Genome-wide association study on coronary artery disease

Fahim Beck

1 Introduction and background

Coronary artery disease (CAD) is one of the major causes of death worldwide. It causes a reduction in blood flow in the arteries of the heart through plaque formation (arteriosclerosis). There are many risk factors (smoking, alcohol, high blood cholesterol, obesity, etc). However, between 40% and 60% of this disease seems to be hereditary (McPherson and Tybjaerg-Hansen, 2016). Hence the interest in carrying out a genome-wide association study for this disease.

Research questions and approaches

The aim of this study is to identify associations among SNPs and the presence of CAD. For a GWAS, the four main steps are: (i) data pre-processing; (ii) new data generation; (iii) statistical analysis; and (iv) post-analytic interrogation.

Dataset

The data are from a GWA study of CAD (PennCATH) of the University of Pennsylvania Medical Center. It includes 3850 individuals enrolled between July 1998 and March 2003. Here, we consider anonymised data that includes 1401 individuals with genotype information over 861,473 SNPs. The clinical data gives information about the age, sex, HDL and LDL cholesterol, triglycerides and CAD status.

2 GWA analysis

2.1 PCA

2.2 Pre-processing / QC steps

2.3 Association / post-association analysis

3 Conclusion

4 References

[1] R. McPherson and A. Tybjaerg-Hansen. Genetics of coronary artery disease. *Circulation Research*, 118(4):564–578, February 2016. doi: 10.1161/CIRCRESAHA.115.306566.

[2] E. Reed, S. Nunez, D. Kulp, J. Qian, M. Reilly, and A. Foulkes. A guide to genome-wide association analysis and post-analytic interrogation. *Statistics in medicine*, 34, September 2015. doi: 10.1002/sim.6605.