

**4th Genomics-Bioinformatics Day on "Expression Data Analysis"
October 5th 1 PM 2004 in the Henry Wellcome Centre for Gene Function.**

There are many exciting research projects currently underway in and around Oxford in the field of expression data analysis. In recognition of this, Jotun Hein, Chris Holmes and Jennifer Taylor are pleased to invite you to the 4th Genomics/Bioinformatics day, this time dedicated to the analysis of expression data. The previous three days were dedicated to “Comparative Genomics”, “Pathogen Analysis” and “Modelling in the Biosciences”, respectively. These days provide a forum to discuss research ideas and questions, and aim to foster collaborations and integrated approaches to shared challenges. If you want to participate, please email mcdermot@stats.ox.ac.uk, so that refreshments and access to the building can be arranged.

1.00pm ***Hein, Holmes & Taylor: Introduction and overview
Bioinformatics, Department of Statistics***

The measurement of mRNA concentration has over the last decade risen to prominence in molecular biology and medicine. It can be used to classify cells as a diagnostic tool, since different cell types and cancers have characteristic levels of gene expression and as a means to probe the dynamics of the cell under different circumstances. However, expression data are also hard to analyze in a statistically rigorous way: It is high dimensional, highly correlated and noisy. We will briefly review the history of the field, discuss some of the informatics challenges and point to current trends and potential areas of future research.

1.25pm ***Sach Mukherjee: “A Theoretical Analysis of Gene Selection”
Department of Engineering Science***

A great deal of recent research has focused on the challenging task of selecting differentially expressed genes from microarray data (‘gene selection’). Numerous gene selection algorithms have been proposed in the literature, but it is often unclear exactly how these algorithms respond to conditions like small sample-sizes or differing variances. This talk introduces a simple notion of accuracy in selection, and goes on to show, via a theoretical analysis, how the choice of gene selection algorithm can critically influence the quality of results. We will also look briefly at recent work which addresses some of the issues raised, by learning an appropriate gene ranking function from data.

1.50pm ***Anne-Mette K. Hein (S. Richardson, H. Causton G. Ambler and P. Green):
“Bayesian hierarchical models” Imperial College London***

We present Bayesian hierarchical models for the analysis of Affymetrix GeneChip data (BGX). The approach we take differs from other available approaches in two fundamental aspects. First, we aim to integrate all steps in the analysis in a common statistically coherent framework, allowing all components and thus associated errors, to be considered simultaneously. Secondly, inference is based on the full posterior distribution of gene expression levels and derived quantities such as fold changes or ranks, rather than on single point estimates. Measures of uncertainty on these quantities are thus available.

2.15pm ***Coffee Break***

2.30pm ***William Cookson group: “Microarrays, asthma and dermatitis”
Wellcome Trust Centre for Human Genetics***

We have used affymetrix microarrays to investigate the characteristics of epithelial cells from the skin and airways. In one experiment we have attempted to identify genes that are expressed during keratinocyte growth and differentiation. In other experiments we hope to identify pathway-specific differences in cellular responses to bacterial components. In a further set of experiments we intend to use microarray expression levels from EBV cell lines as quantitative traits for genetic mapping.

2.55pm ***Cath Willoughby and Pete Underhill: “Whole genome expression profiling
of postimplantation development in the mouse” MRC Harwell***

We overview a recent study that we performed at MRC Harwell to characterise whole genome expression profiles in postimplantation mouse development with a focus on genes from the Del(13)Svea36H deletion. Tissue samples were collected at nine time points between implantation and birth; representing a time span of 19.5 days. We discuss the experimental procedures and the initial statistical analysis to normalise, filter and then cluster the data into coherent groups showing similar time profiles.

3.20pm ***Richard Mott et al.,: “Finding deletion and polysomy using genomic DNA microarrays”***

Wellcome Trust Centre for Human Genetics

The importance of cytogenetically visible rearrangements in human genetic disease has long been recognised and there is now abundant evidence showing that smaller, less readily detectable chromosomal rearrangements can also be clinically important. The full significance and extent to which such cryptic rearrangements contribute to human genetic disease has yet to be determined. One way of elucidating this is by comparative genome hybridization (CGH) to DNA microarrays (array CGH). However, recent array CGH studies noting false negative and false positive results raise important concerns regarding the suitability of the approach for the detection of constitutional chromosomal rearrangements, particularly in a clinical diagnostic environment where a robust assay, providing clear, high quality results of measurable significance is required. Here, we present the results of array CGH studies that address these concerns. Using tiling path arrays for the terminal 2Mb of chromosome 16p and telomere specific arrays we show that probes vary in their usefulness for detecting copy number changes and that sequence composition cannot be used to predict the probe behaviour. Significantly, we present a powerful new method of statistical analysis that overcomes problems of probe specificity and sensitivity by combining fluorescence signals from arrayed probes that map to neighbouring genomic locations. The method uses a modification of the Smith-Waterman algorithm that also provides a measure of robustness. We tested the method using hybridisation data from the 2Mb 16p tiling path array and show that 100% of monosomies >250kb were identified with a high degree of significance and that the boundaries of the deletions were accurately and robustly located. The studies are an important step towards optimising array CGH based approaches and improving their suitability for clinical diagnostic purposes.

3.45pm ***Dr Natalia Bochkina: “Statistical modelling for BAIR project”***
Imperial College London.

BAIR stands for Biological Atlas for Insulin Resistance. It is a collaboration between several Universities (Imperial College London, together with other London Colleges, University of Oxford and University of Cambridge) sponsored by Wellcome Trust. The project is aimed at studying the action of insulin resistance at several levels, transcriptomic, proteomic, metabolic and phenotypic, and thus creating a reference - an atlas - of discovered insulin-related action. I will outline the modelling of the data involved in the BAIR project, and illustrate some of the statistical approaches applied to transcriptomic data.

4.10pm ***Coffee Break***

4.40pm ***Discussion***

This will be determined by the attendees, but natural questions are:

Where is the field moving?

Which opportunities are there to use the diverse expertise available?

What are the researchers needs and how best are they met?