Project team Persyvact2 (2015-2018)

A follow up of the Persyvact exploratory project

PIs of Persyvact 2

Pierre-Olivier Amblard
GIPSA-lab

Michael Blum
TIMC-IMAG

Adeline Samson
LJK

Objectives

- Develop cutting edge data science methodologies to analyse large biomedical data.
- Perform collaborative research by bringing together researchers of different scientific fields of data science (statistics, machine learning, image and signal processing).
- Enhance the international visibility of data science research in Grenoble.

Main statistical challenges with biomed data

- Dimension reduction (PCA, graphs) because of
  - Huge number of individuals
  - Huge number of biological markers (omics data)
- Correlation/repetition
  - Same individuals measured multiple times
  - Correlation between sensors or markers

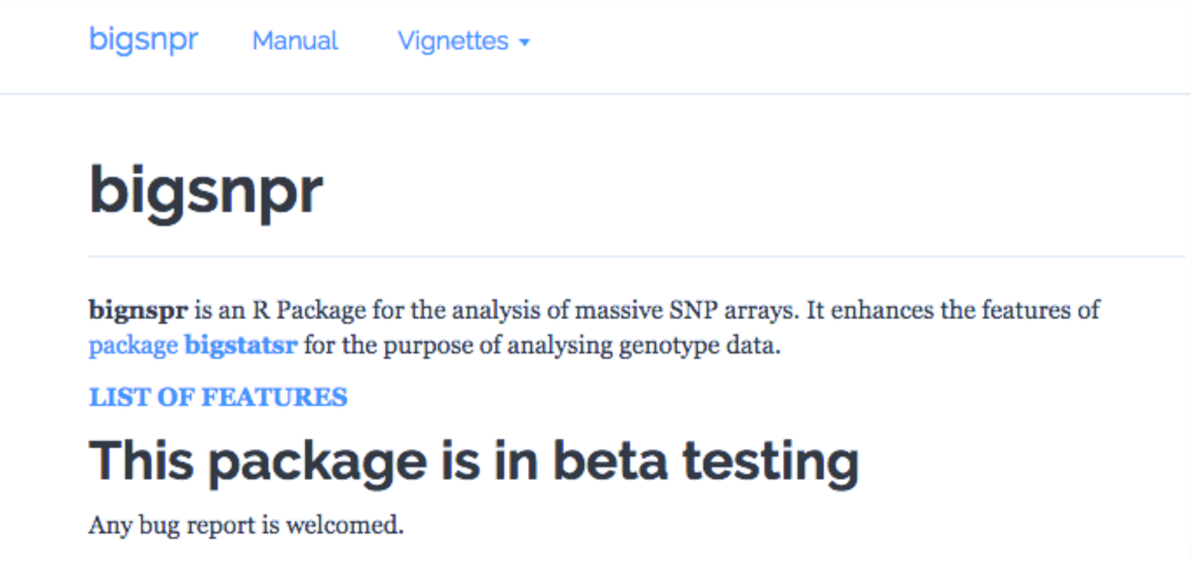
Main Scientific Results and Ongoing Work

- 1. Genome Scans
- 2. Neuroscience
- 3. Pharmacocinetik
- 4. Epigenetic

R packages for genome scans

Personalized medicine: from genomes sequenced in case/control cohorts to risk score for complex diseases.

Delivery of two R packages (bigstatr and bigsnpr) to analyse large data set including genomic data. Produced by Florian Privé, which is a PhD student of the project team.



bigsnpr

bignspr is an R Package for the analysis of massive SNP arrays. It enhances the features of package **bigstatsr** for the purpose of analysing genotype data.

