

## A novel method for imputing summary statistics of untyped SNP using the summary information from the neighboring SNPs

**Yadunath Gautam**  
Central Michigan University, USA

The rapid evolution of genotyping technologies have made it possible to identify regions of the genome harboring genetic variation that predispose to diseases. Given the correlation between genotypes at loci close together, called linkage disequilibrium (LD), it is possible to impute unobserved genotypes. However, genotype imputation is computationally costly (time, shortage). Imputation based on sufficient statistics has the potential to greatly reduce this computational burden. We developed a method of estimating the haplotype frequencies of a set of three genetic markers (SNPs) when one SNP is not observed in the sample. These estimates of haplotype frequencies for case and control data can be used for downstream analysis such as allelic tests of association. We propose new multilocus LD measures defined using Fisher Information matrix and modified optimality criteria from optimal design theory. From a sample of multiple competing SNPs, we select two flanking SNPs with the highest multilocus LD measure with the missing SNP. The multilocus measures are estimated from a phased reference dataset. Different multilocus measures can result in different pairs of SNPs. We compare the accuracy of the imputed test statistics in the case-control study design using the summary data for the pair of flanking SNPs selected from the proposed measures and existing measures. We find that the LD measure based on the mutual information statistic identifies the pair that gives better imputation results than existing measures.

### Biography

Yadunath Gautam is a Ph.D. student in the Department of Mathematics, Central Michigan University. He is the presenter and corresponding author.

[gauta1yn@cmich.edu](mailto:gauta1yn@cmich.edu)