Abstract: The molecular pathways that govern human disease consist of molecular circuits that coalesce into complex, overlapping networks. These network pathways are presumably regulated in a coordinated fashion, but such regulation has been difficult to decipher using only reductionistic principles. The emerging paradigm of “network medicine” proposes to utilize insights garnered from network topology (eg, the static position of molecules in relation to their neighbors) as well as network dynamics (eg, the unique flux of information through the network) to understand better the pathogenic behavior of complex molecular interconnections that traditional methods fail to recognize. As methodologies evolve, network medicine has the potential to capture the molecular complexity of human disease while offering computational methods to discern how such complexity controls disease manifestations, prognosis, and therapy. This review introduces the fundamental concepts of network medicine and explores the feasibility and potential impact of network-based methods for predicting individual manifestations of human disease and designing rational therapies. Wherever possible, we emphasize the application of these principles to cardiovascular disease. (Circ Res. 2012;111:359-374.)

Key Words: network medicine ■ systems biology ■ cardiovascular disease ■ systems pharmacology

For centuries, reductionistic thinking has dominated Western scientific theory and has been rigorously applied to biomedicine. This construct argues that crucial biological factors operate in simple mechanistic association to control disease pathobiology. Consequently, current classification of disease phenotype (pathophenotype) is the result of inductive generalization from clinicopathological evidence founded on the law of parsimony. This paradigm has been helpful to clinicians as it formalizes syndromic patterns that limit the number of potential pathophenotypes. While quite useful in an earlier era, classifying disease in this way overgeneralizes pathophenotypes and does not usually consider individualized nuances in disease expression, susceptibility states, or preclinical disease manifestations. Underlying those variations in predisposition to disease and subsequent disease manifestation is the competing concept that complex biological processes driving human disease rarely result from an abnormality in a single molecular effector. Rather, they are nearly always the net result of multiple pathobiological pathways that interact through an interconnected network.

Network theory offers a tractable structure for distilling relevant insight from the growing sets of molecular “-omics” data. In its simplest form, a network can be generated by mapping a set of molecular entities, called “nodes,” with their positions based on their functional interconnections, called “links” or “edges,” with one another (Figure 1A). In network medicine, a “node” may represent a biological factor (eg, gene, chromosome, metabolite, noncoding RNA, or even a specific disease/phenotype). “Edges” may represent physical interactions, transcriptional induction, enzyme activation, enzymatic conversion of one metabolite to another, or even shared genes/traits among distinct disease manifestations. As a result of rapidly progressing technology and high-throughput data sets, the first representations of a comprehensive network, or “interactome,” of relevant functional molecular interactions in human tissue is now possible.

Currently, data accumulation for network construction has primarily relied on an integration of genetic studies (linkage analyses and genome-wide association studies, GWAS), so-called high-throughput –OMIC analyses (genomic sequencing, transcriptional expression analyses, proteomic analyses, etc), and biochemical studies (ie, metabolomic analyses), coupled with cellular, physiological, and clinical data. As a result, application of network theory has begun to permeate the scientific study of a large number of defined biomedical systems, ranging from simple signaling circuits to complex human diseases, such as cancer. The science of studying disease network behavior is just now emerging and has been termed “network medicine” (Figure 1B).1 In this review, we will define the challenges of building the current human “interactome” and the fundamental principles of biological network theory in medicine. In doing so, we will...
explore how the study of network topology and network dynamics can accelerate discovery of novel disease genes and pathways, offer new insights into the systems-wide effects of drug therapy, and reshape disease classification. In the course of this presentation, we will highlight key studies, cardiovascular and otherwise, that have successfully integrated network theory with experimental validation and thus have demonstrated the power of these approaches for enhancing our understanding of molecular and cellular interconnectedness and its potential regulation in the control of disease phenotype.

Modeling of Biochemical Networks

Although most areas of modern medicine have relied nearly exclusively on reductionist models of human pathogenesis, efforts to establish and apply a detailed understanding of networks for predicting “downstream” effects on cellular or physiological states have been attempted but have not yet been routinely adopted. While catalogs have been curated in recent years describing the global regulation of mammalian genes, detailed biochemical and kinetic data associated with enzymatic pathways controlling cellular metabolism and energy production have been available for decades. As a result, mathematical frameworks exist for quantifying how metabolic variables (eg, metabolite concentrations) depend on biochemical network parameters,2–4 thus setting the stage for attempts at utilizing in silico simulation and modeling to understand complex biochemical processes.

Methodology involves use of either deterministic [eg, systems of ordinary differential equations (ODEs) to assess the effects of perturbations on associated nodes in the network context] or stochastic [eg, accounting for random effects on associated nodes by calculation of effect probabilities after network perturbation] simulation of reaction networks. Such simulation has allowed for the computation of metabolite steady-state concentrations and their stability5 or for the analysis of transcriptional network dynamics, respectively.6 “Sensitivity analysis,” or so-called “metabolic control analysis” (MCA), is used to compute sensitivity coefficients, which generally reflect the ratios between the change in the biochemical network behavior (as monitored by some measure of global system behavior) and the perturbation on specific system nodes.3,4 Resulting parameters with large sensitivity magnitude are deemed to be biologically important and hence considered to be the controlling factors in the functional regulation of a system.7 Various software packages for kinetic modeling of biochemical networks are available and continue to grow in number.8 Although most software shares a common core of functionality, the specific capabilities and user interfaces of different packages differ with respect to their functionality, reliability, efficiency, user-friendliness, and compatibility.

