editorial

A thematic review series: systems biology approaches to metabolic and cardiovascular disorders

Aldons J. Lusis[1]

Department of Medicine/Division of Cardiology, BH-307 Center for the Health Sciences, University of California Los Angeles, Los Angeles, CA 90095-1679

The next series of the *JLR*'s thematic review articles concerns systems-level approaches to cardiovascular and metabolic traits. "Systems-level" means a kind of biologic analysis that looks beyond individual genes or proteins or lipids to the ensemble of multiple elements of a system. Some examples of biologic systems are the transcripts in a cell (a "transcriptome"), the proteins of an organelle (a "proteome"), and the metabolites in a liver (a "metabolome"). Systems-level approaches have been stimulated by the genome project, by the development of experimental techniques that can simultaneously interrogate many elements of a system (such as expression microarrays), and by advances in computational methods for analyzing large data sets (1).

A key concept in systems biology is that of "emergent properties," important features of biologic systems that can best be identified by examining the system as a whole. A simple example of an emergent property is illustrated in **Fig. 1**. In this experiment, the levels of the transcripts in livers of mice from an intercross between two common inbred strains were determined using expression microarrays. The entire data set was then tested to identify transcripts whose levels correlate with one another. One of many such sets of highly correlated transcripts from this experiment is the set of cholesterol biosynthetic enzymes, as shown. In this case, the relationships of these enzymes were already known through the painstaking, decades-long studies of many biochemists. However, even if this information was not known, our global experiment would immediately group all of the enzymes together, implying a functional relationship. Jim Weiss's review in this series will discuss a particularly striking emergent property in metabolism. Thus, both glycolysis and oxidative phosphorylation are capable, under the right conditions, of developing self-sustained oscillations. Such oscillations are observed only when the individual enzymes are coupled into a "network" with other metabolic enzymes to create positive and negative feedback loops. Clearly, in this example, studies of the individual components of the system would not provide mechanistic understanding of the overall dynamics of the system.

A particularly important goal of systems biology is to construct "networks," sets of genes or proteins or metabolites that act in concert in a common biologic process (2). These networks can be experimentally identified by classical methods (e.g., the pathways of cholesterol homeostasis), but systems-level approaches such as expression arrays, chIP-chip studies, and whole-genome yeast two-hybrid experiments can efficiently test millions of possible interactions. Topologically, networks consist of elements (termed "nodes" in network nomenclature) that exhibit functional connections ("edges") (**Fig. 2**). The connections can be identified using coregulation (as in Fig. 1), physical interactions (as in protein complexes), or metabolic relationships (e.g., the intermediates in glycolysis). "Undirected" networks are simply nodes connected by edges, with no causal direction, whereas "directed" networks have edges with a given direction (Fig. 2). An important emergent property observed for most biologic networks is scale-free topology, in which some nodes have many edges (these nodes are termed "hubs") but most nodes have few edges.

A convenient way of organizing the data sets generated using various global platforms is as a series of orthogonal stages, ordered according to the sequential stages of gene expression (genome, transcriptome, proteome) followed by the metabolome (the set of metabolites) and the phenome (the set of physiologic or disease parameters of interest) (**Fig. 3**). Each of these data sets can be used to construct networks or derive other useful information, but none alone tells the whole story. For example, analyses of the transcriptome will miss crucial aspects of protein realization and cellular signaling. Thus, the intersection of these orthogonal data sets is an important challenge. One particularly useful application of such genomic integration is for the identification of genes and pathways con-

