thematic review

Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders

Reverse engineering gene networks to identify key drivers of complex disease phenotypes

Eric E. Schadt[1] **and Pek Y. Lum**

Rosetta Inpharmatics, LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA 98109

Abstract Diseases such as obesity, diabetes, and atherosclerosis result from multiple genetic and environmental factors, and importantly, interactions between genetic and environmental factors. Identifying susceptibility genes for these diseases using genetic and genomic technologies is accelerating, and the expectation over the next several years is that a number of genes will be identified for common diseases. However, the identification of single genes for disease has limited utility, given that diseases do not originate in complex systems from single gene changes. Further, the identification of single genes for disease may not lead directly to genes that can be targeted for therapeutic intervention. Therefore, uncovering single genes for disease in isolation of the broader network of molecular interactions in which they operate will generally limit the overall utility of such discoveries. Several integrative approaches have been developed and applied to reconstructing networks. Here we review several of these approaches that involve integrating genetic, expression, and clinical data to elucidate networks underlying disease. Networks reconstructed from these data provide a richer context in which to interpret associations between genes and disease. Therefore, these networks can lead to defining pathways underlying disease more objectively and to identifying biomarkers and more-robust points for therapeutic intervention.—Schadt, E. E., and P. Y. Lum. **Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J. Lipid Res.* 2006.** 47: **2601–2613.**

Supplementary key words systems biology • networks • genetical genomics

With the completion of the sequencing of genomes from multiple species, the challenge in the life and biomedical sciences now is to decipher the biological function of individual genes, pathways, and, more generally, biological networks that drive complex phenotypes, including common human diseases. The identification of single genes for common diseases has greatly accelerated over the past several years. With access to the complete genome sequence for a diversity of species, large-scale haplotype maps, technologies capable of screening DNA polymorphisms and gene activity on an unprecedented scale, and well-characterized human cohorts, genes explaining an appreciable risk for a number of common human diseases have been identified. Notable examples are TCF7L2, a major disease gene for common forms of type 2 diabetes (1, 2); INSIG2, a major obesity gene potentially explaining 4% of lifetime body mass index (BMI) in the human population (3); CFH, one of the more striking discoveries for age-related macular degeneration, where a number of sequence variations in complement factor H have been found to be strongly associated with this disease in a number of human studies (4–8); and ALOX5, a gene identified in human and mouse populations that predisposes to a number of disease-related traits, including atherosclerosis (9, 10), hyperlipidemia-dependent aortic aneurysm (11), and obesity and bone phenotypes (12). Although these examples represent only a handful of the discoveries that have been made in recent years, they highlight how leveraging large-scale, high-throughput genetic and functional genomic technologies, in addition to well-characterized animal and human populations, can lead directly to the identification of key drivers of disease.

However, despite the identification of a number of novel disease-predisposing genes, progress in uncovering the mechanisms by which these genes lead to disease has been far slower. Even in cases in which genes validating as causal for disease are known to operate in what are thought to be well-understood pathways, it is often unclear whether the connection to disease regarding such genes involves the known pathways, whether these "known"

Manuscript received 11 September 2006.

Published, JLR Papers in Press, October 1, 2006.
DOI 10.1194/jlr.R600026-JLR200

Abbreviations: BMI, body mass index; eQTL, expression quantitative trait loci; QTL, quantitative trait loci.

[1] To whom correspondence should be addressed.
e-mail: eric_schadt@merck.com

pathways are more general than is presently known, or whether the disease-associated genes operate in multiple pathways, some of which are yet to be defined. An example is the gene Tgfbr2, a key component of the transforming growth factor-β signaling pathway, that involves only a modest number of proteins, but whose expression in the liver of mice from an F2 intercross population was shown to associate with thousands of other genes ostensibly unrelated to this classic signaling pathway (13, 14). The gene was subsequently identified and validated as causal for obesity in a segregating mouse population (13), but how variations in this gene lead to obesity is not yet understood.

As the drive to understand the context in which disease-causing genes operate and as ever-bigger data sets monitoring large-scale molecular activity at unprecedented scales increase, it is becoming generally accepted that many biological functions, which are often system and context dependent, need to be studied at the systems level in addition to studying gene function at the level of individual pathways (14, 15). To provide the proper context in which to interpret single-gene discoveries, a systems biology approach is needed, and such approaches will be successful only if they accurately reflect the biological states underlying disease. This type of view can only be achieved via the integration of a diversity of data informing on the complex system under study. Toward this end, there has been significant effort applied to reconstructing and characterizing biological networks based on a diversity of biological data.

After motivating the need to take a systems biology approach to dissecting complex disease traits, we review a number of approaches that have been recently developed to elucidate gene networks associated with complex traits such as common human diseases. Coexpression networks leverage pair-wise interaction data among genes to represent more-general relationships among genes in a comprehensive fashion. Central to this type of network is the manifestation of modules comprised of highly interconnected sets of genes that ostensibly form the functional units of the networks that are associated with complex phenotypes such as disease. To get at causal relationships among genes, and between genes and clinical phenotypes, simple networks comprised of only a handful of genes, clinical phenotypes, and genetic loci are also reviewed. By integrating genetic, expression, and clinical phenotype data, the resulting networks are capable of representing direction along the edges of the network (unlike coexpression networks, which are undirected networks), where directionality among the edges provides the source of causal information used to establish causal relationships both among genes and between genes and clinical phenotypes. Bayesian networks that incorporate genetic and expression data are then reviewed as a generalization of this simpler approach, and again provide for the ability to represent causal relationships both among genes and between genes and clinical traits. We discuss key concepts emerging from patterns of network structure, such as changes in response to different environmental conditions, motivating the need for statistics based on networks that go beyond the single-gene measures of differential expression that have dominated the microarray community since its inception. Finally, we illustrate how networks provide a far richer context within which to characterize genes found to be causal for disease, ultimately providing a framework for identifying key intervention points that can be targeted for disease. These emerging high-dimensional data analysis approaches highlight that evolving statistical procedures on networks will be critical to extracting information related to complex phenotypes such as disease, as research goes beyond the single-gene focus.

A SYSTEMS BIOLOGY APPROACH TO ELUCIDATING DISEASE TRAITS

Diseases such as atherosclerosis, hypertension, obesity, and other such common human diseases involve multiple tissues potentially signaling in complicated and as yet to be defined ways. As highlighted in **Fig. 1**, the GI tract, vasculature, immune system, heart, and brain are all potentially involved in either the onset of diseases such as atherosclerosis or in comorbidities such as myocardial infarction and stroke brought on by such diseases. Further, the risks of comorbidities for diseases such as atherosclerosis are increased by other diseases, such as hypertension, which may, in turn, involve other organs, such as kidney. The role that each organ and tissue type plays in a given disease is largely determined by genetic background and environment, where different perturbations to the genetic background (perturbations corresponding to DNA variations that affect gene function, which, in turn, leads to disease) and/or environment (changes in diet, levels of stress, level of activity, and so on) define the subtypes of disease manifested in any given individual.

Although the physiology of diseases such as atherosclerosis is beginning to be better understood, what has not been fully exploited are the vast networks of molecular interactions at play within the cells comprising tissues related to disease. As shown in Fig. 1, there is a diversity of molecular networks functioning in any given tissue, including genomics networks, networks of coding and noncoding RNA, protein interaction networks, protein state networks, signaling networks, and networks of metabolites. Further, these networks are not acting in isolation within each cell, but instead interact with one another to form complex, giant molecular networks within and between cells that drive all activity in the different tissues, as well as signaling between tissues. Variations in DNA and environment lead to changes in these molecular networks, which, in turn, induce complicated physiological processes that can manifest as disease.