Validation of the predictions arising from such MCA have mostly centered on relatively narrow biochemical networks linked by defined rate and time constants that are consistent among a variety of perturbations. Although in many biological circumstances such rate constants fluctuate and depend on regulation of the catalyzing enzyme, the tight homeostatic regulation that is seen in many biochemical networks, in part, allows for consistent approximations of rate constants without substantially altering the network output. Moreover, probabilistic estimations of dynamically changing rate constants have been attempted.9 Such methodologies have validated important molecular features in metabolic and signaling cascades at the single cell level. Consequently, one of the uses of MCA is to infer the importance of cellular processes or pathways and to provide mechanistic explanations for biological behavior.10–14 As the technology improves for detecting dynamic changes in metabolites via high-throughput metabolomics in a variety of clinically relevant scenarios,15 the potential capabilities of such biochemical kinetic modeling should continue to expand.

Modeling of Transcriptional Gene Regulatory Networks

The success of such biochemical modeling strategies has paved the way for modeling other biological systems. For example, attempts to map transcriptional gene regulatory networks have been pursued. Similar to metabolic networks, the first versions of these networks were constructed from the simple known regulatory pathways mapped via traditional molecular biology studies. With the advent of high-throughput expression array technology and the ability to quantify and correlate alterations in mRNA expression at the transcriptomic level to various stimuli, the ability to construct gene “coexpression networks” has advanced considerably, allowing for mapping of gene modules that are coregulated under a shared stimulus or are associated with a relevant biological condition or genetic polymorphism (Figure 2A). In this context, novel computational procedures have been developed, such as “expression screening,” which can integrate information from thousands of microarray expression data sets to identify genes that are consistently coexpressed with a target pathway, such as those in oxidative phosphorylation16 across biological contexts. By additionally accounting for
predictive algorithms to identify promoter binding sites for transcription factors (TRANSFAC and B-cell interactome) as well as incorporating data from high-throughput assays demonstrating actual physical binding of transcription factors to promoters [chromatin immunoprecipitation followed by microarray analysis (ChIP-ChIP) and ChIP followed by sequencing (ChIP-Seq) as found in databases including the Universal Protein Binding Microarray Resource for Oligonucleotide Binding Evaluation (UniPROBE) and the open access database of transcription factor-binding profiles, JASPAR], further coexpression networks have been verified and, in turn, used to identify novel factors and direct regulatory connections that influence substantially various cellular phenotypes.

Network Modeling of Protein-Protein Interactions and Posttranslational Modifications

More recently, there have been considerable advances made in cataloguing the immense set of human protein-protein interactions (PPI). While curation from the scientific literature alone has revealed substantial obstacles17 (Figure 2B), more promising methodologies using yeast 2-hybrid and 3-hybrid screens as well as immunoprecipitation and high-throughput mass spectrometry have greatly increased the likelihood of attaining a more comprehensive PPI catalog.18–22 To date, novel insight of biological function from these PPI networks has resulted largely from analysis of static networks. There has been thought of establishing more dynamic modeling via incorporation of reported association constants among protein pairs. However, it remains to be seen if mathematical modeling can adequately capture the sheer complexity of these networks, and, more importantly, if such modeling is accurate when considering the potential need to define the precise and changing concentrations of each protein component involved in competitive binding to multiple protein partners in different cellular compartments. Similarly, regulatory relationships that coordinate activating and inhibitory posttranslational modifications (eg, phosphorylation, acetylation, S-nitrosylation, redox modifications, etc) have begun to be described in databases such as PhosphoELM, PhosphoSite, phosphorylation site database (PHOSIDA), NetPhorest, and the CBS prediction database. While phosphorylation sites in the proteome have been reasonably mapped and catalogued, the systematic cataloging of alternative modifications has yet to be completed. When considering the more than 300 posttranslational modifications of the proteome that are now recognized as well as associated protein binding complexes with bioactive inorganic ions (eg, iron, iron-sulfur clusters, selenium, etc), such network mapping re-
 mains an essential but daunting task. For instance, while substantial evidence supports the notion that redox modifications regulate specific enzymatic activities, identification and prediction of redox modifications throughout the proteome presents a complicated challenge, as the exact identity of relevant modifications depends not only on protein site but also on the oxidative state of redox-sensitive amino acids, such as cysteines (e.g., sulfonic versus sulfenic versus thiolate moieties). Nonetheless, the systematic mapping of these modifications by mass spectrometry has begun by both academic and commercial biotechnology laboratories.

Network Modeling of Epigenetic Modifications and Interactions With Noncoding RNA

Epigenetic alterations of the genome (e.g., DNA methylation) and the transcriptome (e.g., repression by noncoding RNA)
have recently been recognized as powerful and ubiquitous gene regulatory mechanisms. These regulatory patterns are increasingly becoming incorporated into more comprehensive gene regulatory networks. For example, the MethDB database contains a catalog of genomic DNA methylation encompassing 5382 methylation patterns or profiles for 48 species, 1511 individuals, 198 tissues and cell lines, and 79 phenotypes. Thus, it may soon be possible to integrate such granularity in this database into other networks to reveal novel functional connections through DNA methylation.

Numerous atlases also have been compiled recently detailing the systems-wide functions of noncoding RNA, such as microRNA (miRNA). MiRNA are conserved, non–protein-coding RNA molecules that have been identified as essential mediators of a variety of genes and cellular processes. Inside the cell, miRNA negatively regulate gene expression primarily by binding to the 3’ untranslated regions of messenger RNA (mRNA) transcripts to repress translation and/or degrade mRNA. Efficient binding relies on Watson-Crick base-pairing between the seven-nucleotide “seed sequence” of a given miRNA and its mRNA target, and several algorithms (TargetScan, miRBase, PicTar, miRDB, and others) have accordingly been developed to predict accurately miRNA targets of each miRNA. In concert, high-throughput analyses using transcriptional expression arrays and/or proteomics have allowed for verification and identification of additional miRNA targets, which have been incorporated into a number of these databases (eg, DIANA and TarBase).