Manuscript received 17 July 2006.
DOI 10.1194/jlr.E600004-JLR200

[1] To whom correspondence should be addressed.
e-mail: jlusis@mednet.ucla.edu

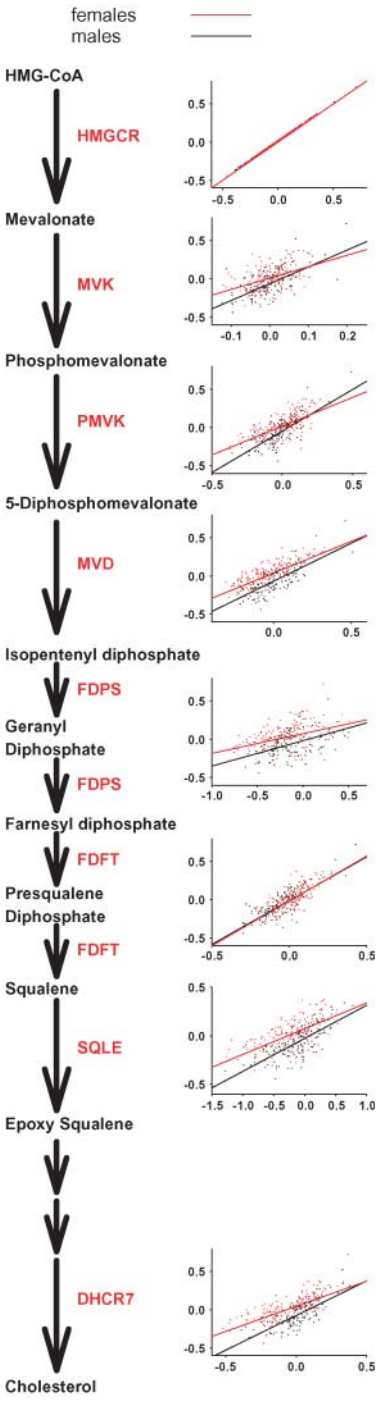

Fig. 1. Transcript levels of the enzymes of the cholesterol biosynthetic pathway are strongly correlated in an intercross between two common strains of mouse. An F2 intercross between strain C3H and C57BL/6J mice was constructed, and livers from the mice were analyzed for transcript levels using microarrays. The levels of the transcripts for each enzyme in the cholesterol biosynthetic pathway (x axis) are plotted against the levels of the transcript for HMG-CoA reductase (y axis). Clearly, transcript levels for all of the enzymes are highly correlated in both female (n = 170; black circles) and male (n = 170; red circles) progeny. This finding illustrates the concept that functionally related genes tend to exhibit coordinated regulation in response to various perturbations, whether genetic (as in this example), environmental, or developmental. These data were generated from a database published by Wang and colleagues (2005), and I thank Susanna Wang for help in preparing the figure (20).

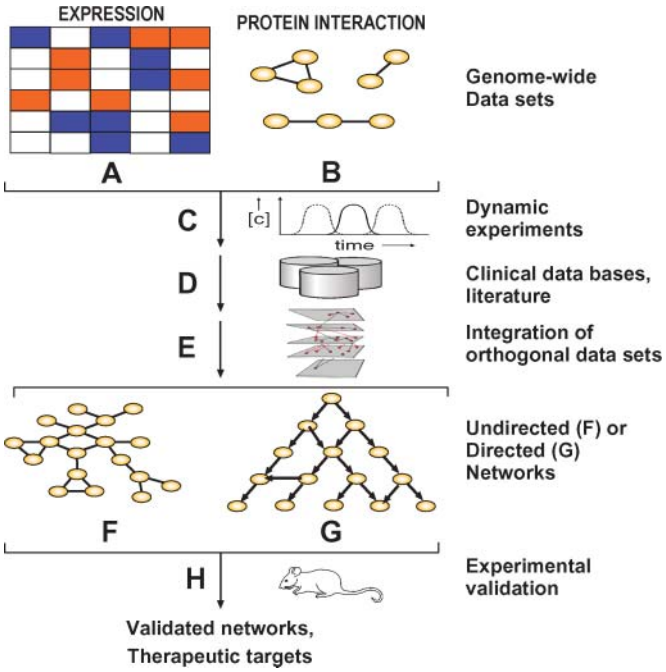

Fig. 2. Integration of genome-wide data sets for the construction of networks and the elucidation of disease mechanisms. Experimental approaches such as gene expression microarrays (A) or global yeast two-hybrid protein interaction experiments (B) can be used to generate data for construction of networks. Such data can be integrated with results from dynamic experiments, clinical data, literature data, or genetic data (C–E). The resulting networks can simply indicate connections, termed "undirected" networks (F), or they can indicate the direction of the interactions, termed "directed" networks (G). Ultimately, such modeling requires experimental validation in transgenic animals or tissue culture (H).

tributing to common diseases. Bayesian inference is a particularly useful statistical approach for the integration of information from heterogeneous data sets, assigning a probability to the predictions rather than only a binary classification (3, 4).

