

Largest resource of human protein-protein interactions can help interpret genomic data

28 November 2016



Credit: NIH

An international research team has developed the largest database of protein-to-protein interaction networks, a resource that can illuminate how numerous disease-associated genes contribute to disease development and progression. Led by investigators at Massachusetts General Hospital (MGH) and the Broad Institute of MIT and Harvard, the team's report on its development of the network called InWeb_InBioMap (InWeb_IM) is receiving advance online publication in *Nature Methods*.

"Modern genetic technologies allow us to routinely sequence the genomes of people with, for example, cancer or psychiatric diseases, but understanding the cellular systems that are affected by disease-causing genetic variations remains a major challenge," says Kasper Lage, PhD, of the MGH Department of Surgery and the Stanley Center for Psychiatric Research at the Broad Institute, project leader and co-corresponding author of the *Nature Methods*

report. "Having more complete maps of the physical interactions of human proteins will enable us to start exploring cellular processes affected in disease at a higher resolution than is currently possible."

While the importance of mapping large-scale protein-protein interaction networks is widely recognized, the most recent experimental efforts have identified fewer than 30,000 direct interactions, representing well under a quarter of the most conservative estimates of the total number of interactions. Lage's team, in collaboration with researchers in Denmark and the U.K., developed a computational framework to integrate data from more than 43,000 published articles, including data from eight established protein-protein interaction databases. They applied stringent quality control in creating InWeb_IM, which consisted of almost 586,000 interactions when the paper was submitted in February 2015 and now includes more than 625,500 interactions.

Co-lead author Taibo Li from Lage's team explains, "Just like people, proteins like to work in groups to carry out their functions, and they do this by physically interacting in protein networks. If you compare protein-protein interaction networks to human social networks, just as platforms like Facebook can infer people who may know each other or share interests based on patterns of interaction with others in the network, constructing networks of protein interactions can infer gene groups and molecular pathways that can improve our understanding of processes that occur in human cells."

Lage adds, "The rapidly declining cost of genome sequencing has far outpaced our ability to interpret the gene variants we identify in patients with undiagnosed diseases. By exploring [interaction networks](#) at the level of proteins and of the genes that may be causing a disease, clinicians may begin to see patterns of genetic data that would

otherwise be difficult to discern, which we illustrate in the article for cancers and autism. For example, around 30 genes appear to be involved in cardiomyopathies, but many individuals with the condition do not have mutations in any of those genes. By looking at interaction partners at the protein level of the 30 cardiomyopathy genes, we can start to identify new candidate genes based on the 'cardiomyopathy network,' potentially leading to new molecular insights into the disease. It is our hope that InWeb_IM can be a resource that contributes to interpreting clinical exome sequencing data and play a part in enabling clinical action in patients with an unknown cause of disease."

The team is continuing to develop ways of using InWeb_IM to explain large-scale genomic datasets; to improve understanding of complex biological systems in a tissue-specific manner by integrating proteomic, transcriptomic and genomic data; and - in collaboration with several groups at MGH - applying that information to the understanding of cardiovascular diseases, birth defects, cancers, reproductive disorders and psychiatric disease. InWeb_IM will be maintained and updated quarterly and is fully accessible to academic users at <http://www.lagelab.org/resources/> or <http://www.intomics.com/inbiomap> .

More information: Taibo Li et al, A scored human protein–protein interaction network to catalyze genomic interpretation, *Nature Methods* (2016).
[DOI: 10.1038/nmeth.4083](https://doi.org/10.1038/nmeth.4083)

Provided by Massachusetts General Hospital
APA citation: Largest resource of human protein-protein interactions can help interpret genomic data (2016, November 28) retrieved 25 April 2021 from <https://medicalxpress.com/news/2016-11-largest-resource-human-protein-protein-interactions.html>

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