

Imaging Genetics: Heritability, Linkage & Association

David C. Glahn, PhD

Olin Neuropsychiatry Research Center &
Department of Psychiatry, Yale University



Yale
UNIVERSITY

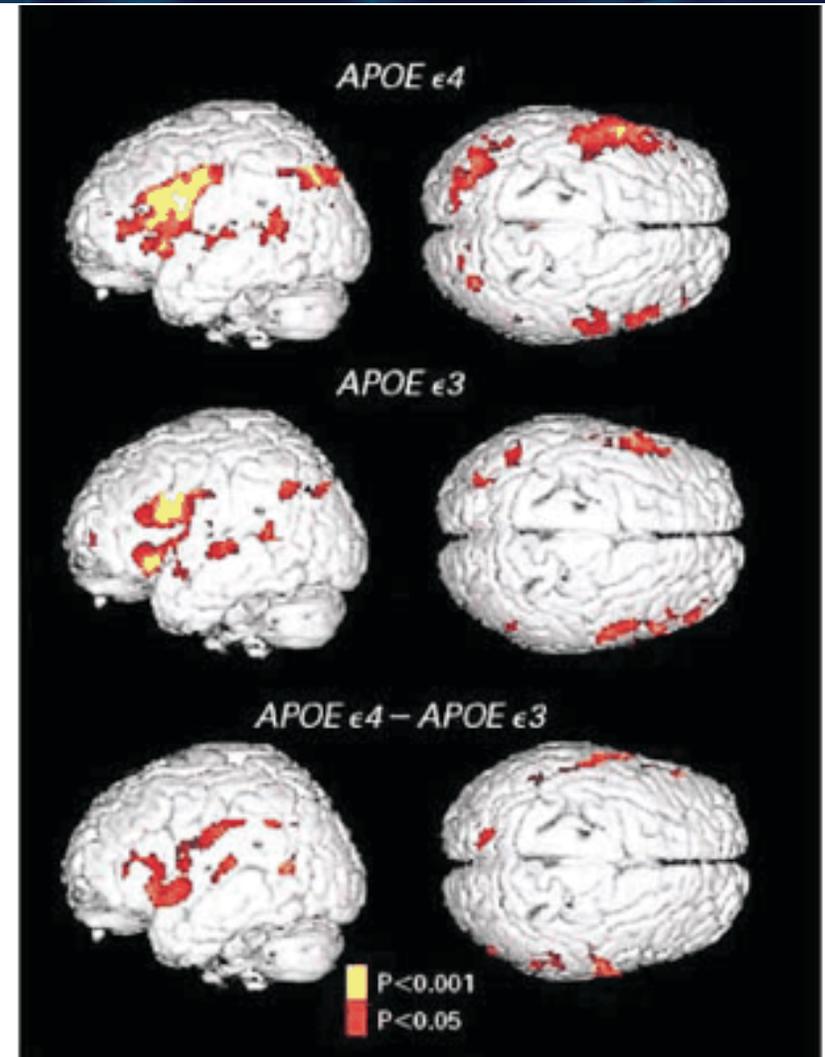


THE
INSTITUTE
OF LIVING

July 17, 2011

Memory Activation & APOE ϵ 4

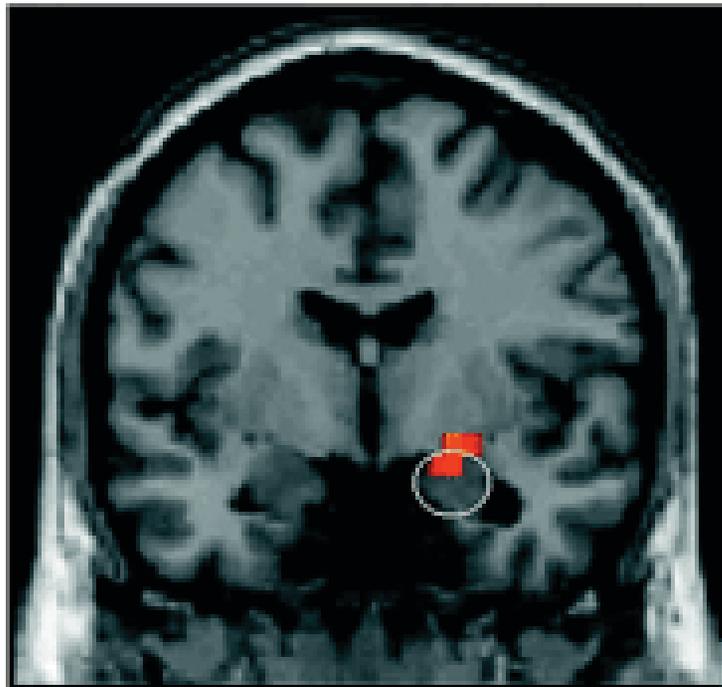
- Risk gene for Alzheimer's
- 16 APOE ϵ 4
- 14 APOE ϵ 3



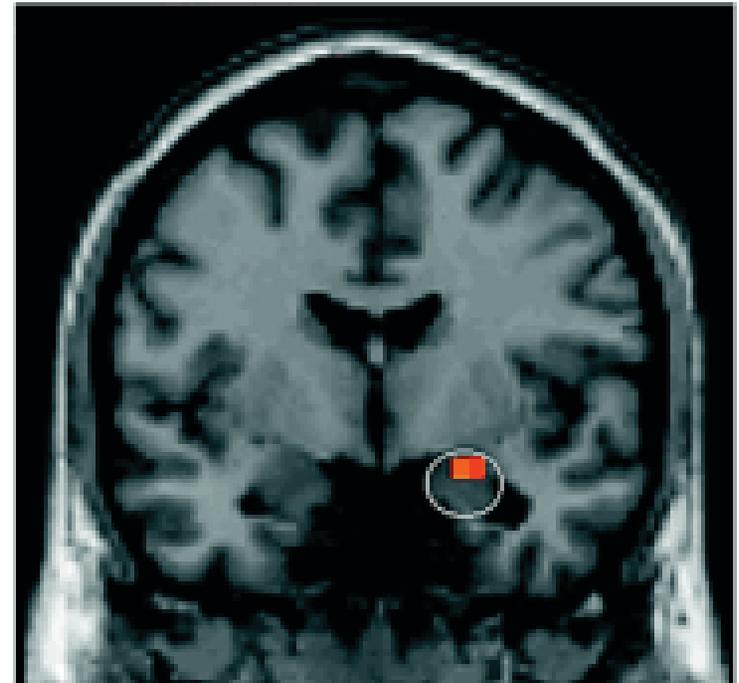
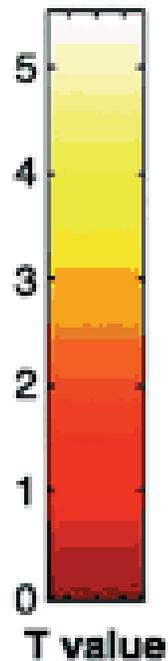
Bookheimer et al., *N Engl J Med*, 2000

Fear Response & Serotonin Transporter Gene (SLC6A4)

Amygdala Response: s Group > l Group



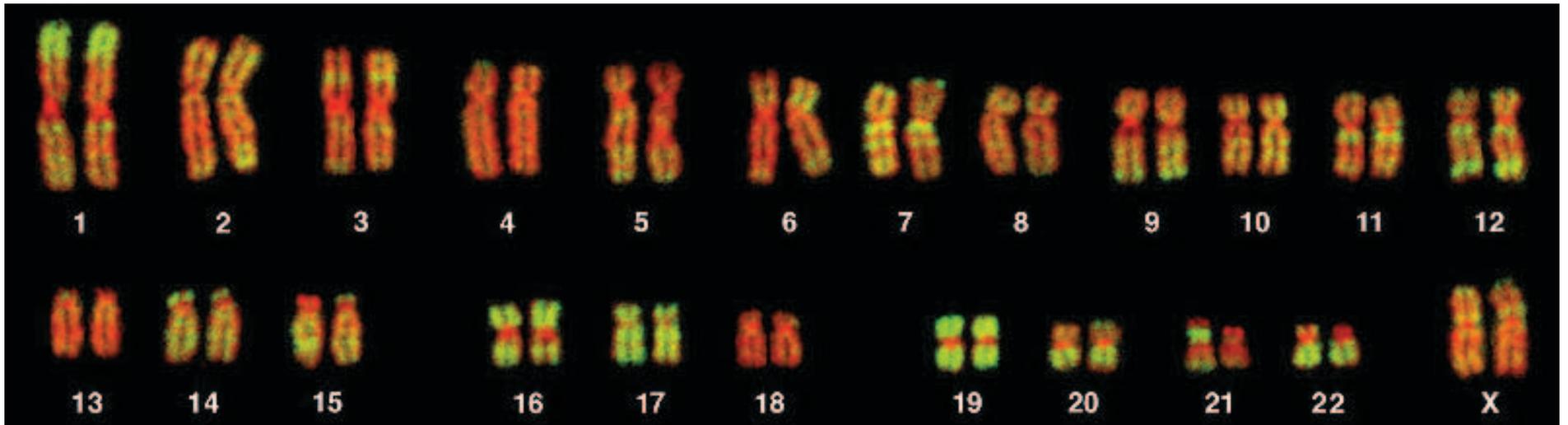
**First Cohort
(N = 14)**



**Second Cohort
(N = 14)**

Hariri et al., *Science*, 2002

Human Genome



23 Chromosomes
~20-25,000 genes
~3 billion base pairs

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

Questions for the Study of Complex Trait Genetics

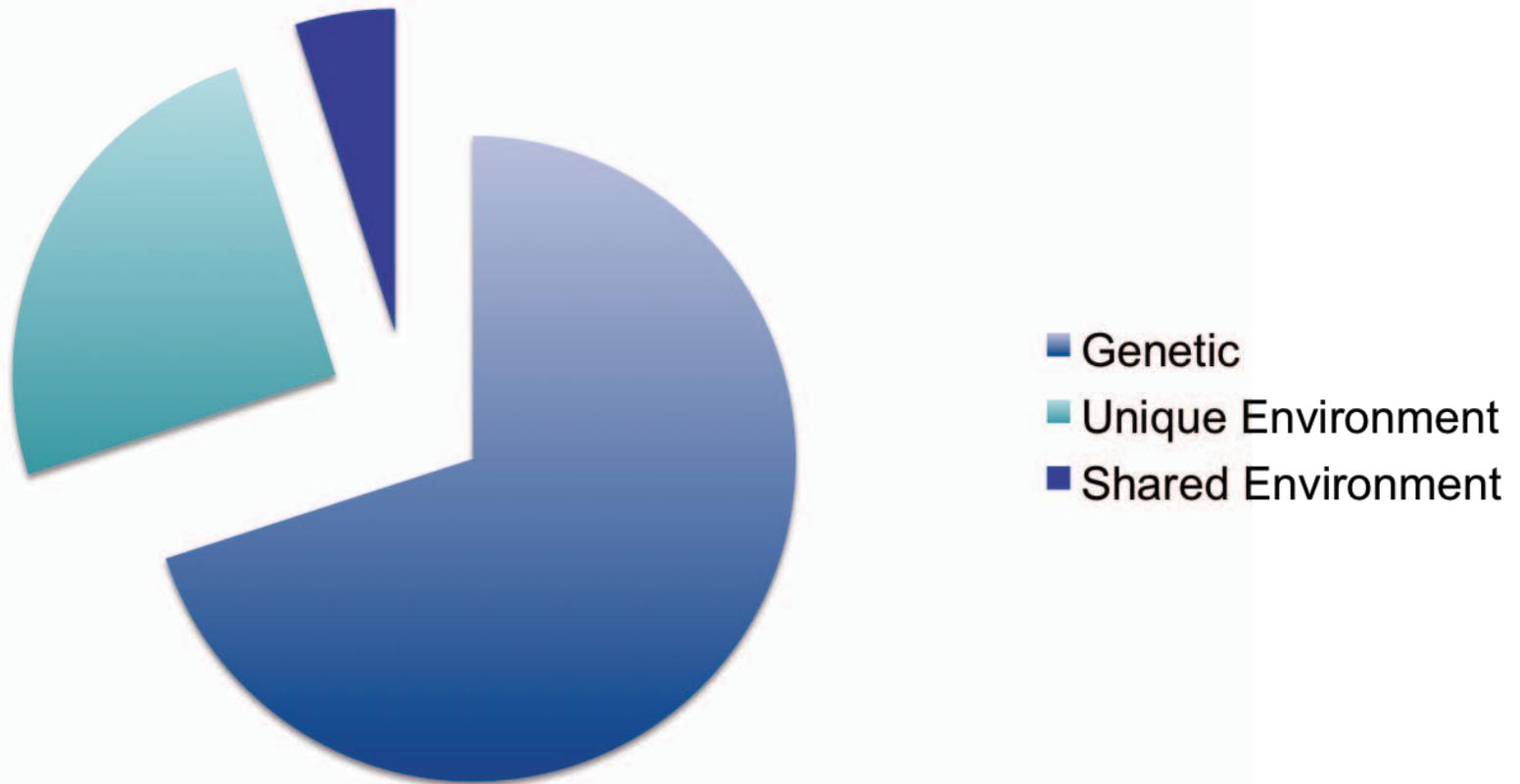
- 1) Is this trait influenced by genetic factors? How strong are these genetic influences?
- 2) Which traits are influenced by the same genes?
- 3) Where are the genes that influence a trait?
- 4) What are the specific genes that influence the trait?
- 5) What specific genetic variants influence the trait and how do they interact with each other and with the environment?

