**Medical Genomics**

# Primer on Medical Genomics
# Part XIV: Introduction to Systems Biology—A New Approach to Understanding Disease and Treatment

NICOLE M. MOREL, PHD, MBA; JOANNE M. HOLLAND, PHD; JAN VAN DER GREEF, PHD;
EDWARD W. MARPLE, MBA; CLARY CLISH, PHD; JOSEPH LOSCALZO, MD, PHD; AND STEPHEN NAYLOR, PHD

The advent of the "-omics revolution" has forced us to reevaluate our ability to acquire, measure, and handle large data sets. Omic platforms such as expression arrays and mass spectrometry, with their exquisite selectivity, sensitivity, and specificity, are unrivaled technologies for detection, quantitation, and identification of DNA, messenger RNA, proteins, and metabolites derived from complex body tissue and fluids. More recently, attempts have been made to capture the utility of these platform technologies and combine them under the umbrella of systems biology, also referred to as pathway, network, or integrative biology. Applied systems biology is the integrated analysis of genetic, genomic, protein, metabolite, cellular, and pathway events that are in flux and interdependent. It necessitates the use of a variety of analytic platforms as well as biostatistics, bioinformatics, data integration, computational biology, modeling, and knowledge assembly protocols. Such sophisticated analyses may provide new insight into the understanding of disease processes and mechanisms of action of pharmaceutical agents. Ultimately, this requires a perspective on how complex systems behave and are modulated. In this regard, systems biology, more appropriately considered as a process containing a series of modules, aims to provide tools and capabilities to carry out such tasks. We describe the essentials required to carry out systems biology experiments, the method in which integrated data in the form of a systems biology correlation network affords new insight into understanding disease, and the vista of developing more efficient biomarkers and therapeutic agents.

*Mayo Clin Proc.* 2004;79:651-658

L-FABP = liver fatty acid–binding protein; mRNA = messenger RNA

Health care spending in the United States alone exceeds $1.4 trillion per year. Nonetheless, for most disease states, a fundamental understanding of causal disease onset, disease mechanism and progression, and optimal treatment is still somewhat limited. In part, this frustrating progress has been encumbered by our inability to fully and rapidly delineate complex cellular metabolic processes and molecular pathways. However, the advent of the "-omics revolution" has afforded tools and technologies that may ultimately help in elucidating such pathways, networks, and processes.[1] It has also necessitated the development of a new "-omics language." Genomics provides us with platforms to measure quantitatively the essential elements (genes) of the cell and includes haplotyping and single nucleotide polymorphism detection.

Transcriptomics affords information about the expression of individual genes at the messenger RNA (mRNA) level. Proteomics focuses on determination of individual protein concentrations present in the biological system being investigated, whereas functional proteomics determines constituent protein-protein, protein-DNA, and protein-RNA interactions and their resulting complexes. Finally, metabolomics (also referred to as metabonomics) involves the characterization and quantitation of small organic molecules in either circulatory or cell-tissue systems. The nomenclature of the –omics revolution and the overlay of systems biology are summarized in Figure 1.

## IMPACT OF -OMICS

The quest to obtain a complete understanding of a disease, thereby providing the medical community with more definitive means of diagnosis and treatment, has monopolized medical research for decades. It has consumed considerable time, energy, and effort, as well as fiscal and scientific resources. However, for the most part, understanding causal relationships in diseases is still poor. There is a paucity of specific and sensitive prognostic, diagnostic, or surrogate biomarkers and even fewer defined therapeutic drug targets.
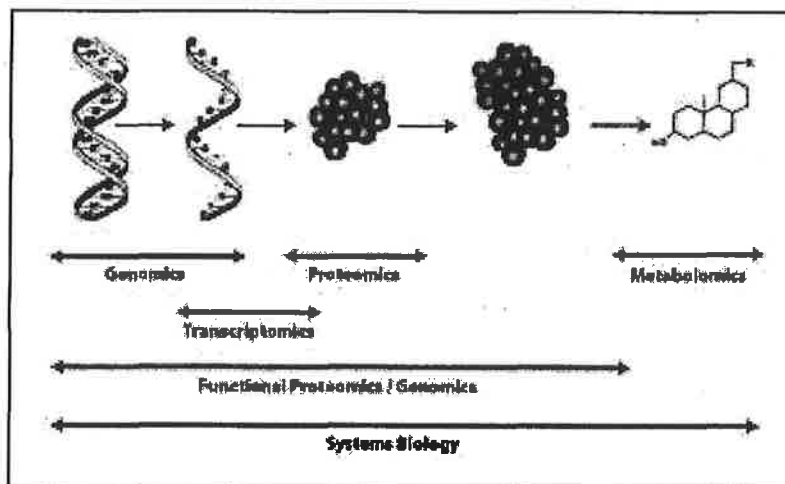
Figure 1. The nomenclature of the "-omics revolution" and the overlay of systems biology.

