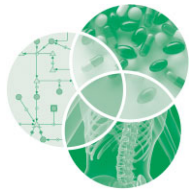
Opinion



Systems biology and the future of medicine

Joseph Loscalzo[1]* and Albert-Laszlo Barabasi[1,2]

Contemporary views of human disease are based on simple correlation between clinical syndromes and pathological analysis dating from the late 19th century. Although this approach to disease diagnosis, prognosis, and treatment has served the medical establishment and society well for many years, it has serious shortcomings for the modern era of the genomic medicine that stem from its reliance on reductionist principles of experimentation and analysis. Quantitative, holistic systems biology applied to human disease offers a unique approach for diagnosing established disease, defining disease predilection, and developing individualized (personalized) treatment strategies that can take full advantage of modern molecular pathobiology and the comprehensive data sets that are rapidly becoming available for populations and individuals. In this way, systems pathobiology offers the promise of redefining our approach to disease and the field of medicine. *WIREs Syst Biol Med* 2011 3 619–627 DOI: 10.1002/wsbm.144

INTRODUCTION

The translation of new knowledge about mechanisms that govern human pathobiology into effective preventive, diagnostic, and therapeutic strategies is a slow and cumbersome process. A major contributor to this translational delay is the use of the traditional characterization and definition of human disease, which dates to the 19th century and is largely based on Oslerian clinicopathological correlation.[1] The Oslerian formalism for human disease links clinical presentation with pathological findings. As a result, disease is defined on the basis of the principal organ system in which symptoms and signs are manifest, and in which gross anatomic pathology and histopathology are correlated. This approach has held sway for over a century, and although there has been continual refinement of the pathological markers used for correlation (e.g., biochemical measurements, immunohistochemistry, flow cytometry, and, more recently, molecular pathological analyses of expressed genes), the general principles remain the same as when the approach was first proposed. Current classification of disease phenotype (pathophenotype) is, then, the result of inductive generalization from clinicopathological evidence predicated on the law of reductive parsimony. This paradigm has been helpful to clinicians as it establishes syndromic patterns that limit the number of potential pathophenotypes they may need to consider. Although quite useful in an earlier era, classifying disease in this way vastly overgeneralizes pathophenotypes, does not usually take into consideration susceptibility states or preclinical disease manifestations, and cannot be used to individualize disease diagnosis or therapy.

MODERN SHORTCOMINGS OF THE OSLERIAN APPROACH TO DISEASE

Based on this history, it is hardly surprising that these conventional pathophenotypes are far too limited to be useful in the postgenomic era. A simple example illustrates this shortcoming. The classic Mendelian disorder, sickle cell disease, is caused by a single point mutation at position 6 of the β-chain of hemoglobin, which changes hemoglobin's oxygen affinity and promotes polymerization under hypoxic conditions. Notwithstanding Mendelian predictions to the contrary, this simple biochemical phenotype and its corresponding monogenotype do not yield a single

*Correspondence to: jloscalzo@partners.org

[1]Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA

[2]Center for Complex Networks Research and Department of Physics, Northeastern University, Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, MA, USA

DOI: 10.1002/wsbm.144

pathophenotype: individuals with sickle cell disease can present with painful crisis, osteonecrosis, acute chest syndrome, stroke, profound anemia, or mild anemia. There are many reasons for these different clinical pathophenotypes, ranging from the presence of disease modifying genes (e.g., hemoglobin F) to environmental influences (e.g., hypoxia).[2] Clearly, even the simplest genetically determined disease is manifestly complex in its expression, a fundamental observation that emphasizes the importance of the genomic and environmental contexts within which disease evolves.

Although conventional reductionist pathophenotyping has guided steady progress in diagnostics and therapeutics for many years, it is fraught with shortcomings, some of which are highlighted by this example, that are particularly problematic for contemporary molecular and genomic analyses. Put another way, in using this sorely outdated approach to defining human disease, we construct nosological silos that focus exclusively on end-stage pathological processes in a single organ largely driven by late-appearing, generic end-stage mechanisms rather than true disease-specific susceptibility determinants viewed in their holistic, systems-based complexity.

