Sponsor: visiting scientist program INP Toulouse

Title: mixOmics Advanced Workshop (**MAW'17**)

Venue and time:
Venue: INRA, Toulouse
Dates: 23-24 October 2017

Instructors:
Dr Kim-Anh Lê Cao, University of Melbourne, Australia
Dr Sébastien Déjean, Universté de Toulouse, France

Tutors:
Two tutors TBA, maybe

Motivation: *(Describe the motivation for the course. Why is it important and timely to present this topic, with reference to current developments in theory and applications? Also, provide an overview of the proposed course content.)*

The mixOmics package has undergone substantial improvements and methodological developments in the last 18 months to address the strong demand from the computational and biological community to integrate multiple (>2) `omics data sets, including microbiome, genotype and longitudinal data. The aim of this advanced workshop is to introduce our new frameworks and encourage discussions, collaborations and suggested improvements on the following themes:

1. N-integration with **DIABLO**: the integration of several `omics data sets measured on the same biological samples but using different types of technological platforms (note: excludes SNP and categorical data). The aim is to identify a correlated multi-`omics molecular signature explaining a phenotype of interest.
Reference: http://biorxiv.org/content/early/2016/08/03/067611

2. P-integration with **MINT**: the integration of independent data sets measured on the same P variables (e.g. genes) but different studies. The aim is to identify a robust molecular signature across those independent studies (note: mostly focused on gene expression data)
Reference https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1553-8

3. Longitudinal `omics analysis with **timeOmics**: time course profiles are modeled using linear mixed model splines, time shift analysis is performed using Fast Fourier Transform, leading to clustering of similar but sometimes delays time trajectories across `omics. The aim is to identify clusters of `omics profiles explaining a phenotype of interest.
Status: in development (methods are implemented in separate packages and will be merged into mixOmics)
Two References:
1- http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134540
2- http://www.nature.com/articles/srep40131

4. Exploratory multivariate analysis with **SNPOmics**: sparse multiple correspondence analysis has been implemented for the identification of Single Nucleotide Polymorphism (SNP). We also hope to present a preliminary supervised version to explain a phenotype of interest.
Status: in development (unsupervised sparse MCA will be implemented in mixOmics for the MAW)
Reference:
http://www.biometrics.org.au/conferences/Hobart2015/talks2015/Wednesday/C_1350_Wed_AnneBernard.pdf

5. **mixMC**: mixOmics for Microbial communities: we have extended sPLS-Discriminant Analysis for 16S / metageonomics data (http://mixomics.org/mixmc-mixomics-for-16s-microbial-communities/). We are

currently expanding the mixMC module to integrate several metagenomics data sets with an unsupervised kernel PCA method (work in collaboration with Jerome Mariette (PhD with Dr Nathalie Villla-Vialaneix).
mixMC: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0160169
multiple kernel learning:
http://maiage.jouy.inra.fr/sites/maiage.jouy.inra.fr/files/u43/RCAM2016_JMariette.pdf

The participants invited to this workshop will be asked to choose the topics that align with their research interest. The number of participants is limited to 12.

Proposed length of course: 2 full days

Target audience of the course:

Academic statisticians working on the development of statistical methodology (including graduate students)
Epidemiologists/statisticians working in the field of epidemiology
Statisticians working on applications from clinical medicine
Statistical geneticists/statisticians working on genomic data
Statisticians working in the pharmaceutical industry
Statisticians working on applications from ecology/forestry/agriculture

Prerequisites:

Since this is an advanced course, we expect the participants to be **expert in R programming language** and familiar with multivariate projection based methods and mixOmics.

Outline: (Provide a detailed description of what will happen in each course session.)

The workshop will be adapted to the research interest of the participants. Potentially three of the four themes described above will be presented.

Each half day will be organized as follows:
- Short presentation of the methods and implementation by one instructor
- Presentation of a case study by one of the participant,
- Computer lab: analysis of the case study and discussions

Last half day
- Detailed discussions, suggestions for improvements.

Learning Outcomes:

Participants will further improve their knowledge and skills in `omics data integration, and will thus be able to analyse complex biological data with our latest integrative tools and methods. Our discussions will enable participants to explore further areas of research. Our practicals will also encourage collaborations.

Instructors:

Dr Kim-Anh Lê Cao (University of Melbourne, Brisbane Australia) 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, Brisbane Australia. Since the beginning of her Ph.D Kim-Anh has initiated a wide range of valuable collaborative and research opportunities in both statistics and molecular biology. Her main research focus is on variable selection for biological data (`omics' data) coming from different functional levels by the means of multivariate dimension reduction approaches. Since 2009, her team has been working on developing a statistical software dedicated to the integrative analysis of `omics' data, to help researchers make sense of biological big data.

Kim-Anh is senior lecturer at the University of Melbourne, and regularly runs statistical training workshops and short series seminars as well as mixOmics multi-day workshops (12 mixOmics workshops, totalling 20 days since 2014). More details on Kim-Anh current research projects: http://sysgen.unimelb.edu.au/research/research-groups/computational-biostatistics-methods-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 (Université de Toulouse, France) 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.

He is also currently involved in training workshops to teach statistics and scientific softwares to scientific and administrative staff. He can talk about fishing rod, pizza-box and soccer ball to explain Principal Component Analysis; do you see why? More details: http://perso.math.univ-toulouse.fr/dejean/

Laptop:

The participants are requested to bring their own laptop. Instructions on the latest R version and packages download will be given prior to the workshop.