Introduction

Sungho Won
Graduate School of Public Health
Seoul National University
Basic Principles of Human Genetics
Genetics

• Science of inheritance

• Concerned with differences/similarities of individuals resulting from the mechanisms of genetic material

• Same root as in *generations*
**GENETICS**

- Father: sperm (1/2 of father's genes)
- Mother: egg (1/2 of mother's genes + cytoplasm)

Fertilized Egg = Zygote (2 sets of genes)

- Embryo
- Child
- Adolescent
- Adult

**ENVIRONMENT**

- Mother's uterus
- Nutrition
- Infection
- Smoking
- Occupational Exposures
Basic Principles

- **Genotype** - Genetic makeup - all the DNA received from the mother and father
  - Generally considered constant over time

- **Phenotype** - Appearance of the individual, what is measured
  - e.g., disease status, blood pressure, weight
  - Changes over time
  - Depends on genotype and environment
  - Phenotype = $F($genotype, environment$)$
Physical Basis of the Human Genotype

- 46 Chromosomes = 23 from father + 23 from mother = 23 pairs of chromosomes
  - 22 pairs of similar autosomes i.e., homologous autosomes
  - 1 pair of sex chromosomes (X,Y)
    - XX = Female
    - XY = Male
- Cytoplasm – mitochondrial DNA

Nucleus: Contains Genetic Material, DNA
Chromosomes

- 46 Chromosomes = 23 from father + 23 from mother = 23 pairs of chromosomes
  - 22 pairs of similar autosomes i.e., homologous autosomes
  - 1 pair of sex chromosomes (X,Y)
  - XX = Female, XY = Male
- Cytoplasm – also contains mitochondrial DNA
Mitosis

Process of cell division to produce somatic cells

Sperm
(father's gamete)

Egg
(mother's gamete)

Fertilized Egg = Zygote

Each diploid daughter cell contains an exact replica of the 23 pairs of chromosomes
Meiosis

Process of cell division that produces gametes (eggs or sperm) by reducing the number of chromosomes from the diploid number of 46 to the haploid number of 23.

Pairing of Homologous Chromosomes

Duplication

Crossover (recombination)

Reduction I

Reduction II

Haploid Gametes
DNA  transcription  protein-coding genes

RNA  transcribed  transcriptionally active regions, TARs

PROTEIN  translated
Definition

- **Gene** – DNA segment that codes for a functional unit; made up of exons interspersed with introns

- **Locus** - Location of a gene (or genes) on a chromosome
  - e.g., the gene(s) that code(s) for the ABO blood type protein is (are) on chromosome #9

- **Alleles** - Different forms of a gene, or different genes, that occupy the same locus
  - e.g., the alleles for the ABO blood type are all found at a specific locus on chromosome #9
Definition

• **Allelic** – "Genes" that occupy the same locus

• **Non-allelic** – "Genes" that occupy different loci (for instance, the gene for albinism)

• **Linkage** – Alleles at different loci that are close together on the same single chromosome tend to stay together because the probability of a crossover is small
Mendelian Segregation

- Mendel's First Law: Law of Segregation
  - For a given locus, each parent transmits one allele, either the maternal allele or the paternal allele, independently to each offspring (zygote)
  - Consider a single autosomal locus with 2 alleles, \( A_1 \) and \( A_2 \)
  - Genotypes = \( A_1 A_1 \) (homozygote, homozygous for \( A_1 \))
    \( A_1 A_2 \) (heterozygote, heterozygous)
    \( A_2 A_2 \) (homozygote, homozygous for \( A_2 \))

<table>
<thead>
<tr>
<th>Parental genotypes</th>
<th>Gametes</th>
<th>P(gamete transmitted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 A_1 )</td>
<td>( A_1 )</td>
<td>1</td>
</tr>
<tr>
<td>( A_1 A_2 )</td>
<td>( A_1 )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td></td>
<td>( A_2 )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>( A_2 A_2 )</td>
<td>( A_2 )</td>
<td>1</td>
</tr>
</tbody>
</table>
Mendelian Segregation

Consider the mating of $A_1 A_2 \times A_1 A_2$

<table>
<thead>
<tr>
<th>Gametes (Probability)</th>
<th>$A_1 (\frac{1}{2})$</th>
<th>$A_2 (\frac{1}{2})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 (\frac{1}{2})$</td>
<td>$A_1 A_1 (\frac{1}{4})$</td>
<td>$A_1 A_2 (\frac{1}{4})$</td>
</tr>
<tr>
<td>$A_2 (\frac{1}{2})$</td>
<td>$A_2 A_1 (\frac{1}{4})$</td>
<td>$A_2 A_2 (\frac{1}{4})$</td>
</tr>
</tbody>
</table>

