

CHAPTER e19

Systems Biology in Health and Disease

Joseph Loscalzo

The field of human biology has progressed over the last three centuries, largely as a result of the reductionist approach to the scientific problems that challenge the discipline. Biologists study the experimental response of a variable of interest in a cell or organism while holding all other variables constant. In this way, it is possible to dissect the individual components of a biologic system and assume that a thorough understanding of a specific component (e.g., an enzyme or a transcription factor) will provide sufficient insight to explain the global behavior of that system (e.g., a metabolic pathway or a gene network, respectively). Biologic systems are, however, much more complex and manifest behaviors that frequently (if not invariably) cannot be predicted from knowledge of their component parts characterized in isolation. Growing recognition of this shortcoming of conventional biologic research has led to the development of a new discipline, *systems biology*, that is defined as the holistic study of living organisms or their cellular or molecular network components to predict precisely their response to perturbations. Concepts of systems biology can be applied readily to human disease and therapy and define the field of *systems pathobiology*, in which genetic or environmental perturbations produce disease and drug perturbations restore normal system behavior.

Systems biology evolved from the field of systems engineering, in which a linked collection of component parts constitute a network whose output the engineer wishes to predict. The simple example of an electronic circuit can be used to illustrate some basic systems engineering concepts. All the individual elements of the circuit—resistors, capacitors, transistors—have well-defined properties that can be characterized precisely. However, they can be linked (wired or configured) in a variety of ways, each of which yields a circuit whose response to voltage applied across it is different from the response of every other configuration. To predict the circuit's (i.e., system's) behavior, the engineer must study its response to perturbation (e.g., voltage applied across it) holistically rather than its individual components' responses to that perturbation. Viewed another way, the resulting behavior of the system is greater than (or different from) the simple sum of its parts, and systems engineering utilizes rigorous mathematical approaches to predict these complex, often nonlinear, responses. By analogy to biologic systems, one can reason that detailed knowledge of a single enzyme in a metabolic pathway or of a single transcription factor in a gene network will not provide sufficient detail to predict the output of that metabolic pathway or transcriptional network, respectively. Only a systems-based approach will suffice.

It has taken biologists a long time to appreciate the importance of systems approaches to biomedical problems. Reductionism has reigned supreme for many decades, largely because it is experimentally and analytically simpler than holism, and because it has provided insights into biologic mechanisms and disease pathogenesis, and has led to successful therapies. However, reductionism cannot solve all biomedical problems. For example, the so-called off-target effects of new drugs that frequently limit their approval likely reflect

the failure of a drug to be studied in holistic context, i.e., the failure to explore all possible actions aside from the principal target action for which it was developed. Other approaches to understanding biology therefore are clearly needed. With the growing body of genomic, proteomic, and metabolomic data sets in which dynamic changes in the expression of many genes and many metabolites are recorded after a perturbation and with the growth of rigorous mathematical approaches to analyzing those changes, the stage has been set for applying systems engineering principles to modern biology.

Physiologists historically have had more of a (bio)engineering perspective on the conduct of their studies and have been among the first systems biologists. Yet, with few exceptions, they, too, have focused on comparatively simple physiologic systems that are tractable using conventional reductionist approaches. Efforts at integrative modeling of human physiologic systems, as first attempted by Guyton for blood pressure regulation, represent one application of systems engineering to human biology. These dynamic physiologic models often focus on the acute response of a measurable physiologic parameter to a system perturbation, and do so from a classic analytic perspective in which all the conventional physiologic determinants of the output parameter are known and can be modeled quantitatively.

Until recently, molecular systems analysis has been limited owing to inadequate knowledge of the molecular determinants of a biologic system of interest. Although biochemists have approached metabolic pathways from a systems perspective for over 50 years, their efforts have been limited by the inadequacy of key information for each enzyme (K_M , k_{cat} , and concentration) and substrate (concentration) in the pathway. With increasingly rich molecular data sets available for systems-based analyses, including genomic, transcriptomic, proteomic, and metabolomic data, biochemists are now poised to use systems biology approaches to explore biologic and pathobiologic phenomena.

