Quantitative Traits: Heritability, Linkage & Association

L. Elliot Hong, MD

Professor and Director
UM Center for Brain Imaging Research
Maryland Psychiatric Research Center
University of Maryland School of Medicine, USA
www.mdbrain.org

Learning Objectives

1. Apply genetic epidemiology concepts in quantitative brain imaging - genetics studies
2. Common vs. rare variant hypotheses
3. Heritability: Falconer's, ACE twin model, family based, missing heritability
4. Linkage vs. association studies

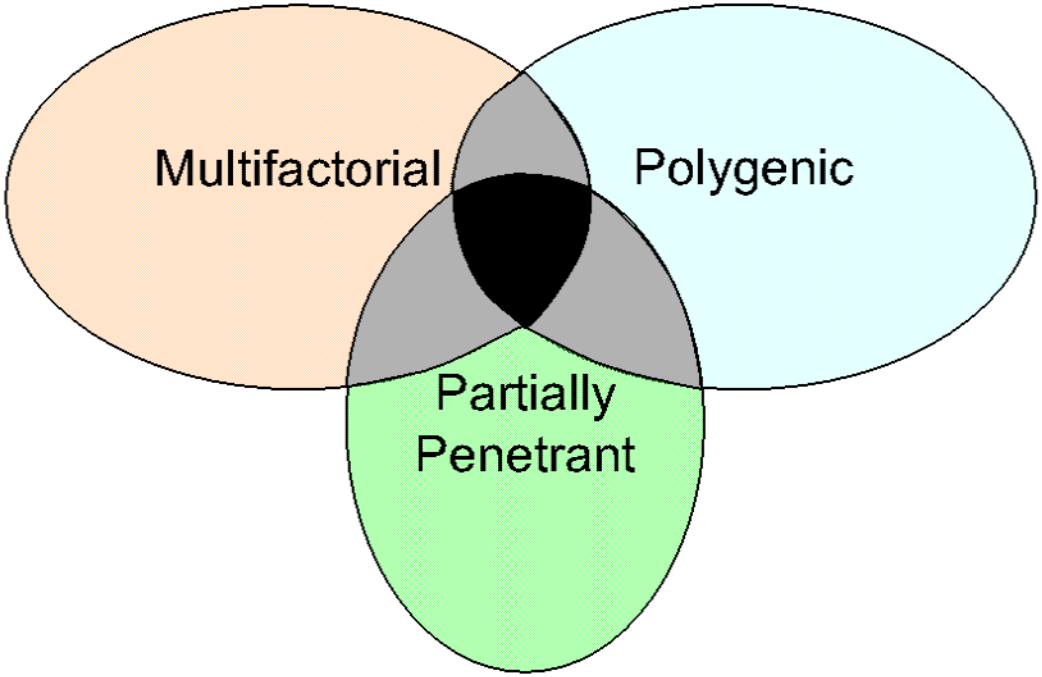
Basic, conceptual discussion. Will have more details on each topic covered by experts in the following lectures

Three concepts used in describing complex diseases

Polygenic: the disease occurs only if several genotypes are present together

Multifactorial: several distinct genes (or sets of genotypes) can independently cause the trait

Partially penetrant: partially or poorly penetrant; nongenetic factors may also be required, or other genes may be required



Complex Trait / Disease

The common disease/common variant hypothesis:

- Each of the liability genes (i.e., variants) has a small effect, but by itself would not cause the trait or the illness.
- Disease cases are due to liability alleles exist in the population at certain frequency.
- The disease is said to be polygenic.
- Genome-wide association studies (GWAS) are based on common variant hypothesis

Candidate Gene Approaches (Hypothesis-driven)

Candidate Genes

Quantitative Trait

Replication

Adopted from TA Pearson

Genome-Wide Association (GWAS) (Discovery or Agnostic)