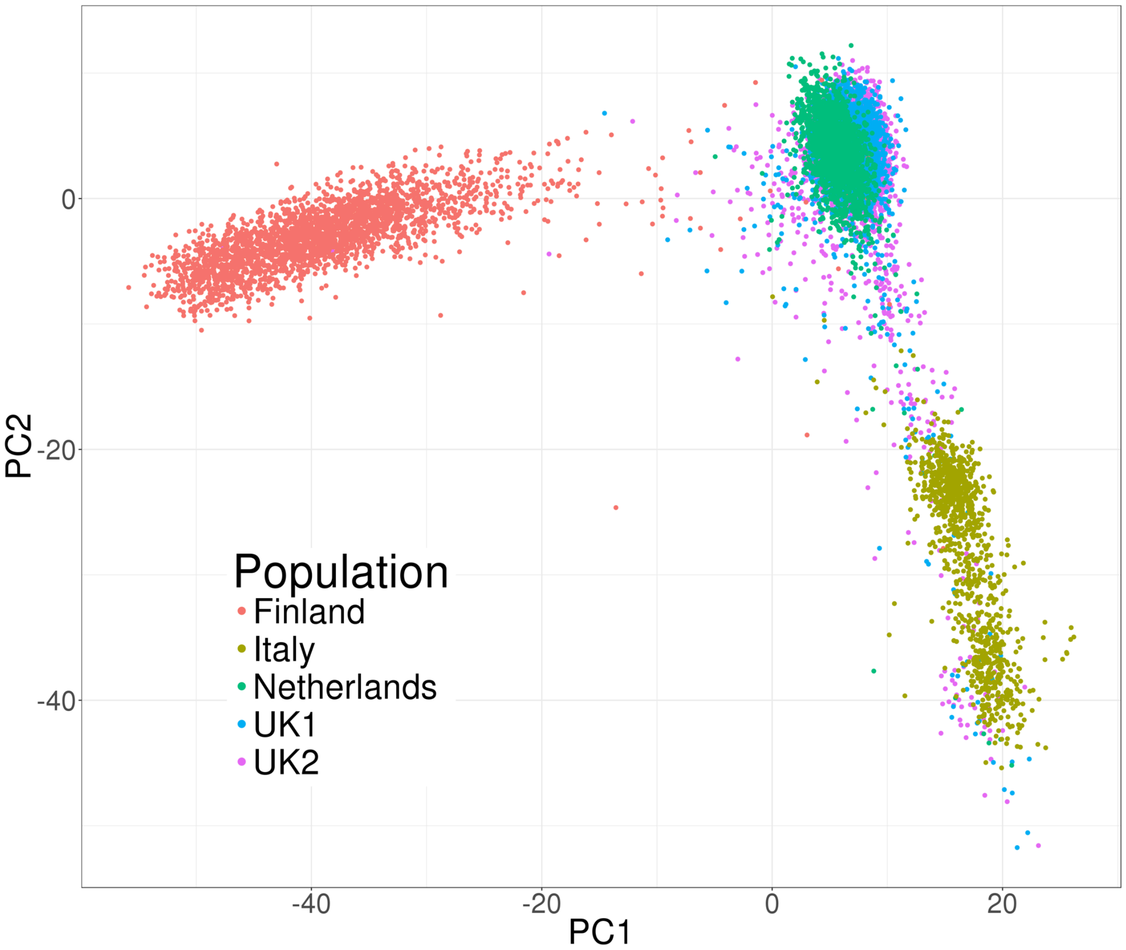
LIST OF FEATURES

This package is in beta testing

Any bug report is welcomed.

