



**Application number/Title:** 45227 - Scalable and Robust methods for biobank data analysis

**Applicant PI:** Dr Seunggeun Lee

**Application Institution:** University of Michigan, Ann Arbor, Michigan, USA

**Keywords provided by the Applicant PI to describe the research project:**  
gwas, gxe, phewas, rare-variant

**Application Lay Summary:**

Large biobanks such as UK-Biobank are important resources for health research. The detailed genetic information coupled with clinical, behavior and environmental measurements provide a great opportunity to discover new genetic variants associated with diseases, such as heart disease, cancer, and type 2 diabetes. In addition, the data will allow identifying genetic variants whose effects are modified by environmental exposures. These discoveries will enable more precise prediction of individual-level disease risk and help optimize disease screening, prevention, and treatment. However, the large size and complex structure of the Biobank data is a major problem for identifying genetic associations and gene-environmental interactions. This application proposes a methodological and computational development which is an important step towards obtaining the rich information of the biobank data. The developed method will be applied to UK-Biobank data to identify novel genetic variants associated with clinical phenotypes, which will contribute to the prevention and treatment of complex genetic diseases.