Despite this vast complexity, the classic approach to elucidating genes that drive disease has focused on single genes or single linearly ordered pathways of genes thought to be associated with disease. This narrow approach is a natural consequence of the limited set of tools that were

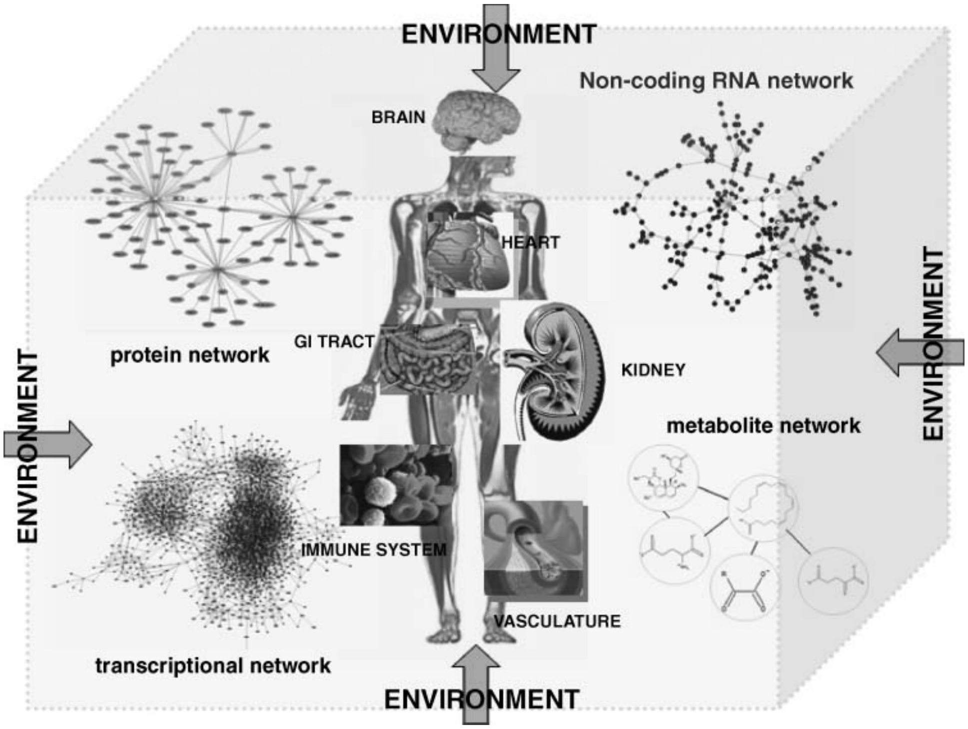

Fig. 1. Diseases such as atherosclerosis and hypertension comprise a diversity of different disease subtypes involving multiple organs and tissue types. Operating within each tissue (and each cell within a given tissue) are a number of molecular networks that ultimately drive the onset of disease. These networks are context specific and sensitive to internal and external environmental conditions as well as genetic background. Variations in the connectivity structure of these networks are induced by variations in the genetic background and environmental conditions, where these variations in turn lead to phenotypic variations, including disease. Studying the molecular networks in all relevant tissues and associating them with clinically relevant phenotype data to identify the networks driving disease are among the goals of systems biology applied to disease research. By taking a more holistic approach, it may be possible to better understand the complex interplay among tissues, molecular networks, and environment that leads to disease.

available for querying biological systems; such tools were not capable of enabling a more holistic approach, resulting in the adoption of a reductionist approach to teasing apart pathways associated with complex disease phenotypes. Although the emerging view that complex biological systems are best modeled as highly modular, fluid systems exhibiting a plasticity that allows them to adapt to a vast array of conditions, the history of science demonstrates that this view, although long the ideal, was never within reach, given the unavailability of tools adequate to carrying out this type of research. The explosion of large-scale, high-throughput technologies in the biological sciences over the past 15 to 20 years has motivated a rapid paradigm shift away from reductionism in favor of a systems-level view of biology (14, 15). With tools now available to take comprehensive looks at entire systems, success in biomedical research in the future will demand a more comprehensive view of the complex array of interactions driving biological systems, including how such interactions are modulated by genetic background, infection, environmental states, lifestyle choices, and social structures more generally (16, 17). This holistic view requires an embracing of complexity in its entirety, where the emerging view of complex systems is one of dynamic, fluid systems able to reconfigure themselves as conditions demand (18–20). Central to the study of complex, integrated systems is the concept of a network as a way of representing the extensive interactions among the different components of a system.

INTERACTION NETWORKS AS A WAY TO ORGANIZE AND CHARACTERIZE DISEASE NETWORKS

Networks provide a convenient framework for representing high-dimensional data in which relationships among the many variables making up such data are the key to understanding the properties that emerge from the complex systems they represent. Networks are simply graphical models comprised of nodes and edges. For gene networks associated with biological systems, the nodes in the network typically represent genes, and edges (links) between any two nodes indicate a relationship between the two corresponding genes. For example, an edge between two genes may indicate that the corresponding expression traits are correlated in a given population of interest (21), that the corresponding proteins interact (22), or that changes in the activity of one gene lead to changes in the activity of the other gene (13). Interaction or association networks, formed by considering only pair-wise relationships between genes, including protein interaction relationships (18), coexpression relationships (23, 24), and other straightforward measures that may indicate association between two genes, have recently gained more widespread use in the biological community.

Barbasi and Albert (25) were among the first to describe key topological features of complex networks in biological systems, and since then, several groups have characterized the topological properties of biological networks and elucidated the plasticity of these networks on the basis of protein interaction, gene deletion lethality, and metabolomic and transcriptome data (18, 19, 25–27). More recently, this type of approach has been applied to coexpression data in yeast (28), mouse (24, 29), and human (30). The key features that emerges from all studies are that coexpression networks across all species examined to date are scale free and hierarchical. The scale-free property basically implies that of all the nodes in the network, most nodes are connected to relatively few nodes, whereas relatively few nodes are highly connected to many other nodes, giving rise to the concept of hub nodes that potentially represent key information control points in the network. The hierarchical property implies that nodes in the network cluster into modules of highly interconnected genes that are not as highly connected with genes outside of the module.

One way to visualize such networks is as a topological overlap map (**Fig. 2A**). Originally described by Barabasi and Oltvai (16), this type of plot represents the connectivity structure of the overall network. The hierarchical structure of the network is well highlighted by this type of plot, in which genes assemble into coherent modules (the blocks along the diagonal in Fig. 2A) that define the functional components of the network (24). These functional units of the network can then be seen to associate with pathways and disease states, providing insights into those parts of the network that may be driving particular subtypes of disease (23, 24, 28). One of the modules highlighted in the Fig. 2A topological overlap map, generated from liver samples in a previously described cross (31, 32), is enriched for genes associated with lipid and cholesterol

