Optimizing parameters for a cardiovascular genomewide association study

Kulle B[1], Wojnowski L[2]
[1]*Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway*
[2]*Department of Pharmacology, Johannes Gutenberg University Mainz, Germany*
wojnowski@uni-mainz.de

Practical experiences with SNP-based genomewide association studies remain limited. We have a long-standing interest in the genetic basis of cardiotoxicity which manifests in some cancer patients following therapies with anthracyclines. Using a gene-candidate approach, we recently identified associations between this phenotype and variants in genes encoding subunits of the NAD(P)H oxidase and anthracycline transporters. In order to prepare a genomewide association study for this phenotype, we performed a number of investigations using a limited number of DNA samples as well as simulations employing the HapMap data. First, a total of 104 DNA samples were hybridized to Affymetrix chips, which contained ~11,000 SNPs distributed throughout the human genome. The resulting genomewide scans were applied to distinguish between carriers and noncarriers of 37 test variants, used as surrogates for monogenic disease traits. The test variants were not contained in the chip and had been determined by other methods. Without adjustment for multiple testing, the procedure detected 24% of the test variants, but the positive predictive value was low (2%). Adjustment for multiple testing eliminated most false-positive associations, but the share of true positive associations decreased to 10-12%. Altogether, the outcome was affected by allelic frequencies of chip SNPs, by the ratio between simulated "cases" and "controls," and by the degree of linkage disequilibrium. Furthermore, the data suggested that major improvement could be achieved by raising the density of the SNP array. Therefore, we recently performed association calculations between simulated disease traits and increasing SNP densities (up to 1 million). The results show that the probability to detect a disease locus in this kind of genomewide scan can be as high as 0.95), but at the expense of a large number of false positive associations (between 60 and 90%). We will discuss strategies to optimize the design of association studies suggested by these results.