Genetics and Genomics

1. Genes and Inheritance

ggibson.gt@gmail.com
http://www.cig.gatech.edu

Course Outline

1. Mon AM  Genes and Inheritance (GG)
2. Mon PM  Molecular Biology of the Genome (CQ)
3. Tue AM  Association Studies and Gene Expression (GG)
4. Tue PM  Epigenetics and Genome Biology (CQ)
5. Wed AM  Evolutionary Genetics (GG) / Functional Genomics (CQ)
**Genotype and Phenotype**

The *genotype* of an organism is the sequence of it’s genes.

The *phenotype* of an organism the way it appears.

In general, genes are not deterministic. Genotypic variation among organisms specifies the information that, in combination with the environment, influences the phenotype.

**Pleiotropy** refers to the ability of single genes to influence multiple phenotypes.

**Penetrance** is the proportion of individuals with a genotype who have the phenotype / disease.

**Expressivity** is the degree / severity of the phenotype in affected individuals.

---

**Mendelian Genetics**

\[
\begin{array}{c c c c}
\text{A} & \text{a} \\
\text{A} & \text{a} \\
\hline
\text{F0: Pure breeding parents} \\

\text{A} & \text{A} \\
\text{a} & \text{a} \\
\hline
\text{F1: Heterozygous offspring} \\

\text{A} & \text{a} \\
\text{A} & \text{a} \\
\hline
\text{F2: Mendelian proportions of} \\
\text{Homozygotes + Heterozygotes}
\end{array}
\]
3 Models of Complex Disease

**CDCV: Common Disease / Common Variant**
The proposition that most disease susceptibility can be attributed to 10 to 20 loci, each of which explain around 5% of disease risk.

**RAME: Rare alleles of Major Effect**
The proposition that diseases are highly heterogeneous, with hundreds or thousands of rare mutations causing individual cases of disease.

**Infinitesimal:**
The proposition that we all carry thousands of very weak susceptibility alleles, and those unlucky enough to have too many are at highest risk, where rare variants or environmental triggers push us over the edge.

Models of the Genetics of Complex Traits

Heritability

Heritability is the proportion of variance in a population that can be attributed to genotypic differences

\[ h^2 = \frac{V_G}{V_P} \text{ where } V_P = V_G + V_E \]

The phenotypes may be discrete, such as disease status; categorical, such as number of digits; or continuous, such as height or a biochemical measure.

1. **Heritability is not a statement about individuals.**
   A heritability of 50% for diabetes does not imply that half the reason why someone is diabetic is genetic, the other half environmental. Rather, it suggests that there would be half as much diabetes in the population if everyone was genetically identical.

2. **Heritability is only a statement about a single population.**
   A heritability of 80% for height does not imply that most of the average difference in height between populations is due to genetic differences. Heritability estimates alone should not be used to draw inferences about genetic divergence between groups.

3. **Heritability is not the same as inheritance.**
   Inheritance is the correspondence between children and their biological parents. It can be due to environmental, including cultural, factors that are shared by family members, or to effects. The only way to confidently interpret heritability is to actually measure the genotypic contribution.

4. **Very low heritability does not imply very little genetic contribution.**
   It may either be due to relatively high environmental variance (hence, a large denominator \( V_E \)), or to an absence of variance in the genes that contribute. Many important genes, including drug targets, are not polymorphic and will only be discovered through other types of approach including model organism research.

---

**Broad Sense Heritability**

\[ V_G = V_A + V_D + V_I + V_{G\times E} \]

Narrow sense heritability is only the additive component whereas Broad sense heritability includes dominance, interaction and genotype-by-environment effects.

---

**Additive**

**Multiplicative**

**Recessive**

**Epistatic**
\[ V_P = V_A + V_D + V_I + V_{GeE} + V_E \]

- Loci are said to have **Additive** effects if the contributions of each individual allele can simply be added algebraically to arrive at a prediction of a phenotype given a genotype.

- **Dominance** refers to the observation that heterozygotes resemble one class of homozygotes more than the other.

