Determining Genetic Effects: Heritability, Linkage & Association

David C. Glahn, PhD
Olin Neuropsychiatry Research Center &
Department of Psychiatry, Yale University

Yale UNIVERSITY

THE INSTITUTE OF LIVING

June XXX, 2010

No Conflict of Interest

- I will not discuss off label use and/or investigational drugs in my presentation
- I have no financial relationships to disclose
- Financial support was provided by:
  - NIMH R01 MH078111 (PI: J Blangero)
  - NIMH R01 MH078143 (PI: DC Glahn)
  - NIMH R01 MH083824 (PI: Glahn & Blangero)
  - NIMH R37 MH059490 (PI: J Blangero)
  - NIMH R01 MH080912 (PI: DC Glahn)
- Many slides were stolen from Laura Almasy

Mendelian & Complex Traits

Mendelian Trait

- A trait influenced by a single gene producing a clear pattern of dominant or recessive inheritance within families.
- Examples: cystic fibrosis, sickle cell anemia, hemophilia

Complex Trait

- A trait influenced by multiple genes and their interactions with each other and with the environment.
- Examples: autism, schizophrenia, Alzheimer's, brain anatomy, BOLD signal

Take Home Message

- “This may be the most important thing I say in this lecture. There is no one size fits all in genetic analysis of complex traits.”

–Laura Almasy

How to do Genetics

Questions for the Study of Complex Trait Genetics

1) Is this trait influenced by genetic factors? How strong are these genetic influences?

2) Which traits are influenced by the same genes?

3) Where are the genes that influence a trait?

4) What are the specific genes that influence the trait?

5) What specific genetic variants influence the trait and how do they interact with each other and with the environment?

Six Types of Samples for Genetics

1. Adoptees: separating the effects of genes and family environment
2. Unrelated individuals: association only, estimation of effect size after variants are identified
3. Parent-child triads: association in the presence of linkage (transmission disequilibrium test), heritability/relative risk
4. Twins: heritability, relative risk, genetic correlations, linkage, association
5. Relative pairs: heritability, relative risk, genetic correlations, linkage, association
6. Pedigrees: heritability, relative risk, genetic correlations, linkage, association

Subject Ascertainment Strategies

1. By phenotype: if you're studying a rare disease, you must ascertain on phenotype. This is also necessary for some study designs (TDT, case/control).

2. Randomly: if you're studying a common disease, you'll find it in a random sample. If you're interested in multiple traits, ascertaining on one improves power only for that one. May also want to study normal variation.

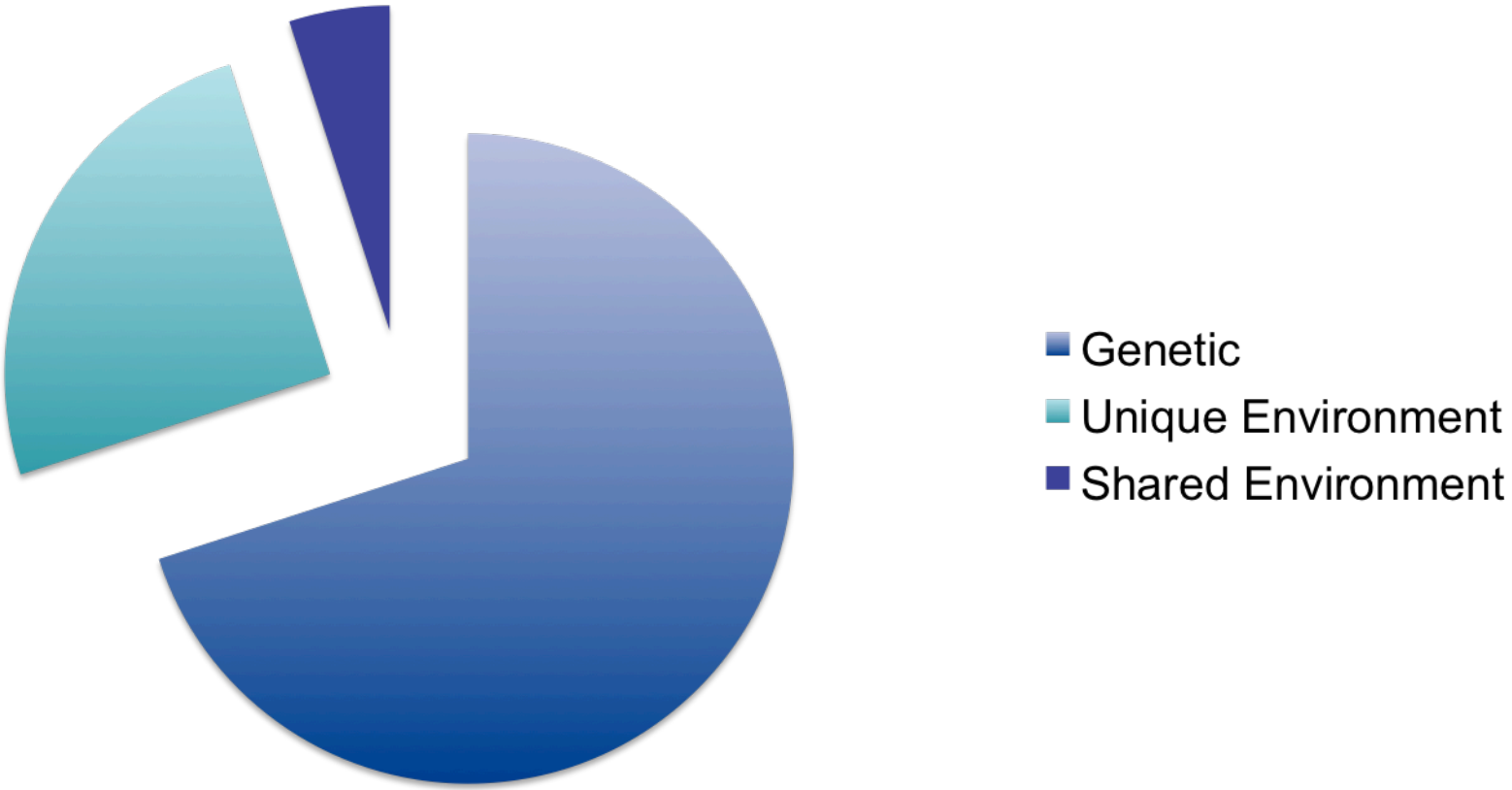
3. Ascertainment also depends on (and limits) method of analysis – TDT, affected relative pair linkage.

