PERSONALIZED MEDICINE

NEW: Network-Enabled Wisdom in Biology, Medicine, and Health Care

Eric E. Schadt[1] and Johan L. M. Björkegren[2,3,4]*

Complete repertoires of molecular activity in and between tissues provided by new high-dimensional "omics" technologies hold great promise for characterizing human physiology at all levels of biological hierarchies. The combined effects of genetic and environmental perturbations at any level of these hierarchies can lead to vicious cycles of pathology and complex systemic diseases. The challenge lies in extracting all relevant information from the rapidly increasing volumes of omics data and translating this information first into knowledge and ultimately into wisdom that can yield clinically actionable results. Here, we discuss how molecular networks are central to the implementation of this new biology in medicine and translation to preventive and personalized health care.

INTRODUCTION

Next-generation technologies that routinely measure biological parameters on a genome-wide scale ("omics" data)—such as DNA variations and epigenetic modifications, RNA and protein concentrations, and a variety of metabolites—are continuously being refined and offered at ever-decreasing costs. The resulting oceans of molecular data (moving quickly from the petabyte to exabyte scale or, even more scary, zetabyte—that's 21 zeros) cannot be deciphered with traditional mathematical analyses carried out on isolated computers. Nor is the traditional representation of biological processes as linear pathways sufficient to represent the hierarchy of levels of molecular and higher-order regulation, and the interplay that defines human physiology and pathology. Instead, the "new biology" requires large warehouse-scale computing and sophisticated algorithms capable of processing and appropriately integrating the vast amounts of molecular data being generated today (and which continue to grow at an exponential pace). We believe that multidimensional networks are required to model human physiology in general and to identify key drivers of pathology in individual patients. (See the documentary film *The New Biology* at http://www.youtube.com/watch?v=sjTQD6E3lH4.)

NEW BIOLOGY

Intricate data-intensive computations that can span days, weeks, or even months give rise to mathematical models that reflect the web of genetic, epigenetic, metabolic, environmental, and biochemical interactions and modes of regulation that define biological processes across many layers of information. This web is best represented by networks that operate at multiple levels of biological hierarchies—from the molecular networks at play within a single cell, to the cellular networks that define the activity of a tissue, to system-level networks that operate across organs in a system, on up to social networks that reflect interactions among individuals in a population and between individuals and their environment in ways that affect disease [questions remain as to the meaning of the disease associations observed in social networks (*1*)]. The architecture of biological networks shares similarities with well-studied ones in other disciplines, such as social and transportation networks. Like these large-scale information networks, molecular networks in biology are sparse and follow a power-law distribution in which most nodes have few interactions (say, one to three), whereas a smaller number, referred to as hub nodes, have many interactions (tens to hundreds or even thousands) (*2*) (Fig. 1).

Mapping the connectivity structure of networks (that is, the topology) is crucial for understanding how biological processes are defined at the molecular level, how they can be disrupted to cause disease, and how we can best assess the risk of and intervene to treat disease. Data-intensive omics technologies have now pushed researchers to adopt a new approach to molecular biology that maximally integrates information from the big data generated from these technologies to uncover biological processes and relationships among these processes that would otherwise remain hidden. Figure 2A summarizes the basic concept of network inference as required for network-enabled wisdom (NEW) in biology.

The omics revolution that drives NEW biology is erupting in biomedicine and nearly every other subfield of biological research. An aspect of human health for which NEW biology provides much-needed assistance is in helping humanity combat the increasing threat of new pandemics—epidemics of infectious diseases (*3*). The expansion of trade and travel has intensified the threat of future pandemics, whose devastating effects can be minimized only through prevention and early detection. Fast and accurate genomic sequencing technologies (*4*) combined with high-performance computing (*5*)—both key drivers of NEW biology—can help to efficiently pinpoint the geographical and sometimes even patient-specific origins of threatening pandemics. Through direct sequencing of viral DNA or bacterial genomes, "disease weather maps" that show the historical, current, and predicted locations of a pandemic spread can be established. These maps could be quite specific and even pinpoint the source of the outbreak down to a given individual. Direct sequencing and comparative genome analysis from resistant and susceptible individuals can help to rapidly identify mechanisms of microbial resistance, which can be lifesaving for affected patients. The actions in response to the outbreaks of cholera in Haiti (*6*) and *Escherichia coli* in Germany (*7*) provide robust examples of the lifesaving capability of NEW biology. In such situations, information about the people and places in which infections and pathogens orig-

[1]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA. [2]Cardiovascular Genomics, Vascular Biology Division, Department of Medical Biophysics and Biochemistry, Karolinska Institutet, 171 77 Stockholm, Sweden. [3]Department of Pathological Anatomy and Forensic Medicine, Faculty of Medicine, University of Tartu, Tartu, Estonia. [4]Clinical Gene Networks AB, Karolinska Science Park, 171 77 Stockholm, Sweden.
*To whom correspondence should be addressed. E-mail: johan.bjorkegren@ki.se

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

inate, respectively, has the potential to lead rapidly to preventive measures that reduce the spread of disease (for example, prophylactic treatment of people who reside in high-risk locations).

A network-based understanding of a pathogen's molecular biology can also take us beyond DNA sequencing to a deeper characterization (for example, genome-wide gene expression and protein analysis) of