■ PROPERTIES OF COMPLEX BIOLOGIC SYSTEMS

To understand how best to apply the principles of systems biology to human biomedicine, it is necessary to review briefly the building blocks of any biologic system and the determinants of system complexity. All systems can be analyzed by defining their static topology (architecture) and their dynamic (i.e., time-dependent) response to perturbation. In the discussion that follows, system properties are described that derive from the consequences of topology (form) for dynamic response (function). Any system of interacting elements can be represented schematically as a network in which the individual elements are depicted as nodes and their connections are depicted as links. The nature of the links among nodes reflects the degree of complexity of the system. *Simple systems* are those in which the nodes are linearly linked with occasional feedback or feedforward loops modulating system throughput in highly predictable ways. By contrast, *complex systems* are nodes that are linked in more complicated, nonlinear networks; the behavior of these systems by definition is inherently more difficult to predict owing to the nature of the interacting links, the dependence of the system's behavior on its initial conditions, and the inability to measure the overall state of the system at any specific time with great precision. Complex systems can be depicted as a network of lower-complexity interacting components or modules, each of which can be reduced further to simpler analyzable canonical motifs (such as feedback and feedforward loops, or negative and positive autoregulation); however, a central property of complex systems is that simplifying their structures by identifying and characterizing the simpler substructures does not yield a predictable understanding of a system's

behavior. Thus, the functioning system is greater than (or different from) the sum of its individual, tractable parts.

Defined in this way, most biologic systems are complex systems whose behaviors are not readily predictable from simple reductionist principles. The nodes, for example, can be metabolites that are linked by the enzymes that cause their transformations, transcription factors that are linked by the genes whose expression they influence, or proteins in an interaction network that are linked by cofactors that facilitate interactions or by thermodynamic forces that facilitate their biochemical association. Biologic systems typically are organized as *scale-free*, rather than stochastic, networks of nodes. Scale-free systems are those in which a few nodes have many links to other nodes (highly linked nodes, or hubs) but most nodes have only a few links (weakly linked nodes). The term *scale-free* refers to the fact that the distribution of nodes in the network is not influenced by the magnitude or scale of the links considered. This is quite different from two other common network architectures: random (Poisson) and exponential distributions. Scale-free networks can be mathematically described by a power law that defines the probability of the number of links per node [$P(k) = k^{-\gamma}$, where k is the number of links per node and γ is the slope of the $\log P(k)$ versus $\log(k)$ plot]; this unique property of most biologic networks is a reflection of their self-similarity or fractal nature (Fig. e19-1).

There are unique properties of scale-free biologic systems that reflect their evolution and promote their adaptability and survival. Biologic networks likely evolved one node at a time in a process in which new nodes are more likely to link to a highly connected node than to a sparsely connected node. Furthermore, scale-free networks can become sparsely linked to one another, yielding more complex, *modular scale-free topologies*. This evolutionary growth of biologic networks has three important properties that affect system function and survival. First, this scale-free addition of new nodes promotes *system redundancy*, which minimizes the consequences

of errors and accommodates adverse perturbations to the system robustly with minimal effects on critical functions (unless the highly connected nodes are the focus of the perturbation). Second, this resulting network redundancy provides a survival advantage to the system. In complex gene networks, for example, mutations or polymorphisms in weakly linked genes account for biodiversity and biologic variability without disrupting the critical functions of the system; only mutations in highly linked (*essential*) genes (hubs) can shut down the system and cause embryonic lethality. Third, scale-free biologic systems facilitate the flow of information (e.g., metabolite flux) across the system compared with randomly organized biologic systems; this so-called “small-world” property of the system (in which the clustered nature of the highly linked hubs defines a local neighborhood within the network that communicates through weaker, less frequent links to other clusters) minimizes the energy cost for the dynamic action of the system (e.g., minimizes the transition time between states in a metabolic network).

These basic organizing principles of complex biologic systems lead to three unique properties that require emphasis. First, biologic systems are *robust*, which means that they are quite stable in response to most changes in external conditions or internal modification. Second, a corollary to the property of robustness is that complex biologic systems are *sloppy*, which means that they are insensitive to changes in external conditions or internal modification except under certain uncommon conditions (i.e., when a hub is involved in the change). Third, complex biologic systems exhibit *emergent properties*, which means that they manifest behaviors that cannot be predicted from the reductionist principles used to characterize their component parts. Examples of emergent behavior in biologic systems include spontaneous, self-sustained oscillations in glycolysis; spiral and scroll waves of depolarization in cardiac tissue that cause reentrant arrhythmias; and self-organizing patterns in biochemical systems governed by diffusion and chemical reaction.

