Organization

Joaquín Dopazo Fátima Al-Shahrour **David Montaner** (CIPF, Valencia, Spain)

Contact

http://bioinfo.cipf.es/docus/courses/coursesCIPF/MicroarraysIII.html

Joaquín Dopazo

E-mail: jdopazo@cipf.es

Centro de Investigación Príncipe Felipe

Avda/ Autopista del Saler, 16 46013 - Valencia, Spain

Registration and deadlines

Registration fee is 350 €, and entitles you to a free issue of the book Data Analysis and Visualization in Genomics and Proteomics. 2005 F. Azuaje and J. Dopazo (Editors) Wiley 2nd *edition.* The course is restricted to a **maximum** of 25 participants. Pre-registration through the web site:

http://bioinfo.cipf.es/docus/courses/coursesCIPF/registration.html

Deadlines:

Pre-registration: February 28th Admittance notified by March 1st Registration and payment: March 8th

Kev topics

- Normalization
- Data transformation
- Clustering
- Gene selection
- Predictors
- Array-CGH
- Functional profiling
- Bioconductor

Software used

CIPF,

Valencia,

Spain

GEPAS

http://www.gepas.org

BABELOMICS

http://www.babelomics.org



http://www.bioconductor.org



Organizers: Joaquín Dopazo Fátima Al-Shahrour **David Montaner**

Lecturers

Fátima Al-Shahrour, Lucía Conde, Joaquín Dopazo, Jaime Huerta, David Montaner, Pablo Mínguez, Ignacio Medina, Susana Vegas, Kasper Henses,

> Centro de Investigación **Príncipe Felipe (CIPF)** Valencia, Spain

III International Course on microarray data analysis

12th-16th March, 2007

Objectives

DNA microarrays constitute, no doubt, a paradigm among post-genomic technologies, which are characterised for producing large amounts of data, whose analysis and interpretation is not trivial. Microarray technologies allows querying living systems in a completely new way, but at the same time pose new challenges in the way hypotheses must be tested and our results ought to be analysed. This course covers the state-of-the-art in topics such as normalization, data transformation, clustering, gene selection, predictors and array-CGH, along with its functional interpretation using information biological (GO, pathways. interactome, etc.) These topics are of major relevance in today's gene expression data analysis. Through sessions of theory and practical examples, the students will acquire the necessary skills to address scientific questions using gene expression microarray data. Special attention will be devoted to important (although not always took into account) aspects in microarray data analysis, such as multiple testing or functional profiling. In addition, some theoretical lessons on basic statistics will be included as part of the programme. Finally, for those who want to go in more depth into analysis possibilities, the last day a short course on Bioconductor (Gentleman et al., 2004) will be taught.

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 basics notions of statistics. Practical sessions will be carried out using the GEPAS (Herrero et al., 2003, 2004, Vaquerizas et al., 2005; Montaner et al., 2006) environment, an integrated web tool for microarray data analysis,

and the Babelomics suite (Al-Shahrour et al., 2005b, 2006) for functional profiling of genomescale experiments. The one-day short course on Bioconductor will be given using the Bioconductor programming environment

Programme

Day 1

Introduction

http://bioinfo.cipf.es/docus/courses/coursesCIPF/MicroarraysIII.html

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.

Normalization (theory and practical exercises)

Getting rid of unwanted variability from sources other than the experimental conditions assayed. Methods for Affymetrix, two-colour and one-colour microarrays

Gene selection (theory and practical exercises)

Methods for selecting genes differentially expressed among two or more experimental conditions, correlated to a continuous variable or correlated to survival. How to deal with the multiple-testing problem.

Day 2

Basic statistical methods

Some theory on basic statistical methods.

Predictors (theory and practical exercises)

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.

Clustering (theory and practical exercises)

Different clustering methods: hierarchical clustering, SOM, SOTA and k-means. Pros and cons. Measures of cluster quality. Cluster visualisation.

Basic statistical methods

International

course on microarray data analysis,

CIPF,

Valencia,

Spain

Some theory on basic statistical methods.

Functional profiling of 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 functionbased hypothesis.

Functional profiling. The Babelomics suite

Different methods for functional profiling of experiments from the Babelomics suite: FatiGO/FatiGO+, Marmite (using text-mining) or TMT (pre-tabulated gene expression results). Methods for finding blocs of functionally-related genes differentially expressed (GSEA, FatiScan).

Dav 4

Basic statistical methods

Some theory on basic statistical methods.

Array-CGH

Estimation of copy number in chromosomal aberrations. Joint study of copy number, gene expression and functional profiling.

Introduction to the programmable GEPAS interface

Using the visual programming interface of GEPAS to build up pipelines of analysis.

Exercises

Do a complete practical exercise (normalization, analysis and functional interpretation) using the tools you learned during the course.

Concluding remarks and final questions

Day 5

One day course on Bioconductor Basic usage of bioconductor tools Day 3