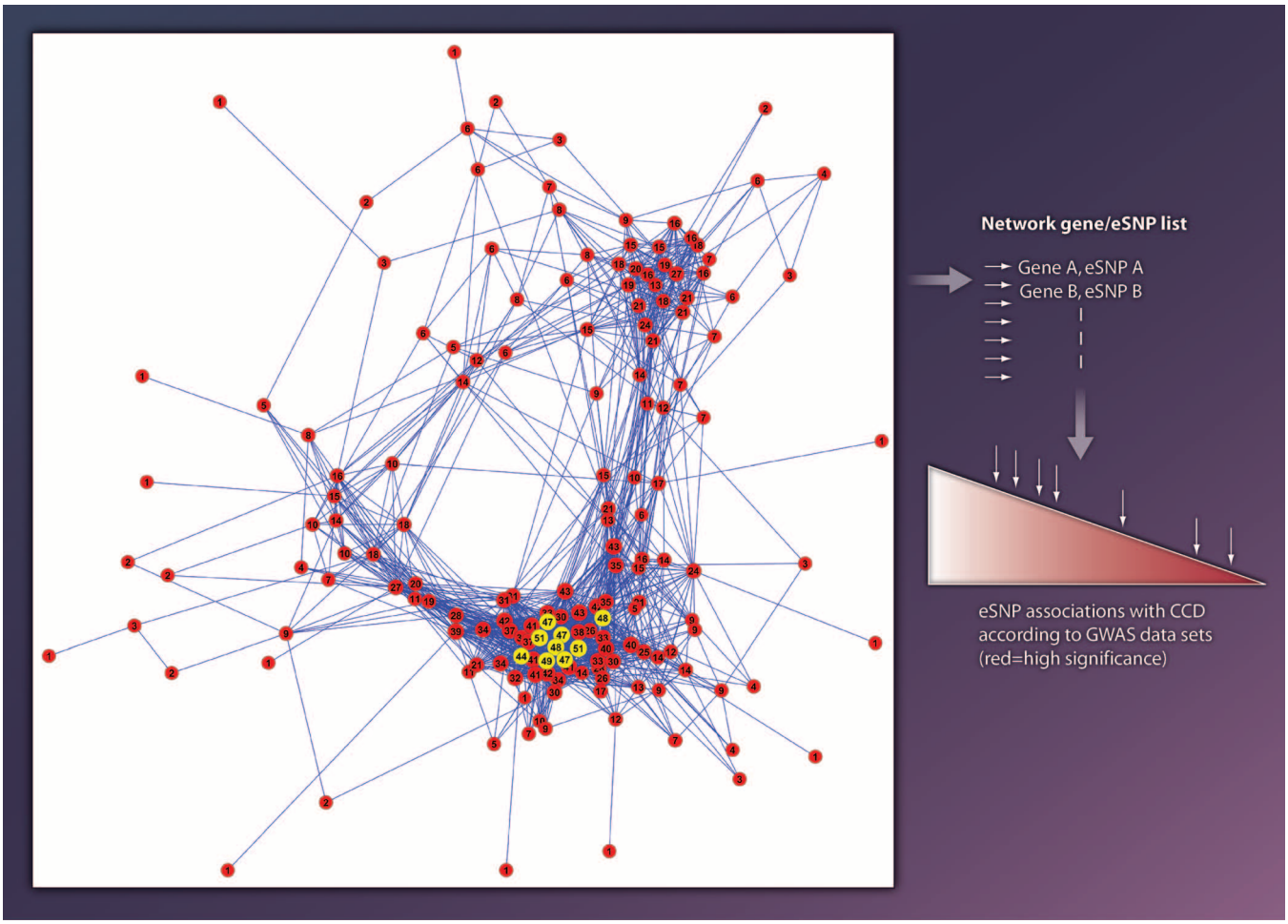

Fig. 1. Network news. Biological networks are sensors and mediators of the combined effects of environmental and genetic CCD risk factors. **Left:** An example of an arterial wall co-expression gene network inferred from genome-wide mRNA profiles isolated from three arterial wall samples from the same patient (three RNA samples per individual). The criterion for edges (that is, interactions between nodes) in this network is supported by the data from at least two arterial wall samples. As in this example, biological networks are sparse, following a power-law distribution in which most nodes (that is, network components, which can be genes, proteins, or metabolites; shown in red) have few interactions; the few nodes that have many interactions are called "hubs" (yellow nodes in this example; >44 edges per node) (*2*). Networks can be inferred from various kinds of genome-wide data sets with the use of computational inference algorithms. One type is the exemplified gene co-expression network. The direction of edges is not revealed in co-expression networks, but the length of the edge is related to the strength of the association between nodes; the longer the edge, the weaker the association. Bayesian network reconstruction is more sophisticated, applying algorithms based on probabilities and conditional dependencies, disclosing networks with edges also holding information about the directions and type of regulation. In this way, Bayesian network reconstruction on the combined data sets of genome-wide DNA sequence variations and gene expression conducted with biological samples from a single individual can be used to distinguish networks that cause disease from those that constitute a reaction to disease (*15*). The reactive or causal role of a disease-related genome-wide gene co-expression network can be investigated by analyzing GWAS data sets to determine enrichment for inherited risk. In the network diagram shown: nodes (genes), the size and number indicate the numbers of neighboring nodes; edges, length is proportional to the strength of the Pearson correlation coefficients between nodes. Visualized using Cytoscape (http://www.cytoscape.org). **Right:** Principal steps for genetic enrichment analysis using GWAS data sets. A network defines a list of functionally associated genes; alternatively, this list can be defined by co-expression clusters of genes (*10*). Next, corresponding DNA variants that affect expression of listed genes (eSNPs) are defined by seeking SNP allele frequencies that correlate with mRNA concentrations. The list of eSNPs is then matched to the GWAS SNP microarray platform using the HapMap (http://www.hapmap.org) or the 1000-genome (http://www.1000genomes.org) platforms. The expanded set of SNPs is then examined for enrichment in disease risk either by searching for the relative number of disease associations [false discovery rate (FDR) = 0.05] or by examining whether the expanded set is shifted toward higher significance [in the figure to the right (increasingly red)] relative to sets of the same number of randomly selected SNPs (x10,000).

