The Intelligent Informatics Systems for Microarray and SNP

Yang Seok Kim

ISTECH Inc.

GenPlex is a comprehensively organized program addressing the key issues of DNA chip data analysis, modularized in four independent components: preprocessing, class comparison, class discovery, and class prediction. The four modules have their own unique analysis methods and are connected in a whole analysis pipeline. The preprocessing module translates various chip platforms raw data into the input file(s) for the three subsequent analysis modules. It includes statistical methodologies like normalizations with useful plottings to enhance data quality by reducing experimental errors. The connectivity between the preprocessing module and the other three modules is the basic gene expression matrix that is a baseline file format for further analysis. The three main analysis components contain diverse algorithms of practical use with insightful visualizations combined. The relevant issues are DEG finding, clustering, and sample classification, respectively. Given gene lists or clusters obtainedfrom each analysis module, their biological meanings are further investigated by retrieving such biological information as Gene Ontology annotations, pathway involvement. Moreover statistical data mining using GO is adopted to find meaningful biological themes on given gene lists. GenPlex will help biologists to capture and extract the meaningful patterns from the high dimensional DNA chip data in a user-friendly environment.

SNPWizard is an enterprise solution for SNP research. It covers whole data management process from experimental design to publishing analyzed results. The solution can be divided into 6 main modules according to their functions: 1) SNP data upload module bearing automatic extraction of related information, 2) intelligent filtering and searching module, 3) GUI module which enable users to comparing their SNP data with published data in graphic environment, 4) analysis module which performs HWE, LD, Haplotype estimation, and QTL analysis, 5) publication module which is designed for easy publication of complex SNP data without any further effort, and 6) authentication module for data security and multi-user usage. Oracle 10g was used for database management system and operating system. The whole system can be accessed by Web interface and DEMO version is available through http://www.istech.info/istech/board/login_form.jsp.