Offspring genotypes $= A_1 A_1 \quad A_1 A_2 \quad A_2 A_2$

$\therefore P(\text{genotype}) = \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4}$
Mendelian Segregation

- In general, for a single autosomal locus with two alleles, $A_1$ and $A_2$, there are 3 possible genotypes ($A_1A_1$, $A_1A_2$, $A_2A_2$) and 9 possible mating types, of which 6 are distinct when ignoring the sex of the parents.

| Mating type | $P(\text{offspring genotype}|\text{mating type})$ |
|-------------|-------------------------------------|
|             | $A_1A_1$ | $A_1A_2$ | $A_2A_2$ |
| $A_1A_1 \times A_1A_1$ | 1 | 0 | 0 |
| $A_1A_1 \times A_1A_2$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 |
| $A_1A_1 \times A_2A_2$ | 0 | 1 | 0 |
| $A_1A_2 \times A_1A_2$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |
| $A_1A_2 \times A_2A_2$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $A_2A_2 \times A_2A_2$ | 0 | 0 | 1 |

- $P(\text{offspring has genotype } u | \text{parents have genotypes } u_F \text{ and } u_M)$ is a transition probability.
- $P(\text{parent with genotype } u \text{ transmits allele } A) = \tau_{u \rightarrow A}$ is a transmission probability.

Elston and Stewart (1971)
Genotype vs Phenotype

- **Genotype** - the specific genetic constitution of an organism
  - The term is often used to pertain to specific traits or loci, e.g., we often use genotype to denote the alleles present at a particular locus
  - The genotype is not observed directly but rather is inferred through the phenotype
  - Usually restricted to the DNA sequence & epigenetic phenomena

- **Phenotype** - the observable expression of the genotype (and the environment)

- **Phenoset** - set of all genotypes consistent with an individual's phenotype
ABO Blood Group Illustration

Proteins in the membrane
Blood Group A
A - Antigen

Proteins in the membrane
Blood Group O
O - Antigen

### Phenotype (blood group) | Genotype
---|---
A | AA or AO
B | BB or BO
AB | AB
O | OO
Dominance and Recessiveness

- Classically refer to the situation in which the expression of one allele in a heterozygote masks the other at a particular locus
  - The allele that is expressed in the phenotype is dominant with respect to that phenotype; the one that is masked is recessive
- Describe the relationship between phenotype and genotype
- Codominance: both alleles are expressed in the heterozygote, or the heterozygote phenotypic expression differs from that of either homozygote (contrast: *additive*)
- Overdominance: the phenotypic expression of the heterozygote does not lie between those of the homozygotes
Relationship between Genotype and Phenotype

• Consider a single autosomal locus with 2 alleles (diallelic) = \(A_1, A_2\)

• "Dominant" and "Recessive" describe relationships between alleles and a specific phenotype

• Consider the phenotype "diseased"

<table>
<thead>
<tr>
<th>Genotype</th>
<th>With respect to &quot;diseased&quot; phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A_1) dominant to (A_2)</td>
</tr>
<tr>
<td>(A_1A_1)</td>
<td>Diseased</td>
</tr>
<tr>
<td>(A_1A_2)</td>
<td>Diseased</td>
</tr>
<tr>
<td>(A_2A_2)</td>
<td>Normal</td>
</tr>
</tbody>
</table>
Genotype and Phenotype

- The phenotype can vary among persons with the same genotype, from no disease to variable degrees of severity (expressivity)

- **Penetrance**: probability of disease given the genotype (i.e. not everyone with the genotype will be affected)

- For the above example, with complete penetrance

| Genotype | P(Disease|Genotype) |
|----------|------------------|
|          | A\(_1\) dominant | A\(_1\) recessive |
| A\(_1\)A\(_1\) | 1                 | 1                |
| A\(_1\)A\(_2\) | 1                 | 0                |
| A\(_2\)A\(_2\) | 0                 | 0                |
# Sickle Cell Anemia

<table>
<thead>
<tr>
<th>SS</th>
<th>Ss</th>
<th>ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>two normal alleles</td>
<td>&quot;trait&quot; (carrier)</td>
<td>disease</td>
</tr>
</tbody>
</table>

- **Disease**: sickling of red blood cells
- **Trait**: sickling of red blood cells in low oxygen tension
- **s** is:
  - recessive with respect to disease
  - dominant with respect to "trait"
Definitions