Question 1: Heritability

**Is this trait influenced
by genetic factors?**

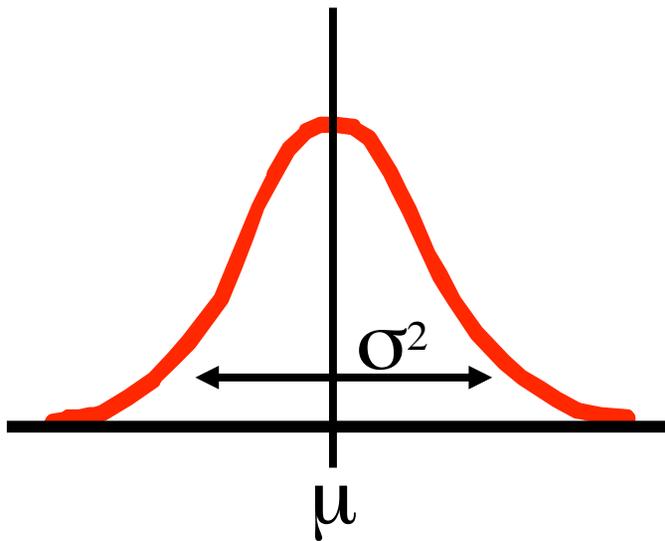
**How strong are these
genetic influences?**

Defining Heritability



Phenotype (P) = Genotype (G) + Environment (E)

Variance Decomposition



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

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

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

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

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

σ_p^2 = total phenotypic

σ_g^2 = genetic

σ_e^2 = environmental

σ_a^2 = additive genetic

σ_d^2 = dominance

Narrow-Sense Heritability (h^2)

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

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

Conceptualizing Heritability

- Heritability estimates vary between 0 and 1
 - 0, genetic factors do not influence trait variance
 - 1, trait variance is completely under genetic control
- If $h^2=0.5$, then 50% of phenotypic variation is due to genetic variation.
 - Not that the trait is 50% caused by genetics
- Stronger heritability does not imply simple genetics

Estimating Heritability with Twins

Falconer's Method

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

r_{MZ} = correlation between monozygotic co-twins

r_{DZ} = correlation between dizygotic co-twins

Variance Component Approach

Modeling the Phenotype:

$$p = \mu + \sum \beta_i x_i + a + d + e$$

μ Population mean

β Regression coefficients

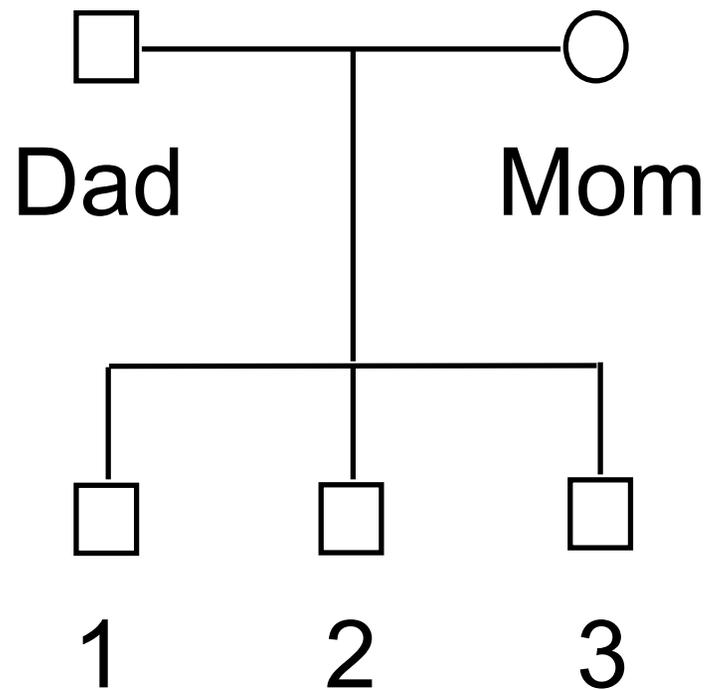
x Scaled covariates

a Additive genetic effects

d Dominance genetic effects

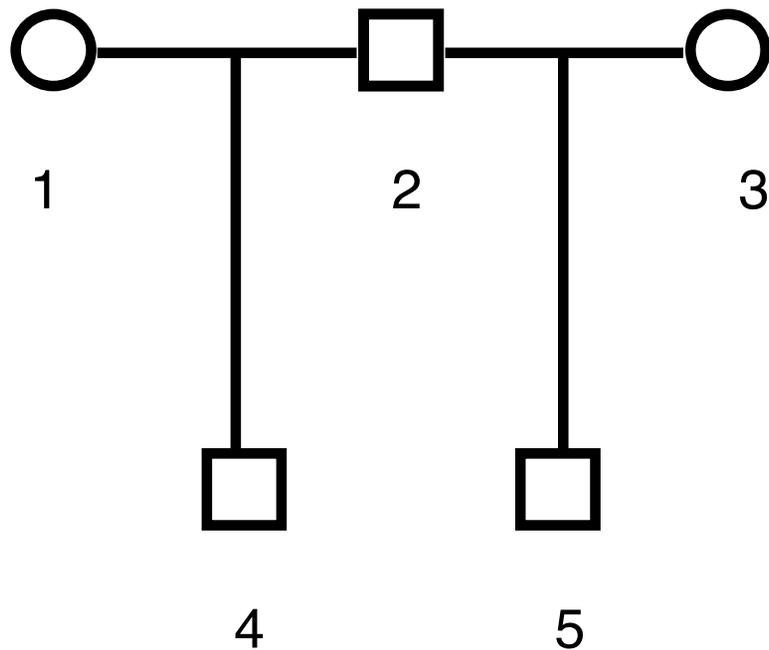
e Random environmental effects

Simple Kinship Matrix



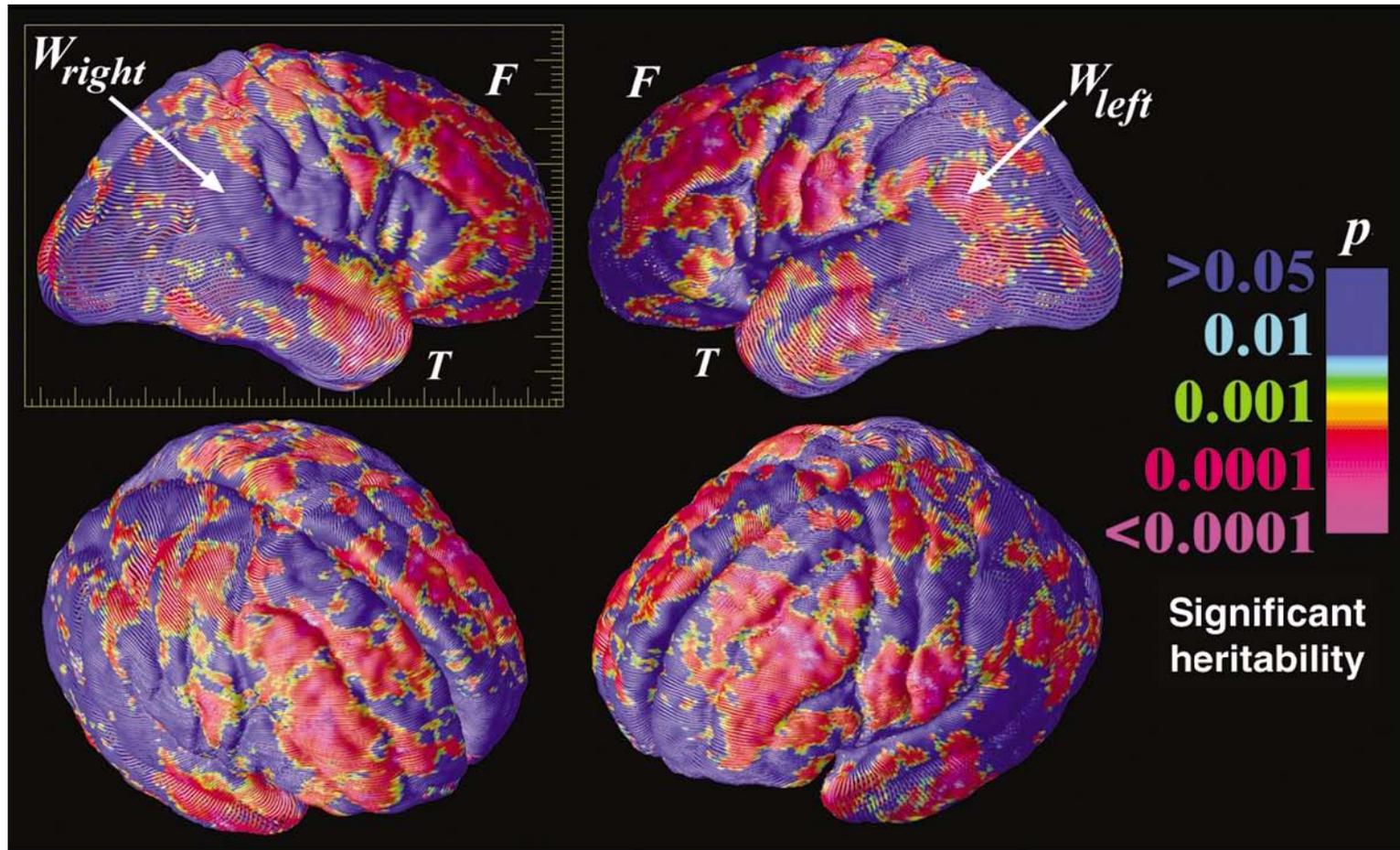
	D	M	1	2	3
D	1	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
M	0	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$
3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1

