Lars Beckmann
Dr. sc. hum.

New haplotype sharing and haplotype assignment methods for mapping genes of complex diseases

Geboren am 11.12.1967 in Melle, Niedersachsen
Diplom der Fachrichtung Mathematik am 10.06.1997 an der Universität zu Köln

Promotionsfach: DKFZ (Deutsches Krebsforschungszentrum)
Doktormutter: Priv.-Doz. Dr. sc. hum. J. Chang-Claude

The major focus of genetic epidemiology is to identify and characterize genes that are involved in the etiology and pathogenesis of common, complex diseases. The findings of relevant genes will lead to progress in prevention on the population level as well as on the individual level, and to improvement in diagnosis and therapy.

The overall topic of the thesis is the exploration and further development of haplotype sharing methods to map genes involved in the etiology of a complex disease. The potential value of haplotypes has attracted widespread interest in the mapping of complex traits. Haplotype sharing methods take the linkage disequilibrium information between multiple markers into account, and have more power to detect predisposing genes, compared with single point linkage or association methods.

The aims of the thesis are threefold. First, the haplotype sharing statistics (HSS) is investigated in the context of complex diseases using simulated data under varying conditions such as different sample sizes, marker characteristics, haplotype assignment, and population characteristics. Secondly, a new approach, which is based on the Mantel statistics for space-time clustering is developed in order to improve the power of haplotype sharing analysis. The new statistic correlate genetic similarity and phenotypic similarity for case-only and case-control studies, and is extended to incorporate

covariates into the analysis. Thirdly, a new approach for haplotype assignment in pedigrees is proposed.

HSS was shown to have higher power compared to the single association method transmission/disequilibrium test (TDT) to map a major gene involved in a complex disease using simulated data from the Genetic Analysis Workshop 12. A major gene was found in a moderate sample from an isolate population in a genome scan approach using highly polymorphic microsatellites. Subsequent haplotype sharing analysis of a candidate gene was performed for a large sample of more than 600 case-parent trios. HSS did not appear to be appropriate for identifying the disease-causing variant when only SNPs within a candidate gene are investigated, and true haplotypes are not available. A TDT performed better than HSS to identify the disease-causing variant within a candidate gene. The simulation results suggest that the power of HSS may be improved with respect to localization of a candidate region or of a causal variant when there is sufficient haplotype decay and when haplotypes can be unambiguously determined.

The new Mantel statistics approach was considered in two scenarios for mapping genes involved in complex diseases. In the first situation involving genes with major effects, the new approach had more power than compared to HSS. However, both methods take only into account the information provided by the haplotypes and the affection status.

In the second situation the Mantel statistics account for the joint effects of genes and non-genetic factors, by considering additional information provided by covariates. These statistics are proposed to map genes, which contribute in interaction with non-genetic risk factors to disease risk. The analysis using simulated data shows that incorporating covariates improves haplotype sharing analysis to localize susceptibility genes.

Every haplotype-based method requires haplotypes that are constructed from genotype data. A 6-rule algorithm to reconstruct haplotype configurations in pedigrees is presented. The algorithm allows exhaustive and fast search of all possible haplotype configurations under the criterion that there are minimum recombinants between

markers. It applies to various pedigree structures with and without consanguinity relationship and allows missing alleles to be imputed during the haplotyping process. Haplotyping examples using both published and simulated data sets showed that the proposed algorithm reconstructs unique or a small number of haplotype configurations, and that one of the configurations is likely to be the true one. In conclusion, this thesis supports the recently evolved high interest in sophisticated haplotype sharing methods: they provide greater power than conventional methods for detecting disease predisposing genes in complex diseases.