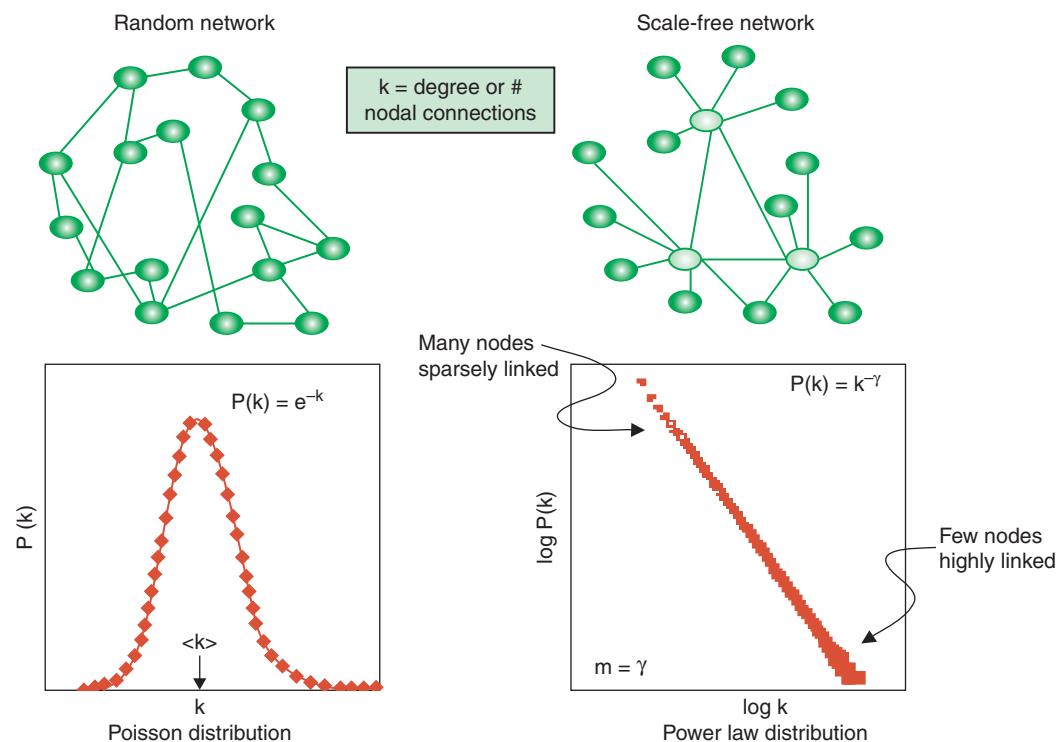


Figure e19-1 Network representations and their distributions. A random network is depicted on the left, and its Poisson distribution of the number of nodal connections (k) is shown in the graph below it. A scale-free

network is depicted on the right, and its power law distribution of the number of nodal connections (k) is shown in the graph below it. Highly connected nodes (hubs) are lightly shaded.

■ APPLICATIONS OF SYSTEMS BIOLOGY TO PATHOBIOLOGY

The principles of systems biology have been applied to complex pathologic processes with some early successes. The key to these applications is the identification of emergent properties of the system under study in order to define novel, otherwise unpredictable (i.e., from the reductionist perspective) methods for regulating the system's response. Systems biology approaches have been used to characterize epidemics and ways to control them, taking advantage of the scale-free properties of the network of infected individuals that constitute the epidemic. Through the use of a systems analysis of a neural protein-protein interaction network, unique disease-modifying proteins have been identified that are common to a wide range of cerebellar neurodegenerative disorders that cause inherited ataxias. Systems biology models have been used to dissect the dynamics of the inflammatory response using oscillatory changes in the transcription factor NF- κ B as the system output. Systems biology principles also have been used to predict the development of an idiotype-anti-idiotype antibody network, describe the dynamics of species growth in microbial biofilms, and analyze the innate immune response. In each of these examples, a systems (patho)biology approach provided insights into the behavior of these complex systems that could not have been recognized with conventional scientific reductionism.

