BIOSTAT 572: Intro Talk

Elisa Sheng

Biostatistics Department, University of Washington

April 10, 2012
**Title:** The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability

**Authors:** Or Zuk\textsuperscript{a}, Eliana Hechter\textsuperscript{a}, Shamil Sunyaev\textsuperscript{a,b}, Eric Lander\textsuperscript{a}

**Affiliations:** \textsuperscript{a}Broad Institute of MIT and Harvard; \textsuperscript{b}Genetics Division, Brigham and Women’s Hospital, Harvard Medical School

**Publication:** Proceedings of the National Academy of Sciences, 2012
Background

**Big Picture:** the genetic revolution in healthcare - use genotype to predict an individual’s risk for disease/traits

**GWASs:** Over 1200 discovered genetic variants, and growing

**The Mystery:** These variants appear to only explain a small portion of the variation in the disease/trait

**Explanations:**

1. haven’t found enough genetic variants yet
2. the portion has been underestimated
Phenotype, $P = \Psi(G, E)$, is a function of genotype ($G$) and environment ($E$).

Heritability, $H^2 = V_G/V_P$, is the proportion of phenotypic variation explained by genetic variants.

- **broad sense** ($H^2$): $V_G$ represents phenotypic variation due to *any* genetic contribution.
- **narrow sense** ($h^2$): $V_G$ represents phenotypic variation due to “additive” contributions of genes.

**explained**: $\pi_{\text{explained}} = \frac{h^2_{\text{known}}}{h^2_{\text{all}}}$

**missing**: $1 - \pi_{\text{explained}}$
Why use narrow sense heritability?

- Not practical to determine $H_{known}^2$
- Easy to compute $h_{known}^2$ using allele frequency ($f$) and effect size ($\beta$):

$$h_{known}^2 = \sum_i f_i(1 - f_i)\beta_i^2$$

- $h^2$ is a commonly used measure
Phantom Heritability

$h^2_{all}$ is the narrow sense heritability that incorporates all of the genetic variants (known and unknown) associated with the phenotype

$h^2_{pop}$ a quantity based on phenotypic correlations in the population, which is not a narrow sense heritability

Typically $h^2_{all}$ is assumed to equal $h^2_{pop}$, however these two quantities are rarely equal (due to genetic interactions).

**phantom heritability:** $\pi_{phantom} = 1 - \frac{h^2_{all}}{h^2_{pop}}$

$\pi_{phantom} > 0 \implies$ inflated missing heritability
Methods of Estimating $h^2_{all}$

Naive Methods (assume $h^2_{all} = h^2_{pop}$):

- the ACE model
- the ADE model
- parent-offspring regression

Better methods (that still have issues):

- Visscher et al., 2006 (confounded by genetic interactions)
- Yang et al., 2010 (inconsistent estimator)
The Brave New Method

Extends the approaches of Visscher et al. and Yang et al.

Has some nice properties:

- It is not confounded by genetic interactions
- It is a consistent estimator
- It provides a way of detecting genetic interactions (by comparison to naive estimates)

Requires that one can detect recent common ancestors between individuals in the population (segments of shared identity-by-decent above a specified threshold)
The Brave New Method

**Theorem:** \( h_{all}^2 = (1 - \kappa_0) \rho' (\kappa_0) \)

- \( \kappa_{i,j} \) is the proportion of genome shared (in large IBD segments) by individuals \( i \) and \( j \)
- \( \kappa_0 \) is the average proportion of large-segment IBD sharing in the population
- For a trait, \( Z = \Psi(G, E) \), let \( \rho(\kappa) \) is the average phenotypic correlation between individuals who share \( \kappa \) of their genomes in large IBD blocks
Summary of Topics

- Explain why commonly used estimators of explained heritability result in phantom heritability
- Develop framework for a plausible model that will illustrate the above
- Apply the model on examples (Crohn’s disease and schizophrenia) and illustrate phantom heritability
- Argue the need for an estimator of $h_{all}^2$ that does not depend on underlying genetic architecture
- Introduce their estimator of $h_{all}^2$
- Simulate data using the previously developed framework and apply their method