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

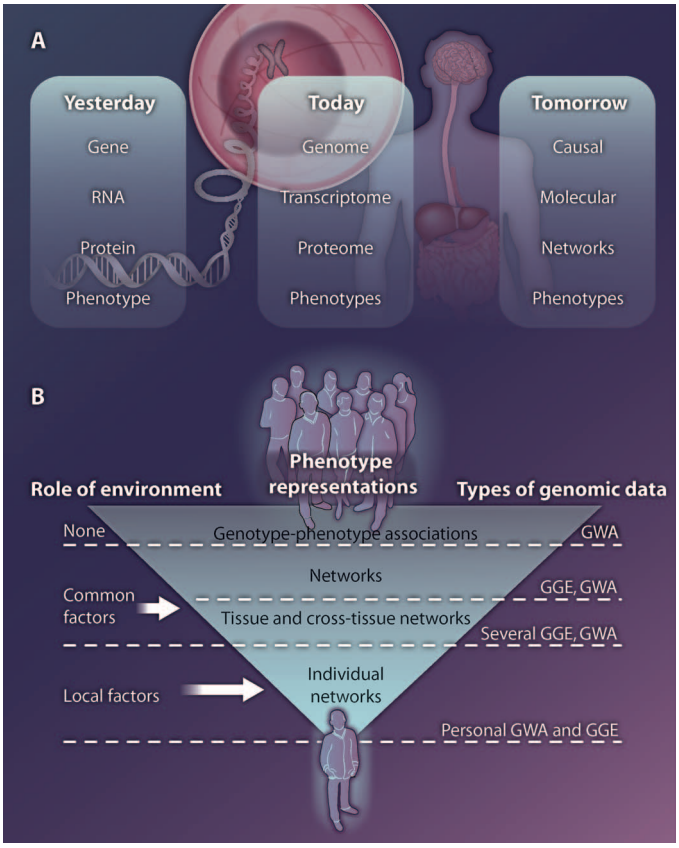

Fig. 2. Tomorrow's human biology. Molecular networks play a fundamental role in the future of biomedicine. (**A**) Basic concept of NEW in biology. Advances in mapping DNA loci related to human diseases and genome-wide profiling of mRNA transcript abundances have occurred on an unprecedented scale (*11*, *22*, *32*–*34*). Of particular interest is the identification of RNA molecules that mediate the flow of information from DNA to disease, because RNA is transcribed directly from a DNA template and is thus the most proximal non-DNA species of all molecules in the cell. Bayesian network reconstruction (*11*, *12*) is a powerful approach for simultaneously considering thousands of molecular or clinical variables and for identifying patterns of causal relationships between these variables in a completely data-driven fashion. We developed a way to overcome the chief limitation of this approach—deriving predictive models from correlation data (*11*, *12*, *35*)—by leveraging DNA variation as a systematic source of perturbation (*32*). The resulting probabilistic causal networks are critical for understanding the behavior of any one gene in the context of human disease, because individual genes operate in molecular networks that define disease-associated biological and pathological events. (**B**) A risk triangle for any given CCD that represents inherited (that is, genetic) risk factors, from the common ones (which are context-independent) to the increasingly individual ones (which are context-dependent). The genetics of gene expression cohorts (GGE) will help us interpret data from GWAS by identifying genes whose corresponding RNA levels associate with genetic loci that also associate with disease (*22*, *23*, *32*, *36*–*38*); gene expression profiles can also be used to infer causal relationships between molecular traits (up to networks) and disease states (*30*, *39*–*41*). The triangle represents the inherited risk for a given CCD in a given individual. CCD risk is shared with the wider population at the base of the triangle, becomes increasingly individual (that is, less represented in the wider population), and is purely individual at its peak. **Left:** Arrows indicate the roles and types of environmental factors that act in concert with inherited risk in the triangle. **Middle:** Ways of representing risk at different levels. **Right:** Types of genomic data needed to map CCD risk.

the virulence of threatening pandemics by providing information about how pathogen networks interact with their environment. Such networks are likely to provide information about the capacity of a given pathogen to spread and actually cause a pandemic—information that cannot be understood solely from isolated genome/DNA analysis. In a recent intriguing study, two bacterial species were cocultured to generate several layers of omics information over time; analysis of the data revealed mechanisms by which the bacteria rapidly adapt to each other's presence. This study illustrates a fundamental strategy for clarifying microbial crosstalk in a minimal ecosystem—a first step toward understanding and possibly manipulating more complex microbial communities, such as the gut microbiome, environmental microbial ecosystems, and organisms cultured in industrial bioreactors (*8*).

NEW MEDICINE

Inherited risk of disease: From few (environmental-independent) to many (environmental-dependent)

For the past decade, various kinds of genomic sequence analyses have been the predominant use of omics in biomedical research. Most of the effort and funding has gone into genome-wide association studies (GWAS), in which the genomes of patients with common complex diseases (CCDs) are screened to seek associations between these diseases and allelic distributions of DNA variants. The ruling analytical strategy of GWA data sets has, however, been based on the notion that genetic risk for CCDs is largely independent of environmental risk—a notion that can be challenged (Fig. 1). Although several common disease–associated DNA sequence variants (that is, with a minor allele frequency of >5%) have been identified in GWAS (*9*) using this strategy, they explain no more than 5 to 10% of the total variation in CCD risk (*10*). Therefore, the huge investments in GWAS have not yet paid off. Fortunately, however, a portion of the unaccounted 85 to 90% disease variation lies hidden in GWAS data sets but can be revealed using NEW strategies by also uncovering environment-dependent disease-risk variants.

So far, most analyses of GWAS data sets have considered DNA variants one by one. This approach has generated a huge multiple-testing problem, which has led to stringent statistical cutoff levels that define a true DNA variant linkage to CCDs (*P* value threshold of less than 5×10^{-8}). Many of the subsignificant hits ($>5 \times 10^{-8}$) may be true risk loci; the question is how to distinguish them from false positives. In addition, current strategies of GWAS analysis build on the idea that disease-linked DNA variants exert their risk-modulating effects independently of both general environmental factors, such as smoking or type 2 diabetes, and local environmental factors present in tissues or cells—an assumption that is highly unlikely and currently changing. One decade ago, CCD risk was divided into inherited factors (genetic) and environmental factors (for example, smoking), with some overlap (gene-environment interactions). However, increasing evidence suggests that most genetic risk variants are dependent on particular environmental contexts to effect risks for CCDs. With NEW strategies, the combined risk-enrichment for groups of functionally associated genes (defined by networks) can greatly increase the amount of CCD risk information that can be extracted from GWAS data sets (Fig. 2B).

The need for patient cohorts with intermediate phenotypes

Organizing genomic sequencing and genome-wide activity data into coherent functional units represented by networks will help to reveal

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

how disease-associated variations in DNA actually contribute to the development of disease. Furthermore, empirically derived networks are necessary for describing the molecular mechanisms and biological processes that drive disease under the influences of inherited risk factors (genetic markers) and environmental risk factors (*11*, *12*). Thus, for biomedical research to take full advantage of omics sciences, a new type of patient cohort must be gathered that is monitored not only for inherited risks reflected in DNA variation profiles but also for a wide array of intermediate phenotypes. These phenotypes arise from the activity of groups of functionally associated genes (for example, in physiological pathways or networks) that are directly affected by disease DNA loci (that is, disease genotypes) and act to mediate the disease risk represented by the DNA loci. Changes to the intermediate phenotypes appear before the full-blown disease phenotype is manifested and can be captured at an omics level through screens of RNA transcripts, proteins, and metabolites in relevant cell types and tissues. Depending on whether the omics data are gathered from healthy or diseased tissues (or at intermediate steps), these networks represent molecular states of physiology transforming into pathology. Using these NEW insights, we can efficiently define key disease processes that drive CCD development.

