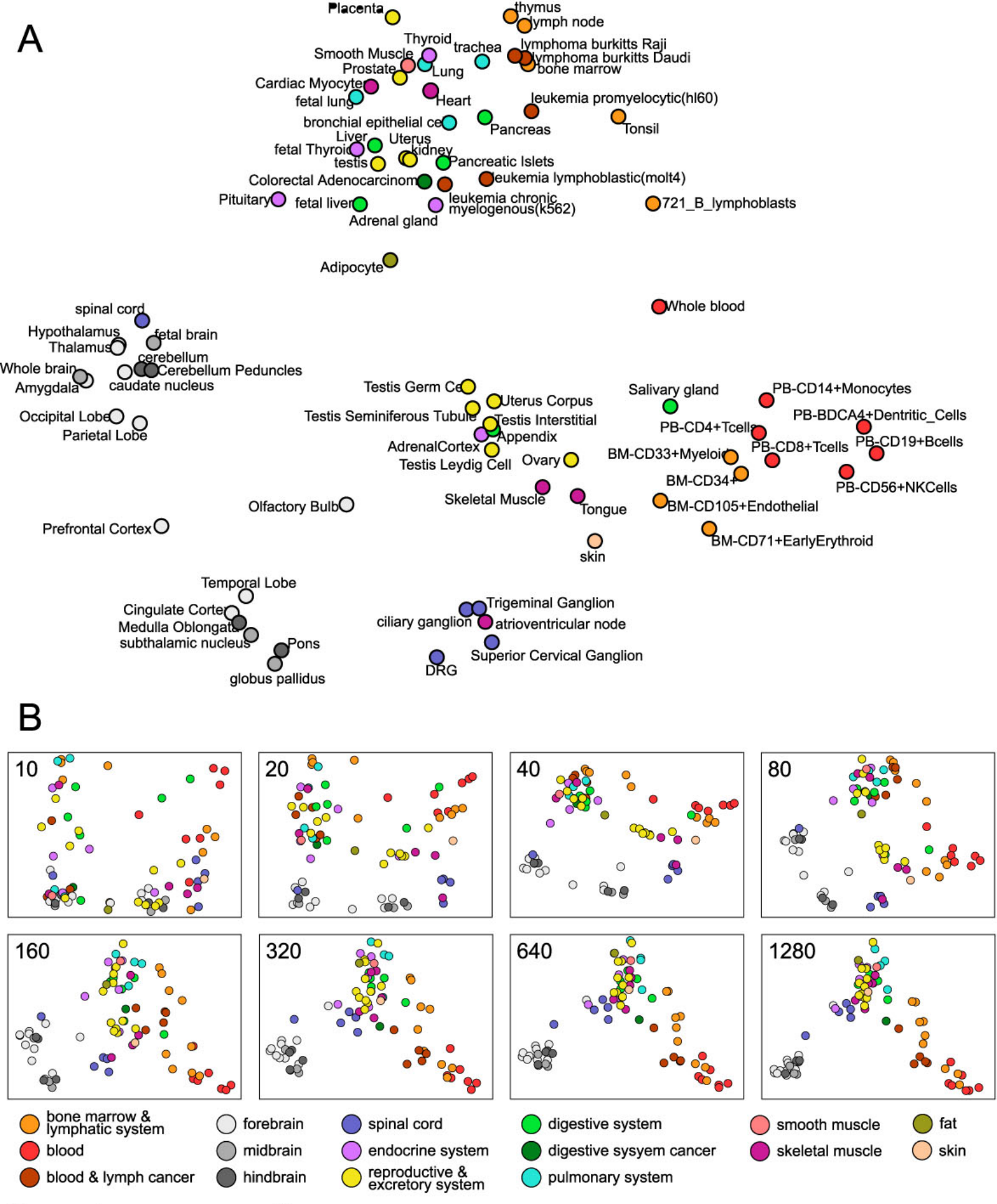
References

S1. H. Kacser, J. A. Burns, The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65 (1973).