Watson and Crick's discovery of the double helical structure of DNA in 1953 led to subsequent studies that focused primarily on the ability to sequence DNA rapidly. On the 50th anniversary of this unique work, it is clear that this area of study has expanded considerably.[2] Indeed, this ultimately resulted in the sequence of the human genome that was reported recently.[3,4] This latter event promises to provide scientists with a new roadmap for understanding disease, as well as yield novel biomarkers, and to produce improved treatments. Although this accomplishment provides a set of blueprints for all the expressed products of the genome, ie, individual proteins and noncoding regulatory sequences, it does not explain how they are assembled or how they function and interact with each other in cells. Furthermore, the unexpectedly small number of approximately 30,000 human genes reported is incapable of expressing the necessary number of protein products needed to effectively carry out all cellular functions and intercellular interactions that comprise the complexity of human biology.[5] Indeed, the fact that the number of new genes discovered was well below that predicted for humans (based on comparison with evolutionarily lower organisms) indicates that humans have increased their complexity primarily by using the same building blocks in more diverse ways, eg, splice variants or posttranslational modifications. Without a more comprehensive understanding of these phenomena, we are still left without the tools for deciphering causal mechanisms of disease.

In the postgenomic era, attention has shifted and advanced to examining the products encoded by the human genome, namely proteins. Uncovering the function and expression pattern of proteins will undoubtedly put us in a better position to draw correlations to disease.[6] However, to approach the study of proteomics in a manner similar to that of the genomics effort is entirely unrealistic from the perspective of routine analysis. In the case of genome function, information is stored primarily in the linear sequence of the DNA. However, protein function depends on considerably more complex criteria, including primary, secondary, tertiary, and quaternary structure. Localized and global 3-dimensional structures, a consequence of protein folding, are inherent properties that ultimately determine protein function. Further adding to this complexity is that, rather than the 4 building blocks used by the genome, the proteome has at least 20 individual component amino acids. Finally, proteins are subject to a wide variety of, often temporally regulated, posttranslational modifications, such as phosphorylation, glycosylation, prenylation, acylation, and ubiquitination. Each of these modifications influences protein function, and many are not readily predicted by genetic sequence information alone. Generating a complete catalog of this type of information and correlating it systematically to disease state are indeed daunting tasks.

An additional complicating factor is the emergence of metabolomics and the ability to characterize metabolite profiles rapidly. The metabolome is the complete set of metabolites that can be found in an organism in all commonly observed physiologically relevant conditions.[7] Obviously, depending on genetic makeup and individual health status, only subsets of the metabolome may be present at any one time; hence, it is sensitive to temporal and perturbation events. Furthermore, the structural complexity of the metabolome is important because the number of subclasses of molecules, eg, fatty acids, amino acids,
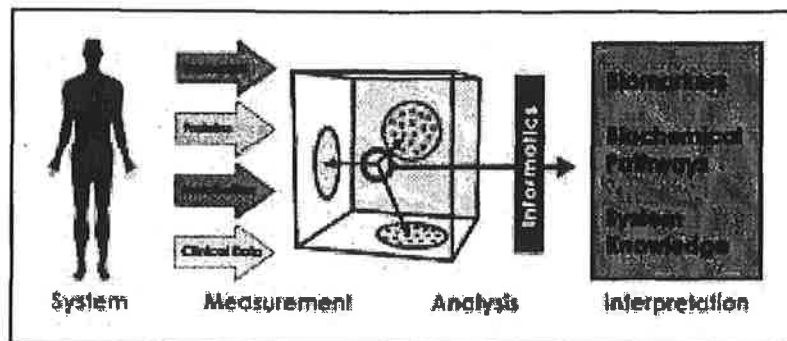
Figure 2. An overview of applied systems biology.

nucleosides, steroids, and vitamins, has extremely different molecular substructures. Although the bioanalytic challenges to determine metabolome structures are considerable, metabolomics is complementary to genomics, transcriptomics, and proteomics and provides key component information and insight into cellular and molecular processes.