Question 1: Heritability

Is this trait influenced by genetic factors?

How strong are these genetic influences?

Heritability



Proportion of the trait variance attributable to genetic effects

Estimating Heritability: Twins

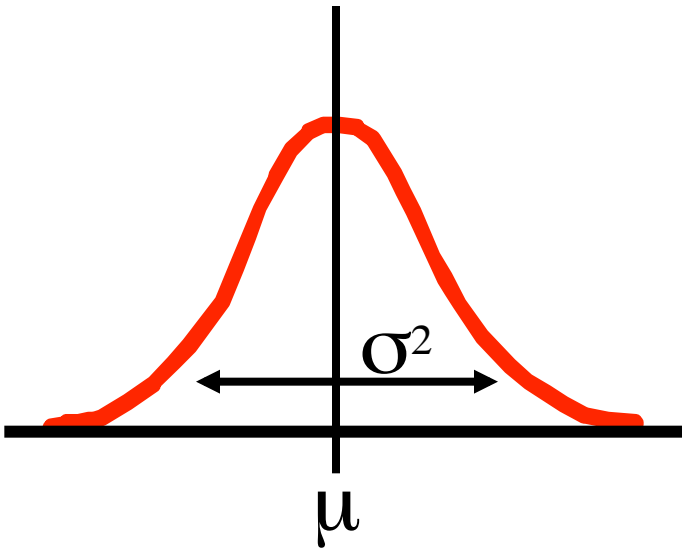
$$h^2 = 2*(r_{MZ} - r_{DZ})$$

r_{MZ}= correlation between monozygotic co-twins

r_{DZ}= correlation between dizygotic co-twins

2 x Difference = σ_a^2 + 3/2 σ_d^2 ≥ h^2

Estimating Heritability: Variance Decomposition



$$\hat{\mu} = \Sigma x_i / n$$

$$\hat{\sigma}^2 = \Sigma(x - \hat{\mu})^2 / n$$

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2$$

$$\sigma_e^2 = \sigma_c^2 + \sigma_{eu}^2$$

σ_p^2 = total phenotypic

σ_g^2 = genetic

σ_e^2 = environmental

σ_a^2 = additive genetic

σ_d^2 = dominance

Almasy & Blangero, *Am J Hum Genet*, 1998

Defining Heritability (h^2)

- Heritability (h^2): the proportion of the phenotypic variance in a trait attributable to the additive effects of genes.

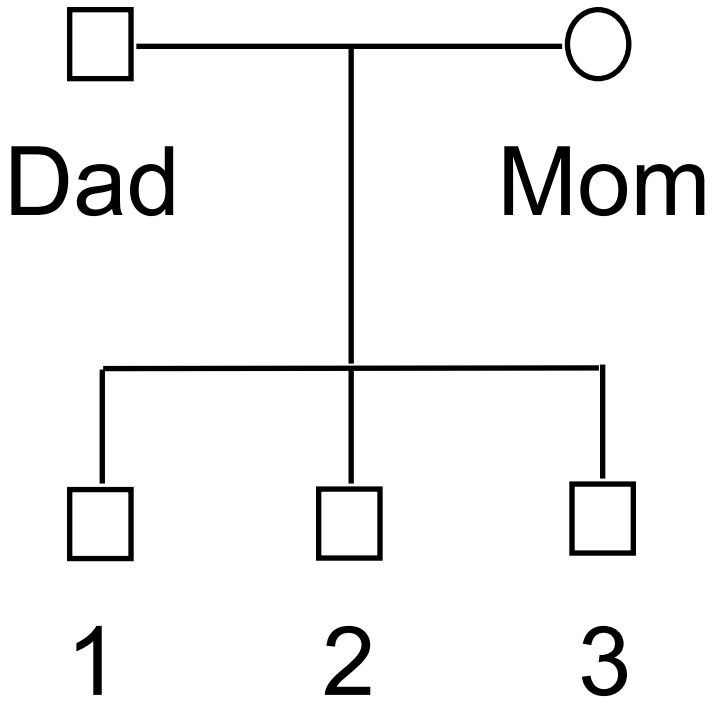
$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Defining Mendelian Relationships