Kinship Matrix



	1	2	3	4	5
1	1	0	0	$\frac{1}{2}$	0
2	0	1	0	$\frac{1}{2}$	$\frac{1}{2}$
3	0	0	1	0	$\frac{1}{2}$
4	$\frac{1}{2}$	$\frac{1}{2}$	0	1	$\frac{1}{4}$
5	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	1

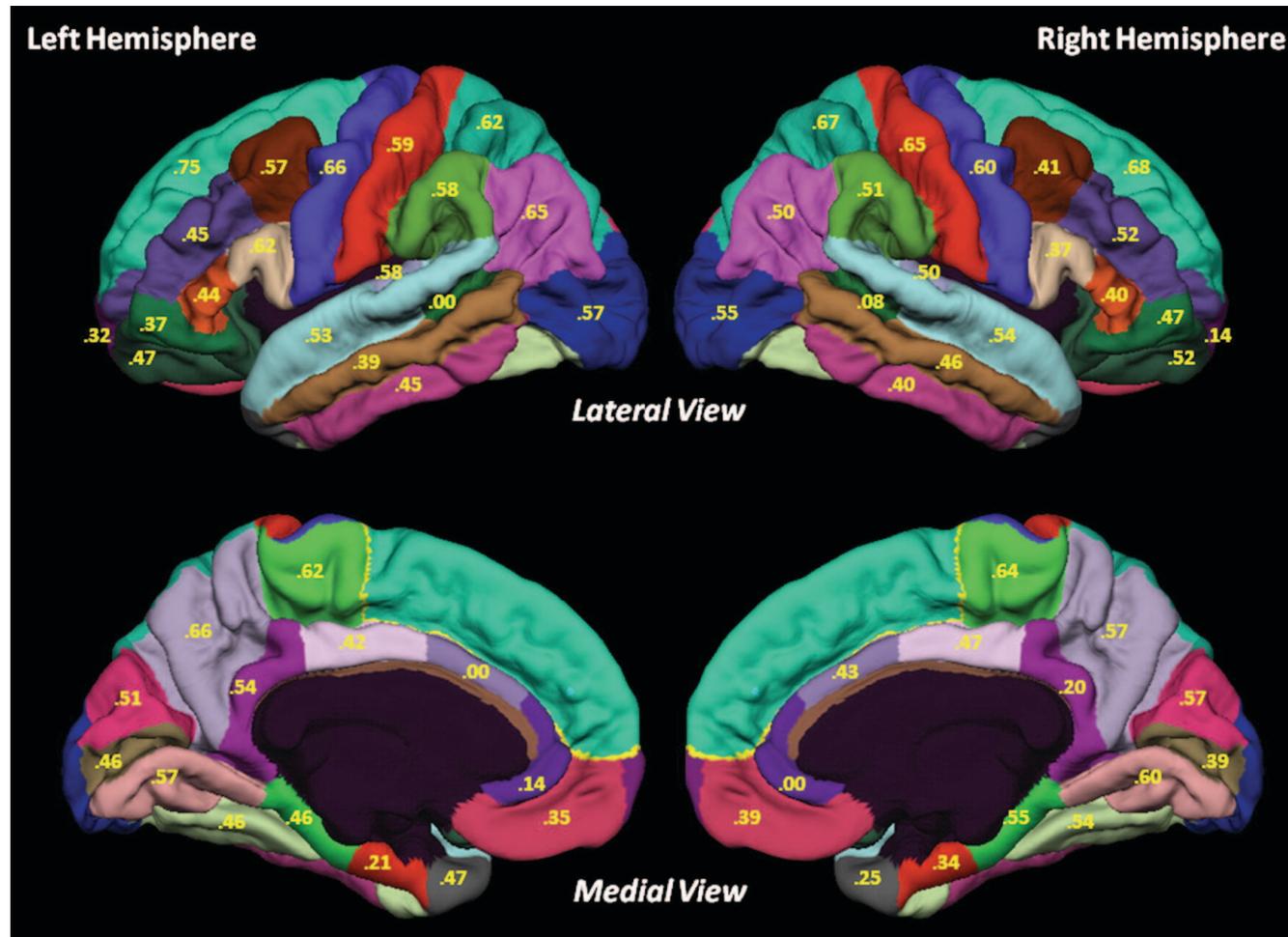
Gray-Matter Thickness Heritability



Thompson et al., *Nat Neuro*, 2001

10 MZ / 10 DZ pairs

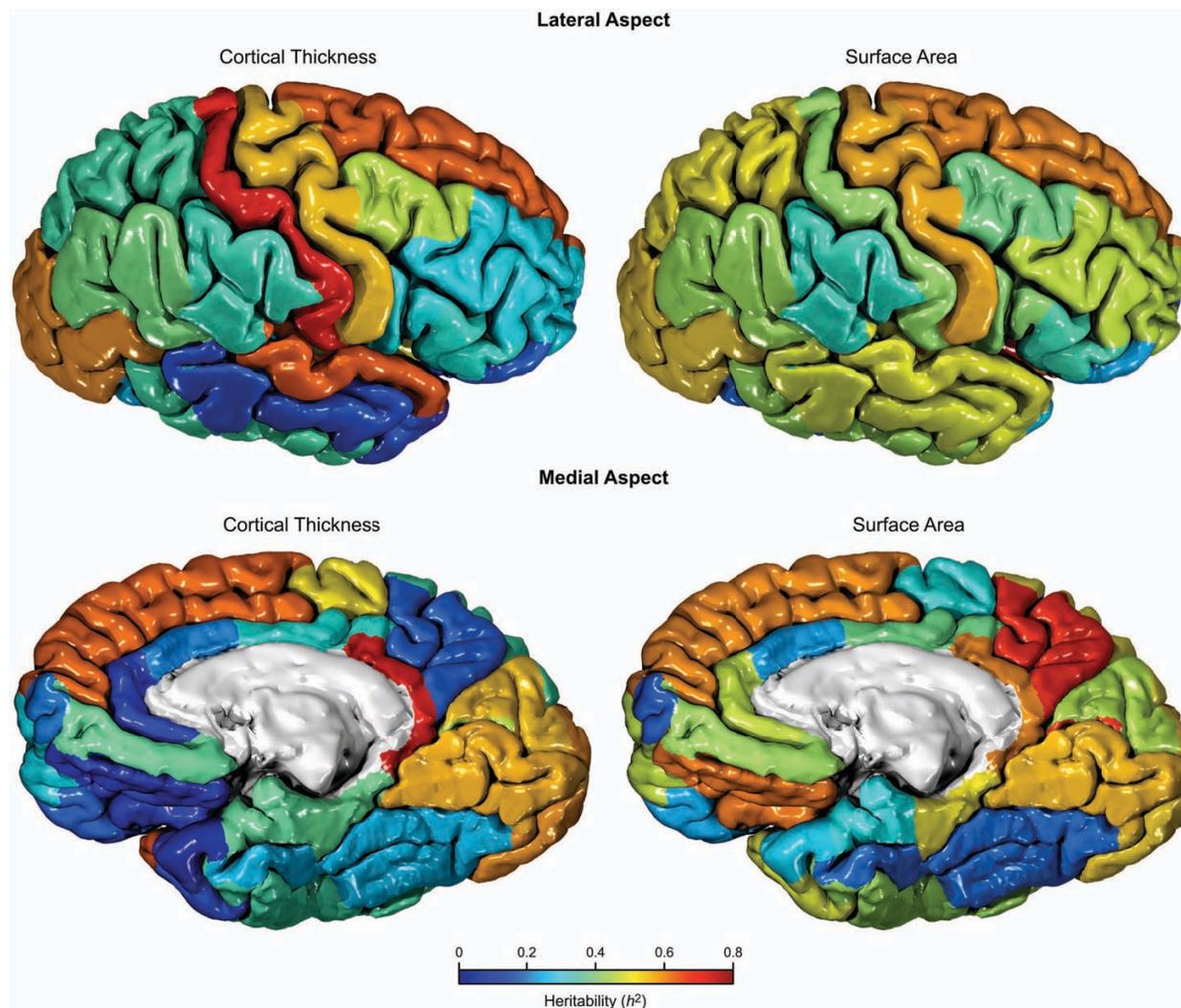
Gray-Matter Thickness Heritability



Kremen et al., *NeuroImage*, 2010

110 MZ / 92 DZ pairs

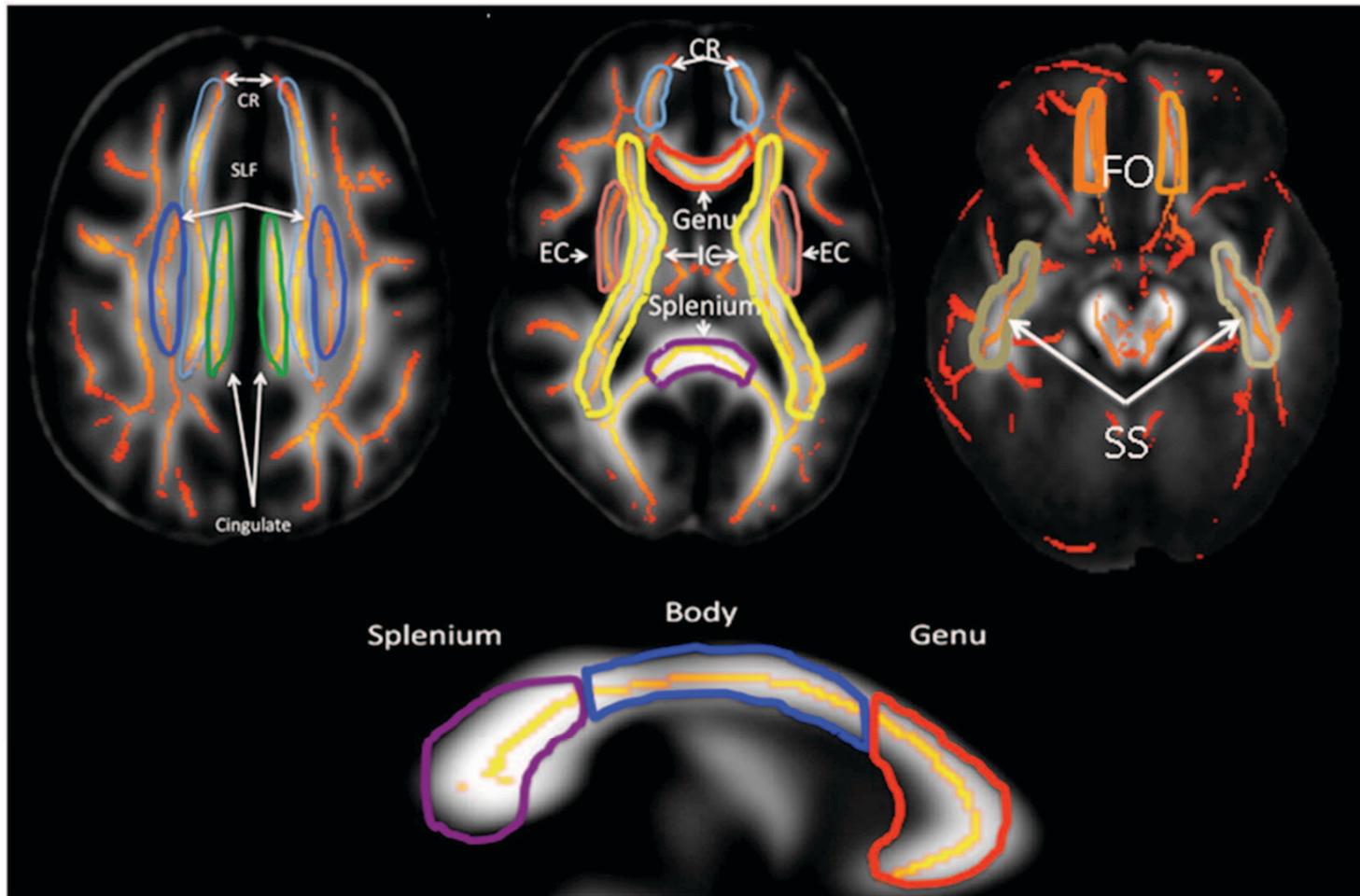
Cortical Thickness & Surface Area



Winkler et al., *NeuroImage*, 2009

486 family members

White-Matter Tracts (DTI) h^2



Kochunov et al., *NeuroImage*, 2010

467 family members

White-Matter Hyperintensity h^2

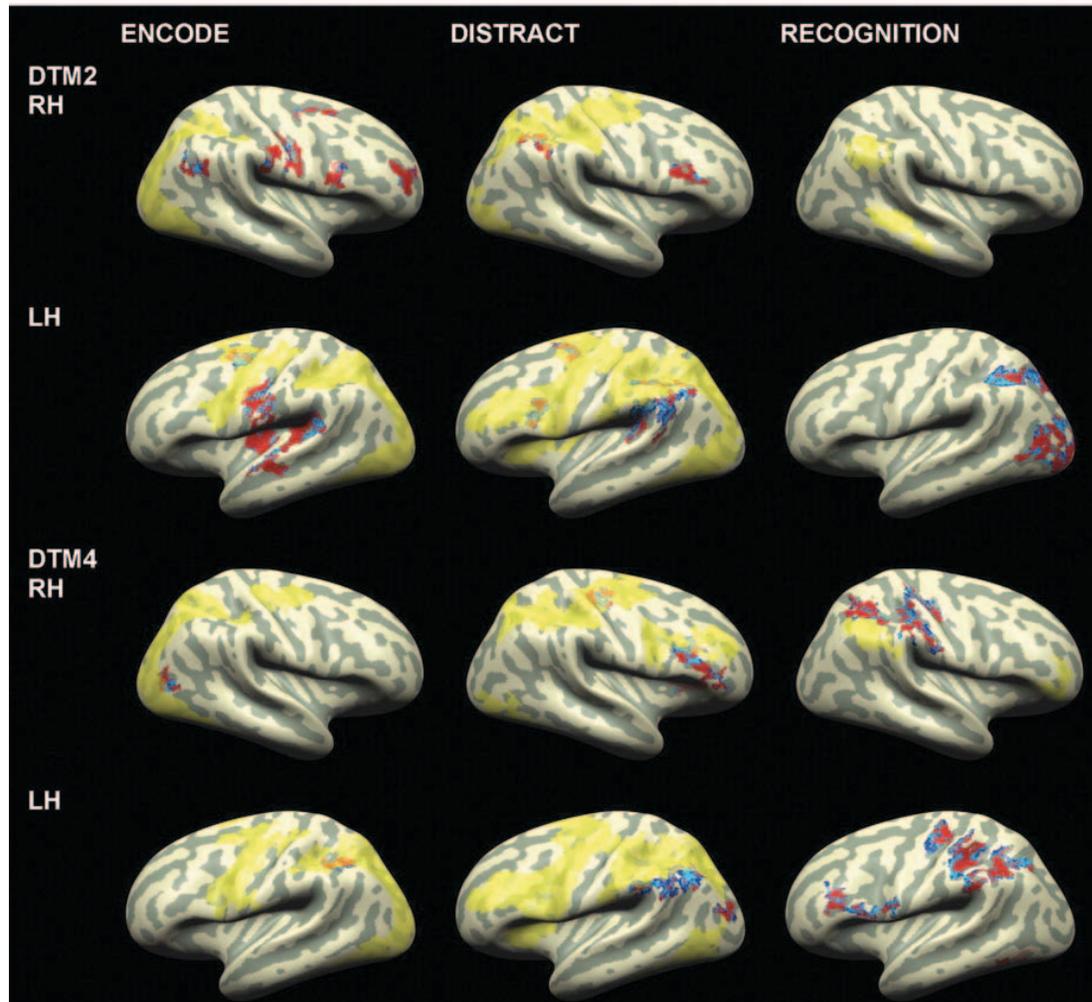
Total Cranial: 0.91
Brain Parenchyma: 0.92
WMH: 0.73



Carmelli et al., *Stroke*, 1998

74 MZ / 71 DZ pairs

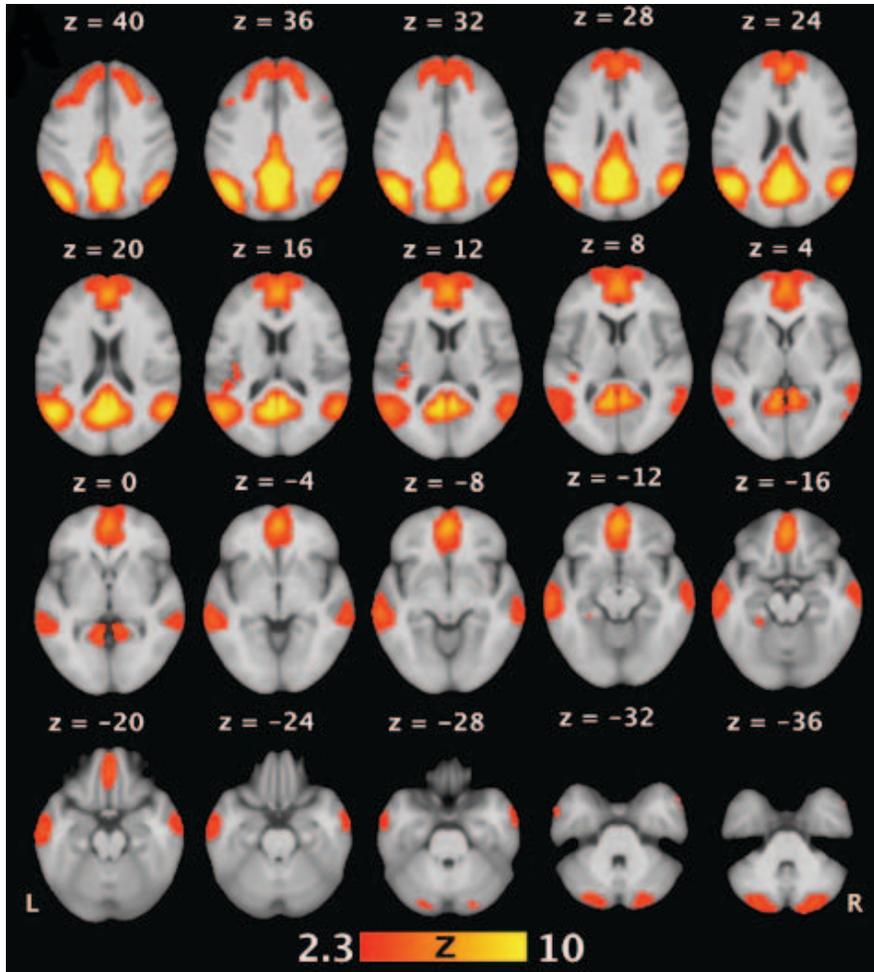
Task Based fMRI Heritability