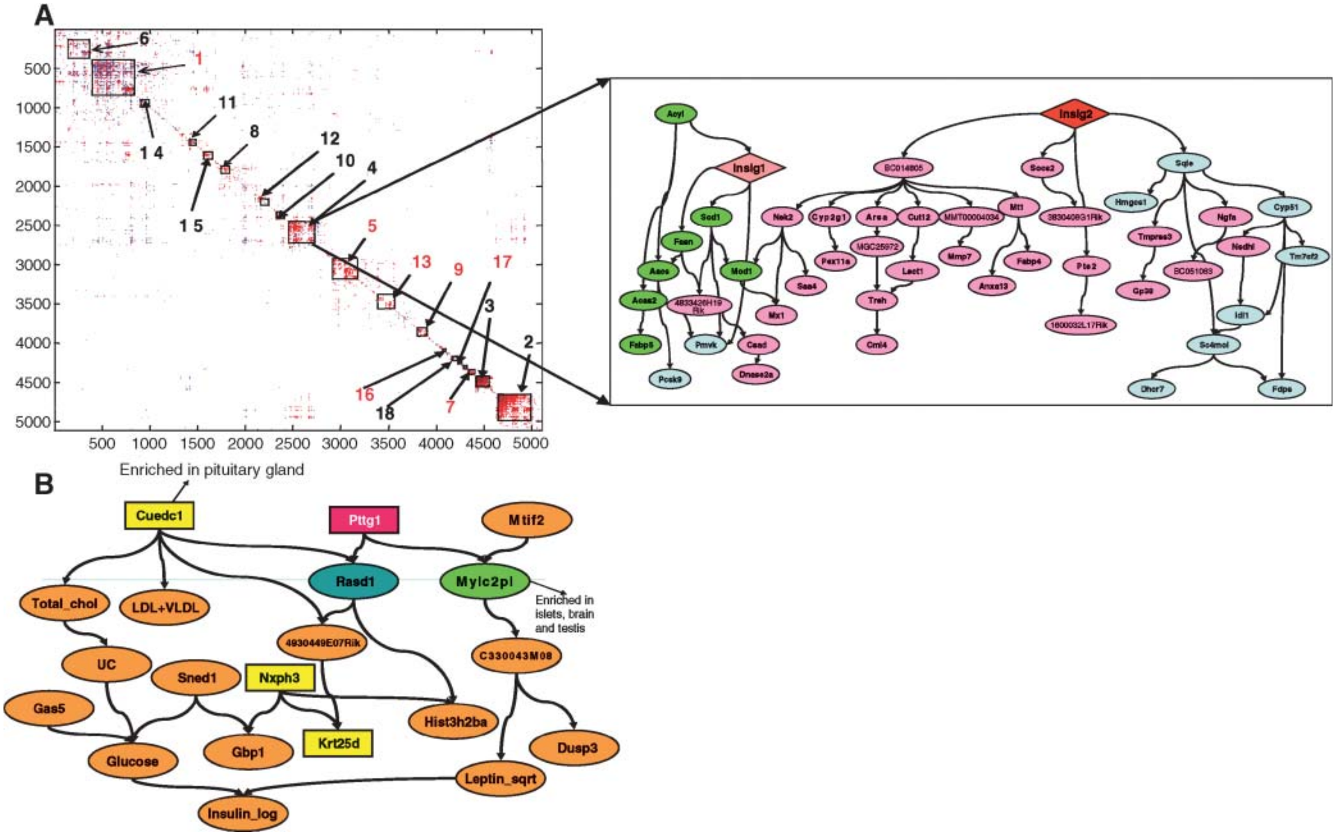

Fig. 2. Coexpression and Bayesian networks constructed from segregating mouse populations. A: The left panel represents a topological overlap map of the liver tissue from female mice in the BXH cross (24, 31, 32) constructed using previously described methods (56). The plot represents 5,000 of the most highly connected genes in the liver tissue of the BXH cross, with red and blue indicating positive and negative correlation, respectively, between the corresponding genes, and white indicating absence of correlation at some prespecified correlation threshold. Genes along the x- and y-axes are clustered according to similarity of correlation measures, which highlights the modular structure of the network, given the appearance of the blocks along the diagonal. The right panel represents an Insig1/Insig2-specific subnetwork from a previously published Bayesian network constructed from the liver data in the BXH cross (33). The genes in this part of the Bayesian network are most significantly enriched for genes in module 4 of the topological overlap map (roughly 60% of the genes in the Bayesian network fall into module 4). B: A previously published Bayesian subnetwork from brain gene expression data in the BXH cross. Here the gene expression traits and clinical traits are considered as nodes in the network.

metabolism. In fact, this module is enriched for genes comprising a previously published network involving an Insig2 subnetwork derived from the same cross, where Insig2 was mapped as a susceptibility gene for cholesterol levels as well as other metabolic traits, including obesity (Fig. 2A) (33). As highlighted in this figure, genes in the network are associated with cholesterol metabolism (light blue nodes) as well as lipid synthesis (green nodes).

This example highlights that the identification of key modules in the network involves known biological pathways and is linked to disease traits, representing the utility of coexpression networks, with respect to the dissection of complex disease traits. Coexpression networks provide a gross characterization of the connectivity properties of the network, thereby organizing vast amounts of data into logical units. For example, the plot shown in Fig. 2A represents nearly 12.5 million correlation measures taken from 5,000 gene expression traits monitored in over 150 liver samples. In contrast to cluster analysis (34), a more routine way of assessing the similarity of expression traits over different genes and experiments, the coexpression networks provide more structure information on the connectivity of the network, as well as highlighting modules of highly interconnected genes, which themselves would cluster closely together in a cluster analysis; but the module is a more highly purified structure with respect to relationships among genes, focusing mainly on genes that are very highly correlated and highly interconnected, with a more limited number of genes falling within a cluster. As discussed below, the highly correlated, highly connected structures within a network module, as well as the weak links that tie these different structures together, highlight potential new strategies for exploring how to target such networks associated with a given disease in order to treat the disease effectively.

DIFFERENTIAL CONNECTIVITY AS A KEY STATISTICAL MEASURE IN BIOLOGICAL NETWORKS

One of the key concepts emerging from the study of networks such as the coexpression network depicted in Fig. 2A is that of differential connectivity with respect to different phenotypic groups within populations that have been profiled. Given tissue samples from two phenotypic groups of individuals (say, those with a given disease vs. those without), a gene is considered differentially connected between the two groups if the number of genes to which the gene is significantly correlated is significantly different. Although this concept has not yet been formally defined in statistical terms in the literature, there have already been striking examples of large-scale changes in the connectivity structure among molecular phenotypes monitored in the same system under different environmental conditions. For example, Luscombe et al. (19) demonstrated that in response to different environmental conditions, portions of the yeast transcriptional network were subject to extensive rewiring. It may be that this type of rewiring at the transcription and/or protein network level may induce dramatic changes in physiological processes as well, as was evidenced by the significant rewiring of arcuate nucleus feeding circuits observed in response to leptin (20).

Coexpression networks, compared with protein interaction networks, in which protein interactions are typically assessed in yeast two-hybrid systems that do not necessarily represent the natural context in which protein interactions normally occur (22), are arguably more coherent, given that the state of the 25,000 or so genes from which such networks obtain are simultaneously measured in each tissue and then monitored in a large number of individuals in a given population, enabling the detection of weak links as well as strong links in the network. As a result, the detection of genes with significantly different connectivity patterns may be more easily achieved in this setting. As an example, the liver expression data published on a previously described cross between B6 and C3H mice on an ApoE-null background (referred to here as the BXH cross) gave rise to striking differences in liver expression between female and male mice, indicating widespread sexually dimorphic patterns of gene expression (32). Given these differences, one could imagine that differential connectivity differences would also manifest themselves in this setting, and this is indeed the case. If these previously published data are examined, it is easy to identify modules in the liver coexpression network that are highly connected in one sex but not the other. **Figure 3A** highlights a set of genes in male mice from the BXH cross that are significantly correlated with many other genes, which themselves are highly interconnected. Figure 3B highlights this very same set of genes for females, where it is quite clear from the image that the genes that were highly interconnected in the males are not highly correlated with these same genes nor are they highly interconnected in the female mice. In fact, only roughly 10% of the genes that were significantly correlated in males are significantly correlated in the females, as shown in the box highlighted in Fig. 3.

One of the most interesting observations related to this pattern of differential connectivity is that genes that are differentially connected between two groups, as shown in Fig. 3, are not necessarily differentially expressed between the two groups. In fact, from the region highlighted in Fig. 3, only 60% of the genes defined in the indicated region were identified as differentially expressed between males and females (32). This suggests that there are whole classes of genes whose expression is not varied between two phenotypic groups of interest, but that nevertheless get wired into the network very differently between the two groups. These observations are particularly important in differences between the sexes in the context of disease, given that striking differences between sexes have been observed for a diversity of complex phenotypes, including obesity (31), diabetes (35), atherosclerosis (36, 37), behavior (38), and drug response (39, 40), among many other phenotypes.

Whether differences in connectivity are the result of genetic, hormonal, or other environmental differences between the sexes, changes in how a gene gets wired into a network is an important concept. In fact, how a gene