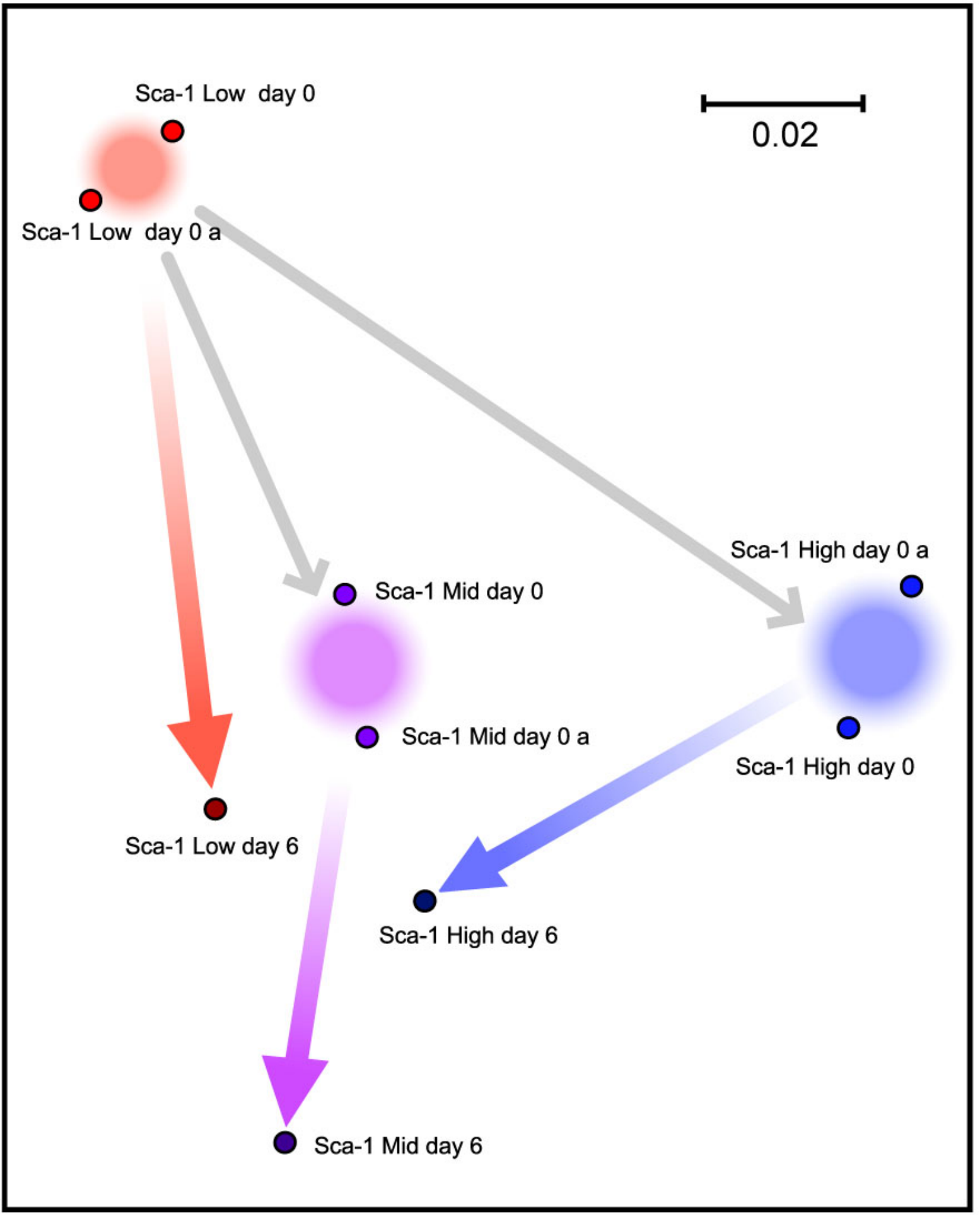
S2. H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, S. Huang, *Nature* **453**, 544 (2008).

S3. A. I. Su et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6062 (2004).

S4. I. Joffe, *Principal Component Analysis* (Wiley, Hoboken, NJ, 2005).



Supplementary figure 1



Supplementary figure 2