With this background, one can rationally catalog the limitations of traditional disease definition as follows:

1. Disease is typically defined by late-appearing manifestations in a dysfunctional organ system, without regard for or knowledge of preclinical pathophenotype or susceptibility factors that precede overt abnormalities. Thus, the focus is not on the specific genetic or environmental susceptibility determinants of the disease phenotype, but, rather, on the late-appearing, intermediate pathophenotypes (generic endopathophenotypes, including inflammation, immunity, fibrosis, thrombosis, hemorrhage, cell proliferation, apoptosis, and necrosis) within a given organ system. As a result, typical therapeutic strategies do not focus on truly unique, targeted disease determinants, but on these same intermediate pathophenotypes (e.g., anti-inflammatory or antithrombotic therapies for acute myocardial infarction).
2. Conventional disease paradigms generally neglect underlying pathobiological mechanisms that may extend beyond the disease-defining organ system, and do not typically consider the molecular (deterministic) and environmental (stochastic) factors that govern disease evolution from susceptibility state to preclinical pathophenotype to overt pathophenotype.
3. Conventional definitions of disease are excessively inclusive of the range of pathophenotypes and are based on the pathophysiological characterizations largely of the premolecular era. These inclusive definitions of disease not only obscure subtle, but potentially important, differences among individuals with common clinical presentations, but also neglect underlying disease mechanisms that cross organ systems and may yield more appropriate and specific therapeutic targets.
4. Yet another dimension to this problem stems from the reductionist approach we use to identify disease mechanisms or therapeutic targets. Disease is rarely (if ever) a simple consequence of an abnormality in a single effector gene product, but, rather, is a reflection of pathobiological processes (deterministic and stochastic) that interact in a complex network to yield pathophenotype, which may be viewed as an emergent property (i.e., discernible only by appreciating the behavior of the network as a whole rather than of its component parts in reductionist isolation) of a pathobiological system.

These shortcomings of conventional disease definition account for many limitations of major recent genome-based efforts to define disease determinants (e.g., the weak effect size of linked alleles observed in genome-wide association studies of complex disease) and to design rational therapies (e.g., the failure of >90% of drug candidates[3]). Thus, solving this problem is not simply an exercise in nosology, but is essential for moving the entire health care enterprise forward to reduce the burden of human disease and suffering.

This background highlights the clear need to reconsider and redefine the determinants of human disease. We begin by stating the obvious: all disease is complex, even simple Mendelian disorders. Pathophenotype reflects the action of a deterministic, defective molecular network within a stochastic environmental context that modulates network function. Defined in this way, disease is the result of the output of a complex modular network of -omic and environmental nodes linked mechanistically to yield pathophenotype.[4] With this background and rationale, we propose a redefinition of all human disease using a combination of approaches to identify systems-based pathobiological mechanisms that render one susceptible to preclinical and overt pathophenotypes. This approach challenges the existing disease paradigm directly, and is justifiable owing to the

largely heuristic strategies that have been used to identify disease mechanisms and treatments to date.

SYSTEMS PATHOBIOLOGY AND NETWORK MEDICINE

A contemporary approach to human disease requires that it be viewed from a systems perspective. In this context, systems pathobiology is defined as the science of integrating genetic, genomic, biochemical, cellular, physiological, and clinical data to create a network that can be used to model predictively disease expression (and response to therapy). In order to understand best disease expression, one needs not only to define the architecture or topology of the disease network (or disease module, see below) within the context of the universe of molecular networks in a cell or organism, but also to explore its dynamic response to perturbations. The characteristically nonlinear responses of these complex systems underlie their emergent properties, which can only be appreciated when the system is viewed in holistic context. From the perspective of disease, the clinicopathological correlations of Oslerian medicine likely reflect these emergent properties that otherwise often defy mechanistic elucidation. In this way, Oslerian approaches can inform systems-based holism as these clnicopathological correlations often give insight into the emergent properties of a disease network. In this way, one begins to define the discipline of network medicine.[4–7]

