BMC Bioinformatics

BioMed Central

Research article

Open Access

False positive reduction in protein-protein interaction predictions using gene ontology annotations

Mahmoud A Mahdavi and Yen-Han Lin*

Address: Department of Chemical Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, S7N 5A9, Canada

Email: Mahmoud A Mahdavi - maa943@mail.usask.ca; Yen-Han Lin* - yenhan.lin@usask.ca

* Corresponding author

Published: 23 July 2007

BMC Bioinformatics 2007, **8**:262 doi:10.1186/1471-2105-8-262

Received: 21 March 2007
Accepted: 23 July 2007

This article is available from: http://www.biomedcentral.com/1471-2105/8/262

Abstract

Background: Many crucial cellular operations such as metabolism, signalling, and regulations are based on protein-protein interactions. However, the lack of robust protein-protein interaction information is a challenge. One reason for the lack of solid protein-protein interaction information is poor agreement between experimental findings and computational sets that, in turn, comes from huge false positive predictions in computational approaches. Reduction of false positive predictions and enhancing true positive fraction of computationally predicted protein-protein interaction datasets based on highly confident experimental results has not been adequately investigated.

Results: Gene Ontology (GO) annotations were used to reduce false positive protein-protein interactions (PPI) pairs resulting from computational predictions. Using experimentally obtained PPI pairs as a training dataset, eight top-ranking keywords were extracted from GO molecular function annotations. The sensitivity of these keywords is 64.21% in the yeast experimental dataset and 80.83% in the worm experimental dataset. The specificities, a measure of recovery power, of these keywords applied to four predicted PPI datasets for each studied organisms, are 48.32% and 46.49% (by average of four datasets) in yeast and worm, respectively. Based on eight top-ranking keywords and co-localization of interacting proteins a set of two knowledge rules were deduced and applied to remove false positive protein pairs. The '*strength*', a measure of improvement provided by the rules was defined based on the signal-to-noise ratio and implemented to measure the applicability of knowledge rules applying to the predicted PPI datasets. Depending on the employed PPI-predicting methods, the *strength* varies between two and ten-fold of randomly removing protein pairs from the datasets.

Conclusion: Gene Ontology annotations along with the deduced knowledge rules could be implemented to partially remove false predicted PPI pairs. Removal of false positives from predicted datasets increases the true positive fractions of the datasets and improves the robustness of predicted pairs as compared to random protein pairing, and eventually results in better overlap with experimental results.

Background

In recent years high throughput technologies have provided experimental tools to identify protein interactions in large scale, generating tremendous amount of protein interaction data [1]. On the other hand, however, computational approaches for protein interaction inference have