Relatives	Covariance	Heritability
Parent-child	$1/2\sigma^2_a$	$b = 1/2\ h^2$
Half siblings	$1/4\sigma^2_a$	$r = 1/4\ h^2$
Full siblings	$1/2\sigma^2_a + 1/4\sigma^2_d$	$r \geq 1/2\ h^2$
Cousins	$1/8\sigma^2_a$	$r = 1/8\ h^2$

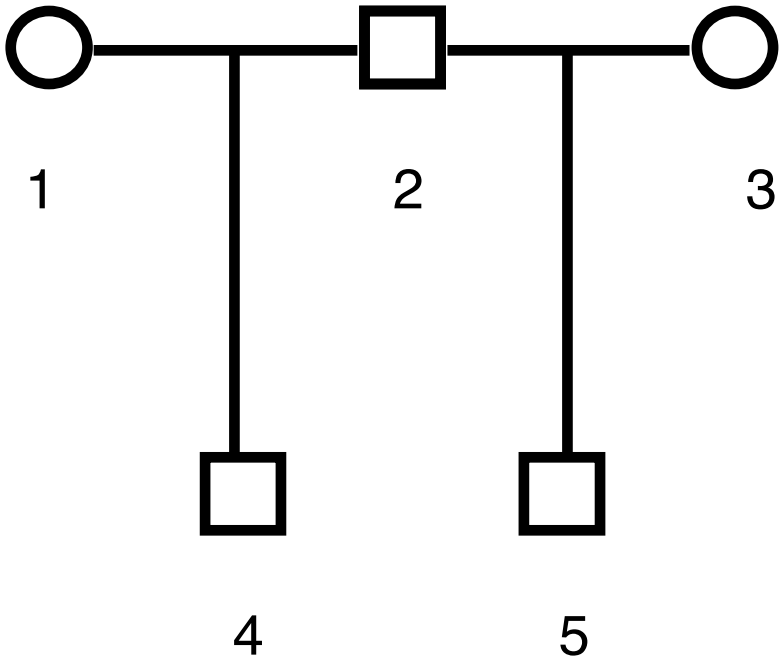
ignoring interactions and shared environmental effects

Simple Kinship Matrix



	D	M	1	2	3
D	1	0	½	½	½
M	0	1	½	½	½
1	½	½	1	½	½
2	½	½	½	1	½
3	½	½	½	½	1

Kinship Matrix



	1	2	3	4	5
1	1	0	0	½	0
2	0	1	0	½	½
3	0	0	1	0	½
4	½	½	0	1	¼
5	0	½	½	¼	1

Limitations of Heritability Estimates

1. Heritability is a population level parameter, summarizing the strength of genetic influences on variation in a trait among members of the population. It doesn't tell you anything about particular individuals.

2. Heritability is an aggregate of the effects of multiple genes. It tells you nothing about how many genes influence a phenotype. A high heritability is not necessarily 'better' if it is due to many, many genes.

Relative Risk

The risk to a relative of an affected individual as compared to a randomly chosen member of the population, λ. 1 - ∞

Heritable vs. Familial

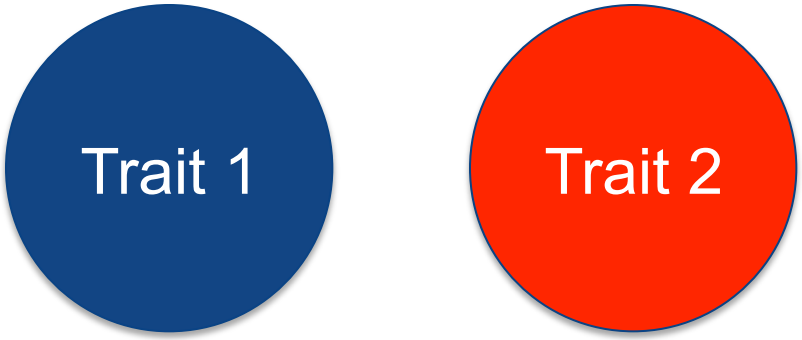
- A trait can be familial without being heritable.
- Genes are not the only thing shared by members of a family (e.g. diet, exercise, environmental exposures, etc).
- Nuclear families vs. extended pedigrees
- Adoption studies

Question 2: Pleiotropy

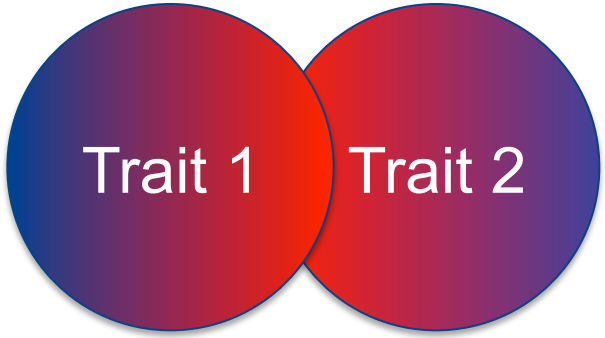
Which traits are influenced by the same genes?