Although the utility of these prediction algorithms and high-throughput methods has been reflected in their prominent use in identifying key mRNA targets of specific miRNA, their use in more comprehensive network-based approaches is just beginning. It is estimated that most messenger RNA transcripts are regulated by miRNA, with each miRNA (>700 identified mammalian miRNA) typically repressing expression of multiple mRNA (at times, >100 targets simultaneously). It has been proposed that miRNA appear to act by regulating multiple interacting targets in the same functional network to generate robust actions in vivo.34 A deeper exploration of the implications of such networks of miRNA regulating networks of mRNA will certainly necessitate advances in computational analyses as well as experimental validation. Nonetheless, the possibility of mapping a comprehensive epigenetic regulatory network for complex human diseases is increasingly becoming realized.

**Topological Characteristics of the Human Interactome**

Building on the advances of mapping biochemical, gene regulatory, protein-protein, and epigenetic networks, we are now able to construct a more expansive network that more comprehensively reflects the multitude of molecular connections affecting human health and disease. Depending on which sets of databases are used, the human “interactome” currently encompasses 25000 protein-coding genes, 1000 metabolites, and an increasing number of posttranslationally modified proteins and functional RNA molecules, together exceeding 100000 participants and substantially greater functional interactions that variably overlap with one another. The resulting interactome displays significant clustering, with pockets of especially dense interconnectedness. As a result, at most, a few links separate each node from any other node. Thus, perturbation of a single node can induce widespread effects throughout multiple modules and, thus, can substantially alter multiple functional systems.

Importantly, the human interactome is a “scale-free” network in which gene interactions do not occur randomly, with the degree distribution following a power-law tail (Figure 3A). Thus, this type of network contains a few highly connected “hubs,” which are nodes that link substantially to multiple other nodes, and typically are essential for maintaining the integrity of the entire network. Hub proteins are probably encoded by evolutionarily conserved genes that function in crucial cellular activities. Some hubs appear to direct intrinsic gene cluster function, whereas others preserve the integrity of the entire interactome by linking distinct clusters to one another. Scale-free networks exhibit properties that offer substantial insight into biological systems. Scale-free properties minimize transition time between states of the system; they facilitate transfer of molecular information across the system; they accommodate perturbations to the system with minimal adverse effects; and they promote biological diversity. Notably, scale-free systems exhibit “overdetermined” features, implying that the complete definition of all components of the network is not necessary to understand how the network functions as a whole. Finally, scale-free systems display emergent behavior, as defined by the fact that intimate knowledge of one node cannot be used to predict its behavior in a network nor the behavior of the entire network (Figure 3B). It is this emergent property of biological networks that emphasizes the significant pitfalls in the reductionistic studies guiding biomedical research presently—that is, erroneously attempting to extrapolate the functional study of a single gene alone to its in vivo role(s) within the context of the endogenous biological network within which it operates.

Disease-relevant genes and related pathways are certainly present within the human interactome but have been incompletely defined. Currently, only 10% of human genes have been associated with a known disease (www.omim.org), and many have been found to correlate more often with nonhub nodes. Disease genes may also exist as hubs or essential genes, but since dysfunction of essential genes often results in embryonic lethality, the association of disease genes and hubs has been more difficult to prove, especially for chronic diseases of adulthood. According to the “local hypothesis,” disease genes also tend to interact directly with other disease genes in “disease modules” that induce a common pathophenotype (Figure 3C). Indeed, expansive analysis of gene expression arrays has confirmed the consistent and pervasive nature of these disease “signatures” of gene expression even across vastly different tissue sources. These disease modules can be challenging to map. Nonetheless, they may substantially overlap with “topological modules”—identifiable by unbiased network-clustering tools—and with already-defined “functional modules”—clusters of nodes of similar or related function that have been defined previously by independent experimental validation (Figure 3C). Accord-
ingly, human disease manifestation may result from several perturbations of a single disease module which may overlap with the same components as related, but perhaps more easily identifiable, topological and functional modules. Consequently, the topographical position of a single disease gene in the human interactome can offer unique insight into its novel associated partners, their connected roles in pathogenesis, and connected modules that regulate disease manifestation and response to therapeutic interventions.

Network Topology and Human Disease

Network medicine can accurately predict novel disease genes, based on their topographical position in relation to known disease-relevant factors (Figure 4A). First, the so-called “linkage” method enables the identification of unique factors important in human disease via their direct connection to known disease genes. In cardiovascular disease, successful examples have included identification of a network of candidate susceptibility genes for coronary artery disease based on disease genes selected from GWAS followed by analyses of directly connected neighboring genes via a protein-protein interaction database. Ghazalpour et al identified a cadre of genes that are associated with obesity in the mouse based on their links in a coexpression network with disease genes genetically associated with an obese phenotype. Second, the so-called “disease module” method (with some variants called “neighborhood” or “clustering” methods) relies on network mapping of genes altered in transcriptional coexpression analyses of diseased tissue, thus defining previously unrecognized factors carried within these “disease modules” that influence pathogenesis. For cardiovascular diseases, relevant modules have been defined for atherosclerosis, in-stent restenosis, type 2 diabetes mellitus, and cardiac hypertrophy and failure, among others. We have recently identified a disease module important in the progression of pulmonary hypertension, allowing for the identification of a set of microRNA that control the pathophenotype based on the predominance of their targets in that gene module (Figure 4B). Finally, delineation of modules that topologically neighbor but do not overlap with a given disease gene or module has facilitated the delineation of new disease pathways. Based on the mean topological distance from a given set of disease nodes, so-called “random walk” algorithms have been useful in defining such neighboring modules and linking them to diseases such as diabetes mellitus. Linkage methods and mapping of disease modules have become much more prevalent, probably due to the relative ease of performing these analyses. Conversely, diffusion methods appear to carry the best predictive power, probably stemming from their ability to integrate a greater amount of relevant topographical information into the final analysis.
Network medicine may also provide insight into the presumed but otherwise unappreciated molecular links among separate human diseases. Goh et al first demonstrated the existence of these widespread connections after constructing a comprehensive map of disease modules in the human interactome. Similarly, analyses of coexpression networks have revealed novel molecular connections between Alzheimer disease and cardiovascular illnesses. More recently, Rende et al constructed a cardiovascular disease functional linkage network (CFN) that carries significant interconnections among modules representing cardiovascular diseases with other complex disorders such as infection by *Listeria monocytogenes*, myasthenia gravis, hemorrhagic diatheses, and protein S deficiency. Gene coexpression analyses have also identified shared candidate genes that link normal myocardial development to myocardial hypertrophy and failure. Thus, such analyses can be used to identify links among developmental and pathological contexts in the same tissue. As a result, these molecular interconnections may drive the clinical cosegregation of specific disease phenotypes. Furthermore, identification of these shared links may lead to novel, shared therapeutic targets for multiple diseases or suggest a “repurposed” use of medications already approved for separate disease conditions.

