

R. Clifton Bailey Statistics Seminar Series

**A pooling strategy to effectively
use genotype data in quantitative
traits genome-wide association
studies**

Wei Zhang

National Institutes of Health

Johnson Center G19 – Gold Room
[4400 University Drive, Fairfax, VA 22030](https://www.nih.gov/locations/fairfax)

May 4, 2018

11:00 A.M. - 12:00 Noon

Abstract: The goal of quantitative traits genome-wide association studies (GWAS) is to identify associations between a phenotypic variable, such as a vitamin level, and genetic variants, often single-nucleotide polymorphism (SNPs). When funding limits the number of assays that can be performed to measure the level of the phenotypic variable, a subgroup of subjects is often randomly selected from the genotype database and the level of the phenotypic variable is then measured for each subject. Because only a proportion of the genotype data can be used, such a single random sampling method may suffer from

substantial loss of efficiency, especially when the number of assays is relative small and the frequency of the less common variant (minor allele frequency) is low. We propose a pooling strategy in which subjects from the remaining study subjects to form independent pools; blood samples from subjects in each pool are mixed; and the level of the phenotypic variable is measured for each pool. We demonstrate that the proposed pooling approach produces considerable gains in efficiency over the simple random sampling method for inference concerning the phenotype-genotype association, resulting in higher precision and power. The methods are illustrated using genotypic and phenotypic data from the Trinity Students Study, a quantitative GWAS.