

New study maps protein interactions for a quarter of the human genome

May 17 2017



Credit: CC0 Public Domain

Harvard Medical School researchers have mapped the interaction partners for proteins encoded by more than 5,800 genes, representing over a quarter of the human genome, according to a new study published online in *Nature* on May 17.

The [network](#), dubbed BioPlex 2.0, identifies more than 56,000 unique

[protein](#)-to-protein interactions—87 percent of them previously unknown—the largest such network to date.

BioPlex reveals protein communities associated with fundamental cellular processes and diseases such as hypertension and cancer, and highlights new opportunities for efforts to understand human biology and disease.

The work was done in collaboration with Biogen, which also provided partial funding for the study.

"A gene isn't just a sequence of a piece of DNA. A gene is also the protein it encodes, and we will never understand the genome until we understand the proteome," said co-senior author Wade Harper, the Bert and Natalie Vallee Professor of Molecular Pathology and chair of the Department of Cell Biology at Harvard Medical School. "BioPlex provides a framework with the depth and breadth of data needed to address this challenge."

"This project is an atlas of human protein interactions, spanning almost every aspect of biology," said co-senior author Steven Gygi, professor of [cell biology](#) and director of the Thermo Fisher Center for Multiplexed Proteomics at Harvard Medical School. "It creates a social network for each protein and allows us to see not only how proteins interact, but also possible functional roles for previously unknown proteins."

Bait and prey

Of the roughly 20,000 protein-coding genes in the human genome, scientists have studied only a fraction in detail. To work toward a description of the entire cast of proteins in a cell and the interactions between them—known as the proteome and interactome, respectively—a team led by Harper and Gygi developed BioPlex, a high-

throughput approach for the identification of protein interplay.

BioPlex uses so-called affinity purification, in which a single tagged "bait" protein is expressed in human [cells](#) in the laboratory. The bait protein binds with its interaction partners, or "prey" proteins, which are then fished out from the cell and analyzed using mass spectrometry, a technique that identifies and quantifies proteins based on their unique molecular signatures. In 2015, an initial effort (BioPlex 1.0) used approximately 2,600 different bait proteins, drawn from the Human ORFeome database, to identify nearly 24,000 protein interactions.

In the current study, the team expanded the network to include a total of 5,891 bait proteins, which revealed 56,553 interactions involving 10,961 different proteins. An estimated 87 percent of these interactions have not been previously reported.

Guilt by association

By mapping these interactions, BioPlex 2.0 identifies groups of functionally related proteins, which tend to cluster into tightly interconnected communities. Such "guilt-by-association" analyses suggested possible roles for previously unknown proteins, as these communities often commingle proteins with both known and unknown functions.

The team mapped numerous protein clusters associated with basic cellular processes, such as DNA transcription and energy production, and a variety of human diseases. Colorectal cancer, for example, appears to be linked to protein networks that play a role in abnormal cell growth, while hypertension is linked to protein networks for ion channels, transcription factors and metabolic enzymes.

"With the upgraded network, we can make stronger predictions because

we have a more complete picture of the interactions within a cell," said first author Edward Huttlin, instructor of cell biology at Harvard Medical School. "We can pick out statistical patterns in the data that might suggest disease susceptibility for certain proteins, or others that might suggest function or localization properties. It makes a significant portion of the human proteome accessible for study."

Launching point

The entire BioPlex network and accompanying data are publicly available, supporting both large-scale studies of protein interaction and targeted studies of the function of specific proteins.

Although the network serves as the largest collection of such data gathered to date, the authors caution it remains an incomplete model. The current pipeline expresses bait proteins in only one cell type (human embryonic kidney cells) grown under one set of conditions, for example, and distinct interactions may occur in different cell types or microenvironments.

As the network increases in size and more human proteins are used as baits, scientists can better judge the accuracy of each individual protein interaction by considering its context in the larger network. Isolating the same protein complex several times, each time using a different member as a bait, can provide multiple independent experimental observations to confirm each protein's membership. Moreover, by using prey proteins as bait, many protein interactions can be observed in the opposite direction as well. Both of these scenarios greatly reduce the likelihood that particular interactions were identified due to chance. The team continues to add to BioPlex, with a target goal of around 10,000 bait proteins, which would cover half of the [human genome](#) and would further increase the predictive power of the network.

"We certainly aren't seeing all the interactions, but it's a launching point. We think it's important to continue to build this map, to see how much of it is reproduced in other cell types under different conditions, to see whether the interactions are similar or dynamic," Gygi said. "Because whether you're interested in cancer or neurodegenerative disease, basic development or evolutionary fitness—you can make new hypotheses and learn something from this network."

More information: Architecture of the human interactome defines protein communities and disease networks, *Nature* (2017).
[nature.com/articles/doi:10.1038/nature22366](https://doi.org/10.1038/nature22366)

Provided by Harvard Medical School

Citation: New study maps protein interactions for a quarter of the human genome (2017, May 17) retrieved 19 April 2024 from <https://phys.org/news/2017-05-protein-interactions-quarter-human-genome.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.