

Missing Heritability: New statistical and algorithmic approaches

Or Zuk

Broad Institute of MIT and Harvard

The completion of the human genome project set a stepping stone in building catalogs of common human genetic variation. These catalogs, in turn, enabled the search for associations between common variants and complex human traits and diseases, by performing Genome-Wide Association Studies (GWAS). GWAS have been successful in discovering thousands of statistically significant, reproducible, genotype-phenotype associations. However, the discovered variants (genotypes) explain only a small fraction of the phenotypic variance in the population for most human traits. In contrast, the *heritability*, defined as the proportion of phenotypic variance explained by *all* genetic factors, was estimated to be much larger for those same traits using indirect population-based estimators. This gap is referred to as 'missing heritability'.

Mathematically, heritability is defined by considering a function F mapping a set of (Boolean) variables, (x_1, \dots, x_n) representing genotypes, and additional environmental or 'noise' variables ϵ , to a single (real or discrete) variable z , representing phenotype. We use the variance decomposition of F , separating the linear term, corresponding to additive (narrow-sense) heritability, and higher-order terms, representing genetic-interactions (epistasis), to explore several explanations for the 'missing heritability' mystery. We show that genetic interactions can significantly bias upwards current population-based heritability estimators, creating a false impression of 'missing heritability'. We offer a solution to this problem by providing a novel *consistent* estimator based on unrelated individuals. We also use the Wright-Fisher process from population genetic theory to develop and apply a novel power correction method for inferring the relative contributions of rare and common variants to heritability. Finally, we propose a novel algorithm for estimating the different variance components (beyond additive) of heritability from GWAS data.

I will give the needed genetics background, and discuss the statistical methods and algorithms used.