Learning more from GWAS

The hallmark of NEW in biomedicine is its focus on molecular processes defined by interactions among all molecules that participate in the development of a CCD (*13*, *14*). This is in marked contrast to associations between isolated DNA sequence variations or single-target genes and CCDs. In contrast to such isolated molecules, complete disease networks ideally consist of all relevant genes and thus represent the entire complexity of a given disease and capture all pathological perturbations that alter disease-related processes. Thus, the network senses alterations in the microenvironments and the presence of certain DNA variants and reacts accordingly, leading to increased or decreased disease risk and development. In this way, disease networks both sense and mediate the effects of microenvironmental contexts and the genetic variations relevant to those contexts (*15*). We do not need to understand the exact structure (or topology) of the network, nor do we need to have monitored every single molecular component of the network to leverage it as a sensor. Instead, we need only enough components and relationships among the components to recognize the patterns of network states that reflect a disease state. And because disease networks represent many genes and gene products, they reflect a much greater proportion of the diverse contributions to CCD risk, contributions that can be detected by reexamining GWAS data sets for the enrichment of certain networks in the context of specific CCDs (*16*).

NEW in biomedicine will also enable us to tackle another aspect of CCD risk that has been largely neglected—the experimentally supported notion that most disease-linked DNA sequence variations exert their effects in specific contexts (*10*, *17*). For instance, the effects of most DNA sequence variants linked to type 2 diabetes in Caucasians are manifested only in patients with a body mass index above 26 (*18*). Similarly, DNA sequence variations linked to certain types of high blood pressure exert their negative effects only in the context of low physical activity (*19*).

Defining inherited risk dependent on environments

To unearth the 85 to 90% of CCD risk that is not explained by current analyses of GWAS data sets, it will be important to address contexts defined by macro- and microenvironmental factors (*20*). Macroenvironmental factors—mainly those that affect the individual through exposure to toxins, food intake, and other life-style factors—vary over time and alter the microenvironments within distinct tissues and cell types (Fig. 3A). The predominant microenvironment determines which DNA variants promote disease risk and the extent to which they do so. Thus, the effects of most disease-linked DNA variants probably vary with the phenotypic situation, as reflected by exposures to shifting macro- and microenvironments.

The differences in microenvironmental contexts from person to person and also over time in different organs are highly relevant to NEW medicine. If most DNA variants exert their risk-modifying effects on CCDs only within a certain set of microenvironments, how are we to define all of the combinations of DNA variants and microenvironments that increase CCD risk? The answer to this conundrum highlights the importance of adopting NEW medicine. Because these combinations might be rare (that is, found in a limited number of patients) and also have limited effect on risk over time, the only way to detect these risk scenarios is by integrating various omics measures to understand how they relate to one another, using DNA variation as a source of systematic perturbation to make causal inferences among the molecular phenotypes of interest (Fig. 3B). From the relationships defined by different layers of omics data, the molecular networks of disease can be inferred using relatively simple (Fig. 1) as well as sophisticated mathematical algorithms. With an atlas of human networks in hand, GWAS data sets can be reanalyzed to confirm DNA variants linked to CCDs via the network given sets of environmental contexts and to "rank" the degree of CCD linkage as indicated by the GWA (*21*–*23*).

From this perspective, an essential question that arises is whether the architecture (that is, the node connectivity structure) of these disease networks is largely unaffected by shifting contexts or whether shifts in context induce weak links that affect the network activity, predisposing to disease. Previous studies indicate that the architecture of biological networks is conserved through evolution (*24*–*26*), suggesting that the overall wiring diagram may be somewhat robust to changes in the contexts. If so, context-dependent risk variants will more likely induce changes to the network activity potentially by increasing or decreasing the number of edges and nodes in the network, thereby perturbing the molecular processes and biological functions defined by that network.

Regardless of how context-driven or non–context-driven genetic perturbations transform physiological molecular networks into those that drive disease, reconstructing networks from a new type of CCD cohort characterized by several layers of omics measurements in and between affected organs (Fig. 4) (*27*) is, in our view, critical to elucidating how physiological molecular processes are turned into pathological ones and thereby reveal the full complexity of CCDs. Networks specific to interactions between organs are particularly interesting from the CCD perspective because they are likely more important in later phases of disease development, when pathological changes are spreading across the borders of individual organs (*27*) (Fig. 4; see overlapping networks). Insights into DNA variants and microenvironments that are central for the cross-tissue spread of disease will be essential for providing the correct therapy for such events and improving the survival of patients with severe disease.

NEW HEALTH CARE

NEW strategies for equal and cost-effective health care

Like education, health care should be a fundamental right of all citizens. However, even in countries such as Sweden, where almost all

