



The Impact of Network Medicine in Gastroenterology and Hepatology

Citation

Baffy, György. 2013. "The Impact of Network Medicine in Gastroenterology and Hepatology." *Clinical Gastroenterology and Hepatology* 11 (10) (October): 1240–1244. doi:10.1016/j.cgh.2013.07.033.

Published Version

doi:10.1016/j.cgh.2013.07.033

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:34864844>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

The impact of network medicine in gastroenterology and hepatology

György Baffy, MD, PhD

Affiliations: Department of Medicine, VA Boston Healthcare System and Brigham and Women's Hospital, Harvard Medical School, Boston, MA

Correspondence: Dr. György Baffy, VA Boston Healthcare System, 150 S. Huntington Ave., Room 6A-46, Boston, Massachusetts 02130. Email: gbaffy@bics.bwh.harvard.edu.
Phone: 857-364-4327. Fax: 857-364-4179.

Word count: 2,894 (main text, references, and figure legends)

Conflicts of interest: No relevant conflicts of interest exist.

Abstract

In the footsteps of groundbreaking achievements made by biomedical research, another scientific revolution is unfolding. Systems biology draws from the chaos and complexity theory and applies computational models to predict emerging behavior of the interactions between genes, gene products, and environmental factors. Adaptation of systems biology to translational and clinical sciences has been termed network medicine and is likely to change the way we think about preventing, predicting, diagnosing, and treating complex human diseases. Network medicine finds gene-disease associations by analyzing the unparalleled digital information discovered and created by high-throughput technologies (dubbed as ‘omics’ science) and links genetic variance to clinical disease phenotypes through intermediate organizational levels of life such as the epigenome, transcriptome, proteome, and metabolome. Supported by large reference databases, unprecedented data storage capacity, and innovative computational analysis, network medicine is poised to find links between conditions that were thought to be distinct, uncover shared disease mechanisms and key drivers of the pathogenesis, predict individual disease outcomes and trajectories, identify novel therapeutic applications, and help avoid off-target and undesirable drug effects. Recent advances indicate that these perspectives are increasingly within our reach for understanding and managing complex diseases of the digestive system.

Ever wondered why it is so difficult to recall a ‘textbook case’ of a patient with inflammatory bowel disease (IBD) or nonalcoholic fatty liver disease (NAFLD)? It is because while some disorders have distinct etiology and predictable outcomes, IBD and NAFLD cannot be attributed to a single cause and their course varies from being a minor dysfunction to an occasionally life-threatening illness^{1,2} (**Figure 1a**). These conditions are determined by interactions between multitudes of genes, gene products, and environmental factors (**Figure 1b**). In other words, complex diseases are the byproducts of biological systems in which a large number of intrinsically or extrinsically perturbed constituents interact in nonlinear or chaotic ways, span across successive levels of organization, and show emergence (i.e., system properties cannot be explained or predicted from knowing the parts)³. As a result, complex diseases often pose diagnostic challenges, defy prognostication, and require a wide repertoire of (often suboptimal) therapeutic interventions^{1,2}.

Although biomedical research has provided a wealth of information in the past century, it has not resolved the issue of complexity and another revolution was necessary to begin the process of organizing and analyzing data in an entirely new way. Systems biology is a rapidly evolving field that focuses on the complex interactions between biological components and applies mathematical and computational models to predict emerging behavior⁴⁻⁶. Adaptation of this approach to translational and clinical sciences has been termed network medicine⁷. This new discipline aims to improve our understanding of complex diseases, including those of the digestive system, marking the beginning of systems gastroenterology and systems hepatology.

Technological primer: Systems biology strategies for finding gene-disease associations

Big data science: Cataloging the parts

A number of recent advances in biomedical technologies began to tackle complexity with increasing success. A significant challenge is to enumerate the components of various biological systems that contribute to the wide variability of complex disease phenotypes. For instance, there are over 50 million commonly occurring DNA sequence variants (single-nucleotide polymorphisms or SNPs) in the human genome that combine in a myriad of ways to affect disease development and outcomes in response to genetic and environmental perturbations.