A unique application of systems biology to biomedicine is in the area of drug development. Conventional drug development involves identifying a potential target protein and then designing or screening compounds to identify those that inhibit the function of that target. This reductionist analysis has identified many potential drug targets and drugs, yet only when a drug is tested in animal models or humans are the systems consequences of the drug's action apparent; not uncommonly, so-called off-target effects may become apparent and be sufficiently adverse for researchers to cease development of the agent. A good example of this problem is the unexpected outcomes of the vitamin B–based regimens for lowering homocysteine levels. In these trials, plasma homocysteine levels were reduced effectively; however, there was no effect of this reduction on clinical vascular endpoints. One explanation for this outcome is that one of the B vitamins in the regimen, folate, has a panoply of effects on cell proliferation and metabolism that probably offset its homocysteine-lowering benefits, promoting progressive atherosclerotic plaque growth and its consequences for clinical events. In addition to these types of unexpected outcomes exerted through pathways that were not considered *ab initio*, conventional approaches to drug development typically do not take into consideration the possibility of emergent behaviors of the organism or the metabolic pathway or the transcriptional network of interest. Thus, a systems-based analysis of potential drugs (drug-target network analysis) can benefit the development paradigm both by enhancing the likelihood that a compound of interest will not manifest unforeseen adverse effects and by promoting novel analytic methods for identifying unique control points in metabolic or genetic networks that would benefit from drug-based modulation.

■ SYSTEMS PATHOBIOLOGY AND HUMAN DISEASE CLASSIFICATION

Perhaps most important, systems pathobiology can be used to revise and refine the definition of human disease. The classification of human disease used in this and all medical textbooks derives from the correlation between pathologic analysis and clinical syndromes that began in the nineteenth century. Although this approach has been very successful, serving as the basis for the development of many effective therapies, it has major shortcomings. Those shortcomings include a lack of sensitivity in defining preclinical disease, a primary focus on overtly manifest disease, failure to recognize

different and potentially differentiable causes of common late-stage pathophenotypes, and a limited ability to incorporate the growing body of molecular and genetic determinants of pathophenotype into the conventional classification scheme.

Two examples will illustrate the weakness of simple correlation analyses grounded in the reductionist principle of simplification (Occam's razor) in defining human disease. Sickle cell anemia, the "classic" Mendelian disorder, is caused by a Val6Gln substitution in the β chain of hemoglobin. If conventional genetic teaching holds, this single mutation should lead to a single phenotype in patients who harbor it (genotype-phenotype correlation). This assumption is, however, false, as patients with sickle cell disease manifest a variety of pathophenotypes, including hemolytic anemia, stroke, acute chest syndrome, boney infarction, and painful crisis, as well as an overtly normal phenotype. The reasons for these different phenotypic presentations include the presence of disease-modifying genes or gene products (e.g., hemoglobin F, hemoglobin C, glucose-6-phosphate dehydrogenase), exposure to adverse environmental factors (e.g., hypoxia, dehydration), and the genetic and environmental determinants of common intermediate pathophenotypes (i.e., variations in those generic pathologic mechanisms underlying all human disease—*inflammation, thrombosis/hemorrhage, fibrosis, cell proliferation, apoptosis/necrosis, immune response*).

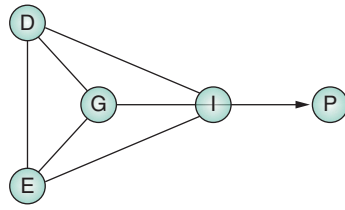
A second example of note is familial pulmonary arterial hypertension. This disorder is associated with 50 different mutations in three members of the transforming growth factor β (TGF- β) superfamily: bone morphogenetic protein receptor-2 (BMPR-2), activin receptor-like kinase-1 (Alk-1), and endoglin. All these different genotypes are associated with a common pathophenotype, and each leads to that pathophenotype by molecular mechanisms that range from haploinsufficiency to dominant negative effects. As only approximately one-fourth of individuals in families that harbor these mutations manifest the pathophenotype, other disease-modifying genes (e.g., the serotonin receptor 5-HT_{2B}, the serotonin transporter 5-HTT), genomic and environmental determinants of common intermediate pathophenotypes, and environmental exposures [e.g., hypoxia, infective agents (HIV), anorexigens] probably account for the incomplete penetrance of the disorder.

On the basis of these and many other related examples, one can approach human disease from a systems pathobiology perspective in which each "disease" can be depicted as a network that includes the following modules: the primary disease-determining elements of the genome (or proteome, if posttranslationally modified), the disease-modifying elements of the genome or proteome, environmental determinants, and genomic and environmental determinants of the generic intermediate pathophenotypes. [Figure e19-2](#) graphically depicts these genotype-phenotype relationships for the six common disease types with specific examples for each type. [Figure e19-3](#) shows a network-based depiction of sickle cell disease using this kind of modular approach.