- **Incomplete penetrance** - refers to the failure of a phenotype to be evident even though the genotype usually producing the phenotype is present
  - *Variable age of onset* e.g., Huntington's Disease, also called "age dependent penetrance"

- The term **penetrance function** is used to describe the probability (mass or density) that an individual has a particular phenotype given a particular genotype

- **Sporadic** - refers to the presence of a phenotype even though the genotype usually producing it (individual or parental) is absent

- **Variable expressivity** - refers to variation in the expression of a trait, e.g., in severity or more generally

- **Imprinting** – the phenotype depends on the sex of the transmitting parent
  - *Paternal (maternal) imprinting*: father’s (mother’s) allele is not expressed
Definitions

- **Syntenic** - loci on the same chromosome (pair)
- **Paralogous** – loci that arose from a common ancestor locus, by duplication
- **Orthologous** – same locus in different species, by evolution
- **Pleiotropy** - the occurrence of multiple diverse effects due to segregation at a single locus; effects on multiple traits due to a single “allele”, e.g., phenylketonuria (PKU)
  - If a single locus produces an effect on several traits, that locus is said to be pleiotropic (or have pleiotropic effects)
- **Phenocopy** - an individual (condition) whose phenotype under a particular environmental condition is identical to the one of another individual whose phenotype is determined by the genotype, e.g., Vanessa genus of butterflies, deafness, thalidomide syndrome vs. phocomelia
Definitions

- **Genetic heterogeneity** - refers to a phenomenon in which a single phenotype or genetic disorder may be caused by any one of a multiple number of alleles or non-allele (locus) mutation

- Allelic heterogeneity – most diseases. a phenomenon in which a single phenotype or genetic disorder may be caused by any one of a multiple number of alleles or non-allele (locus) mutations. For example, there are over 1000 known mutant alleles of the CFTR gene that cause cystic fibrosis.

- Locus heterogeneity - variations in completely unrelated gene loci cause a single disorder. For example, retinitis pigmentosa has autosomal dominant, autosomal recessive, and X-linked origins. However, only one mutant locus is needed for the phenotype to manifest.

Retinitis Pigmentosa
(망막세포변성)

Definition

- **Epistasis** - interaction (non-additivity) between alleles situated at different loci, e.g., where an allele at one locus masks or prevents the expression of an allele at a different locus, or otherwise affects expression in a non-additive manner

  - Interaction: In general, the statistical interaction of factors on a measured response depends on the scale of measurement
  - e.g., for an additive model, multiplicative effects represent interaction

- Epistasis is *inter*-locus interaction (between alleles at different loci) whereas dominance is *intra*-locus interaction (between alleles at the same locus)

  - Note: dominance and epistasis depend on the scale used to measure the trait (the effect of a binary trait may be measured on the penetrance scale or a function of that, e.g. logit, probit)
Definitions

- **Linkage** - Location of genetic loci sufficiently close together on a chromosome that they do not segregate independently

- **Crossing-over** - a reciprocal event in a pair of homologous chromosomes, takes place when the chromosomes have already duplicated at the beginning of the first meiotic division

- **Recombination Fraction** (θ) - the probability that an odd number of crossover events will take place between two loci; the proportion of recombinant gametes
Definitions

- **Genetic distance** – measures the amount of crossing over

- **Interference** – occurs when crossing-over in an interval influences the amount of crossing over in a nearby region

- **Map functions** – relate the recombination fraction to the genetic distance
  - **Haldane** map function – does not allow for interference
  - **Kosambi** map function – one of many map functions that allow for positive interference
**Measures of Linkage Disequilibrium (LD)**

<table>
<thead>
<tr>
<th>Locus 1</th>
<th>Allele 1</th>
<th>Allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele 1</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Allele 2</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>a + c</td>
<td>b + d</td>
</tr>
</tbody>
</table>

- $D'$ defined as $D/D_{\text{max}}$, where $D = ad - bc$
  - $D_{\text{max}} = \min[(a + b)(b + d), (c + d)(a + c)]$ for $D > 0$
  - $D_{\text{max}} = \min[(a + b)(a + c), (c + d)(b + d)]$ for $D < 0$

- $\log \text{OR} = \log]\frac{ad}{bc}\]$

- $r^2 = \frac{D^2}{(a + c)(b + d)(a + b)(c + d)} = \chi^2 \text{ stat}/N$

- LD is a special case of gametic phase disequilibrium

Zabetian et al, 2003
**Traits**

- Phenotypic traits of interest may be
  - Discrete or continuous
  - Univariate or multivariate

