

Talk 1

**Title: Finding differences between two multivariate samples**

**Tarn Duong**

School of Mathematics and Statistics,  
University of New South Wales, Sydney, Australia

**Abstract:**

A common question in flow cytometry data analysis is where two multivariate samples are different. Differences between the samples can indicate differences in biological responses. Early attempts to answer this question are based on probability binning (a binary division algorithm) combined with classical chi-squared tests. Subsequently these have been found to be sub-optimal for flow cytometry data which typically have moderate dimension (around 3 to 20) and large sample sizes (in the order of 10 000 and 100 000).

Our proposal is to use the Patient Rule Induction Method (PRIM), a bump-hunting algorithm, combined with generalized chi-squared tests. We believe that our proposed method is more appropriate since PRIM is suited to moderate dimensions and generalized chi-squared tests are suited to large sample sizes. We briefly outline our theoretical reasoning and follow it with some case studies.

---

Talk 2

**Title: Estimating regions of large differences in multivariate samples with a k Nearest Neighbours method**

**Inge Koch**

School of Mathematics and Statistics  
University of New South Wales, Sydney, Australia

**Abstract:**

We propose an extension of the k nearest neighbour estimator in density estimation to regions of density differences in a multivariate setting. Rather than estimating the density difference everywhere our method focuses on regions where the difference is large.

The density in high density regions is estimated separately for each of the two multivariate samples: the regional estimates (in regions where the density is high) combine a preprocessing step which partitions the data into disjoint clusters with a knn estimator which is applied separately to each cluster and uses different tuning parameters for each cluster. The results are combined to yield a density estimate in high density regions. Differences of the two regional density estimates will allow the determination of a region where the two samples differ mostly.

---

Talk 3

**Title: Feature Significance for Multivariate Densities**

**Matt Wand**

School of Mathematics and Statistics,  
University of New South Wales, Sydney, Australia

**Abstract:**

Features such as modes and ridges are often of interest in the underlying density of multivariate samples. Modes, in particular, may correspond to sub-populations of possible interest. The SiZer-type technology, pioneered by Chaudhuri and Marron (Journal of the American Statistical Association, 1999), is a useful framework for assessing statistical significance of such features. We describe extensions of this technology to general higher dimensions, with particular emphasis on the trivariate case. Recently advances in three-dimensional graphics in the R environment are shown to have an important role. Examples involving flow cytometric data will be presented.

---

Talk 4

**Title: Quantitation of Phosphoprotein Biomarkers in Whole Blood as a Measure of Disease State and Therapeutic Response**

**Gary Means**  
Principal scientist

**Abstract:**

Phosphoproteins integral to signal transduction are being evaluated as markers of biochemical coverage in preclinical studies. Stimulation of whole blood with a variety of ligands generates changes in specific phosphoproteins, transcripts and secreted analytes. The application of 7-color FACS protocols to selectively monitor changes in phosphoproteins in particular cell types has also led to the identification of disease-related differences in pathway responses which exist between patient and normal donor whole blood samples. Some of these cell-intrinsic changes in signaling pathway activation correlate with condition-specific differences in analyte secretion.

---

Talk 5

**Title: Modeling flow data dose response**

**Cheng Su**  
Principal Biostatistician II

**Abstract:**

In an ex-vivo study of diseased individuals, whole blood samples were stimulated and the response, as measured by protein phosphorylation, was monitored using flow cytometry. The dose response to stimulant was estimated using hierarchical mixture models which characterize the individual events in each treated sample by a mixture distribution and the proportions of the mixtures by a sigmoid dose response model.

---

Talk 6

**Title: Data Standards for Flow Cytometry**

**Ryan Brinkman**  
BC Cancer Research Centre  
University of British Columbia, Canada

**Abstract:**

First developed in 1984, the Flow Cytometry Standard (FCS) specification has kept pace with many years of technological evolution. ISAC has adopted FCS for the common representation of FCM data, and this standard is supported by all analytical instrument and third party software suppliers. Scientists can choose among instruments and software with no major compatibility issues for the raw fluorescence values that FCS captures. Previous versions of the FCS included data, metadata and analysis components within the same file. However, metadata and analysis components, if included at all, are not recorded in a standardized fashion or in sufficient detail for use by independent parties. This is assuming that independent parties can access experimental results, as important data sets supporting publications are almost invariably unavailable. Finally, if metadata annotation takes place at any time subsequent to data capture, the all-inclusive format of FCS necessitates the generation of a new version of the file, which replicates the (hopefully unmodified) primary data.

