

Sample Use Case

John J. Rux

Given a list of genes and/or proteins resulting from high-throughput genomic and proteomics studies, generate hypotheses describing the molecular basis for a biological phenotype by comparing sample set X to sample set Y, where X and Y could be normal vs. disease, treated vs. untreated, time course experiments, etc.

Use case 1: Current approach:

Use existing web-based tools and/or build and query a local database to look for co-occurrence of genes and proteins in existing interaction databases (e.g. BIND), pathways (KEGG, Biocarta) and functional classifications (e.g. GO, GenMAPP). Use stand-alone applications to visualize expression levels in context of pathways and functional classifications. Search literature documenting these functions/interactions. After reviewing literature and consulting with colleagues, search literature again with broader context keywords and concepts to look for gene and protein synonyms. Re-query databases with synonyms to fill in known interactions missed in previous searches.

Use case 2: Near-term improvements gained by using standardized CaBIG tools.

After a researcher X is notified that her Rproteomics, etc. analyses have been completed. She queries the CaBIG pathway resources to look for known relationships within the gene/protein list. She is provided with both graphical and tabular output to illustrate the pathways and functional classifications that include these proteins/genes. The expression and/or abundance values are indicated graphically. The occurrence of potentially overlapping networks is detected and displayed. Links are provided to external data sources such as NCBI's Entrez.

Use case 3: Wish list.

Want to concurrently analyse both gene and protein datasets to visualize networks based on both transcription and translation events. Need tools to correlate biological processes across multiple data sets. Integrate/merge existing pathway and molecular interaction databases so that given a set of genes of interest the presence of overlapping networks can be identified. Need a much richer amount of biological network and pathway data to be available for searching. The relationships between the networks should be documented in terms of molecular interactions including both the translational and transcriptional levels of regulation. Literature mining tools should allow complex queries to provide references to related studies that can be sorted by genes, proteins, pathway, GO terms, MeSH terms, etc.