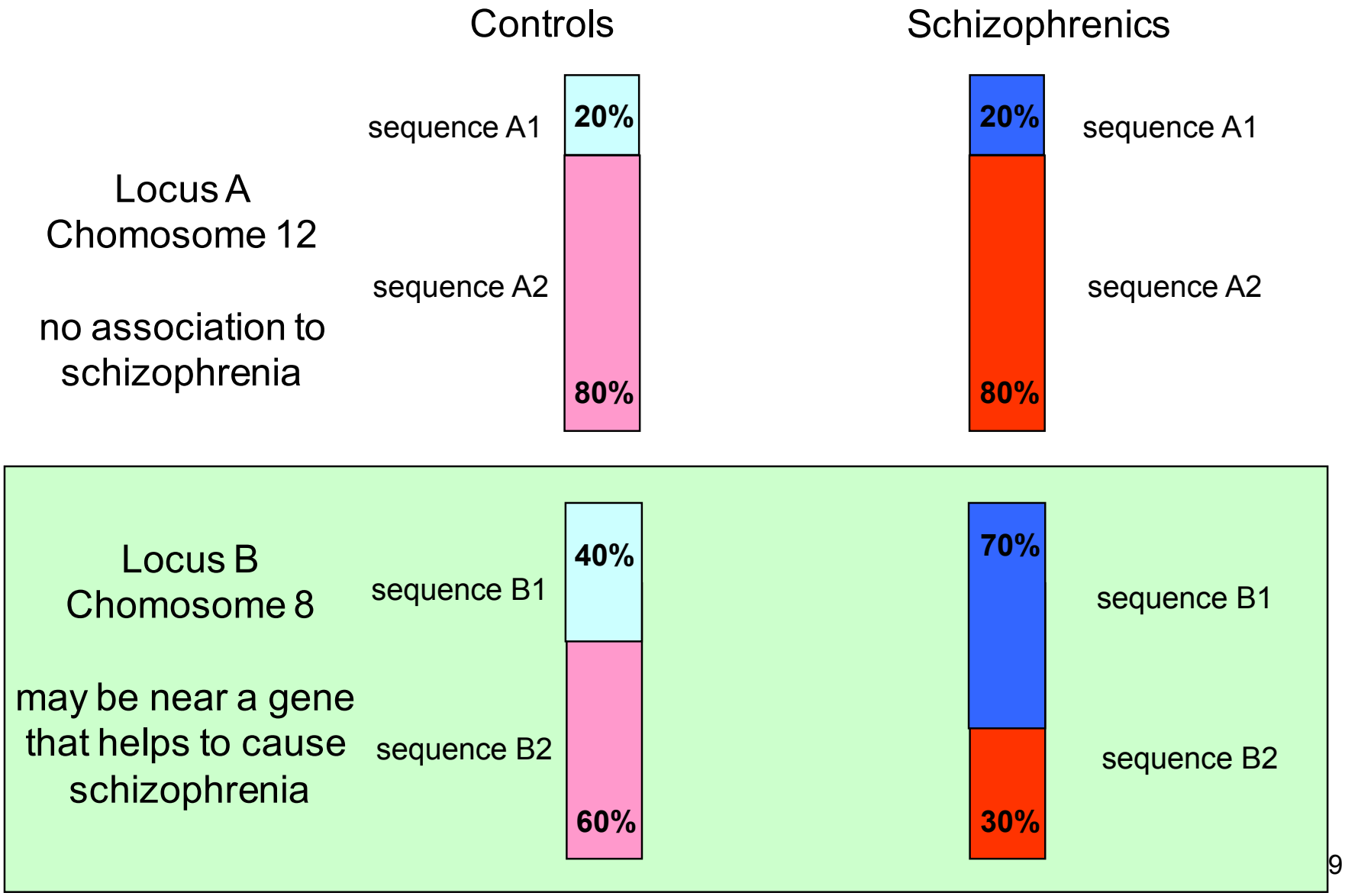
Entire Genome

Quantitative Trait

Replication

Adopted from TA Pearson

Association study to identify candidate loci



Complex Trait / Disease

The common disease/common variant hypothesis:

- GWAS is to repeat the association study many times across the genome.
- Variants in GWAS were selected based on common variants in the population (about 5% minor allele frequency)
- With the newest chips of 4.8 million SNPs: about 1% MAF
- Imputation can increase the coverage further
- GWAS is based on the assumption that common genetic variation plays a sizable role in the heritable variations of common disease

Study Designs for Association Analysis

- Family-Based
  - Parent-child Trio
  - Discordant sibpairs
- Advantages
  - Uses the non-transmitted alleles as the control (trio)
- Disadvantages
  - Ascertainment & Power