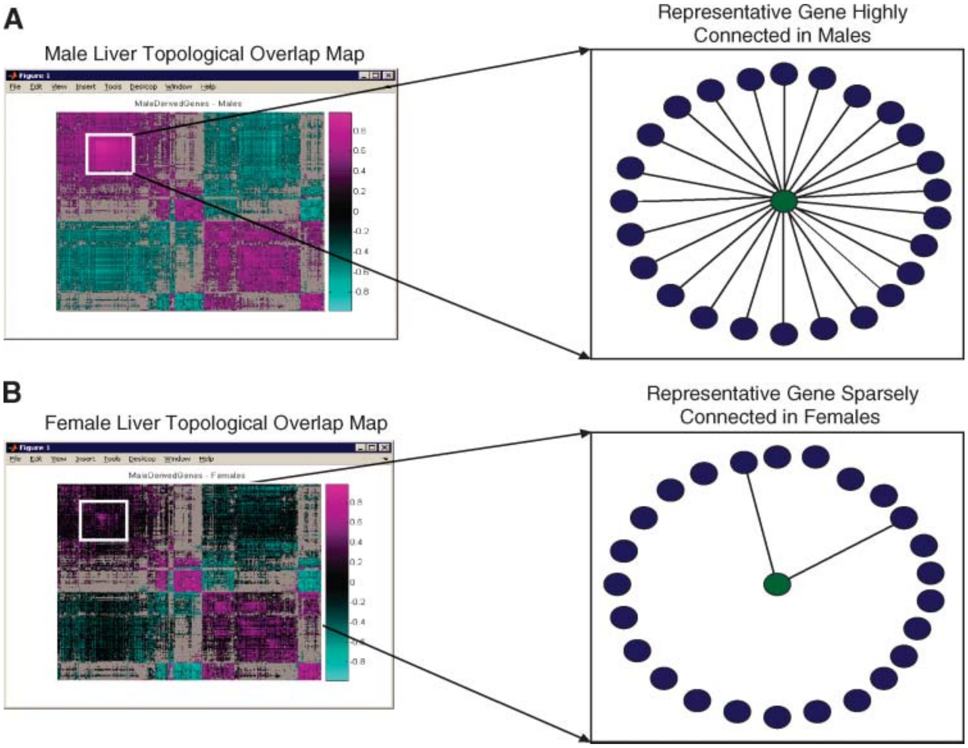

Fig. 3. Differential connectivity between males and females in the BXH liver gene expression data. A: On the right is a portion of a topological overlap map from the male BXH liver data, comprised of a set of genes identified as highly interconnected in males, but not females. Color scale on the right side of the left panel indicates degree of correlation. Depicted on the right side is a representative gene (green node) from the set of highly interconnected genes falling in the indicated white box, where the representative gene is connected to many other genes. B: The same plot as in A, but for females, where the gene order is the same as that determined for males in A. Despite having more animals in the female group (162 vs. 158), the correlations in the module highlighted by the white box are very significantly reduced compared with males. The representative gene shown in A is highlighted at left (also green node) and is seen in the female liver data to be connected to fewer than 10% of the genes connected to this gene in the males.

gets wired into different biological networks within and between different tissues may turn out to be at least as important as differential regulation between disease and nondisease groups, with respect to elucidating how networks drive disease. Certainly this type of idea is worthy of further pursuit, and, in fact, it raises awareness that down the road, in order to be effective at leveraging networks to elucidate disease, statistical methods need to address data at the level of the network as opposed to the level of single genes varying among groups. Considering large sets of genes to increase power to make inferences on single genes has been beautifully leveraged in current analysis methods (41–43). However, what will be needed in the future are methods that leverage this type of information to make inferences on whole networks of genes associated with different phenotypic states in systems of interest.

INTEGRATIVE GENOMICS APPROACHES TO INFERRING CAUSAL NETWORKS

Coexpression networks are informative as discussed for gross characterizations of the properties of biological networks, identification of highly connected (hub) nodes, and identification of functional modules that aid in the characterization of subnetworks associated with disease. Despite these and other advantages, coexpression networks do not provide explicit details on the connectivity structure among genes in the network, including the representation of causal associations among genes and between genes and phenotypes. Causal associations among genes or between genes and traits have classically been established using time series experiments, gene knockouts, or transgenics that overexpress a gene of interest, RNAi-based knockdown or viral-mediated overexpression of genes of interest, and chemical activation or inhibition of genes of interest. However, recently, a number of studies have demonstrated that establishing such causal associations is now possible by leveraging naturally occurring variations in DNA, given that gene expression and other molecular phenotypes in a number of species have been shown to be significantly heritable and at least partially under the control of specific genetic loci (44–54). By examining the effects that naturally occurring DNA have on variations in gene expression traits in human or experimental populations, other phenotypes (including disease) can be examined with respect to these same DNA variations and ultimately ordered with respect to genes to infer causal control (**Fig. 4**) (12, 13, 55, 56).

It is important to note here that when we use the term causality in the present context, it is perhaps meant in a more nonstandard sense than that to which most researchers in the life sciences may be accustomed. In the molecular biology or biochemistry setting, claiming a causal relationship between, say, two proteins probably implies that one protein has been determined experimentally to physically interact with or to induce processes that directly affect another protein, and that this in turn leads to a phenotypic change of interest. In such instances, the causal factors relevant to this activity are known, and careful experimental manipulation of these factors subsequently allows for the identification of genuine causal relationships. However, in the present setting, the term causality is used from the standpoint of statistical inference, where statistical associations between changes in DNA, changes in expression (or other molecular phenotypes), and changes in complex phenotypes like disease are examined for patterns of statistical dependency among these variables that support directionality among them, where the directionality then provides the source of causal information (highlighting putative regulatory control as opposed to physical interaction). The gene networks described here, therefore, are necessarily probabilistic structures that use the available data to infer the correct structure of relationships both among genes and between genes and clinical phenotypes. The mathematical theory of how causal inferences can be made by examining dependency patterns from raw data is well established, and Judea Pearl (57, 58), among the very first to develop the mathematical and computational methods for this purpose, provides an excellent description and treatment of this underlying theory.

A number of groups have now published on related strategies for identifying key drivers of complex traits by examining genes located in regions of the genome genetically linked to a complex phenotype of interest, and then looking for colocalization of *cis*-acting expression quantitative trait loci (QTL) for those genes residing in the region linked to the phenotype (13, 44, 49, 50, 52, 59, 60). Those genes with *1*) expression values that are significantly correlated with the complex phenotype of interest (including disease), *2*) transcript abundances controlled by QTL that colocalize with the phenotype QTL, and *3*)

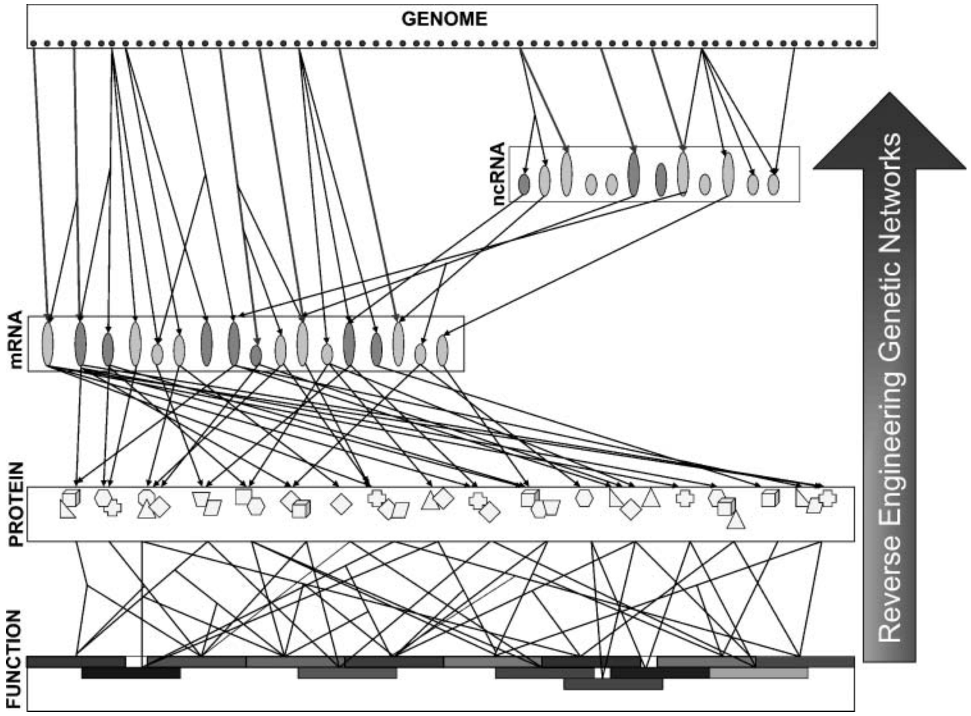

Fig. 4. Simplified view on strategy for reverse engineering gene networks using genetics in a fixed environment. The top layer represents DNA in the genome, where, in any given population, we can associate changes in the DNA with changes in the levels of transcription of both protein coding and noncoding (RNA) genes. DNA variations that fall within the region of the structural gene and associate with that gene's expression are referred to as *cis*-acting expression quantitative trait loci (eQTL), as opposed to *trans*-acting eQTL in which the DNA variation does not fall in the genomic region supporting the corresponding structural gene region (72). The proximity of transcribed sequences to the DNA provides for increased power to detect regions of the genome affecting transcript abundance levels. Changes in RNA are then shown to induce changes in proteins, where a complex web of protein interactions can form and give rise to varied cellular functions that in turn lead to disease. The gene network reconstruction methods discussed in the text leverage the ultimate source of perturbations (changes in DNA) in a system under fixed environmental conditions to order nodes in the network.

physical locations supported by the phenotype and expression QTL are natural causal candidates for the complex phenotype of interest. In these cases, the DNA variation serves as a causal anchor (given that variations in DNA lead to changes in transcription and other molecular trait activities), making it possible to partition the thousands of gene expression traits that may be correlated with a given phenotype of interest into sets of genes that are supported as causal for, reacting to, or independent of the given phenotype. The key to the success of this approach is the unambiguous flow of information, from changes in DNA to changes in RNA and protein function (Fig. 4). That is, given that two traits are linked to the same DNA locus, there are a limited number of ways in which such traits can be related with respect to that locus (13, 61, 62), whereas in the absence of such genetic information, many indistinguishable relationships would be possible, so that additional data would be required to establish the correct relationships.