Koten et al., *Science*, 2009

10 MZ pairs / 10 sibs

Resting State fMRI Heritability



$$h^2 = 0.424$$

Glahn et al., *Proc Nat Sci USA*, 2010

333 family members

Limitations of Heritability Estimates

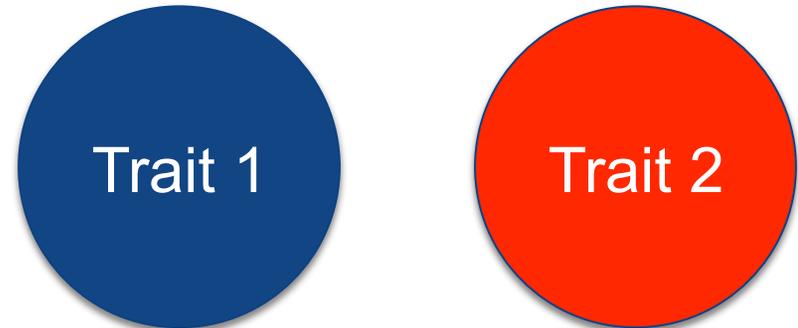
1. Heritability is a population level parameter, summarizing the strength of genetic influences on variation in a trait among members of the population. It doesn't tell you anything about particular individuals.
2. Heritability is an aggregate of the effects of multiple genes. It tells you nothing about how many genes influence a phenotype. A high heritability is not necessarily 'better' if it is due to many, many genes.

Question 2: Pleiotropy

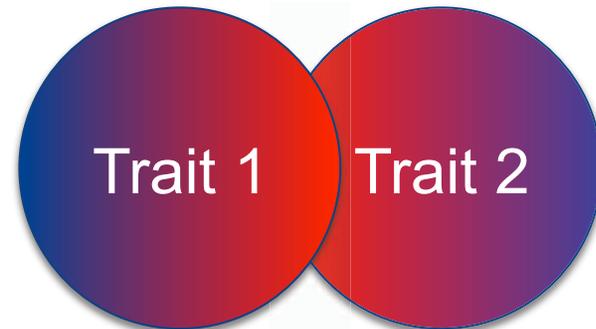
**Which traits are influenced
by the same genes?**

Levels of Pleiotropy

No Pleiotropy



Partial Pleiotropy



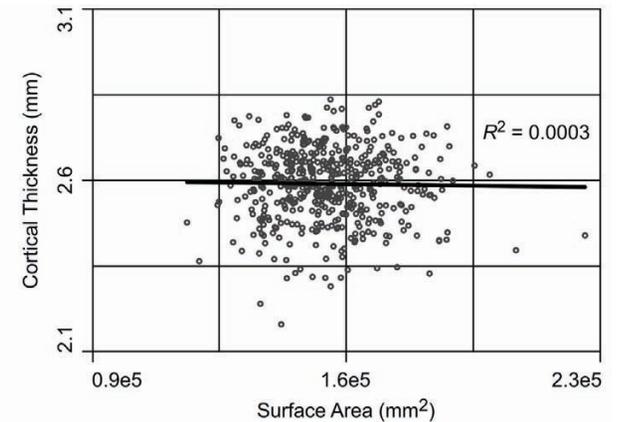
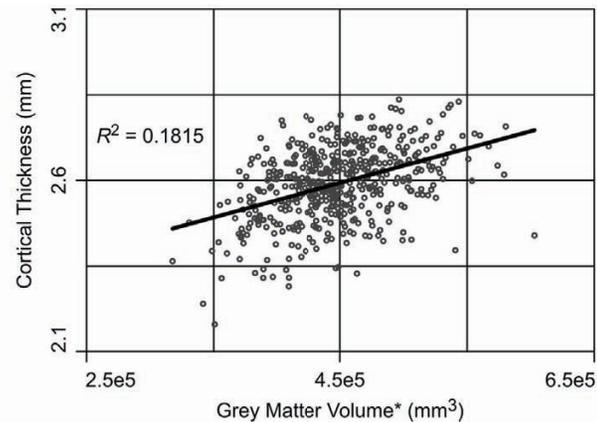
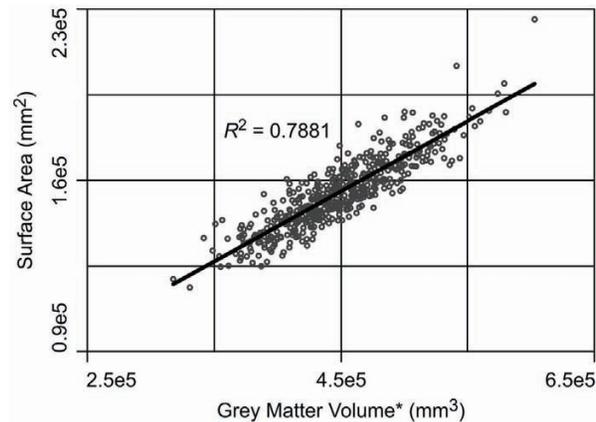
Full Pleiotropy