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

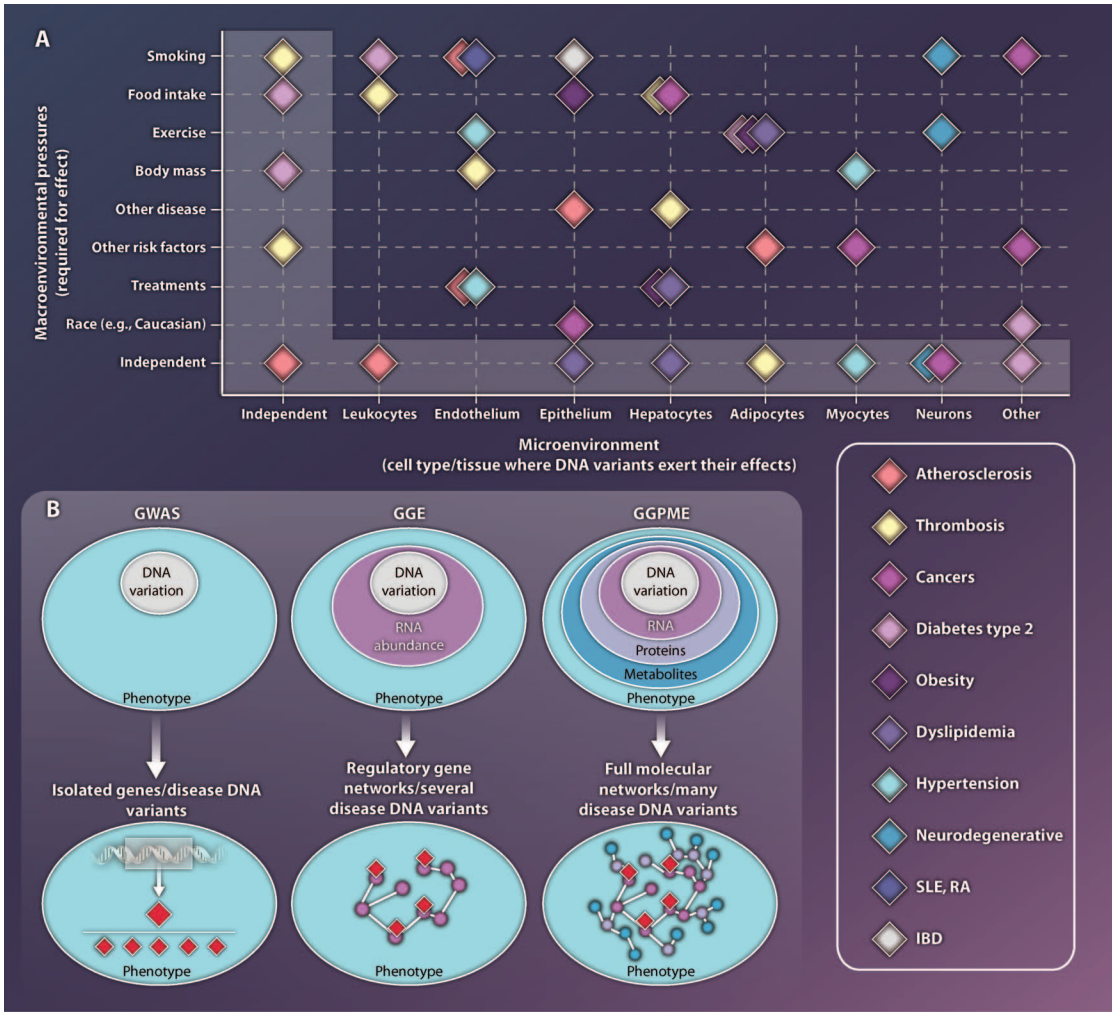

Fig. 3. Context-dependent inherited risk and the importance of intermediate phenotypes in clinical research. (**A**) The diagram plots sets of genetic risk variants according to their dependencies on contexts in the macroenvironment (such as life-style factors and exposures to environmental toxins) that, over time, transform microenvironments (in cell and tissue types) to activate DNA variants that then exert their risk-promoting effects on certain CCDs. Some genetic risk variants are environment-independent, and their disease associations may be detected early in life (light purple shading, left and bottom parts of the graph); the effects of other genetic risk variants are age-related (such as epigenetic changes), because exposures to macroenvironments alter microenvironments over time (dark purple background); environment-dependent DNA risk variants become increasingly important for CCD development later in life. Color-coded key for CCDs that are affected by DNA risk variants is shown below the graph at the right. SLE, systemic lupus erythematosus; RA, rheumatoid arthritis; IBD, inflammatory bowel diseases. (**B**) The importance of intermediate phenotypes in clinical studies of CCDs. **Left:** Traditional GWAS are based on the idea that genetic variations follow Mendelian inheritance and are relatively infrequent and context-independent, even for complex biological events and diseases. Most CCD-linked variants will not be discovered in this way, because the genetic perturbation (top circle, gray center) is too weak to be sensed by the disease phenotype (top circle, blue outer area). Some DNA risk variants that are context-independent (those that are common in the population) can be revealed with a GWAS design alone [blue bottom circle, large red diamonds shown "above the surface" (horizontal line)]; however, such studies do not explain the full variation in disease phenotypes (blue bottom circle, smaller red diamonds below the surface). **Middle:** In genetics of gene expression (GGE) studies (top circle), the apprehending of an intermediate phenotype of mRNA abundance (a measure of, for example, gene expression) (top circle, intermediate purple area) from patients and control individuals allows additional DNA variants to be identified from GWAS data sets (top circle, gray center)—in particular those that are context-dependent—thereby explaining more of the variations in disease phenotypes. This is achieved because intermediate gene expression data provide a more proximal sensor of DNA variation than does the clinical phenotype alone (top circle, outer blue area; compare with GWAS design alone). A GGE design thus allows for the inference of disease networks [in bottom circle, nodes (purple)] that harbor several DNA risk variants (shown as red diamonds in the network) linked to disease where these networks act to drive disease phenotypes (blue background). **Right:** The top circle depicts genetics of gene (gray center), protein (intermediate purple area), and metabolite (intermediate dark turquoise area) expression (GGPME) studies, which provide an even richer collection of proximal sensors of DNA variation that inform the clinical phenotype (top circle, outer blue area). Bottom circle: The identification of several layers of genome-wide measurements that sense the flow of DNA information allows inference of complex full disease networks (in bottom circle) with all disease-linked DNA risk variants (red diamonds) in contrast to networks whose effects are reflected in changes in mRNA concentrations alone. Dark purple, light purple, and dark turquoise network nodes (circles) are derived from genome-wide RNA, protein, and metabolite measurements, respectively; the associated phenotypes are depicted by the blue background.

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

health care is provided in the public sector through taxation, equal access has not fully been achieved. Today's health care distribution is inefficient and, therefore, costly, because we are constantly battling (reacting to) manifest diseases. Most patients who seek treatment for CCDs already have symptoms and so require reactive care. Of course, such care is appropriate for some conditions, such as bone fractures. However, for most CCDs, care begins far too late to cure the disease, which thus becomes a condition that requires lifelong treatment. Often, care is focused on preventing the disease from getting worse rather than on restoring health.

Besides the obvious benefits of preventive care (treatment dispensed before disease symptoms surface) for individual patients, there are savings to be made for the entire health care system (*28*). The extent of savings from today's preventive measures can be debated (*29*). However, with NEW health care evolving as a result of omics sciences, preventive measures have the potential to vastly reduce health care costs for societies, enabling more equal access to treatment.