- Case-Control
- Advantages
  - Ascertainment & Power
- Disadvantages
  - Sensitivity to assumptions
  - Matching

- With few exceptions, most brain-related traits are assumed to be from multiple genetic factors working in aggregate
- The debate is on the frequency of the gene variants

Complex Trait / Disease

The common disease/rare variant hypothesis:

- Rare in the population, but each of the liability genes or variants can be highly penetrant, or have a large effect.
- Allelic heterogeneity: each disease-causing gene has many rare variations that disrupt that gene in about the same way
- Disease cases are due to liability alleles at very low frequencies, detectable by sequencing.
- The disease is said to be multifactorial and polygenic.

Schork et al 2009 Curr Opin Genet Dev

Complex Trait / Disease II

The common disease/rare variant hypothesis:

- Where do the rare variants come from?

  - Old. Being selected against, or

  - New. Not enough time to spread

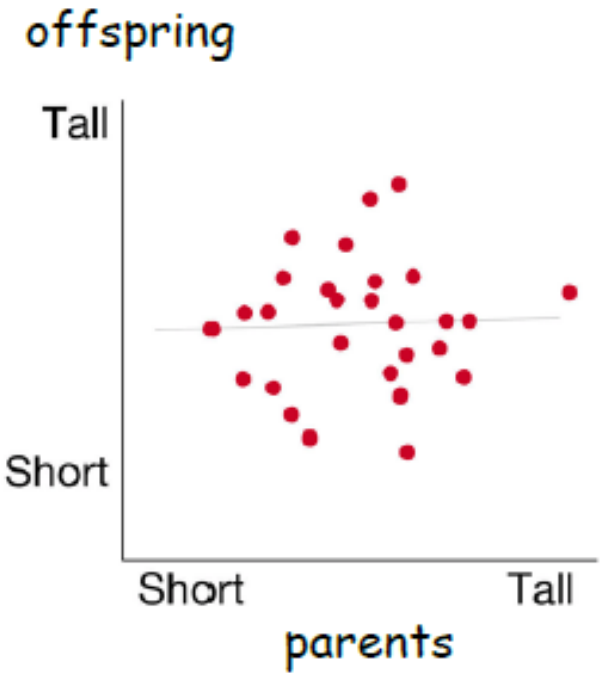
- Examples: Breast cancer genes BRCA1 and BRCA2. Hundreds of rare, disease-causing mutations have been identified.

- Motivation for sequencing rather than GWAS

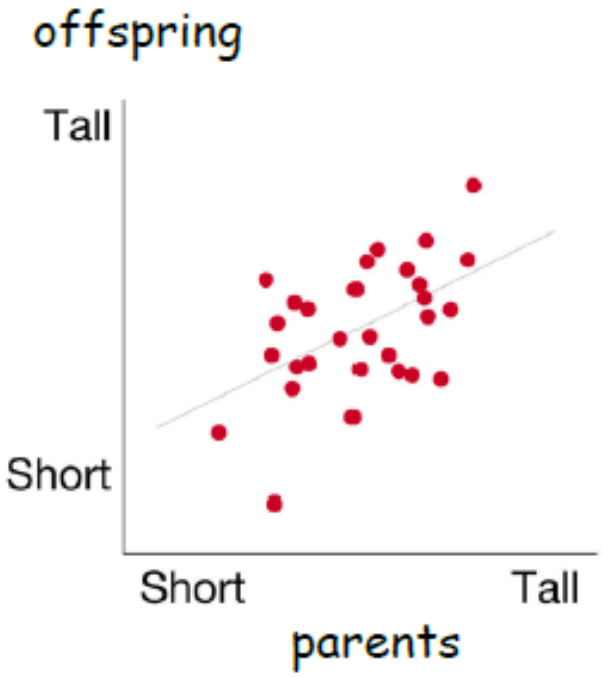
Heritability (h^2)