## SYSTEMS BIOLOGY

The use of high throughput and efficient -omic platforms has principally been applied in global gene expression and whole genome functional analysis. Such approaches have provided valuable information content, but often these data sets are somewhat unidimensional.[8] This problem is compounded by the fact that many genetic, biochemical, and metabolic disorders, ie, cancer, Alzheimer disease, and atherosclerosis, result from complex multimolecular interaction effects. Such diseases cannot be explained readily by an alteration in a single gene, gene product, or metabolite. Hence, it is necessary to use a more holistic, multidimensional systems level approach to reconcile the genotype-to-phenotype gap in disease onset, manifestation, and progression.

More recently, a new paradigm known as systems biology (also referred to as pathway, network, integrative, or new biology) has emerged.[9-13] Applied systems biology integrates genomic, transcriptional, protein, metabolite, and clinical data by concomitantly measuring differential expression, regulation, and abundance of molecular compound subclasses in healthy vs perturbed systems, eg, disease states (Figure 2).[13] The acquired data sets are manipulated using a variety of data preprocessing, statistical, computational, and informatics mining tools, as well as modeling and knowledge assembly interrogation.[14,15] The final data output allows both linear and nonlinear correlations of individual components to be determined and correlation pathways and multidimensional networks to be unraveled and ultimately modeled and simulated.[14]

There appears to be a general consensus across academic, biotechnology, and pharmaceutical industry sectors that systems biology will indeed "revolutionize" our understanding of specific diseases and radically improve treatment of such diseases. However, there is substantially less agreement about a unique definition of systems biology. Hood[1] (widely regarded as the founding pioneer of modern systems biology) recently defined systems biology as "the study of all the elements in a biological system (all genes, mRNAs, proteins, etc) and their relationships one to another in response to perturbations." Others have argued that systems biology is even more all-encompassing than defined by Hood. Assorted definitions are captured in the following amalgam: "systems biology is the science of integrating genetic, genomic, biochemical, cellular, physiological and clinical data to create a system network that can be used to predictively model a biological event(s)." Part of the difficulty in defining systems biology lies in the fact that it is perhaps more appropriate to consider systems biology a process, rather than a new subdiscipline of modern biology. As noted previously, it is the integration of -omic as well as clinical, physiological, and imaging measurements that is central to the paradigm shift of systems biology. However, the process of systems biology analyses can be considered as consisting of a series of at least 8 different input modules and a minimum of 2 output modules (Figure 3).

1.  Biological Question and Experimental Design—one needs to select the biological system (ie, body fluid, organelle, cell type, tissue, organ or organism) to be studied, along with the appropriate hypothesis or discovery-driven question to be answered. The necessary controls (eg, normal vs diseased), sample histories, and outcomes as well as statistically significant sample numbers to be analyzed also need to be determined.

2.  Data Acquisition—omic, clinical, physiological, and imaging data are acquired on a variety of analytic platforms. Data sets are obtained on both control and perturbed (eg,
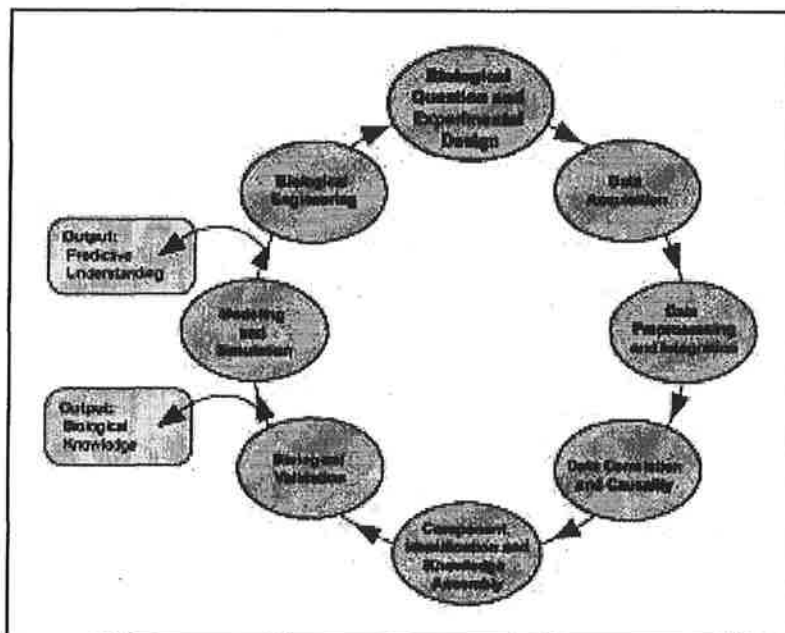
Figure 3. Modules of systems biology, a process consisting of numerous interchange-
able modules. The order of the process is determined by the biological question being
investigated.