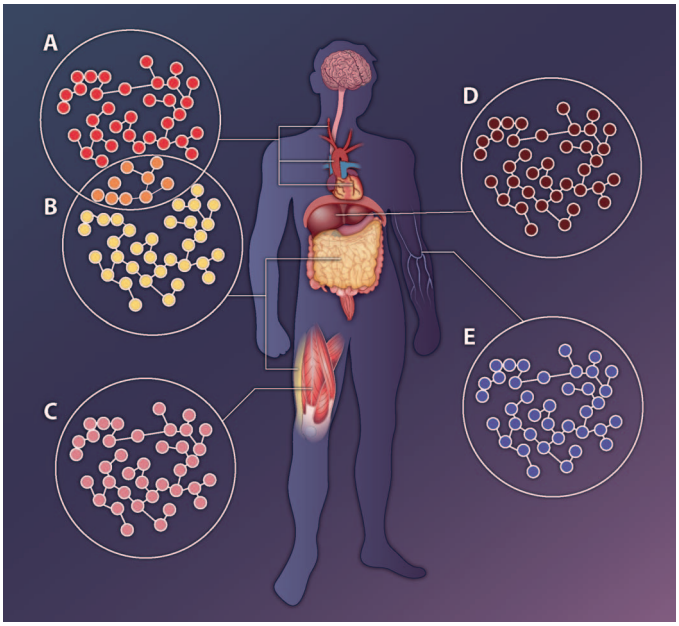

Fig. 4. Tissue-specific and cross-tissue molecular networks. Tissue-specific networks: (**A**) red nodes, carotid or coronary lesions, atherosclerotic arterial wall, control arterial wall; (**B**) yellow nodes, subcutaneous or omental visceral fat; (**C**) pink nodes, skeletal muscle; (**D**) brown nodes, liver; (**E**) blue nodes, blood cells (for example, leukocytes, such as monocytes/macrophages). Orange nodes are part of networks shared across tissues [in this example, cross-tissue communication is shown for arterial wall (red) and fat (yellow) samples, but such crosstalk occurs among many tissues]. Sampling of patient tissues and, if possible, control individuals is central to a systems approach to CCDs. Here, the focus is on cardiovascular and metabolic diseases: Samples from several tissue and organ locations are necessary to study control individuals and cohorts of patients with cardiovascular and related metabolic diseases (such as obesity, diabetes, and dyslipidemia). The tissue samples are then used to isolate DNA, RNA, proteins, and possibly metabolites for genome-wide data generation; strict protocols for tissue isolation and immediate processing are crucial for data quality and meaningful downstream analyses as is careful clinical characterization of the CCD phenotypes. Molecular intermediate phenotypes from several disease-relevant tissues are then used for disease-network inference. Different organs and tissues have distinct mRNA footprints, and genome-wide RNA abundance measurements in each tissue allow detection of DNA variants that affect gene expression (that is, general and tissue-specific eSNPs). Tissue-specific causal networks can be inferred from computer-supported integrative analysis of omics data sets, including DNA-variation data. The disease impact of inferred networks is also determined by examining the network's relative enrichment with inherited risk for CCDs using existing GWAS cohorts (Fig. 1). Some genes specialize in cross-tissue communication (*27*), so that some parts of networks (in this example, shown as orange nodes) are shared among different tissues and likely are responsible for related molecular activities. Genes in cross-tissue networks do not appear to belong to tissue-specific networks (and visa versa). These networks are believed to be particularly important for cardiovascular disease, cancers, and metabolic diseases, which involve many organs, particularly in late disease stages. In the future clinic, where NEW strategies rule, disease networks (tissue-specific and cross-tissue) will be used for early disease detection and for monitoring effects of preventive therapies.

Principals for NEW healthcare

NEW healthcare begins by providing each and every individual with his or her DNA profile early in life. This profile serves as a map of the total risk for disease at a young age (Fig. 3A, context-independent risk), maximizing the ability to implement practices and treatments designed to maintain good health. Individuals with a strong family history of CCDs may require more frequent health examinations in early life stages (late 30s to early 40s). The DNA map will also help people avoid certain macroenvironments that can lead to microenvironmental changes that trigger context-dependent DNA variants linked to CCDs (Fig. 3A, context-dependent risk). Emerging technologies are also making it possible to identify epigenetic changes in the genome at the time of sequencing (*30*), and given recent evidence that such imprints can be passed on to future generations (*31*), these data generated from appropriate cell types will be invaluable in assessing risk of disease in early life.

However, depending on how individuals end up living their lives—an outcome governed both by conscious choices and by chance—the risk scenarios and the preventive measures needed to avert them will vary. Therefore, a DNA profile alone will not suffice to generate a full individual risk profile for CCDs. As mentioned above, most risk for CCDs as we age reflects the context-dependent effects of DNA variants. Thus, as the individual grows older, his or her DNA profile may need to be regenerated multiple times to pick up shifts in epigenetic patterns, and such profiles will need to be complemented with other omics measures that provide an estimate of the current state of the body in terms of risk for CCDs. The precise nature and required resolution of this activity measure remains to be established. Although the entire volume of blood (~5 liters) circulates throughout the body almost once per minute and, in theory, provides information about the molecular status of all organs including the brain, this information is likely too unfiltered to be of diagnostic use. A more realistic approach would be to sort the various types of blood cells divided into well-defined categories and to decipher their transcriptomes. Indeed, there is already evidence that blood leukocytes play a key role in the risk of developing at least two major CCDs, namely, obesity/type 2 diabetes and atherosclerosis/cardiovascular disease (*32*). Combined with the DNA profile, cell type–specific blood mRNA profiles could provide snapshot of the risk status of the individual.

In the future, activity profiles will be more comprehensive than they are today, with 1000 to 2000 markers that indicate the current status of the molecular physiology of all organs and whether there are any signs of molecular pathology (such as tumor growth, atherosclerosis,

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

inflammation, or immune responses). Today's markers (for example, liver enzymes, C-reactive protein, and leukocyte count) can provide some of this information, but the markers of tomorrow will be more tissue- and disease-specific. If a DNA profile and an activity screen of some sort that includes blood components suggest that pathological processes are likely occurring in an organ, the next step might be imaging with labeled markers (such as oxygen) to detect areas of increased cell turnover and metabolism in the indicated tissue. If the blood markers and DNA profile instead indicate a systemic disease that involves many organ systems, a second noninvasive step would be to extend the analyses to whole-genome scans, which will allow for the real-time delineation of detailed molecular networks that indicate the type and status of the systemic disease.

Early diagnosis of subclinical molecular pathologies will require preventive treatments that squelch or slow disease development. With NEW health care, all treatments will be adjusted to the individual circumstances (DNA profile) and needs (RNA/protein profile of diseased tissue), and clinical outcomes of treatment will be monitored to assess efficacy and side effects. Defining disease networks will guide treatment (type and dosage) and continued monitoring will reveal whether disease networks are gradually replaced by normal organ physiology. In such a scenario, networks form the cornerstone of personal and preventive medicine.

