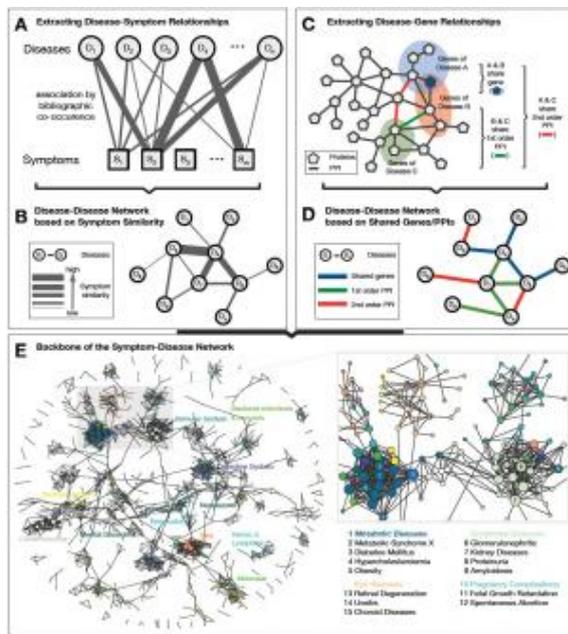


Diseases, symptoms, genes, and proteins linked together in giant network



Construction of the human symptoms-disease network consists of (a) extracting disease-symptom relationships from the PubMed bibliographic literature database, (b) constructing a disease network in which nodes represent diseases and links represent similarities between diseases, and (c) integrating both disease-gene associations and protein-protein interaction databases to obtain shared genes/protein-protein interactions between diseases. (d) The resulting disease network in which links represent shared genes/protein-protein interactions. (e) Highly clustered regions of diseases belong to the same broad disease category. Credit: ©2014 Zhou, et al. Nature Communications

(Medical Xpress) – "The first indication that you're sick is typically one or more symptoms: perhaps a cough, fever, abdominal pain, etc. Symptoms are high-level clinical manifestations of a disease that, at a lower level, is caused by molecular-level components, such as genes and proteins. Understanding the complex ways in which symptoms, diseases, and their underlying molecular mechanisms are related can provide a valuable tool for medical researchers when designing better treatments.

However, this area of research is still very new and not well understood. In a new study published in *Nature Communications*, researchers XueZhong Zhou, et al., have constructed a human symptoms-disease network (HSDN) that reveals the numerous and sometimes surprising connections between symptoms, diseases, genes, and proteins.

"Symptoms are the clinical manifestations that are closer to everyday activities and generally can be perceived by medical lay persons," coauthor Amitabh Sharma at Northeastern University, the Dana-Farber Cancer Institute, and Brigham and Women's Hospital, all in Boston, Massachusetts, told *Phys.org*. "The human symptoms-disease network represents a [data source](#) that has great potential in better research applications and clinical care. The HSDN definitely boosts the translational medicine and precision medicine field, where the data source can be used to identify the clinical phenotypes hidden in the large-scale [electronic medical records](#) to elaborate the clinical features of diseases."

The HSDN is a giant network, consisting of more than 4,000 diseases and 300 symptoms. The data was extracted from millions of PubMed bibliographic records with at least one disease or symptom term in the metadata field.

In the network, nodes represent diseases and links represent symptom similarities between diseases. For example, insulin resistance and metabolic syndrome are two diseases that share many of the same symptoms, such as obesity and hypertension, and therefore have a strongly weighted link between them.

"Diseases, symptoms, genes, and proteins linked together in giant network." Medical Xpress. 18 Jul 2014.

<http://medicalxpress.com/news/2014-07-diseases-symptoms-genes-proteins-linked.html>

Overall, the network is very dense, with 94% of the nodes being connected to more than 50% of all other nodes (i.e., they have at least one shared symptom). The most highly connected disease is hyponatremia, an electrolyte disorder associated with a number of common symptoms that occur in many diseases, such as headache, nausea, and fatigue.

After constructing the network, the researchers then integrated genetic data from three genotype-phenotype databases as well as protein data from five protein-protein interaction databases. In the resulting networks, two diseases are connected if they share an associated gene or protein interaction, respectively. The integrated networks showed that diseases with more similar symptoms are more likely to have both common gene associations as well as shared protein interactions.

These associations among symptoms, diseases, genes, and proteins reveal a large amount of information, some that is widely known and some that is just beginning to be discovered in ground-breaking research.

Confirming what is widely known about disease categories, the network shows highly interconnected communities of diseases, such as those that involve the respiratory tract, digestive system, cardiovascular system, etc. In particular, the network shows that the three main disease risks—namely, infectious diseases, chronic inflammation diseases, and neoplasms (tumors)—are all highly interconnected.

As an example of less well-known associations, the network shows that Parkinson's disease has very similar symptoms, as well as correlated genes and protein interactions, with substance-related diseases such as mercury and manganese poisoning. In just the past few years, research has, in fact, suggested similarities between these diseases.

As another example, the network reveals that Alzheimer's disease shows high symptom similarity with epilepsy. Again, researchers have recently found that an antiepileptic drug (levetiracetam) can reverse deficits in learning and memory in mice with Alzheimer's disease, and might help do the same in humans.

Another major area where the network may be very useful is in comparing genetic and infectious diseases. For example, the network shows that Epstein-Barr virus, which causes mononucleosis, shares symptoms with several other diseases, including T-cell lymphoma, Hodgkin disease, and non-Hodgkin lymphoma, all of which have correlations between genes and protein interactions. The results suggest that symptom similarity scores could provide clues to understanding how viral/bacterial infections may affect genes and protein interactions, increasing susceptibility to [infectious diseases](#).

In the future, the researchers plan to further expand the network by incorporating even more big data, from sources including electronic health records and clinical terminology systems. They predict that advances in the field of automated text mining will play a vital role in accumulating and analyzing this large amount of data.

"We believe that a symptoms-level view of disease phenotypes can shed new light on the different aspects of disease manifestation," Zhou said. "Understanding the human symptoms-disease network in the future could help in revealing the underlying network behind the diseases, and this will eventually lead to clinical cures of the diseases. We are focusing on translating the human [symptoms](#)-disease network knowledge into wisdom that can yield clinically actionable results like predicting and controlling human [disease](#). We believe it would contribute significantly to the new taxonomy of diseases and improved clinical care, which will be more elaborated and patient-oriented in this new information era."

More information: XueZhong Zhou, et al. "Human symptoms-disease network." *Nature Communications*. DOI: [10.1038/ncomms5212](https://doi.org/10.1038/ncomms5212)

Â© 2014 Medical Xpress

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.