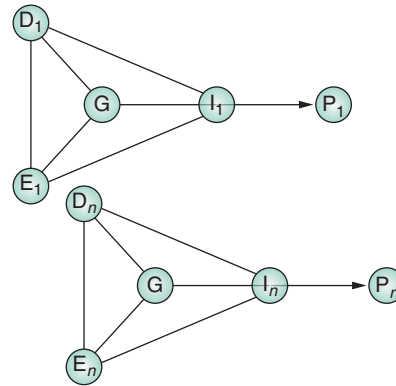
Goh and colleagues developed the concept of a human disease network ([Fig. e19-4](#)) in which they used a systems approach to characterize the disease-gene associations listed in the Online Mendelian Inheritance in Man database. Their analysis showed that genes linked to similar disorders are more likely to have products that associate and greater similarity between their transcription profiles than do genes not associated with similar disorders. In addition, proteins associated with the same pathophenotype are significantly more likely to interact with one another than with other proteins not associated with the pathophenotype. Finally, these authors showed that the great majority of disease-associated genes are not highly connected genes (i.e., not hubs) and are typically weakly linked nodes within the functional periphery of the network in which they operate.

**Classic mendelian disorder:
Single phenotype**

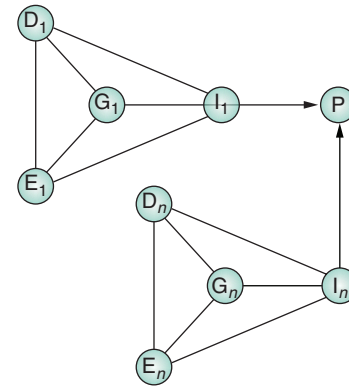
Example:?

**Classic mendelian disorder:
Multiple phenotypes**

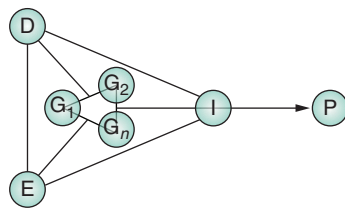
Example: Sickle cell disease

**Classic mendelian disorder:
Multiple mutations, single phenotype**

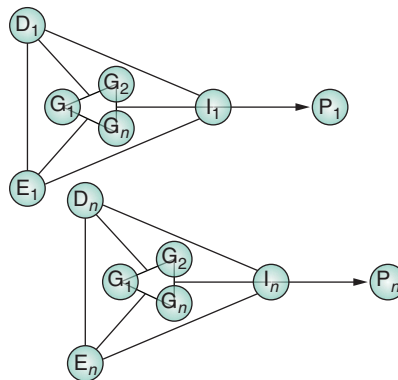
Example: Hypertrophic cardiomyopathy

**Polygenic disorder:
Single phenotype**

Example: Essential hypertension

**Polygenic disorder:
Multiple phenotypes**

Example: Ischemic heart disease

**Environmental disorder**

Example: Subacute bacterial endocarditis

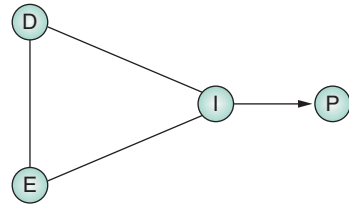


Figure e19-2 Examples of modular representations of human disease. G, primary human disease genome or proteome; D, secondary human disease

genome or proteome; E, environmental determinants; I, intermediate phenotype; P, pathophenotype. (Reproduced with permission from Loscalzo et al.)

This type of analysis validates the potential importance of defining disease on the basis of its systems pathobiologic determinants. Clearly, doing this will require a more careful dissection of the molecular elements in the relevant pathways (i.e., more precise molecular pathophenotyping), less reliance on overt manifestations of disease for their classification, and an understanding of the dynamics (not just the static architecture) of the pathobiologic networks that underlie pathophenotypes defined in this way.

