CONTROL ID: 3469554

PRESENTER: Seunggeun Lee

PRESENTER (INSTITUTION ONLY): Seoul National University

TITLE: Scalable and accurate rare variant test for big biobank data analysis

ABSTRACT BODY:

Abstract Body: Large-scale biobanks have emerged as a powerful resource for complex disease studies and precision medicine. Genomic information coupled with clinical, behavioral, and environmental measurements enables to discover novel genetic associations and disease mechanism across the entire phenome. However, the scale and complex structure of biobank data have remained substantial challenges. For example, case-control imbalance, which is common in biobank data, can greatly increase type I error rates. Adjustment of sample relatedness requires a large amount of computation and memory usages. In this talk, I will introduce our new methods of rare variant tests for biobank size data. I will first introduce a robust approach that adjusts for case-control imbalance using saddlepoint approximation and efficient resampling while providing a scalable computation. And then I will discuss the mixed model gene-based test, SAIGE-GENE, which utilizes state-of-the-art optimization strategies to reduce computational and memory cost of mixed effect model, and hence is applicable to hundreds of thousands of samples. Through the analysis of UK-Biobank whole-exome sequence data of 50,000 samples and imputed genotype data of 408,910 White British samples, we show that our approaches efficiently analyze large sample data with type I error rates well controlled.

AUTHORS/INSTITUTIONS: S. Lee, Data Science, Seoul National University, Seoul, KOREA (THE REPUBLIC OF)|