Data reduction for Genome Scans

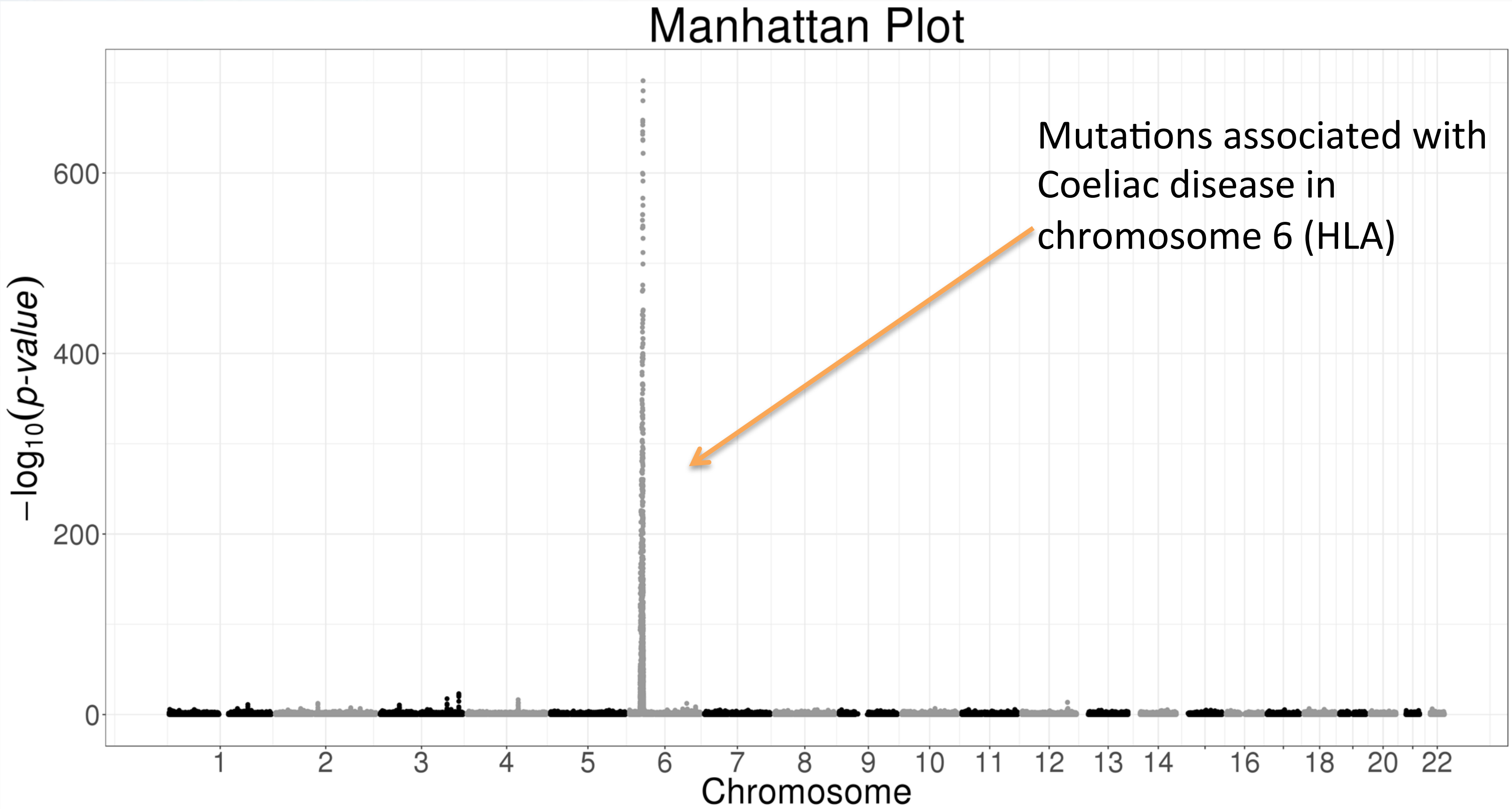
R package provides fast computation of principal component analysis for large datasets comprising more than 1,000 individuals and more than 1,000,000 SNP markers (mutations).



PCA for a case-control Coeliac data

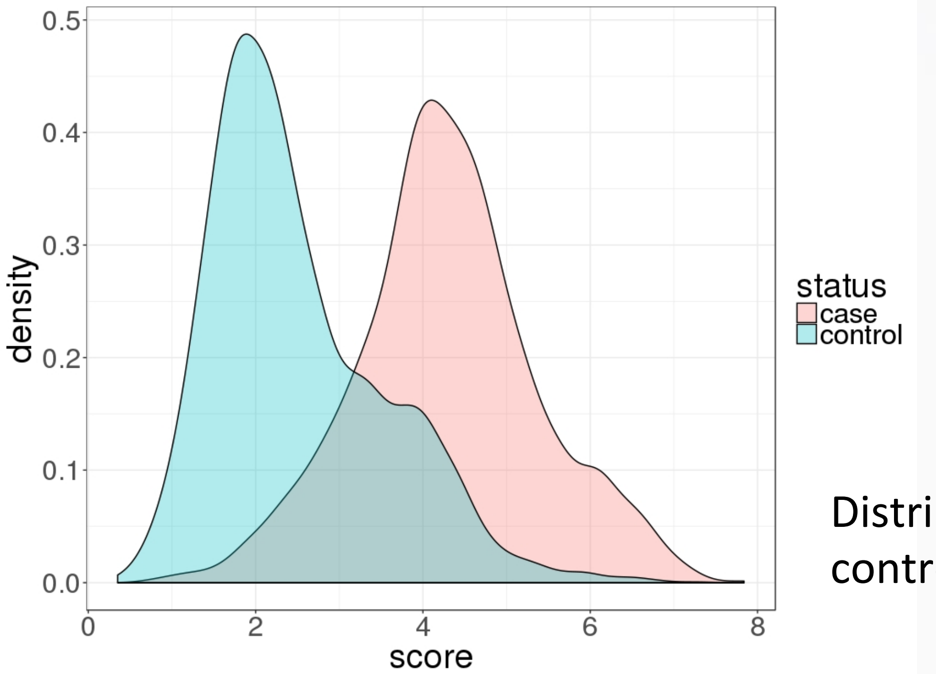
Finding associations in Genome Scans

R package bigsnpr can detect mutations associated with disease



Mutations associated with Coeliac disease in chromosome 6 (HLA)

