DISEASOME AND GENETIC NETWORKS

Kwang-Il Goh

Department of Physics, Korea University, Seoul 136-713, Korea

Keywords: Diseasome, human disease network, interactome, cellular network, systems biology.

Contents

Summary

Recent advances in identifying the molecular components responsible for the proper functioning of life and the genetic mutations underlying the malfunctioning thereof—that is, the human disease—have resulted in a greater understanding of physiology and pathology in human biology. In particular, the importance of cellular networks through which the concerted collective actions of various biological components are mediated to fulfill the emergent phenomenon of life is increasingly appreciated. Little is known, however, about how such connectivity and interrelatedness between molecular parts translate into the relationship between their malfunctioning in the genomic and organismic scale. In this regard, a new concept called *the human diseasome* is introduced recently, as the combined set of all known human disorders and the human disease genes the mutations of which confer those disorders.

Here we will survey some important aspects of diseasomefrom the perspective of the human diseasome network. The use of the diseasome in combination with the interactome network and other disease-related factors is also reviewed.

1. Introduction

Illness does not always come alone. Complication is a word so common that is part of our daily vocabulary. Until very recently, however, it had not been a well-posed question that how different mal-conditions of a body relate generically with one another. On the other hand, the knowledge on specific genetic mutations giving rise to specific disease phenotype has been accumulating steadily. Thanks to the decades-long efforts to map human disease loci, first genetically and later physically, followed by the positional cloning and genome-wide association studies, we now have an impressive list of the disorder-disease gene associations. As of 2005, the Online Mendelian Inheritance in Man (OMIM) database enumerated the list of 1,286 human genetic disorders implicated by mutations in 1,777 human disease genes. This progress helped more accurate diagnosis and prognosis, and offered promise for the development of novel therapeutics. The answer to the previous question, however, remained elusive.

Recently a conceptually different approach has been proposed to answer it, by exploring whether human genetic disorders and the corresponding disease genes might be related to one another at a higher level of cellular and organismal organization, that is, at an *omic* level. Early clue to the validity of this approach is provided by examples of genetic disorders that arise from mutations in more than a single gene (referred to as the locus heterogeneity in genetics). For example, Zellweger syndrome is caused by mutations in any of eleven genes, all associated with peroxisome biogenesis. Conversely, there are many examples of different mutations in the same gene giving rise to phenotypes currently classified as different disorders (called the allelic heterogeneity). For example mutations in *TP53* have been linked to eleven clinically distinguishable cancer-related disorders as of 2005. Add to this, it is becoming increasingly understood that the protein products of disease genes in general do not function in isolation, but are part of highly interconnected cellular networks, dubbed as the "interactome," comprised of interlaced array of protein-protein, DNA-protein and RNA-protein physical interactions, as well as metabolic reactions and protein modification relationships. Given the potential complexity encoded by the genetic heterogeneity and the highly interlinked internal organization of the cell dictated by the interactome, it should be possible to improve the single gene-single disorder approach by developing a conceptual framework to link systematically all genetic disorders (the human "disease phenome") with the complete list of disease genes (the "disease genome"), leading to the notion of the "diseasome", the combined set of all known disorder-disease gene associations.

On this new wave of disease genetics, the complex network theory, a branch of mathematics and theoretical physics, plays an instrumental role, providing conceptual insights as well as offering visual and computational methodology. It is thus natural to place the diseasomics onto network biology, a new platform of biological sciences which fully embraces the complexity and connectivity of biological processes, in line with the recent resurrection of the holistic approach to biological systems, advocated in the name of "systems biology." The main goal of this topical review article is to provide the readers an overview and motivate them to engage in this exciting new field. This topical review is organized as follows. In Section 2, the human diseasome is introduced in detail and its network properties such as network topology and modularity are discussed. There the so-called disease module hypothesis will also be introduced and investigated. In Section

3, we will assess the centrality properties of the human disease genes from the network biology perspective. There we will distinguish two classes of human disease genes, the essential genes responsible for diseases-in-utero and the non-essential disease genes that contribute to later onset diseases and compare their functional attributes such as protein interaction degree and coexpression to understand their distinctive properties. In Section 4, we will review various topics related with the diseasome, such as the disease gene prediction, overlaying drug-target network, and augmenting the diseasome with additional layers of disease genetic and non-genetic factors to make it more complete. Finally, Section 5 will conclude the topical review with conclusion and perspective.