Knowledge of two broad categories of interrelated networks within a cell or organism is essential for understanding the determinants of disease expression; these are *molecular networks* and *phenotypic networks*. Molecular networks include protein interaction networks,[8,9] metabolic networks,[10–12] and regulatory networks, including transcription factor networks[13] and noncanonical RNA networks.[14,15] Phenotypic networks include coexpression networks in which genes are linked when they manifest similar expression patterns in two different diseases,[16] and genetic networks in which genes are linked that together define a phenotype which is distinct from that defined by either gene alone.[17,18]

Organizing Principles of Biological Networks

Proceeding from a topological description of these networks to an appreciation of their role in defining human disease requires recognition of a few important organizing principles derived from network theory.[19–21] In brief, any network can be viewed as a collection of linked nodes, the distribution of which can range from random to highly clustered. Biological networks are not random collections of nodes and links, but evolve as clustered collections of genes, regulatory RNAs, proteins, or metabolites. Biological and pathobiological networks are *scale-free*; contain few highly connected nodes (*hubs*)[22] and *bottlenecks* (nodes that link different highly connected clusters to each other, gaining, as a result, high 'betweenness centrality')[23]; manifest the small-world effect[24] and *disassortativity* (highly connected nodes, or hubs, typically avoid linking to one another)[25]; and contain motifs with predictable functional consequences (feedback loops, oscillators, etc.).[26] All of the biological networks relevant to disease manifest these properties, as well, which gives us a starting point from which to begin to identify those subnetworks or modules that are responsible for a specific pathobiological process or a specific disease.

Disease Modules and Their Identification

At the molecular level, reductionist approaches to disease have assumed that abnormalities in any gene, protein, or regulatory RNA molecule could be responsible for a disease; however, at the current time, only ~10% of human genes are known to be associated with a disease.[27] Armed with this knowledge, one can intuitively surmise that mutations in hub genes or proteins are more likely to yield disease than those in less connected, peripheral genes or proteins in the network. However, from a systems perspective, only genes or proteins that are peripherally located in the molecular network are likely to account for complex disease in adults owing to the fact that hubs are more likely to correspond to essential genes, loss of the function of which can lead to embryonic lethality.[4,28] Although not typically hubs, disease genes and proteins do cluster in the same network neighborhood, as shown by Goh and colleagues who reported a 10-fold increase in the number of physical interactions observed between gene products associated with the same disease than would be expected by chance.[28] In addition, genes linked to diseases with similar pathophenotypes have a higher likelihood of interacting with each other than those not linked to the pathophenotype.[29,30] Taken together, these observations support the notion that disease-related components of a network are likely to comprise a subnetwork or *disease module*. A disease module is defined as a group of network components that contribute to a cellular or organismic phenotype the disruption of which leads to a particular pathophenotype. Most precisely, a disease module represents a subnetwork in the overall molecular network that reflects a unique set of interactions, either proximate

or remote, that contribute to an abnormal phenotype when one or more of its components is(are) dysfunctional. Importantly, a specific gene, protein, or metabolite can participate in several disease modules, indicating that disease modules themselves can overlap within the global network, and consistent with the clustering of disease determinants described above.[28]

Disease modules can be identified either using bioinformatics or experimental methodologies.[31] Bioinformatics-based approaches comprise two broad categories that exploit knowledge of relevant molecular networks or of functional and structural similarities among elements of the global network. The molecular network-based approach for identifying disease modules begins with constructing the global 'human interactome,' a network that includes all known (macro)molecular interactions in human cells or tissues, regardless of cell or tissue type, respectively. In subsequent steps, the network is systematically reduced using genetic and biochemical data to identify the functional modules (i.e., subnetworks) involved in the pathophenotype(s) of interest. Bioinformatics sources for this network's construction include literature-curated and systematic high-throughput human protein–protein interaction datasets,[8,9,32–34] literature-curated and predicted human protein–DNA interaction datasets,[35,36] and human metabolic pathways[37,38] leading to potential metabolic coupling,[39] which, together, define the 'host interactome.' The search for a disease module within the network is then based on two complementary hypotheses:

