Life's Complexity Pyramid

Zoltán N. Oltvai, *et al.*

Science **298**, 763 (2002);
DOI: 10.1126/science.1078563

This inevitably leads to widespread haploinsufficiency at several gene loci, only a fraction of which provide the nascent tumor cell with some degree of selective advantage. Do tumor suppressor genes exist for which haploinsufficiency is more strongly selected for than complete inactivation? Only accurate and quantitative genome-wide expression profiling by microarray or proteomic analysis will enable such gene-dosage defects to be identified. Analyzing targeted hypomorphic alleles in experimental animals should facilitate the identification of modifier genes, their tissue-specific dosage thresholds, and their interaction with more penetrant tumor suppressor genes and environmental mutagens.

References

1. A. G. Knudson Jr., *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820 (1971).
2. S. B. Gruber *et al.*, *Science* **297**, 2013 (2002).
3. K. H. Goss *et al.*, *Science* **297**, 2051 (2002).
4. K. Spring *et al.*, *Nature Genet.* **32**, 185 (2002).
5. S. Venkatachalam *et al.*, *EMBO J.* **17**, 4657 (1998).
6. W. J. Song *et al.*, *Nature Genet.* **23**, 166 (1999).
7. R. Smits *et al.*, *Gastroenterology* **119**, 1045 (2000).
8. M. L. Fero *et al.*, *Nature* **396**, 177 (1998).
9. M. Kucherlapati *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9924 (2002).
10. M. Swift *et al.*, *N. Engl. J. Med.* **325**, 1831 (1991).
11. N. G. Howlett *et al.*, *Science* **297**, 606 (2002).
12. H. Meijers-Heijboer *et al.*, *Nature Genet.* **31**, 55 (2002).
13. K. Inoue *et al.*, *Genes Dev.* **15**, 2934 (2001).
14. H. Miyoshi *et al.*, *Cancer Res.* **62**, 2261 (2002).
15. Y. Zhu *et al.*, *Science* **296**, 920 (2002).
16. C. Wetmore *et al.*, *Cancer Res.* **60**, 2239 (2000).
17. R. H. Zurawel *et al.*, *Genes Chr. Cancer* **28**, 77 (2000).
18. B. Kwabi-Addo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11563 (2001).