**Network Dynamics and Human Disease**

Although network topology certainly provides unique insights, more substantial advances in our comprehension of biological function in complex systems depend on understanding dynamic signal flux through these networks. To do so, as with any complex circuity, direct measurement or inference of three key components is essential: the initial concentrations of each component or node, the overall structure of the network (eg, the interconnectivity), and the strength and direction of the nodal interactions. Although great advances have been made in defining the first two of these components in a variety of human disease conditions, accurate determination of the dynamically changing strength of interactions is a challenging task. Nonetheless, rapidly advancing methodologies have been proposed to define or infer the details of those interconnections.

As discussed previously, understanding the dynamics of biochemical networks has been fruitful given the previously defined details regarding the network structure and the enzymatic rate constants of the reactions that interconnect within the network. Especially in the setting of the study of single-celled prokaryotes (eg, bacteria) or eukaryotes (eg, yeast) where initial metabolite concentrations are readily measured, precise calculations of sensitivity coefficients reflect the predicted effects on the biochemical network after specific perturbations. Application of such modeling has been powerful when performed in the context of bacterial pathogenesis and has emphasized the importance of global metabolic alterations that contribute to bacterial growth, antibiotic susceptibility, and infective capabilities in humans. Modeling of metabolic networks in human tissue and human pathogenesis has been more challenging. First, complete metabolomic profiles have only recently become available from clinically relevant tissue samples, thus increasing the difficulty in inferring accurately initial or final concentrations of metabolites in a given condition. Second, the human network of metabolic reactions is more complex than that of single-celled organisms. Finally, owing to the complexity of overlapping pathways and inputs in human tissue, rate constants are variable and are subject to alterations in enzyme activity, expression, and time-dependent relationships. Nonetheless, various mathematical models have been developed to reduce the impact of these constraints, leading to significant insight into the impact of metabolic dysfunction in human disease.

Defining and applying the dynamic flux through gene regulatory networks presents an even greater challenge. Gene coexpression networks alone falter as compared to more sophisticated approaches to simulate reaction networks, which stems from several causes. First, as generated from expression array data, coexpression networks are relatively large and tend to encompass numerous subsets of molecular networks. Coexpression alone is typically insufficient to map the direct regulatory connections among all relevant genes, and thus gene regulatory maps tend to remain incomplete. Furthermore, unlike the previously described biochemical networks in which each connection can typically be described by a precise rate constant, connections functionally linking gene regulatory networks together have not been as precisely defined, as binding affinity of transcription factors to gene promoters do not necessarily always quantitatively reflect the relative level of transcriptional activation. Finally, modeling dynamic changes in gene expression generally requires stochastic, rather than deterministic, approaches given the frequently low concentrations of transcription factors and the limited number of promoter binding sites (as few as 2).

Despite these obstacles, methodologies have been established to infer better the exact nature of the relationships and the rudimentary dynamics of gene regulatory networks. One method involves applied human genetics analysis. Presently, chromosomal linkage mapping and GWAS are the most prevalent analytic methods used to reveal common and rare genetic mutations and polymorphisms linked with human disease. These are powerful methods by which to identify novel genetic associations with human disease, but, when used alone, these approaches are limited in their ability to prove a causal or directed pathogenic mechanism. In contrast, when coupled with an appreciation of where those genes lie in the interactome network, these genetic alterations can be integrated to infer the direction and dynamics of information involving associated factors throughout the network to determine causality among those molecular relationships. For example, when used in an iterative and systematic process to map how variations in DNA associate with relative transcript abundances in a relevant coexpression network, statistical algorithms have been successful in inferring an independent, causative, or reactive function of disease-gene candidates relative to a specific human disease phenotype. This approach has been used for the identification of 3 new genes in the susceptibility to human obesity and potentially could be applied to human mutations that segregate with overall cardiovascular health risk (Figure 4D). More broadly, inference of network dynamics through genetic study may help to...
Figure 4. Discovery of novel disease genes using network-based approaches. A, Identification of novel disease genes using network-based approaches. Adapted from Chan et al.167 with permission. B, A network-based strategy for identifying miRNA important in pulmonary hypertension (PH) elucidates a target network of 7 miRNA groups at the crucial intersection of hypoxia, inflammation, and TGF/BMP signaling. Such an approach also reveals genes that may represent points of coordinated miRNA regulation in PH. Circles indicate predicted gene targets; triangles, miRNA; blue lines, predicted associations of miRNA and targets; dotted gray lines, gene interactions documented in the consolidated interactome. Circle size is proportional to the number of miRNA groups (among these 7) predicted to target that particular gene. Node colors represent functional pathways (ie, endothelin signaling, mitochondrial metabolism, etc). Adapted from Parikh et al.29 with permission. C, A human “diseasome” displays molecular interconnections among separate categories of human disease. An edge denotes the involvement of a shared disease factor(s) or pathways common to multiple different diseases. Adapted from Goh et al.36 with permission. D, Schadt et al used a likelihood-based causality model selection test (LCMS), using...
answer the question of “missing heritability” in human traits. Specifically, despite efforts to identify additional genetic loci to explain the vastly differing patterns of heritability in human disease, recent computational analyses by Zuk et al\textsuperscript{56} suggest that variations in heritability need not involve missing genetic variants but could adequately be explained by complex variations in genetic interactions among different molecular pathways. Thus, these findings emphasize the intriguing potential of studying network dynamics to reveal a widespread “network-based” origin of disease heritability.

