

Statistics for Microarray Data Analysis

Short Course, IBC2004

Gordon Smyth¹, Jean Yang², James Wettenhall¹, Matt Ritchie¹

1. Walter and Eliza Hall Institute of Medical Research

2. Division of Bioinformatics, University of California, San Francisco

Sunday 11 July 2004

8.30am – 5.30pm

Radisson Plaza Hotel

Description

Microarray technology, which provides a way to globally measure differential gene expression, promises to be extremely useful for the diagnosis, treatment, and prevention of complex disease as well as for the elucidation of biological mechanisms. These studies yield tens of thousands of simultaneous gene measurements from each biological sample. Issues in measurement and calibration of the microarrays need to be addressed appropriately in order to obtain valid datasets. To gain insight into genes and their function, patterns of expression and expression changes must then be discerned from high-dimensional data in which the number of observations is small relative to the number of variables. The purpose of the one-day short course in Statistics for Microarray Data Analysis is to introduce statisticians and other researchers to statistical issues in the design and analysis of microarray studies of current interest to biologists and biomedical researchers.

Experience with statistical methods and in data analysis is a pre-requisite, but no previous exposure to microarray data is assumed. The course will include the opportunity for participants to apply statistical methods to several datasets that will be provided.

Participants will be expected to bring their own laptops and to have downloaded the specified software and datasets prior to arriving in Cairns, or to purchase CDs with the same on site. Those paying the reduced fees will be able to share a laptop if necessary.

Schedule

8:30 – 9:15 Lecture 1 (Gordon, Jean)

Introduction to gene expression studies

Introduction to microarray technology and platforms

Pre-processing for spotted two-colour arrays: image analysis, data exploration, within-array normalization, between-array normalization

9:15 – 10:15 Lab 1 (James, Matt)

Examine GenePix and SPOT data using limmaGUI. Setup targets and spot-types files. Input data. Examine effects of different background corrections on MA-plot. Command-line data entry and normalization using limma. [If time, use of spotTools and gpTools to generate summary one-page pictures.]

10:15 – 10:30 Break

10:30 – 11:00 Lecture 2 (Gordon)

Details of Affy platform.

Pre-processing for spotted single-channel arrays: normalization

Pre-processing for Affy arrays: normalization, data exploration

11:00 – 11:30 Lab 2 (James, Jean, Matt)

Introduction to affy package. CEL files, CDF package, AffyBatch and exprSet objects. Image plots of robust regression weights.

11:30 – 13:00 lunch

13:15 – 14:15 Lecture 3 (Gordon)

Differential expression

Linear models with affy data

Linear models with two-colour data

Single channel analysis of two-colour data

Design of two-colour experiments

14:15 – 15:15 Lab 3 (James, Matt)

Differential expression and linear modelling using limma.

15:15 – 15:30 break

15:30 – 16:15 Lecture 4 (Jean)

Classification

Clustering

Annotation

16:15 – 17:00 Lab 4 (Jean with help)

Clustering and classification.

17:00 – 17:30 Lecture 5 (Gordon)

Assessing methodology: truth construction

Software overview