How important will systems-level approaches be? For well over 50 years, since chemists developed the tools for studies of biological molecules, research in biology has focused on individual genes, painstakingly connecting them to other genes and to functions. Almost the entire body of mechanistic understanding of cellular and developmental biology has been built up in this reductionistic way. Not surprisingly, many biologists are skeptical about the new systems biology that seeks a global view of all of the elements and their interactions and their responses to environmental changes. The goal of this series of reviews is to illustrate how systems-based approaches can complement traditional approaches to the biology of lipids and diseases involving lipids. The most elegant systems biology studies to date have been in model organisms, such as bacteria, yeast, *Caenorhabditis elegans*, and flies (5–7). However, a number of studies involving various global data sets for mammals (primarily mouse and human) (**Table 1**) have already contributed importantly to our understanding of metabolic and cardiovascular diseases.

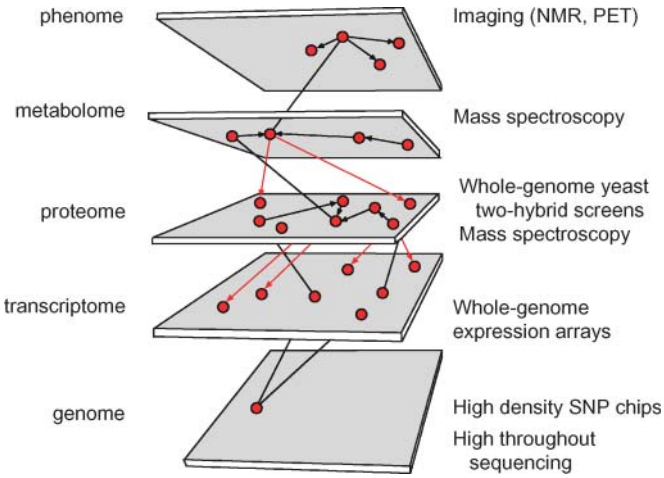

Fig. 3. Genome-wide data sets and online resources and databases used in systems-based approaches. Each data set can be viewed as a stage that is orthogonal to the others. A major challenge of systems biology is to intersect these data sets. The lines between the stages illustrate the effects of a hypothetical genetic perturbation. PET, positron emission tomography; SNP, single nucleotide polymorphism.

A particularly important goal of systems biology is to identify genes involved in diseases. The genes underlying approximately half of the Mendelian disorders listed in McKusick's Online Mendelian Inheritance in Man have now been identified (3), but the search for genes and pathways involved in common complex diseases, such as the metabolic syndrome, diabetes, and cardiovascular diseases, is only beginning. The first review of this series, by Karen Reue and Laurent Vergnes (UCLA), deals with the integration of genomic resources, particularly the genome sequences of human, mouse, and other model organisms, for the identification and functional analysis of genes involved in lipid metabolism. This review outlines the various genomic approaches to this problem, including studies with model organisms, systematic gene modification projects, and comparative genomics. For example, a number of new genes involved in lipid metabolism have recently been identified based on sequence similarities with known genes and their functions determined using mice with gene modifications (8).

In the second article in the series, Andrew Watson (UCLA) will review "lipidomics," a systems-based study of all lipids, the molecules with which they interact, and their functions. Lipidomics can be considered a branch of metabolomics. Lipids, of course, play crucial structural roles in biologic systems, but the appreciation of their key regulatory functions in inflammation and chronic diseases is greatly expanding. For example, certain recently identified oxidation products of phospholipids appear to play a key role in atherosclerosis (9), and one such oxidized lipid was shown to significantly perturb the expression of >1,000 genes in human endothelial cells (10). It is noteworthy that the National Institutes of Health has established a large collaborative effort, the LIPID MAPS consortium, to identify, characterize, and quantitate all the lipids in specific cells. This includes a bioinformatics focus to organize and integrate the data, consisting of rapidly expanding lists of lipids (presently over 8,000), lipid related proteins, and lipid related changes in the transcriptome (11, 12).