Several methods are available to analyze genotype-phenotype associations in complex diseases. Genome-wide association (GWA) studies explore the relationship between molecular and clinical phenotypes by comparing frequencies of DNA sequence variants between two groups of individuals (usually diseased subjects and healthy controls)⁸. Identification of allelic polymorphism near the *IL28B* gene predicting therapeutic responsiveness in chronic hepatitis C was an important example for the utility of GWA approach⁹⁻¹¹. Next generation sequencing (parallel sequencing followed by reassembly of genomic DNA fragments) has recently made genomic analysis extremely fast and affordable, permitting whole genome sequencing and generating digital information at an extraordinary pace^{5, 12}. DNA microarrays quantify gene expression simultaneously for hundreds of genes and correlate the findings with clinical characteristics such as tumor

progression, recurrence, and drug response¹³. The search for disease-associated genes is narrowed by computational methods of prioritization such as gene set enrichment analysis to identify candidate genes at the top or the bottom of all examined genes ranked by expression relative to a reference microarray or published database¹⁴.

Further complexity results from epigenetic variability (e.g., histone acetylation and DNA methylation), post-translational protein modification, and a large number of regulatory metabolites that account for intermediate molecular phenotypes and link genetic risk variants to clinically manifest disease^{12, 15}. These post-genomic organizational scales are analyzed by additional high-throughput methodologies, including liquid chromatography, mass spectrometry, and proton MR spectrometry (**Figure 2a**).

Storing and organizing biomedical information amassed by these methods is another fundamental task that must be undertaken to realize the vision of systems biology. The scale and quality of data require high-power computational tools to model and understand complex system behavior in health and disease¹⁶. After the introduction of terabytes, petabytes, and exabytes, digital information created this way is soon to be measured in zettabytes (ZB), denoting one sextillion or 10^{21} bytes.

Systems biology also requires reference libraries that contain validated information, allow comparison of experimental data, and follow standardized terminology. A number of online repositories are freely accessible such as the Online Mendelian Inheritance in Man, listing all diseases with a known genetic component¹⁷, Gene Ontology, a

compendium of controlled vocabulary for describing genes and gene products¹⁸, or even PubMed, which currently includes over 22 million abstracts ('bibliome'). Digestive disease databases are also becoming available: IBDsite includes metagenomic and metatranscriptomic data about the commensal bacteria involved in IBD¹⁹, while the Library of Molecular Associations collects information on molecular associations in chronic liver disease²⁰.

Network-based analysis of complex diseases: mapping the interactions

Once components of a biological system are identified, the next step is to analyze their interactions and understand their emerging role in the development of complex diseases. The mathematical graph (network) theory is a convenient framework for visualizing and interpreting the structural and functional relationships between genes, gene products, and environmental factors⁷. In simplest terms, networks consist of nodes connected with links and arranged in a particular physical layout (topology). Nodes in biological networks can be genes, coding and non-coding RNAs, proteins, metabolites, cells, individuals, diseases, drugs, and so on. Links represent a variety of relationships between these nodes: genes regulated by the same microRNA, proteins associated with the same disease, metabolites involved in the same pathway, individuals exhibiting the same clinical symptoms, and drugs used to treat the same diseases or causing the same adverse reactions.

Networks can be described by scientific formulas and modeled by mechanical circuitries to simulate real-life situations in biological systems^{21, 22}. Biological networks commonly follow a scale-free topology in which most nodes have limited connections, while a few nodes (hubs) have a progressively high number of links²³. Modules in scale-free networks

represent clusters of highly interconnected nodes with similar or sequentially arranged functions (e.g., metabolic pathways or signaling cascades)²⁴.

Networks can be created by using curated databases, computational predictions, or systematic experimental strategies⁶ (**Figure 2b**). A key assumption in these studies is that groups of nodes (e.g., genes, miRNAs, proteins, and metabolites) associated with a specific disorder represent a disease module and additional, unknown disease components may be identified by their topographical position to disease-relevant nodes⁵ (**Figure 2c**). A frequently used approach follows the concept of guilt-by-association, in which new candidate genes are identified based on their association with genes already known to be involved in a given condition. For instance, Goh et al. created a comprehensive human disease network in which additional genes implicated in the same or similar disease can be identified either via direct links or by being in the same network neighborhood²⁵. Thus, network-based approaches enhance the search for disease genes and give new meaning to clinical syndromes and comorbidities.

The beginnings of systems gastroenterology: Early findings and implications

Network-based approaches to studying complex diseases of the digestive system are rapidly gaining hold. For instance, analyzing gene expression profiles and human-viral protein interactions ('virhostome') to understand the pathogenesis and identify therapeutic targets in chronic infection with the hepatitis C virus (HCV) is a major goal of systems hepatology²⁶. Systematic expression profiling of microRNAs (miRNAs) in HCV-infected and healthy liver tissue identified 38 regulatory modules within a miRNA-

mRNA network, predicting a number of potential targets for the therapy of chronic hepatitis C beyond the already known miR-122²⁷.

