

# *ExGen:MARIE* Workshop 2005

## Statistics for Microarrays in R

ICMS, Edinburgh 21-22 September 2005.

We are pleased to announce the course “Statistics for Microarrays in R” which is going to take place in ICMS, Edinburgh, 21-22 September 2005. The course will be taught by Dr Ernst Wit and a team of experts from the MARIE consortium. This course is sponsored by a BBSRC Exploiting Genomics grant. Registration is **FREE** and **no prior knowledge of R required**.

More information and to register, please visit:

<http://exgen.ma.umist.ac.uk/course/>

### Course description

Gene expression profiling has become a routine technique that can be useful to many applied life scientists at some stage of their research. The relative ease with which it is possible to generate thousands of gene profiles stands in no comparison to the difficult and often treacherous path of getting reliable conclusions or even interesting research directions from these data. Typically, microarray data is very noisy and consists of only a small number of independent replicates. This problem is compounded further because traditional computational and statistical techniques are not designed to deal with the large numbers of variables, i.e. genes.

The aim of the course is to present practitioners of microarray technology with state-of-the-art methodology applicable to all stages of the microarray analysis: experimental design, data normalisation, and interpretation of the data. Although some of lectures will involve the use of mathematical concepts, the emphasis will lie on practical analysis in front of a computer and no previous knowledge of R is presupposed.

The course provides practical training in key aspects of microarray technology. The course is taught in three 2.5 hour slots, each with a separate theme. The course is conducted via a mix of lecture-style explanations of statistical concepts and hands-on explorations of real microarray data using an R interface.

**Recommended reading** for the course:

*Statistics for Microarrays: Design, Analysis and Inference* by E. Wit and J. D. McClure, J. Wiley and Sons, 2004.

## Course outline

### *1<sup>st</sup> day*

*12:00 Arrival participants and lunch*

#### ***13:00 – 15:30 DESIGN OF MICROARRAY EXPERIMENTS***

- Understand the difference between technical and biological variation and explain how pooling mRNA samples reduces biological variation of gene expression.
- Understand how considerations about optimal design can have a real impact on the practical set-up of microarray experiments.

#### ***16:00-18:30 NORMALIZATION AND QUALITY CONTROL***

- Deal with missing data, such as data imputation.
- Understand the pros and cons of sequential normalization techniques.
- Perform spatial, dye, within- and across-condition normalizations.
- Perform multiple normalizations to detect sensitivity of analysis

*19:30 Social dinner*

### *2<sup>nd</sup> day*

#### ***9.00 - 12.00 DIFFERENTIAL EXPRESSION, CLUSTERING AND PREDICTION***

- Distinguish between different test-statistics for differential expression and understand when each of them is appropriate.
- Understand different error rates involved in multiple testing, such as the False Discovery Rate (FDR), the Family-wise Error Rate (FWER) and the False Positive Rate (FPR).
- Understand and use different methods for clustering genes and samples.
- Explain the importance of gene filtering and the variance-bias trade-off for classifying of samples.
- Understand and use methods (LDA and k-NN) for classifying microarray samples and understand how to assess associated misclassification rates.
- Use penalized regression (LASSO) and k-NN regression to relate gene profiles with continuous outcomes associated to the mRNA sample.