

# Genotype imputation accuracy with different reference panels in admixed populations

Guan-Hua Huang\* (黃冠華) and Yi-Chi Tseng  
Institute of Statistics, National Chiao Tung University

## Abstract

Genome-wide association studies have successfully identified common variants that are associated with complex diseases. However, the majority of genetic variants contributing to disease susceptibility are yet to be discovered. It is now widely believed that multiple rare variants are likely to be associated with complex diseases. Using custom-made chips or next-generation sequencing to uncover rare variants' effects on the disease can be very expensive in current technology. Many researchers thus turn to use the genotype imputation approach to predict the genotypes at these rare variants that are not directly genotyped in the study sample. One important question in genotype imputation is how to choose a reference panel that will produce high imputation accuracy in a population of interest. Using whole genome sequence data from the Genetic Analysis Workshop 18 data set, this report compares genotype imputation accuracy among reference panels representing different degrees of genetic similarity to a study sample of admixed Mexican Americans. Results show that a reference panel that closely matches the ancestry of the study population can increase imputation accuracy, but it can also result more missing genotype calls. Having a reference panel with larger size can reduce imputation error and missing genotype, but the improvement can be limited. We also find that, for the admixed study sample, the simple selection of a single best reference panel among HapMap African, European or Asian population is not appropriate. The composite reference panel combining all available reference data should be used.