diseased, drug or toxin treated, knockout animal) cohorts of samples.

3. Data Preprocessing and Integration—data files are smoothed, aligned, normalized, and ultimately merged into composite files.

4. Data Correlation and Causality—merged data files are compared, and ultimately correlation and causal networks are produced. Tools for data visualization are applied at this juncture.

5. Component Identification and Knowledge Assembly—statistically significant components differing between control vs perturbed cohorts are identified. Ultimately, correlation or causal networks are interrogated against all known knowledge using a variety of tools, from simple text mining to semantic Web protocols.

6. Biological Validation—to ensure that the correlation or causal networks have biological relevance, findings must be related back to the biology of the system being investigated. Strategies such as RNA interference, high throughput cellular bioassays, and knock-in or knockout organisms can all be used to provide a biofocusing of data back to the relevance of the biological question posed originally.

7. Modeling and Simulation—data and correlation networks can be used as a framework for further modeling and simulation studies of the biological processes being investigated.

8. Biological Engineering—once a system has been modeled and understood, one can reengineer the pathway or network to produce a better outcome (eg, disease-resistant crop) or to identify the optimal therapeutic agent.

9. Output: Biological Knowledge—relevant biological knowledge is produced to answer, in part or completely, the original question posed.

10. Output: Predictive Understanding—based on modeling and simulation studies of the biological system being investigated; it should now be possible to predict the outcome of specific changes in the pathway or network.

The specific process flow and the connectivity of the modules are determined by the initial biological question posed. For example, a discovery-driven question may require the modular flow aforementioned and shown in Figure 3. However, a hypothesis-focused question may necessitate a restructuring of modules, eg, "Knowledge Assembly" is carried out before "Data Acquisition." In any event, the systems biology process attempts to capture and portray the staggering complexity and interconnectivity of cellular and organism events that occur on the microsecond-hour-day time scale. This unique modular approach with its combination of tools, technologies, and resulting systems knowledge potentially holds the key to dissecting disease onset, progression, and treatment as well as accel-

erating the bottlenecks of the therapeutic drug discovery and development process.

## SYSTEMS BIOLOGY AND DISEASE
### Overview

Despite the advances made to date and the tools currently available for the study of genomics, transcriptomics, proteomics, and metabolomics, numerous disease states can be explained only by complex multimolecular interactions, rather than by alteration of a single gene, gene product, or metabolite. Thus, to garner a more complete and relevant understanding of disease, we must obtain a comprehensive perspective of the biological system, thereby uncovering the interdependent and dynamic cellular pathway and network events that undergo change as a function of disease onset and progression.