As discussed above, a major application of genome-wide data sets will be to model networks. One important characteristic of networks is "topology," the overall architecture of the interactions between nodes. A second is "dynamics," the temporal interactions involved in perturbations of the network or the maintenance of homeostasis. Interaction maps represent possible networks, but not all edges will be present at the same time in a particular cell, just as edges will differ between cell types and subcellular locations. Studies in yeast, for example, have shown that environmental responses involve multiple temporal stages and surprisingly large-scale topological changes (13). An understanding of the dynamic nature of a network will thus

TABLE 1. Some global online data sets and resources for metabolic and cardiovascular disorders

"Omic" Space	Online Resources and Databases
Phenome	Genomics of Lipid-Associated Disorders Database, gold.tugraz.at/main.jsp
	Diabetes Genome Anatomy Project, www.diabetesgenome.org
	Online Mendelian Inheritance in Man, www.ncbi.nlm.nih.gov
	Unified Medical Language System, umlsinfo.nlm.nih.gov
	Mouse Genome Database, www.informatics.jax.org/
	Rat Genome Database, rgd.new.edu/
Metabolome	LIPID Metabolites and Pathway Strategy (Lipid Maps), www.lipidmaps.org/
	Biological magnetic resonance database, www.brmb.wisc.edu/metabolomics/
	Lipid Bank, lipidbank.jp/
	European Nutrigenomics Organization, www.nugo.org/metabolomics/
Proteome	Database of Interacting Proteins, dip.doe-mbi.ucla.edu/
	Biomolecular Interaction Network Database, bind.ca/
	Proteomes of Higher Eukaryotic Organisms, ebi.ac.uk/IPI/IPIhelp.html
Transcriptome	Gene Expression Omnibus, www.ncbi.nlm.nih.gov/geo
	miRNA Registry, www.sanger.ac.uk/software/rfan/mirna/
Genome	UCSC, genome.ucsc.edu/
	Ensembl, www.ensembl.org/
	National Center for Biotechnology Information, www.ncbi.nlm.gov/genomes/
	The SNP Consortium, snp.sshl.org/
	Perlegen Mouse SNP Database, mouse.perlegen.com/mouse/download.html

require comprehensive time-course data sets with tools for quantifying kinetic parameters such as concentrations, interaction strengths, and fluxes (14) (Fig. 2). In the third review, Jim Weiss (UCLA) will review studies of the dynamics of cardiac and smooth muscle metabolism (15). These studies illustrate how systems-level approaches can reveal emergent properties of a system that lead to testable hypotheses. Thus, studies of cardiac metabolism indicate that, in a normal state, energy supply and demand are closely balanced, but under stress (as in postischemic injury), this coordination can be lost, leading to cell-wide oscillations in ATP levels.

There have been a number of successes in the identification of genes underlying common diseases, and there will likely be hundreds of additional genes identified in the next few years. But how will this information be used in the identification of targets and the development of safe, efficient treatments? In the fourth review, Eric Schadt (Merck) will discuss how systems-based approaches can be used to define targets for therapeutic intervention (16). He will first describe how directed networks can be constructed using data from molecular profiling, genotyping, and clinical studies. He will then discuss how such networks can be intersected with orthogonal experimental data, such as data from transgenic mice or small interfering RNA experiments in cultured cells, to identify candidates for drug intervention.

Proteomics involves the systematic analysis of proteins and their interactions. Although the technologies for examining proteins on a genome-wide basis are less powerful than the microarray technologies for RNA expression, they are already providing important complementary information. Peipei Ping and Thomas Drake (UCLA) will review the various levels of proteomic investigation as they apply to metabolic and vascular diseases, including organelle topography, protein-protein interaction networks, protein interactions with DNA and lipids, protein processing, and disease marker discovery. One important application of proteomics is the identification of disease markers (17, 18).

In the final review, Erwin Kurland (State University of New York at Stony Brook) will focus on metabolomics, the systematic profiling of metabolites using nuclear magnetic resonance, tandem mass spectrometry, calorimetry, and stable isotopes. Topics relevant to lipid metabolism, such as energy partitioning and metabolite flux between tissues, will be described. This review will emphasize how metabolomics can help clarify very complex disorders such as diabetes that involve multiple tissues [for example, (19)].

The author expresses his gratitude to the investigators mentioned in this article for participating in this project. They are all leaders in their respective areas of research, and the author is confident that their reviews will generate interest in this promising new approach in biology.