1. disease modules are frequently associated with common, highly interconnected local groups of nodes that can be identified by network clustering algorithms; and
2. the nodes of a disease module correspond to cellular components of similar or closely related functions associated within a specific neighborhood of the network.[39] (Note that the same node can belong to multiple disease modules, and a series of new clustering algorithms[40,41] can be used to identify systematically such overlapping modules.)

The disease modules could contain dozens of interconnected nodes, only some of which may be truly relevant to the pathophenotype. In order to identify the most relevant pathways within each module, algorithms that incorporate flow- (or diffusion-) based pathway prioritization are used that assign to each path within the module a quantitative value reflecting its topological and functional proximity to the validated disease components. Prior work indicates that the highest ranked pathways not only show a statistically significant enrichment in disease-related components, but also are of potential relevance to the underlying *causative* mechanisms of the disease.[42]

Owing to the current limited knowledge of the human interactome, the molecular network-based approach may fail to identify genes whose interactions with the genes involved in the disease modules have not yet been validated. For this reason, an alternative approach is used that can identify systematically additional potential disease genes, enhancing the procedure for finding the 'seed' genes of the disease module. This functional and structural similarity search-based approach once again begins with the list of genes known to be involved in the pathophenotype of interest. In subsequent steps, genes that show sufficient functional, structural, or contextual similarity to 'seed' genes are considered candidate genes that may be relevant to the pathophenotype. An iterative process next ensues in which the new candidates are functionally characterized and validated until available evidence fails to support further expansion of the disease module. Accordingly, the functional and structural similarity search-based approach proceeds according to the following algorithm. The initial seed components identified above are clustered into nascent modules using publically available bioinformatics resources as in the molecular network-based strategy, *viz.*, protein–protein interactions, gene regulatory relationships (including miRNAs), and metabolic pathways, using high-throughput and literature-curated data, as well as precompiled information from Gene Ontology (GO),[43] Kyoto Encyclopedia of Genes and Genomes (KEGG),[37] and literature-based annotations in Genome Recognition Analysis Internet Link (GRAIL)[44]; the nascent modules can also be delineated with tissue-specific expression microarray, miRNA,[45] and metabolomic data. The nascent disease module is next expanded by deriving a set of candidate genes with which to augment the disease module using information in the bioinformatics resources, relying especially on GO, KEGG, and GRAIL patterns of (tissue-specific) coexpression, and protein–protein interaction datasets.

These bioinformatics approaches or variants thereof have been used to identify disease modules for a wide range of disorders, including Alzheimer disease,[46] Parkinson disease,[47] type 2 diabetes mellitus,[48] asthma,[49] cardiovascular diseases,[50] and a variety of malignancies.[51,52] What often limits the successful mapping of a disease module, however, is the limited coverage of available cellular maps from