Separate from human genetic studies, an alternative methods by which to study network dynamics in human disease involves “reverse engineering” regulatory networks.\textsuperscript{57} In this approach, a discrete and typically small module of relevant genes is systematically perturbed with measurement of the consequent changes of their associated genes. The regulatory relationships among the genes can then be inferred, yielding novel predictions of direct regulatory connections that can be verified experimentally. Indeed, such processes have been validated in small sets of discrete gene networks active in organisms, such as bacteria,\textsuperscript{58} yeast,\textsuperscript{59} and \textit{Drosophila},\textsuperscript{60} as well as for mapping networks of small noncoding RNA in bacteria.\textsuperscript{61} This approach has recently been applied to human disease by Ergun et al, who inferred a gene regulatory network important in prostate cancer that led to validation of the androgen receptor, in the context of this network, as a marker to detect relatively aggressive primary prostate cancers (Figure 4E).\textsuperscript{62} The advantage of such reverse engineering is its unbiased approach to network mapping,\textsuperscript{63} as a preconceived framework of regulatory connections is not necessary; however, this approach is hampered by the technical challenges of applying the algorithms to larger and much more complicated regulatory networks. Additionally, because inferred relationships can only be depicted dichotomously as direct “activating” or “inhibiting” conditions, more complex network relationships that involve connections beyond transcriptional regulation (ie, posttranslational modifications, protein-protein oligomerization, etc) require different types of representation, and thus the accuracy of any simulation of network perturbation could suffer. Future iterations of analysis of the dynamics of gene regulatory networks may benefit from utilizing the known topology of the existing consolidated interactome coupled with statistical algorithms that incorporate the results of perturbations of genetic coexpression networks via genetic associations and/or alternative stimuli and thus iteratively annotate and optimize the regulatory map. Consequently, a more sophisticated and comprehensive network representation is possible that could be applied to further systems-level discoveries of how pathogenic or therapeutic stimuli may alter overall disease manifestation.

**Network Medicine and Drug Development (Systems Pharmacology)**

In addition to aiding discovery of novel molecular aspects of human disease, network medicine has the potential to revitalize the process of drug discovery.\textsuperscript{64} Historically, drug design has been based on key observations of pharmacological or genetic manipulation of putative targets or pathways that, when perturbed, alter a quantifiable (patho)phenotype. Unwanted pharmacological side effects have been attributed to an interaction of a medication with alternative factors in separate pathways. However, in complex human diseases, both therapeutic and adverse consequences of drug treatment may depend on shared effectors and pathways that, when manipulated in distinct combinations involving specific cellular or tissue contexts, may drive disparate pathophenotypes. It has been estimated that approximately 400 to 600 proteins of the presumed 100,000 participants in the consolidated interactome are targets of the 1200 Food and Drug Administration (FDA)-approved drugs\textsuperscript{65}; however, a large proportion of these established drugs do not target actual disease-associated proteins but merely neighboring proteins.\textsuperscript{66} Taken together, these drugs may only modestly affect disease phenotype. In addition, there likely exist a larger number of potential therapeutic targets that await identification.

Furthermore, despite the rapid advances in our knowledge regarding complex biological processes, pharmaceutical research and development productivity has been in decline since the mid-1990s (Figure 5A). In striking statistics drawn from the pharmaceutical industry (2000 to 2008), only \textasciitilde25 to 30 new molecular entities per year are identified as viable therapeutic agents. Clinical development of these compounds necessitates an average of 13.9 years, and new agents carry a probability of success of approximately 2.01\% per year.\textsuperscript{67} Reasons for such declining productivity include a more intense regulatory environment, the exhaustion of “easy” or more obvious targets, an increasing attrition rate for developing drugs secondary to their unanticipated side effects, and, from the perspective of this review, an intrinsically flawed reductionistic approach to drug development centered on the perceived need to identify a single drug target that entirely reverses a given disease phenotype (ie, an Erlichian “magic bullet”). Conversely, the failures of single-target drugs in treating complex human disease have already suggested their severe limitations when used alone and have prompted the exploration of potent combinatorial drug therapies for a variety of disease phenotypes (ie, tuberculosis, cancer, HIV/AIDS, heart failure, etc). Furthermore, given the advances in pharmacogenomics and improved understanding of how drugs may be differentially effective based on individual genotype, development of optimal drug regimens based on