Levels of Pleiotropy

No Pleiotropy

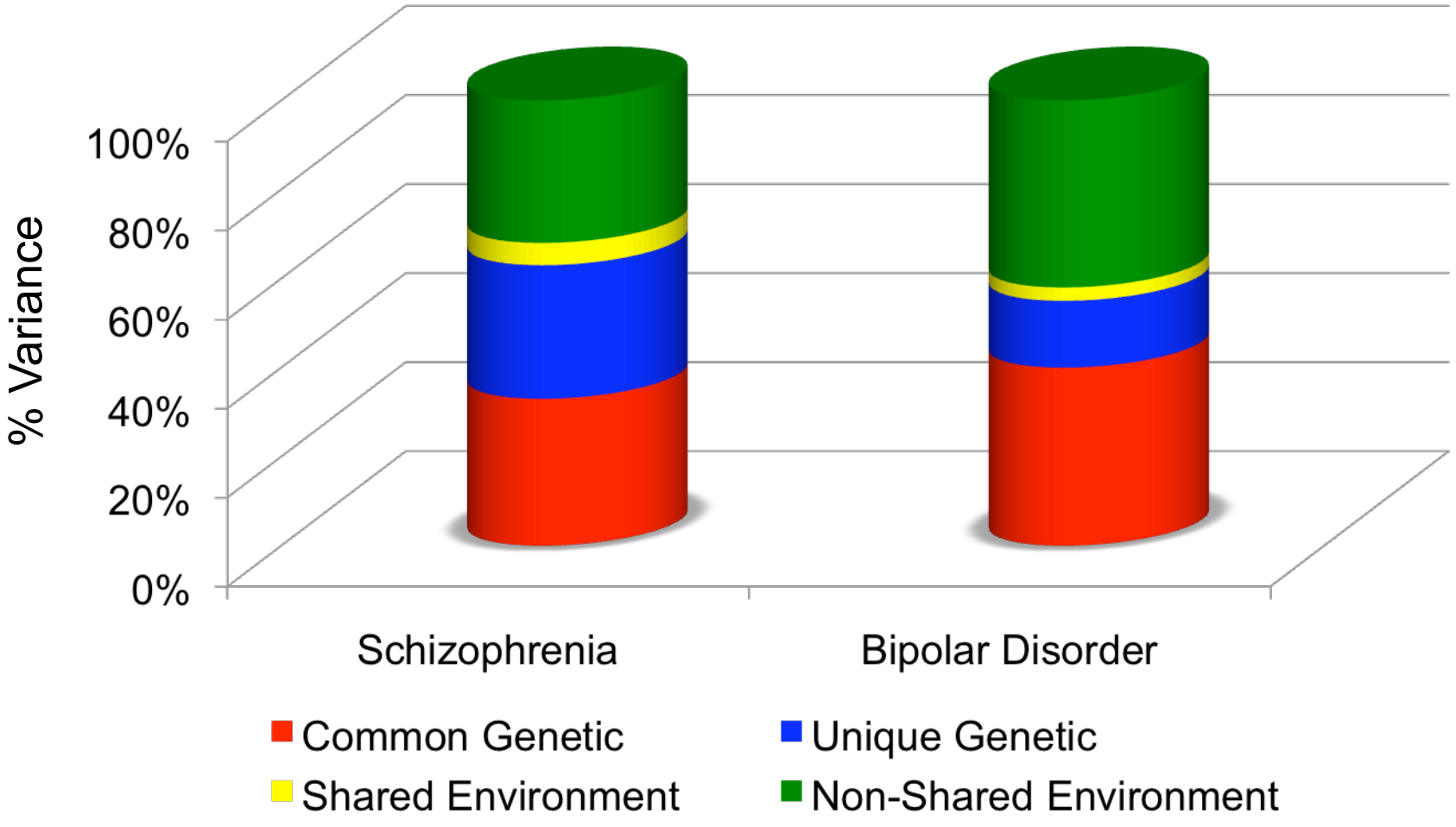


Partial Pleiotropy



Full Pleiotropy

Common Genetic Determinants of Schizophrenia & Bipolar Disorder



Lichtenstein et al., *Lancet*, 2009

Estimating Pleiotropy: Genetic Correlation

- Genetic correlation (ρ_g): a measure of the overlap in genetic effects between traits.
- ρ_g varies from -1 to 1
- 0 = no pleiotropy; -1 or 1 = complete pleiotropy

Question 3: Localization

Where are the genes that influence a trait?

Two Common Methods for Gene Localization

Linkage analyses: test for co-segregation of phenotype and genotype within families - a function of physical connections of genes on chromosomes

Association analyses: test for deviations of phenotype-genotype combinations from that predicted by their separate frequencies - a function of linkage disequilibrium created by population history

Association Studies

Pro: Can be done with unrelated individuals. Statistical methods easy and fast. May be able to detect loci of smaller effect with a given sample size.

Con: Requires disequilibrium, which may not be present, making power difficult to estimate. Susceptible to allelic heterogeneity.

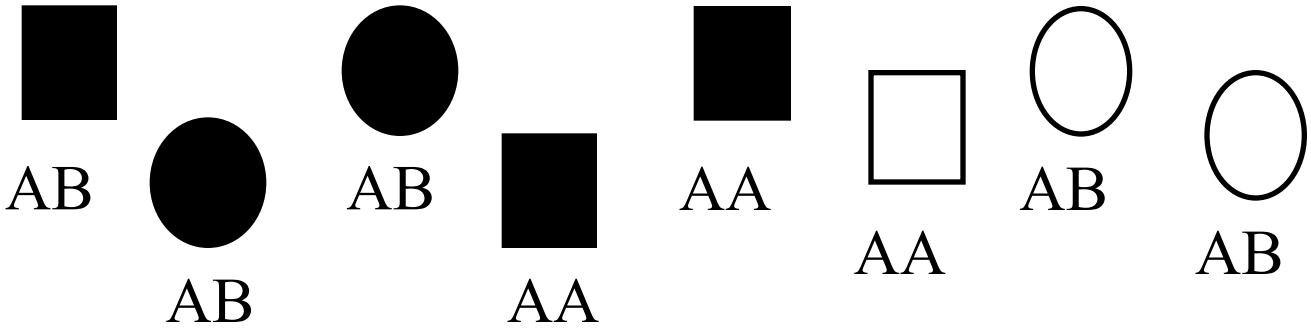
Linkage Disequilibrium (LD)