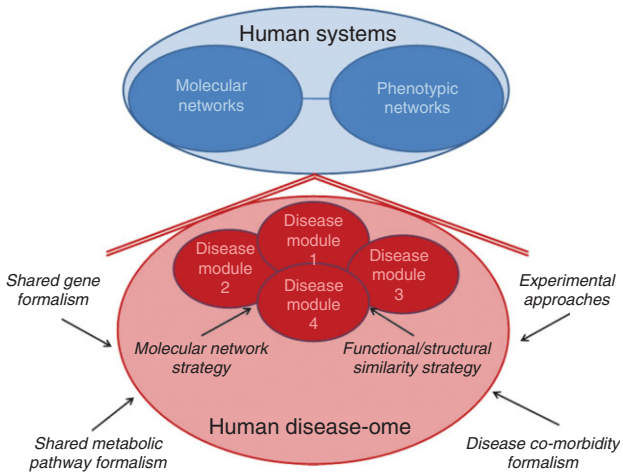


FIGURE 1 | The human systems biology universe. Human systems comprise molecular and phenotypic networks, which are related to, but distinct from, each other, as indicated by the separate linked ovals. The human disease-ome represents a collection of subnetworks, the disease modules, which are identified by one of two strategies, the molecular network-based strategy or the functional and structural similarity-based strategy. The assembly of disease modules into the disease-ome can be determined by bioinformatics-based approaches—the shared gene formalism, the shared metabolic pathway formalism, or the disease comorbidity formalism—or by laboratory-based experimentation.

bioinformatics resources. Under these circumstances, brute-force experimentation designed to identify the universe of interaction partners is required and has also been used successfully for several diseases, including spinocerebellar ataxia,[53] Huntington disease,[54] and schizophrenia.[55]

The Disease-ome and Its Identification

A natural extension of developing a systems-based approach to the identification of disease modules is the growing recognition that these modules are often interdependent (Figure 1). Systematic mapping of overlapping disease modules and their pathophenotypes leads to the construction of the disease-ome, or a network of disease nodes linked to one another by their common molecular underpinnings. There are (at least) three different representations of the disease-ome that have gained traction of late, each of which reflects shared mechanisms or shared (intermediate) pathophenotypes between the incorporated diseases. The first, the *shared gene formalism*, recognizes that diseases which share a gene or genes likely have a common genetic basis. This concept has been developed by Goh et al.,[28] who utilized the Online Mendelian Inheritance in Man (OMIM) database to build a disease network the links within which reflect genes shared between diseases. In this network, they found 867 of 1284 diseases with an associated gene linked to at least one other disease, with 516 of them belonging to a single disease cluster. The second, the *shared metabolic pathway formalism*, recognizes that enzymatic defects that affect the flux of a proximate reaction in a metabolic pathway may affect downstream fluxes in the same pathway, leading to pathophenotypes that are known to be associated with the downstream reactions. The corollary to this formalism is that connections that reflect shared metabolic pathways are more likely to be relevant to expression of metabolic diseases than are connections based on shared genes (not in the same metabolic pathway). Using this formalism, metabolic disease network maps can be constructed in which two diseases are linked if the enzymes associated with them catalyze adjacent (sequential) reactions.[56] The third, the *disease comorbidity formalism*, links diseases based on their co-occurrence in excess of the play of chance, leading to the construction of phenotypic disease network maps.[57] In a recent example of this approach, Rzhetsky et al.[58] constructed a phenotypic disease network incorporating 657 diseases from 1.5 million Medicare patients in which two diseases are linked if their comorbidity exceeds a predefined threshold. Importantly, the phenotypic disease network does not depend on molecular or genetic mechanism, nor on environmental perturbations. Alternative approaches to developing disease networks have been proposed recently, including the identification of topological modules in the human protein interaction network and their linkage through expression data to diseases in which they are either up- or down-regulated[59]; the linkage of genetic determinants and environmental exposures to specific diseases, thereby implicating environmental perturbations of gene function in disease pathogenesis[60]; and the associations of diseases in a disease network if they have a miRNA or miRNA cluster in common.[45]

OTHER APPLICATIONS OF SYSTEMS BIOLOGY TO MEDICINE

Other areas of medicine in which systems approaches will likely prove useful include drug development, behavioral influences on disease propensity, and metagenomics. Until very recently, drug development has persisted in its reductionist quest for the Ehrlichian 'magic bullet' for each disease target, and as molecular modeling and pharmacological target identification have become more refined, the quest has become ever more targeted. Furthermore, so-called off-target effects, which often lead to a drug's withdrawal from