- Discrete traits (qualitative or quantitative) are categorical and can be dichotomous (binary) or polytomous,
  - e.g., affected vs. unaffected status; no disease, mild disease, severe disease; blood group designation (Type A, B, AB, O); number of children

- Continuous traits (quantitative) are measured on at least an interval scale
  - e.g., measurements of blood pressure, anthropometric measures, lipid, and lipoprotein levels
Phenotypic differences can be caused by allelic variation at:

- Single locus = monogenic (unilocal)
- Few loci = oligogenic (paucilocal)
- Many loci = polygenic (multilocal)
- Any of the above + environment = multifactorial

**Multifactorial** - determined by multiple genetic and/or nongenetic factors; literally, "many factors"

The term **polygenic** literally implies multiple factors that are exclusively genetic, though it is also used for a model involving the environment plus multiple genetic factors with a particular type of effect, e.g., the simple additive polygenic model (on a particular scale)
### Single Locus Model

- **Single Autosomal Locus:** In general, there can be $K$ alleles at a locus in the population, but only 2 in any one individual.
  - $K$ alleles possible: $A_1, A_2, \ldots, A_K$
  - Possible genotypes (pairing of any 2 alleles)

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$\ldots$</th>
<th>$A_K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$A_1A_1$</td>
<td>$A_1A_2$</td>
<td>$A_1A_K$</td>
<td></td>
</tr>
<tr>
<td>$A_1$</td>
<td></td>
<td>$A_2A_1$</td>
<td>$A_2A_2$</td>
<td>$A_2A_K$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$A_K$</td>
<td>$A_KA_1$</td>
<td></td>
<td>$A_KA_K$</td>
<td>$A_KA_K$</td>
</tr>
</tbody>
</table>

- Number of distinct homozygotes
- Number of distinct heterozygotes,
  - if $A_iA_j + A_jA_i$ are equivalent
  \[ = \frac{K(K - 1)}{2} \]
  - if $A_iA_j + A_jA_i$ are not equivalent
  \[ = K(K - 1) \]
- Total, distinct genotypes
  - if $A_iA_j + A_jA_i$ are equivalent
  \[ = K + \frac{K(K - 1)}{2} = \frac{K(K + 1)}{2} \]
Allele Frequency

- In a population of N persons, each with 2 alleles at a particular locus, the total number of alleles is 2N
  - (Relative) frequency of $A_1$ allele = (Number of $A_1$ alleles)/2N

- Population allele frequencies can change from generation to generation due to:
  - Abnormal locus segregation (segregation distortion, meiotic drive - all gametic contributions are not equally probable)
  - Mutation
  - Selection - depends on fertility of parents and viability of offspring
  - Migration
  - Chance, in small populations - genetic drift

- Allele frequencies can also change with age within a generation, and could be sex dependent
Hardy-Weinberg Proportions: One Autosomal Locus

- Assume panmixia
  - Normal locus segregation
  - Equal allele frequencies for males and females
  - No mutation, selection, migration
  - Large population
  - Random mating with respect to genotype of interest - an individual has the same chance of mating with any individual of the opposite sex, so that genotypes of mates are randomly paired
Hardy-Weinberg Proportions

• Then, because genotypes of parents are randomly paired, and parents randomly transmit one of their two alleles, there is random pairing of alleles and the genotype frequencies in a population depend only on the allele frequencies.

• For an autosomal locus, let $q_i$ be the frequency of the allele $A_i$, the genotype frequencies are said to be in Hardy-Weinberg proportions when:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_iA_i$</td>
<td>$q_i^2$</td>
</tr>
<tr>
<td>$A_iA_j$</td>
<td>$2q_iq_j$</td>
</tr>
</tbody>
</table>

• For 2 alleles, $A_1$ and $A_2$, with frequencies $q$ and $(1 - q)$:

| Genotype | Frequency |\
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>$\psi_{A_1A_1} = q^2$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$\psi_{A_1A_2} = 2q(1-q)$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$\psi_{A_2A_2} = (1-q)^2$</td>
</tr>
</tbody>
</table>
Hardy-Weinberg Proportions

• Note:

• Genotype frequencies at one autosomal locus attain Hardy-Weinberg proportions after a single generation of random mating, regardless of genotype frequencies in the parents

• If allele frequencies do not change over generations, they are in equilibrium

• Hardy-Weinberg proportions + Equilibrium = Hardy-Weinberg Equilibrium (HWE)

