

Hidden Markov models for admixture mapping

Chun Zhang
University of California, Davis

Abstract:

Some complex human genetic diseases have different risks in different populations (e.g. Lupus). The genes associated with these diseases can be mapped by using admixed populations from the two founding populations, such as the African-American population. This gene mapping method is called admixture mapping. It requires recent admixture and a set of marker loci with large allele frequency differences between the two founding populations. By utilizing admixture linkage disequilibrium, admixture mapping is potentially more powerful than association study and family data based linkage analysis. This method tests for linkage by detecting association of the allele ancestry with the disease, but the difficulty is that the marker data is not fully informative for allele ancestry indicators.

We will demonstrate two designs: case-control design and case-only design. For case-control design, we will describe the hidden Markov models for haplotype and genotype data, and simulation studies for evaluating parameter estimation and examining how the factors of interest affect the power of admixture mapping. A likelihood ratio test for linkage is used for both models. For haplotype data, we develop an iterative two-step procedure for estimating the population-specific allele frequencies. While for genotype data, we propose plugging in the allele frequencies in the founding populations directly and using simulation studies to examine if our method is sensitive to the misspecification of population-specific allele frequencies. For case-only design, we propose a nonparametric test of excess area to detect linkage. We utilize the hidden Markov models for haplotype and genotype data and estimate the log odds of ancestry across the marker loci by posterior decoding algorithm. The test of excess area is based on the log odds estimates from the hidden Markov models. Results of the simulation studies will be presented and conclusions based on simulation results will be reached. We will further discuss assumptions and limitations of our models and methods as well as future work.