A Novel Method Combining Linkage Disequilibrium Information and Imputed Functional Knowledge Selection

Ryan Rochat

Mentors:
C. Charles Gu, Ph.D.
and
Victor G. Dávila-Román, M.D.

Abstract

The application of high density SNP maps in genetic studies has introduced problems of prohibitive genotyping costs and inflated false discovery rates. To attenuate this effect, specific SNPs are often pre-selected for genotyping. In general, there are two ways to reduce the dimensionality of data and facilitate selection. The first is to focus on the biological property inherent to each SNP (Zhao et al. Human Mutation and select those that are likely to bear functionality relevant to the disease of interest. The second approach focuses on the underlying data structure of linkage disequilibrium (LD) among SNPs and selects by factor analysis a representative subset of SNPs et American Journal of Human Genetics. The first method ignores the data structure between SNPs, whereas the second method neglects the biological significance of the SNPs in question. The information considered and neglected by these two methods makes them complimentary. Therefore, in an attempt to enhance the selection of we propose a method combining the information obtained from each of these methods in a weighted factor analysis model. By considering both the functional-and information-based approaches, it is possible to design an amalgamated procedure that can more efficiently and accurately select a subgroup of for use in genetic studies.