the market if unexpected and severe, are a reflection of the failure to consider any pharmacological agent in holistic context, perturbing a molecular network, not just a single specific target. These limitations to conventional drug development have, no doubt, contributed to the limited number of annually approved new drugs despite ever more powerful molecular approaches to their identification. Network-based systems approaches have begun to correct this serious limitation and will clearly redirect the industry's efforts; these include developing drug target networks in order to seek commonalities of targets,[61] exploring side-effect similarities among approved drugs,[62] optimizing the consequences of perturbation of metabolic networks,[63] and using Bayesian approaches to identify optimal therapeutic strategies.[64] The effect of social networks on behavior and its consequences for disease expression adds yet another dimension to systems approaches to human disease, as demonstrated recently for obesity.[65] Lastly, the complexity of the microbiome, its interactions with the human host genome, and the pathobiological consequences of its perturbation is an area rife for exploitation using systems approaches.[66–68]

CONCLUSIONS

The range of possible applications of systems pathobiology to medicine is vast, yet, is in its very early stages. The advantages of using an holistic, network-based approach to characterize human disease will, at last, begin to move medicine from a field of simple associations rooted in semi-empiric reductionism in search of the 'cure' for each disease to one that recognizes the power of the molecular networks upon which human biology is based as a highly rational paradigm by which to identify disease cures. The emergent behavior of these networks dictates that reductionist approaches will, by definition, fail to ascertain the complexity implicit in these scale-free systems, and will, therefore, fail to appreciate commonalities among diseases, unique treatment approaches that will likely require combinations of therapies, and the many molecular consequences that environmental or pharmacological perturbations can evoke. Although there are clear examples of the successful application of systems principles to medicine reviewed in this article, the breadth of the success of this approach has yet to be realized but will, no doubt, revolutionize the science and practice of medicine.

ACKNOWLEDGMENTS

This work was supported in part by NIH grants HL61795, HL81587, HL70819, and HL48743. The author wishes to thank Stephanie Tribuna for expert technical assistance.

REFERENCES

1. Osler W. *The Principles and Practice of Medicine*. New York: Appleton; 1892.
2. Sebastiani P, Ramoni MF, Nolan V, Baldwin CT, Steinberg MH. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat Genet* 2005, 37: 435–440. doi:10.1038/ng1533.
3. Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Revs Drug Discov* 2009, 8:959–968. doi:10/1038.nrd2961.
4. Loscalzo J, Kohane I, Barabasi A-L. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol* 2007, 3:124. e-published. doi:10/1038/msb100163.
5. Zanzoni A, Soler-Lopez M, Aloy P. A network medicine approach to human disease. *FEBS Lett* 2009, 583: 1759–1765. doi:10.1016/febslet.2009/03.001.
6. Pawson T, Linding R. Network medicine. *FEBS Lett* 2008, 582:1266–1270.
7. Barabasi AL. Network medicine—from obesity to the diseasome. *N Engl J Med* 2007, 357:404–407. doi:10.1056/NEJMe078114.
8. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Druze M, Ayivi-Guedehoussou N, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005, 437:1173–1178. doi:10.1038/Naure04209.
9. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005, 122:957–968. doi:10.1016/j.cell.2005.08.029.
10. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. *Nature* 2000, 407:651–654. doi:10.1038/35036627.
11. Fell DA, Wagner A. The small world of metabolism. *Nat Biotechnol* 2000, 18:1121–1122. doi:10.1038/81025.

12. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 2007, 104:1777–1782. doi:10.1073/pnas.010772104.

13. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al. The transcriptional landscape of the mammalian genome. *Science* 2005, 309:1559–1563. doi:10.1126/science.1112014.

14. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, 120:15–20. doi:10.16.j.cell.2004.12.035.

15. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. *Nat Biotechnol* 2004, 22:326–330. doi:10.1038/nbt936.

16. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 2003, 302:249–255. doi:10.1176/science.1087447.

17. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Rev Genet* 2007, 8:437–449. doi:10.1038/nrg2085.

18. Beltrao P, Cagney G, Krogan N. Quantitative genetic interactions reveal biological modularity. *Cell* 2010, 141:739–745. doi:10.1016.j.cell.2010.05.019.