Genetic Correlation (Pleiotropy)

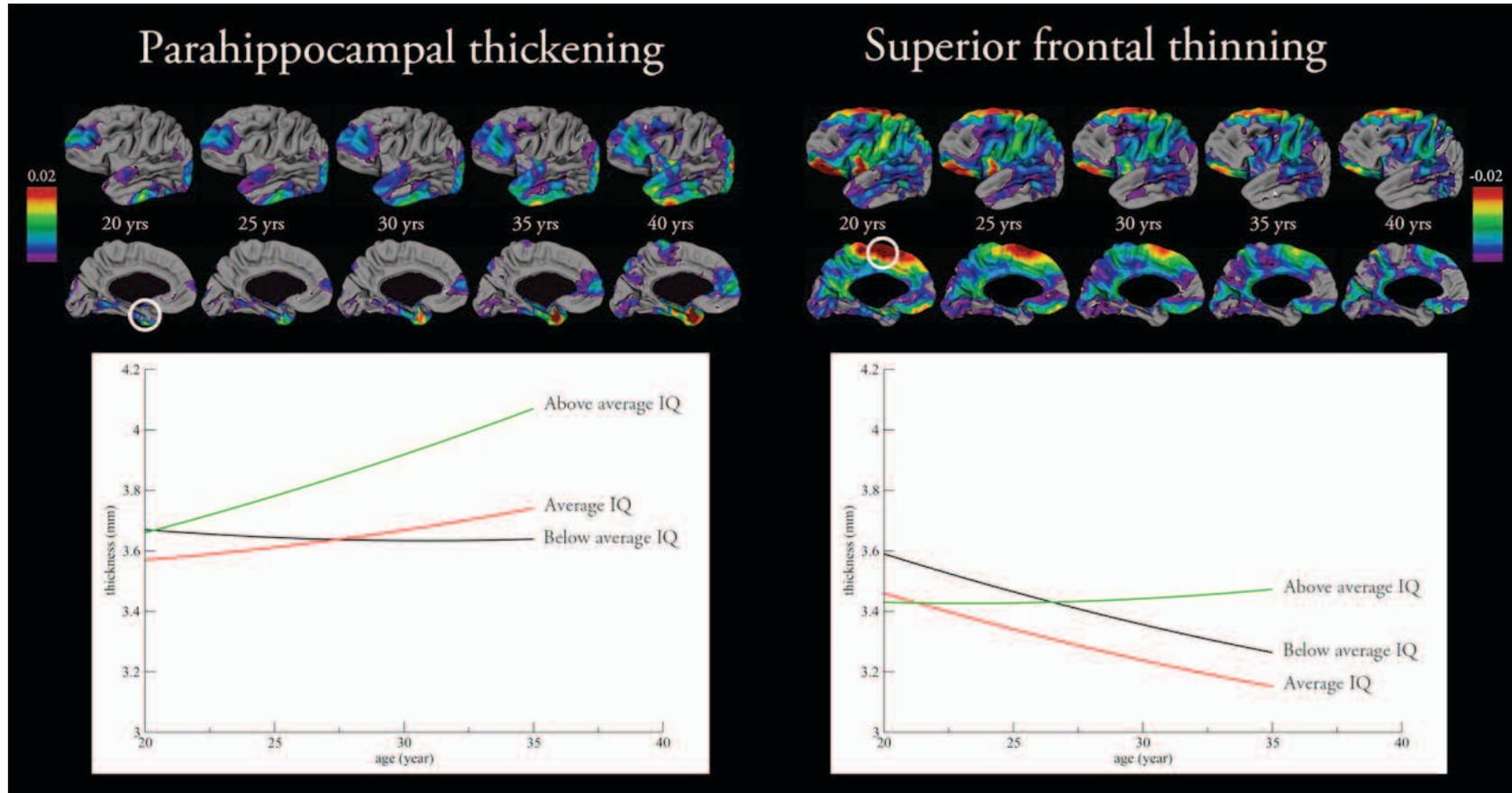
- Genetic correlation (ρ_g): a measure of the overlap in genetic effects between traits.
- ρ_g varies from -1 to 1
- 0 = no pleiotropy; -1 or 1 = complete pleiotropy

Cortical Thickness & Surface Area



Schmitt et al., *Cerebral Cortex*, 2008 107 MZ / 47 DZ / 228 others
Winkler et al., *NeuroImage*, 2009 486 family members

Cortical Thinning & IQ



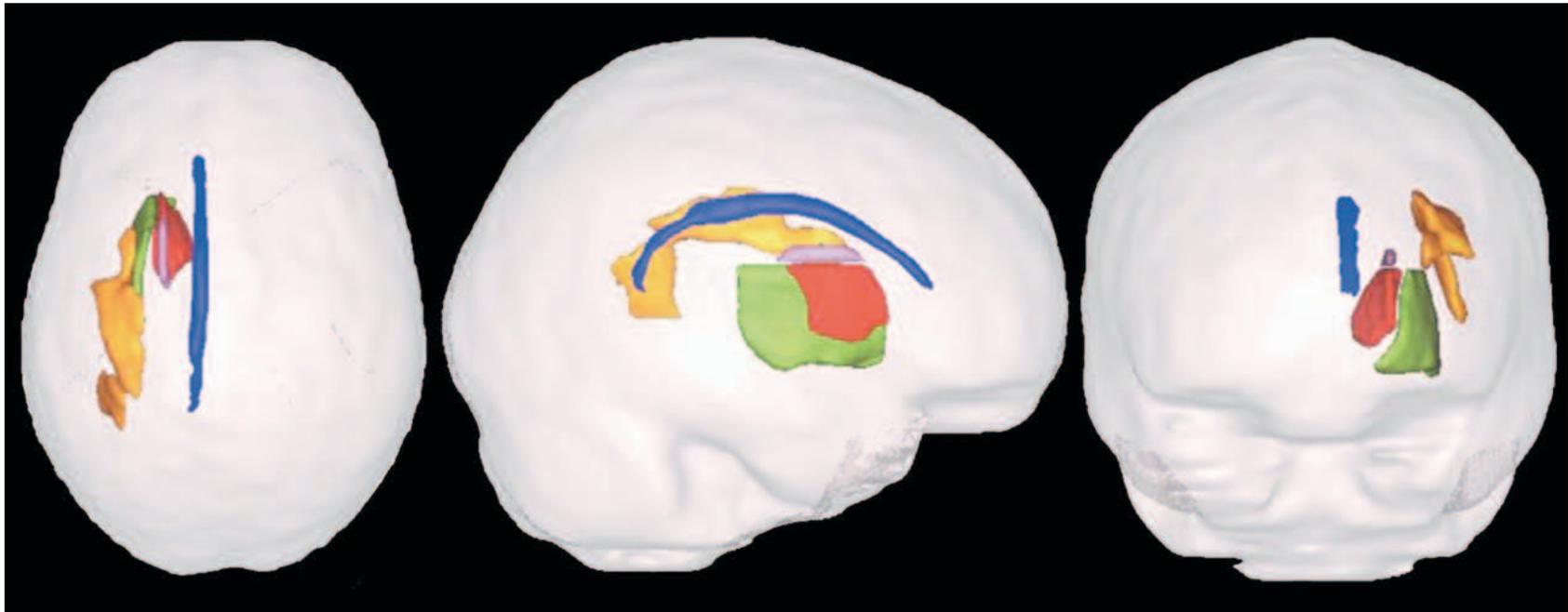
$$\rho_g = 1.0/1.0$$

$$\rho_g = 1.0/0.7$$

Brans et al., *J Neurosci*, 2010

77 MZ / 84 DZ pairs

White Matter Tracts & Working Memory



Superior longitudinal fasciculus – Spatial DRT: $\rho_g = 0.593$

Karlsgodt et al., *J Neurosci*, 2010

467 family members

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.

Genome Wide Association

- Tests for correlation between genotype and phenotype
- Association analyses work when:
 - 1) your genotyped marker is a functional polymorphism
 - 2) your genotyped marker is in disequilibrium with a functional polymorphism
- Currently: 700K to 1.5Mil SNPs

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**

Association, LD & Haplotypes

- Unlikely that functional variants are typed
- Association studies are dependent on LD between a genotyped marker and a functional variant.
- May change with complete sequencing

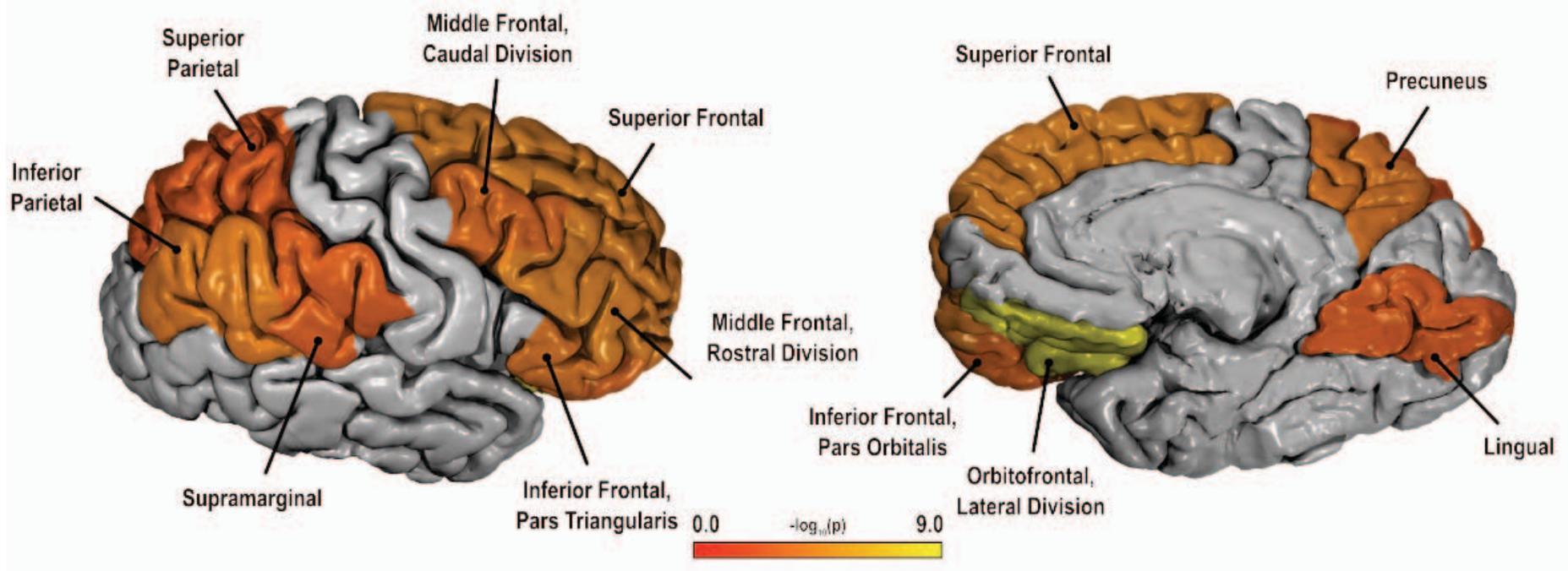
Multiple testing: p-values

- A p-value of 0.05 implies that 5% of the time we will reject the null hypothesis (i.e. conclude that we have an association) when the null hypothesis is actually correct
- If we test 100 SNPs and each time we use a p-value of 0.05 as our cutoff for significance, we would expect 5 of those SNPs to be significant ($p < 0.05$) just by chance

Multiple testing

- The simplest correction is the Bonferroni: multiply each p-value by the total number of tests, or divide the significance threshold required by the number of tests ($0.05 / \#$).
- For 550K SNPs, requires $p=1.2 \times 10^{-7}$
- This maintains an experiment-wide significance threshold, but may be too conservative when the tests are correlated, as they might be if some of the markers tested are in LD with each other.