---

**Figure 4 (Continued).** Quantitative trait loci (QTL) data, to infer relationships between RNA levels and complex traits that are best supported by the data. Possible relationships (M1-M5) are shown here among QTL (L), RNA levels of a gene (R), and a complex trait (C). Adapted from Schadt et al,\textsuperscript{65} with permission. \textit{E}, Strategy of reverse engineering to infer a gene regulatory network in nonrecurrent primary and metastatic prostate cancer. In phase 1, high-throughput microarray data obtained from cancer cell lines or patient samples are incorporated by the MNI algorithm to infer a model of regulatory interactions between genes (blue-filled circles indicate genes; arrows indicate regulatory effects). In phase 2, test condition expression data are filtered using the reconstructed network to distinguish the genes affected by a condition (red-filled circles) from the remaining genes that exhibit changes in expression. Adapted from Ergun et al,\textsuperscript{62} with permission.
their individualized effects is under further investigation and of premium priority. Rapidly evolving methods of "systems pharmacology" facilitate a more rational and efficient approach to identifying the most relevant "network-based" targets based on their presence in relevant disease modules (Figure 5B). Given the expansive small molecule libraries that have been mapped to direct targets through large scale chemical biology repositories, network theory presents the opportunity not only to discern previously unidentified targets for disease therapy and to map novel pathogenic pathways but also to predict how those targets can be manipulated, in isolation or in combination. A generalized methodology has recently been proposed for a systems-based method for drug discovery that involves defining a disease module and iteratively refining the map based on its emergent properties (Figure 5C). In doing so, drugs can be predicted and/or produced that rationally target key disease nodes followed by iterative refinement in silico that account for its effects on the emergent behavior of the disease module. Therefore, contrary to traditional drug design, systems pharmacology may offer elegant and cost-effective solutions to identifying multiple drug combinations ("rational polypharmacy") that maximize the therapeutic efficacy, minimize adverse side effects, and minimize drug-drug interactions. As a result, we may expect a much lower attrition rate as well as decreased time from drug conception to approval, and substantially higher rate of success for all compounds that enter clinical testing (provided regulatory bodies can adapt to this change in drug development strategy and clinical trial design strategies [adaptive] follow suit).

Identification and Targeting Novel Drug-Host Interactions

Application of systems pharmacology to the identification of novel drug-host interactions in complex human disease has been reported with success. To identify a larger pool of potential therapeutic targets in human disease that are not limited by preexisting knowledge of disease genes, systems pharmacology has been employed to generate a quantitative framework within which to compare and contrast diseases by an integrated analysis of disease-related mRNA expression data and the human protein interaction network. For example, Lamb et al have devised a process ("Connectivity Map") of identifying drug-dependent gene expression profiles that correlate with known disease-dependent profiles, thus allowing for a method for in silico screening of drugs that may have salubrious effects in multiple disease contexts. A variant of that approach has been pursued by Hu and Agarwal, in which gene expression data are not prospectively collected but can be directly obtained from an existing database, Gene Expression Omnibus. A "guilt by association" method has also been proposed by Chiang and Butte, whereby predictions of novel associations between drugs and diseases can be made by assuming that if two diseases are treated by the same drug, alternative drugs treating only one of them might also treat the other. More recent studies have attempted to make predictions on a broader scale. Suthram et al identified 4620 functional modules in the human protein network and provided a quantitative metric to record their responses in 54 human diseases leading to 138 significant similarities between diseases (Figure 6A). Interestingly, 14 of the significant disease correlations also shared common drugs, support-
ing the hypothesis that similar diseases can be treated by the same drugs, allowing for predictions for new uses of existing drugs. Finally, the authors identified 59 modules that are significantly enriched for genes that are known to be drug targets and are dysregulated in at least half of the studied diseases, representing a common disease-state “signature” and highlighting the importance of these core modules as prime therapeutic opportunities. Intriguingly, Gottlieb et al have described a novel method (PREDICT) for the large-scale prediction of drug indications. This approach is predicated on the idea that similar drugs are indicated for similar diseases. Thus, by quantifying and considering the known similarities among drugs (ie, chemical structure, predicted side effects, etc), networks of drug targets, and clinical pathophenotypic similarities, a more comprehensive compendium can be developed of all relevant drug indications for human disease. Consequently, these types of studies may form the computational foundation for future development of personalized therapies, in which variations in gene expression signatures and specific phenotypes from individual patients could be considered in order to dictate the course of care.

Advances in “repurposing” or “repositioning” existing drugs for alternative diseases have also benefited from network methodology. High-throughput methods of literature curation have proven effective in defining novel drug-disease connections that can guide repurposing of known medica-
tions. For example, Li et al have constructed disease-specific (eg, Alzheimer disease, in this case) drug-protein connectivity maps derived solely from literature database searches as a way to uncover a more comprehensive set of known drugs acting in the same disease context. More recently, commercial bioinformatics companies specializing in the analysis of scientific literature, such as Ingenuity, GeneGo, and Adriane Genomics, have offered proprietary pathway- and network-based solutions to the pharmaceutical industry for drug repurposing. Beyond literature curation, other groups have further developed a systematic computational approach to predict novel therapeutic indications on the basis of comprehensive testing of molecular signatures in drug-disease pairs. For instance, Siroti et al integrated gene expression measurements from 100 diseases and gene expression measurements on 164 drug compounds, yielding predicted therapeutic potentials for these drugs. Previously validated drug and disease relationships were predicted, as well as a number of novel indications for these 164 drugs. Importantly, one such prediction was experimentally validated for the antiulcer drug cimetidine as a candidate therapeutic in the treatment of lung adenocarcinoma, demonstrating its efficacy in vivo using mouse xenograft models. In a separate set of experiments, another prediction regarding the potential for the anticonvulsant topiramate in treatment of inflammatory bowel disease was successfully validated in a rodent model of inflammatory bowel disease. As a result, the use of existing gene coexpression networks in combination with the known molecular interactions of existing drugs has yielded more comprehensive insight into the pleiotropic effects of various medicinal compounds.

