NAME OF PRESENTING AUTHOR: Seon-Kyenog Jang

EMAIL ADDRESS OF PRESENTING AUTHOR: jangx303@umn.edu

Estimating SNP-based heritability of complex traits in diverse ancestry: Method comparisons

Seon-Kyeong Jang[1], Saonli Basu[2], Zhatong Lin[2], Scott Vrieze[1], Trans-Omics Precision Medicine

[1] Department of Psychology, University of Minnesota, Minneapolis, MN, USA

[2] Department of Biostatistics, University of Minnesota, Minneapolis, MN, USA

KEYWORDS: Heritability, admixed, population structure, linear mixed model, HE regression

ABSTRACT:

SNP-based heritability (h^2_{SNP}) indicates the degree to which a set of genotyped variants and their tagging variants account for phenotypic variance of a given trait. It can be estimated by linear mixed model or Haseman-Elston (HE) regression using unrelated individuals. To date, the existing methods can only be applied to population with relatively little recent admixture. Extending the methods to the genomes of diverse ancestry enrich our understanding of genetic etiology of complex traits and potentially improve power by incorporating larger sample. Here we are considering three approaches to estimating h^2_{SNP} of two anthropometric traits (height and BMI) and two smoking traits (age of smoking initiation and cigarettes per day) in a sample with population substructure; 1) linear mixed model with PC adjustments (GCTA-REML), 2) HE regression with PC adjustments in a second-order estimating equation (Adj-HE), and 3) HE regression with kinship estimates adjusted for PCs (PC-relate-HE). We used TOPMed whole-genome sequences of 43,724 and >26,000 unrelated samples for anthropometric and smoking traits, respectively, of which 30% consisted of non-European ancestry. Using variants common in European, African and Hispanic admixed ancestries, h^2_{SNP} was estimated similarly across two methods (GCTA-REML and Adj-HE): about 0.40 and 0.18 (SE 0.008~0.01) for height and BMI, and 0.04 and 0.08 (SE 0.013) for age of smoking initiation and cigarettes per day. We further test the performance of all three methods in admixed populations separately and consider the influence of ancestry-specific common variants along with variance differences among cohorts.