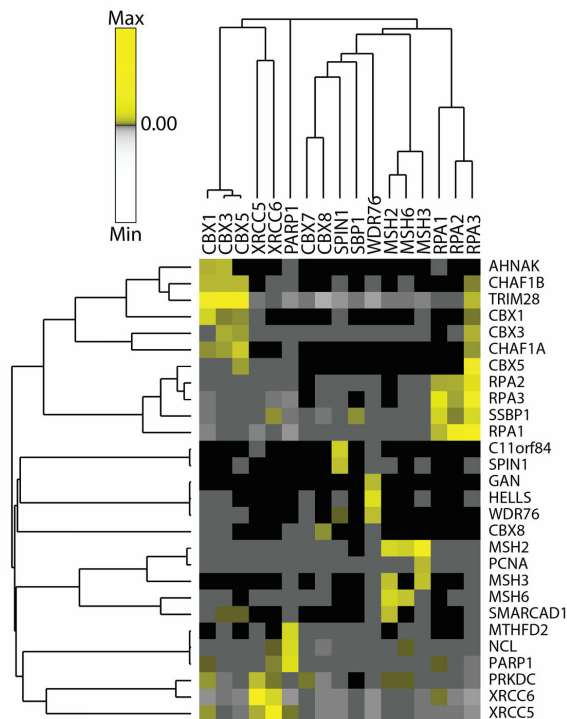


New method of scoring protein interactions mines large data sets from a fresh angle

8 March 2019



A cluster map showing the profiles of bait proteins (rows) that associate with human DNA repair and epigenetic proteins (columns) based on high topological scoring (TopS) values. Yellow (high TopS score) indicates a higher protein interaction preference. Credit: the Washburn Lab, Stowers Institute for Medical Research.

Researchers from the Stowers Institute for Medical Research have created a novel way to define individual protein associations in a quick, efficient, and informative way. These findings, published in the March 8, 2019, issue of *Nature Communications*, show how the topological scoring (TopS) algorithm, created by Stowers researchers, can—by combining data sets—identify proteins that come together.

The approach is similar to looking at the activities

and interactions of all the individuals in a community and then selecting out the most meaningful interactions, some which may be very rare. The researchers are looking for the biological equivalent of two individuals who may be the only two in the entire community that participate in an important interaction.

Not only does this help researchers identify how proteins perform [biological functions](#) or carry out [biological processes](#), the algorithm can be applied to previously generated [biological data](#) and potentially other areas of science to glean new information.

"It's a form of big data analysis that we are applying to proteomics data to identify and understand protein interaction networks," says Michael Washburn, Ph.D., director of the Stowers Proteomics Center. "It's complementary to a lot of techniques already in use so it can be used to ask and answer new questions."

Protein data sets can be challenging to examine for meaningful information because they are so large. "You have thousands of proteins to look at," says Mihaela Sardi, Ph.D., a senior research specialist at Stowers. Understanding how a wide variety of proteins come together to do something, like repair DNA, is a difficult problem. "We wanted to simplify the problem."

That meant instead of taking an overall view of everything, they hunted for less common events. Researchers did this by looking for bait (proteins already known to be involved in processes of interest) and prey (proteins that could interact with bait proteins) to see how they interacted in human DNA repair and yeast chromatin remodeling complexes. Through TopS, data is analyzed in a parallel fashion, meaning that data from several biologically-related baits are considered at the same time. A key attribute of TopS is the ability to evaluate the preference of a prey [protein](#) for a bait

relative to other baits. "Instead of calculating a score by concentrating only information of a single bait, we now aggregate information from the entire data set," explains Sardi.

Washburn and Sardi believe that TopS can be applied to a wide range of data sets beyond proteomics, in both basic research and beyond. Sardi sees potential in using it for healthcare data, where physicians might be able to compare a patient's health to others, like being able to tell if a patient's disease is "really advanced compared to others or not," she says.

The team has also published these findings on Github, a computer code repository, because they want to offer other researchers the opportunity to test the algorithm and see how they can apply it to their own projects.

"We're excited to see how far this can go. It's a potentially high impact tool and we want to see what other creative and innovative people can come up with," says Washburn. "We think this is a really valuable potential tool for a lot of people out there who struggle with the challenge of sorting through very large-scale data."

Other contributors from the Stowers Institute included Joshua M. Gilmore, Ph.D., Brad D. Groppe, Arnob Dutta, Ph.D., and Laurence Florens, Ph.D. Dutta is currently an Assistant Professor at the University of Rhode Island, Groppe is now working at Thermo Fisher Scientific, and Gilmore is a scientist with Boehringer Ingelheim.

This research was funded by the Stowers Institute and a grant from the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM112639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Lay Summary of Findings

Researchers from the Stowers Institute for Medical Research have created a topological scoring (TopS) algorithm, which allows scientists to look at big sets of data in new ways, to help them uncover

more details about how proteins interact and understand more precisely how certain activities on the cellular level happen. The findings appear in the March 8, 2019, issue of *Nature Communications*. Study lead and Director of the Stowers Proteomics Center Michael Washburn, Ph.D., sees potential in applying this algorithm to large [data sets](#) in other areas of scientific research, and beyond.

More information: Mihaela E. Sardi et al, Topological scoring of protein interaction networks, *Nature Communications* (2019). [DOI: 10.1038/s41467-019-09123-y](https://doi.org/10.1038/s41467-019-09123-y)

Provided by Stowers Institute for Medical Research

APA citation: New method of scoring protein interactions mines large data sets from a fresh angle (2019, March 8) retrieved 25 June 2019 from <https://phys.org/news/2019-03-method-scoring-protein-interactions-large.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.