Networking metabolites and diseases

Pascal Braun*†‡, Edward Rietman*§, and Marc Vidal*†

**Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; †Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115; and §Physical Sciences Inc., 20 New England Business Center, Andover, MA 01810*

Biological systems are increasingly viewed and analyzed as highly complex networks of interlinked macromolecules and metabolites. Network analysis has been applied to interactome maps of protein–protein, protein–DNA, and protein–RNA interactions as well as transcriptional, metabolic, and genetic data. Such network views of biological systems should facilitate the detection of nonlinear long-range effects of perturbations, for example, by mutations, and help identification of unanticipated indirect causal connections.

Diseasome and Drug-Target Network

Recently, Goh *et al.* (1) constructed a "diseasome" network in which two diseases are linked to each other if they share at least one gene, in which mutations are associated with both diseases. In the resulting network, related disease families cluster tightly together, thus phenotypically defining functional modules. Importantly, for the first time this study applied concepts from network biology to human diseases, thus opening the door for discovering causal relationships between disregulated networks and resulting ailments.

Subsequently Yilderim *et al.* (2) linked drugs to protein targets in a drug–target network, which could then be overlaid with the diseasome network. One notable finding was the recent trend toward the development of new compounds directly targeted at disease gene products, whereas previous drugs, often found by trial and error, appear to target proteins only indirectly related to the actual disease molecular mechanisms. An important question that remains in this emerging field of network analysis consists of investigating the extent to which directly targeting the product of mutated genes is an efficient approach or whether targeting network properties instead, and thereby accounting for indirect nonlinear effects of system perturbations by drugs, may prove more fruitful. However, to answer such questions it is important to have a good understanding of the various influences that can lead to diseases.

Metabolic Connections

One group of diseases that was very poorly connected in the original diseasome network was the family of metabolic diseases. In this issue of PNAS, Lee *et al.* (3) hypothesize that metabolic diseases may instead be connected via metabolites and common reactions. To investigate this hypothesis Lee *et al.* first constructed a metabolic network from data available in two manually curated databases detailing well known metabolic reactions, the involved metabolites, and catalyzing enzymes. In addition, gene–disease associations were identified by using the Online Mendelian Inheritance in Man (OMIM) database (www.ncbi.nlm.nih.gov/sites/entrez?db=omim&itool=toolbar). In a last step, a metabolic disease network (MDN) was constructed by connecting two diseases if their associated genes are linked in the metabolic network by a common metabolite or metabolites used in a common reaction.

17% of all diseases in the network show significant comorbidity.

Metabolites are not only linked by common reactions, but on a larger scale by coupled fluxes within a metabolic network, which may also influence disease phenotypes. An increase in the concentration of one metabolite may increase several fluxes across reaction pathways that use this compound, which may lead to diverse phenotypes and distinct diseases. The fluxes within the metabolic network are calculated by using the Flux Coupling Finder method described by Nikolaev *et al.* (4) and Burgard *et al.* (5), which is based on the assumption that pools of metabolites are conserved. To functionally validate the network, coexpression correlations are measured for genes linked by adjacent reactions and those linked by fluxes. Interestingly, the average coexpression correlation for flux-coupled genes (0.31) is higher than that for genes simply catalyzing adjacent reactions (0.24) (compared with 0.10 for all gene pairs in the network).

If the links between diseases identified in the MDN are functionally and causally relevant it should be expected that linked diseases occur more frequently in the same individual. To test this hypothesis, Lee *et al.* (3) measured the co-occurrence of diseases in patients by using detailed Medicare information of 13 million patients and 32 million hospital visits within a 3-year period. A comorbidity index was computed to measure the degree to which one disease will increase the likelihood of a second disease in the same patient. The average comorbidity for all genes is 0.0008 (Pearson correlation coefficient), which increases 3-fold to 0.0027 when disease pairs that are metabolically linked are analyzed, which is highly statistically significant ($P < 10^{-8}$). When diseases are analyzed that are directionally coupled by a flux (see ref. 3 for details), the correlation increases to 0.0062. Thus, whereas 17% of all diseases in the network show significant comorbidity, this fraction nearly doubles to 31% for metabolically linked diseases. Further analysis reveals that comorbidity effects can be detected up to three links (metabolites, reactions) apart from each other with statistical significance, but not farther away.

