

Summer School on Multivariate data analysis methods for biological data using the R package mixOmics

Are you struggling with understanding and analyzing large and complex biological data sets? The Toulouse GenoToul Biostat platform, the Laboratory of Plant-Microbe Interactions and the Plant Science Research Laboratory are organizing a Summer School within the EU COST Action "The quest for tolerant varieties: phenotyping at plant and cellular level (FA1306) (http://www.cost.eu/COST_Actions/fa/FA1306?). The course will provide both theory and hands-on training focusing on multivariate data analysis and dimension data reduction for biological studies.



The objective of this Summer School is to introduce the fundamental concepts of multivariate dimension reduction methodologies. Those methods are particularly useful for data exploration and integration of large data sets, and especially in the context of systems biology, or in research areas where statistical data integration is required. Each methodology that will be presented during the course will be applied on biological "omics" studies including transcriptomics, metabolomics and proteomics data sets using the R package mixOmics (<http://mixomics.org/>).

Organized by GenoToul Biostat platform (<http://perso.math.univ-toulouse.fr/biostat/>)
Laboratory of Plant-microbe Interactions (LIPM, http://www6.toulouse.inra.fr/lipm_eng/)
Plant Science Research Laboratory (LRSV, <https://www.lrsv.ups-tlse.fr/?lang=en>)

Dates Monday 12 September until Wednesday 14 September 2016

Practical information

The course fee is 300 € for the academia and 600 € for the industry. It covers tuition, course material, coffee/tea, lunches and 1 diner downtown on Monday 12 September. Please note that PhD students and early postdocs can apply for a bursary from the COST Action FA1306. There is room to reimburse course fee accommodation and travel for a maximum of 12 people.

Location: Toulouse, France

Maximum number of participants: 30

Deadline for application: 15 June 2016

Decision for attendee selection and grant allowance: 30 June 2016

Registration: 4 to 15 July 2016

Apply for this Summer School at:

<https://www.surveymonkey.com/r/FLSYJZP>

In addition, send a CV to E Jamet (jamet@lrsv.ups-tlse.fr) and do not forget to mention if you want to apply for a grant.

Contact in Toulouse

E Jamet

LRSV

jamet@lrsv.ups-tlse.fr

Prerequisite and requirements

The audience is expected to have a good working knowledge in R (e.g. handling data frames and perform simple calculations). Attendees are requested to bring their own laptops, having installed the software RStudio <http://www.rstudio.com/> and the R package mixOmics (instructions provided prior to the training).

More details on the covered topics

1. *Key methodologies in mixOmics and their variants*
 - A. Exploration of one data set and how to estimate missing values
 - B. Identification of biomarkers to discriminate different treatment groups
 - C. Integration of two data sets and identification of biomarkers
 - D. Repeated measurements design
 - E. Introduction to the integration of more than two data sets
2. *Review on the graphical outputs implemented in mixOmics*
 - A. Sample plot representation
 - B. Variable plot representation for data integration
 - C. Other useful graphical outputs
3. *Case studies and applications*

The following statistical concepts will be introduced: covariance and correlation, multiple linear regression, classification and prediction, cross-validation, selection of diagnostic or prognostic markers, cross-validation, l_1 and l_2 penalties in a regression framework. Each methodology will be illustrated on a case study (we will alternate theory and application).

Note that mixOmics is not limited to biological data only and can be applied to other type of data where integration is required.

Target group

The course is intended for data analysts in the fields of bioinformatics, computational biology and applied statistics with a good statistical knowledge and **a good working knowledge in R**. It will be particularly useful to those interested in:

1. *Exploring large data sets.*
2. *Selecting features with methods implementing LASSO-based penalisations.*
3. *Using graphical techniques to better visualise data.*
4. *Understanding and/or applying multivariate projection methodologies to large data sets.*

Results

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.*

Course leaders



Pr **Kim-Anh Lê Cao** (University of Queensland Diamantina Institute) was awarded her PhD in 2008 at Université de Toulouse, France. She then moved to Australia as a postdoctoral fellow at the University of Queensland. Kim-Anh is currently leading the Computational Biostatistics Methods group at UQDI and is the head of the UQDI Biostatistics facility that provides statistical support to researchers from her institute. Her research interests focus on the development of innovative statistical approaches for the analysis and the integration of large biological data sets for studies in cancer, and diseases involving the immune system, including arthritis, chronic infections, and diabetes. She has been teaching Statistics to undergraduate and postgraduate (UQ Bioinformatics Master's program) students for 4 years. Together with the mixOmics team, Kim-Anh continues to develop methodologies to analyse complex biological studies.

More details on Kim-Anh's current research projects: <http://www.di.uq.edu.au/dr-kim-anh-le-cao>



Dr **Sébastien Déjean** was awarded his PhD in Applied Statistics in 2002 at Université de Toulouse, France after spending 4 years in a Biometry lab at INRA (French National Institute for Agronomic Research). He then has been working at the Toulouse Mathematics Institute as a research engineer. He works in close collaboration with researchers working in different areas, such as high-throughput biology, chemistry and information retrieval. Sébastien is an expert in statistical data analysis and he contributes to the development of several R packages including mixOmics as a core member.

More details on Sébastien's current research projects: <http://perso.math.univ-toulouse.fr/dejean/>

Additional teachers

Dr David Rengel (data analyst, LIPM)
 François Bartolo (post-doc, Toulouse Institute of Mathematics)
 Harold Duruflé (PhD student, LRSV/Toulouse Institute of Mathematics)

Local organizing committee

Dr François Bartolo (Toulouse Institute of Mathematics)
 Dr Sébastien Dejean (Toulouse Institute of Mathematics, Toulouse Biostat platform)
 Catherine deprey (LRSV)
 Pr Christophe Dunand (LRSV)
 Harold Duruflé (LRSV)
 Dr Elisabeth Jamet (LRSV)
 Dr Nicolas Langlade (LIPM)
 Dr David Rengel (LIPM)



Garonne quay in Toulouse