A variation on the theme of repurposing drugs includes the notion of rationally “rewiring” molecular networks such that cells resistant to certain medications can become responsive to pharmacological exposure. Recently, Lee et al undertook such a project, utilizing an expansive systems-level approach using comprehensive time-dependent measurements of signaling networks, gene expression profiles, and cell phenotypic responses in combination with computational network modeling of breast cancer cells resistant to genotoxic drugs. In doing so, the authors constructed and validated a molecular roadmap to reprogram sequentially oncocentric signaling pathways and resensitize breast cancer cells to those drugs. Such results indicate the feasibility of not only studying, but more importantly manipulating, large-scale network dynamics toward a desired cellular phenotype.

Perhaps the most striking successes of network pharmacology to date are derived from studies of bacterial metabolism. Owing to the fine detail of both the topology and rate constants connecting metabolic networks in the bacterial interactome, in silico predictions have been possible regarding the effects on metabolite flux and consequent phenotypic alterations after perturbation of specific nodes in those pathways. As a result, by combining phenotypic and genetic experiments with microarray analyses to identify relevant bacterial networks important in antibiotic response, Kohanski et al have demonstrated that all classes of bactericidal antibiotics, regardless of their specific target, promote the generation of lethal hydroxyl radicals. Kohanski et al have further perturbed these bacterial networks leading to mistranslation and misfolding of membrane proteins central to aminoglycoside-induced oxidative stress and cell death, thus expanding our understanding of the common mechanism of killing induced by known bactericidal antibiotics (Figure 6B). Furthermore, potent new antimicrobial agents have been identified and validated de novo based on their predicted system-based metabolic responses on bacterial survival.

Our understanding of systems-wide metabolic alterations in diseased mammalian cells has also led to novel potential therapeutic targets. For instance, by mapping the dynamic fluxes of labeled metabolites in mammalian cells, Munger et al noted an especially prominent increase in fatty acid biosynthesis during human cytomegalovirus (HCMV) infection. Subsequent pharmacological targeting of fatty acid biosynthesis suppressed the replication of HCMV and influenza A, thereby confirming that fatty acid synthesis is essential for the replication of two divergent viruses and, importantly, that systems-level metabolic flux profiling can identify metabolic targets for antiviral therapy.

In the field of cancer biology, Otto Warburg first described a half a century ago that tumors exhibit enhanced glycolysis coupled with a reduction in oxidative phosphorylation, despite adequate oxygen availability. This “Warburg effect” was proposed to be a key driver of tumor progression. Since that initial observation, we now recognize that there exists a systems-wide “reprogramming” of metabolic pathways associated with alterations in bioenergetics and mitochondrial function in transformed cells; for this reason, metabolic interventions are now emerging as potential therapeutic antineoplastic strategies. Accordingly, the fine detail of systems-level profiling available in cancer biology has allowed for a growing number of network-based studies reporting novel metabolic pathways important in driving cellular transformation and thus identifying crucial metabolic targets for pharmacological repression. Conversely, in the future, pharmacological rescue of metabolic dysfunction could be attempted in genetic metabolic diseases as a novel alternative to gene therapy.

Identification of “Off-Target” Effects and Drug-Drug Interactions

In addition to guiding the rational targeting of specific pathways in complex diseases, network-based approaches have better informed us about the systems properties of a full drug interaction network that encompass currently unwanted but otherwise unanticipated (so-called “off-target”) effects and drug-drug interactions. For example, Yeh et al created a network map of drug-drug interactions using a unique bioluminescence technique to provide quantitative measurements of pairwise interactions among 21 antibiotics that affect growth rate in Escherichia coli. Interestingly, the investigators reported that the drug interaction network revealed 2 classes of drug that interact either purely synergistically or purely antagonistically—network properties that correspond directly to the cellular functions affected by the drugs. This network approach introduced a new conceptual framework for understanding the functional mechanisms of drugs and their cellular targets, which could be expanded to other
FDA-approved medications or small molecule libraries. Interestingly, Torella et al have extended these studies of antibiotic synergy by utilizing mathematical modeling of the in vivo dynamics of bacterial infection to determine the optimal strategies for utilizing synergism and antagonism to suppress better the evolution of multi-drug resistance.\(^9\)

Analysis of the consolidated interactome can yield even greater insight into these pleotropic and complicated molecular interactions among various drugs. Recently, such a pharmacological approach was applied to the long-QT syndrome (LQTS) to determine whether a network-based analysis could predict arrhythmic side effects of known FDA-approved medications (Figure 6C).\(^9\) It is already established that mutations in any of 12 genes encoding ion channels are linked with congenital LQTS and potentially fatal ventricular arrhythmias, such as torsades de pointes. In an elegant study, Berger et al reasoned that drugs targeting proteins located in the network “neighborhood” of bona fide LQTS genes should also reveal arrhythmic side effects. On construction of a protein-protein interactome, a LQTS “neighborhood” of genes was identified by use of a mean first passage time algorithm to reveal proteins that interact directly or with close neighbors of seed LQTS genes. Importantly, the known targets of certain FDA-approved drugs that are associated with LQTS as an adverse event were enriched in the LQTS neighborhood. By screening the remaining cache of FDA-approved medications for those that target the LQTS neighborhood, additional drugs were identified that probably influence QT length. These predictions were then validated by cross-referencing with the Adverse Events Reporting System of the FDA (FDA-AERS). Thus, identification and utilization of disease-selective neighborhoods within the human interactome can provide specific insight into predicting adverse event susceptibility for new and “repurposed” drugs. Recently, Tatonetti et al have developed comprehensive, network-driven databases of drug effects and of drug-drug interaction side effects and demonstrated novel predicted adverse effects among cardiovascular drug classes.\(^10\)