Risk score in Genome Scans



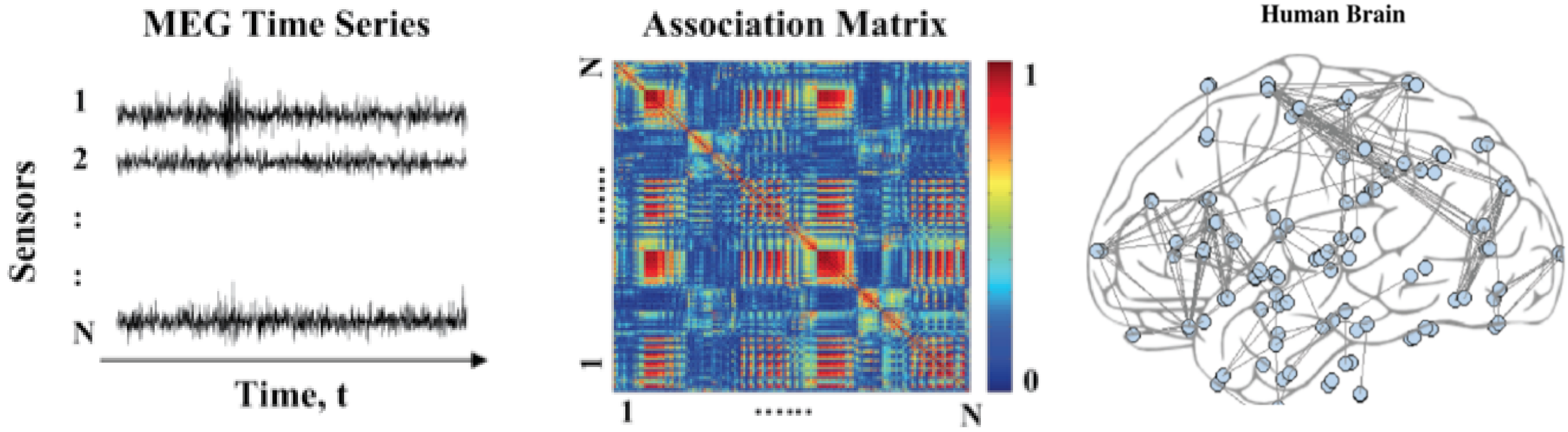
Distribution of risk scores for control and cases (Coeliac)

- Ongoing work: Use one of the largest health ressource to provide risk score computation for several complex diseases.

