BIOSTAT 572: Final Presentation

Elisa Sheng

Biostatistics Department, University of Washington

May 22, 2012

The Paper

Title: The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability

Authors: Or Zuk[a], Eliana Hechter[a], Shamil Sunyaev[a,b], Eric Lander[a]

Affiliations: [a]Broad Institute of MIT and Harvard; [b]Genetics Division, Brigham and Women's Hospital, Harvard Medical School

Publication: Proceedings of the National Academy of Sciences, 2012

Narrow Sense Heritability (h^2)

- Z = phenotype
- G_i = genotype at SNP$_i$ (0,1, or 2)
- ϵ = environment

$$Z = \alpha + \sum_i \beta_i G_i + \epsilon + \text{interactions}$$

Narrow Sense Heritability:

$$h^2 = \frac{Var(\sum_i \beta_i G_i)}{Var(Z)} = \frac{Var\left(\sum_{\text{known}} \beta_i G_i + \sum_{\text{unknown}} \beta_i G_i \right)}{Var(Z)}$$

Phantom Heritability

Explained Heritability:

$$\frac{Var(\sum_{\text{known}} \beta_i G_i)}{Var(\sum_i \beta_i G_i)} \equiv \frac{h^2_{known}}{h^2_{all}}$$

Phantom Heritability:

- the explained heritiability is incorrect (usually too small) due to using a biased estimator of h^2_{all}
- particularly a problem when broad sense heritability is used to estimate h^2_{all} for a trait affected by genetic interactions

An Example of Phantom Heritability

Suppose

$$Z = \sum_i \beta_i G_i + \sum_{\mathbf{i}=(i_1,i_2)} \gamma_{\mathbf{i}} G_{i_1} G_{i_2} + \epsilon.$$

The narrow and broad sense heritabilities are then given by:

$$h^2 = \frac{Var(\sum_i \beta_i G_i)}{Var(Z)} \quad \text{and} \quad H^2 = \frac{Var(Z-\epsilon)}{Var(Z)}.$$

An Example of Phantom Heritability

- Suppose $G_i \sim Bin(2, \pi_i)$ for all i, and $\epsilon \sim N(0, 1)$.
- Let $\beta_i = 2$ and $\pi_i = 0.5$ for $i = 1, \ldots, 10$.
- Let $\gamma_i = 0.1$ for $i = 1, \ldots, \binom{10}{2}$.
- Suppose we only know 5 of the SNPs.

Using simulated genotype data, we can compute Z.

An Example of Phantom Heritability

Estimates of H^2, h^2 and h^2_{known} :

$$\hat{H}^2 = 0.99, \quad \hat{h}^2 = 0.48, \quad \text{and} \quad \hat{h}^2_{known} = 0.22$$

Explained Heritability:

$$\frac{Var(\sum_{\text{known}} \beta_i G_i)}{Var(\sum_i \beta_i G_i)} \equiv \frac{h^2_{known}}{h^2_{all}}$$

Explained heritability (using $\hat{h}^2$ for h^2_{all}) $= 0.46$

Explained heritability (using $\hat{H}^2$ for h^2_{all}) $= 0.22$

Phenotype and Relatedness

For two arbitrary individuals 1 and 2, consider the two variables:

- $Z_1 Z_2$, the product of their phenotypes, and
- $R_{1,2}$, their **relatedness**, the proportion of the genome where they share common ancestors[1].

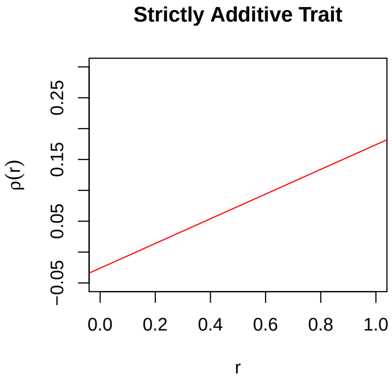
$$R_{1,2} \ \rightarrow \ \text{genotype} \ \rightarrow Z_1 Z_2$$

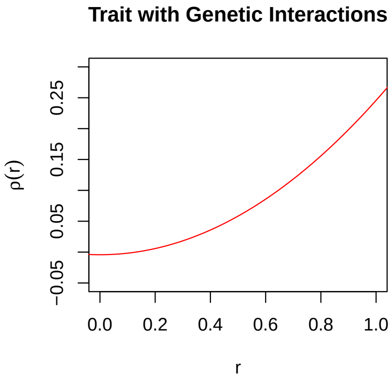
The **phenotypic correlation** between individuals 1 and 2 with relatedness $R_{1,2} = r$ is

$$\rho(r) \equiv E(Z_1 Z_2 | R_{1,2} = r).$$

[1] to use software to detect relatedness, common ancestors are within 20 generations

Phenotype and Relatedness





Phenotype and Relatedness

The Question: Is there a linear relationship between phenotypic correlation and relatedness, regardless of the existence of genetic interactions?

The Anwser: ...