REFERENCES AND NOTES

1. N. A. Christakis, J. H. Fowler, The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).
2. A. L. Barabási, Z. N. Oltvai, Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
3. B. L. Pike, K. E. Saylors, J. N. Fair, M. Lebreton, U. Tamoufe, C. F. Djoko, A. W. Rimoin, N. D. Wolfe, The origin and prevention of pandemics. *Clin. Infect. Dis.* **50**, 1636–1640 (2010).
4. E. E. Schadt, S. Turner, A. Kasarskis, A window into third-generation sequencing. *Hum. Mol. Genet.* **19**, R227–R240 (2010).
5. E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, G. P. Nolan, Computational solutions to large-scale data management and analysis. *Nat. Rev. Genet.* **11**, 647–657 (2010).
6. C. S. Chin, J. Sorenson, J. B. Harris, W. P. Robins, R. C. Charles, R. R. Jean-Charles, J. Bullard, D. R. Webster, A. Kasarskis, P. Peluso, E. E. Paxinos, Y. Yamaichi, S. B. Calderwood, J. J. Mekalanos, E. E. Schadt, M. K. Waldor, The origin of the Haitian cholera outbreak strain. *N. Engl. J. Med.* **364**, 33–42 (2011).
7. D. A. Rasko, D. R. Webster, J. W. Sahl, A. Bashir, N. Boisen, F. Scheutz, E. E. Paxinos, R. Sebra, C. S. Chin, D. Iliopoulos, A. Klammer, P. Peluso, L. Lee, A. O. Kislyuk, J. Bullard, A. Kasarskis, S. Wang, J. Eid, D. Rank, J. C. Redman, S. R. Steyert, J. Frimodt-Møller, C. Struve, A. M. Petersen, K. A. Krogfelt, J. P. Nataro, E. E. Schadt, M. K. Waldor, Origins of the *E. coli* strain causing an outbreak of hemolytic–uremic syndrome in Germany. *N. Engl. J. Med.* **365**, 709–717 (2011).
8. Y. Nakanishi, S. Fukuda, E. Chikayama, Y. Kimura, H. Ohno, J. Kikuchi, Dynamic omics approach identifies nutrition-mediated microbial interactions. *J. Proteome Res.* **10**, 824–836 (2011).
9. T. A. Manolio, Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363**, 166–176 (2010).
10. E. E. Eichler, J. Flint, G. Gibson, A. Kong, S. M. Leal, J. H. Moore, J. H. Nadeau, Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).
11. S. A. Monks, A. Leonardson, H. Zhu, P. Cundiff, P. Pietrusiak, S. Edwards, J. W. Phillips, A. Sachs, E. E. Schadt, Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**, 1094–1105 (2004).
12. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M. F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
13. S. Hägg, J. Skogsberg, J. Lundström, P. Noori, R. Nilsson, H. Zhong, S. Maleki, M. M. Shang, B. Brinne, M. Bradshaw, V. B. Bajic, A. Samnegård, A. Silveira, L. M. Kaplan, B. Gigante, K. Leander, U. de Faire, S. Rosfors, U. Lockowandt, J. Liska, P. Konrad, R. Takolander, A. Franco-Cereceda, E. E. Schadt, T. Ivert, A. Hamsten, J. Tegnér, J. Björkegren, Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: The Stockholm Atherosclerosis Gene Expression (STAGE) study. *PLoS Genet.* **5**, e1000754 (2009).
14. A. J. Lusis, J. N. Weiss, Cardiovascular networks: Systems-based approaches to cardiovascular disease. *Circulation* **121**, 157–170 (2010).
15. E. E. Schadt, Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
16. V. Emilsson, G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink, J. Zhu, S. Carlson, A. Helgason, G. B. Walters, S. Gunnarsdottir, M. Mouy, V. Steinthorsdottir, G. H. Eiriksdottir, G. Bjornsdottir, I. Reynisdottir, D. Gudbjartsson, A. Helgadottir, A. Jonasdottir, A. Jonasdottir, U. Styrkarsdottir, S. Gretarsdottir, K. P. Magnusson, H. Stefansson, R. Fossdal, K. Kristjansson, H. G. Gislason, T. Stefansson, B. G. Leifsson, U. Thorsteinsdottir, J. R. Lamb, J. R. Gulcher, M. L. Reitman, A. Kong, E. E. Schadt, K. Stefansson, Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
17. G. Gibson, The environmental contribution to gene expression profiles. *Nat. Rev. Genet.* **9**, 575–581 (2008).
18. V. Lyssenko, A. Jonsson, P. Almgren, N. Pulizzi, B. Isomaa, T. Tuomi, G. Berglund, D. Altshuler, P. Nilsson, L. Groop, Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.* **359**, 2220–2232 (2008).
19. T. Rankinen, T. Church, T. Rice, N. Markward, A. S. Leon, D. C. Rao, J. S. Skinner, S. N. Blair, C. Bouchard, Effect of endothelin 1 genotype on blood pressure is dependent on physical activity or fitness levels. *Hypertension* **50**, 1120–1125 (2007).
20. D. Thomas, Gene–environment-wide association studies: Emerging approaches. *Nat. Rev. Genet.* **11**, 259–272 (2010).
21. D. B. Goldstein, Genomics and biology come together to fight HIV. *PLoS Biol.* **6**, e76 (2008).
22. M. F. Moffatt, M. Kabesch, L. Liang, A. L. Dixon, D. Strachan, S. Heath, M. Depner, A. von Berg, A. Bufe, E. Rietschel, A. Heinzmann, B. Simma, T. Frischer, S. A. Willis-Owen, K. C. Wong, T. Illig, C. Vogelberg, S. K. Weiland, E. von Mutius, G. R. Abecasis, M. Farrall, I. G. Gut, G. M. Lathrop, W. O. Cookson, Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
23. E. E. Schadt, C. Molony, E. Chudin, K. Hao, X. Yang, P. Y. Lum, A. Kasarskis, B. Zhang, S. Wang, C. Suver, J. Zhu, J. Millstein, S. Sieberts, J. Lamb, D. GuhaThakurta, J. Derry, J. D. Storey, I. Avila-Campillo, M. J. Kruger, J. M. Johnson, C. A. Rohl, A. van Nas, M. Mehrabian, T. A. Drake, A. J. Lusis, R. C. Smith, F. P. Guengerich, S. C. Strom, E. Schuetz, T. H. Rushmore, R. Ulrich, Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
24. V. F. Hinman, A. T. Nguyen, R. A. Cameron, E. H. Davidson, Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13356–13361 (2003).
25. H. B. Fraser, A. M. Moses, E. E. Schadt, Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2977–2982 (2010).
26. H. B. Fraser, E. E. Schadt, The quantitative genetics of phenotypic robustness. *PLoS One* **5**, e8635 (2010).
27. R. Dobrin, J. Zhu, C. Molony, C. Argman, M. L. Parrish, S. Carlson, M. F. Allan, D. Pomp, E. E. Schadt, Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol.* **10**, R55 (2009).
28. J. T. Cohen, P. J. Neumann, M. C. Weinstein, Does preventive care save money? Health economics and the presidential candidates. *N. Engl. J. Med.* **358**, 661–663 (2008).
29. J. E. Fielding, C. G. Husten, J. H. Richland, Does preventive care save money? *N. Engl. J. Med.* **358**, 2847 (2008).
30. B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
31. M. D. Anway, A. S. Cupp, M. Uzumcu, M. K. Skinner, Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466–1469 (2005).
32. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M. F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
33. M. Morley, C. M. Molony, T. M. Weber, J. L. Devlin, K. G. Ewens, R. S. Spielman, V. G. Cheung, Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
34. E. E. Schadt, S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb, G. Cavet, P. S. Linsley, M. Mao, R. B. Stoughton, S. H. Friend, Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
35. J. Zhu, M. C. Wiener, C. Zhang, A. Fridman, E. Minch, P. Y. Lum, J. R. Sachs, E. E. Schadt, Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput. Biol.* **3**, e69 (2007).
36. H. B. Fraser, X. Xie, Common polymorphic transcript variation in human disease. *Genome Res.* **19**, 567–575 (2009).
37. S. Kathiresan, O. Melander, C. Guiducci, A. Surti, N. P. Burtt, M. J. Rieder, G. M. Cooper, C. Roos, B. F. Voight, A. S. Havulinna, B. Wahlstrand, T. Hedner, D. Corella, E. S. Tai, J. M. Ordovas, G. Berglund, E. Vartiainen, P. Jousilahti, B. Hedblad, M. R. Taskinen, C. Newton-Cheh, V. Salomaa, L. Peltonen, L. Groop, D. M. Altshuler, M. Orho-Melander, Six new loci associated with blood