In the MDN, several highly connected hubs, e.g., hypertension and hemolytic anemia, are linked to many different co-occurring diseases not unexpected for such complex diseases that can result from many different genetic alterations or variants. Importantly, though, most of the connections to the different linked diseases are mediated by diverse connections in the metabolic network. Thus, in the future such insights may be helpful for finer classification of the complex hub disease. Furthermore, depending on the onset of the complex (hub) disease in relation to the associated diseases, such relationships may potentially be used to systematically stratify patients and develop targeted treatments acting on the underlying metabolic links.

Returning to the starting point of their study, Lee *et al.* (3) next investigated whether metabolic diseases are better linked through the metabolic network than they are in the previously described gene–disease network. When purely metabolic diseases are consid-

Author contributions: P.B., E.R., and M.V. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 9880.

‡To whom correspondence should be addressed. E-mail: pascal_braun@dfci.harvard.edu.

ered, the comorbidity is, in fact, best predicted by metabolic links. Interestingly, when all diseases linked to metabolic enzymes are considered, which involves many diseases that are merely related to metabolic diseases through multifunctional enzymes, the gene and metabolic networks are nearly equally predictive of comorbidity, indicating that as a general approach information from many different biological dimensions should be integrated to identify the most relevant connections.

Together, all these findings support the initial hypothesis that metabolic diseases are linked by metabolic networks. Practically, alteration of one metabolite or one reaction can have numerous repercussions in the network, each of which can manifest as different diseases that frequently occur together in affected patients.

Perspective

However, Lee *et al.*'s study (3) also well illustrates some of the challenges associated with this type of analysis, which leads to statistically very significant, but still rather small, correlation values of <0.01. One contributing factor is very likely the imperfect information about underlying networks and linkages, which in this case include missing disease–gene associations and incompletely defined metabolic networks. When considering a more global picture, network analysis is restricted by still very incomplete knowledge about, for example, information fluxes in the protein interactome network, which are mediated by protein interactions and enzyme–substrate relationships, and many other network dimensions that are too numerous to list. A second limitation, although necessary for a first analysis of this kind, is the restriction to one dimension of the biological system (metabolic reactions), whereas *in vivo* effects on many different levels act together to yield a given phenotype.

Thus, most importantly, this work by Lee *et al.* (3) defines a program for, and constitutes an important step toward, linking data from diverse areas of systems biology. Data gained by metabolic profiling, mapping of enzyme–substrate and interactome networks, and many other activities need to be combined into a single high-dimensional systems network model, which can then be used to explore network effects of disease causing genetic or environmental alterations. Construction of such models, however, will require much more comprehensive data for nearly all aspects of biology and may even necessitate the development of novel mathematical and statistical tools to deal with them (6). Ultimately, it should be expected that this type of integrated network analysis will profoundly alter our view of biological systems, our understanding of the way mutations lead to disease phenotypes, and how these insights are used in drug discovery. Exciting times lie ahead of us.

1. Goh KI, *et al.* (2007) The human disease network. *Proc Natl Acad Sci USA* 104:8685–8690.
2. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M (2007) Drug-target network. *Nat Biotechnol* 25:1119–1126.
3. Lee DS, *et al.* (2008) The implications of human metabolic network topology for disease comorbidity. *Proc Natl Acad Sci USA* 105:9880–9885.
4. Nikolaev EV, Burgard AP, Maranas CD (2005) Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. *Biophys J* 88:37–49.
5. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale meta-bolic network reconstructions. *Genome Res* 14:301–312.
6. Lespinats S, Verleysen M, Giron A, Fertil B (2007) DD-HDS: A method for visualization and exploration of high-dimensional data. *IEEE Trans Neural Netw* 18:1265–1279.