Zheng et al. analyzed molecular networks during HCV-associated progression of hepatocellular carcinoma (HCC) by combining published gene expression profiles and high-throughput protein interaction data with gene enrichment analysis²⁸. They created networks from pairwise comparison of gene expression data corresponding to successive pathological stages of hepatocarcinogenesis (normal, cirrhosis, dysplasia, early HCC, and advanced HCC) and identified hub proteins in each network (e.g., LCK in normal-to-cirrhosis, MMP and TIMP in cirrhosis-to-dysplasia, and CDC2 in early-to-advanced HCC). The largest change in network topology was seen between dysplasia and early cancer, pointing to this transition as a critical step in hepatocarcinogenesis²⁸.

To find biomarkers that predict the clinical behavior of HCC more reliably, Zhang et al. conducted computational data mining from 3 published sets of microarray analysis on HCC and corresponding normal liver tissue²⁹. Networks created from up-regulated and down-regulated genes common to all 3 sets were analyzed to identify hub genes as candidate markers. These hub genes encode key oncogenic signaling pathways (e.g., *EGFR*, *FOS*, *MAPK*, and *SMAD2*) and their use in the HCC classifier improves predictive accuracy²⁹. This study argues for the utility of combining topological features with gene co-expression data to find successful biomarkers and reminds us that high connectivity genes are not necessarily expressed at high levels.

Network-based approaches to find opportunities for drug repositioning, (i.e., identifying new molecular targets for existing drugs by computational methods) is a cost-saving strategy for drug discovery³⁰. Dudley et al. compared gene expression signatures of IBD against reference mRNA expression data of cultured cells exposed to 164 drug compounds in the Connectivity Map, a publicly available compendium of drug-gene associations³¹. Drugs were scored based on the degree of discordance between disease-affected and drug-affected gene expression patterns to infer potential therapeutic indications. By following this strategy, the anticonvulsant topiramate was identified as a ‘repurposed’ drug for IBD and its efficacy was experimentally verified in TNBS-induced colitis, illustrating the promise of computational drug repositioning through systems pharmacology³¹.

The utility of creating networks by using arbitrary free-text resources has recently been demonstrated by Sookoian et al., who have set out to find molecular mechanisms shared between alcoholic and nonalcoholic fatty liver disease³². The authors searched for co-occurrences of genes and proteins in PubMed abstracts and used computational gene prioritization (ranking based on similarity scores of known genes involved in these conditions) to create functional molecular maps for shared disease pathways. This entirely *in silico* analysis concluded that insulin signaling is impaired in both conditions, but only NAFLD is linked to cardiovascular disease. Moreover, both conditions are strongly associated with cancer-related pathways, indicating that fatty liver, regardless of the cause, is an emerging mechanism of oncogenic activation³².

Conclusions

Rooted in the mathematical graph theory and supported by the philosophy of chaos and complexity, network medicine may seem at first a challenging and intimidating concept. However, this rapidly evolving field holds a number of promises and is likely to change the way we think about preventing, predicting, diagnosing, and treating human disease. As in other fields of network medicine, systems gastroenterology is poised to find links between conditions that were thought to be distinct, uncover common underlying disease mechanisms and key drivers of the pathogenesis, predict individual disease outcomes and trajectories, identify novel therapeutic applications, and help avoid off-target and undesirable drug effects.

Unmistakably, network medicine has several limitations. Computational analysis may only be as good as the data on which it is based. Big data statistics may become too inclusive with false positive discoveries and the resulting gene lists do not always result in useful biological insights. Thus, observational or experimental validation of *in silico* findings remains an important step of the network-based discovery process. Analyzing perturbed genetic networks compared to their normal counterparts could reveal important elements of the pathomechanism, but may not readily identify the environmental challenge.

Despite these and other concerns and cautions, network medicine has generated tremendous excitement. Improved biomarker discovery may reduce the cost of surveillance programs by filtering at-risk populations before undertaking expensive

testing. With the recent progress of next generation sequencing technologies, analysis of entire microbial communities is becoming a routine exercise and the human microbiome – outnumbering our own genome by more than 2 magnitudes – may soon become part of routine network analysis as outlined above. Some experts envision a world of network-enabled wisdom in which every individual is provided with a DNA profile early in life, updated from time to time to recalibrate the risk of disease¹⁶. The road of network medicine is increasingly traveled and we should buckle up for an exhilarating ride.