biobank uk
Improving the health of future generations

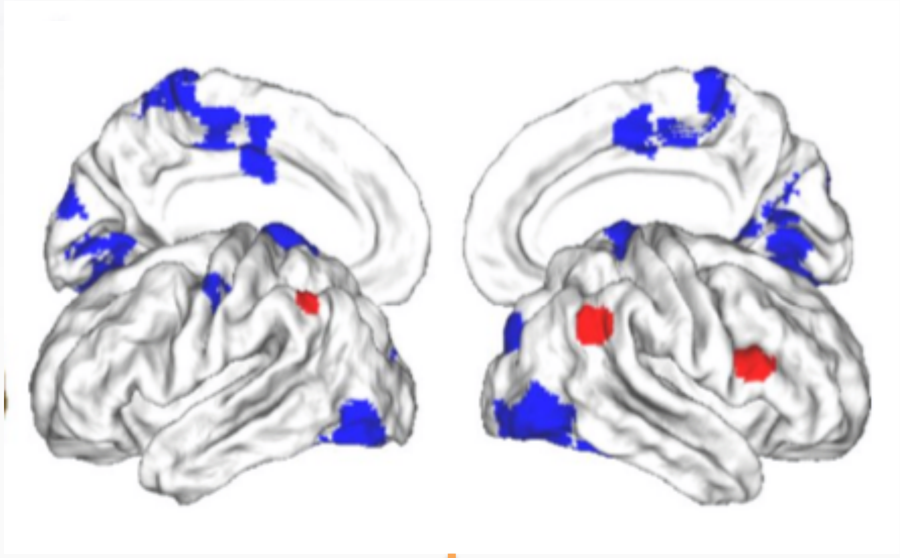
Data reduction in Neuroscience

Using graphs to explore data



Signatures of diseases

Graph analysis of brain network in patients with stroke



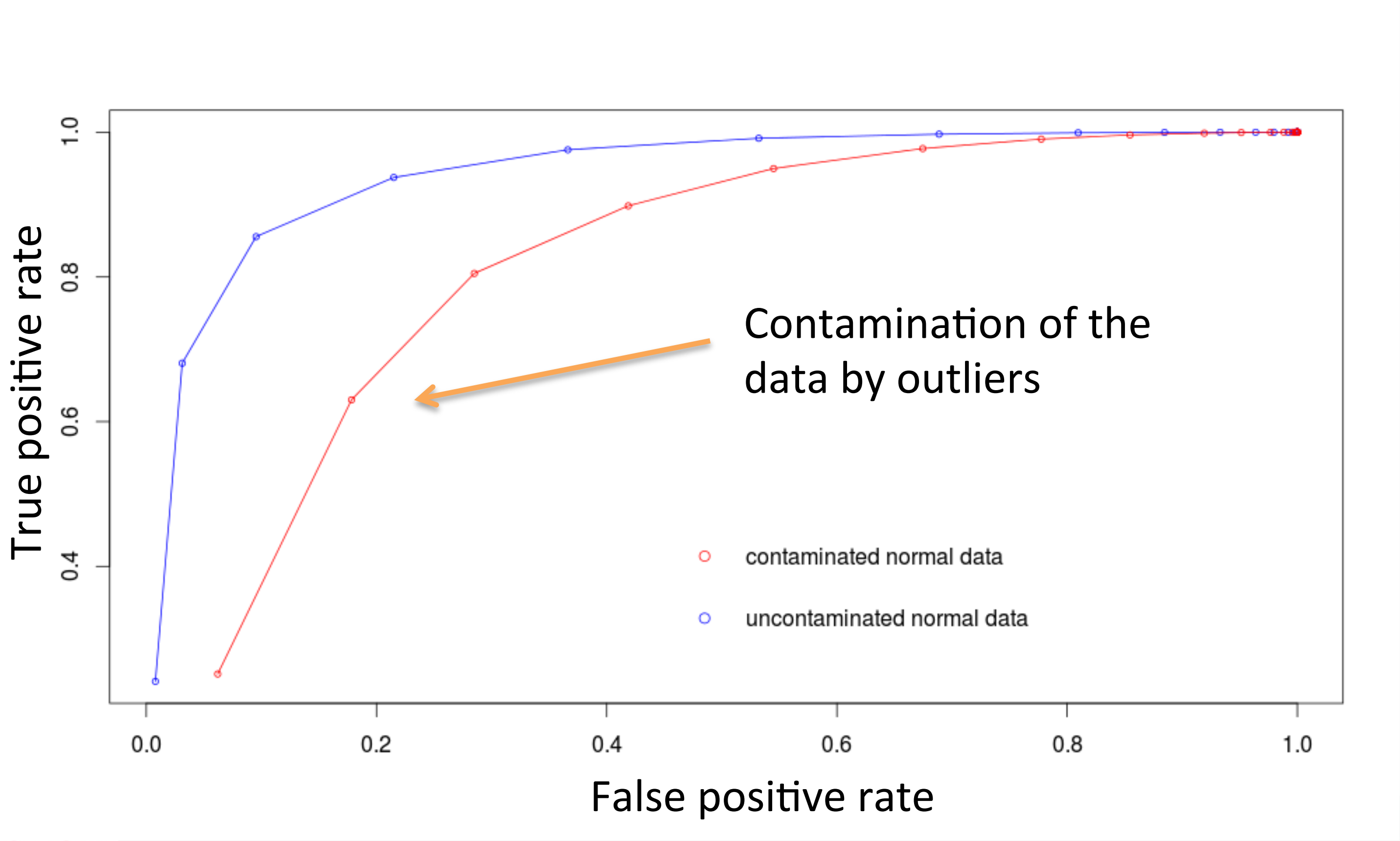
🟦 Significant decrease of connectivity

🟥 Significant increase of connectivity

→ Hub disruption index to quantify change of connectivity

Robustness of graph reconstruction

Robustness of graph lasso reconstruction: PhD work of Karina Ashurbekova, a PhD student of the project team.

