IDP provides efficient help to address the flood of information

Although all body cells have the same genetic code, they nevertheless behave differently. This is due to the fact that different cells have different gene expression profiles. DNA microarrays can be used to analyse the expression of thousands of genes simultaneously. The Institute of Data Analysis and Process Design (IDP) at the Zurich University of Applied Sciences (ZHAW) has developed intelligent software to analyse gene expression products automatically. This software considerably reduces the workload. The IDP has recently become a member of BioLAGO e.V. and specialises in statistical data analysis, modelling and optimisation of processes and systems.

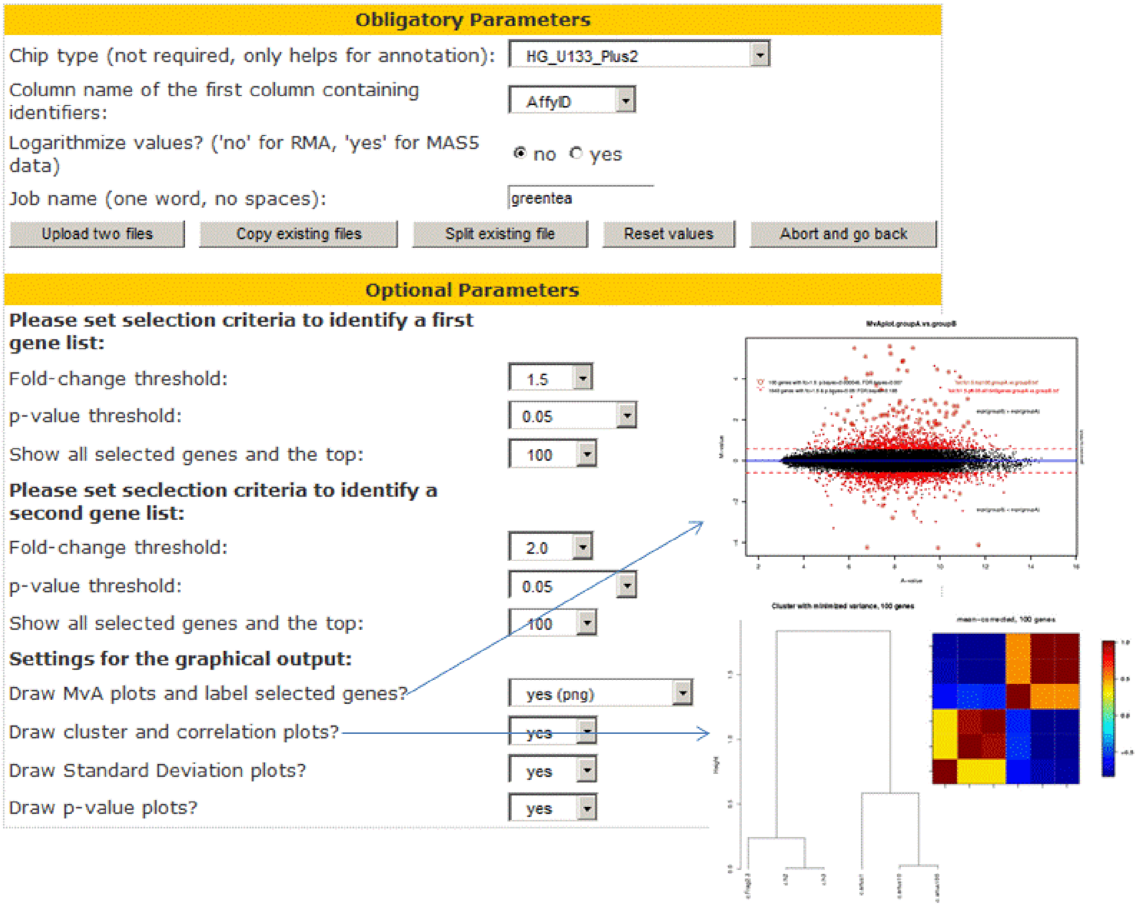
Professors, lecturers and scientific staff at the IDP
© Institute for Data Analysis and Process Design

