Use of R in Genome-wide Association Studies

Jing Hua Zhao[1*] , Qihua Tan[2]

1. MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK
2. Odense University Hospital, Denmark

Keywords: Bioinformatics, Biostatistics, Genome-wide Association Studies, Complex Traits

Recent GeneChip and sequencing technologies have made it possible to use ~1 million or more single-nucleotide polymorphisms (SNPs) in large-scale genetic epidemiological studies. They are the most common genetic variants in human genome and their association with complex traits in relation to the environment is the subject of genome-wide association studies (GWASs), through which important variants have been successfully identified for complex traits ranging from anthropometric measurements, etiology and progression of common diseases, drug response to diversity and evolution of human populations. However, there is still a considerable scope for advancing these initiatives.

In this presentation, we provide an overview of the background and issues in design and analysis for such studies, as well as their connection with international collaborative projects and consortium work. We give real examples to illustrate how the R statistical and programming environment has been used, and discuss the extent to which this could be further developed. We believe that GWAS makes a strong case of being a motivation and inspiration for development of analytical and computational tools and that it also facilitates a vigorous inter-disciplinary collaboration between researchers in substantive areas such as biology, genetics, mathematical statistics and computing.

References

Altschuler D, MJ Daly, ES Lander (2008). Genetic mapping in human disease. *Science*, 322, 881–888.

Donnelly P. (2008). Progress and challenges in genome-wide association studies in humans. *Nature*, 456, 728–731.

Elston RC, M Ann Spence. (2006). Advances in statistical human genetics over the last 25 years. *Statistics in Medicine*, 25, 3049–3080.

Pearson TA, TA Manolio. (2008). How to interpret a genome-wide association studies. *JAMA*, 299, 1335–1344.

Zhao JH, Q Tan (2006). Integrated analysis of genetic data with R. *Human Genomics*, 2, 258-265.