This concept can be illustrated by considering diabetes mellitus.[16] In type 1 diabetes, treatment with exogenous insulin restores homeostatic carbohydrate and lipid levels. However, in type 2 diabetes, which accounts for approximately 90% of all reported cases of diabetes, defects in insulin secretion and mode of action as well as increased hepatic glucose output are the principal causes of disease. Several therapeutic agents are commercially available to treat one or more of these disease factors in type 2 diabetes, including 5 different classes of hypoglycemic agents— sulfonylureas, meglitinides, biguanides, α-glucosidase inhibitors, and thiazolidinediones.[17] However, it is widely recognized that currently available oral antidiabetic agents have considerable limitations in terms of efficacy and/or safety, and they have not prevented the continued emergence of the global epidemic of diabetes. In part, the limited progress in our understanding and treatment of diabetes is due to its complexity and the multifactorial components involved in disease onset and progression. Type 2 diabetes is a metabolic syndrome of interrelated metabolite problems. Also, it is modulated by environmental factors such as diet, excess body weight, and sedentary lifestyle, and there appears to be a significant genetic predisposition for the disease.[16] Clearly, genetics and protein biochemistry as well as subclasses of metabolites (ie, lipids) all play an interconnected role in the onset, modulation, and progression of the disease. Hence, acquiring the genomic, proteomic, and metabolomic data simultaneously over time and the spatial localization and interconnecting these data to discern the causal relationships of changes in expression of all components as they affect each other may provide important new insight into the disease process. Such an applied systems biology approach should afford connectivity networks of components and causality networks, provided time point measurements are also acquired. These interactive pathways and networks of the disease state should provide substantially more insight into understanding the predisposition, onset, and progression of this complex disease.

### Apolipoprotein E*3 Transgenic Mouse Model for Atherosclerosis

Recently, the first-ever systems biology analysis of a mammalian disease model was reported.[18] As part of a proof of principle study, apolipoprotein E*3 (APOE*3)-Leiden transgenic mice were subjected to an applied systems biology analysis. The APOE*3-Leiden mutation is a knock-in mouse that expresses a *human* mutant apolipoprotein E that is normally associated with familial dysbetalipoproteinemia in humans. These mice are highly susceptible to diet-induced hyperlipoproteinemia and atherosclerosis because of diminished hepatic low-density lipoprotein receptor recognition, but when fed a normal chow diet, they display only mild type 1 (macrophage foam cells) and 2 (fatty streaks with intracellular lipid accumulation) lesions at 9 months.[19]

Liver tissue was obtained from individual isogenic wild-type mice (n=10) and APOE*3-Leiden mice (n=10) that were fed a normal chow diet and killed at 9 weeks of age (see Figure 4 for a process flow diagram of the analysis). Initially, data sets were acquired on individual mice for mRNA expression levels, soluble lipoprotein abundances, and lipid differential profile analyses. Hepatic mRNA expression analysis showed a 25% decrease in the apolipoprotein A-1 gene and a 43% increase in liver fatty acid–binding protein (L-FABP) expression between transgenic and wild-type control mice, whereas there was no change in peroxisome proliferator–activated receptor-α expression. Additional proteomic and metabolomic analysis allowed rapid identification of early protein and metabolite markers of disease pathology, including a 44% increase in L-FABP in transgenic animals compared with controls, as well as an increase in triglyceride levels and select bioactive lysophosphatidylcholine species.

To carry out a systems biology analysis, we merged the individual transcriptomic, proteomic, and metabolomic data sets for each mouse liver (3 files per mouse) into single data files (Figure 4). The resulting 20 concatenated data files, now containing transcript, protein, and metabolite information, were then aligned and normalized so that a variety of univariate and multivariate statistical analyses could be performed. Finally, in the hope of discovering hitherto unknown relationships that might exist among measured genes, proteins, and metabolites, a correlation network analysis was carried out. The association between any 2 components $i$ and $j$ (eg, a protein and a metabolite) is determined by calculating their Pearson correlation coefficient, $C_{ij}$. This was done for each set of pairs within the
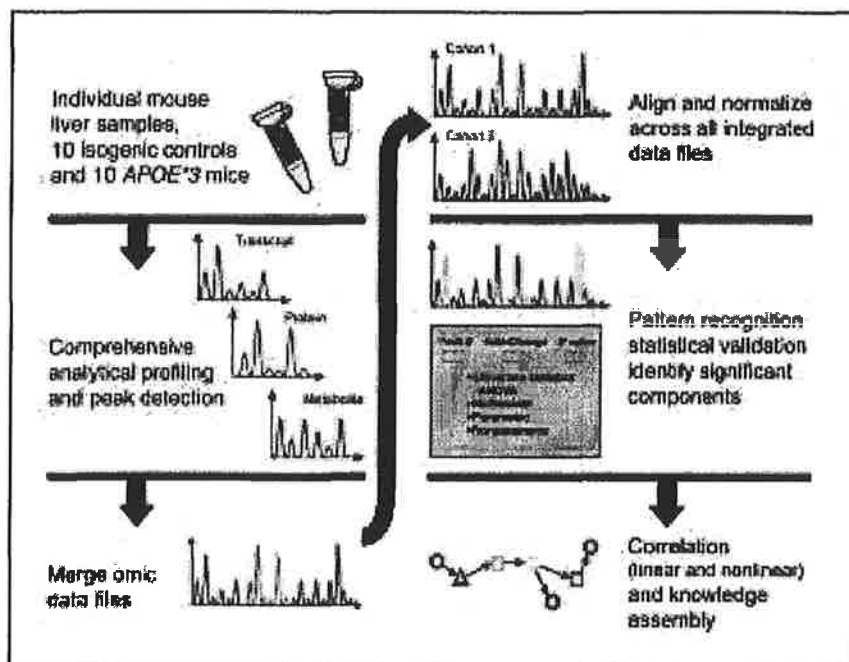
Figure 4. Step-by-step description of the process carried out to produce the correlation network shown in Figure 5, for the apolipoprotein E*3 (APOE*3)-Leiden transgenic mouse systems biology experiment described in the text. ANOVA = analysis of variance.