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197 (2008).
38. D. A. Smirnov, M. Morley, E. Shin, R. S. Spielman, V. G. Cheung, Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**, 587–591 (2009).
39. M. Mehrabian, H. Allayee, J. Stockton, P. Y. Lum, T. A. Drake, L. W. Castellani, M. Suh, C. Armour, S. Edwards, J. Lamb, A. J. Lusis, E. E. Schadt, 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 (2005).
40. A. Ghazalpour, S. Doss, S. S. Sheth, L. A. Ingram-Drake, E. E. Schadt, A. J. Lusis, T. A. Drake, Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* **6**, R59 (2005).
41. X. Yang, J. L. Deignan, H. Qi, J. Zhu, S. Qian, J. Zhong, G. Torosyan, S. Majid, B. Falkard, R. R. Kleinhanz, J. Karlsson, L. W. Castellani, S. Mumick, K. Wang, T. Xie, M. Coon, C. Zhang, D. Estrada-Smith, C. R. Farber, S. S. Wang, A. van Nas, A. Ghazalpour, B. Zhang, D. J. Macneil, J. R. Lamb, K. M. Dipple, M. L. Reitman, M. Mehrabian, P. Y. Lum, E. E. Schadt, A. J. Lusis, T. A. Drake, Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat. Genet.* **41**, 415–423 (2009).

Acknowledgments: We thank S. Ordway for editorial assistance. **Funding:** Swedish Heart Lung Foundation (J.L.M.B.), Swedish Research Council (J.L.M.B.), and PROCARDIS (precocious coronary artery disease) consortium, a European Commission Framework 6 integrated project (J.L.M.B., Clinical Gene Networks). **Competing interests:** E.E.S. was the Chief Scientific Officer and serves on the scientific advisory board of Pacific Biosciences and is a consultant for Numedii and GNS Healthcare. J.L.M.B. is acting Chief Executive Officer of and shareholder in Clinical Gene Networks AB.

Submitted 11 January 2011
Accepted 16 November 2011
Published 4 January 2012
10.1126/scitranslmed.3002132

Citation: E. E. Schadt, J. L. M. Björkegren, NEW: Network-enabled wisdom in biology, medicine, and health care. *Sci. Transl. Med.* **4**, 115rv1 (2012).

Downloaded from http://stm.sciencemag.org/ on January 1, 2017

NEW: Network-Enabled Wisdom in Biology, Medicine, and Health Care
Eric E. Schadt and Johan L. M. Björkegren (January 4, 2012)
Science Translational Medicine **4** (115), 115rv1. [doi: 10.1126/scitranslmed.3002132]

Editor's Summary

Editor's Summary

Multidimensional networks are required to model human physiology in general and to identify key drivers of pathology in individual patients.

The following resources related to this article are available online at http://stm.sciencemag.org.
This information is current as of January 1, 2017.

Article Tools Visit the online version of this article to access the personalization and article tools:
http://stm.sciencemag.org/content/4/115/115rv1

Related Content The editors suggest related resources on *Science*'s sites:
http://stm.sciencemag.org/content/scitransmed/5/202/202ed15.full
http://stm.sciencemag.org/content/scitransmed/6/225/225ed5.full
http://science.sciencemag.org/content/sci/347/6224/1257601.full
http://stm.sciencemag.org/content/scitransmed/7/297/297ed10.full
http://stm.sciencemag.org/content/scitransmed/7/300/300ps17.full
http://stm.sciencemag.org/content/scitransmed/8/322/322fs3.full
http://stm.sciencemag.org/content/scitransmed/8/322/322ra9.full
http://stm.sciencemag.org/content/scitransmed/8/322/322pc1.full

Permissions Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl

Science Translational Medicine (print ISSN 1946-6234; online ISSN 1946-6242) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Copyright 2017 by the American Association for the Advancement of Science; all rights reserved. The title *Science Translational Medicine* is a registered trademark of AAAS.

Downloaded from http://stm.sciencemag.org/ on January 1, 2017