5V733 - GENOM Project **Gabriela LOBINSKA, Amandine SANDRI**
Teacher: Sophie GARNIER Master BIM-BMC - 2018/2019
UMR S 1166 Equipe 1
Date: 15/01/19

Partie 1 : Stratégie et contrôle Qualité

Le présent projet consiste en une étude d'association génétique sur génome entier menée dans le but de repérer les déterminants génétiques dans la maladie de l'artherosclérose.

Nous avons mener les tests d'association sans retirer les SNPs en déséquilibre de liaison dans un premier temps puis en faisant varier le seuil de LD de **0.2**, **0.5** et **0.8** ensuite. Enfin, nous avons observé les différences de performance statistique entre les tests d'association en fonction de l'intégration ou non des covariables.

Avant de réaliser les analyses statistiques d'association sur les données issues du procédé de génotypage, il est nécessaire de réaliser un contrôle qualité [1] afin de s'assurer que nous avons des données de bonne qualité, c'est-à-dire relativement complètes et homogènes, afin d'obtenir des résultats avec un haut seuil de confiance. :

— s'affranchir des SNPs et des individus pour lesquels plus de **2%** des données sont manquantes en raison d'un mauvais génotypage

— vérifier que l'équilibre de Hardy-Weindberg est bien respecté en éliminant les SNPs dont la p-value associée au test statistique est inférieure à **e-5**

— vérifier l'absence de parenté entre les individus en calculant les coefficients **IBS** et **IBD**

— s'affranchir des SNPs trop rares en éliminant ceux dont la fréquence allélique est inférieure à **5%**

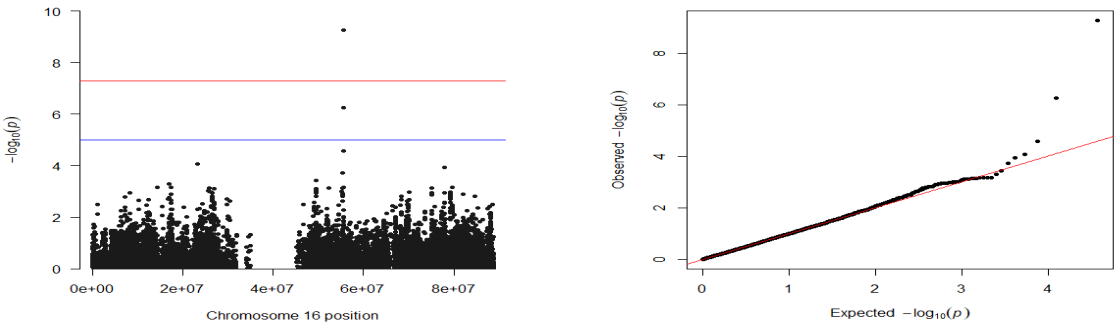
— S'assurer de la non contamination des données en analysant la distribution du taux d'hétérozygotie et en éliminant les individus dont le taux d'hétérozygotie dévie de plus de **trois écarts types de la moyenne**.

L'ensemble des filtres ont réduit la quantité de SNPs initiale de **40%**.

Résultats

Les signaux issus des tests d'association et leur p-value respective sont recensés dans le tableau 1. Trois SNPs présents sur le chromosome 16 ressortent très sigificativment. Le SNP rs1532625 est systématiquement apparu en sortie de toutes les analyses et deux autres SNPs rs247617, rs9989419 sont apparus après analyses pour lesquelles aucun seuil de LD n'a été appliqué. Ces variants se situent sur le gène de la protéine de transfert des esters de cholestérol, abrégée en CETP.

Les trois signaux significatifs sont représentés au niveau du genome-wide plot (manhattan plot)[2] et du QQ-plot présentés sur les figure 1a et figure 1b.



(a) Manhattan plot (b) QQ-plot

Figure 1: Graphiques issus de l'étude d'association sur Chromosome 16 sans ajout de covariables et sans contrainte de LD

Seuil de LD	Taux d'apolipoprotéine	Taux d'HDL
0,2,	rs1532625 : p-value = 5,442e-9	rs1532625 : p-value = 2,150 e-16
Sans LD	rs247617 : p-value=5,102e-12	rs247617 : p-value=1.983e-20 rs1532625 : p-value=2.547e-16 rs9989419 : p-value=1,540e-9

Table 1: P-Values des signaux de SNPs significatifs pour deux phénotypes

Conclusion

Les travaux réalisés ont permis de constater l'implication du chromosome 16 dans la maladie de l'artherosclérose. Les variants se situent au niveau d'un gène dont les mutations sont liées au risque cardiovasculaire. Nous avons pu constater qu'en augmentant le seuil de LD appliqué pour éliminer les SNPs trop corrélés nous obtenons moins de signaux significatifs. Cette stratégie semble appropriée lorsque les donées contiennent un nombre de SNPs plus grand que le notre qui est de l'ordre de 60 000 par chromosome. Des expérimentations approfondies telle qu'une étape d'imputation suivie d'une métanalyse seraient nécessaires pour déterminer précisément les variants causaux et les mécanismes biologiques en jeu

References

[1] Andries T. Marees Hilde de Kluiver. *ECE 100*[online]. A tutorial on conducting genome wide association studies: Quality control and statistical analysis, 2017 [cited Decembre 2018]. Available from World Wide Web: (https://onlinelibrary.wiley.com/doi/epdf/10.1002/mpr.1608).

[2] Cathy C. Laurie, Kimberly F. Doheny. from Genet Epidemiol. 2010 September; 34(6): 591–602. doi:10.1002/gepi.20516