It is common in studies leveraging microarray data to speculate on the possible functions implied by sets of differentially expressed mRNAs. However, functional interpretation of mRNA data is made difficult by the multiple protein products of each mRNA, the extremely large number of possible protein-protein interactions, the difficulty in precisely defining functional categories, and the fact that cell functions are controlled by phosphorylation, GTP transfer, and other signaling pathways, rather than by protein abundance alone. Further, given the large number of possible products, including posttranslational products of each mRNA, as well as the large number of possible protein-protein interactions, the number of possible combinations may be too large to fit easily into somewhat arbitrary, obviously overlapping functional categories. Therefore, the direction of the analysis arrow shown in Fig. 4 highlights that reconstructing gene networks by tracing the complex network of protein interactions back through to the mRNA levels and ultimately up to the genes responsible for the regulatory control of gene expression (i.e., identification of the key regulators of expression levels) is ostensibly an easier network reconstruction problem, given that the number of nodes is necessarily constrained to be no greater than the number of expressed genes. In contrast, reconstructing networks based on the complex, heavily intertwined network of protein interactions, states, and signaling pathways that lead to ill-defined, heavily overlapping functional categories seems a far more challenging problem.

Integrating genetic and gene expression data has been successfully applied to identifying a number of novel genes for diseases such as obesity, as well as to identifying key regulator-target pairs (13, 21, 55). In one particular study involving a segregating mouse population in which liver expression profiles were generated in 111 animals spanning a range of metabolic phenotypes, not only were scores of genes predicted as causal for obesity, suggesting that entire networks of genes lead to obesity, but three of the genes predicted as causal (C3ar1, Tgfbr2, and Zfp90) were subsequently validated in the same study (13). A more general form of this approach was developed to examine the effects that single-gene knockouts have on the transcriptional network, where such single-gene perturbation signatures can then be intersected with networks associated with disease constructed from independent sets of data (12). With this approach, the aim is to leverage the matching patterns of gene changes in the perturbed networks to isolate genes causal for disease. This approach was successfully applied to identify Alox5 as a susceptibility gene for obesity and bone traits (12).

The relatively simple approaches described above really serve to partition the gene networks associated with disease into causal, reactive, and independent pieces with respect to a phenotype of interest, so that genes supported as causal for the phenotype can be identified. These concepts have been generalized to varying degrees by several groups to allow for the more general reconstruction of gene networks by the integration of genetic and gene expression data (21, 55, 63, 64). The reconstruction of biological networks has achieved moderate success in the past, where predictive networks associated with cell cycle, circadian rhythm, and development have emerged that capture many of the fundamental attributes of living systems. The key to elucidating biological networks is a systematic source of perturbations, where the more harsh perturbations involving the complete knockout of gene activity or extreme overexpression of a gene's activity are now being balanced with naturally occurring DNA variations that more subtly perturb biological systems. These sources of perturbation are also highly multifactorial, and, in combination with environmental variation, explain a majority of complex phenotypic variations in natural populations (13, 21, 44, 49, 50, 52, 59, 65).

Zhu et al. (21) were among the first to formally incorporate genetic data into the reconstruction of gene expression networks using Bayesian network reconstruction methods. With Bayesian network reconstruction methods taking gene expression data as the only source of input, many relationships between genes in such a setting will be Markov equivalent. This means that one cannot statistically distinguish whether a given gene causes another gene to change or vice versa. To break this symmetry, Zhu et al. incorporated genetic information to establish more reliably the correct direction among expression traits. This method has been applied to networks comprised only of expression traits, as well as to networks comprised of both expression and disease traits, where the aim has been to identify those portions of the network that are driving a given disease trait. Figure 2A represents an example of a subnetwork resulting from the application of this method in the BXH cross. In this published study, *Insig2* was identified as a key driver of a number of metabolic traits, including cholesterol levels (33). *Insig2* was subsequently shown to explain a significant percentage of lifetime BMI in the human population, validating its role as a key gene in metabolic processes (3). The network shown in Fig. 2A is a subnetwork from a much larger network reconstructed from the liver gene expression data of the BXH cross, and highlights the context that these networks can provide for specific genes of interest. For example, not only does *Insig2* associate with cholesterol metabolism genes, it also associates with genes such as *Mod1*, which itself has been previously predicted as a susceptibility gene for obesity in a completely independent cross (13).

Figure 2B highlights a subnetwork from the same BXH cross referenced above, but focused, in this case, on the brain gene network, in which clinical phenotypes as well as gene expression traits were incorporated (56). This subnetwork, which is part of the larger brain gene network constructed in this study, again provides a context for a given gene of interest (*Pttg1* in this case), highlighting not only genes that a given gene is predicted to influence, but clinically relevant phenotypes as well (insulin, glucose, and leptin levels, as shown in Fig. 2B). The contexts provided by the networks in these cases enable the prioritization of putative points of therapeutic intervention for diseases of interest, given that targets in the network can be identified that involve genes identified in multiple disease areas (such as *Insig2*), genes specific to a given disease subtype, genes that may be involved in toxicity-associated pathways, as well as genes involved in pathways associated with adverse events (14).

In addition to the work of Zhu et al., Kulp and Jagalur (55) recently described a method that integrates interacting expression and genotype data to identify key regulator–trait pairs, facilitating direct identification of quantitative trait genes for gene expression QTL. Stylianou et al. (64) also recently devised a method based on structural equation modeling to incorporate a number of clinical phenotypes and associated QTL into a graphical model to identify interacting networks of many genes regulating obesity-related traits. Finally, Koller et al. (63) have recently described a novel probabilistic method for integrating genotypic and expression data to uncover regulatory relationships among genes that would not otherwise be uncovered by looking at genes independent of the network in which they operate. These new network-based methods not only support earlier work in this area on inferring causal associations by integrating genotypic and gene expression data, but they also continue to take us closer to building large-scale networks that are predictive and that will enhance our ability to elucidate disease and other complex phenotypes underlying living systems.

FROM NETWORKS TO THE IDENTIFICATION OF KEY INTERVENTION POINTS

The primary goal in elucidating the network associated with a disease involves not only finding the key drivers of the disease but also developing a better understanding of how to better target the disease, given the structure of the underlying network. Networks aid in the identification of disease subtypes and provide a broader context in which to assess the best genes to target for therapeutic intervention of disease, taking into account nodes in the network associated with multiple diseases, subtypes of disease, and toxicity pathways, as well as other clinical phenotypes that may be adversely affected (14, 52). One of the complexities encountered in dissecting common human diseases is the degree of disease heterogeneity represented in almost every population studied. Different pathways associated with different molecular processes may underlie a particular disease subtype, so that any given population represents a mixture of these different subtypes (12, 24, 52, 66, 67), and networks provide a rational way to identify pathways specific to different subtypes (**Fig. 5A**). The topological overlap map shown in Fig. 5A highlights functional units of the network (the gene modules) that are enriched for genes operating in pathways known to be associated with disease-related traits. Such modules would be expected to be enriched not only for genes associated with disease but also for genetic loci associated with disease, and this type of information can be used to aid in the stratification of disease populations into specific disease subtypes. Ultimately, identification of subnetworks associated with different disease subtypes leads to the identification of targets that are specific to a given subtype, as well as to biomarkers for the disease subtypes that can be used in classifying patients into treatment groups (14).

