

BiologicalNetworks: enabling systems-level studies of host-pathogen interactions

Kozhenkov S (1), Sedova M (1), Dubinina Y (1), Gupta A (1), Ray A (3), Ponomarenko J (1,2,*), Baitaluk M (1)

(1) San Diego Supercomputer Center, (2) Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, 92093, USA. (3) Keggs Graduate Institute, Claremont, CA

Teaser

The approach at cross-scale data integration to study host-pathogen interactions is proposed and demonstrated on a study of the *Influenza* infections. The methods and data are available through the application (www.biologicalnetworks.org) and web interface (<http://flu.sdsc.edu>).

Motivation

Understanding of immune response mechanisms of pathogen-infected host requires multi-scale analysis of genome-wide data. Data integration methods have proved useful to the study of biological processes in model organisms, but their systematic application to the study of host immune system response to a pathogen and human disease is still in the initial stage. To study host-pathogen interaction on the systems biology level, an extension to the previously described BiologicalNetworks (1) and IntegromeDB (2) is proposed.

Materials and Methods

Data integration and mapping to the internal database is fully automated and based on Semantic Web technologies and Web Ontology Language. The system represents a general-purpose graph warehouse with its own data definition and query language, augmented with data types for biological entities. The list of integrated databases is at <http://www.biologicalnetworks.net/Database/tut5.php>. Bioinformatics methods were implemented that allow to reconstruct/modify phylogenetic trees, get multiple sequence alignments, identify phylogenetically conserved transcription factor binding sites, and other.

Results

The developed system has been applied to the systems-level analysis of the influenza virus-host interactions, including host molecular pathways that are induced/repressed during the infections, co-expressed genes, and conserved transcription factor binding sites. Previously unknown to be associated with the influenza infection genes were identified and suggested for further investigation as potential drug targets.

Discussion

The developed methods and data integration and querying tools allow simplifying and streamlining the process of integration of diverse experimental data types, including molecular interactions and phylogenetic classifications, genomic sequences and protein structure information, gene expression and virulence data for pathogen-related studies. The data can be integrated from the databases and user's files for public use.

References:

- 1) Baitaluk M, et al. (2006) *Nucleic Acids Res.* **34**: W466.
- 2) Baitaluk M, Ponomarenko J (2010) *Bioinformatics*, doi:10.1093/bioinformatics/btq231.

The US National Institute of Health grants R01GM084881 and R01GM085325.