h^2 is the regression (slope) of offspring on parents

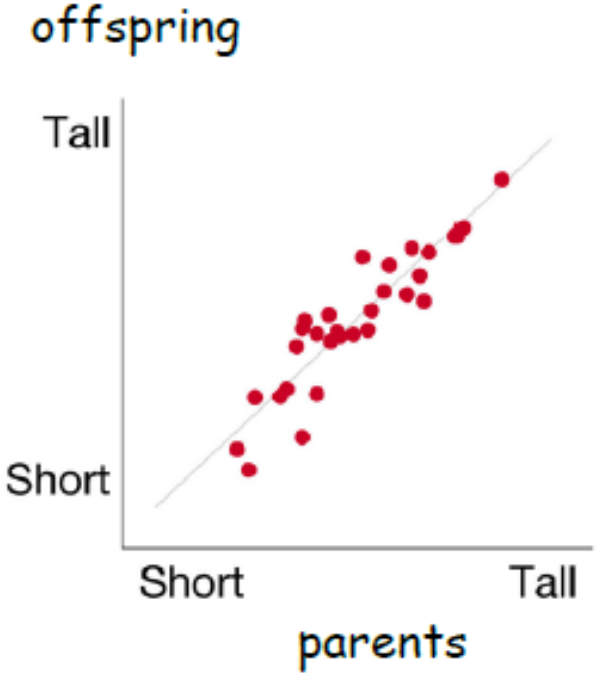
$h^2 \approx 0$



$h^2 \approx \frac{1}{2}$



$h^2 \approx 1$



From Alan R. Rogers, University of Utah

How to find the genetic basis of quantitative imaging variation?

BROAD-SENSE HERITABILITY (H^2)

The proportion of a trait's variation that is due to genetics. The rest of it due to "environmental" factors.

$$V_P = V_G + V_E$$

V_P	V_G	V_E
phenotypic variance	genetic variance	environmental variance

$$H^2 = V_G / V_P$$

Twin based heritability

Falconer's formula $$h^2_{pop}(ACE) = 2(r_{MZ} - r_{DZ})$$

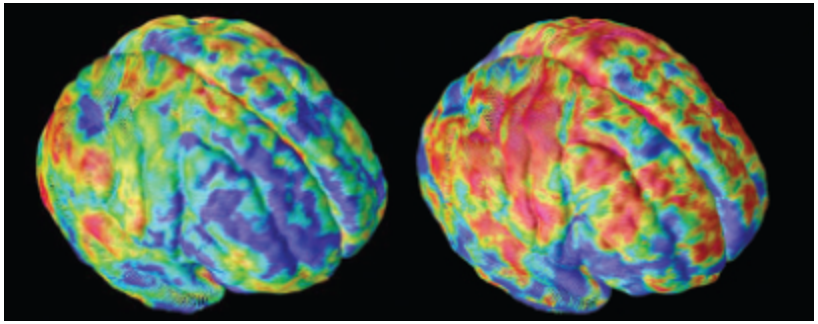
This is the broad sense heritability.

rmz is the monozygotic (MZ) identical twin correlation

rdz is the dizygotic (DZ) fraternal twin correlation

Heritability Estimate

$$\text{Heritability} = 2(r_{mz} - r_{dz})$$



Heritability of Brain Volume = 2(.90 - .45) = .90

Thompson et al 2001 Nat Neuroscience

Twin based heritability

The ACE model:

1. Assumption: shared environment is MZ = DZ.
2. Caveat: If there are differences in the shared environment due to in-utero or early development effects (likely MZ > DZ), then the ACE model will over-estimate h2.
3. Many additional steps are taken to minimize or eliminate this bias

$$h^2_{pop}(ACE) = 2(r_{MZ} - r_{DZ})$$

Additive vs. Dominance Genetic Variance

- H2 is problematic because all genetic factor is represented by Vg.
- Genetic variances includes Vg = Va + Vd

- Va: trait variance due to the effects of each allele adding together
- Vd: variance due to interactions between alleles

They can be:

1) Interaction of alleles of the same gene

2) Epistasis: interaction of the alleles between difference genes

Additive vs. Dominance Genetic Variance
continues….

- Vd is not directly inherited from parent to offspring. It is due to the interaction of genes from both parents
- The end results: when interaction is not accounted for, there is a risk of inaccurate estimate

How to find the genetic basis of quantitative imaging variation?

