A Novel Approach to Use Sequence Data for Linkage Analysis

Gao T. Wang1, Di Zhang1, Biao Li1, Hang Dai1, Suzanne M. Leal1, \*

1Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

\*Correspondence:

Suzanne M. Leal

Department of Molecular and Human Genetics

Baylor College of Medicine

One Baylor Plaza, 700D

Houston, TX 77030

713-798-4011

[sleal@bcm.edu](mailto:sleal@bcm.edu)**Abstract**

Traditionally, linkage analysis has been the main approach to elucidate causes of Mendelian disorders in families with multiple affected individuals. Recent advances in next generation sequencing (NGS) technology has made it standard approach to prioritize and screen for causal mutations in whole-genome or whole-exome sequenced individuals suffering from Mendelian diseases. Combined linkage and sequence analysis is gaining popularity in the human genetics community, yet there are few applications of directly performing linkage analysis using sequence data. Inspired by the “aggregation analysis” commonly practiced in rare variants complex disease association studies, we have developed a collapsed haplotype pattern method to generate markers from sequence data for linkage analysis. We demonstrate with a range of simulation studies of two-point linkage analysis that our method is substantially more powerful over linkage analysis using single nucleotide variants. We developed the XXX software package that uses the method described here to perform two-point linkage analysis on sequence data. Additionally XXX can output marker data in formats compatible with XXX, XXX, and XXX software, reviving these linkage analysis tools for use in NGS era. To illustrate the possibilities we examined a data set … and detected linkage of XXX (disease) to XXX (gene). **Introduction**

Long version (4 paragraphs) 1. Review of linkage methods, 2. Review of variants prioritization and screening, 3. The motivation to use sequence data for linkage analysis and 4. Introduce our method.

Short version (2 paragraphs) 1. The motivation to use sequence data for linkage analysis and 2. Introduce our method.

**Methods**

*Linkage analysis with Collapsed Haplotype Markers*

Outline. The idea of haplotype pattern: definition of units, determination of genetic distance.

Mendelian error checking, genetic haplotyping, missing data imputation.

Collapsed haplotype pattern method. coding themes, compute combined MAF, handling of recombination events, handling of missing data.

*Implementation*

*Simulation studies*

*Analysis of xxx dataset*

**Results**

*Simulation studies*

*Analysis of xxx dataset*

**Discussion**

The use of MAF reference. The use of genetic map distance. Uncertainty in genetic haplotyping

**Web Resources**

**Acknowledgements**

We would like to thank … This work is supported by National Institute of Health grants …

**References**

**Tables**

**Figure legends**