- Linkage disequilibrium is the non-random association of alleles at two or more loci
- LD = the population frequency of allelic combinations – the expected combinations from random formation of haplotypes
- Level of LD is influenced by a number of factors including genetic linkage, selection, the rate of recombination, the rate of mutation, genetic drift, non-random mating, and population structure.
- LD is unpredictable

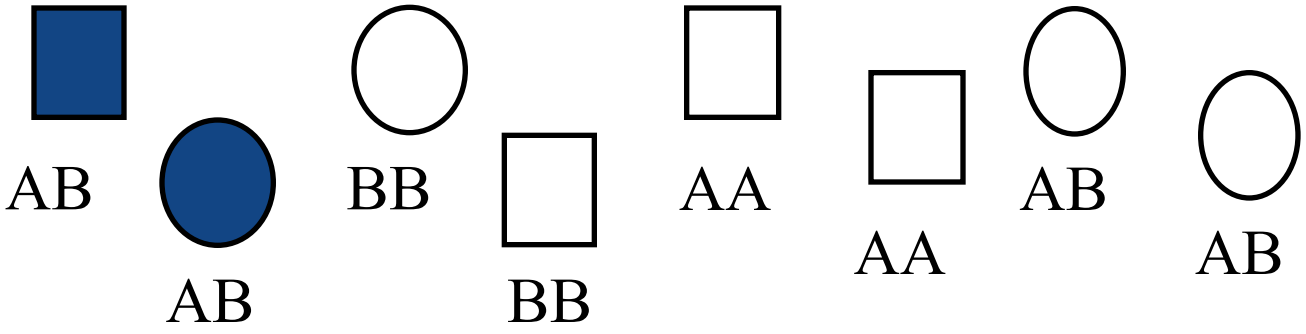
Limitations of Association

- Since LD need not be present, negative association results have implications only for the marker you have tested, lack of association does not exclude the gene or region.
- Population Stratification: If the sample contains multiple populations that differ in the trait of interest, any locus whose allele frequencies differ between the populations will show association

Example: Hypertension

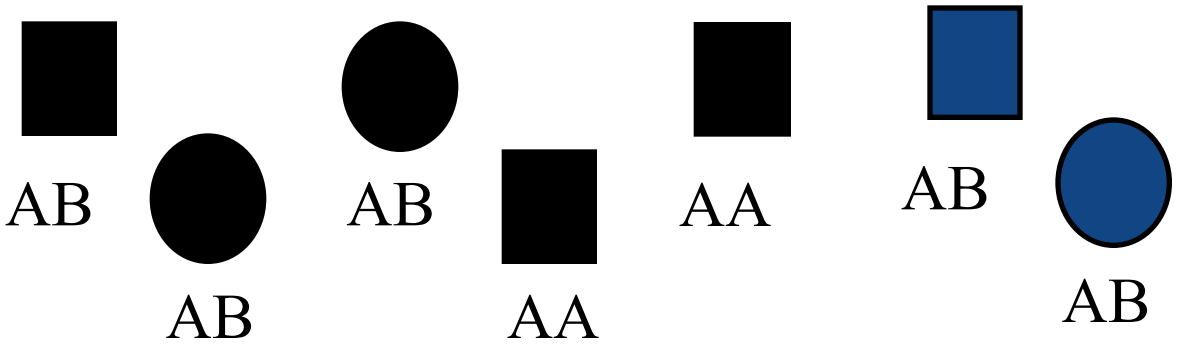


African Americans 70% A, 30% B

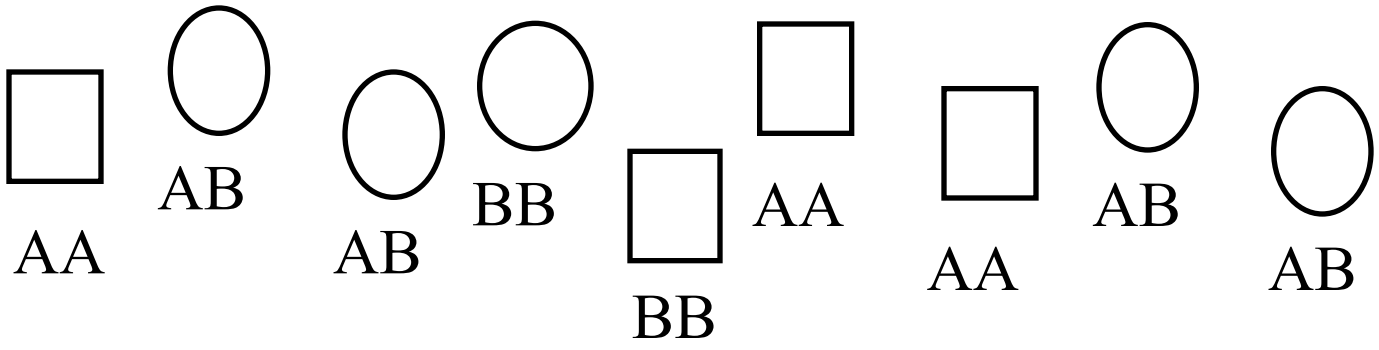


European Americans 50% A, 50% B

Example: Hypertension



Affected
64% A,
36% B



Unaffected
56% A,
44% B

Minimizing Limitations of Association

1. Match cases and controls carefully or try to obtain subjects from a single well defined population.

2. Use one of a variety of statistical approaches designed to deal with population stratification (e.g. TDT, genomic control)

Linkage Studies

Pro: Power to find a gene can be more easily quantified. Guaranteed to work if the sample size is large enough. Not influenced by allelic heterogeneity.