As yet another potential consideration, one can argue that disease reflects the later-stage consequences of the predilection of an organ system to manifest a particular intermediate pathophenotype in response to injury. This paradigm reflects a reverse causality view in which a disease is defined as a tendency to heightened inflammation, thrombosis, or fibrosis after an injurious perturbation. Where the process is manifest (i.e., the organ in which it occurs) is less important than that it occurs (with the exception of the organ-specific pathophysiologic consequences that may require acute attention). For example, from this perspective, acute myocardial infarction (AMI) and its consequences are a reflection of thrombosis (in the coronary artery), inflammation (in the acutely injured myocardium), and fibrosis (at the site or sites of cardiomyocyte death). In effect, the major therapies for AMI address these intermediate pathophenotypes (e.g., antithrombotics, statins) rather than any organ-specific disease-determining process. This paradigm would argue for a systems-based analysis that would first identify the intermediate pathophenotypes to which a person

is predisposed, then determine how and when to intervene to attenuate that adverse predisposition, and finally limit the likelihood that a major organ-specific event will occur. Evidence for the validity of this approach is found in the work of Rzhetsky and colleagues, who reviewed 1.5 million patient records and 161 diseases and found that these disease phenotypes form a network of strong pairwise correlations. This result is consistent with the notion that underlying genetic predispositions to intermediate pathophenotypes form the predicate basis for conventionally defined end organ diseases.

Regardless of the specific nature of the systems pathobiologic approach used, these analyses will lead to a drastic revision of the way human disease is defined and treated. This will be a lengthy and complicated process but ultimately will lead to better disease prevention and therapy and probably do so from an increasingly personalized perspective. The analysis of pathobiology from a systems-based perspective is likely to help define specific subsets of patients more likely to respond to particular interventions based on shared disease mechanisms. This approach is being applied to certain conditions, for example, the responsiveness of lung cancer patients with mutations in the epidermal growth factor receptor (EGFR) to erlotinib, an agent that targets EGFR. Although it is unlikely that the extreme of “individualized medicine” will ever be practical (or even desirable), complex diseases can be mechanistically subclassified and interventions may be tailored to those settings in which they are more likely to work.