PERSPECTIVES: SYSTEMS BIOLOGY

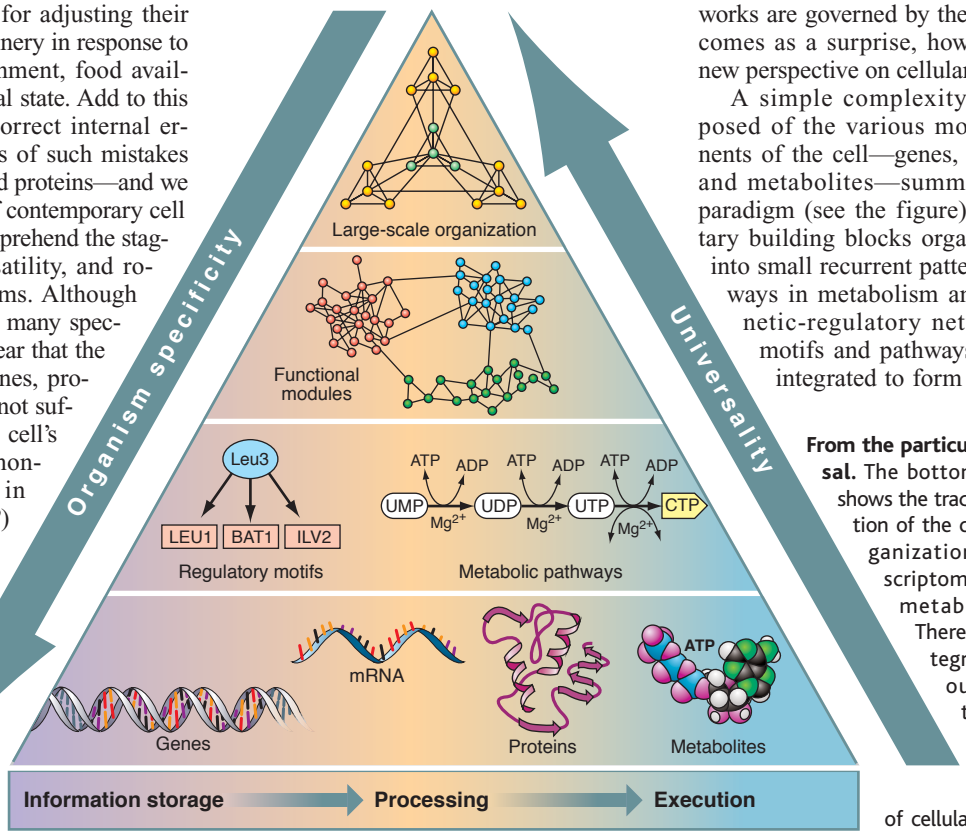
Life's Complexity Pyramid

Zoltán N. Oltvai and Albert-László Barabási

Cells and microorganisms have an impressive capacity for adjusting their intracellular machinery in response to changes in their environment, food availability, and developmental state. Add to this an amazing ability to correct internal errors—battling the effects of such mistakes as mutations or misfolded proteins—and we arrive at a major issue of contemporary cell biology: our need to comprehend the staggering complexity, versatility, and robustness of living systems. Although molecular biology offers many spectacular successes, it is clear that the detailed inventory of genes, proteins, and metabolites is not sufficient to understand the cell's complexity (*1*). As demonstrated by two papers in this issue—Lee *et al.* (*2*) on page 799 and Milo *et al.* (*3*) on page 824—viewing the cell as a network of genes and proteins offers a viable strategy for addressing the complexity of living systems.

According to the basic dogma of molecular biology, DNA is the ultimate depository of biological complexity. Indeed, it is generally accepted that information storage, information processing, and the execution of various cellular programs reside in distinct levels of organization: the cell's genome, transcriptome, proteome, and metabolome. However, the distinctness of these organizational levels has recently come under fire. For example, although long-term information is stored almost exclusively in the genome, the proteome is crucial for short-term information storage (*4, 5*) and transcription factor–controlled information retrieval is strongly influenced by the state of the metabolome. This integration of different organizational levels increasingly forces us to view cellular functions as distributed among groups of heterogeneous components that all interact within large networks (*6, 7*). There is clear evidence for the existence of such cellular networks: For example, the proteome organizes itself into a protein interaction network and metabolites are interconverted through an intricate metabolic web (*7*). The finding that the structures of these networks are governed by the same principles comes as a surprise, however, offering a new perspective on cellular organization.

A simple complexity pyramid composed of the various molecular components of the cell—genes, RNAs, proteins, and metabolites—summarizes this new paradigm (see the figure). These elementary building blocks organize themselves into small recurrent patterns, called pathways in metabolism and motifs in genetic-regulatory networks. In turn, motifs and pathways are seamlessly integrated to form functional mod-



CREDIT: KATHARINE SUTLIFF/*SCIENCE*

From the particular to the universal. The bottom of the pyramid shows the traditional representation of the cell's functional organization: genome, transcriptome, proteome, and metabolome (level 1). There is remarkable integration of the various layers both at the regulatory and the structural level. Insights into the logic of cellular organization can be achieved when we view the cell as a complex network in which the components are connected by functional links. At the lowest level, these components form genetic-regulatory motifs or metabolic pathways (level 2), which in turn are the building blocks of functional modules (level 3). These modules are nested, generating a scale-free hierarchical architecture (level 4). Although the individual components are unique to a given organism, the topologic properties of cellular networks share surprising similarities with those of natural and social networks. This suggests that universal organizing principles apply to all networks, from the cell to the World Wide Web.

Z. N. Oltvai is in the Department of Pathology, Northwestern University, Chicago, IL 60611, USA. E-mail: zno008@nwu.edu A.-L. Barabási is in the Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: alb@nd.edu

ules—groups of nodes (for example, proteins and metabolites) that are responsible for discrete cellular functions (*6*). These modules are nested in a hierarchical fashion and define the cell's large-scale functional organization (*8*).

The papers by Lee *et al.* (*2*) and Milo *et al*. (*3*) offer key support for the cellular organization suggested by the complexity pyramid (see the figure). Using 106 tagged transcription factors of the budding yeast *Saccharomyces cerevisiae*, Lee *et al*. have systematically identified the genes to whose promoter regions these transcription factors (regulators) bind. After establishing transcription factor binding at various confidence levels, they uncovered from 4000 to 35,000 genetic-regulatory interactions, generating the most complete map of the yeast regulatory network to date. The map allows the authors to identify six frequently appearing motifs, ranging from multi-input motifs (in which a group of regulators binds to the same set of promoters) to regulatory chains (alternating regulator-promoter sequences generating a clear temporal succession of information transfer). A similar set of regulatory motifs was recently uncovered in the bacterium *Escherichia coli* by Alon and co-workers (*9*). In their new study, Milo, Alon and colleagues provide evidence that motifs are not unique to cellular regulation but emerge in a wide range of networks, such as food webs, neural networks, computer circuits, and even the World Wide Web (*3*). They identified small subgraphs that appear more frequently in a real network than in its randomized version. This enabled them to distinguish coincidental motifs from recurring significant patterns of interconnections.