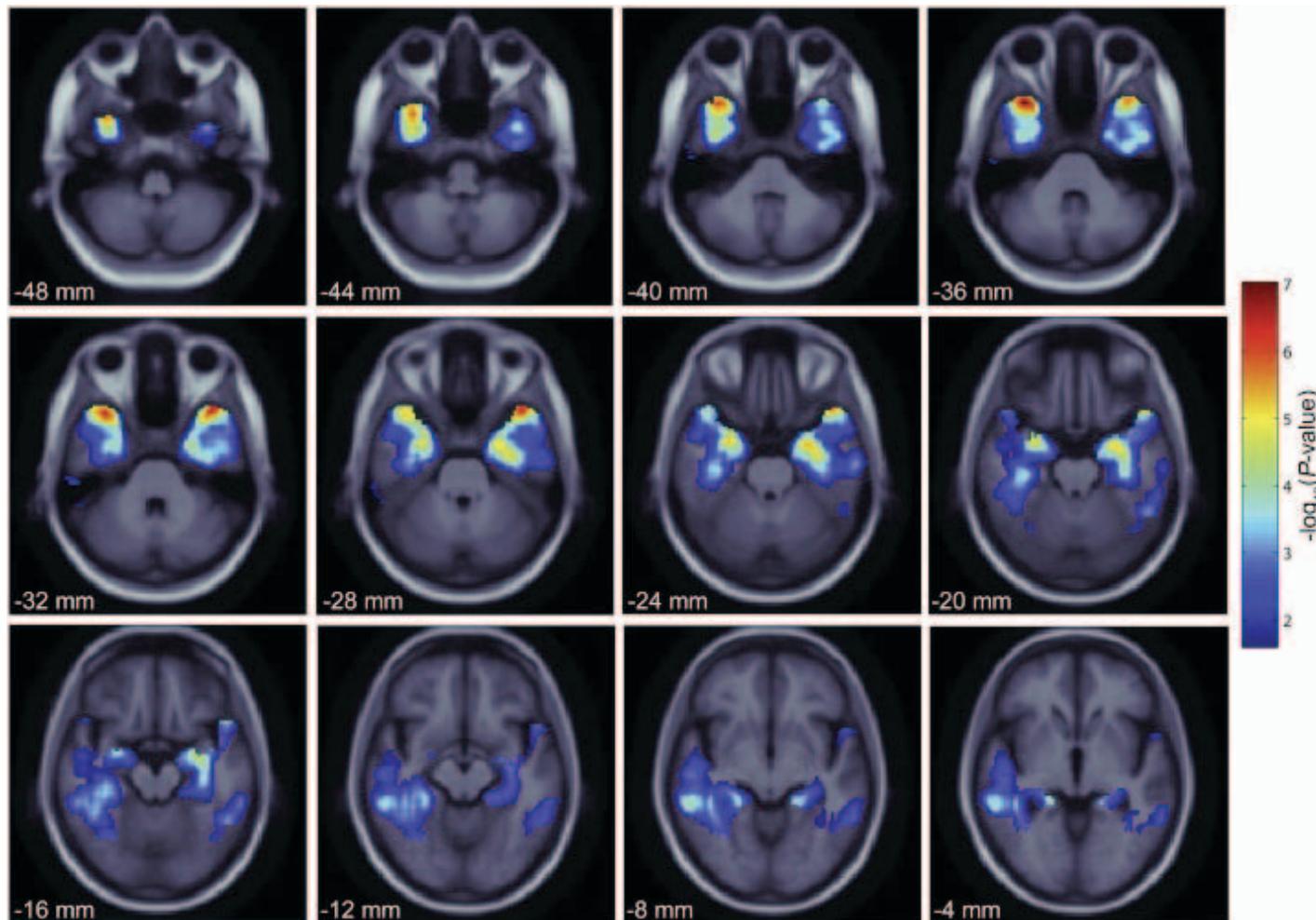
Cortical Thickness GWA



Association found at rs2342227 ($p=4.0 \times 10^{-10}$), chr 13q31, near SLIT- & NTRK-like family, member 6. SLITRNs are expressed predominantly in neural tissues and have neurite-modulating activity. Associated with Tourette's Syndrome

Glahn et al., *ACNP*, 2009

Temporal Lobe GWA



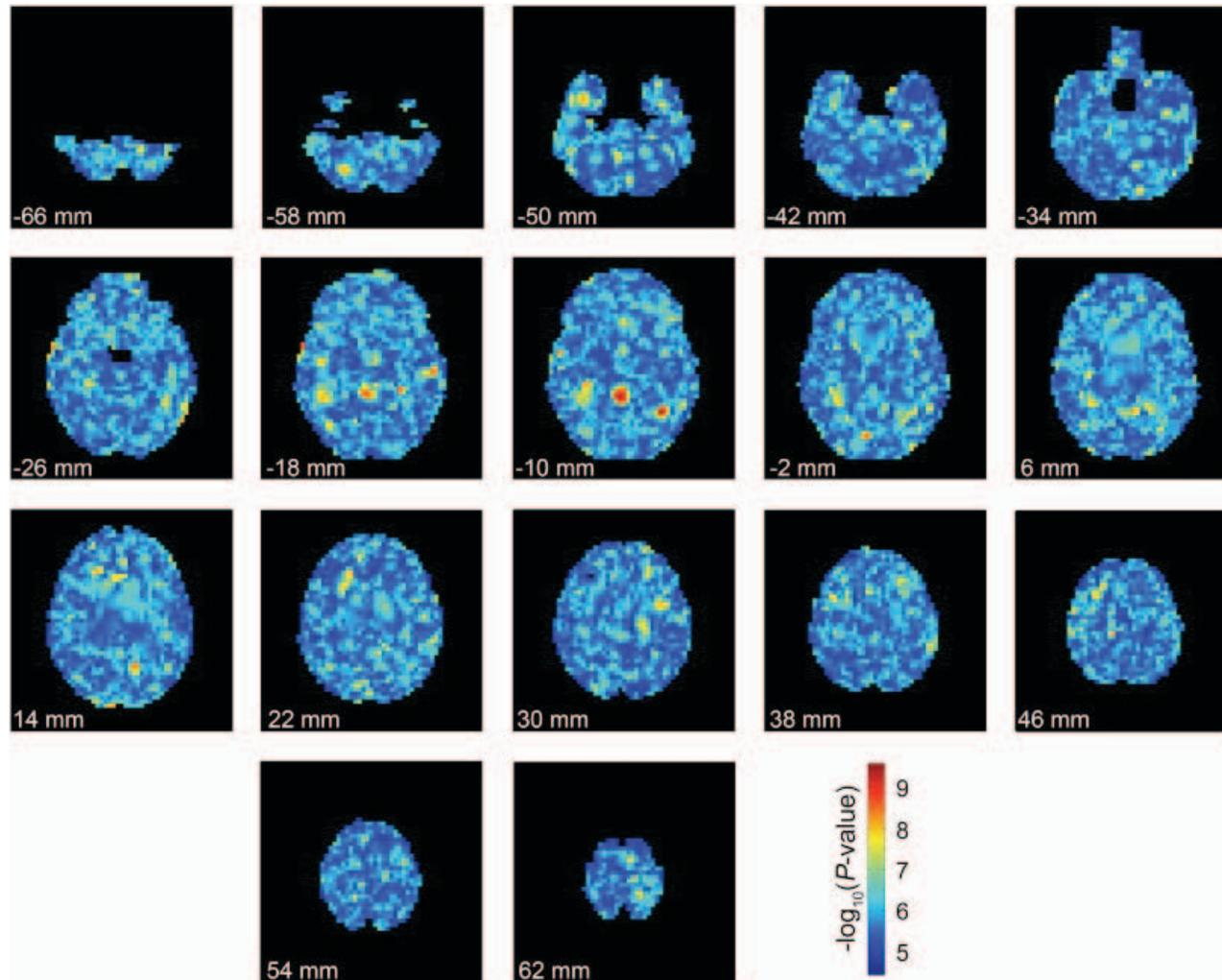
n=742

rs10845840
 $p=1.3 \times 10^{-7}$

rs2456930
 $P=3.1 \times 10^{-7}$

Stein et al., *NeuroImage*, 2010

VBM & GWA

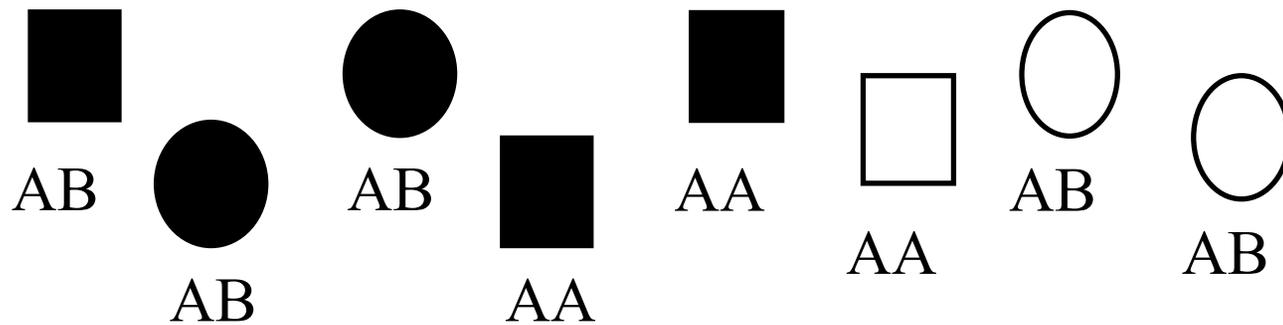


Stein et al., *NeuroImage*, 2010

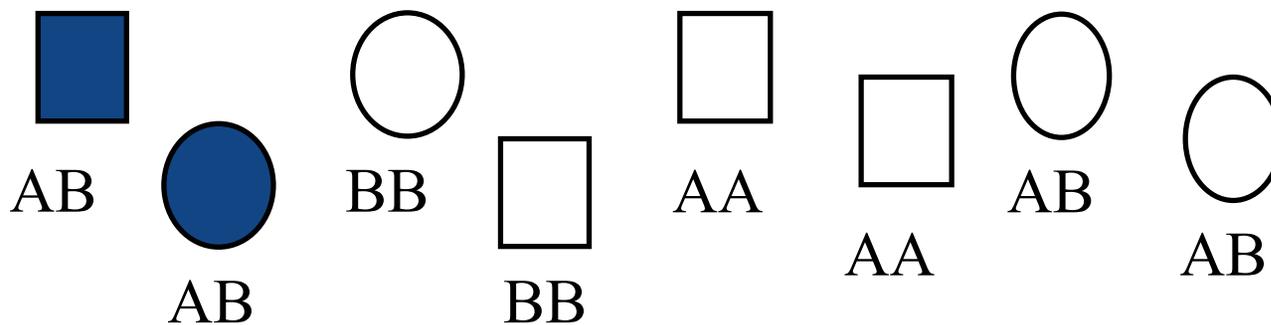
Limitations of Association

- A QTL may be in equilibrium with the other polymorphisms surrounding it. Disequilibrium need not be present.
- 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