19. Barabasi AL, Oltvai Z. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004, 5:101–113. doi:10.1038/nrg1272.

20. Albert R, Barabasi AL. Statistical mechanics of complex networks. *Rev Mod Phys* 2002, 74:47–97. doi:10.1103/RevModPhys.74.47.

21. Caldarelli G. *Scale Free Networks*. Oxford: Oxford University Press; 2007.

22. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science* 1999, 286:509–512. doi:10.1126/science.286.5439.509.

23. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 2007, 3:e59. doi:10.137/journal.pcbi.0030059.

24. Watts DJ, Strogatz SH. Collective dynamics in 'small-world' networks. *Nature* 1998, 393:440–442. doi:10.1038/30918.

25. Pastor Satorras R, Vazquez A, Vespignani A. Dynamical and correlation properties of the internet. *Phys Rev Lett* 2001, 87:258701. e-published. doi:10.1013/PhysRevL.87.258701.

26. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. *Science* 2002, 298:824–827. doi:10.1126/science.298.5594.824.

27. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's online Mendelian inheritance in man (OMIM). *Nucleic Acids Res* 2009, 37:D793–D796. doi:01.1039/nar/qkn665.

28. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U S A* 2007, 104:8685–8690. doi:10.1073/pnas.0701361104.

29. Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN. Analysis of the human protein interactome and comparison with yeast, worm, and fly interaction datasets. *Nat Genet* 2006, 38:285–293. doi:10.1038/ng1747.

30. Xu J, Li Y. Discovering disease-genes by topological features in human protein-protein interactions networks. *Bioinformatics* 2006, 22:2800–2805. doi:10.1093/bioinformatics/btl467.

31. Barabási AL, Guhlbace N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature Rev Genet* 2010, 12:56–68. doi:10.1038nrg2918.

32. Vankatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, et al. An empirical framework for binary interactome mapping. *Nat Methods* 2009, 6:83–90. doi:10.1038/nmeth.1280.

33. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, et al. The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 2010, 38:D525–D531. doi:10.1093/nar/qkp878.

34. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G. MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res* 2010, 38:D532–D539. doi:10.1093/nar/qkp983.

35. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003, 31:374–378.

36. Mani KM, Lefebvre C, Wang K, Lim WK, Basso K, Dalla-Favera R, Califano A. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol Syst Biol* 2008, 4:169. e-published. doi:10.1038/msb.2008.2.

37. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010, 38:D355–D360. doi:10.1093/nar/qkp896.

38. Schellenberger J, Park JO, Conrad TM, Palsson BO. BiGG: a biochemical genetic and genomic knowledge base of large scale metabolic reconstrjuctions. *BMC Bioinformatics* 2010, 11:213. e-published. doi:10.1186/1471-21-5-11-213.