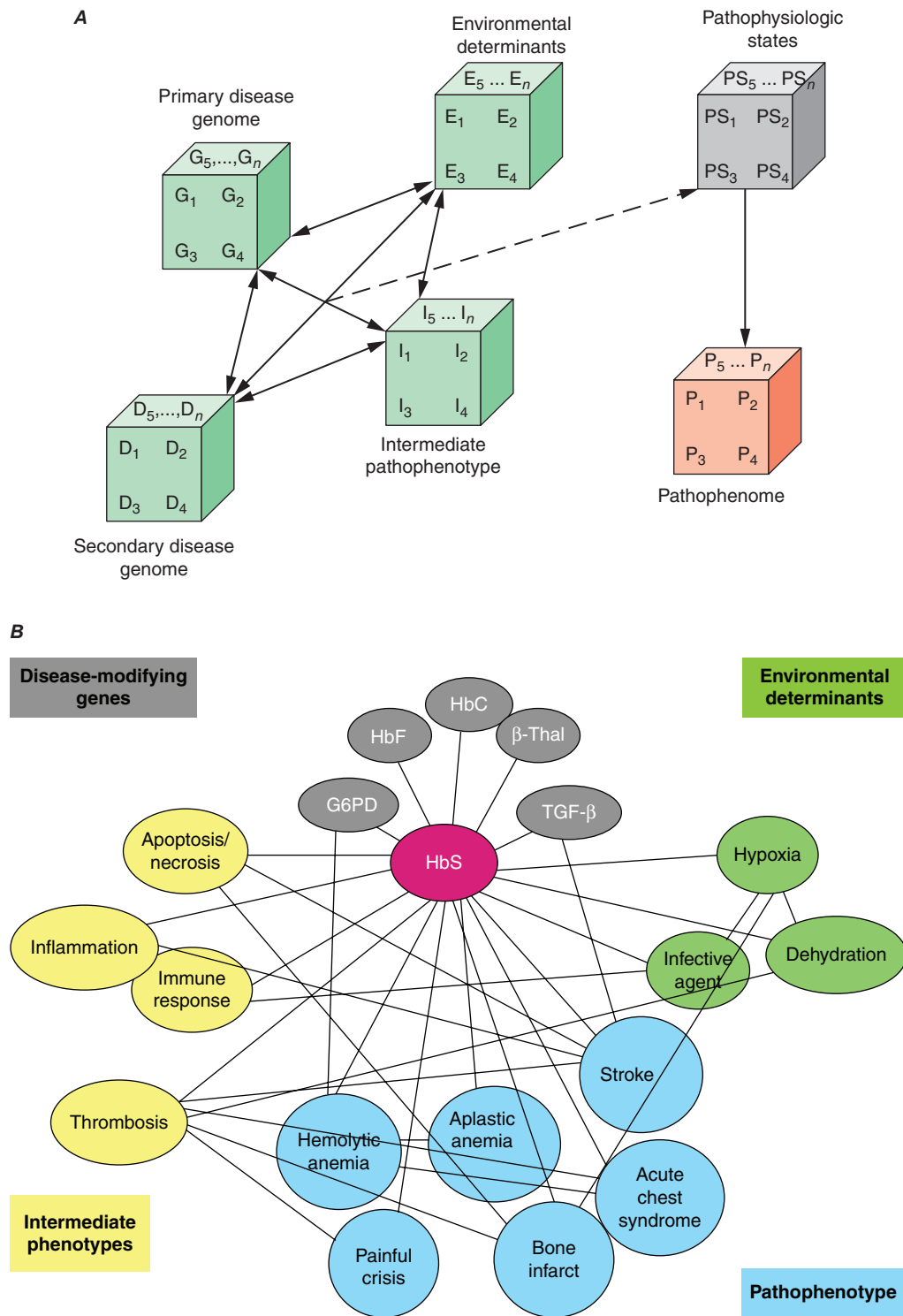


Figure e19-3 **A**. Theoretical human disease network illustrating the relationships among genetic and environmental determinants of the pathophenotypes. Key: G, primary disease genome or proteome; D, secondary disease genome or proteome; I, intermediate phenotype; E, environmental determinants; PS, pathophysiological states leading to

P, pathophenotype. **B**. Example of this theoretical construct applied to sickle cell disease. Key: Red, primary molecular abnormality; gray, disease-modifying genes; yellow, intermediate phenotypes; green, environmental determinants; blue, pathophenotypes. (Reproduced with permission from Loscalzo et al.)

