

A comparative analysis of the methods and databases used in the study of protein-protein interactions

Eleni Rapsomaniki

Abstract:

High-throughput data (e.g. from yeast-two-hybrid and co-immunoprecipitation experiments) are believed to be of particularly low quality, with a high false-positive rate. Bioinformatics evidence (such as the measurement of co-expression and annotation similarity) is widely used to identify potential false-positive interactions in low-quality data sets.

We have extended existing approaches to the evaluation of protein interaction data by covering more species, by using a wider range of bioinformatics evidence, and by using sophisticated statistical techniques to handle missing data. One interesting result of our analysis is that, in contrast to popular belief, interactions stored in many databases are often more likely to be false-positives than interactions found uniquely in a single database.