NARROW-SENSE HERITABILITY (h^2)

Only additive genetic variance is used

$$V_P = V_A + V_D + V_E$$

- V_P: phenotypic variance
- V_A: genetic variance
- V_D: dominance genetic variance
- V_E: environmental variance

$$h^2 = V_A / V_P$$

Heritability = 0 ≠ Not genetic

Every living adult has a brain. The h2 of having a brain is ??

If 100% genetic and no variance, then no heritability to calculate, according to current genetic epidemiology

Twin based heritability - Issues

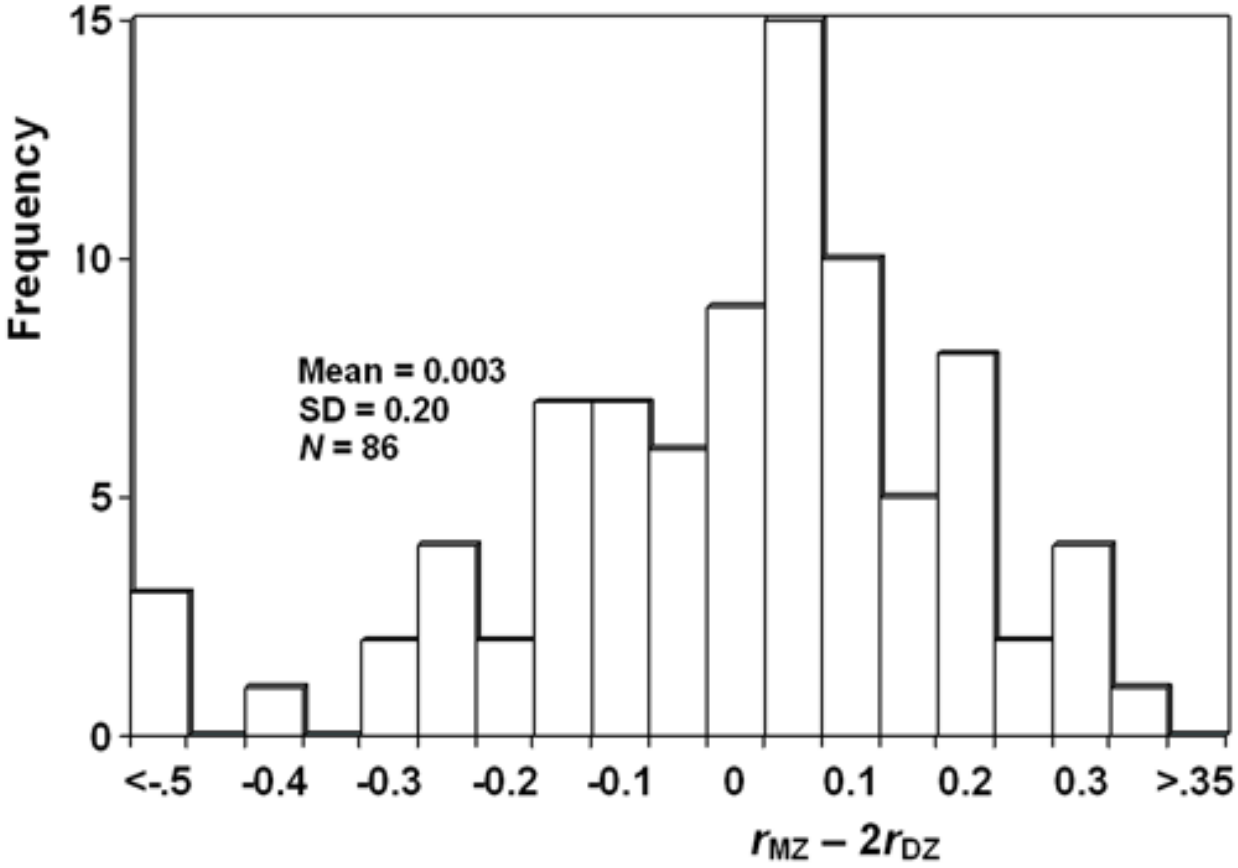
Potentially problematic assumptions, chiefly among them:

1. The environments of identical twins = the environments of non-identical twins. But possibly mz > dz
2. Genetic risks are combined together in a simple way. But possible there are interactions.
3. Ascertainment bias: Co-twin with disease is more likely to participate in twin studies as compared to unaffected co-twin.

