Poster L-27

graPT : A genome resourcing analyzer for Predictive Toxicology

Authors:
Jung Hoon Woo (*Seoul National University Biomedical Informatics, Interdisciplinary Program in Bioinformatics, Seoul*)
Ju Han Kim (*Seoul National University Biomedical Informatics, College of Medicine, Seoul National University*)

Short Abstract: Predictive toxicology have been main issues of emerging technology, toxicogenomics. We have developed predictive methods and tools, graPT which is based on the gene-toxicant, toxicant-disease relationships to provide association between toxic endpoints and unknown chemicals.

Long Abstract:

Introduction and Motivation

As genomics technologies have been gradually integrated into conventional toxicology, the new era so-called ‘toxicogenomics’ is emerged in the field of study. Predictive toxicology (-predict toxic endpoints with unknown chemical exposure) have been main issues of conventional study, accordingly it also have become challenge of toxicogenomics. Comparing microarray (-explain thousands of transcripts' changes with specific condition) gene expression patterns between model organisms stimulated with toxicant or environmental stress and control have been widely used strategy for prediction of toxicity of new and existing chemicals. However it is not feasible to get directly meaningful information from the high-dimensional gene expression data. Microarray Experiments, as well as other genomic analysis, often result in large set of genes containing up to several hundred of genes. To identify function of genes or to get significant information, such as prediction of toxic endpoints, is not easily feasible for that reason. We have developed predictive methods and tools, graPT which is based on the gene-toxicant, toxicant-disease relationships to provide association between toxic endpoints and unknown chemicals.

Material

We localized CTD (http://ctd.mdibl.org/) data and CHE (http://database. healthandenvironment.org/) Toxic data for using ‘toxicant-gene’ and ‘toxicant-disease’ relationship. Furthermore we localized Entrez Gene (http://www.ncbi.nlm.nih.gov/entrez/), RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/) and MeSH (http://www.nlm.nih.gov/mesh/) data for annotating each gene, toxicant and disease data type. We only use ‘toxicant-gene’ relationship for significance testing. About information of three data types is included for annotate the result of microarray experiment.

Method

Input data generation – Basically, a goal of DNA microarray experiment in toxicology is finding differentially expressed genes (DEGs) between cases and controls. The input of operation would be unique ID of DEGs, and it could be determined by any existing statistical ranking functions.

Toxicant term statistics - Frequencies of toxicant terms within the dataset are calculated and compared with reference frequencies (with genomic frequencies or with or with the frequencies of these toxicants terms in the complete list of genes on array used of experiment). Perform hypergeometric tests between whole toxicants and DEGs may produce list of statistically significant toxicant terms associated with target one.
Mapping disease (or symptom) to target toxicant – Use toxicant-disease relationship information for this step. If significant toxicant terms generated from previous testing have relationship with any symptom (or disease), then operator makes associations between the symptoms and experimental toxicant.

Result

The microarray analysis generates set of genes so called DEGs that are stimulated (could be up-regulated or down regulated) by experimental chemical. The set of genes be an input of graPT. Performing statistical test based on Hypergeometric distribution, this system permits the automatic ranking of all toxicant terms, as well as the evaluation of the significance of their occurrence within the dataset. High ranked toxicants are closely related with experimental one based on the genomic information, toxic endpoints could be predicted with assigned symptoms (or diseases) on highly ranked toxicants in consequence.
Additionally, there are three (-gene, toxicant and disease) distinct data types in this system. Those data types are cross-referenced. Researcher can query about affected genes by specific toxicant or disease and vice versa.

Conclusion

As mentioned above, ‘predictive toxicology’ is the main purpose of toxicogenomics experiments. In this study we have developed the system based on ‘gene-toxicant’ and ‘toxicant-disease’ relationships perform statistical testing to give predictive toxic endpoints of unknown chemicals, directly.