Genome-Wide Association Study (GWAS)

Background: Every human being, and more generally every living thing, has his own DNA sequence. We call this sequence the genotype. Genotype can determine almost every character of a person, and changes between every individual. Every mutation in this sequence can lead to a change in the way to be of a person. DNA is used by the cells to obtain proteins, which in turn synthesize the other substances present in our body. We call the ensemble of this substances the metabotype. The metabotype in turn affect what we call the phenotype, which is the set of observables characteristics, like the colour of the eyes, the colour of the skin and many others, even all the diseases can be grouped under the phenotype characters. So, we can see that a variation of the genotype can affect the metabotype, that affects the phenotype, taking in some cases to diseases.

Goal: The goal of GWAS is to identify if there is any significant connection between one or more known mutations in the genotype and the risk factor of a determinate disease.

Methods: To reach our goal we did many regressions between each known mutation site (SNP) of the genotype (in our case only of the chromosome 5, for some reasons) and any metabotype feature (taken from the urine).

Dataset: We obtained our data thanks to the COLAUS project. Colaus is a large stud based on the population of Lausanne, it is more than 6’000 persons.

Mathematical tools: We used Matlab to create all the plots, to manipulate all the matrices containing the data and mainly to effectuate our about 200 million regressions.