Microarray Data Analysis

Joaquin Dopazo, Javier Santoyo and Ana Conesa

(Department of Bioinformatics and Functional Genomics Node (INB) Centro de Investigacion Principe Felipe (CIPF) 46013, Valencia, Spain)

3 Day course: 16.09 – 18.09. 2009

Held at The National e-Science Centre,

246b Kelvin Building, University of Glasgow, G12 8QQ

Register here: www.sbforum.org > Home > Activities > Events
(http://www.sbforum.org/events.php?e_id=61)

Back for the second time, this three day training event on microarray analysis introducing new concepts of transcriptomics with next generation sequencing.

Introduction

DNA microarrays constitute, no doubt, a paradigm among post-genomic technologies, which are characterised by producing large amounts of data, whose analysis and interpretation is not trivial. Microarray technologies allow the querying of living systems in a completely new way, but at the same time present new challenges in the way hypotheses must be tested and our results ought to be analysed.

Since the first papers published in the late nineties, the number of questions that have been addressed through this technique have both increased and diversified. Initial interest was focused on genes coexpressing across sets of experimental conditions, implying essentially the use of clustering techniques. More recently, however, the interest has switched to finding genes differentially expressed among distinct classes of experiments, or correlated to diverse parameters. There is also much interest in robust methods for building predictors of clinical outcomes. Also, CGH-arrays (Albertson and Pinkel, 2003) are recently becoming an alternative for studying the relationship between chromosomal alterations affecting the copy number (which are behind many diseases) and gene expression. In addition, there is also a clear demand for methods that allow automatic transfer of biological information to the results of microarray experiments.

This course covers the state-of-the-art in the above mentioned topics, which are of major relevance in today's gene expression data analysis. Through sessions of theory and practical examples, students will acquire the experience necessary to address scientific questions using gene expression array datasets and solve them. Special attention will be devoted to important (although not always taken into account) aspects in microarray data analysis, such as multiple testing or functional annotation. The course is designed to be a mixture of theoretical and practical sessions. The latter will require some familiarity with the use of web-based tools and knowledge of basic notions of statistics. Practical sessions will be carried out using the GEPAS (Herrero et al., 2003, 2007; Vaquerizas et al., 2005) environment, an integrated web tool for microarray data analysis, and the Babelomics suite (Al-Shahrour et al., 2005) for functional annotation of genome-scale experiments.

Programme

Day 1

10:00-13:00	**Presentation and Introduction**
	Structure of the course. Why microarrays? Pre- and post-genomics hypothesis testing: a note of caution. Design of experiments. Data preprocessing and normalization. Unsupervised analysis (clustering). Supervised analysis (gene selection, predictors). Functional profiling.
13:00- 14:00	*Lunch*
14:00-18:00	**Normalization**
	Getting rid of unwanted variability from sources other than the experimental conditions assayed. Methods for Affymetrix, two-color and one-color microarrays. Theory and practical exercises using GEPAS.

Day 2

10:00-13:00	**Gene Selection**
	Methods for selecting genes differentially expressed among two or more experimental conditions, correlated to a continuous variable or correlated to survival time. How to deal with the multiple-testing problem. Theory and practical exercises using GEPAS.
13:00-14:00	*Lunch*
14:00-17:00	**Predictors**
	Gene selection in the context of class prediction. How to deal with the selection bias problem. Different methods for class prediction. Estimating the error of classification. Interpretation of confusion matrices.
17:00-18:00	**Clustering**
	Different clustering methods: hierarchical clustering, SOM, SOTA and k-means. Pros and cons. Measures of cluster quality. Cluster visualization.

Day 3

10:00-12:00	**Gene Annotation Data Bases Overview**
	Most popular gene and protein annotation repositories be revised. These are some of the data that the Babelomics suit imports for the functional profiling of genomic experiments.
12:00-13:00	**Functional profiling of genomic experiments**
	Understanding the biological roles played by the genes in the experiments. Using different types of information for the functional profiling of microarray experiments: Gene Ontology, InterPro motifs, transcription factor binding sites, gene expression in other experiments, text-mining, etc. New trends in the analysis of microarray data: testing pathway-based or function-based hypothesis. Building up Functional Annotation using Blas2GO.
13:00-14:00	*Lunch*
14:00-16:30	**Functional profiling of genomic experiments ctd.**
16:30-17:00	Concluding remarks and final questions

References

Albertson, D.G. And Pinkel, D. Genomic microarrays in human genetic disease and cancer. Hum Mol Genet, 2003 12 Spec No 2, R145-52

Al-Shahrour F, Minguez P, Vaquerizas JM, Conde L, Dopazo J: BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. Nucleic Acids Res 2005, 33:W460-464

Herrero J, Al-Shahrour F, Diaz-Uriarte R, Mateos A, Vaquerizas JM, Santoyo J, Dopazo J: GEPAS: A web-based resource for microarray gene expression data analysis. Nucleic Acids Res 2003, 31:3461-3467.

Herrero J, Vaquerizas JM, Al-Shahrour F, Conde L, Mateos A, Diaz-Uriarte JS, Dopazo J: New challenges in gene expression data analysis and the extended GEPAS. Nucleic Acids Res 2007, 32:W485-491

Vaquerizas JM, Conde L, Yankilevich P, Cabezon A, Minguez P, Diaz-Uriarte R, Al-Shahrour F, Herrero J, Dopazo J: GEPAS, an experiment-oriented pipeline for the analysis of microarray gene expression data. Nucleic Acids Res 2005, 33:W616-620

http://bioinfo.cipf.es http://www.gepas.org http://www.babelomics.org

Register here: www.sbforum.org > Home > Activities > Events
(http://www.sbforum.org/events.php?e_id=61)