

Genome-wide association studies - first experiences with study planning, data management and statistical analysis

Wichmann HE¹, Gieger C¹, Heid I¹, Illig T¹, Meitinger T²

¹*Institute of Epidemiology, GSF, Neuherberg*

²*Institute of Human Genetics, GSF, Neuherberg*

Association studies are a powerful approach in genetic epidemiology. Over the past years, both the technical and statistical foundations have been laid for whole-genome association analyses. Recent progress has made SNP-based genome screens an option. Several screens based on Affymetrix and Illumina DNA chips have been performed or are underway. Stepwise study designs have become an important element of genome-wide screens. Recent genome-wide association studies have provided proof of principle and yielded numerous loci showing a strong association with disease and disease-related phenotypes. There is broad consensus that time is ripe for launching genome-wide association studies. Although the KORA cohort is a random sample from the general population, it has been demonstrated repeatedly that the phenotypes available for this population provide an excellent basis for genetic research. So far, KORA samples have been used in more than 40 genetic case control studies as population-based controls. New genes could be identified or replicated in candidate gene studies for diverse phenotypes such as obesity and atopic eczema.

Many quantitative traits available for the KORA cohort lend themselves to genome-wide analyses. We could show the feasibility to detect association signals in candidate genes for quantitative traits such as prolonged QT and BMI in candidate gene approaches and genome-wide analysis for these traits using Affymetrix 100K arrays.

We just have started the KORA 500K Chip Project together with several partners. In this project, 1,500 subjects of the KORA survey S3 which also have participated in the follow-up investigation F3 10 years later, will be genotyped by Affymetrix 500 K chips. These 500,000 genotypes are the basis for genome wide association studies with respect to several endpoints such as Type 2 Diabetes, Metabolic Syndrome, hypertension, body mass index (BMI), QT interval, left ventricular hypertrophy (LVH), inflammatory parameters, lipids and nicotine addiction. In the first step, the 1,500 subjects of the KORA S3/F3 cohort will be genotyped and the resulting SNPs and haplotypes will be compared with the selected phenotypes. Based on strength of association and knowledge about potential pathways, SNPs will be chosen for a second stage analysis and genotyped in all 4400 persons of KORA S3. SNPs still showing strong association with the phenotype(s) of interest will finally be analysed in well characterized replication samples.

The genotypes will be stored in the BC Gene database. Statistical analysis will start with central screening of all 500,000 SNPs on dichotomous phenotypes using chi-square test. In the further steps logistic and linear regression will be used to analyse the quantitative and qualitative traits adjusting for environmental confounders.