

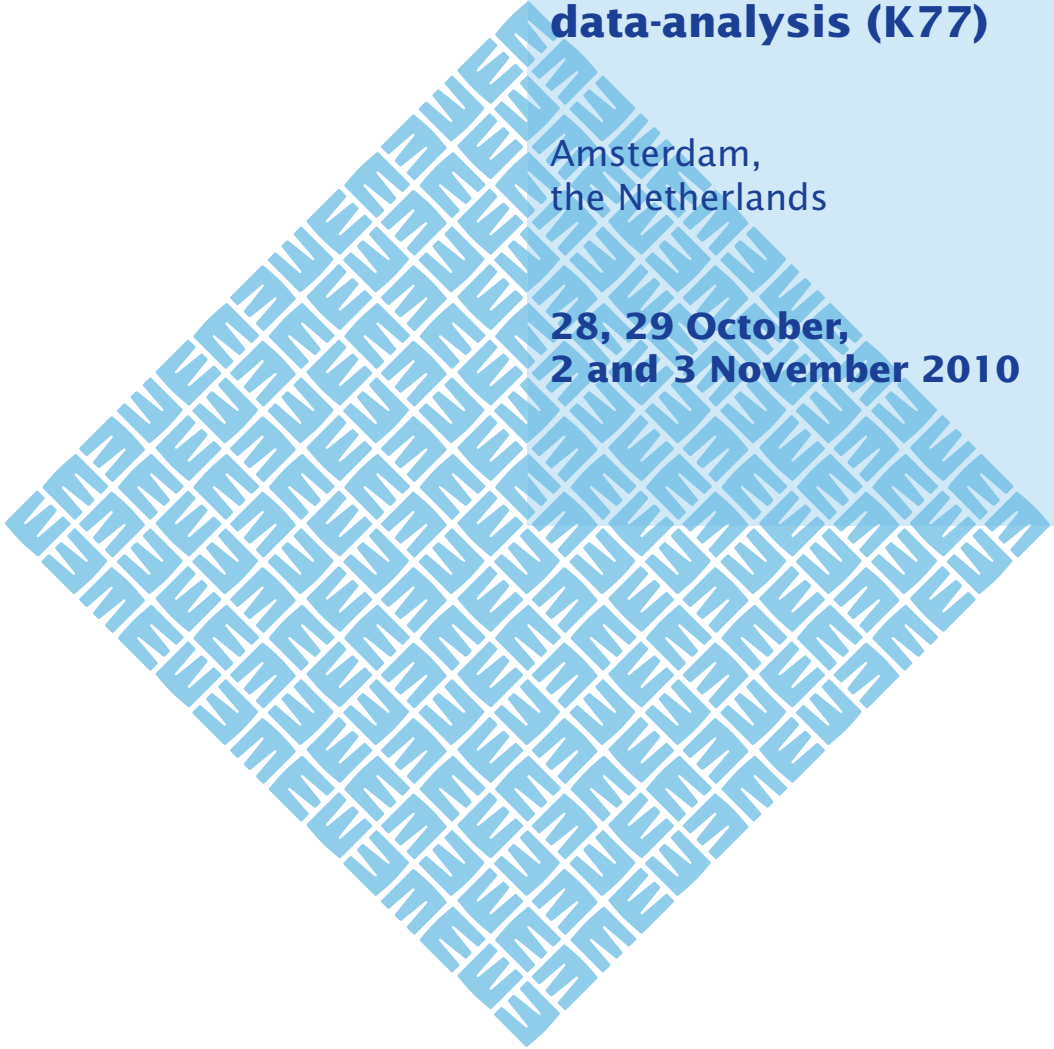


International course on

**Statistics for genomic
data-analysis (K77)**

Amsterdam,
the Netherlands

**28, 29 October,
2 and 3 November 2010**



Course description and topics

With the advent of genomic techniques, many molecular aspects of the cell are measured. Typical for data generated with such techniques is the high-dimensionality - for each of a limited number of patients (for instance), thousands of 'genes' are measured. This turns the traditional paradigm of statistics, in which a limited number of characteristics of many patients is registered, upside down. Traditional statistical methods may be inappropriate for the analysis of high-dimensional genomic data, which requires new statistical methodology. This is introduced through the analysis of gene expression and DNA copy number data. Central to this are the canonical questions:

- How to design an experiment? The principles of experimental design are illustrated in genomic experiments.
- How to get from raw intensity to biological signal? Several normalization techniques for both single and dual colour arrays are treated.
- How to find differentially expressed genes? The knowledge of hypothesis testing is briefly refreshed. Terminology and problems surrounding simultaneous testing of many hypotheses (genes) are introduced. The analysis of complex designs is introduced using the Bioconductor package Limma.
- How to identify (new) subgroups of samples? This is handled by means of hierarchical clustering methods and principal components analysis.
- How to predict clinical outcome? Several prediction methods and concepts like variable selection, cross-validation and evaluation measures are discussed.