• It is possible for allele frequencies to change over generations, yet the genotype frequencies satisfy Hardy-Weinberg proportions in each generation

• Therefore Hardy-Weinberg proportions ⇨ equilibrium

Li (1976)
Heritability
Two alleles are identical by descent if they are copies of the same ancestral allele.
Identity in State versus Identity by Descent

- Alleles physically identical are identical in state (IIS) i.e., the same
- Alleles that are copies of the identical allele carried by a common ancestor are identical by descent (IBD)
- Alleles that are IBD are also IIS (ignoring the possibility that mutation occurs, a rare event)
- Both alleles at a given locus within a particular individual may be IBD; i.e. in the offspring of a consanguineous mating, such as marriage of first cousins
- We speak of the number or proportion of alleles shared IBD by two individuals
  - Monozygotic twins have two alleles IBD at every locus
  - If the two parents are unrelated, a parent and child share exactly one allele IBD at every locus
  - Siblings may have 0, 1, or 2 alleles IBD at a given locus
  - In a large random mating population, two unrelated persons share 0 alleles IBD at every locus
IBD Probabilities

• What are the probabilities $f_2$, $f_1$, or $f_0$ of sharing 2, 1, or 0 alleles, respectively, IBD at a random autosomal locus for different types of relatives?

• Assume a large random mating population (no consanguinity):
  • For identical twins: $f_2 = 1$, $f_1 = 0$, $f_0 = 0$
  • For parent-offspring: $f_2 = 0$, $f_1 = 1$, $f_0 = 0$
  • For siblings: $f_2 = \frac{1}{4}$, $f_1 = \frac{1}{2}$, $f_0 = \frac{1}{4}$

• UNILINEAL RELATIVES - related by only "one line" of genetic descent; can have at most one allele IBD, implying that $f_2 = 0$
For full sibs:

\[
\begin{array}{ccccc}
\text{A}_1\text{A}_2 & \times & \text{A}_3\text{A}_4 \\
\hline
\text{A}_1\text{A}_3 & 2 & 1 & 1 & 0 \\
\text{A}_1\text{A}_4 & 1 & 2 & 0 & 1 \\
\text{A}_2\text{A}_3 & 1 & 0 & 2 & 1 \\
\text{A}_2\text{A}_4 & 0 & 1 & 1 & 2 \\
\end{array}
\]

\[
P(\text{sharing 2}) = f_2 = \frac{4}{16} = \frac{1}{4}
\]

\[
P(\text{sharing 1}) = f_1 = \frac{8}{16} = \frac{1}{2}
\]

\[
P(\text{sharing 0}) = f_0 = \frac{4}{16} = \frac{1}{4}
\]
An offspring of a consanguineous mating can have his/her two alleles at a locus IBD.

This can also occur under certain patterns of non-random mating, for all individuals in a population.

The probability that two alleles are IBD at a locus is called the inbreeding coefficient \( f \), for that population.

Then, for an autosomal locus where \( q_i \) is the frequency of the allele \( A_i \), we have:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_iA_i )</td>
<td>( (1-f)q_i^2 + f \cdot q_i )</td>
</tr>
<tr>
<td>( A_iA_j )</td>
<td>( (1-f)q_iq_j )</td>
</tr>
</tbody>
</table>

For 2 alleles, \( A_1 \) and \( A_2 \), with frequencies \( q \) and \( 1 - q \):

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( \psi_{A_1A_1} = (1-f)q^2 + f \cdot q )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( \psi_{A_1A_2} = 2 \cdot (1-f) \cdot q \cdot (1-q) )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>( \psi_{A_2A_2} = (1-f)(1-q)^2 + f \cdot (1-q) )</td>
</tr>
</tbody>
</table>
Genetic Components of Variance

• Additive genetic variance
  • due to the additive effects of alleles
• Dominant genetic variance
  • intra-locus interaction of alleles
• Epistatic genetic variance
  • inter-locus interaction of alleles
Genetic Components of Variance

- Environmental variance \( = \sigma_e^2 \)
- Total variance \( = \sigma_g^2 + \sigma_e^2 \)
  \[= \sigma_a^2 + \sigma_d^2 + \sigma_e^2\]
Partitioning the Genotypic Variance at One Locus

\[ \sigma_g^2 = 2pq[a-d(p-q)]^2 + 4p^2q^2d^2 = \sigma_a^2 + \sigma_d^2 \]