References

1. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;474:307-317.
2. Sookoian S, Pirola CJ. The genetic epidemiology of nonalcoholic fatty liver disease: toward a personalized medicine. *Clin Liver Dis* 2012;16:467-485.
3. Bar-Yam Y. *Dynamics of complex systems*. Westview, 1997.
4. Piro RM. Network medicine: linking disorders. *Hum Genet* 2012;131:1811-1820.
5. Chan SY, Loscalzo J. The emerging paradigm of network medicine in the study of human disease. *Circ Res* 2012;111:359-374.
6. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell* 2011;144:986-998.
7. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2010;12:56-68.
8. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008;299:1335-1344.

9. Ge D, Fellay J, Thompson AJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009;461:399-401.
10. Tanaka Y, Nishida N, Sugiyama M, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 2009;41:1105-1109.
11. Suppiah V, Moldovan M, Ahlenstiel G, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 2009;41:1100-1104.
12. Lee JS, Kim JH, Park YY, et al. Systems biology approaches to decoding the genome of liver cancer. *Cancer Res Treat* 2011;43:205-211.
13. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 1999;21:33-37.
14. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-15550.
15. Dumas ME. Metabolome 2.0: quantitative genetics and network biology of metabolic phenotypes. *Mol Biosyst* 2012;8:2494-2502.
16. Schadt EE, Bjorkegren JL. NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 2012;4:115rv111.
17. Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). *Hum Mutat* 2011;32:564-567.
18. Gene Ontology C. The Gene Ontology project in 2008. *Nucleic Acids Res* 2008;36:D440-444.

19. Merelli I, Viti F, Milanesi L. IBDsite: a Galaxy-interacting, integrative database for supporting inflammatory bowel disease high throughput data analysis. *BMC Bioinformatics* 2012;13 Suppl 14:S5.
20. Buchkremer S, Hendel J, Krupp M, et al. Library of molecular associations: curating the complex molecular basis of liver diseases. *BMC Genomics* 2010;11:189.
21. Albert R, Barabási AL. Statistical mechanics of complex networks. *Reviews of Modern Physics* 2002;74:47-97.
22. Alon U. *An introduction to systems biology : design principles of biological circuits*. Chapman & Hall/CRC, 2007.
23. Barabási AL, Albert R. Emergence of scaling in random networks. *Science* 1999;286:509-512.
24. Ravasz E, Barabasi AL. Hierarchical organization in complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2003;67:026112.
25. Goh KI, Cusick ME, Valle D, et al. The human disease network. *Proc Natl Acad Sci U S A* 2007;104:8685-8690.
26. Ireton RC, Gale M, Jr. Systems biology analyses to define host responses to HCV infection and therapy. *Curr Top Microbiol Immunol* 2013;363:143-167.
27. Peng X, Li Y, Walters KA, et al. Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics* 2009;10:373.

28. Zheng S, Tansey WP, Hiebert SW, et al. Integrative network analysis identifies key genes and pathways in the progression of hepatitis C virus induced hepatocellular carcinoma. *BMC Med Genomics* 2011;4:62.
29. Zhang Y, Wang S, Li D, et al. A systems biology-based classifier for hepatocellular carcinoma diagnosis. *PLoS One* 2011;6:e22426.
30. Sirota M, Dudley JT, Kim J, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011;3:96ra77.
31. Dudley JT, Sirota M, Shenoy M, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med* 2011;3:96ra76.
32. Sookoian S, Pirola CJ. Systems biology elucidates common pathogenic mechanisms between nonalcoholic and alcoholic-fatty liver disease. *PLoS One* 2013;8:e58895.

Figure legends

Figure 1. Phenotype distribution and determinants in complex disease

A. Disease course is usually predictable in cases of self-limiting (e.g., viral rhinitis), uniformly fatal (e.g., Huntington chorea), or chronic conditions with no significant consequences (e.g., skin tags). By contrast, complex diseases have a wide spectrum of severity that remains difficult to predict. **B.** Phenotypes in complex diseases are determined by interactions between genes (genotypes) and gene products (intermediate phenotypes) in response to intrinsic (genetic) and extrinsic (environmental) perturbations. Specific groups of genes (*circles*) are associated with specific disease phenotypes (*squares*) that may both overlap, indicating pleiotropy (same genes associated with various diseases) or ‘metadisease’ (various gene disorders associated with same disease). Studies of gene-disease associations in complex disease may elucidate shared etiology and pathogenesis of disparate disease phenotypes involving multiple organs (syndrome).