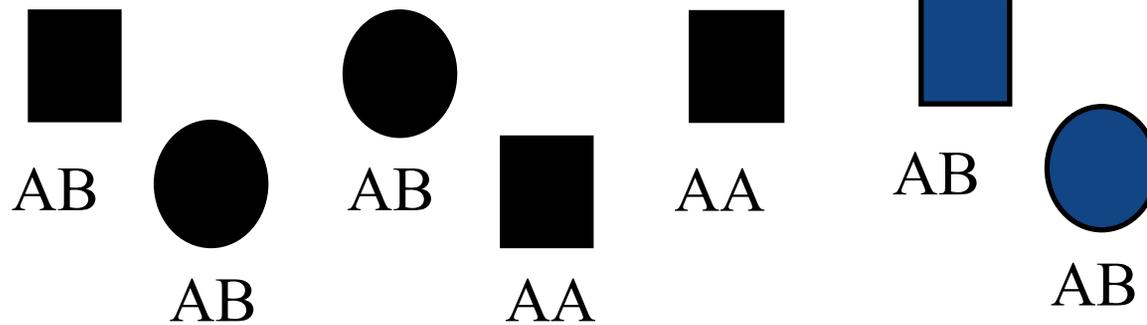


African
Americans
70% A,
30% B

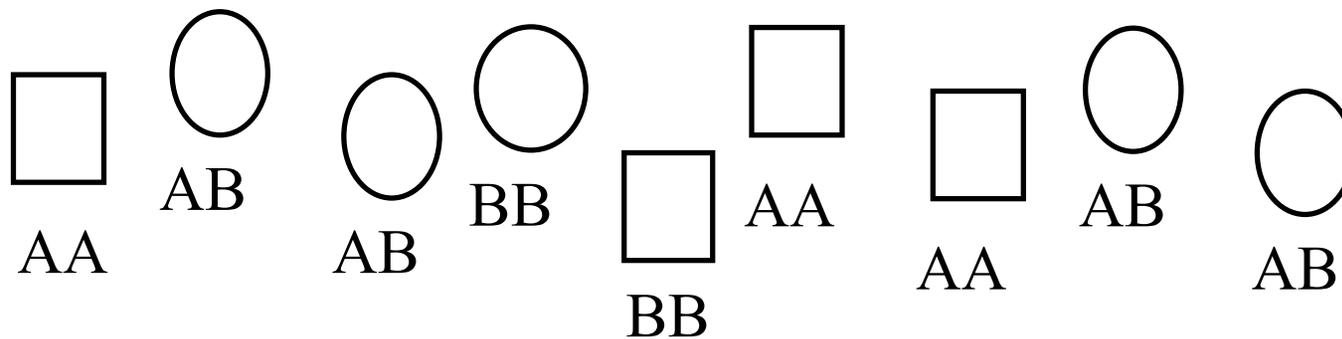


European
Americans
50% A,
50% B

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.

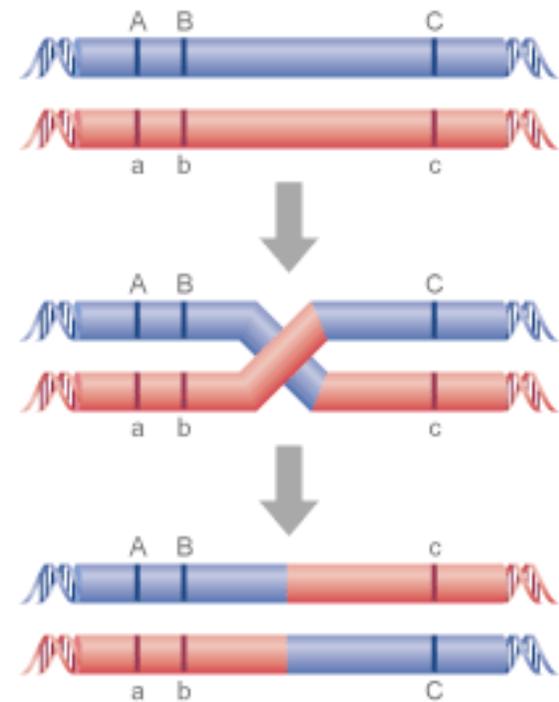
Con: Need a large sample of related individuals. “Large enough” may be too large to be practical.

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)

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

NR denotes the number of non-recombinant offspring,

R denotes the number of recombinant offspring.

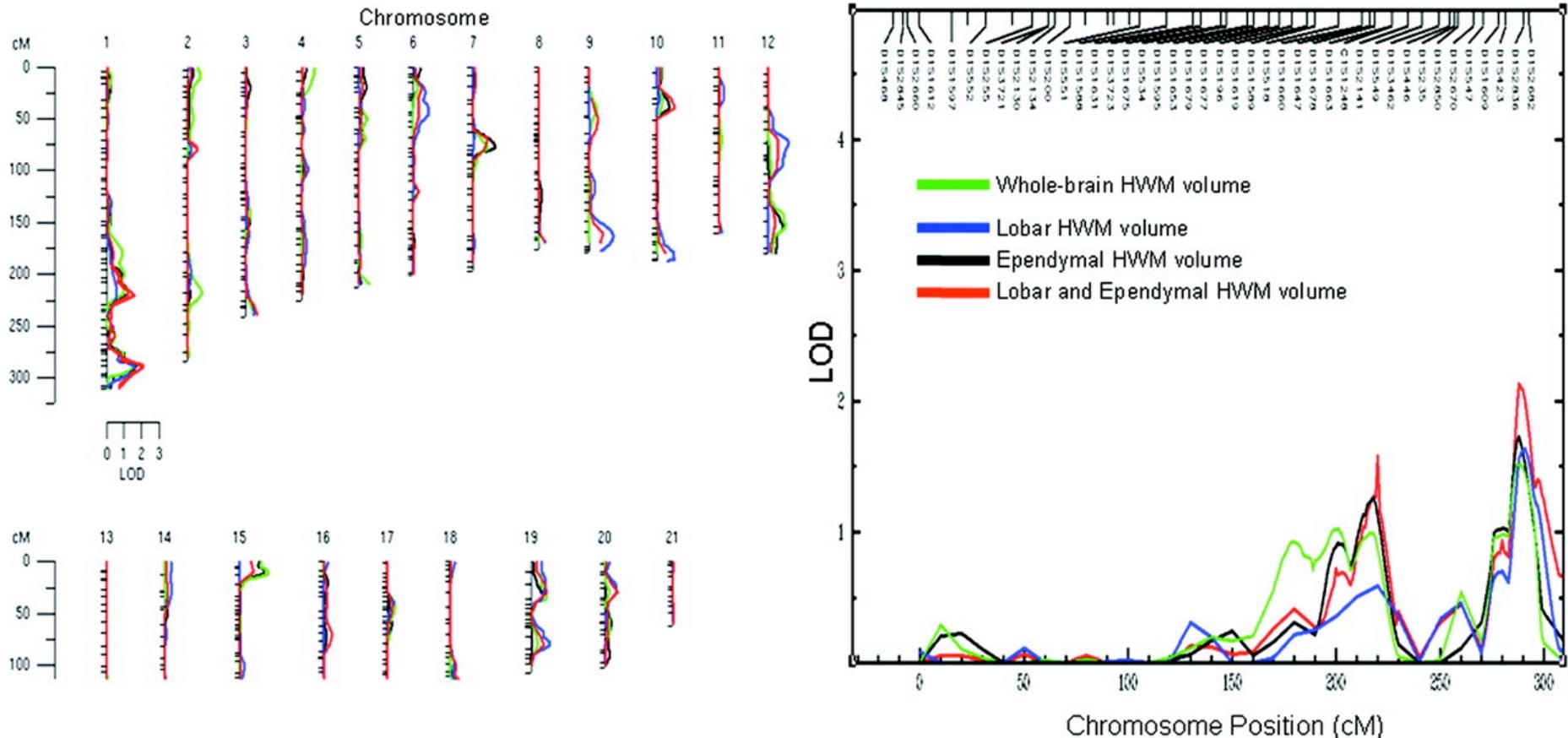
Theta = recombinant fraction = $R / (NR + R)$

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 with White-Matter Hyperintensities

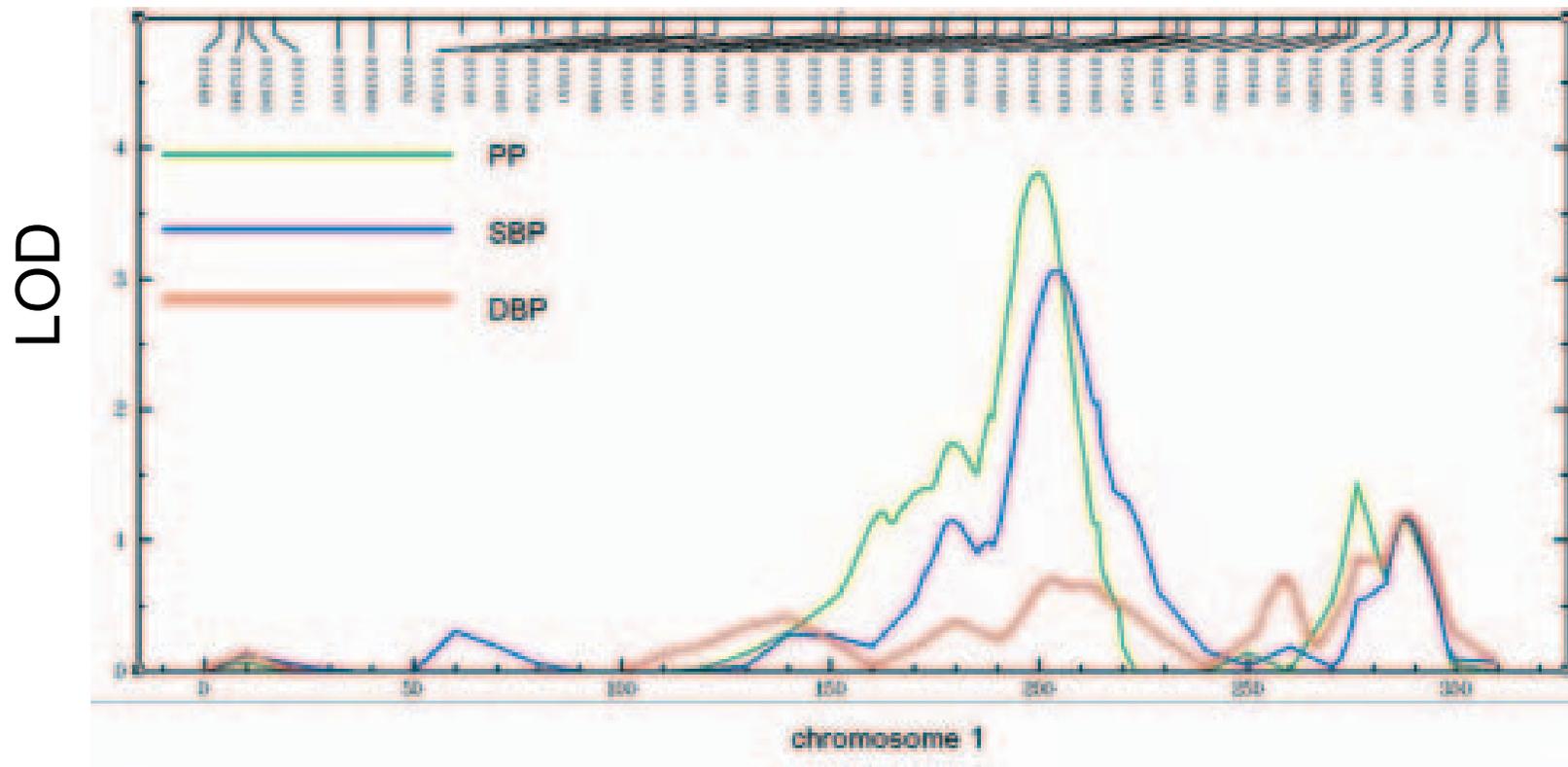


Bivariate linkage for subcortical & ependymal HWM volumes (log of odds=2.12) on chr 1 at 288 cM.