In addition to disease subtypes, biological networks bring to light the need to target multiple genes driving common forms of diseases such as obesity, diabetes, and atherosclerosis. One simple example of a combination therapy for targeting multiple pathways to treat disease involves statins and the drug ezetimibe. Statins lower cholesterol by inhibiting 3-hydroxy-3-methylglutaryl-CoA reductase, the rate-limiting enzyme for cholesterol synthesis. On the other hand, ezetimibe inhibits Niemann-Pick type C1 gene-like 1 gene, an intestinal cholesterol transporter. Thus, these two classes of drugs affect cholesterol levels by attacking completely independent pathways, and these drugs taken in combination may have a more pronounced impact on cholesterol levels than either drug taken alone (68). Beyond this simple example is the idea that common diseases such as obesity, diabetes, and atherosclerosis are emergent properties of the network, and so are likely not to be treatable in a highly efficacious manner by targeting only a single gene. Instead, disrupting the different parts of the network driving the disease will likely require the simultaneous targeting of multiple genes. Yeh et al. (69) constructed drug interaction networks by looking at all pairs of a number of antibiotic and anticancer drugs, monitoring the activity of these drug combinations using a sensitive bioluminescence assay to detect subtle variations in viability and growth. Interestingly, they found that the drugs could mostly be classified into two sets, where in one set, all drugs acted synergistically and in the other, they acted antagonistically, supporting the idea that gene network studies may reveal what combinations provide synergistic effects that may significantly enhance efficacy of treatment.

The types of gene networks discussed herein suggest that many more than two targets will need to be hit at a time in order to destabilize networks driving disease. The modules identified in the type of coexpression network depicted in Fig. 2A are comprised of sets of genes that are

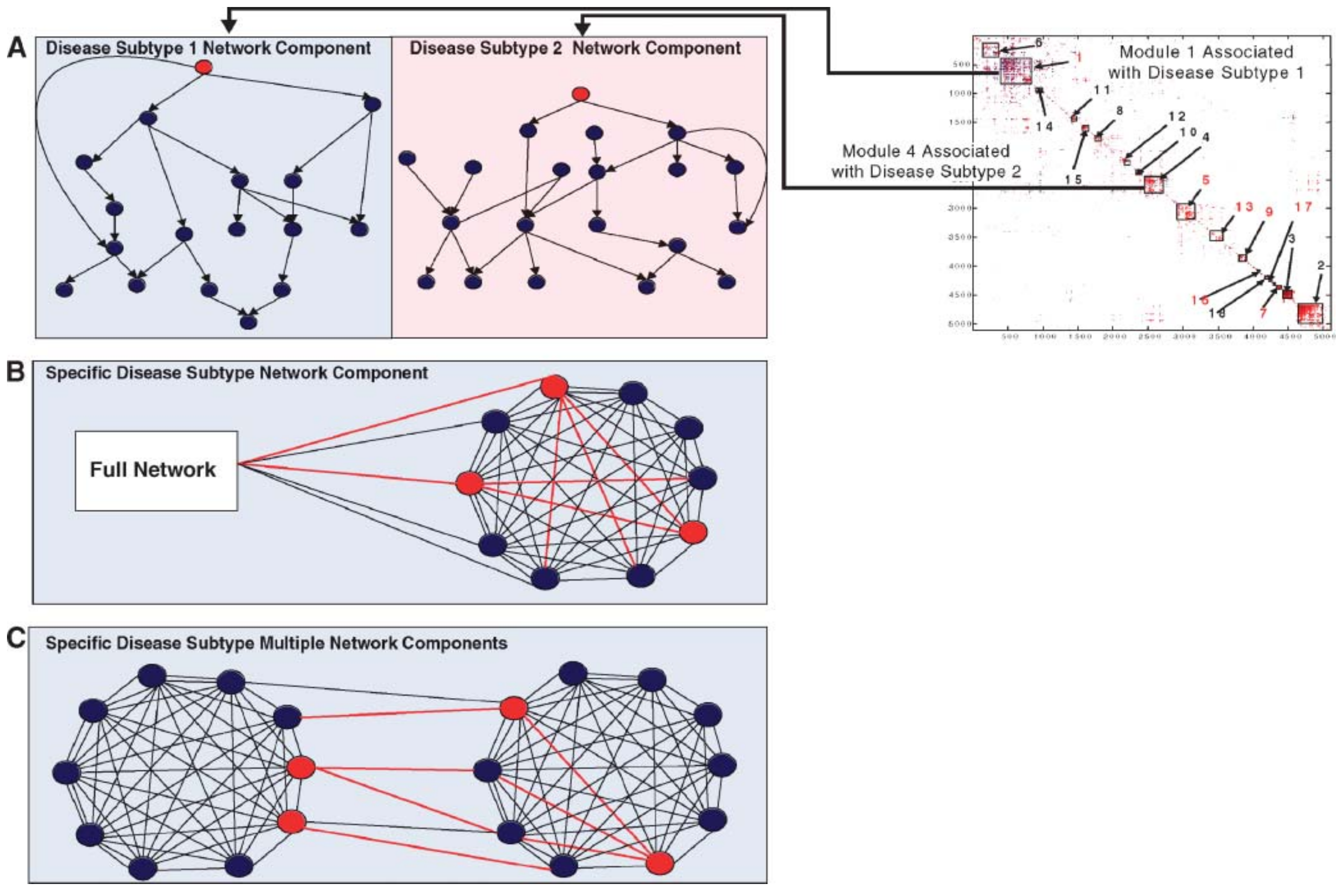
A
Disease Subtype 1 Network Component
Disease Subtype 2 Network Component
Module 1 Associated with Disease Subtype 1
Module 4 Associated with Disease Subtype 2
6
1
11
14
8
12
10
4
15
5
13
9
17
3
2
16
18
7
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
B
Specific Disease Subtype Network Component
Full Network
C
Specific Disease Subtype Multiple Network Components

Fig. 5. Highlighting the need to target multiple genes to treat common diseases such as obesity, diabetes, and atherosclerosis. A: Integrating molecular profiling, genetic and clinical data can result in the identification of multiple disease subtypes and in elucidation of pathways underlying the different subtypes. The topological overlap map from Fig. 2 is displayed on the right, with modules 1 and 2 highlighted as representing two different functional components of the network driving two different subtypes of disease (red and blue boxes on the left). The red nodes indicate putative targets within each subtype that could be developed to treat disease. B: A hypothetical clique structure from one of the disease-associated modules highlighted in the topological overlap map in panel A. Given the highly interconnected nature of this structure, targeting a single gene may not lead to an efficacious treatment of disease, given the ability of such interconnected structures to compensate for such perturbations. Instead, it may be necessary to target multiple genes to inhibit the ability of other genes in the subnetwork to compensate. Three genes are highlighted as being targeted in this case (red nodes). The red edges indicate connections that are weakened or eliminated, leading to destabilization of the structure and subsequent loss of the coherence needed to maintain a disease state. C: Similar to B, but there are two clique structures whose interaction drives one of the disease subtypes. In this case the coherence needed to maintain the disease state is achieved via relatively weak links between the clique structures. Simultaneously targeting multiple genes in these structures may destabilize the weak links, and, as a result, destroy the coherence needed to maintain the disease state.

completely connected, in which every gene in each of the completely connected sets is significantly correlated to every other gene in the set (in graph theory terms, the set forms a clique structure), as depicted in Fig. 5B. If these types of structures turn out to be central to the onset of disease, targeting a single gene probably will not destabilize the structure enough to lead to an effective treatment, given the redundancy inherent in such structures, which provides them the ability to compensate for perturbations to single genes in the structure. To sufficiently destabilize such structures, multiple genes would likely have to be pharmacologically targeted to treat disease effectively, as indicated in Fig. 5B.

The type of clique structure depicted in Fig. 5B may also associate with other clique structures via weaker links (weaker than the links making up the clique structures), giving rise to what are known as clique communities that together may drive systems into a particular disease state. That is, clique communities rather than individual cliques may turn out to induce the type of coherence needed to drive a system into a disease state. In such cases, it may again be necessary to target not only multiple genes within a given clique, but also multiple genes within multiple cliques to destabilize the weak links between these structures and so destroy, in turn, the coherence needed to drive systems into disease states (Fig. 5C). Given the many different parts of entire networks that may be involved in complex diseases such as obesity, diabetes, and atherosclerosis, it may be necessary to target 10 or 20 or even 50 or more genes simultaneously to treat disease most effectively. Classic small-molecule designs will probably not be feasible in this setting, whereas if RNAi-based therapeutics prove effective (via the use of miRNAs or si/shRNAs), such a technology would offer a way to simultaneously target a number of genes. Already there are technologies developing that not only allow for the simultaneous targeting of a number of genes using RNAi-based approaches, but that also enable the efficient delivery of double-stranded RNA molecules to the targeted cell types of choice (70). Because the type of targeting depicted in Figs. 5B and C aims to destabilize portions of the network driving disease, complete gene knockdown or activation probably will not be required. Instead, suppressing or activating multiple genes enough to destabilize the links needed to achieve the coherence necessary to push a system into a disease state would be needed, and again, this may be something RNAi-based technologies are well suited to deliver. It is of note that if such treatment paradigms become possible, the most effective treatments will come from investigators who understand disease at the systems level, which no doubt will require the reconstruction of predictive gene networks in the spirit of what we have reviewed here.

