

BIostatistical ASPECTS OF GENOME-WIDE ASSOCIATION STUDIES

Andreas Ziegler¹, Inke R. König¹, John R. Thompson²

*1 Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Ratzeburger Allee
160, 23538 Lübeck, Germany*

2 Department of Health Sciences, University of Leicester, Leicester, UK

To search the entire human genome for association is a novel and promising approach to unravelling the genetic basis of complex genetic diseases. In these genome-wide association studies (GWAs), several hundreds of thousands of single nucleotide polymorphisms (SNPs) are analyzed at the same time, posing substantial biostatistical and computational challenges (Ziegler et al., 2008). In this paper, we discuss a number of biostatistical aspects of GWAs in detail. We specifically consider quality control issues and show that signal intensity plots are a sine qua condition non in today's GWAs. Approaches to detect and adjust for population stratification are briefly examined. We discuss different strategies aimed at tackling the problem of multiple testing, including adjustment of p-values, the false positive report probability and the false discovery rate. Another aspect of GWAs requiring special attention is the search for gene-gene and gene-environment interactions. We describe multistage approaches to GWAs. We illustrate the approaches by using a recently published GWA on coronary artery disease (Samani et al. 2007).

References

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., et al (2007). Genome-wide association analysis of coronary artery disease. *N Engl J Med* 357, 443-453.

Ziegler, A., König, I.R., Thompson, J.R. (2008). Biostatistical aspects of genome-wide association studies. *Biom J* (in press).