Kochunov et al., *Stroke*, 2009

459 family members

Linkage with White-Matter Hyperintensities/Blood Pressure



Kochunov et al., *Stroke*, 2010

357 family members

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.

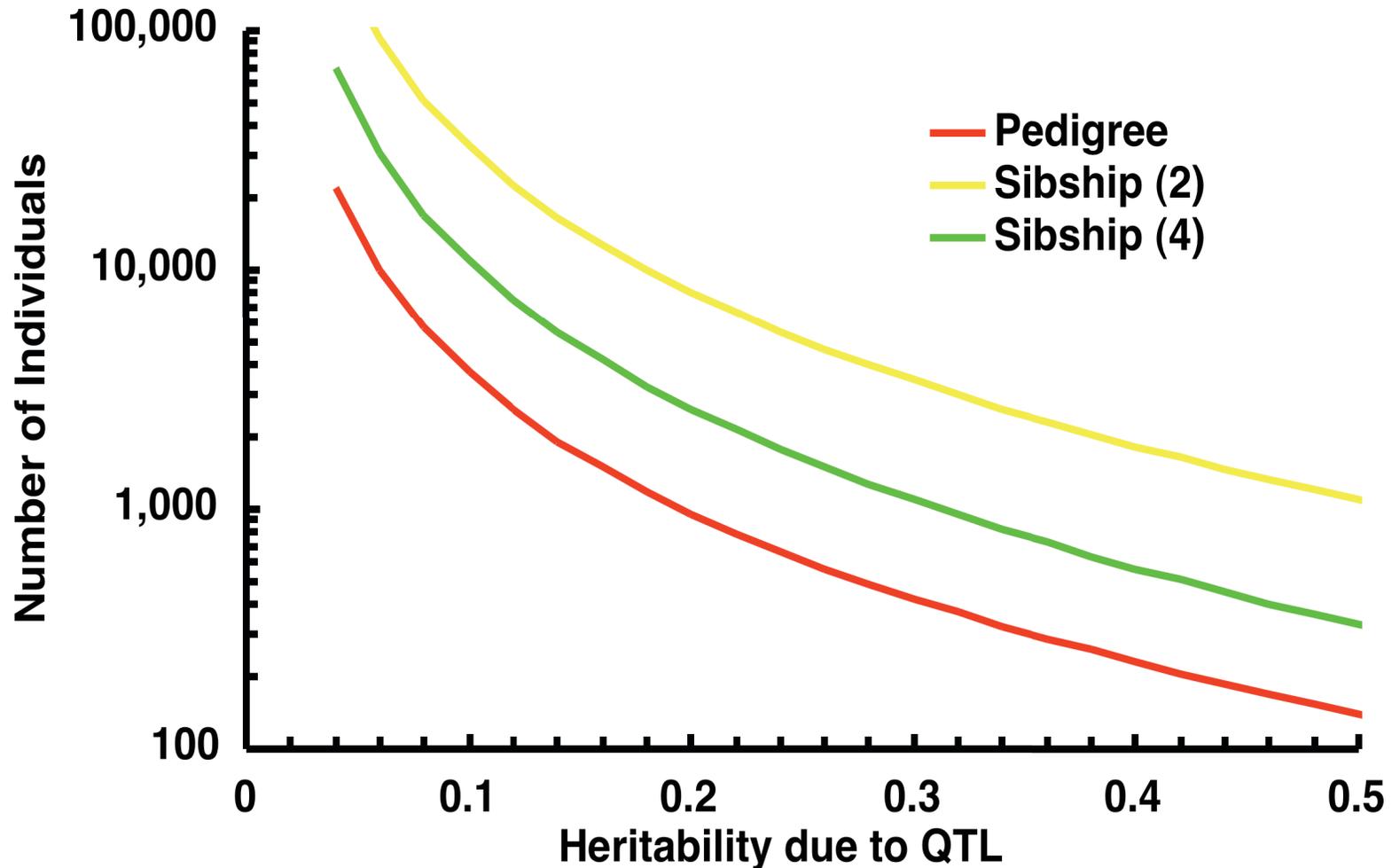
Linkage: you're testing for an excess of the parental type. That parental type (i.e. the alleles traveling together) could be different in every family (i.e. linkage equilibrium) and you would still get linkage. Can detect cumulative effect of multiple variants (including rare variants).

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



Question 4: Identification

What specific genes influence the trait?

Identifying a Causal Gene

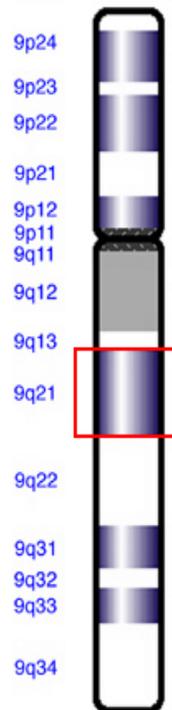
- Once a significant QTL is identified, additional genetic tests are needed to determine the exact identity of the gene
 - Association: identifies a genomic region of ~500kb (250kb to either side of the association) determined by the general extent of linkage disequilibrium
 - Linkage: detect the cumulative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

VPS13A: A Candidate Gene for Brain Structure/Function

- VPS13A is a 73 exon gene localizing to chromosome 9q21

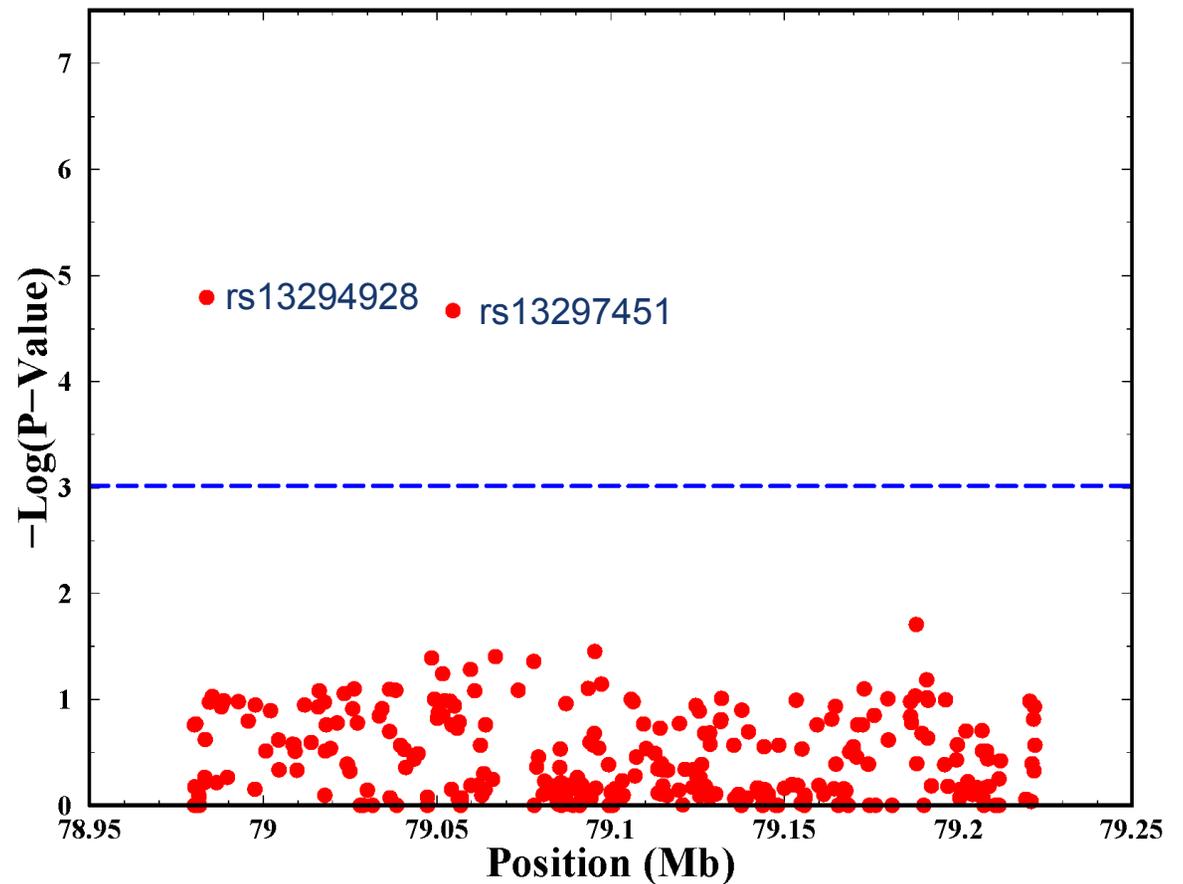
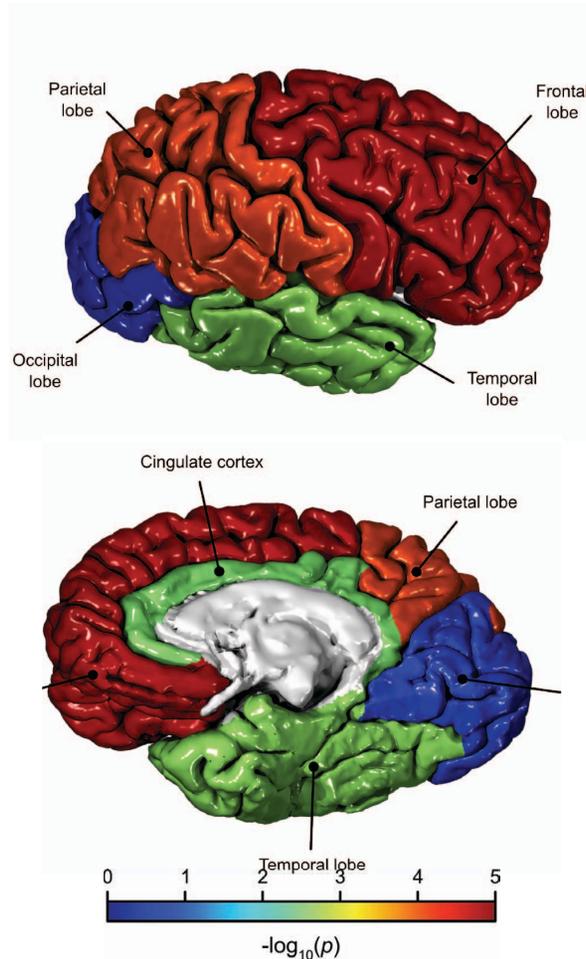


3174 amino acid protein



- Involved in chorea-acanthocytosis
 - Schizophrenia
 - OCD
 - Frontal lobe functioning

Association Analysis: *VPS13A* SNPs and Frontal Lobe Volume



Bellis et al., *ASHB*, 2009

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



Coffee Break