- **Epistasis** refers to a locus-by-locus interaction, such as when alleles at two loci antagonize or synergize with one another.

- \( V_E \) is the *environmental* variance

---

**Dominance ratio**

- **BB**
  - Mean phenotype: \( a = 18-10 = 8 \)
  - Expected mid-value: \( (18+2)/2 = 10 \)

- **Bb**
  - Mean phenotype: \( d = 16-10 = 6 \)

- **bb**
  - Mean phenotype: \( -a \)

- **Number of “b” alleles**

---
From Mendelian to Quantitative genetics

Estimating Heritability
Twin Studies

Identical / Maternal

\[ r_{mz} = A + C \]

Dizygotic / Fraternal

\[ r_{dz} = \frac{1}{2}A + C \]

A = Additive Genetic component; C = Common Environment (smaller if reared apart)
E = unique environment = 1 - \( r_{mz} \)
\( r_{dz} \) should be greater than \( r_{mz} \) since C is larger where the womb/upbringing is shared

Mendelian Pedigree Studies

Recessive
1 in 4 affected if both parents are carriers

Dominant
1 in 2 affected if one parent is heterozygous

Sex-linked
Usually only males affected, can skip a generation

AUG TCC CAA CGA
AUG TCC TAA CGA
AUG TCC CAA CGA
AUG TCC CAA TGA
AUG TCC CAA TGA
AUG TCC TAA TGA

Mendelian Pedigree
Monogenic Disorders

Approximately 1 in 3,700 Americans have Cystic Fibrosis.

Assuming $p^2 = 0.00027$, then $p = 0.016$, the mutant allele frequency.

That is, 1 in 30 people are carriers (which is 120 times as many people as have CF), that is, 3% of Caucasians are carriers, and less than 0.03% sufferers.

It is very likely that someone in this class is a carrier of a CF mutation.

A CF carrier has a 1 in 30 chance of marrying another carrier by chance, and 1 in 4 of their children will be expected to have CF, and only half of all 2-child families will have an affected child.

There are hundreds of similar conditions (rare recessives with $p \approx 0.01$), so we are all carriers for multiple Mendelian disease genes. Collectively, as many as 1 in 25 couples should expect to be dual carriers for a recessive Mendelian disorder, corresponding to an approximate 1% affected rate in all children.

Around 1 in 400 children have an inherited inborn Error of Metabolism, namely an enzyme deficiency affecting amino acid, lipid or other organic molecule biosynthesis. For example:

- Phenylketonuria 1/15,000 mental retardation syndrome
- Galactosemia 1/40,000 liver dysfunction and cataracts
- Gaucher’s Disease 1/60,000 facial dysmorphology, liver disease
- Zellweger Syndrome 1/50,000 seizures, low muscle tone
- Lesch-Nyhan Syndrome 1/380,000 self-inflicted injury, gout / kidney disease

Online Mendelian Inheritance in Man (OMIM)
In each generation, slightly deleterious mutations add ~ 0.1% of the standing environmental variance to the heritability of traits, also reducing viability.


Assume there are 100 mutations at 1% frequency, each of which increases the risk of disease 2.5-fold over a baseline environmental risk of 1%.
Whence 0 alleles have a risk of 1%, 1 of 2.5%, 2 of 6%, 3 of 15%, 4 of 39%, 5 or more is highly penetrant.
The missing heritability problem is that variants discovered by GWAS only explain a minor fraction of the expected heritability. This may be because:

- The effect sizes are much smaller than previously thought (GRR 1.1 rather than 2)
- Narrow sense heritability has been over-estimated in pedigree studies
- It is rare, not common, variants, that contribute most of the variation
- Epigenetic inheritance accounts for much of the resemblance among relatives
- Broad sense heritability is prevalent, but hard to detect
- Genotyping chips do not tag causal variants effectively enough
Omnigenetics

Estimate that ~4% of all common SNPs influence height

Many core biochemical processes associate to different degrees with various traits


Polygenic Heterogeneity and Personalized Medicine