Human Population Genomics
Heritability & Environment

Heritability of Disorders in Twin Studies

*Psychological medicine*, 41 (1), 33-40 PMID:

## Heritability & Environment

<table>
<thead>
<tr>
<th>Disease</th>
<th># of Loci</th>
<th>Heritability Explained</th>
<th>Heritability Estimated</th>
<th>Measure of Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age related macular degeneration</td>
<td>5</td>
<td>50%</td>
<td>46-71%</td>
<td>Sibling recurrent risk</td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td>32</td>
<td>20%</td>
<td>50-60%</td>
<td>Genetic risk (liability)</td>
</tr>
<tr>
<td>Systemic Lupus Erithematosus</td>
<td>6</td>
<td>15%</td>
<td>44-66%</td>
<td>Sibling recurrent risk</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>18</td>
<td>6%</td>
<td>26%</td>
<td>“</td>
</tr>
<tr>
<td>HDL Cholesterol</td>
<td>7</td>
<td>5.2%</td>
<td></td>
<td>“</td>
</tr>
<tr>
<td>Height</td>
<td>40</td>
<td>5%</td>
<td>81%</td>
<td>Phenotypic Variance</td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>4</td>
<td>1.5%</td>
<td></td>
<td>“</td>
</tr>
</tbody>
</table>

## Heritability & Environment

<table>
<thead>
<tr>
<th>Trait/Disease</th>
<th>Estimated heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcoholism</td>
<td>50-60%</td>
</tr>
<tr>
<td>Alzheimers</td>
<td>58-79%</td>
</tr>
<tr>
<td>Asthma</td>
<td>30%</td>
</tr>
<tr>
<td>Bipolar Disorder</td>
<td>70%</td>
</tr>
<tr>
<td>Depression</td>
<td>50%</td>
</tr>
<tr>
<td>Hair Curliness</td>
<td>85-95%</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>8%</td>
</tr>
<tr>
<td>Height</td>
<td>81%</td>
</tr>
<tr>
<td>Obesity</td>
<td>70%</td>
</tr>
<tr>
<td>Longetivity</td>
<td>26%</td>
</tr>
<tr>
<td>Sexual Orientation</td>
<td>60%</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>81%</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>88%</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>26%</td>
</tr>
</tbody>
</table>

Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).

Where is the missing heritability?

Some plausible explanations

- **Rare variants** not captured in genotyping microarrays
- **Many variants** of small effect
- **Structural variants** not captured in short read sequencing
- **Epistatic effects**: non-linear gene-gene interactions
- ???

Global Ancestry Inference

Modeling population haplotypes – VLMC

\[ G_1, \ldots, G_N; \quad G_i = g_{i1} \ldots g_{in}; \quad g_{ij} \in \{0, 1, 2\} \]

\[ G_i = H_{i1} + H_{i2}, \text{ where,} \]

\[ H_i = h_{ij1} \ldots h_{ijn}; \quad h_{ijk} \in \{0, 1\} \]
Phasing

**Haplotype Phasing**

Haplotypes
- ATCCGA
- AGACGC

Genotype
- A\[T\][C]CG\[C\]
- G\[A\]

- High throughput cost effective sequencing technology gives genotypes and not haplotypes.

Possible phases:
- ATACGA
- AGACGA
- AGCCGC
- ATCCGC
- ....

Browning & Browning, 2007
Identity By Descent
IBD detection

IBD = F

IBD = T

FastIBD: sample haplotypes for each individual, check for IBD

Parente

Rodriguez et al. 2013

Browning & Browning 2011
The genetics of Mexico recapitulates Native American substructure and affects biomedical traits, Moreno-Estrada et al. Science, 2014.
Fixation, Positive & Negative Selection

How can we detect negative selection?

How can we detect positive selection?
How can we detect positive selection?

**Ka/Ks ratio:**
Ratio of nonsynonymous to synonymous substitutions

Very old, persistent, strong positive selection for a protein that keeps adapting

**Examples:** immune response, spermatogenesis
How can we detect positive selection?

**Fig. 3.** Low diversity and many rare alleles at the Kell blood antigen cluster. On the basis of three different statistical tests, the 115-kb region (containing four genes) shows evidence of a selective sweep in Europeans (28).

**Fig. 4.** Excess of high-frequency derived alleles at the Duffy red cell antigen (FY) gene (34). The 10-kb region near the gene has far greater prevalence of derived alleles (represented by red dots) than of ancestral alleles (represented by gray dots).

**Fig. 5.** Extreme population differences in FYO allele frequency. The FYO allele, which confers resistance to P. vivax malaria, is prevalent and even fixed in many African populations, but virtually absent outside Africa (39).

**Fig. 6.** Long haplotype surrounding the lactase persistence allele. The lactase persistence allele is prevalent (~77%) in European populations but lies on a long haplotype, suggesting that it is of recent origin (6).
Positive Selection in Human Lineage

Positive Selection in Human Lineage

Mutations and LD

Slide Credits: Marc Schaub
Long Haplotypes – EHS, iHS tests

Less time:
- Fewer mutations
- Fewer recombinations
• Study of genes known to be implicated in the resistance to malaria.

• Infectious disease caused by protozoan parasites of the genus *Plasmodium*

• Frequent in tropical and subtropical regions

• Transmitted by the *Anopheles* mosquito
Application: Malaria

Image source: NIH - http://history.nih.gov/exhibits/bowman/images/malariacycleBig.jpg

Slide Credits: Marc Schaub
Application: Malaria

Malaria Endemic Countries, 2003

Note: This map shows countries with endemic malaria. In most of these countries, malaria risk is limited to certain areas.


Slide Credits: Marc Schaub
Results: G6PD, TNFSF5


Slide Credits: Marc Schaub
• Allison (1954): Sickle-cell anemia is limited to the region in Africa in which malaria is endemic.
Lactose Intolerance

Figure 1: Old world distribution of frequency of lactase persistence (lactose digesters) taken from available published data. Red indicates the proportion of lactose digesters in a given population with yellow representing maldigesters. Charts with a green central circle indicate that the overall published frequency for a country is comprised of different ethnic groups with very different phenotype frequencies. Data compiled by Ingram 2007.


Slide Credits: Marc Schaub
Lactose Intolerance

Positive Selection in Human Lineage

Fu W, Akey JM. 2013.