

# Multipoint Fine-scale Linkage Disequilibrium Mapping: Importance of Modeling Background LD

*Andrew L. Strahs and Mary Sara McPeck*

## Abstract

In linkage disequilibrium (LD) mapping, use of information on multiple markers simultaneously is expected to lead to greater power to detect association and smaller confidence intervals (CIs) for the location of the variant of interest than would be obtained from single-point analysis. Among the important challenges facing case-control LD mapping methods are (i) even when an appropriate control sample is available, there may be background LD in the control sample which must be taken into account in the analysis, especially when fine-scale data are collected, and (ii) in practice, genotype rather than haplotype data are often available, limiting the applicability of some methods. Furthermore, in cases when genotype data can, in principle, be incorporated, it can be computationally challenging. We focus on simultaneous solution of these problems in the context of the Decay of Haplotype Sharing (DHS) method. We develop a computationally efficient method that allows for genotype or haplotype data on many loci and incorporates background LD based on a Markov model of order  $\eta$ . The case of a Markov model of order 2 is implemented in free software. In addition, we demonstrate that failure to adequately model background LD can potentially have a major effect on the analysis, and we develop and apply methods for assessing the adequacy of the model for background LD.

**Keywords:** Decay of Haplotype Sharing; linkage disequilibrium; fine-scale mapping; background linkage disequilibrium; cystic fibrosis; hidden Markov model

## 1 Introduction

Linkage disequilibrium (LD) has been shown to be useful for fine-mapping of trait-associated variants [6, 10, 11, 15]. While early approaches generally treated each marker separately, haplotype-based LD mapping methods have the potential to provide considerable additional information when dense marker data are available in a region. There are several approaches that combine results across loci in various ways without explicitly modeling dependence among loci [4, 7, 17, 23, 31, 32]. Among approaches that explicitly model dependence across loci, Service *et al.* [29] and MacLean *et al.*