PROTEIN INTERACTIONS EXTRACTED FROM GENOMES AND PAPERS

ALFONSO VALENCIA

Structural and Computational Biology Programme,
Spanish National Cancer Research Centre (CNIO)

To assess the feasibility of extracting protein interactions from text we have recently organized the BIOCREATVE II challenge (http://biocreative.sourceforge.net) in collaboration with the MINT and INTACT databases. The competition was divided in four sub-tasks: a) ranking of publications by their relevance on experimental determination of protein interactions, b) detection of protein interaction partners in text, c) detection of key sentences describing protein interactions, and d) detection of the experimental technique used to determine the interactions. 20 teams participated in the competition that used full text and the information on interactions, sentences and experimental vocabularies provided by the associated databases. The results showed quite promising results and clearly pointed to the main challenges setting the path for future research. Furthermore BioCreative has channel the collaboration of several teams for the creation of the first text mining meta-server (the complete set of BioCreative papers to be published in a special issue of Genome Biology).

Regarding the extraction of information on protein interactions from genomic information along the years my group and others have contributed to the developed of a set of methods based on the concept of concerted evolution between interacting protein families. In a new effort we have recently developed a completely new approach that uses the full power of co-evolution to integrate information from complete collections of protein families.