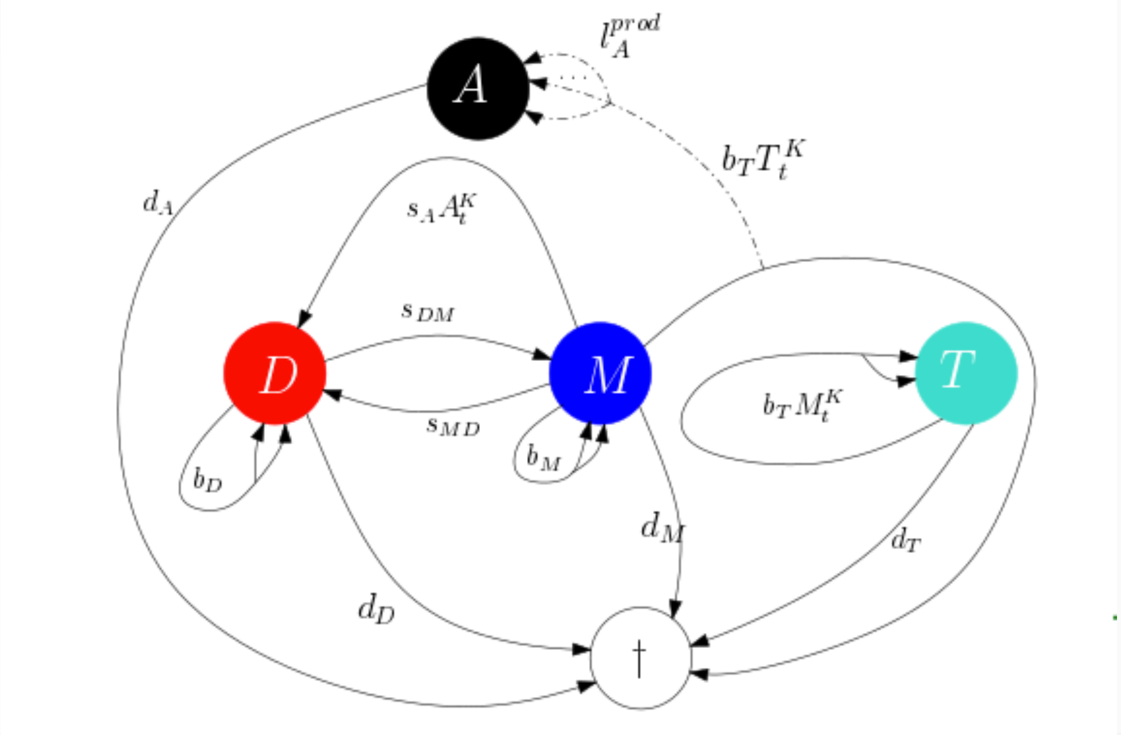


Contamination of the data by outliers

Pharmacokinetic

Stochastic model in personalized medicine: treatment optimization for skin cancer immunotherapy.

