Exploring the genetics of age at menopause in women through association studies in the CoLaus population

Introduction

Menopause is the period in a woman's life when menstruations cease following a natural decline in reproductive hormones. The consequence is not only the end of fertility but also increased risks for cardiovascular diseases or osteoporosis. Knowing the genetic determinants of the age of menopause onset, women could be diagnosed for having a risk of premature (before the age of 40) or early (between 40 and 45 years) menopause. This prevention would allow women to adapt family planning. Over the long term, a gene therapy might even be possible.

The main goal of our project is to find single-nucleotide polymorphisms (SNPs) and related genes, which might have a significant impact on our phenotype, the onset age at menopause. To achieve this objective, we use data from the CoLaus study thanks to which 500'000 SNPs of over 6'000 individuals were sequenced.

Method

MATLAB

To handle such huge data, we use the numerical computing environment and programming language known as MATLAB. Thanks to this software, we create functions returning us P-values or graphics such as the Manhattan plot.

Linear regression

The model of the linear regression predicts a dependent variable Y *(in our case the onset age at menopause)* by one or more explanatory variables X *(each SNP genotyped)*. Each SNP will be assigned a regression coefficient β *(± years from the mean onset age at menopause)*. Because we assume that the relation between Y and X is not perfectly linear, the equation also contains an error variable. Linear regression is used in the candidate gene approach and in genome-wide association studies in order to characterize the association between genotype and phenotype.

Candidate gene approach

We focus on SNPs already determined as having an impact on other phenotypes, which could be correlated with the age at menopause. In order to do so, we select candidates of interest. Therefore, we say that candidate gene approach is based on *a priori* knowledge.

Genome-wide association study (GWAS)

This statistical method of analysis scans the entire genome for common genetic variation. It searches for association between many SNPs and a feature of interest and is applied to randomly chosen individuals. Every SNP will be attributed a P-value estimating the significance of the relation with the phenotype. The results can be visualized on a Manhattan plot.

Results & Analysis

Our purpose is to find SNPs with a low associated P-value, which consequently could have an impact on the age at the menopause. We have done two different analyses:

- Candidate Gene Approach
- GWAS

Candidate Gene Approach

SNP Tested			Linear regression			
SNP	**Gene**	**Diseases**	**β**	**SE**	**p-value**	**P-value adjusted**
rs13182402-G	ALDH7A1	**Osteoporosis**	-0.4942	0.2645	**0.0617**	**1**
rs784288-A	MECOM		-0.1494	0.1685	0.3752	1
rs2046210-A	ESR1	**Breast - cancer**	0.0497	0.1704	0.7707	1
rs3803662-T	TOX3		0.0006	0.1752	0.9972	1
rs8170-A	MERIT40		-0.0405	0.2563	0.8744	1
rs161645-A	NUDT12	**Depression**	-0.4038	0.1740	**0.0203**	**0.43**
rs1803274-T	BCHE	**Cardiovascular disease**	0.1844	0.1960	0.3468	1
rs7671266-T	WDR1		-0.0502	0.1943	0.7962	1
rs314280-T	LIN28B	**Menarche**	0.1668	0.1669	0.3178	1

We choose SNPs which have been highlighted to have an impact on a certain trait or disease. We test them on our phenotype (onset age at menopause) to see if they have also an impact on it. Unfortunately, no one showed a significant P-value when adjusted to the number of tests.

GWAS

Here is the Manhattan plot obtained by the GWAS analysis. Normally significance threshold is fixed at 5×10^{-8} for these types of analysis. Because we have no results with such P-value, we choose to analyse the six lowest P-values found.

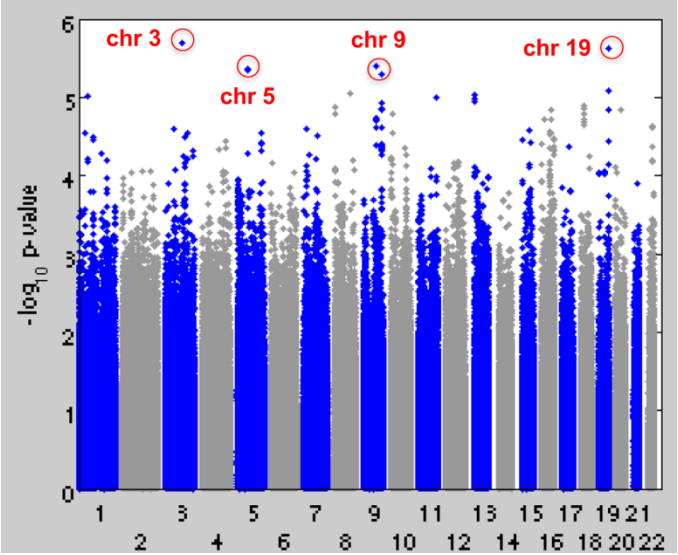

Figure 1 : Manhattan plot with lowest P-values (circled in red)

For each selected SNP, we made a zoom on the plot to be informed of their genomic context.

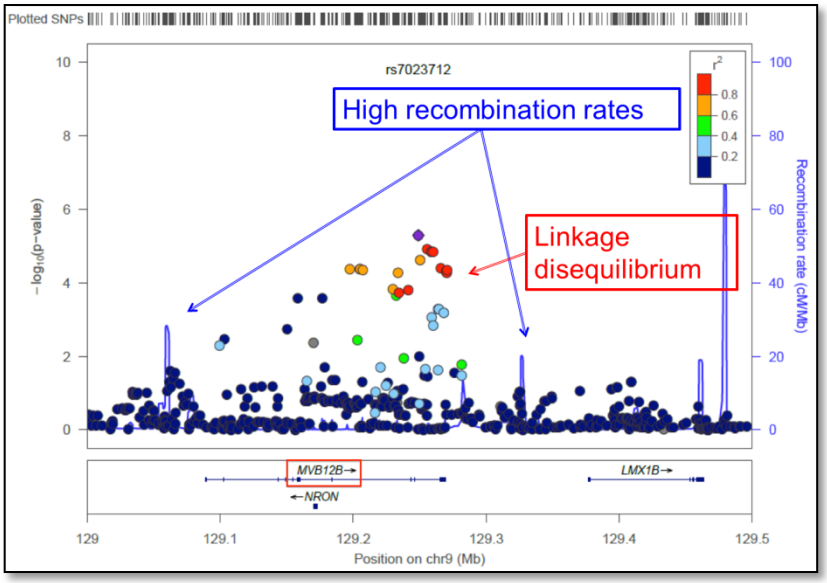

Figure 2 : Zoom for chromosome 9

On this type of graph, we can be informed on the linkage disequilibrium between the SNP of interest and other SNPs, recombination peaks and the genomic context. Here, we can see that a lot of SNPs are in linkage disequilibrium, which means they are likely to be transferred together during the meiosis. We also look for functions of the genes to know if they related to menopause. Unfortunately, we find no gene having potentially an impact on age at menopause.

Conclusion

Neither our 6 SNPs nor genes functions showed statistically significant association with the onset age at menopause. Nonetheless, the candidate gene approach highlighted a nominal significance for one gene affecting depression. On the other hand, the GWAS showed some suggestive genes like the MVB12B. In order to confirm their impact on the menopause, we would have to replicate the study on other cohorts and to analyze extreme values. If our results are conclusive, we will then proceed to a much deeper analysis on genes functions. To finish with, a functional validation of our hypothesis *(cell mutagenesis; impacts on reproductive function)* could confirm our theories.