CONCLUSIONS

The network methods reviewed here provide a convenient framework for moving beyond examining genes one at a time to understand the complexity of common human diseases. Whether constructing interaction networks, Bayesian networks, networks based on structural equation models, or networks based on any of the other myriad of network reconstruction methods, the advantage the network view affords is the ability to consider more of the raw data informing on complex phenotypes simultaneously, taking into account dependency structures between all of the different fundamental components of the system, and providing a framework to integrate a diversity of data types (something Bayesian network reconstruction methods are particularly good at). Researchers in the life and biomedical sciences will have few other options available to them in considering the vast amounts of data being generated to elucidate how the hundreds of thousands or even millions of fundamental components within the cell interact to give rise to complex phenotypes. Networks provide one of the only frameworks of which we are aware to systematically and simultaneously take all of the fundamental components into account. Statistical inferences on networks will come to define much of the future research needed in this field to adequately leverage what network models can provide.

Of course, the types of approaches reviewed here represent only the first steps being taken to reconstruct meaningful gene networks, and even in this review, we have largely restricted attention to genetic, gene expression, and disease phenotype data. Ultimately, it will be necessary to integrate many different lines of experimental data simultaneously. Protein-protein interactions, protein-DNA interactions, protein-RNA interactions, RNA-RNA interactions, protein state, methylation state, and especially differential methylation states have now been shown to act transgenerationally (71), and interactions with metabolites, among many other important interactions, are all-important pieces that define complex phenotypes that emerge in living systems. What a given protein and RNA do will give way to what a network of protein, RNA, DNA, and metabolite interactions do, where such networks of interaction are defined by the context in which they operate, with environment playing a critical role. Understanding a particular network state that drives disease (or other complex phenotypes that define living systems) will require not only knowledge of DNA and environmental variation and the changes these variation components induce in the network, but also information on the previous states of the network that led to the current state, where environmental stresses interacting in complex ways with genetic background not only influence the current state of the network but also can lead to longer lasting effects on the network that act transgenerationally.

While this more comprehensive reconstruction of biological networks is still outside the scope of what is presently doable, the types of approaches reviewed here represent solid first steps toward this ultimate goal. Even though the number of networks that can be reconstructed from the fundamental components of living systems is truly daunting, as work progresses in this area, we will learn the rules that necessarily constrain the possible ranges of molecular interactions, and as a result, will begin to capture the more conserved network motifs that form the framework upon which all other interactions are based. The complexity revealed by a systems biology-motivated approach to elucidating complex phenotypes such as disease should be embraced, given the potential to develop a better understanding of the true diversity of disease and the constellation of genes that need to be targeted to effectively treat disease.

REFERENCES

1. Florez, J. C., K. A. Jablonski, N. Bayley, T. I. Pollin, P. I. de Bakker, A. R. Shuldiner, W. C. Knowler, D. M. Nathan, and D. Altshuler. 2006. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* **355:** 241–250.
2. Grant, S. F., G. Thorleifsson, I. Reynisdottir, R. Benediktsson, A. Manolescu, J. Sainz, A. Helgason, H. Stefansson, V. Emilsson, A. Helgadottir, et al. 2006. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* **38:** 320–323.
3. Herbert, A., N. P. Gerry, M. B. McQueen, I. M. Heid, A. Pfeufer, T. Illig, H. E. Wichmann, T. Meitinger, D. Hunter, F. B. Hu, et al. 2006. A common genetic variant is associated with adult and childhood obesity. *Science.* **312:** 279–283.
4. Edwards, A. O., R. Ritter 3rd, K. J. Abel, A. Manning, C. Panhuysen, and L. A. Farrer. 2005. Complement factor H polymorphism and age-related macular degeneration. *Science.* **308:** 421–424.
5. Haines, J. L., M. A. Hauser, S. Schmidt, W. K. Scott, L. M. Olson, P. Gallins, K. L. Spencer, S. Y. Kwan, M. Noureddine, J. R. Gilbert, et al. 2005. Complement factor H variant increases the risk of age-related macular degeneration. *Science.* **308:** 419–421.
6. Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, et al. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science.* **308:** 385–389.
7. Li, M., P. Atmaca-Sonmez, M. Othman, K. E. Branham, R. Khanna, M. S. Wade, Y. Li, L. Liang, S. Zareparsi, A. Swaroop, et al. 2006.

CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat. Genet.* **38:** 1049–1054.