These problems, raised by the high-dimensional character of genomic data for these questions, are treated theoretically as well as in practical settings.

The course focus is on gene expression and DNA copy number arrays, but most methodology applies also to the analysis of data from (among others) microRNA expression and SNP arrays.

Learning objectives

1. The participant realizes that a fruitful experiment requires a good design.
2. The participant is familiar with R and is able to load genomic data into R, install and activate libraries required for the analysis.
3. The participant knows how to normalize gene expression data and DNA copy number data.
4. The participant is capable of analysing a comparative microarray experiment using Limma and identifying differentially expressed genes. Therein he/she knows how to deal with the fact that multiple hypotheses are tested simultaneously.
5. The participant knows how to perform a cluster analysis, judge and present visually its results.
6. The participant is able to build and evaluate a binary predictor correctly.
7. The participant knows the pitfalls of existing analyses and is able to critically judge the statistical analysis of genomic data as performed by others.

Programme

Day 1	Introduction to R Experimental design Quality control & normalization
-------	---

Day 2	Hypothesis testing Multiple testing Regression
Day 3	Class discovery Class prediction
Day 4	Copy number data analysis

Format

This course is intensive and covers the basic concepts and methods required for the analysis of high-dimensional data. Discussed statistical methods will be practiced on experimental microarray data with the statistical software R, which is both open-source and freeware. With every lecture followed by a hands-on session on the computer, students can quickly put theory into a practical context. The lecturers will set aside ample time for questions and answers.

Also, medical papers that use the presented methodology will be discussed. The course is concluded with a final assignment, during which students have the chance of performing a complete analysis using their newly acquired skills.

Course pre-requisites

The following concepts are assumed known by students at the start of this course:

- Basic statistics: mean, variance, standard deviation.
- Probability distributions: normal, Student's t.
- Statistical testing: t-test, Wilcoxon test, significance level, p-value, null & alternative hypotheses.
- Measures of association: Pearson correlation, linear regression.

These are typically seen during a basic statistics course, e.g. "Introduction to biostatistics and clinical epidemiology" (R02).

This course also assumes that students know what gene expression microarrays are. For those who do not, but still wish to attend, we refer to the articles in Nature Genetics, Vol 21 No. 1s, January 1999 - supplement "The Chipping Forecast".

Target group

The course is tailored for researchers (such as pathologists, psychological biologists, human geneticists, oncologists, neuro-geneticists) whose research involves experiments that generate genomic data.

Lecturers

R. X. de Menezes, PhD

Department of Epidemiology & Biostatistics, VU University Medical Center

M. A. van de Wiel, PhD

Department of Epidemiology & Biostatistics, VU University Medical Center and

Department of Mathematics, Faculty of Exact Sciences, VU University Amsterdam

W. N. van Wieringen, PhD

Department of Epidemiology & Biostatistics, VU University Medical Center and

Department of Mathematics, Faculty of Exact Sciences, VU University Amsterdam

Course material

On the first day of the course you will receive a packet which contains copies of all lecture presentations, assignments, and feedback on these assignments.

Place and time

The course will take place on 28, 29 October and 2, 3 November 2010 at the Open Universiteit, Studiecentrum Amsterdam, Amstelveenseweg 390, 1076 CT Amsterdam. Course programme from 9.30 - 16.30 hrs.

Costs

The participation fee is € 1,095 and includes course material, coffee/tea, lunch.

Enrolment and cancellation

Please submit your application using the enclosed form or the website: http://www.epidm.nl/aanmelden/aanmelden_cursussen.asp. Once we have received your application, we will send a notice of acceptance and an invoice with instructions for payment.

Spaces are filled on a first-come first-serve basis.

A maximum of 28 participants may enrol.

- If you cancel before 16 September 2010, the administration costs are € 75
- If you cancel between 16 September and 14 October 2010, 50% of your participation fee will be refunded.
- No refund will be made for cancellation requests postmarked after 14 October 2010.

Only written cancellation requests will be accepted.

Organization

For further information, please contact:

Ms. Eva Stokx (course secretary) or
Ms. Marjolein Stuij, MSc (coordinator)

EpidM, Department Epidemiology and Biostatistics
VU University Medical Center
Van der Boechorststraat 7
1081 BT Amsterdam

Telephone +31 (0)20 44 48188
Fax +31 (0)20 44 48181
E-mail epidm@vumc.nl
Website www.epidm.nl