2. The Human Diseasome

2.1. Bipartite Graph Representation of the Diseasome

The diseasome can best be represented by a bipartite graph consisting of two disjoint sets of nodes. First set is the disease nodes and the other is the disease gene nodes. A disorder and a gene are then connected by a link if mutations in that gene are implicated in that disorder. A bipartite graph is a graph in which the links always connect the two nodes, each from two disjoint sets of nodes, as is the case for the diseasome. The first version of the diseasome (Figure 1) was created based on the list of human disorders, disease genes, and associations between them obtained from the OMIM database as of December 2005. OMIM initially focused on monogenic disorders and has only recently expanded to include complex traits and the genes mutations of which confer susceptibility for these common disorders, so the current disorder-disease gene associations are biased towards those transmitted in a Mendelian manner. Despite such potential biases and the evident incompleteness of the disease association records, OMIM represents the most up-to-date and reliable repository of known disease genes and the disorders they confer.

The obtained diseasome network is naturally and visibly clustered according to major disorder classes (Figure 1), which are classified manually based on the physiological system affected in the disorder. The most obvious example of this aggregation is the large cancer cluster, which is tightly interconnected due to the many genes associated with multiple types of cancer (*TP53*, *KRAS*, *ERBB2*, *NF1*, *etc*.) and includes several diseases with strong predisposition to cancer, such as Fanconi anemia and ataxia telangiectasia. One can also observe a couple of neurological disorder cluster in the middle-left part of the map, as well as the ophthalmologic disorder cluster on the right, among others. By contrast, metabolic disorders do not form a single distinct cluster, but represent the least connected disorder class relative to the size of the class. They are under-represented in the giant component and over-represented in the small connected components. It is, however, later shown that the metabolic disorders are more connected metabolically via adjacent metabolic pathways, rather than genetically via sharing common disease gene mutations.

Starting from the diseasome bipartite graph, two biologically relevant network projections can be obtained. First, in the *human disease network* (HDN), nodes represent disorders, and two disorders are connected to each other if they share at least one gene in which mutations are associated with both disorders. Second, in the *disease gene network* (DGN), nodes represent disease genes, and two genes are connected if they are associated with the same disorder.
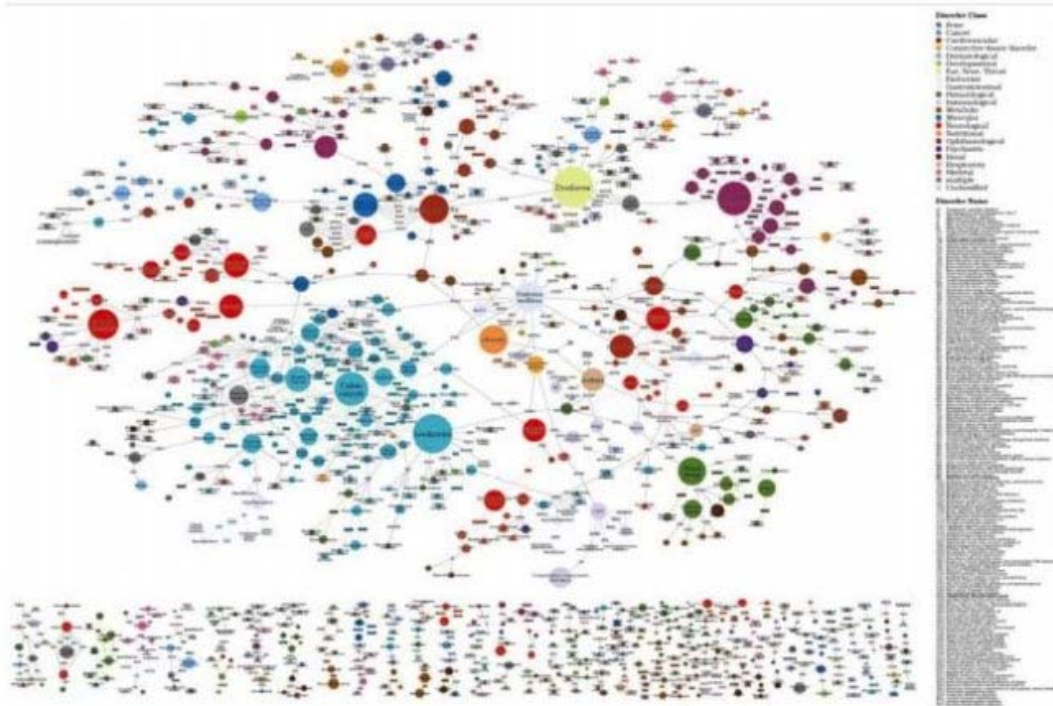
Figure 1. The bipartite graph representation of the human diseasome, comprised of disjoint disease node set (circles) and disease gene node set (boxes) connected by disorder-disease gene association listed in the OMIM database as of the year 2005. Simple single disorder-single gene association pairs are omitted in the figure to emphasize genetically heterogeneous associations. Disorder nodes are colored according to the manually curated disorder class listed in the top-right legend. Disease gene nodes linked to disorders in single disorder class are colored following the class, while those connecting disorders in more than one class are colored light grey. Adapted from Goh, K.-I., *et al*. (2007). Copyright (2007) National Academy of Sciences U.S.A.