During the twelve years since its establishment in 2001, the IDP has carried out as many as 400 projects in the fields of laboratory analytics, health and medicine as well as environment and ecology, to name some of the IDP's main fields of application and activities. The institute analyses high-content-screening data, mass spectral data, data obtained with different assay platforms as well as gene expression data (DNA microarrays and RNA-Seq). One project focuses on the development of the RACE (Remote Analysis Computation for gene Expression data) software suite, which is a collection of bioinformatics tools designed for the semi-automated analysis of DNA microarrays. It can, for example, identify genes that determine the number of hairs on the sex combs of male Drosophila legs. “We always start with so-called CEL files, which contain raw data produced with Affymetrix GeneChips, to compute gene expression levels,” said Dr. Beate Sick, professor of statistical data analysis and project leader at ZHAW. RACE provides researchers with quantitative and scalable results as it tests the entire DNA of one or several cells for its expression strength under different conditions. The raw data are classified into different categories: “good”, “suspicious” and “bad”. The raw data quality check assures researchers that subsequent results will also be reliable. DNA microarrays can be used to simultaneously determine transcript levels for known genes in the genome of the organism under investigation. For genes encoding proteins, the expression level can be directly assessed from the quantity of mRNA present in a cell. The quantity of mRNA corresponds to gene activity. DNA microarrays can therefore be used to determine the gene expression profiles in healthy and tumour tissue.

Identification of faulty chips

The software reliably identifies 90% of all faulty chips. “100 percent reliability will never be guaranteed because the data are extremely high-dimensional and heterogeneous,” said Dr. Oliver Dürr, lecturer in statistical data analysis and project leader at the ZHAW. The more data that can be used for the identification of faulty chips, the greater the reliability of the software. “The aim of this software is to cover the typical 80% of cases that arise in the analysis of Affymetrix microarray data. Affymetrix microarrays are used for genome-wide gene expression analyses in medical research into diverse diseases, to name but one area of application. The technology is extremely popular in tumour research aimed at identifying genes that can be targeted for disease diagnosis and therapy. The Gene Ontology (GO)-term analysis tool supports the biological interpretation and annotation of results; it encompasses the three principle GO categories “cellular component”, “biological process” and “molecular function”.

Potential application: gene analysis in intestinal diseases



Obligatory Parameters

Chip type (not required, only helps for annotation): HG_U133_Plus2

Column name of the first column containing identifiers: AffyID

Logarithmize values? ('no' for RMA, 'yes' for MAS5 data) ● no ○ yes

Job name (one word, no spaces): greentea

Optional Parameters

Please set selection criteria to identify a first gene list:

Fold-change threshold: 1.5

p-value threshold: 0.05

Show all selected genes and the top: 100

Please set seclection criteria to identify a second gene list:

Fold-change threshold: 2.0

p-value threshold: 0.05

Show all selected genes and the top: 100

Settings for the graphical output:

Draw MvA plots and label selected genes? yes (png)

Draw cluster and correlation plots? yes

Draw Standard Deviation plots? yes

Draw p-value plots? yes

MvAplot groupA vs groupB
A-value
Cluster with minimized variance, 100 genes
mean-corrected, 100 genes

Screen shot of RACE
© Institute for Data Analysis and Process Design

RACE has already been used in various research and development projects, including the analysis of gene expression associated with chronic bowel inflammation in mice. “This project, which was carried out by the Dutch Organisation for Applied Scientific Research, focussed on the assessment of disease development in response to three weekly administrations of trinitrobenzoic acid (TNBA),” said Dr. Oliver Dürr. The intestines of the mice became inflamed and the structure of the intestine changed. The pathological features of the tissue corresponded with the results obtained from mRNA analysis, which showed that chemokines, which play a key role in acute inflammation, were activated in the TNBA-treated mice.

RACE helps save a lot of time

The advantage of the software is that it is restricted to the most important applications. It was also possible to reduce the time required for the manual analysis of such data from around 12 to under 2 hours. “Due to the software’s transparency and reproducibility, it also enables less experienced users to carry out complex analyses,” Prof. Beate Sick says. The calculations are carried out with the software R, a free software environment for statistical computing and graphics. R provides a wide variety of statistics and graphical techniques and is highly extensible. Oliver Dürr also points out that RACE can also be adapted to new techniques like RNA-Seq which use high-throughput sequencing technologies to get information on the differential expression of genes.

Broad acceptance in hospitals and universities

RACE is used by numerous academic microarray centres and hospitals; around 5 to 50 analyses are carried out every day. The IDP is part of a Swiss university of applied science, which is why the institute is specifically focused on applied research, which in the ideal case scenario is carried out in cooperation with or on behalf of the industry. “We hope that our BioLAGO membership and BIOPRO will help us enhance our presence in life sciences projects,” said Dürr going on to add that he and his team are generally interested in projects where data analysis plays a key role.

Further information:

Dr. Oliver Dürr
Institute of Data Analysis and Process Design (IDP)
Rosenstrasse 3
CH-8401 Winterthur
Tel: +41 (0)58 934 67 47
Fax: +41 (0)58 935 67 47
E-mail: oliver.duerr(at)zhaw.ch

Article

25-Mar-2013
Nyamsuren Davaasambuu
BioLAGO

Zurich University
of Applied Sciences

zh
aw

School of
Engineering

IDP Institute of
Data Analysis and
Process Design