investigated data set (summing all samples). After removing associations weaker than $|Cij|<0.8$, we had a nonlinear kernel principal component analysis network containing only highly correlated entities. This method is similar to relevance networks introduced by Butte et al.[20] A correlation network containing approximately 10% of the overall data set acquired in the APOE*3 analysis is shown in Figure 5. This correlation network represents identified proteins and metabolites, as well as genes that encode the identified proteins or those whose expression was significantly different in wild-type vs APOE*3-Leiden mice. The network analysis shows a high degree of correlation among apolipoprotein A-I gene, L-FABP (both gene and protein expression), and lipids, such as diacylglycerol and lysophosphatidylcholine, because these components appear as a cluster within the network. Components of this cluster also show a high correlation to proteins involved in metabolism (eg, pyruvate kinase) and signal transduction (protein kinase C μ).

After obtaining the correlation network, we subjected it to a variety of knowledge assembly inquiries, including text mining. Known biological connections and pathway constituents were detected, which had been reported previously in fatty acid metabolism studies. However, less than 10% of the correlations corresponded to known biological connectivities reported previously. Clearly, such an approach affords a powerful new tool in discovery biology. Nonetheless, it also indicates the necessity for additional biological validation of such findings (Figure 3). Ultimately, this study demonstrates the utility of an integrated approach for characterization of a highly complex system. By generating high content analytic output and comparing principal component factors derived from composite data sets, we were able to rapidly elucidate identities and the relative abundances of major lipoprotein metabolism mediators that define the APOE*3-Leiden phenotype compared to the isogenic control.

