Vivien Andrea Rothe, geb. Marquard
Dr. sc. hum.

Mantel Statistic Using Haplotype Sharing in Population-Based Association Studies: Properties and Further Development

Geboren am 20.4.1981 in Hamburg
Diplom in der Fachrichtung Statistik am 9.5.2005 an der Universität Dortmund

Promotionsfach: DKFZ
Doktormutter: Frau Prof. Dr. sc. hum. J. Chang-Claude

In this thesis, we investigate several properties of the Mantel Statistic Using Haplotype Sharing (MSUHS) and examine possible extensions to improve power. The MSUHS is a flexible haplotype-based method, usually applied to candidate genes or candidate regions in population-based case-control association studies. It is used to identify the association of a putative disease locus with the disease under study, by incorporating the information of neighboring markers via the haplotype-sharing approach.
Here, three major goals are of particular interest. The first concerns the type I error and the power in the presence of genotyping errors. The second and third goal concern the improvement of the MSUHS by using different measures of genetic similarity or by combining the MSUHS with the newly developed haplotype reconstruction algorithm SHARE.
Although it is known that genotype errors influence the performance of family-based association methods and the haplotype frequency estimation, so far, few studies have investigated the impact on haplotype-based methods.
Our aim is to study the effect of differential and non-differential genotyping errors on the power and type I error of the MSUHS. Therefore, we conduct a case-control simulation study, where genotyping errors are incorporated following two different misclassification models with varying error rates. We compare the performance of the MSUHS with that of the single-point Armitage trend test and that of the haplotype-based score test. The performance of all three test statistics is found to depend on the error rate and whether differential or non-differential genotyping errors are present. We show that in the presence of a realistic amount of genotyping errors (with a mean error rate per locus of 0.2 to 0.5%, all three examined association test statistics perform well.
The MSUHS correlates the phenotypic similarity of any two individuals with their corresponding genetic similarity. The task that arise is the definition of mathematical measures for the similarity between individuals. So far, the most common measure for the genetic similarity has been the number of intervals surrounding a marker x, which are identical by state.
We investigate different similarity measures that take the physical distance in bp or the genetic distance in cM or LDUs into account. Furthermore, we investigate a measure that gives more weights on the sharing of rare alleles than on common ones. In a case-control simulation study based on data provided by the Genetic Analysis Workshop 15, we incorporate nine different similarity measures and find no power gain compared with the most common measure. But the application of Yu's measure may improve the fine-mapping properties.
Furthermore, the MSUHS requires the phase information of individuals. But molecular haplotyping methods are still labor-intensive and costly, so that haplotype reconstruction algorithms are generally needed to infer the haplotype pairs of individuals. We propose a new algorithm, SHARE, which is a two-staged algorithm that performs reconstruction in unrelated individuals in the population and allows for multiple phylogenies by clustering of haplotypes. We

compare SHARE with the currently most common method fastPHASE, and find that SHARE is as good as fastPHASE regarding the quality of haplotype estimation, even in the presence of missing data, as well as in time and computational feasibility.
Additionally, our aim is to evaluate alternatives for the MSUHS by incorporating additional cluster information given by SHARE. We propose four different extensions of the MSUHS using an additional covariate that describes the similarity of two clusters containing the individuals' haplotype pairs. In case-control simulation scenarios, we compare the type I error and the power of the proposed extensions with that of the original MSUHS. The results show that all investigated extensions hold the nominal significance level of 0.05 and yield high power, even in the presence of allelic heterogeneity.
In summary, we are able to show, that the Mantel Statistic Using Haplotype Sharing is a reasonable test statistic for population-based case-control association studies. In the presence of a realistic amount of genotyping errors, it holds the nominal significance level and achieves high power of about 90% to detect the putative disease locus. Furthermore, we show that the MSUHS is a very flexible approach, that can be extended in a straightforward way to account for allelic heterogeneity, either by using a different genetic similarity measure or by incorporation of an additional cluster similarity measure.