Microarray gene expression studies have revealed the complexity of cellular function. In even the simplest biological system, multiple biological pathways are simultaneously active and these are regulated by RNA, genes, proteins and metabolites. Therefore in order to fully understand gene expression data, we need to investigate it in this biological context. Integrating heterogeneous types of biological data generated from diverse technology platforms is challenging but we need to develop methods for analysis of these data in order to understand cellular function and in turn develop predictive models of cells.

We have recently published a series of papers on the application of multivariate analysis approaches for the visualization and exploratory analysis of microarray data with other heterogeneous "omics" data sets. In this talk, I will describe the application of co-inertia analysis to the cross-platform integration of microarray data, but also to the more interesting problem of understanding the relationship between gene and protein expression data (Fagan, et al, 2007). I will present our approach to discovering which transcriptional pathways have been activated or repressed in microarray gene expression data (Jeffery, 2007), and discuss how we can extend these approaches to build methods for exploratory data analysis of multiple biological datasets.

Culhane, A.C., Perriere, G. and Higgins, D.G. (2003) Cross-platform comparison and visualisation of gene expression data using co-inertia analysis. BMC Bioinformatics, 4:59.
Jeffery IB, Higgins DG, Culhane AC. (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics. 7:359.
Fagan, A., Culhane, A.C. and Higgins, D.G. (2007) A multivariate analysis approach to the integration of proteomic and gene expression data. Proteomics, In press
Jeffery IB, Madden SF, McGettigan PA, Perriere G, Culhane AC, Higgins DG. (2007) Integrating transcription factor binding site information with gene expression datasets. Bioinformatics. 23(3):298-305.