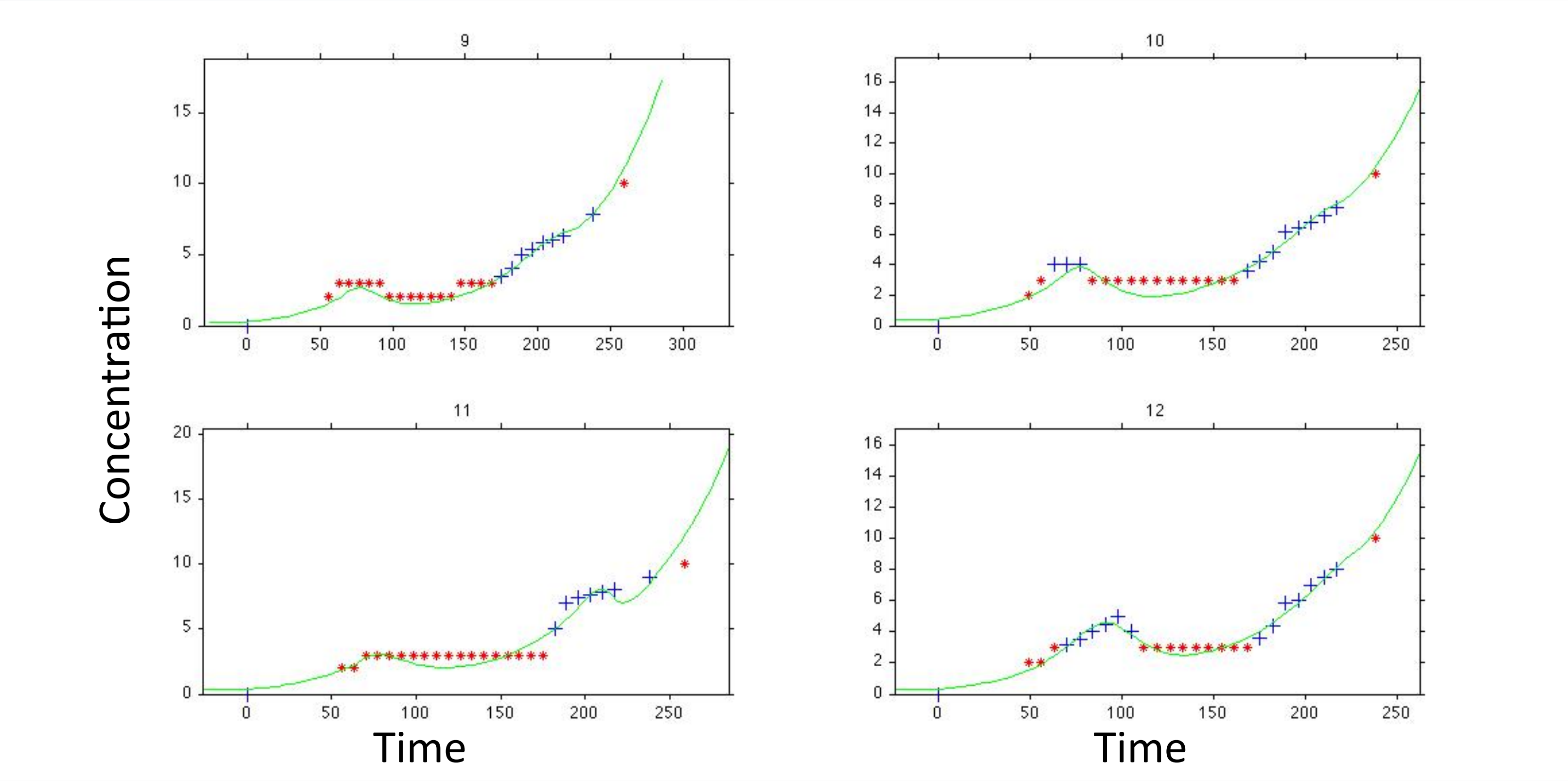
PhD work of Modibo Diabaté, a PhD student partly funded by the project team.



Treatment optimization=
Minimization of T cell extinction probability

Pharmacokinetic

Good fit to the data and biological validation of computational method for treatment optimization is ongoing

