

## **BioISI Research Seminar**

# **Quantifying overall (Cross-)correlation in High-Dimensional omics data**



**Claus-Dieter Mayer, PhD**

Biomathematics & Statistics Scotland (BioSS), Aberdeen

**When: Thursday, May 10 - 1h00**

**Where: Building C1, Room 1.3.20**

Extremely high-dimensional data sets from gene expression (microarrays, RNAseq experiments) or metabolomic studies are commonly generated in biological and medical experiments. Variables (genes, metabolites) measured in these experiments typically interact with each other in gene regulatory networks or metabolic pathways, leading to correlation between variables. From a purely statistical point of view this can be a nuisance. In a highly multiple testing setting methods to control the family wise error rate (FWER) or the false discovery rate (FDR) usually assume independence or only weak correlation of variables. From a biological point of view, however, strong correlations are often of particular interest because they indicate the activation of important processes. For either case it is useful to have a method that measures and quantifies the overall correlation either in the whole data set or in relevant subsets (e.g. pathways). Often it is also of interest to quantify the cross correlation between two such omics data sets, e.g. to study how strongly linked methylation status and gene expression are in samples taken from the same subjects. We will focus on methods that are computationally fast and are easy to interpret. We will show how these numbers can help as exploratory tools before conducting a network analysis for a single omics data set or an integrative analysis of two or more such data sets. We will also discuss if and how measurements of overall correlation could be useful in controlling the family wise error rate (FWER) or the false discovery rate (FDR) that arise in when testing in such high-dimensional data.