Genome Medicine

METHOD Open Access

Rule-based induction method for haplotype comparison and identification of candidate disease loci

Sirkku Karinen[1†], Silva Saarinen[2†], Rainer Lehtonen[2,3], Pasi Rastas[3,4], Pia Vahteristo[2], Lauri A Aaltonen[2] and Sampsa Hautaniemi[1*]

Abstract

There is a need for methods that are able to identify rare variants that cause low or moderate penetrance disease susceptibility. To answer this need, we introduce a rule-based haplotype comparison method, Haplous, which identifies haplotypes within multiple samples from phased genotype data and compares them within and between sample groups. We demonstrate that Haplous is able to accurately identify haplotypes that are identical by descent, exclude common haplotypes in the studied population and select rare haplotypes from the data. Our analysis of three families with multiple individuals affected by lymphoma identified several interesting haplotypes shared by distantly related patients.

Background

One of the most important goals in biomedical research is to identify genes that predispose humans to diseases, such as cancer. To facilitate the identification of these genes, a number of genome-wide approaches have been suggested, such as genetic linkage and genome-wide association (GWA) methods [1]. The linkage methods have revealed several high penetrance disease susceptibility loci [1], whereas GWA studies have been useful in the 'common disease - common allele' model [2]. However, neither approach is well suited to tackle moderate penetrance susceptibility because such a condition rarely results in large pedigrees, with few or no phenocopies, convenient for linkage analysis. As the GWA approaches cannot detect these presumably rare alleles, there is clearly a need for methods that are able to identify loci where such variants could be located. Evolutionarily recent, and thus rare, mutations are usually conveyed in a pedigree by a shared haplotype. Therefore, detection of such haplotypes can lead to the identification of rare or moderate penetrance variants behind disease susceptibility.

We introduce here Haplous, a novel computational approach that uses phased genotype data, such as genome-wide SNPs, to identify and prioritize genomic regions likely to be inherited from a common ancestor. The central idea of our approach is to use haplotypes, instead of single alleles, and rank them based on expert-defined rules that determine the haplotypes shared in heterozygous and homozygous forms. As the identification of haplotypes has been recognized as useful for revealing disease predisposing genes, several haplotype association methods have been developed [3-10]. These methods include detection of haplotype diversity and statistical association tests. Haplotypes can be detected with fixed or variable length sliding window [7,11,12], haplotype blocks [13], haplotype clustering [9], a cladistic approach [10] or considering non-contiguous haplotypes [5,8]. Some haplotype analysis methods are feasible in genome-wide settings [9,11,13], although several are intended only for smaller datasets [6,12]. Many haplotype-based approaches essentially aim to identify identical by descent (IBD) regions between samples [14-16].

To our knowledge, Haplous is the first method that uses a rule-based approach to identify rare haplotypes shared by multiple individuals from genome-wide data.

* Correspondence: sampsa.hautaniemi@helsinki.fi
† Contributed equally
[1]Research Programs Unit, Genome-Scale Biology, and Institute of Biomedicine, Biochemistry and Developmental Biology, University of Helsinki, Haartmaninkatu 8, Helsinki, FIN-00014, Finland
Full list of author information is available at the end of the article