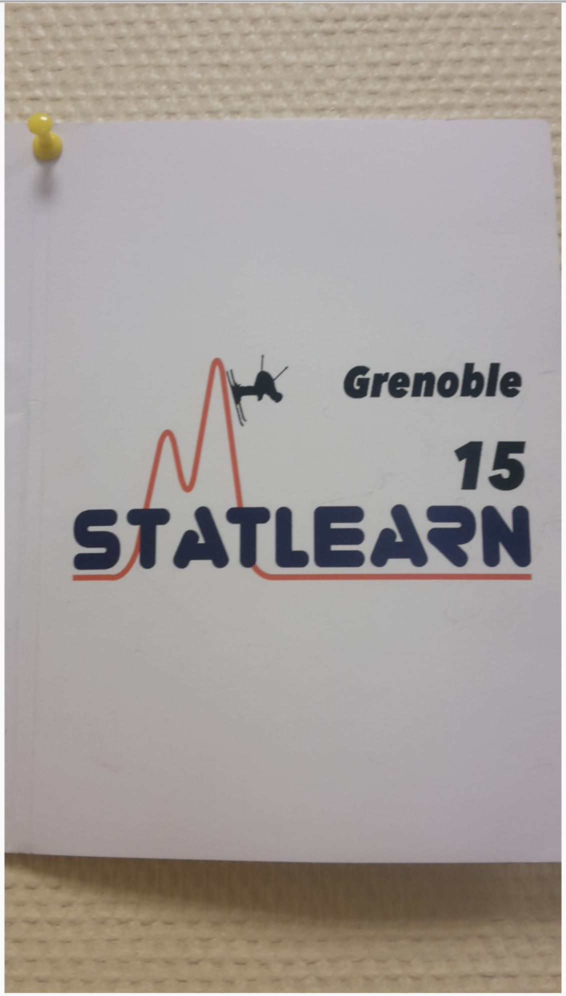


Scientific events

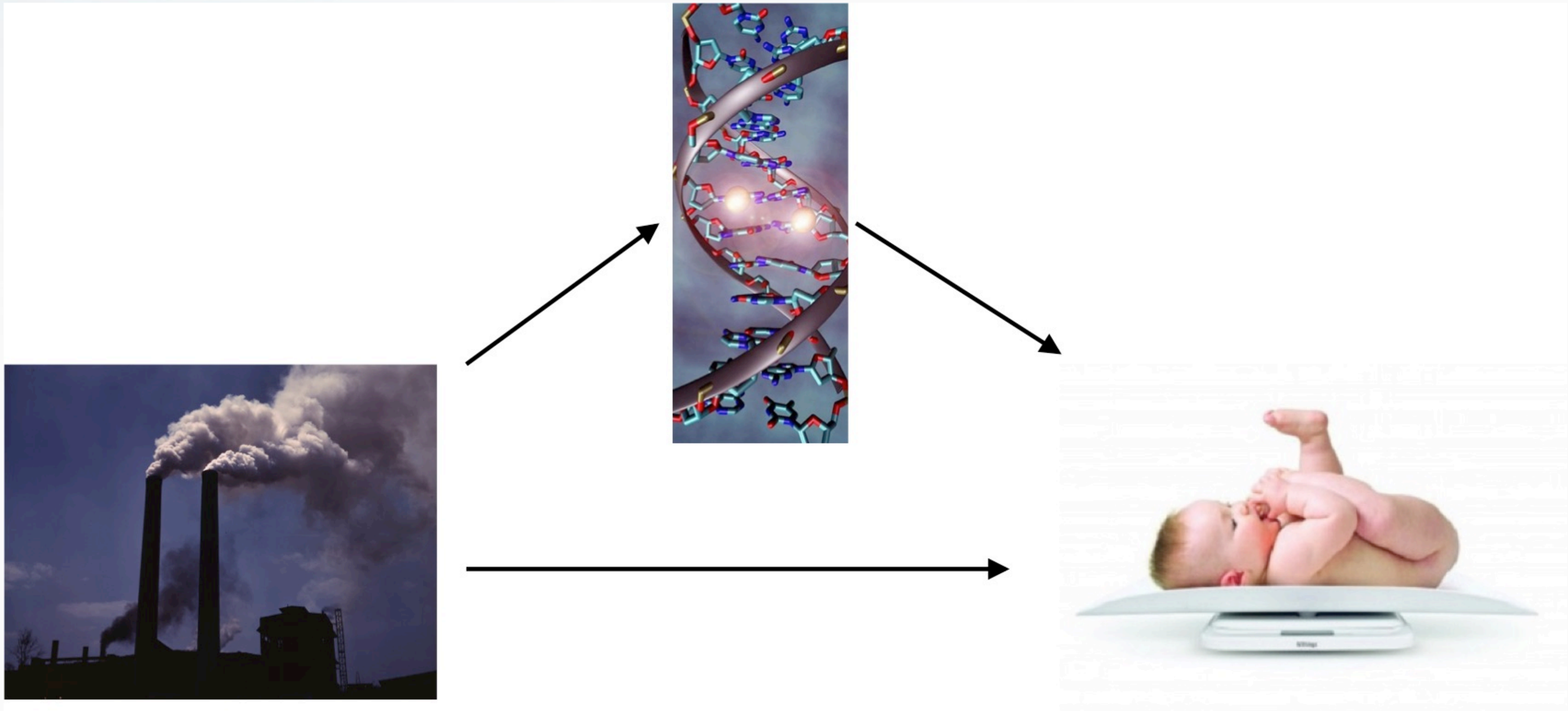
Organization of an international workshop in statistical learning

150 participants

Speakers from academia (ENS Paris, UCL London, Univ Bologna, Univ Valencia, KU Leuven) and the private sector (Facebook, Microsoft, Xerox).



Epigenetic & High-Dimension Mediation Data Challenge



Includes epidemiologists from the Institute for Advanced Biosciences (IAB, Grenoble)

A data challenge on epigenetics

Collaboration

Active learning

Creativity (beaten by my PhD student)

Aussois, June 7-9 2017

To be continued

Data Institute
Univ. Grenoble Alpes

To promote applications of data science in social science, biology, earth science…