

Systems biology course: "Statistical analysis of ~omics data"

8-11 December 2008,
Wageningen, The Netherlands

Organized by the Graduates School VLAG and EPS and Wageningen UR Plant Breeding



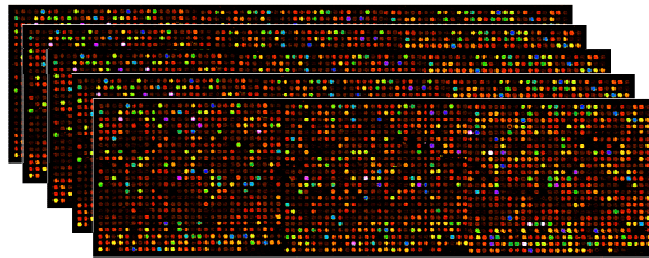
Introduction

Background

Nowadays increasing numbers of complete genomic sequences are available and high-throughput methods have been developed to study gene expression (transcriptomics), proteins (proteomics) and metabolite levels (metabolomics). Because high-throughput methods generate large datasets, special analysis and visualisation techniques are required to extract relevant information for elucidating the function of genes, proteins and metabolites, the interactions between these molecules and the underlying regulatory mechanisms. Statistical analysis of these data is non-trivial since in many cases the number of genes/metabolites outweighs the number of samples by hundred or thousand folds.

Target group

Basic background knowledge in Cell biology, Molecular biology, Biochemistry, Statistics and Excel. Without this knowledge, it is advisable to carry out review and remedial work for these topics regarding transcriptomics, metabolomics and statistics. No programming skills will be necessary for this course. However, some experience with computers and navigating the Internet is necessary



Course contents

Course design

In order to successfully interpret experimental results generated by these high-throughput methods we will teach in this course the principles underlying processing, analysis and visualisation of large datasets derived from transcriptomics and metabolomics experiments. The emphasis will be on statistical aspects and analysis. Relevant software will be mentioned and some will be used during hands-on exercises. During the course students are provided with a syllabus, handouts, exercises and an overview of relevant literature and Internet links

Programme topics

Day 1: Transcriptomics data processing

We will discuss the various techniques used to make transcriptome data (microarray-data) ready for analysis. The focus will be on dealing with sources of technical and biological variation in data generation (pre-processing, normalisation and experimental design). In addition, the principles of different approaches to find differentially expressed genes will be explained (fold change, t-test, ANOVA, methods that control the False Discovery Rate). Transcriptomics data processing will be exercised.

Day 2: Transcriptomics data analysis

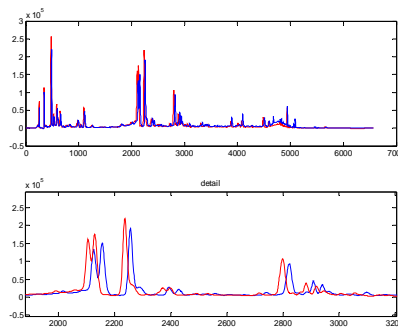
Principles of the different approaches in transcriptomics data analysis will be explained and exercised. In addition, we will discuss statistical approaches for multivariate data analysis. For the interpretation of large data sets, these analysis methods are very often linked to visualisation techniques. Therefore data visualisation techniques will be integrated in this module (clustering, PCA). Ways of relating large numbers of genes or metabolites to phenotypic data will be discussed.

Day 3: Metabolomics data processing

Metabolomics data can be obtained using different analytical techniques. In this course we will focus on GC-MS and LC-MS. We will explain the principles of data structure and the steps that are needed to process metabolomics data for analysis (baseline extraction, peak selection, chromatogram alignment). In exercises chromatograms and MS spectra will be compared manually and processed metabolomics data will be analysed in Excel and through clustering and PCA analysis.

Day 4: Classification methods and Systems biology

After identifying mass differences, the masses need to be assigned to compounds. We will explain the paths towards compound identification. Different statistical approaches for classifying biological samples by their expression or metabolic profile will be treated, as well as cross validation procedures for estimating the quality of the classifier, and the selection of genes or metabolites useful in predicting a phenotype of interest.



Organisation

Course coordinators

Dr. Chris Maliepaard – Department of Plant Breeding
Mrs. Ingeborg van Leeuwen-Bol – Graduate School VLAG

Lecturers:

Dr. Peter Schaap, Fungal genomics
Dr. Sander van der Krol, Plant Physiology
Prof. Fred van Eeuwijk, Biometris

General Information

Date & duration

The course will be held from 8-11 December 2008 (4 days)

Study load

The study load of this course is 1.4 ECTS credits

Language

The course will be conducted in English

Location and accommodation

Lectures will be given at:

Wageningen University

Building 352; Rooms: C90 (lectures) PC93/94 (practicals),

Arboretumlaan 4,

Wageningen

A number of rooms have been blocked for course participants at the WICC, but only until November 1st, 2008. Accommodation costs are about 75 (single room) per night incl. breakfast, excl tax.

Participants have to book their own hotel room.

Hotel reservation is handled by WICC, Phone: +31-317-490133, <http://www.wicc.nl/>.

Mention booking code: "SB'08".

Contact information

More information about the course contents can be obtained from:

Dr. Chris Maliepaard (chris.maliepaard@wur.nl)

For organisational matters please contact :

Mrs. Ingeborg van Leeuwen-Bol (ingeborg.vanleeuwen-bol@wur.nl)

Registration and course fee

The course fee includes coffee/tea during the breaks, lunches, course material and one course dinner.

VLAG/EPS PhD students:	€200
PhD students:	€400
University staff/Non-profit	€700
Industry	€1400

Cancellation may be free of charge until October 8th, 2008. After this date the charge will be 25% of the fee paid or due. Substitution may be made until the start of the course

Please register by completing the on-line [registration form](#). In total 40 seats are available. There will be no selection procedure. However, due to the space limitation it is advisable to register early.