Twin based heritability – may not be that bad

Across a wide variety of traits rmz – 2 * (rdz) = 0.

Consistent with predominantly additive genetic variance and the absence of a large bias from common environmental effects



Hill et al 2008 Plos Genetics

Family Based Heritability (h^2)
Variance components

$$V_P = V_a + V_c + V_u + X_i$$

V_P	V_a	V_c	V_u	X_i
Phenotypic variance	Additive genetic variance	Common environmental variance	Unique environmental variance	Covariates

$$h^2 = V_A / V_P$$

Genetic vs. Environmental Correlations in Families -- Information for Family Based Heritability (h^2)

Relationship	Genetic Correlation	Environmental Correlation		
		Children	Adult	Family
Parent-child	0.5	0	0	1
Sib-sib	0.5	1	0	1
Spouse-Spouse	0	0	1	1
Half-sibs, parernal	0.25	0	0	0
Half-sibs, maternal	0.25	1	0	1
Cousins	0.125	0	0	0

Linkage Analysis

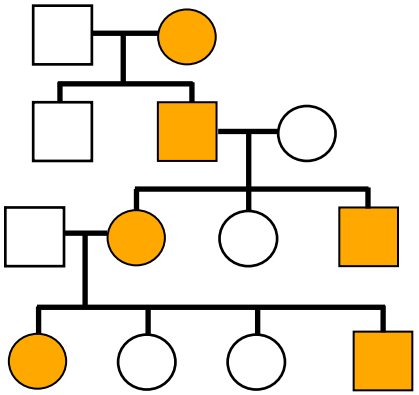
- Traces the segregation of the trait through a family
- Traces the segregation of the chromosomes through a family
- Statistically measures the correlation of the segregation of the trait with the segregation of the chromosome

Linkage vs. Association

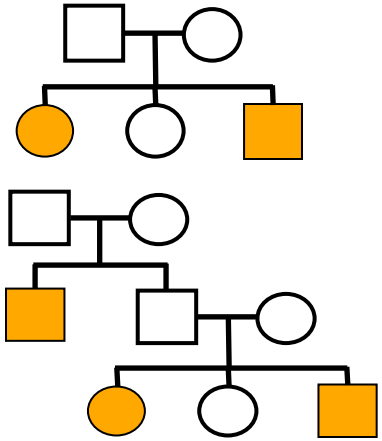
Linkage Analysis

Shared within families

Large Families



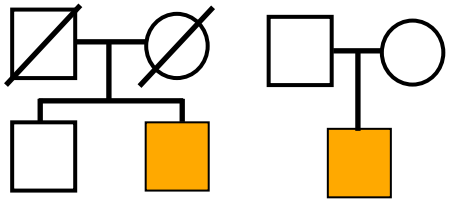
Small Families



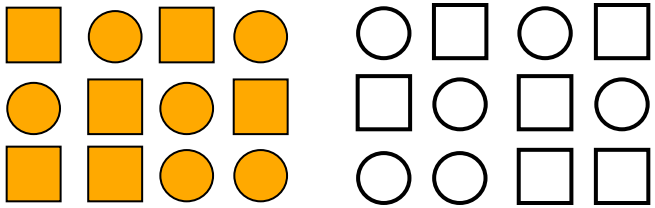
Association Studies

Shared across families

Family-Based



Case-Control



Adapted from Jonathan L. Haines Ph.D.

Association vs. Linkage

Association

- Population based
- candidate gene search and verification
- easier to perform

Linkage

- Family-based
- Mono- or oligogenic disorders
- Mendelian inheritance
- Position identification

Missing Heritability

- Heritability of common traits and major disorders are in the range of 50% to 90%
- GWAS reveal small variances of these traits are explained genetic variants (from 0.01% to 1-2% for single SNP, to 10-30% in comprehensively studied cases)
- The resting the heritability is ‘missing’

