Databases, Web-servers and Prediction Tools

The management of experimental data is organized in a way as to make them available as soon as possible and over a long period of time for molecular network modeling and, thus, for generations of scientific hypotheses and the next steps of model-based experimental design. Therefore, a data warehouse has been developed and used for service-oriented data and knowledge management: data collection, storage, pre-processing and standardized primary analysis of experimental data including genome, transcriptome and proteome data as well as other biochemical, microbiological, clinical data.

Web-servers and Prediction Tools

VirMiner http://147.8.185.62/VirMiner/ is a software tool which provides comparatively comprehensive phage information in metagenomic data: 1) identifies phage contigs using reliable pre-trained model; 2) gets full functional annotation for these phage contigs; 3) predicts possible phage-host relationships using existing tools; 4) if users upload two groups of metagenomic samples, downstream analysis for comparison among different groups would be done.

COMAN (**CO**mprehensive **M**etatranscriptome **AN**alysis web-server, http://sbb.hku.hk/COMAN/) is an integrated web server dedicated to comprehensive functional analysis of metatranscriptomic data, translating massive amount of reads to data tables and high-standard figures. It is expected to facilitate the researchers with less expertise in bioinformatics in answering microbiota-related biological questions and to increase the accessibility and interpretation of microbiota RNA-Seq data.

MESSI (**M**etabolic **E**ngineering Target **S**election and Best **S**train **I**dentification Tool, http://sbb.hku.hk/MESSI/index.php is a webserver for predicting efficient chassis and regulatory components for yeast bio-based production. The server provides an integrative platform for users to analyze ready-to-use public high-throughput metabolomic data, which are transformed to metabolic pathway activities for identifying the most efficient S. cerevisiae strain for the production of a compound of interest.

NutriChem, available at http://cbs.dtu.dk/services/NutriChem-2.0, is a database generated by text mining of 21 million MEDLINE abstracts for information that links plant-based foods with their small molecule components and human disease phenotypes. NutriChem contains text-mined data for 18478 pairs of 1772 plant-based foods and 7898 phytochemicals, and 6242 pairs of 1066 plant-based foods and 751 diseases. In addition, it includes predicted associations for 548 phytochemicals and 252 diseases.

GRN2SBML automatically encodes gene regulatory networks derived from several inference tools in systems biology markup language. Providing a graphical user interface, the networks can be annotated via the simple object access protocol (SOAP)-based application programming interface of BioMart Central Portal and minimum information required in the annotation of models registry. Additionally, we provide an R-package, which processes the output of supported inference algorithms and automatically passes all required parameters to GRN2SBML. Therefore, GRN2SBML closes a gap in the processing pipeline between the inference of gene regulatory networks and their subsequent analysis, visualization and storage. GRN2SBML is freely available under the GNU Public License version 3 and can be downloaded

from http://www.leibniz-hki.de/en/grn2sbml.html .

Systematically extracting biological meaning from omics data is a major challenge in systems biology. Enrichment analysis is often used to identify characteristic patterns in candidate lists. FungiFun is a user-friendly Web tool for functional enrichment analysis of fungal genes and proteins. The novel tool **FungiFun2** uses a completely revised data management system and thus allows enrichment analysis for 298 currently available fungal strains published in standard databases. FungiFun2 offers a modern Web interface and creates interactive tables, charts and figures, which users can directly manipulate to their needs. FungiFun2, examples and tutorials are publicly available at https://elbe.hki-jena.de/fungifun/ .

The CASSIS suite (https://sbi.hki-jena.de/cassis/): detection of secondary metabolite gene clusters in eukaryotic genomes. **CASSIS (Cluster ASSignment by Islands of Sites)** is a tool to predict secondary metabolite gene clusters around a given anchor/backbone gene. A gene cluster is a small group of genes, which are tightly co-localized, co-regulated, and participate in the same metabolic pathway. **SMIPS (Secondary Metabolites by InterProScan)** is a tool for genome-wide prediction of anchor/backbone genes. Anchor genes encode enzymes, which play a major role in the biosynthesis of secondary metabolites. SMIPS identifies three most common classes of the anchor genes: polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), and dimethylallyltryptophan synthases (DMATS).