AUTOMATIC FUNCTIONAL ANNOTATION OF PUBCHEM BIOASSAYS

Rajarshi Guha[1], Julien Gobeill[2,3], Patrick Ruch[3]
[1] School of Informatics, Indiana University, United States of America
[2] Swiss-Prot research group, Swiss Institute of Bioinformatics, Geneva, Switzerland
[3] University of Applied Science & University of Geneva, Switzerland

BACKGROUND and OBJECTIVES

The PubChem Bioassay collection is a set of 1293 assays that cover a wide range of compounds of different sizes (from 20 molecules to 200,000 molecules), and techniques (enzymatic, phenotypic etc.).

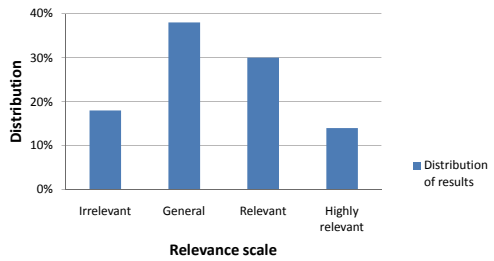
We designed an automated approach to associate each bioassays with a set of functional categories, in order to help search and navigate the assays.

DATA and METHODS

The description section of each assay record was sent to the Gene Ontology Categorizer web services. Designed to help text to GO curation, the categorizer typically outputs a ranked list of GO descriptors.

A sample (size = 30) of the resulting association was manually assessed, using a four-value scale : *highly relevant, relevant, relevant but general, irrelevant*.

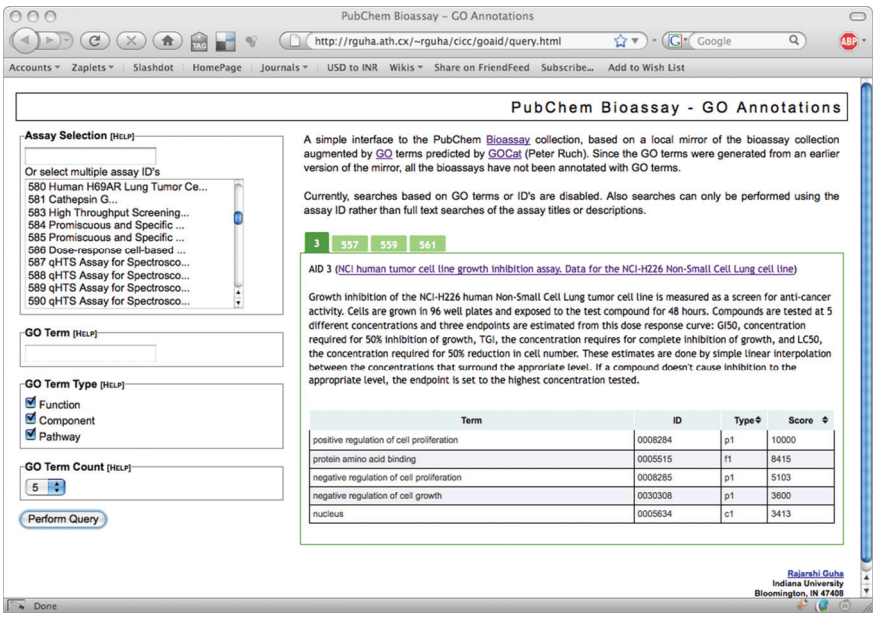
RESULTS and NEXT STEPS



1. Relevant assignments massively overcome irrelevant associations: 82% vs 18%.

2. With the availabilty of formalized annotations, we plan to extend the interface to support more sophisticated similarity based searches as well as more intuitive visualizations of inter-assays relationships.

RESOURCE AVAILABILITIES



The database and query interface are available at http://goassay.rguha.net/index.html
The GO Categorizer is available at http://eagl.unige.ch/GOCat