A Linear Relationship?

$$Z = \alpha + \sum_i \beta_i G_i + \text{interactions} + \epsilon$$

$$\begin{aligned}
\rho(r) &\propto E\left[(\sum_i \beta_i G_{1i})(\sum_i \beta_i G_{2i})|R_{1,2} = r\right] + f(r) \\
&\propto E\left(\sum_i \beta_i^2 G_{1i} G_{2i} | R_{1,2} = r\right) + f(r) \\
&\propto \sum_i \beta_i^2 E(G_{1i} G_{2i} | R_{1,2} = r) + f(r)
\end{aligned}$$

A Linear Relationship?

For a general trait (which may include genetic interactions),

$$\rho(r) \propto \sum_i \beta_i^2 E(G_{1i} G_{2i} | R_{1,2} = r) + f(r).$$

For a strictly additive trait,

$$\rho(r) \propto \sum_i \beta_i^2 E(G_{1i} G_{2i} | R_{1,2} = r).$$

A Linear Relationship?

Assumption 1: The population is unstructured.

Assumption 2: The probability of sharing a common ancestor is constant over the genome.

These assumptions translate into the following mathematical properties:

$$E(G_{1i}G_{2i}|R_{1,2} = r_0) = E(G_{1i})E(G_{2i}) = \pi_i^2$$

$$E(G_{1i}G_{2i}|R_{1,2} = r) = rE(G_{1i}G_{2i}|R^i_{1,2} = 1) + (1-r)E(G_{1j}G_{2j}|R^i_{1,2} = 0)$$

note: r_0 is the average level of relatedness in the population

A Linear Relationship?

Claim 1: Based on the mathematical properties,

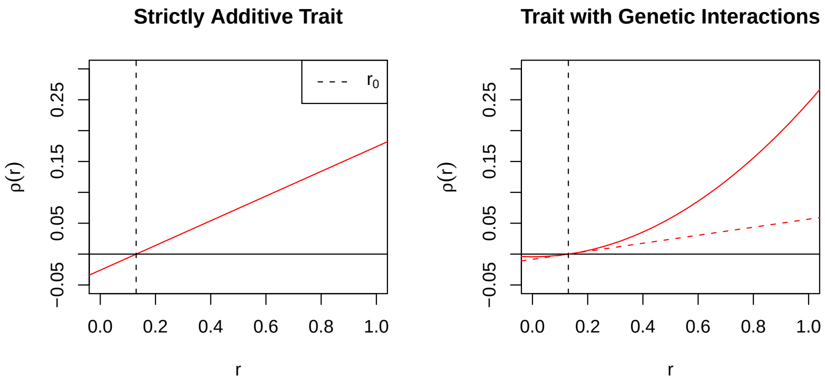
$$\sum_i \beta_i^2 E(G_{1i} G_{2i} | R_{1,2} = r) = \frac{r}{1 - \frac{r_0}{2}} h_{all}^2.$$

Claim 2: At the population average level of relatedness, r_0,

$$\rho'(r_0) = \frac{1}{1 - \frac{r_0}{2}} h_{all}^2$$

.

A Linear Relationship!



Example: An Additive Trait

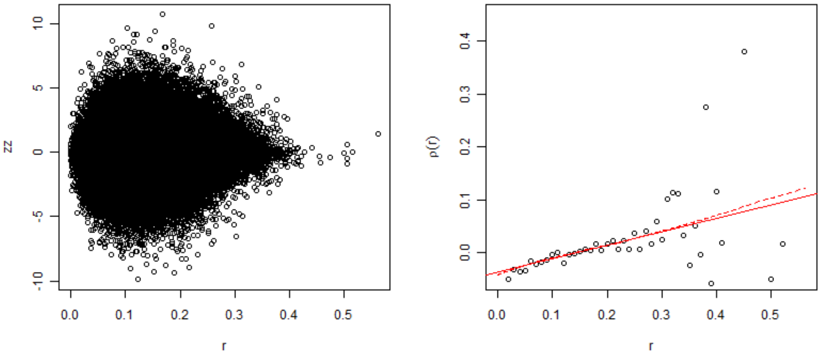
- Simulated a population of size 311,073, based on 28 founders over 23 generations
- Randomly sampled 1000 subjects
- Computations are based on 1000 SNPs
- Population average level of relatedness: $r_0 = 0.135$
- Trait is strictly additive:

$$Z = \sum_i \beta_i G_i + \epsilon$$

- $\beta_i \sim N(0, 0.022)$ and $\epsilon_i \sim N(0, 0.755)$

$$h^2_{all} = 0.245$$
$$\hat{h}^2_{all} = 0.234$$

Example: An Additive Trait



Example: A Trait with Interactions

- Trait has genetic interactions:

$$Z = \sum_i \beta_i G_i + \sum_{\mathbf{i}=(i_1,i_2)} \gamma_{\mathbf{i}} G_{i_1} G_{i_2} + \epsilon$$

- $\beta_i = 1$ and $\pi_i = 0.5$ for $i = 1, \ldots, 1000$.
- $\gamma_i = 1$ for $i = 1, \ldots, \binom{1000}{2}$.
- $\epsilon \sim N(0, 1)$.

$$h^2_{all} = 0.566$$

$$\hat{h}^2_{all} = 0.574$$

Example: A Trait with Interactions