---

Talk 7

**Title: Representing Flow Cytometry Experiments within FuGE**

**Josef Spidlen**  
BC Cancer Research Centre

**Abstract:**

The lack of bioinformatics standards in flow cytometry (FCM) prevents collaborative opportunities to recreate experimental methods and results. Flow cytometry data file standard (FCS) has been adopted as the common representation of FCM data without major compatibility issues for fluorescence values that FCS captures. However, experimental metadata and analysis components are not recorded in a standardized fashion for use by independent parties. Flow cytometry community has identified the need of a shared object model to capture experimental workflow including data and metadata in broad context.

Flow Informatics and Computational Cytometry Society (FICCS) Object Model Working Group (FICCS OMWG) has been established during the FICCS meeting in Seattle, WA, USA, in September 2006. The group committed to evaluate possibilities of building an FCM-specific object model based on existing technologies and models; possibly reusing and sharing common concepts with other 'omics' sciences. Representatives of the working group met for a development workshop in Dallas, TX, USA, in October 2006. The main purpose of the workshop was to closely evaluate whether the Functional Genomics Experiment (FuGE) Object Model is useful as the core for the FCM object model, and, based on the results, specify methodology and strategy to develop the first initial draft of the model. The timeline and resulted action items were specified aiming to present an initial draft solution at the FICCS meeting in Seattle, WA, USA, in November 2006.

FICCS OMWG chose FuGE as the core for the FCM object model. The main areas to start extending FuGE were identified as:

- (i) *Flow cytometry protocol package.* FCM-specific classes shall be extended from FuGE to support FCM-specific actions including computational and data processing actions.
- (ii) *Data package.* Event though FCS is supposed to represent FCM data, an FCM-specific data class was identified as essential to capture associations between parameters (FCS) and conceptual molecules (FuGE).
- (iii) *Material package.* Physical characteristics of material used as FCM sample are describable using general FuGE concepts. FCM extension mainly captures the role of various subcomponents in conjugated samples (i.e., reporter and detector), which is crucial to associate detected fluorescence values with properties of biological interests.

Several FCM use cases will be modeled within the FCM-extended FuGE in order to provide immediate feedback and validation.

---

Talk 8

**Title: The use of flow in vaccine development**

**Stephen De Rosa**

UW/Fred Hutchinson Cancer Research Center

**Abstract:**

---

Talk 9

**Title: Statistical positivity criteria and multiplicity adjustment methods for the analysis of gated ICS assay data**

**Yunda Huang, Zoe Moodie, Steve De Rosa, Steve Self**

Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

**Abstract:**

To maximize the multi-parameter capability of ICS assays in HIV-1 vaccine development and evaluation, it is generally not sufficient to claim an overall difference in the percentage of positive staining cells between antigen stimulated and negative control wells. We therefore model gated ICS assay data as multivariate binomial distributed, and conduct separate Fisher's exact tests to determine the positivity of each cytokine subset response. In this way, we are able to isolate specific significances and to take into account the sparseness of such data.

In addition, since the number of tests increases linearly with the number of antigens and exponentially with the number of cytokines of interest, we adopt resampling-based multiplicity adjustment methods for multivariate binary data developed by Westfall and Young to control the family wise error rate of these tests. We confirmed in our simulation studies that the power of the tests can be greatly improved using such methods that properly incorporate the discreteness and the dependence structure of the distributions. Results from recent ICS assay validation studies will also be presented.

---

Talk 10

**Title Automated Population Discovery in Multidimensional FACS Data**

**Yu Qian**

UT Southwestern, Dallas, TX, USA

**Abstract:**