**Venue:** P638, QUT Gardens Point, Brisbane, Australia

**Date:** Friday 4th November 2016

**Title:** Multivariate data analysis methods for biological data using the R package mixOmics

**Overview**

Multivariate dimension reduction approaches are useful exploratory tools to get a first understanding of large and complex data sets. These approaches are extremely efficient to compute and highly flexible as they can answer a variety of biological integrative questions. Our latest developments in that exciting area of research include feature selection and statistical integration of several ‘omics data sets.

The workshop will introduce the fundamental concepts of multivariate dimension reduction methodologies for data exploration, identification of biomarkers and integration of large data sets; especially in the context of systems biology, or in research areas where statistical data integration is required. Each methodology that will be introduced will be applied on biological ‘omics studies including transcriptomics, metabolomics and proteomics data sets using the R package mixOmics ([http://mixomics.org/](http://mixomics.org/)).

**Goals**

After completion of this workshop, participants will be able to

1. Understand fundamental principles of multivariate projection-based dimension reduction technique,
2. Perform statistical integration and feature selection using recently developed multivariate methodologies,
3. Apply those methods to high throughput biological studies, including their own studies.

**Key statistical concepts introduced**

Covariance and correlation, linear combination of features, supervised classification, prediction, cross-validation, selection of diagnostic or prognostic markers, $l_1$ and $l_2$ penalties.

**Topics covered**

1. *Key multivariate methodologies available in mixOmics*
   A. Exploration of one data set with Principal Component Analysis
   B. Identification of biomarkers to discriminate different treatment groups with PLS-Discriminant Analysis
   C. Integration several ‘omics data sets and identification of correlated biomarkers

2. *Review on the graphical outputs implemented in mixOmics*
   A. Sample plot representation
   B. Variable plot representation for data integration analyses

3. *Case studies and applications*
A. Three case studies will be analysed using the methods presented above

**Important prerequisite and software requirement**
We expect the audience to have a good working knowledge in R programming (e.g. handling and subsetting data frames, perform simple calculations and graphical outputs) in order to fully appreciate the workshop.
Software requirements are described on our webpage: [http://mixomics.org/2016/09/requirements-mixomics-workshops/](http://mixomics.org/2016/09/requirements-mixomics-workshops/)

**Target Audience**
The mixOmics workshop is aimed to data analysts in the fields of bioinformatics, computational biology and applied statistics with a good statistical knowledge and a good working knowledge in R.

The tutorial will be useful to those interested in:
- exploring large data sets
- selecting features with methods implementing LASSO-based penalisations
- using graphical techniques to better visualise data
- understanding and/or applying multivariate projection methodologies to large data sets.

**Presenter**
**Dr Kim-Anh Lê Cao** (The University of Queensland Diamantina Institute, Brisbane Australia) is an expert in multivariate statistical methods and novel developments. Since 2009, her team has been working on developing mixOmics dedicated to the integrative analysis of `omics' data, to help researchers mine and make sense of biological big data ([http://www.mixOmics.org](http://www.mixOmics.org)).

**Tutors**
**Dr Florian Rohart** (University of Queensland Diamantina Institute, Brisbane Australia) is a core developer for mixOmics and develops cutting-edge multivariate methods for horizontal and vertical integration.

**Mr Nicholas Matigian** (Biostatistics facility, The University of Queensland Diamantina Institute) is specialized in high-throughput data analysis and provide assistance and support to UQDI researchers.