Alternatively, network approaches offer substantial insight into the elucidation of the unanticipated “off-target” effects of otherwise presumed direct targeted therapies. Pharmacological examples of this scenario are plentiful, such as the well-documented clinical failure of the cholesteryl ester transfer protein (CETP) inhibitor torcetrapib.\(^10\) Despite its efficacy in augmenting HDL levels and reducing triglyceride levels, the use of torcetrapib surprisingly incurred risk of substantial hypertension, prompting a halt to its development following phase III clinical trials.\(^102,103\) Based on in silico predictions of protein-ligand binding, Xie et al identified a panel of off-targets for torcetrapib and other CETP inhibitors from the human structural genome and mapped those targets to a network of biological pathways curated from the literature.\(^101\) The predicted protein-ligand network was consistent with experimental results revealing that the side effect of hypertension by torcetrapib could be predicted by its targeting of the peroxisome proliferator-activated receptor (PPAR) system. Importantly, second-generation CETP inhibitors are not predicted to have the same magnitude of action on the PPAR system and thus a lower likelihood of inducing hypertension (Figure 6D). Furthermore, using these networks as a guide, the adverse drug effects of second generation CETP inhibitors could be minimized by fine-tuning multiple off-target interactions using single or multiple therapies.

Similarly, network approaches may also be effective in predicting drug-drug interactions. Hansen et al have described their preliminary experience with studying the antidiabetic thiazolidinedione, rosiglitazone, which has been withdrawn from the market due to serious adverse side effects.\(^6\) According to the FDA-AERS, rosiglitazone is associated with significant thrombotic complications as well as peripheral edema and total body volume overload. Consequently, the authors have constructed protein-protein interaction networks based on Gene Ontology terms that relate to thrombosis and fluid regulation. To determine the degree of overlap, these subnetworks were then cross-referenced with direct gene targets of PPAR\(\gamma\), a well-established direct target of rosiglitazone. Via utilization of FDA-AERS, drugs were then identified, which, when used in combination with rosiglitazone, should increase or decrease associated complications of thrombosis or edema by analyzing alterations in the effectors shared as targets of rosiglitazone and as members in the above subnetworks. These authors continue to develop algorithms to predict whether combinations of drugs may target the same pathway and whether they could potentiate therapeutic or adverse consequences.

Conclusions

Network medicine offers insight into the integrative nature of human pathogenesis by considering the dense interconnections among disease genes rather than merely considering those genes in isolation. It allows for novel predictions of disease genes, the interconnected actions of those genes in pathogenesis, the molecular relationships among separate disease conditions, and the concerted response to therapeutic interventions. Although still in its early stages, this evolving discipline offers a potentially powerful set of concepts and tangible methods by which to improve rapidly our molecular understanding of the networked nature of human disease and to design medical therapies rooted in the complexity of these biological connections. Nonetheless, despite benefiting from the overdetermined property of biological networks, the fidelity of network medicine is limited by the currently incomplete nature of the interactome maps. Yet, while many predictions derived from network methodology await confirmation, specific fundamental network principles have been consistently supported by independent experimental proof in human and animal models of disease. Such methods are even beginning to permeate the commercial biotechnology sector, as exemplified by companies such as Proteostasis (www.proteostasis.com), which was founded on the concept of using systems pharmacology to target efficiently the “proteasome” gene network. In the very near future, we expect a more widespread expansion of network methodology in biomedicine. Consequently, “systems pathophysiology” may become more of a reality as we explore not only intracellular molecular networks but also the even less well understood spatio-temporal interconnections among distinct cells, tis-
A Network-Based Definition of Human Disease

Figure 7. Network-based classification of human disease. In this model, a specific disease module (white nodes), associated with myocardial ischemia (module 1) and containing a known disease gene (arrow), is functionally linked to 2 disease modules (modules 2 and 3; gray nodes) by shared nodes (black) and edges. These shared nodes may regulate intermediate (patho) phenotypes important in other disease processes, such as inflammation or thrombosis. Influenced by a combination of environmental and genetic events that also regulate these disease networks, individualized manifestations of myocardial dysfunction result (early myocardial infarction, ventricular tachycardia, or ischemic cardiomyopathy). By incorporating an understanding of network function in these biological circuits, these manifestations may be better predicted, treated, and prevented. Adapted from Chan et al,107 with permission.

Finally, by integrating the dense interconnections among disease networks, a new framework has been proposed to reclassify human disease and thus allow for a more expansive consideration of the pathogenic mechanisms that are truly at play (Figure 7).106 By incorporating the relationships of individual disease genes, their connections within a disease module, and that module’s shared links with separate disease modules, insights can be garnered regarding the coordinated actions of several disease modules as well as specific molecular links that represent “intermediate” phenotypes (endophenotypes) shared among various disease conditions (ie, inflammation, thrombosis, etc). Furthermore, genetic, environmental, and even behavioral “nodes” can be represented accurately by this approach. As a result, this type of disease classification could predict a highly individualized clinical pathophenotype that ultimately represents a more accurate molecular schema on which to base future discovery in personalized medicine.

Acknowledgments
We thank Stephanie Tribuna for expert administrative assistance and Sol Chan for assistance with figures.

Sources of Funding
This work was supported in part by the NIH, the Pulmonary Hypertension Association, Gilead Sciences, and the Lerner, Harris, and Watkins funds (S.Y.C.); and NIH grants HL061795, HL48743, HL107192, HL70819, and HL108630 (J.L.).

Disclosures
None.

References


The Emerging Paradigm of Network Medicine in the Study of Human Disease
Stephen Y. Chan and Joseph Loscalzo

Circ Res. 2012;111:359-374
doi: 10.1161/CIRCRESAHA.111.258541

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/111/3/359

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/