Metabolomics – linking genotypes and phenotypes

Introduction

The project we worked on, is about metabolomics that we want to explain with phenotype and genotype. For that, we analysed the data collected from the research project CoLaus, which contains information from Lausanne inhabitants. We used three types of data: phenotypes, metabolite features and genotype. Phenotypes data contains blood creatinine, urine creatinine, coffee consumption, smoking severity, alcohol consumption, age, sex, menaupausal status and gen_pc (principal component of the genotype). Metabolite features are gathered from NMR spectroscopy of urine samples and genotype data contains SNPs of chromosome 19.

Goals

The aim of this project was to model the relationship between phenotypes and metabolites, and metabolites and genotypes and finally compare our results with others published research.

Method

First of all, we wanted to explain the metabolite features with the phenotypes. To achieve this we used the stepwise regression. Each metabolite feature was explained by the phenotypes that found relevant by stepwise regression, as following:

$Y_i = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_9 x_9 + \varepsilon$

- Y_i: Metabolite feature (1276)
- β_i=coefficient of explanatory variable (phenotypes)
- x_i=phenotype (9)

We obtained a regression model for each feature. The models show which phenotypes significantly influence each of the different metabolite features. The residuals are the part of metabolite features not explained by this model and we tried to explain them with the genotype data using single linear regression:

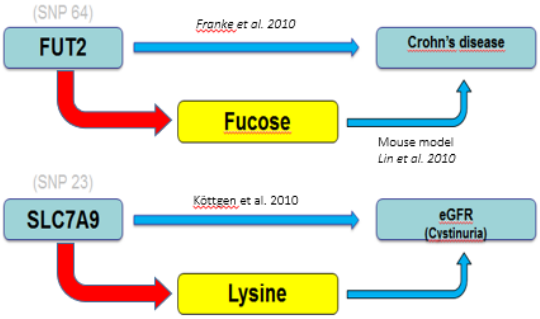
$\mathbf{R_i = \beta_j SNP_j + \varepsilon'}$

- R_i: Residual from phenotype regression (1276)
- B_j=coefficient of explanatory variable (SNPs)
- SNP_i : SNPs from chromosome 19 (33,360)

We found **109 SNPs** that are associated *significantly* with at least one metabolite feature (GWAS = p-value significance of $5x10^{-8}$). According to Linkage disequilibrium we found out that these SNPs were located on two loci.

We focused on two SNPs: rs8101881 and rs492602. Thanks to the NCBI gene bank, we found that these SNPs are respectively maps onto the SLC7A9 and FUT2 genes. Among the metabolite features that these SNPs explain significantly, we chose the most significant ones. The BMRB data bank revealed that this metabolite features defined respectively Lysine and Fucose. Our data analysis demonstrate that SLC7A9 is strongly linked to Lysine and FUT2 is strongly linked to Fucose.

Published research (*Franke et al. 2010*) demonstrates that mutation in gene SLC7A9 can cause dysfunction in renal filtration involved in Cystinuria disease. Another paper (Köttgen et al. 2010) suggests that a mutation in the FUT2 gene can cause Crohn's disease. An analysis of Fucose and Lysine in urinary samples can help detecting those diseases. More generally, the analysis of urinary sample allows to predict a probability of having some disease in a simple and fast way.



The link we found is illustrated by the red arrow,
And the blue arrows represent the links proved by published research

Students: Jonas Garessus, Julia Mermier, Luca Pellandini
Supervisor : Reyhan Sönmez