8. Maller, J., S. George, S. Purcell, J. Fagerness, D. Altshuler, M. J. Daly, and J. M. Seddon. 2006. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat. Genet.* **38:** 1055–1059.
9. Dwyer, J. H., H. Allayee, K. M. Dwyer, J. Fan, H. Wu, R. Mar, A. J. Lusis, and M. Mehrabian. 2004. Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. *N. Engl. J. Med.* **350:** 29–37.
10. Mehrabian, M., H. Allayee, J. Wong, W. Shi, X. P. Wang, Z. Shaposhnik, C. D. Funk, and A. J. Lusis. 2002. Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91:** 120–126.
11. Zhao, L., M. P. Moos, R. Grabner, F. Pedrono, J. Fan, B. Kaiser, N. John, S. Schmidt, R. Spanbroek, K. Lotzer, et al. 2004. The 5-lipoxygenase pathway promotes pathogenesis of hyperlipidemia-dependent aortic aneurysm. *Nat. Med.* **10:** 966–973.
12. Mehrabian, M., H. Allayee, J. Stockton, P. Y. Lum, T. A. Drake, L. W. Castellani, M. Suh, C. Armour, S. Edwards, J. Lamb, et al. 2005. Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat. Genet.* **37:** 1224–1233.
13. Schadt, E. E., J. Lamb, X. Yang, J. Zhu, S. Edwards, D. Guhathakurta, S. K. Sieberts, S. Monks, M. Reitman, C. Zhang, et al. 2005. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37:** 710–717.
14. Schadt, E. E., A. Sachs, and S. Friend. 2005. Embracing complexity, inching closer to reality. *Sci. STKE.* **295:** pe40.
15. Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray. 1999. From molecular to modular cell biology. *Nature.* **402:** C47–C52.
16. Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5:** 101–113.
17. Zerhouni, E. 2003. Medicine. The NIH roadmap. *Science.* **302:** 63–72.
18. Han, J. D., N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. Walhout, M. E. Cusick, F. P. Roth, et al. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature.* **430:** 88–93.
19. 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.
20. Pinto, S., A. G. Roseberry, H. Liu, S. Diano, M. Shanabrough, X. Cai, J. M. Friedman, and T. L. Horvath. 2004. Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science.* **304:** 110–115.
21. Zhu, J., P. Y. Lum, J. Lamb, D. GuhaThakurta, S. W. Edwards, R. Thieringer, J. P. Berger, M. S. Wu, J. Thompson, A. B. Sachs, et al. (2004). An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.* **105:** 363–374.
22. Rual, J. F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature.* **437:** 1173–1178.
23. 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 of human endothelial cell responses to oxidized lipids. *Proc. Natl. Acad. Sci. USA.* **103:** 12741–12746.
24. Ghazalpour, A., S. Doss, B. Zhang, S. Wang, C. Plaisier, R. Castellanos, A. Brozell, E. E. Schadt, T. A. Drake, A. J. Lusis, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* Epub ahead of print. August 18, 2006; doi:10.1371/journal.pgen.0020130.
25. Barabasi, A. L., and R. Albert. 1999. Emergence of scaling in random networks. *Science.* **286:** 509–512.
26. Gunsalus, K. C., H. Ge, A. J. Schetter, D. S. Goldberg, J. D. Han, T. Hao, G. F. Berriz, N. Bertin, J. Huang, L. S. Chuang, et al. 2005. Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. *Nature.* **436:** 861–865.
27. Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. 2002. Hierarchical organization of modularity in metabolic networks. *Science.* **297:** 1551–1555.
28. Carlson, M. R., B. Zhang, Z. Fang, P. S. Mischel, S. Horvath, and S. F. Nelson. 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics.* **7:** 40.
29. Li, H., H. Chen, L. Bao, K. F. Manly, E. J. Chesler, L. Lu, J. Wang, M. Zhou, R. W. Williams, and Y. Cui. 2006. Integrative genetic analysis of transcription modules: towards filling the gap between genetic loci and inherited traits. *Hum. Mol. Genet.* **15:** 481–492.
30. Zhang, B., and S. Horvath. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* Vol. 4: No. 1, Article 17 Accessed October 20, 2006, at http://www.bepress.com/sagmb/vol4/iss1/art17.
31. 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.
32. Yang, X., E. E. Schadt, S. Wang, H. Wang, A. P. Arnold, L. Ingram-Drake, T. A. Drake, and A. J. Lusis. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16:** 995–1004.
33. Cervino, A. C., G. Li, S. Edwards, J. Zhu, C. Laurie, G. Tokiwa, P. Y. Lum, S. Wang, L. W. Castellini, A. J. Lusis, et al. 2005. Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. *Genomics.* **86:** 505–517.
34. Hughes, T. R., M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, et al. 2000. Functional discovery via a compendium of expression profiles. *Cell.* **102:** 109–126.
35. Legato, M. J., A. Gelzer, R. Goland, S. A. Ebner, S. Rajan, V. Villagra, and M. Kosowski. 2006. Gender-specific care of the patient with diabetes: review and recommendations. *Gend. Med.* **3:** 131–158.
36. Nathan, L., and G. Chaudhuri. 1997. Estrogens and atherosclerosis. *Annu. Rev. Pharmacol. Toxicol.* **37:** 477–515.
37. Regitz-Zagrosek, V. 2006. Therapeutic implications of the gender-specific aspects of cardiovascular disease. *Nat. Rev. Drug Discov.* **5:** 425–438.
38. Di Marco, F., M. Verga, M. Reggente, F. Maria Casanova, P. Santus, F. Blasi, L. Allegra, and S. Centanni. 2006. Anxiety and depression in COPD patients: the roles of gender and disease severity. *Respir. Med.* **100:** 1767–1774.
39. Goldberg, A. C. 2004. A meta-analysis of randomized controlled studies on the effects of extended-release niacin in women. *Am. J. Cardiol.* **94:** 121–124.
40. Smesny, S., T. Rosburg, S. Klemm, S. Riemann, K. Baur, N. Rudolph, S. Grunwald, and H. Sauer. 2004. The influence of age and gender on niacin skin test results—implications for the use as a biochemical marker in schizophrenia. *J. Psychiatr. Res.* **38:** 537–543.
41. Kendziorski, C. M., M. Chen, M. Yuan, H. Lan, and A. D. Attie. 2006. Statistical methods for expression quantitative trait loci (eQTL) mapping. *Biometrics.* **62:** 19–27.
42. Storey, J. D., J. Y. Dai, and J. T. Leek. The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics.* Epub ahead of print. August 23, 2006; doi: 10.1093/biostatistics/kxl019.
43. Storey, J. D., W. Xiao, J. T. Leek, R. G. Tompkins, and R. W. Davis. 2005. Significance analysis of time course microarray experiments. *Proc. Natl. Acad. Sci. USA.* **102:** 12837–12842.
44. Brem, R. B., G. Yvert, R. Clinton, and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science.* **296:** 752–755.
45. DeCook, R., S. Lall, D. Nettleton, and S. H. Howell. 2006. Genetic regulation of gene expression during shoot development in Arabidopsis. *Genetics.* **172:** 1155–1164.
46. Hubner, N., C. A. Wallace, H. Zimdahl, E. Petretto, H. Schulz, F. Maciver, M. Mueller, O. Hummel, J. Monti, V. Zidek, et al. 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* **37:** 243–253.
47. Jin, W., R. M. Riley, R. D. Wolfinger, K. P. White, G. Passador-Gurgel, and G. Gibson. 2001. The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. *Nat. Genet.* **29:** 389–395.
48. Klose, J., C. Nock, M. Herrmann, K. Stuhler, K. Marcus, M. Bluggel, E. Krause, L. C. Schalkwyk, S. Rastan, S. D. Brown, et al. 2002. Genetic analysis of the mouse brain proteome. *Nat. Genet.* **30:** 385–393.

49. Monks, S. A., A. Leonardson, H. Zhu, P. Cundiff, P. Pietrusiak, S. Edwards, J. W. Phillips, A. Sachs, and E. E. Schadt. 2004. Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75:** 1094–1105.
50. Morley, M., C. M. Molony, T. M. Weber, J. L. Devlin, K. G. Ewens, R. S. Spielman, and V. G. Cheung. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature.* **430:** 743–747.
51. Oleksiak, M. F., G. A. Churchill, and D. L. Crawford. 2002. Variation in gene expression within and among natural populations. *Nat. Genet.* **32:** 261–266.
52. Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb, G. Cavet, et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* **422:** 297–302.
53. Stranger, B. E., M. S. Forrest, A. G. Clark, M. J. Minichiello, S. Deutsch, R. Lyle, S. Hunt, B. Kahl, S. E. Antonarakis, S. Tavare, et al. 2005. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1:** e78.
54. Yvert, G., R. B. Brem, J. Whittle, J. M. Akey, E. Foss, E. N. Smith, R. Mackelprang, and L. Kruglyak. 2003. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat. Genet.* **35:** 57–64.
55. Kulp, D. C., and M. Jagalur. 2006. Causal inference of regulator-target pairs by gene mapping of expression phenotypes. *BMC Genomics.* **7:** 125.
56. Lum, P. Y., Y. Chen, J. Zhu, J. Lamb, S. Melmed, S. Wang, T. A. Drake, A. J. Lusis, and E. E. Schadt. 2006. Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J. Neurochem.* **97 (Suppl.):** 50–62.
57. Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, San Mateo, CA.
58. Pearl, J. 2000. Causality. Cambridge University Press, New York, NY.
59. Chesler, E. J., L. Lu, S. Shou, Y. Qu, J. Gu, J. Wang, H. C. Hsu, J. D. Mountz, N. E. Baldwin, M. A. Langston, et al. 2005. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* **37:** 233–242.
60. Cheung, V. G., R. S. Spielman, K. G. Ewens, T. M. Weber, M. Morley, and J. T. Burdick. 2005. Mapping determinants of human gene expression by regional and genome-wide association. *Nature.* **437:** 1365–1369.
61. 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.
62. Schadt, E. E. 2006. Novel integrative genomics strategies to identify genes for complex traits. *Anim. Genet.* **37(Suppl.):** 18–23.
63. Lee, S., D. Pe'er, A. M. Dudley, G. M. Church, and D. Koller. 2006. Identifying regulatory mechanisms using individual variation reveals key role for chromatin modification. *Proc. Natl. Acad. Sci. USA.* **103:** 14062–14067.
64. Stylianou, I. M., R. Korstanje, R. Li, S. Sheehan, B. Paigen, and G. A. Churchill. 2006. Quantitative trait locus analysis for obesity reveals multiple networks of interacting loci. *Mamm. Genome.* **17:** 22–36.
65. Bystrykh, L., E. Weersing, B. Dontje, S. Sutton, M. T. Pletcher, T. Wiltshire, A. I. Su, E. Vellenga, J. Wang, K. F. Manly, et al. 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat. Genet.* **37:** 225–232.
66. Ghazalpour, A., S. Doss, S. S. Sheth, L. A. Ingram-Drake, E. E. Schadt, A. J. Lusis, and T. A. Drake. 2005. Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* **6:** R59.
67. van't Veer, L. J., H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* **415:** 530–536.
68. Ballantyne, C. M., N. Abate, Z. Yuan, T. R. King, and J. Palmisano. 2005. Dose-comparison study of the combination of ezetimibe and simvastatin (Vytorin) versus atorvastatin in patients with hypercholesterolemia: the Vytorin Versus Atorvastatin (VYVA) study. *Am. Heart J.* **149:** 464–473.
69. Yeh, P., A. I. Tschumi, and R. Kishony. 2006. Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* **38:** 489–494.
70. Li, C. X., A. Parker, E. Menocal, S. Xiang, L. Borodyansky, and J. H. Fruehauf. 2006. Delivery of RNA interference. *Cell Cycle.* **5:** 2103–2109.
71. Anway, M. D., A. S. Cupp, M. Uzumcu, and M. K. Skinner. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science.* **308:** 1466–1469.
72. Doss, S., E. E. Schadt, T. A. Drake, and A. J. Lusis. 2005. Cis-acting expression quantitative trait loci in mice. *Genome Res.* **15:** 681–691.