39. Park J, Lee DS, Christakis NA, Barabasi AL. The impact of cellular networks on disease comorbidity. *Mol Syst Biol* 2009, 5:262. e-published. doi:10.1038/msb.2009.16.
40. Palla G, Derenyi F, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 2005, 435:814–818. doi:10.1038/nature03607.
41. Ahn YY, Bagrow JP, Lehmann S. Link communities reveal multi-scale complexity in networks. *Nature* 2010, 466:761–764. doi:10.1038/nature09182.
42. Naviakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. *Bioinformatics* 2010, 26:1057–1063. doi:10.1093/ bioinformatics/btq076.
43. Harris MA, Clark J, Ireland A, Lomas J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004, 32:D258–D261. doi:10.1093/nar/qkh036.
44. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Guo W, Cui Q. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 2009, 5:e1000534. doi:10.1371/journal.pgen.1000534.
45. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. An analysis of human microRNA and disease associations. *PLoS One* 2008, 3:e3420. doi:10.10307/journal.pone.0003420.
46. Ray M, Ruan J, Zhang W. Variations in the transcriptome of Alzheimer's disease reveal molecular networks involved in cardiovascular disease. *Genome Biol* 2008, 9:R148. doi:10.1186/gb-2008-9-10-r148.
47. Moran LB, Graeber MB. Towards a pathway definition of Parkinson's disease: a complex disorder with links to cancer, diabetes, and inflammation. *Neurogenetics* 2008, 9:1–13. doi:10.1007/s10048-007-0116-v.
48. Liu M, Liberzon A, Kong SW, Wai WR, Park PJ. Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genet* 2007, 3:e96. e-published. doi:10.1037/journal/pgen.0030096.
49. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H. A protein interaction network associated with asthma. *J Theo Biol* 2008, 252:722–731. doi:10.1016/j.jtbi.2008.02.011.
50. Lusis AJ, Weiss JN. Cardiovascular networks: a systems-based approach to cardiovascular disease. *Circulation* 2010, 121:157–170. doi:10.1161/ CIRCULATIONAHA.108.847699.
51. Lee E, Jung H, Radivojac P, Kim JW, Lee D. Analysis of AML genes in dysregulated molecular networks. *BMC Bioinformatics* 2009, 10:S2. doi:10.1186/1471-2105-10-S9-S2.
52. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol* 2007, 3:140. e-published. doi:10.1038/msb4100180.
53. Lim J, Hao T, Shaw C, Patel AJ, Szabo C, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 2006, 125:801–814. doi:10.1016/j.cell.2006.03.032.
54. Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenberg KS, Knoblich M, Haenig C, et al. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell* 2004, 15:853–865. doi:10.1016/j.molcel.2004.09.016.
55. Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S, Bonnert TP, Whiting PJ, Brandon NJ. Disrupted in schizophrenia 1 interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psych* 2007, 12:74–86. doi:10.1038/sj.mp.4001880.
56. Lee DS, Park K, Kay KA, Christakis NA, Oltvai ZN, Barabási AL. The implications of human metabolic network topology for disease comorbidity. *Proc Natl Acad Sci U S A* 2008, 105:9880–9885. doi:10.1073/pnas.0802208105.
57. Hidalgo CA, Blumm N, Barabasi AL, Christakis A. A dynamic network approach for the study of human phenotypes. *PLoS Comp Biol* 2009, 5:1–11. doi:10.137/journal.pcbi.1000353.
58. Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci U S A* 2007, 104:11684–11689. doi:10.1073/pnas.0704820104.
59. Suthram S, Dudley AJ, Chiang P, Chen R, Hastie TJ, Butte AJ. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput Biol* 2010, 6:e1000662. e-published. doi:10.1371/ journal.pcbi.1000662.
60. Liu YI, Wise PH, Butte AJ. The "etiome": identification and clustering of human disease etiological factors. *BMC Bioinformatics* 2009, 10(suppl 2):S14. e-published. doi:10.1186/1427-1-2105-10-S2-S14.
61. Yildrim MA, Goh KI, Cusik ME, Barabasi AL, Vidal M. Drug-target networks. *Nat Biotechnol* 2007, 25: 1119–1126. doi:10.1038/nbt1338.
62. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side effect similarity. *Science* 2008, 321:263–266.
63. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 2002, 99:15112–15117. doi:10.1073/pnas.232349399.
64. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. *Nat Rev Drug Disc* 2009, 8:286–295. doi:10.1038/nrd2826.

65. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med* 2007, 357:370–379.

66. Uetz P, Dong YA, Zeretzke C, Atzler B, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E, et al. Herpesviral protein networks and their interaction with the human proteome. *Science* 2006, 311:239–242. doi:10.1126/science.1116804.

67. Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazques A, Holthaus AM, Ewence AE, Li N, Hirozane-Kishikawa T, Hill DE, et al. Epstein-Barr virus and virus-human protein interaction maps. *Proc Natl Acad Sci U S A* 2007, 104:7606–7011. doi:10.1073/pnas.0702332104.

68. Turnbaugh PJ, Gordon JI. An invitation to the marriage of metagenomics and metabolomics. *Cell* 2008, 134:708–713. doi:10.1016.j.cell.2008.08.025.

