

Studies offer guide as protein interaction mapping comes of age

January 6 2009

During the past 20 years, researchers have identified thousands of cell protein interactions, with the ultimate goal of inventorying all that occur within cells of various organisms - a comprehensive catalogue known as the interactome. Such information will be critical to understanding the basic mechanics of cellular life, and how malfunctions in these processes contribute to cancer.

Unfortunately, the data collected by different teams of researchers has been somewhat inconsistent. One group's "map" of protein interactions in yeast cells, for example, may only partially overlap the map produced by another group. Because science depends on investigators' ability to reproduce and build on one another's work, such variability presents a considerable obstacle. The value of interactome maps -- and the potential of further research -- will be at issue as long as the accuracy and thoroughness of the underlying data is uncertain.

To recapture momentum, the field needs to be clear about the strengths and weaknesses of different methods of tracking protein interactions, researchers say, and reach a consensus on questions such as, How reliable is the data produced by different techniques? What portion of the interactome of different organisms has been mapped so far? Why do existing experimental techniques fail to detect certain interactions? What can be done to improve the quality of data collected?

In a series of four papers published in the January issue of the journal *Nature Methods*, investigators in Dana-Farber Cancer Institute's Center

for Cancer Systems Biology (CCSB) start to answer those questions by examining the accuracy and thoroughness of current interactome maps and the techniques by which they are compiled. The studies -- in a special issue of the journal on the interactome -- provide a set of ground rules for future research and demonstrate the power of such research when backed by well-proven experimental techniques. The CCSB's director, Marc Vidal, PhD, is the senior author of the papers.

Framework for study

The first study, lead-authored by the CCSB's Kavitha Venkatesan, PhD, offers a framework for gauging the quality of current maps of the interactome in human cells. The maps draw on three sources of information about protein interactions: high throughput yeast two-hybrid (HT-Y2H) procedures, which use robotic equipment to screen thousands of proteins to see which bind to each other (the binding switches on a "reporter" gene that can be chemically detected); compilations of published studies on small numbers of protein interactions; and studies that predict interactions based on computational techniques. While each approach is useful, it isn't clear whether small-scale experiments provide better data than high volume screenings (as some studies have suggested), whether the interactions detected in experiments actually occur in living cells, and whether existing maps depict a small- or large-sized chunk of the entire interactome.

All experimental techniques generate some false positives -- in which interactions are "detected" that haven't really taken place -- and false negatives - in which interactions that have occurred fail to be found. To weed them out, the new framework examines experimental methods from the standpoint of precision, sensitivity, and completeness. "The framework approach takes as standards interactions reported in multiple studies of high quality, and then verifies those standards against results obtained by other techniques," says Venkatesan.

Using the framework, the Dana-Farber team found that each technique captures only 20-30 percent of all the interactions within cells. That led them to determine that the human interactome contains about 130,000 interactions, a small minority of which have been mapped so far.

The second study offers researchers a tool kit for determining whether a newly discovered interaction is indeed real, and not a false positive reading from a particular type of experiment. The kit is a set of four, high-capacity protein interaction tests that have been weighted in relation to a common set of benchmark data. When scientists identify two proteins as likely interactors, the pair can be tested in the tool kit to obtain a "confidence score" about whether they do, in fact, interact.

"This general approach will allow researchers to systematically and objectively assign confidence scores to all individual protein-protein interactions in cells," says lead author Pascal Braun, PhD. "Such a universally interpretable quality standard is critical for constructing accurate interactome maps."

The third study uses the quality control framework from the first study to compile a new, expanded map of the interactome of the worm *Caenorhabditis elegans* (*C. elegans*), a scientific favorite whose cells have roughly the same number of genes as human cells do. The previous version of the map was assembled from studies involving about 2,000 proteins. For the new map, lead author Nicolas Simonis, PhD, of the CCSB and his associates screened some 10,000 protein pairs, documenting 3,864 high quality interactions. The framework enabled the researchers to estimate that the worm's genome includes about 116,000 interactions, meaning that 96 percent of its interactome remains uncharted.

Trust, but verify

Interactome maps are constructed from a variety of sources -- new experiments and data from earlier studies. As Michael Cusick, PhD, and co-authors show in the fourth *Nature Methods* paper, the information in some of those much-used databases is not as reliable as one would hope.

The team focused on databases built from published studies that involve just a few protein interactions -- an approach sometimes thought to be more accurate than mass-screening techniques. Researchers typically cull information from several such studies to draw conclusions about which proteins interact. In examining such studies closely, however, the researchers found that the results overlap rather infrequently. Of some 12,000 interactions that have been identified in yeast cells, 75 percent were reported in one study only. When Cusick and colleagues reviewed 100 of these shakily supported interactions, they could independently substantiate only 25 percent of them.

The authors suggest that the lower-than-expected quality of this data has less to do with the skill of the scientists who handle the data than with the inherent difficulty of extracting information from long, text-heavy documents. "Often, these studies use different reporting guidelines, which makes it difficult to compile results in a uniform way," Cusick remarks. One solution is the molecular interaction experiment initiative, or MIMIx, which standardizes reporting of protein interactions in published manuscripts.

"Interaction mapping is a complex field," Cusick states. "By teasing apart the process of interaction discovery and verification, we've identified where problems are coming from and offered solutions to minimize inconsistencies in the future. This will be critical as efforts continue to map the entire interactome of various species, including humans."

Source: Dana-Farber Cancer Institute

Citation: Studies offer guide as protein interaction mapping comes of age (2009, January 6)
retrieved 26 April 2024 from

<https://medicalxpress.com/news/2009-01-protein-interaction-age.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.