- If \( d = (a-a)/2 = 0 \), \( \sigma_g^2 = 2pqa^2 = \sigma_a^2 \)
Heritability

- Traits are familial if members of the same family share them, for whatever reason.
- Traits are genetically heritable only if the similarity arises from shared alleles and genotypes.
- To quantify the degree of heritability, one must distinguish between two sources of phenotypic variation:
  - hereditary (i.e., genetic) and environmental.
- \( \text{Phenotype} = \text{Heredity} + \text{Environment} + \text{Genotype} \times \text{Environment} \)
- \( V(\text{Phenotype}) = V(\text{Heredity}) + V(\text{Environment}) + V(\text{Genotype} \times \text{Environment}) \)

\[
\sigma_P^2 = \sigma_H^2 + \sigma_E^2 + \sigma_{GE}^2
\]

Note: interaction depends on scale of measurement.
Heritability in the Broad Sense

- The proportion of the total phenotypic variance that is due to the hereditary variance:

\[
h^2 = \frac{\sigma_H^2}{\sigma_P^2} = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_E^2 + \sigma_{GE}^2}
\]

where

\[
\sigma_H^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2
\]

\[
\sigma_A^2 = \text{additive genetic variance}
\]

\[
\sigma_D^2 = \text{dominant genetic variance}
\]

\[
\sigma_I^2 = \text{epistatic genetic variance}
\]

- \(h^2\) is commonly misinterpreted as a measure of how "important" genes are in influencing a trait
Heritability in the Broad Sense

• Limitations of $h^2$:
  • Not a fixed characteristic of a trait, but depends on the population in which it was measured and the set of environments in which that population developed
  • If genotype and environment interact to produce phenotype, no partition of variation can actually separate causes of variation
  • High $h^2$ does not mean that a trait cannot be affected by its environment
  • Therefore $h^2$ has limited meaning and use in humans other than as a parameter to allow for familial correlations
Heritability in the Narrow Sense

- We have
  \[ \sigma_H^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 \]
  where \[ \sigma_A^2 = \text{additive genetic variance} \]
  \[ \sigma_D^2 = \text{dominant genetic variance} \]
  \[ \sigma_I^2 = \text{epistatic genetic variance} \]

- \[ h^2 = \frac{\sigma_A^2}{\sigma_P^2} \] is heritability in the narrow sense

- This is a key parameter in plant and animal breeding

- In human genetics it is sometimes used to measure possible genetic effects not due to major gene segregation (\( H \) or \( H_p \), "polygenic heritability", the heritability after excluding any major gene effect from the phenotype)
Simple Additive Polygenic Model

- Polygenic inheritance is due to the cumulative effects of alleles at a large number of segregating loci
- It is parameterized as a normal distribution based on an "infinite" number of loci with small, equal, and additive allele effects at each locus
- Both discrete and continuous traits can be modeled using this approach
- Simple Polygenic Model: completely autosomal, additive (no dominance or epistasis, though the model is sometimes extended to allow for these)
Simple Additive Polygenic Model

• Consider many autosomal loci, 2 alleles at each with genotypes AA Bb CC dd ..., or Aa Bb Cc DD ..., etc, and each capital letter increases the value of a continuous phenotype or the risk of being affected.

• Assume the value of the continuous phenotype or the risk of being affected depends only on the number of capital letters, i.e., that segregation at each locus has equal effect - a "cumulative action of alleles".

• In a random mating population with allele frequencies at each locus strictly between 0 and 1, as the number of loci increases, the proportion of capital letters will tend to be normally distributed in the population.

• For practical purposes, even as few as 3 or 4 loci can lead to a reasonable approximation to normality.
Example of Additive Polygenic Model

- Three-locus system with loci A, B, C with equifrequent alleles A a, B b, C c, respectively.
- The continuous phenotype is determined by the number of uppercase alleles.

![Diagram showing genotypes and number of uppercase letters](chart.png)
Additive Polygenic Model: Continuous Trait

- Assume an additive polygenic effect that we will call $A \sim N(\mu, \sigma_A^2)$ and, for any given $A$, the phenotypic distribution is $N(A, \sigma_E^2)$, $P = A + E$, $E \sim N(0, \sigma_E^2)$
- Assume $A$ and $E$ uncorrelated, then $P \sim N(\mu, \sigma_A^2 + \sigma_E^2)$

Heritability of a continuous trait

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2}$$

Examples: height, arm span
Additive Polygenic Model: Binary Trait

- Additive genetic liability = \( A \sim N(\mu, \sigma_A^2) \)
- \( P(\text{affected}|A) = \text{risk function} \)
- If the risk function is cumulative \( N(T, \sigma_E^2) \), we have:

\[
L = A + E
\]

\[
E \sim N(0, \sigma_E^2), \quad L = A + E, \quad A \text{ and } E \text{ independent}, \quad L \sim N(\mu, \sigma_A^2 + \sigma_E^2)
\]

