A data mining platform for systems biology and biomarker discovery

G I G A

Olivier Stern[1], Raphaël Marée[2], Christophe Van Huffel[3], Jean-François Laes[4], Carine Michiels[5], Lionel Flamant[5], Véronique Mainfroid[6], Daisy Flamez[7], Louis Wehenkel[1], Pierre Geurts[1]

(1) Department of Electrical Engineering and Computer Science, GIGA-Research, ULg, (2) GIGA Management, Bioinformatics Platform, ULg, (3) KeyMarker project - BioWin, Namur, (4) DNAvision SA, Gosselies, (5) Unité de Recherche en Biologie Cellulaire, FUNDP, Namur, (6) Eppendorf Array Technologies SA, Namur, (7) Laboratory of Experimental Medicine, ULB

Université de Liège

Motivation

The general goal of this research is to develop a bioinformatic strategy to discover new candidate biomarkers for some specific disease, by integrating biological knowledge available in public databases and experimental data related to this disease obtained from high-throughput instrumentations, such as transcriptomic (microarray), proteomic (mass spectrometry), and genomic (SNP) data. The resulting strategy will be implemented in a generic and flexible software platform that will allow biologists to easily instantiate this strategy on their own datasets. This research is part of the Keymarker project (Biowin), whose general goal is to identify biomarkers for molecular imagery.

EXPERIMENTAL DATA

TREATMENT
(NORMALISATION, BACKGROUND CORRECTION…)
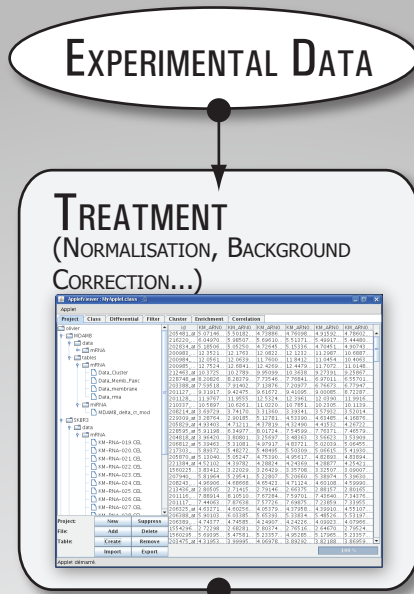
A key aspect of the software is that it integrates all the steps of the analysis in a common platform, making easier the interactions between them. The platform is accessible as a web service as well as a standalone application. Future modules will be added to allow the integration of different experimental data sources (transcriptomic, proteomic, and genomic).
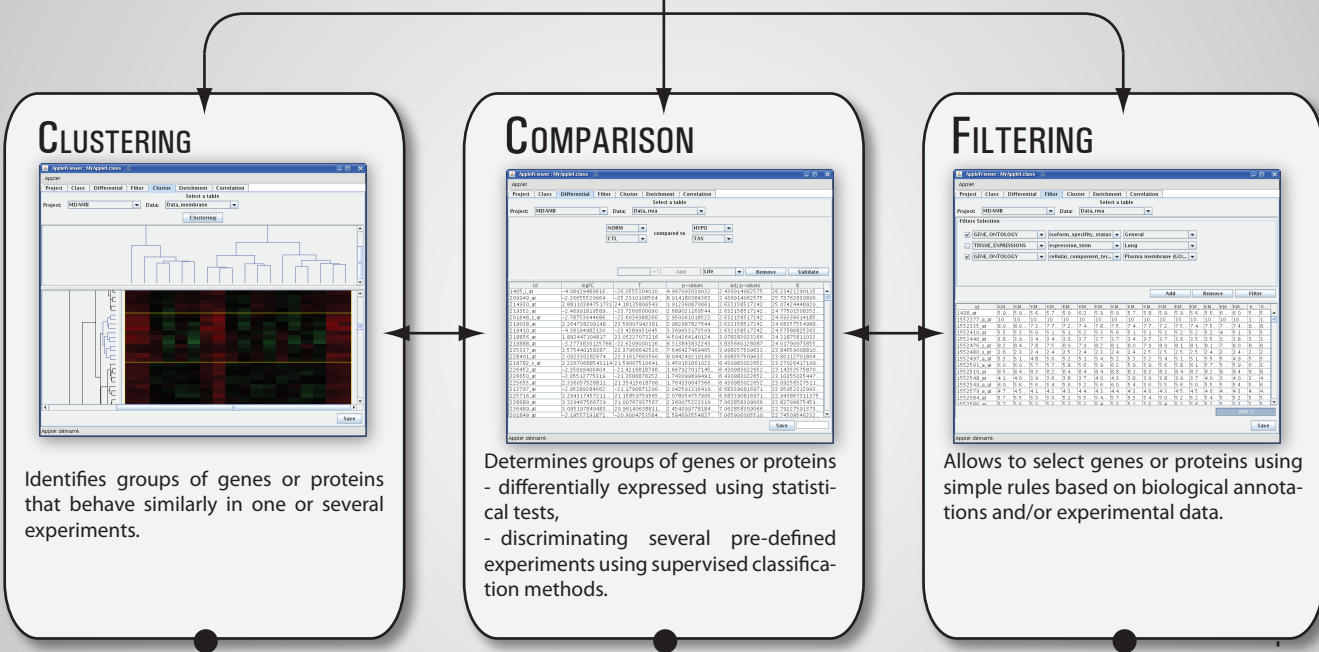
CLUSTERING

Identifies groups of genes or proteins that behave similarly in one or several experiments.

COMPARISON

Determines groups of genes or proteins
- differentially expressed using statistical tests,
- discriminating several pre-defined experiments using supervised classification methods.

FILTERING

Allows to select genes or proteins using simple rules based on biological annotations and/or experimental data.
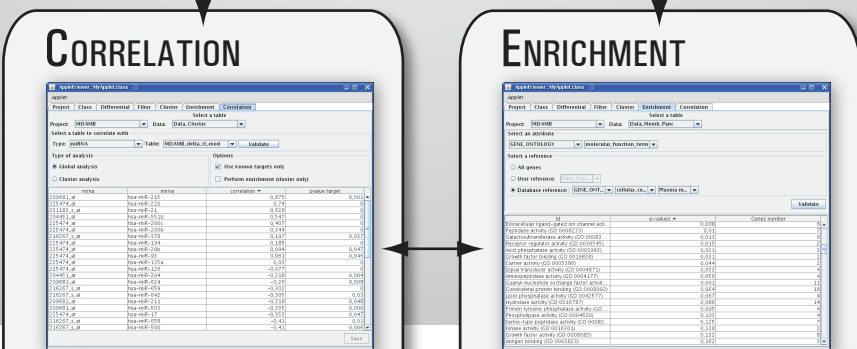
CORRELATION

Analyses jointly mRNA data with other biological data such as transcription factors or micro-RNA. The method determines, from expression patterns, whether there is a link between two particular elements.

ENRICHMENT

Characterizes the groups of relevant biomarkers highlighted by the other modules (by exploiting existing annotations of the corresponding genes or proteins in biological databases).

BIOLOGICAL DATABASES

Keymarker
powered by
BioWin
THE HEALTH CLUSTER OF WALLONIA