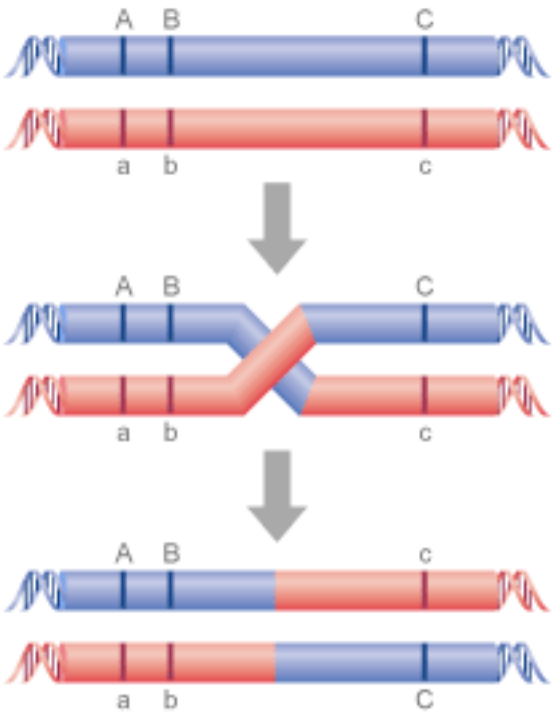
Con: Need a large sample of related individuals. “Large enough” may be too large to be practical.

Genetic Linkage Defined

Genetic loci that are physically close to one another tend to stay together during meiosis.

Independent assortment occurs when the genes on different chromosomes are separated by a great enough distance on the same chromosome that recombination occurs at least half of the time.

An exception to independent assortment develops when genes appear near one another on the same chromosome. When genes occur on the same chromosome, they are usually inherited as a single unit. Genes inherited in this way are said to be linked, and are referred to as "linkage groups."



Measuring Linkage: Lod Score

LOD = $\log_{10}$(probability of birth sequence with a given linkage value/probability of birth sequence with no linkage)

$$LOD = \log_{10}((1-\theta)^{NR} \times \theta^{R})/0.5^{NR+R}$$

NR denotes the number of non-recombinant offspring,
R denotes the number of recombinant offspring.
Theta = recombinant fraction= R / (NR + R)

A LOD score >=3.0 is considered evidence for linkage
A LOD score of 3 indicates 1000 to 1 odds that the linkage being observed did not occur by chance
A LOD score <=-2.0 is considered evidence to exclude linkage

Linkage vs. Association

Association: you're testing for an excess of a specific combination of alleles at two loci. The same alleles must be traveling together at a population level. Detects effects of common variants.