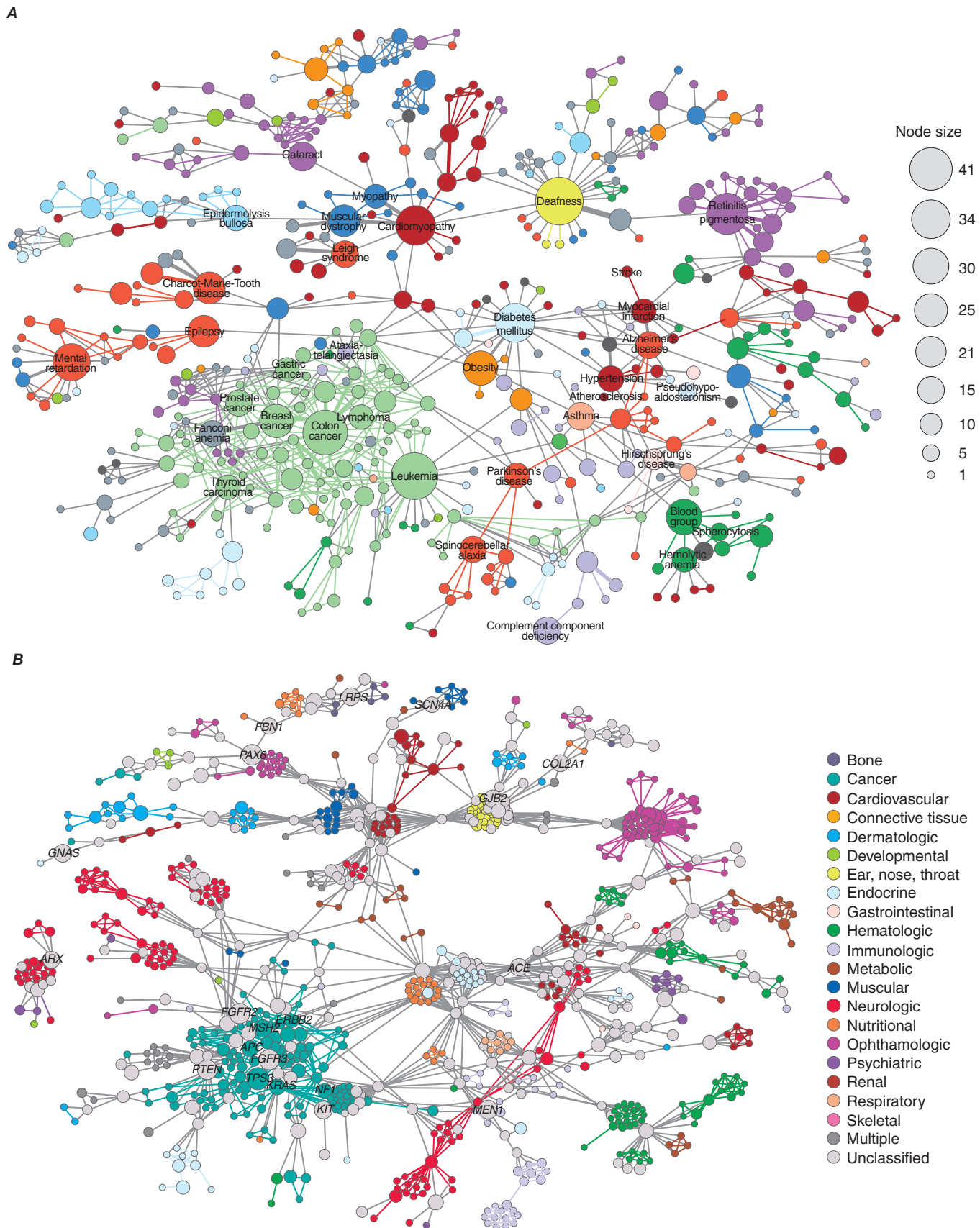


Figure e19-4 A. Human disease network. Each node corresponds to a specific disorder colored by class (22 classes, shown in the key to **B**). The size of each node is proportional to the number of genes contributing to the disorder. Edges between disorders in the same disorder class are colored with the same (lighter) color, and edges connecting different disorder classes are colored gray, with the thickness of the edge proportional to the number

of genes shared by the disorders connected by it. **B. Disease gene network.** Each node is a single gene, and any two genes are connected if implicated in the same disorder. In this network map, the size of each node is proportional to the number of specific disorders in which the gene is implicated. (From Goh et al. Reproduced with permission from the National Academies Press.)

FURTHER READINGS

- BARABÁSI AL et al: Network medicine: A Network-based approach to human disease. *Net Rev Genet* 12:56, 2011
- BRYNILDSEN MP, COLLINS JJ: Systems biology makes it personal. *Mol Cell* 34:137, 2009
- BURGGREN WW, MONTICINO MG: Assessing physiological complexity. *J Exp Biol* 208:3221, 2005
- DANIELS BC et al: Slowness, robustness, and evolvability in systems biology. *Curr Opin Biotechnol* 19:389, 2008
- DE BOER RJ, PERELSON AS: Size and connectivity as emergent properties of a developing immune network. *J Theor Biol* 149:381, 1991
- DEZSO Z, BARABASI AL: Halting viruses in scale-free networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 65:055103 (e-published), 2002
- GOH KI et al: The human disease network. *Proc Natl Acad Sci USA* 104:8685, 2007
- GUYTON AC et al: Systems analysis of arterial pressure regulation and hypertension. *Ann Biomed Eng* 1: 254, 1972
- HART E et al: How affinity influences tolerance in an idiotypic network. *J Theor Biol* 249:422, 2007
- HE X, ZHANG J: Why do hubs tend to be essential in protein networks? *PLoS Genet* 2:e88, 2006
- LIM J et al: A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 126:801, 2006
- LOSCALZO J: Homocysteine trials—clear outcomes for complex reasons. *N Engl J Med* 354:1629, 2006
- et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol Syst Biol* 3:1, 2007
- PASTOR-SATORRAS R, VESPIGNANI A: Epidemic spreading in scale-free networks. *Phys Rev Lett* 86:3200, 2001
- RZHETSKY A et al: Probing molecular overlap among complex human phenotypes. *Proc Natl Acad Sci USA* 104:11694, 2007
- SCHADT EE et al: A network view of disease and compound screening. *Nature* 8:286, 2009
- WEISS JN et al: Network perspectives of cardiovascular metabolism. *J Lipid Res* 47:2355, 2006
- XAVIER JB, FOSTER KR: Cooperation and conflict in microbial biofilms. *Proc Natl Acad Sci USA* 104:876, 2007
- YILDIRIM MA et al: Drug-target network. *Nat Biotechnol* 25:1119, 2007
- YUE H et al: Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: A case study of an NK-kB signaling pathway. *Mol Biosyst* 2:640, 2006
- ZAK DE, ADEREM A: Systems biology of innate immunity. *Immunol Rev* 227:264, 2009