Critical assumption: The same transformation induces normality in both
Logistic and Normal Distributions

Cumulative Probability

-3 -2 -1 0 1 2 3

Normal
Logistic

x, SD units
Pedigree Relationships

Examples:

- Spouses: 1 and 2
- Parent-offspring: 12 and 22
- (Full)siblings: 9 and 10
- Half siblings: 17 and 21
- Maternal grandparent-grandchild: 5 & 16
- Paternal grandparent-grandchild: 8 & 24
- Avuncular: 7 and 15
- First cousins: 14 and 18
- Second cousins: 23 and 30
- First cousins once removed: 19 and 25
- Dizygotic (fraternal) twins: 26 and 27
- Monozygotic (identical) twins: 28 and 29
- Founders: 1, 2 (total of 8 in this pedigree)
- Non-founder: 4
Degree of Relationship

- Let $\pi = \text{mean probability two relatives share an allele IBD, } f_2 + \frac{1}{2} f_1$
- Then in a randomly mating population, degree of relationship (an integer) = $\log_{\frac{1}{2}} \pi$

<table>
<thead>
<tr>
<th>Relationship</th>
<th>$f_2$</th>
<th>$f_1$</th>
<th>$\pi$</th>
<th>Degree of Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monozygotic twins</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dizygotic twins and sibs</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
<tr>
<td>Parent-offspring pairs</td>
<td>0</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
<tr>
<td>Grandparent-grandchild Avuncular pairs</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
<td>2</td>
</tr>
<tr>
<td>First cousin great-grandparent-grandchild pairs</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{8}$</td>
<td>3</td>
</tr>
</tbody>
</table>

- Note: The "first degree relatives" are a very heterogeneous group that are the same on average, but have very different distributions, even at a single locus
Covariance Between Relatives

- To understand the way a trait is inherited, one can look at the individual components of phenotypic variance

- To estimate these components we need to determine the covariances between different sets of relatives

- These can be expressed in terms of the components we have defined plus other environmental terms whose expectations can be made to be zero in well-designed experiments

- In the case of human data (in which experimentation is not possible), the magnitudes of these environmentally caused covariances are often in doubt
Covariance Between Relatives

- Suppose we have two related individuals from a random mating population, each with its own phenotypic measure

\[ P = A + D + I + E + G \times E \]

- The covariance between two relatives will be made up of the covariance due to their genetic similarities and the covariance due to their environmental \((+ G \times E)\) similarities

- Covariance due to their environmental \((+ G \times E)\) similarities can be made zero in planned experiments

- Otherwise, it is a nuisance term which we can express as \(Cov (c.e.)\), where \(c.e.\) stands for common environment, and is specific to the type of relationship
Locus-Specific Genetic Variance

- In general, if there is no epistatic variance (i.e. $\sigma_i^2 = 0$), the genetic covariance between two relatives in a random mating population is

$$ (f_2 + \frac{1}{2} f_1)\sigma_A^2 + f_2\sigma_D^2 $$

where $f_i$ is the probability the relatives share $i$ alleles IBD at a random locus.

- If the $f_i$ are for a specific locus with additive genetic variance $\sigma_a^2$ and dominance variance $\sigma_d^2$, then the locus-specific genetic variance is

$$ (f_2 + \frac{1}{2} f_1)\sigma_a^2 + f_2\sigma_d^2 $$

- $\frac{\sigma_a^2 + \sigma_d^2}{\sigma_p^2}$ is the locus-specific broad heritability.

- $\frac{\sigma_a^2}{\sigma_p^2}$ is the locus-specific narrow heritability.

- $f_2 + \frac{1}{2} f_1 = \pi$, mean probability of sharing an allele IBD.
Kinship coefficient

- Kinship coefficient: probability that two genes sampled at random from each individual are identical-by-descent
- Inbreeding coefficients: the kinship coefficient between individuals’ parents

\[
\pi_{3,4} = \frac{1}{4} P(A_{31} \equiv A_{43}) + \frac{1}{4} P(A_{31} \equiv A_{44}) + \frac{1}{4} P(A_{32} \equiv A_{43}) + \frac{1}{4} P(A_{32} \equiv A_{44})
\]

\[
= P(A_{31} \equiv A_{43})
\]

\[
= \frac{1}{2} P(\text{IBD}(3,4) = 2) + \frac{1}{4} P(\text{IBD}(3,4) = 1)
\]
Kinship coefficient