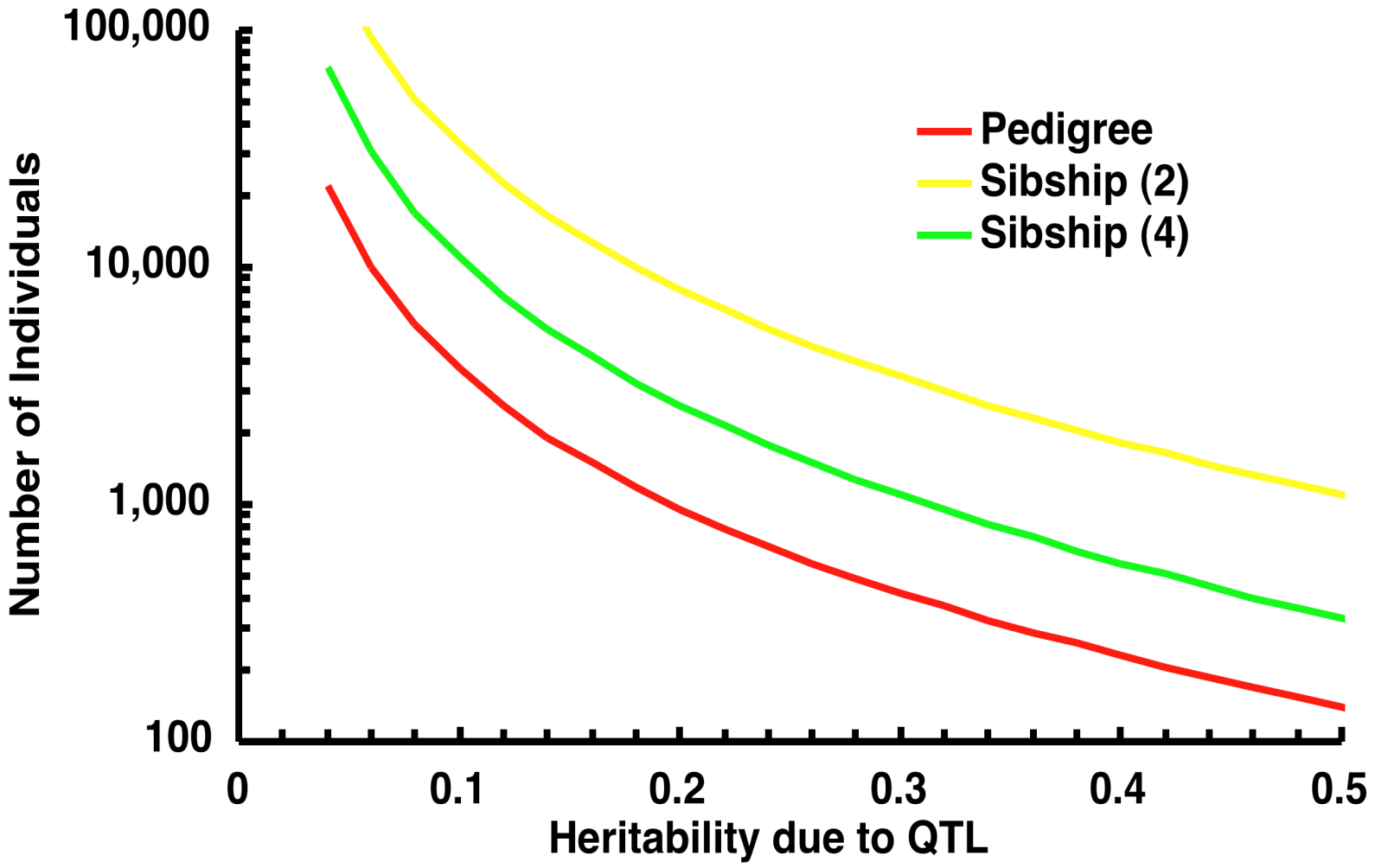
Linkage: you're testing for an excess of the parental type. That parental type (i.e. the alleles traveling together) could be different in every family (i.e. linkage equilibrium) and you would still get linkage. Can detect cumulative effect of multiple variants (including rare variants).

Determining Linkage Power

The power to map a QTL in a human linkage study is a function of:

1. locus-specific heritability (genetic signal-to-noise ratio)
2. Sample size
3. Pedigree size and complexity

Sample size required for 80% power to detect linkage to a QTL at a LOD of 3



Determining Association Power

The power to find association in a human study is a function of:

1. QTN-specific heritability (not QTL)
2. r^2 between the QTN and a genotyped marker
3. Sample size

Approaches to Genotyping

Candidate genes: genotype only markers in genes potentially related to the trait.

- Pro: fast and easy, may be able to be more thorough with a higher density of markers
- Con: must get lucky in choice of genes, lower potential for something really novel

Genome screen: genotype anonymous markers spanning the genome at regular intervals

- Pro: can identify previously unknown genes, covers all of the possibilities
- Con: slower and more expensive, may have lower marker density which could translate to less power

Question 4: Identification

What specific genes influence the trait?