## SYSTEMS BIOLOGY AND EFFECTIVE DRUG TREATMENT

The ability of systems biology to affect our understanding of disease mechanism and progression can be extended to understanding the mechanism of action for existing treatments of disease. For example, signaling events within or between cells are not restricted to linear pathways but are well recognized to be components of complex and dynamic networks.[21] Thus, the true mechanism of action of many therapeutics has alluded investigators. Deciphering the complexities of intercellular and intracellular signaling relationships is a mission uniquely feasible through systems biology because the fundamental strength of this approach is its ability to analyze an entire system comprehensively.
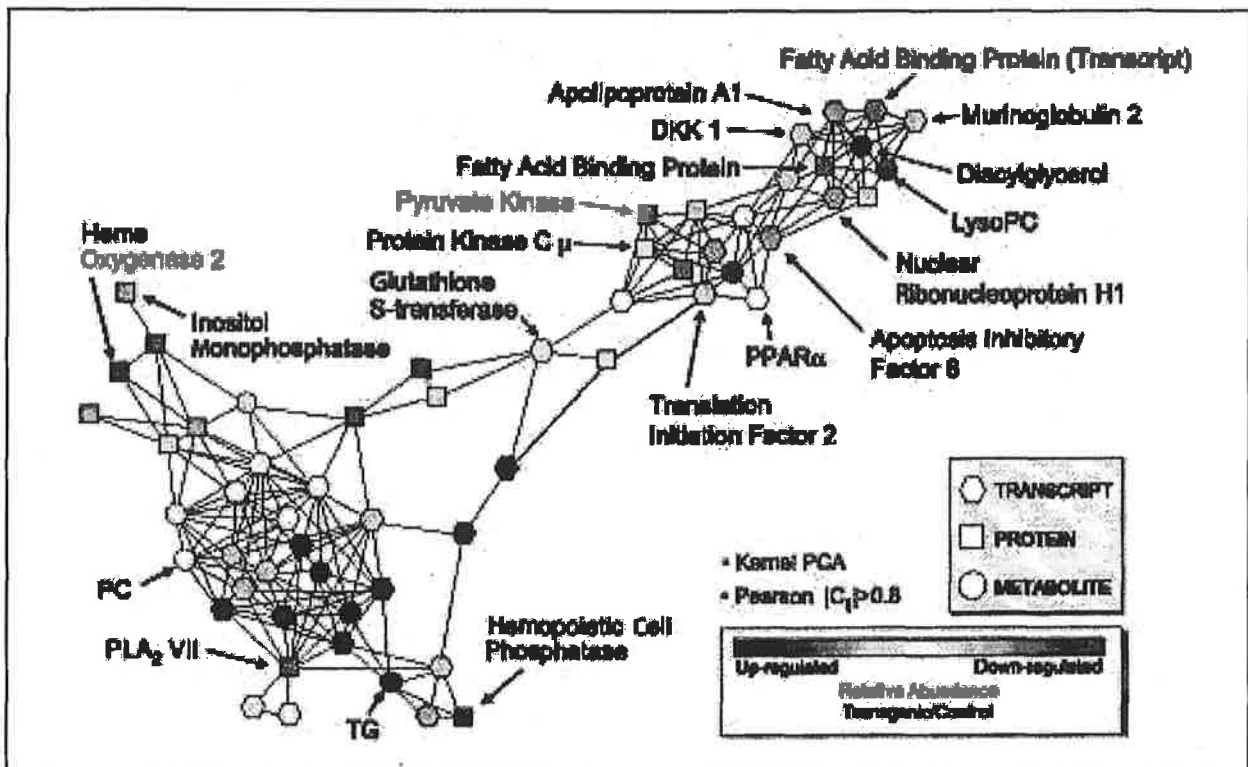
Figure 5. Systems biology correlation network of select expressed genes, lipoproteins, and lipids. The colored shading inside each node indicates the relative amount in the transgenic apolipoprotein E*3 (APOE*3)-Leiden mice compared to wild-type controls (red = higher level, green = lower level), and a black line connecting 2 entities indicates a high level of correlation (a Pearson correlation coefficient of >0.8 was considered significant). From Clish et al,[18] with permission. DKK = Dick Kopf gene; LysoPC = lysophosphatidylcholine; PC = phosphatidylcholine; PCA = principal component analysis; PLA = platelet-activating factor; PPAR = peroxisome proliferator–activated receptor; TG = triacylglycerols.

A comparative analysis of treated and untreated diseased samples (ie, drug perturbation studies) will facilitate an understanding of signaling networks associated with drug treatment. Specifically, through a comprehensive analysis of the biological system after treatment with small molecules, we may define molecular consequences affected by these molecules. By strategically selecting a panel of perturbating agents (perturbagens) for parallel studies, we may be able to discriminate between cellular changes associated with therapeutic benefits and those associated with adverse effects of various agents—information relevant for development and regulatory purposes. The discovery of novel sites of interaction within the signaling pathway where molecular targets are free of the connection to adverse effects provides the prospect for therapeutics with enhanced efficacy, reduced adverse effects and improved therapeutic indices, and diminished onset to efficacy.[22]

One intriguing prospect (as discussed previously) is that systems biology is poised to affect therapy through its ability to define causal mechanisms of disease expeditiously. Causative targets may be well-characterized proteins that have previously lacked a recognized connection to the disease in question (eg, the discovery of the connection of phosphodiesterase type 5 enzyme to erectile dysfunction[23]). They may also be novel molecular targets such as endogenous (wild-type) or mutant proteins that formerly have been uncharacterized and therefore do not have an established connection to disease (eg, the identification of the leukemia-associated protein encoded by the BCR-ABL gene[24]). Either form of target can be uncovered through a systems biology analysis of the differences between normal and diseased samples.