Figure 2. Network-based discovery of gene-disease associations in complex disease

A. Dysfunction of complex biological systems entailing hierarchical organizational scales, each with a large number of various components interacting in an unpredictable (chaotic) manner, is difficult to analyze. Big data or ‘omics’ research aims to enumerate components at successive system levels such as genes (genome), heritable non-DNA elements (epigenome), coding and non-coding RNA (transcriptome), native and post-translationally modified proteins (proteome), metabolites (metabolome), clinical phenotypes (phenome), and environmental factors (exposome). **B.** Studies of gene-

disease associations commonly compare interacting system components between affected populations (or experimentally perturbed systems) and healthy controls or previously established reference via experimental or *in silico* strategies, followed by standard steps of network-based analysis. **C.** Network topology may predict disease genes (candidate genes) based on the relationship of a known disease gene (*filled circle*) to candidate genes (*open circles*) that are directly linked, map within the same module, or occur in the network neighborhood.

Figure 1

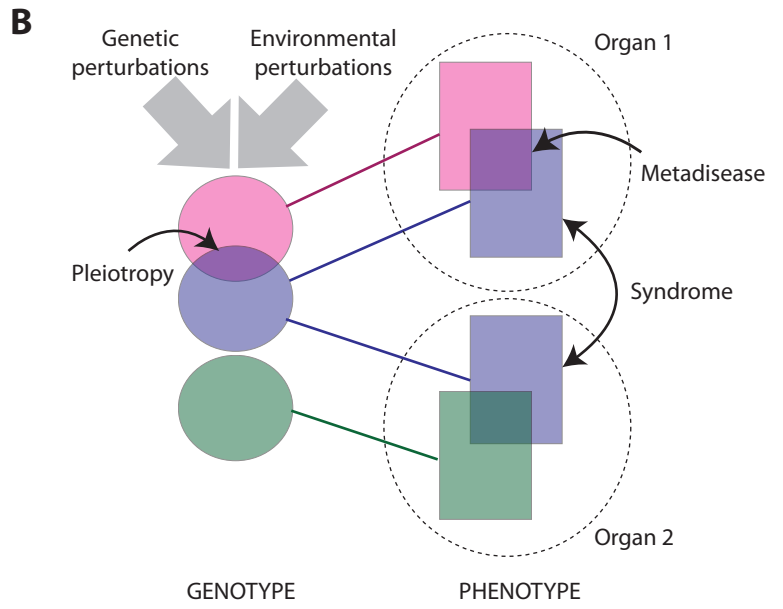
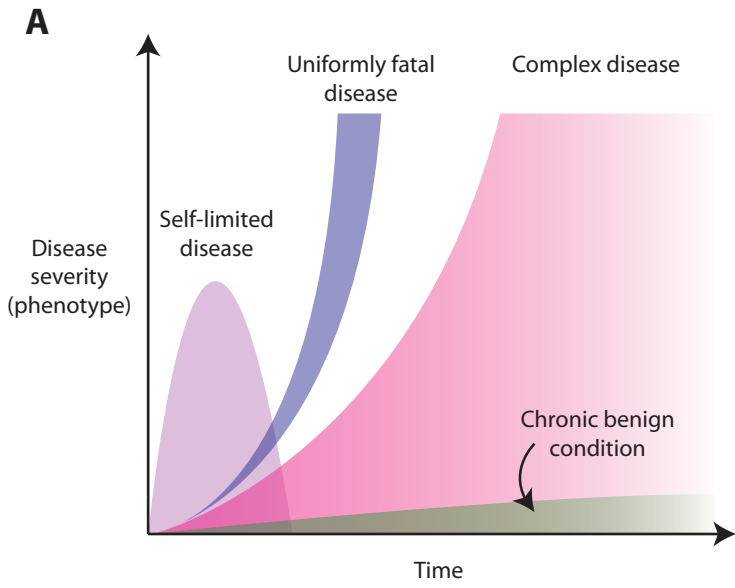


Figure 2