- Dynamic algorithm

\[
\pi_{ij} = \begin{cases} 
0 & i, j \text{ are founders} \\
\frac{1}{2} & i = j, i \text{ is a founder} \\
\frac{1}{2} \left( \pi_{\text{mother}(i)j} + \pi_{\text{father}(i)j} \right) & i \neq j \\
\frac{1}{2} \left( 1 + \pi_{\text{mother}(i)\text{father}(j)} \right) & i = j, i \text{ is a nonfounder}
\end{cases}
\]
Linear Mixed Model vs Heritability

- **Linear Mixed Model**

  \[ Y = X\beta + G + \varepsilon, \quad G \sim MVN(0, \sigma^2 g \Phi), \quad \varepsilon \sim MVN(0, \sigma^2 \varepsilon I) \]

  \[ Y = (y_i)_{n \times 1}, \quad X = (x_{ij})_{n \times p}, \quad G = (g_i)_{n \times 1}, \quad \varepsilon = (\varepsilon_i)_{n \times 1} \]

- **Additive polygenic effect for individual** \( i \): \( g_i = g_{i1} + g_{i2}, \ g_{ij} \sim N(0, \sigma^2 g / 2) \)
  - \( \text{cov}(y_i, y_{i'}) = \text{cov}(g_i, g_{i'}) = E(g_{i1}g_{i'1} + g_{i1}g_{i'2} + g_{i2}g_{i'1} + g_{i2}g_{i'2}) \)
  - \( = P(g_{i1} = g_{i'1}) \text{var}(g_{i1}) + P(g_{i1} = g_{i'2}) \text{var}(g_{i1}) + P(g_{i2} = g_{i'1}) \text{var}(g_{i2}) + P(g_{i2} = g_{i'2}) \text{var}(g_{i2}) \)
  - \( = 2\pi_{ii'} \{2 \text{var}(g_{i1})\} = 2\pi_{ii'} \text{var}(g_{i1} + g_{i2}) \)

- **Therefore,**

  \[ \Phi = (2\pi_{ii'})_{n \times n}, \quad \pi_{ii'} : \quad \text{kinship coefficient} \]

- **Locus-specific narrow-sense heritability**

  \[ H^2 = \frac{\sigma^2 g}{\sigma^2 g + \sigma^2 \varepsilon} \]
Linear Mixed Model vs Heritability

- There can be dominant (interacting) polygenic effect and in such a case
  \[ g_i = g_{i1} + g_{i2} + d_i \cdot I(\text{genotype}_{i1} = \text{genotype}_{i2} = 1), \quad d_i \sim N(0, \sigma_d^2) \]
  \[ Y = X\beta + G + D + \varepsilon, \quad G \sim MVN(0, \sigma_g^2 \Phi), \quad \varepsilon \sim MVN(0, \sigma_e^2 I) \]
  \[ D \sim MVN(0, \sigma_d^2 A), \quad (A)_{ij} = f_{2(ij)} \]

- Broad-sense heritability can be obtained by
  \[ H^2 = \frac{\sigma_g^2 + \sigma_d^2}{\sigma_g^2 + \sigma_d^2 + \sigma_e^2} \]
Heritability Estimates in Korean Population

- Narrow-sense heritability

<table>
<thead>
<tr>
<th>Traits</th>
<th>Cohort</th>
<th>HTK</th>
<th>ASF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.76(0.04)</td>
<td>0.66(0.09)</td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td>0.43(0.05)</td>
<td>0.41(0.08)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td>0.37(0.05)</td>
<td>0.27(0.08)</td>
</tr>
<tr>
<td>TG</td>
<td></td>
<td>0.47(0.05)</td>
<td>0.50(0.08)</td>
</tr>
<tr>
<td>TCHL</td>
<td></td>
<td>0.72(0.04)</td>
<td>0.50(0.07)</td>
</tr>
<tr>
<td>HDL</td>
<td></td>
<td>0.43(0.05)</td>
<td>0.47(0.08)</td>
</tr>
<tr>
<td>LDL</td>
<td></td>
<td>0.37(0.05)</td>
<td>0.23(0.08)</td>
</tr>
<tr>
<td>SBP</td>
<td></td>
<td>0.53(0.05)</td>
<td>0.21(0.08)</td>
</tr>
<tr>
<td>DBP</td>
<td></td>
<td>0.43(0.05)</td>
<td>0.41(0.08)</td>
</tr>
</tbody>
</table>
Questions??