## SUMMARY
With the evolution of our understanding of the complexities of cellular processes, we have been forced to recognize that our attempts to unravel the biological mechanisms of disease have been restricted by our narrow view of cell and molecular biology. Rather than focusing on one aspect of

molecular pathways, we now realize the importance of pathway and network connectivities. The ability of systems biology to generate a broader, integrated view of the biological system will yield a more complete understanding of disease, ultimately providing us with improved tools for treatment. One can further understand how this approach can be readily implemented for identifying prognostic, diagnostic, and surrogate biomarkers.[25] Finally, systems biology analysis of appropriately designed cohorts may also provide the key to understanding the differences in patients who do or do not respond to any particular therapy. This information may be used for a more efficient (and therefore more cost-effective) design of clinical trials. The potential benefits in each of these arenas offer the promise of dramatically affecting medicine, thereby positively influencing the outcome of human health. We are definitely at the beginning of an exciting and productive journey in medicine, in terms of penetrating new insight into understanding and treating disease.

## REFERENCES

1. Hood L. A personal view of molecular technology and how it has changed biology. *J Proteome Res.* 2002;1:399-409.
2. Roberts L, Davenport RJ, Pennisi E, Marshall E. A history of the Human Genome Project. *Science.* 2001;291:1195.
3. Lander ES, Linton LM, Birren B, et al, International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860-921.
4. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome [published correction appears in *Science.* 2001; 292:1838]. *Science.* 2001;291:1304-1351.
5. Haberman A. Potential of systems biology and pathways studies. *Genet Eng News.* 2002;22:26-32.
6. Abbott A. And now for the proteome. *Nature.* 2001;409:747.
7. van der Greef J, Davidov E, Verheij E, et al. The role of metabolomics in systems biology. In: Harrigan GG, Goodacre R, eds. *Metabolite Profiling: Its Role in Biomarker Discovery and Gene Function Analysis.* Boston, Mass: Kluwer Academic Publishers; 2003:171-198.
8. Cockett M, Dracopoli N, Sigal E. Applied genomics: integration of the technology within pharmaceutical research and development. *Curr Opin Biotechnol.* 2000;11:602-609.
9. Ideker T, Thorsson V, Ranish JA, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science.* 2001;292:929-934.
10. Kitano H. *Foundations of Systems Biology.* Cambridge, Mass: MIT Press; 2001.
11. Kitano H. Systems biology: a brief overview. *Science.* 2002;295: 1662-1664.
12. Hood L. Systems biology: integrating technology, biology, and computation. *Mech Ageing Dev.* 2003;124:9-16.
13. Davidov E, Holland J, Marple E, Naylor S. Advancing drug discovery through systems biology. *Drug Discov Today.* 2003;8:175-183.
14. Neumann E, Thomas J. Knowledge assembly for the life sciences. *Drug Discov Today.* 2002;7(20, suppl):S160-S162.
15. Kitano H. Computational systems biology. *Nature.* 2002;420:206-210.
16. Saltiel AR. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *Cell.* 2001;104:517-529.
17. Chakrabarti R, Rajagopalan R. Diabetes and insulin resistance associated disorders: disease and the therapy. *Curr Sci.* 2002;83: 1533-1538.
18. Clish CB, Davidov E, Oresic M, et al. Integrative biological analysis of the APOE*3-Leiden transgenic mouse. *OMICS.* 2004;8:3-13.
19. Lutgens E, Daemen M, Kockx M, et al. Atherosclerosis in APOE*3-Leiden transgenic mice: from proliferative to atheromatous stage. *Circulation.* 1999;99:276-283.
20. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A.* 2000;97:12182-12186.
21. Marinissen MJ, Gutkind JS. G-protein-coupled receptors and signaling networks: emerging paradigms. *Trends Pharmacol Sci.* 2001;22:368-376.
22. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov.* 2002;1:727-730.
23. Terret NK, Bell AS, Brown D, Ellis P. Sildenafil (VIAGRA™), a potent and selective inhibitor of type 5 cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. *Bioorg Med Chem Lett.* 1996;6:1819-1824.
24. Capdeville R, Buchdunger E, Zimmermann J, Matter A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov.* 2002;1:493-502.
25. Naylor S. Biomarkers: current perspectives and future prospects [editorial]. *Expert Rev Mol Diagn.* 2003;3:525-529.