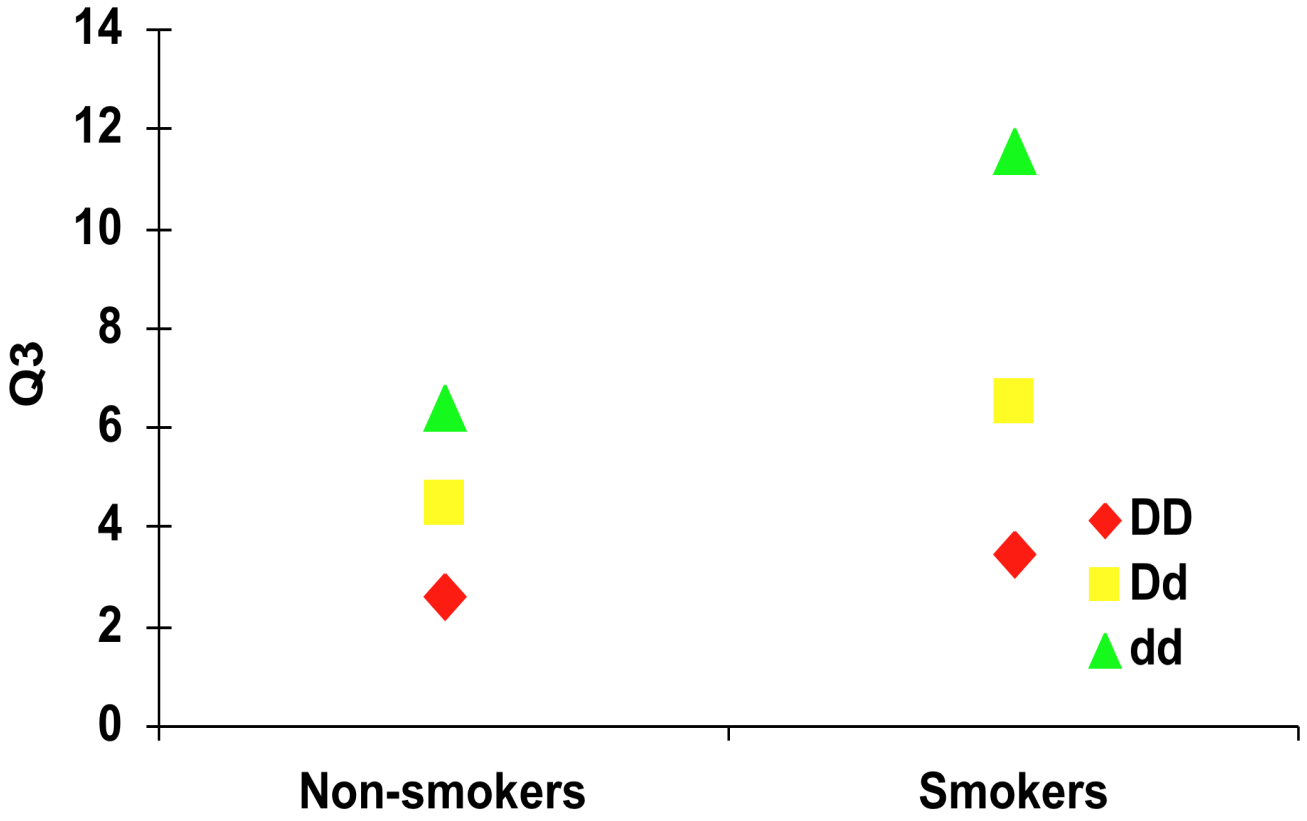
Identifying a Causal Gene

- Once a significant QTL is identified, additional genetic tests are needed to determine the exact identity of the gene
  - Association: identifies a genomic region of ~500kb (250kb to either side of the association) determined by the general extent of linkage disequilibrium
  - Linkage: detect the cumulative additive genetic signal of all functional variants within a much larger genomic region (e.g. 10-15Mb)

Question 5: Characterization

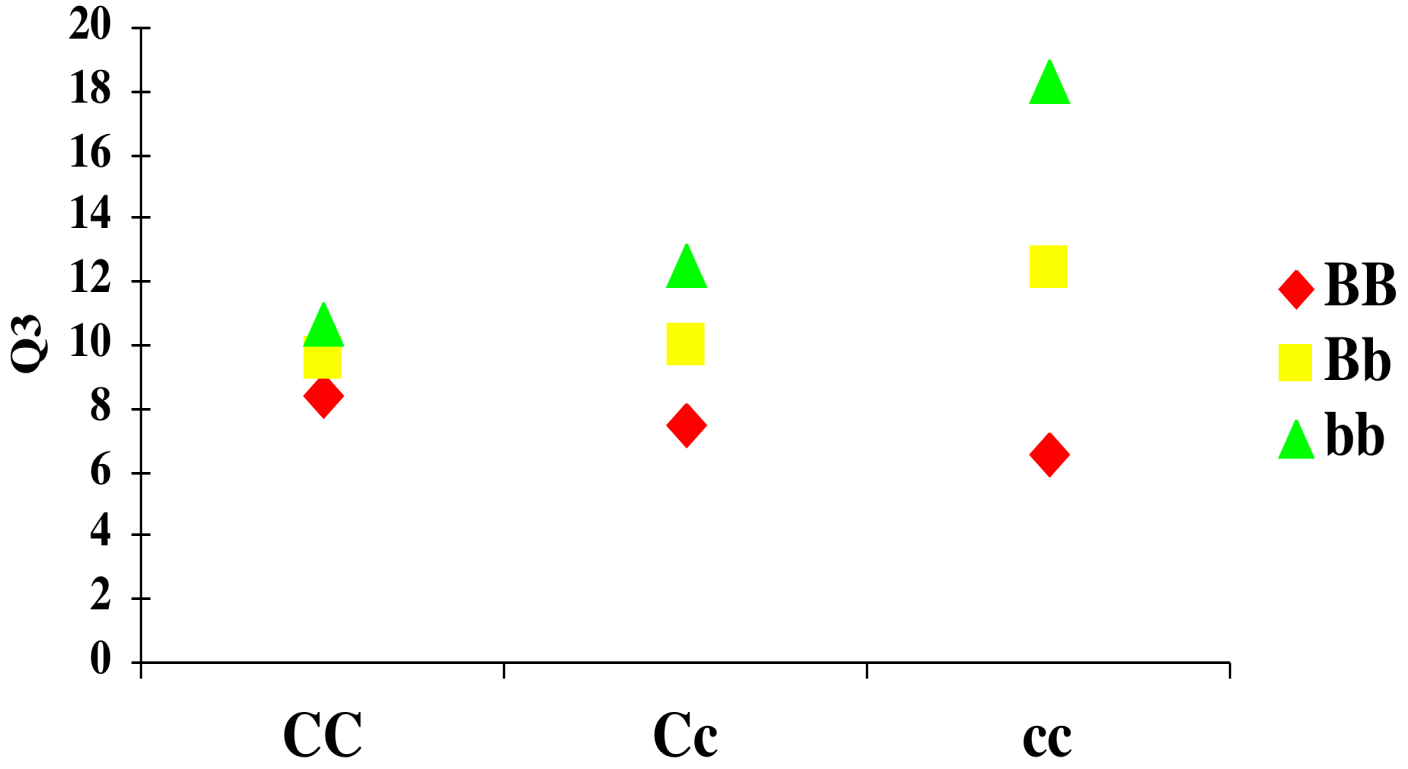
What specific genetic variants influence the trait and how do they interact with each other and with the environment?

Gene x Environment Interaction



The effect of a particular genotype differs in different environments OR the effect of the environment depends on genotype

Epistasis: Gene x Gene Interaction



The effect of a genotype at one locus depends on the genotype you have at another locus.