RUTGERS
School of Arts and Sciences

Department of Statistics and Biostatistics
Rutgers, The State University of New Jersey
Busch Campus
501 Hill Center
101 Frelinghuysen Road
Piscataway, NJ 08854

Web: stat.rutgers.edu
Email: office@stat.rutgers.edu
Phone: 848-445-2690
Fax: 848-445-3428

RUTGERS UNIVERSITY
DEPARTMENT OF STATISTICS AND BIOSTATISTICS
CENTER FOR INTEGRATIVE PROTEOMICS RESEARCH
www.stat.rutgers.edu

Seminar

Speaker: **Prof. Yijuan Hu**
Department of Biostatistics and Bioinformatics
Rollins School of Public Health, Emory University

Title: **Integrative analysis of sequencing and array genotype data for discovering disease associations with rare mutations**

Time: **3:20pm – 4:20pm, Wednesday, February 11, 2015**

Place: **552 Hill Center**

Abstract

In the large cohorts that have been used for genome-wide association studies (GWAS), it is prohibitively expensive to sequence all cohort members. A cost-effective strategy is to sequence subjects with extreme values of quantitative traits or those with specific diseases. By imputing the sequencing data from the GWAS data for the cohort members who are not selected for sequencing, one can dramatically increase the number of subjects with information on rare variants. However, ignoring the uncertainties of imputed rare variants in downstream association analysis will inflate the type I error when sequenced subjects are not a random subset of the GWAS subjects. In this article, we provide a valid and efficient approach to combining observed and imputed data on rare variants. We consider all commonly used gene-level association tests, all of which are based on the score statistic for assessing the effects of individual variants on the trait of interest. We show that the score statistic based on the observed genotypes for sequenced subjects and the imputed genotypes for non-sequenced subjects is unbiased. We derive a robust variance estimator that reflects the true variability of the score statistic regardless of the sampling scheme and imputation quality, such that the corresponding association tests always have correct type I error. We demonstrate through extensive simulation studies that the proposed tests are substantially more powerful than the use of accurately imputed variants only and the use of sequencing data alone. We provide an application to the Women's Health Initiative (WHI). The relevant software is freely available.

***** Refreshments will be served @3:00pm in Room 502 Hill Center***