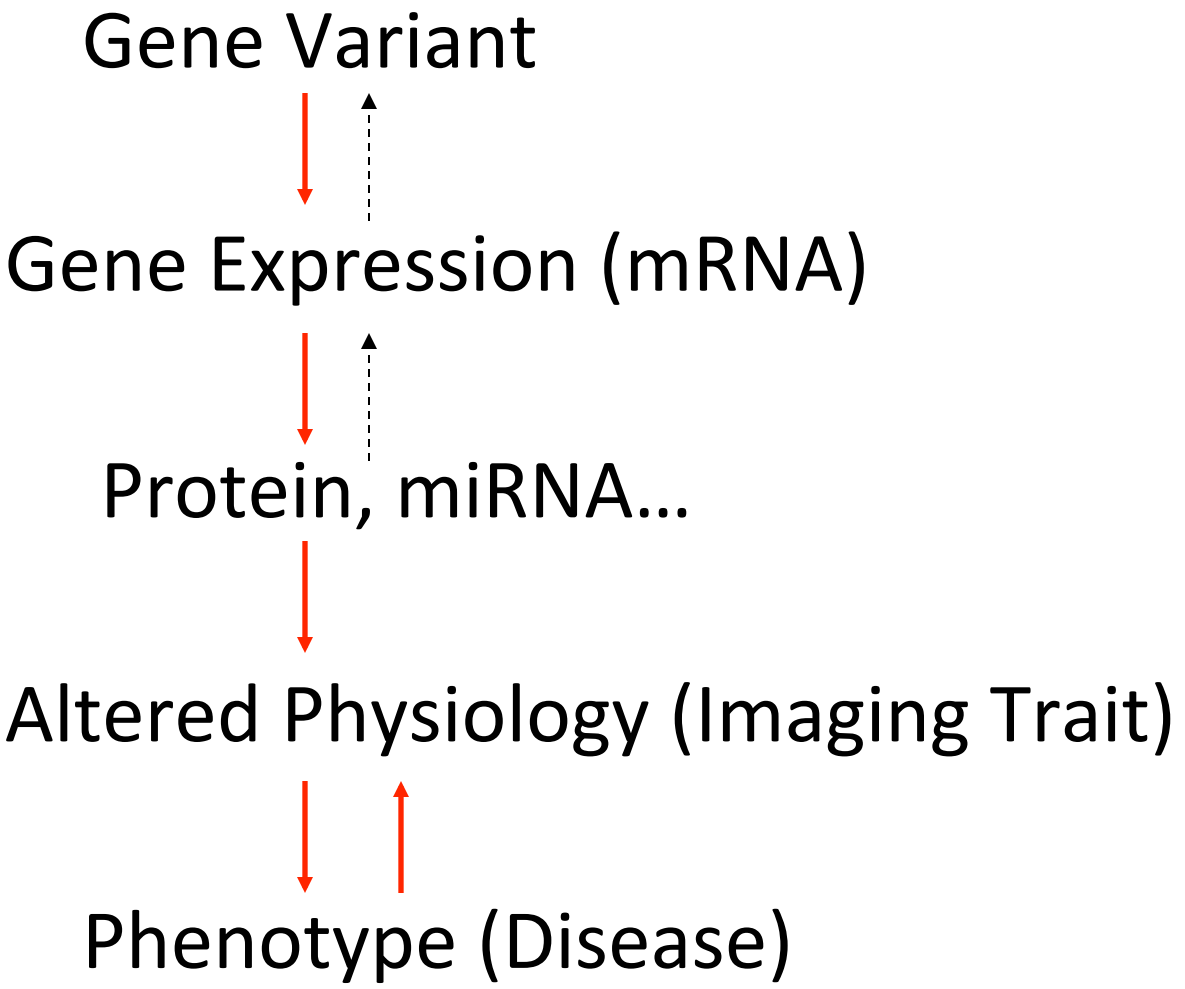
Zuk et al 2011 PNAS
Zuk et al 2013 PNAS

Missing Heritability – possible hiding places

- More common variants are there to be discovered
- Rare variants not shared by individuals – the 'missing heritability' may come from rare variants
- 'Over-estimated h2' due to interactions – recall how twin based h2 is calculated.
- Gene x gene or gene x environment interactions

Zuk et al 2011 PNAS
Zuk et al 2013 PNAS

The Genetic Origin of Neuroimaging Quantitative Trait



Adopted from TA Pearson