-
-
-

TO ACCESS ALL THE **18 PAGES** OF THIS CHAPTER,
Visit: http://www.eolss.net/Eolss-sampleAllChapter.aspx

Bibliography

Barabasi, A.-L. (2007) Network medicine—From obesity to the "diseasome." *New England J. Med.* **357**(4), 404-407: [A perspective paper on the importance of network concept in human disease and medicine].

Ergun A., Lawrence C. A., Kohansiki M. A., Brennan T. A., & Collins J. J. (2007) A network biology approach to prostate cancer. *Mol. Syst. Biol.* **3**, 82: [See annotation of Shlomi et al. (2009)].

Goehler H., et al. (2004) A protein interaction network links GIT1, an enhancer of Huntington aggregation, to Huntington's disease. *Mol. Cell* **15**, 853: [See annotation of Shlomi et al. (2009)].

Goh K.-I., Cusick M.E., Valle D., Childs B., Vidal M., Barabasi, A.-L. (2007) The human disease network. *Proc. Natl. Acad. Sci. USA* **104**(21), 8685-8690: [The original research paper in which the concept of the diseasome has first been introduced and studied].

Hidalgo C.A., Blumm N., Barabasi A.-L., Christakis N.A. (2009) A dynamic network approach for the study of human phenotypes. *PLoS Comput. Biol.* **5**(4), e1000353: [The original research paper in which global comorbid relations between human diseases are studied].

Lee D.-S., Park J., Kay K.A., Christakis N.A., Oltvai Z.N., Barabasi, A.-L. (2008) The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. USA* **105**(29), 9880-9885: [The original research paper in which the comorbid relations between human diseases are studied in terms of metabolic connections].

Lim J., et al. (2006) A protein-protein interaction network for human inherited ataxia and disorders of Purkinje cell degeneration. *Cell* **125**, 801: [See annotation of Shlomi et al. (2009)].

Oti M., Huynen M.A., Brunner H.G. (2008) Phenome connections. *Trends Genet.* **24**(3), 103-106: [A perspective paper on the importance of modular view of genetic diseases].

Park J., Lee D.-S. Christakis N. A., Barabasi A.-L. (2009) The impact of cellular networks on disease comorbidity. *Mol. Syst. Biol.* **5**, 262: [The original research paper that highlights the role of cellular network-level interactions on the disease comorbidity].

Rzhetsky A., Wajngurt D., Park N., Zheng T. (2007) Probing genetic overlap among complex human phenotypes. *Proc. Natl. Acad. Sci. USA* **104**(28), 11694-11699: [The original research paper in which genetic overlap in 161 neurological disorders are quantified with 1.5 million patient records].

Shlomi T., Cabili M., Ruppin E. (2009) Predicting metabolic biomarkers of human inborn errors of metabolism. *Mol. Syst. Biol.* **5**, 263: [This and preceding three original research papers demonstrate the use of systems approach to disease gene and disease biomarker identification].

Yildirim M.A., Goh K.-I., Cusick M.E., Barabasi, A.-L., Vidal M. (2007) Drug-target network. *Nature Biotechnol.* **25**(10), 1119-1126: [The original research paper in which the drugome has first been introduced].

Biographical Sketch

Kwang-Il Goh was born in Daejon, Korea on May 30, 1975. He finished his BA, MS, and PhD degrees in statistical physics at Seoul National University in Seoul, Korea in 1998, 2000, and 2004, respectively.

He did postdoctoral training at the University of Notre Dame and Dana-Farber Cancer Institute, Harvard Medical School in the USA from 2005 to 2007. Since 2007, he is an Assistant Professor at the Physics department of Korea University in Seoul, Korea. His main research interests lies in the application of statistical physics concepts and ideas in the study of complex systems, with a focus on the biological systems such as the disease network, stochastic cellular dynamics, and human dynamics, as well as in general theoretical aspects of complex network theory such as network structural analysis, network modeling, and network dynamics.

Prof. Goh is currently a member of the Korean Physical Society, the American Physical Society, and the Korean Society of Bioinformatics.