An important attribute of the complexity pyramid is the gradual transition from the particular (at the bottom level) to the universal (at the apex). Indeed, the precise repertoire of components—genes, metabolites, proteins—is unique to each organism. For example, 43 organisms for which relatively complete metabolic information is available share only ~4% of their metabolites (*7*). Key metabolic pathways are frequently shared, however, and—as demonstrated in this issue (*2, 3*) and elsewhere (*9*)—so are some of the motifs. An even higher degree of universality is expected at the module level; although quantitative evidence is lacking, it is generally believed that key properties of functional modules are shared across most species. The hierarchical relationship among modules, in turn, appears to be quite universal, shared by all examined metabolic (*8*) and protein interaction networks. Finally, the scale-free nature (*7*) of the network's large-scale organization is known to characterize all intracellular relationships documented in metabolic, protein interaction, genetic, and protein domain networks. The Milo *et al*. study now raises the possibility that the complexity pyramid might not be specific only to cells. Indeed, scale-free connectivity with embedded hierarchical modularity has been documented for a wide range of nonbiological networks. Motifs are now known to be abundant in networks as different as ecosystems and the World Wide Web.

These results highlight some of the challenges systems biology will face in the coming years. Lately, we have come to appreciate the power of maps—reliable depositories of molecular interactions. Yet existing maps are woefully incomplete; key links between different organizational levels are missing. For example, we lack the systematic tools to map out lipid-protein or metabolite–transcription factor interactions in vivo. The topological relationships among pathways, motifs, modules, and the full network will also have to be studied in much more detail. Most important, maps must be complemented with detailed measurements of cellular dynamics, recording the timing of processes that take place along the links. This topic is increasingly studied within isolated motifs and modules (*10*) but has received relatively scant attention at the whole-network level. Despite all of these recent challenges, an initial framework offering a rough roadmap appears to have been established. As we seek further insights, we increasingly understand that our quest to capture the system-level laws governing cell biology in fact represents a search for the deeper patterns common to complex systems and networks in general. Therefore, cell biologists, engineers, physicists, mathematicians, and neuroscientists will need to equally contribute to this fantastic voyage.

References

1. H. Kitano, *Science* **295**, 1662 (2002).
2. T. I. Lee *et al.*, *Science* **298**, 799 (2002).
3. R. Milo *et al.*, *Science* **298**, 824 (2002).
4. D. Bray, *Nature* **376**, 307 (1995).
5. U. S. Bhalla, R. Iyengar, *Science* **283**, 381 (1999).
6. L. H. Hartwell *et al.*, *Nature* **402**, C47 (1999).
7. H. Jeong *et al.*, *Nature* **407**, 651 (2000).
8. E. Ravasz *et al.*, *Science* **297**, 1551 (2002).
9. S. S. Shen-Orr *et al.*, *Nature Genet.* **31**, 64 (2002).
10. J. Hasty *et al.*, *Nature Rev. Genet.* **2**, 268 (2001).

PERSPECTIVES: ARCHAEOLOGY

Climate and Human Migrations

Tom D. Dillehay

Archaeological records are affected by a variety of natural and cultural processes at a variety of spatial and temporal scales (*1*). A given cultural phenomenon may appear across a broad range of environments, or may be limited to a narrow range of environments and time periods. Paleoecological studies can help to discriminate between these cases. But most reconstructions of early human ecosystems are based on the excavation and interpretation of individual archaeological sites. Paleoecological studies of long-term climatic change are also often limited in scope (*2*).

Integrative studies of multiple sites, multiple records, and larger areas over long time periods can dramatically change the interpretation (*3*–*7*). On page 821of this issue, Núñez *et al*. (*8*) demonstrate the power of such a comprehensive approach. They closely integrate paleoecological and archaeological analysis to study the long-term interaction between hunter-gatherers and changing environments over the last 15,000 years in the Atacama desert of northern Chile.

The authors examine why initial human occupation occurred about 2000 years later in this hyperarid region than in more humid forested regions in south central Chile (*9*), and several centuries later than in less arid areas in the central and southern Andes. They also ask why a long "Silencio Arqueologico" (a cultural hiatus in the archaeological record) took place between 9500 and 4500 calendar years before the present (cal yr B.P.).

The possible reasons for these variations in human presence considered by Núñez *et al.* include migration lags, inhospitable late Pleistocene environments, biased survey and visibility, and rapid and long-term abandonment of the region. The study illustrates the importance of integrating local environmental and archaeological information in studying regional human ecosystems and in comparing the findings with other regions at a larger scale.

The authors assume that high-altitude ancient lakes (paleolakes), mid-altitude grasslands (puna), and low-altitude wetlands best indicate changes in habitat ex-

The author is in the Department of Anthropology, University of Kentucky, Lexington, KY 40506, USA. E-mail: dilleha@uky.edu