REFERENCES

1. Ideker, T., T. Galitski, and L. Hood. 2001. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2:** 343–372.
2. Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5:** 101–113.
3. Giallourakis, C., C. Henson, M. Reich, X. Xie, and V. K. Mootha. 2005. Disease gene discovery through integrative genomics. *Annu. Rev. Genomics Hum. Genet.* **6:** 381–406.
4. Troyanskaya, O. G., K. Dolinski, A. B. Owen, R. B. Altman, and D. Botstein. 2003. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc. Natl. Acad. Sci. USA.* **100:** 8348–8353.
5. Dolinski, K., and D. Botstein. 2005. Changing perspectives in yeast research nearly a decade after the genome sequence. *Genome Res.* **15:** 1611–1619.
6. Ashrafi, K., F. Y. Chang, J. L. Watts, A. G. Fraser, R. S. Kamath, J. Ahringer, and G. Ruvkun. 2003. Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature.* **421:** 268–272.
7. Tong, A. H., G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, et al. 2004. Global mapping of the yeast genetic interaction network. *Science.* **303:** 808–813.
8. Vergnes, L., A. P. Beigneux, R. Davis, S. M. Watkins, S. G. Young, and K. Reue. 2006. Agpat6 deficiency causes subdermal lipodystrophy and resistance to obesity. *J. Lipid Res.* **47:** 745–754.
9. Berliner, J. A., and A. D. Watson. 2005. A role for oxidized phospholipids in atherosclerosis. *N. Engl. J. Med.* **353:** 9–11.
10. Gargalovic, P. S., M. Imura, B. Zhang, N. M. Gharavi, M. J. Clark, J. Pagnon, W-P. Yang, A. He, A. Truong, S. Patel, et al. 2006. Identification of inflammatory gene modules based on variations in human endothelial cell responses to oxidized lipids. *Proc. Natl. Acad. Sci. USA.* In press.
11. Dennis, E. A., H. A. Brown, R. Deems, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. H. Raetz, W. Shaw, S. Subramaniam, D. W. Russell, et al. 2005. The LIPIDS MAPS approach to lipidomics In Functional Lipidomics. L. Feng and G. Prestwich, editors. CRC Press/Taylor & Francis Group, London. 1–15.
12. Fahy, E., S. Subramaniam , H. A. Brown, C. K. Glass, A. H. Merrill Jr, R. C. Murphy, C. R. Raetz, D. W. Russell, Y. Seyama, W. Shaw, et al. 2005. A comprehensive classification system for lipids. *J. Lipid Res.* **46:** 839–861.
13. Luscombe, N. M., M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann, and M. Gerstein. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature.* **431:** 308–312.
14. Albert, R. 2005. Scale-free networks in cell biology. *J. Cell Sci.* **118:** 4947–4957.
15. Weiss, J. N., Z. Qu, P. S. Chen, S. F. Lin, H. S. Karagueuzian, H. Hayashi, A. Garfinkel, and A. Karma. 2005. The dynamics of cardiac fibrillation. *Circulation.* **112:** 1232–1240.
16. Schadt, E. E. 2005. Exploiting naturally occurring DNA variation and molecular profiling data to dissect disease and drug response traits. *Curr. Opin. Biotechnol.* **16:** 647–654.
17. Ping, P., T. M. Vondriska, C. J. Creighton, T. K. Gandhi, Z. Yang, R. Menon, M. S. Kwon, S. Y. Cho, G. Drwal, M. Kellmann, et al. 2005. A functional annotation of subproteomes in human plasma. *Proteomics.* **5:** 3506–3519.
18. Mayr, M., J. Zhang, A. S. Greene, D. D. Gutterman, J. K. Perloff, and P. Ping. Proteomic based development of biomarkers in cardiovascular disease: mechanistic, clinical, and therapeutic insights. *Mol. Cell. Proteomics.* Epub ahead of print. May 29, 2006; doi:10.1074/mcp.R600007-MCP200.
19. Xu, J., V. Chang, S. B. Joseph, C. Trujillo, S. Bassilian, M. F. Saad, W. N. Lee, and I. J. Kurland. 2004. Peroxisomal proliferator-activated receptor alpha deficiency diminishes insulin-responsiveness of gluconeogenic/glycolytic/pentose gene expression and substrate cycle flux. *Endocrinology.* **145:** 1087–1095.
20. Wang, S., N. Yehya, E. E. Schadt, H. Wang, T. A. Drake, and A. J. Lusis. 2006. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet.* **2:** e15.