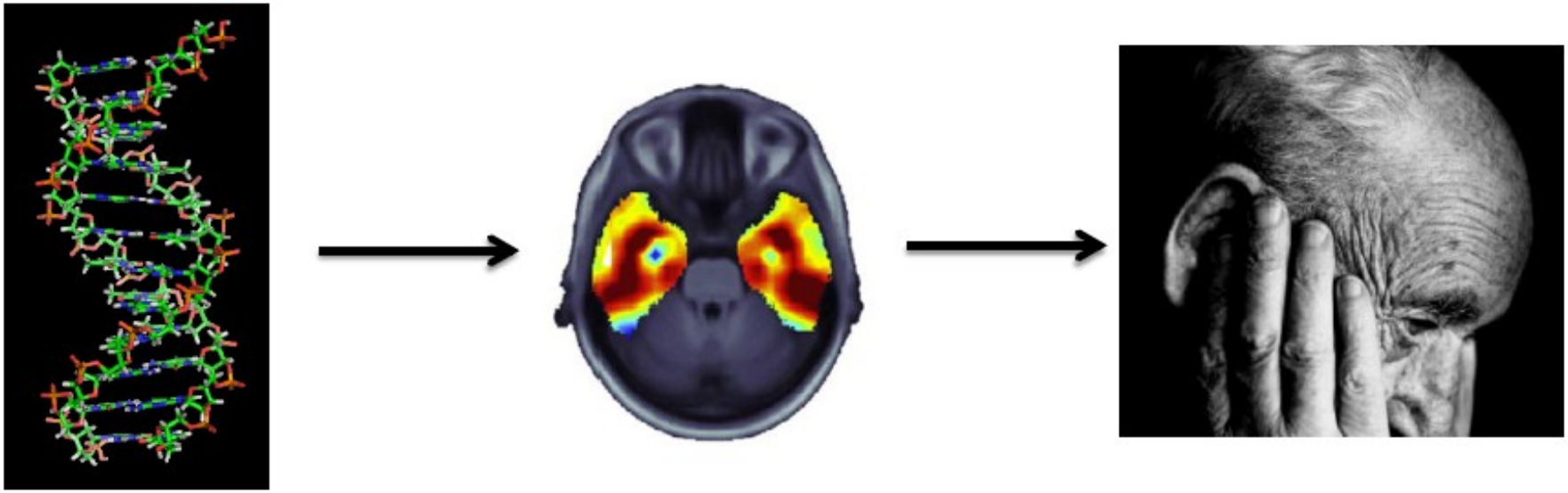
Neuroimaging and Genetics

Carrie E. Bearden, Ph.D.
Departments of Psychiatry and Biobehavioral Sciences and Psychology
Semel Institute for Neuroscience and Human Behavior
University of California, Los Angeles
cbearden@mednet.ucla.edu

What can we learn from identifying the genetic influences on brain structure?



1. **New brain biology:** Understand the development of the human brain
2. **Characterize known disease variants:** A way to localize the effect or develop mechanisms of known disease associated genetic variants
3. **Discover new disease variants:** If genes have a stronger influence on brain structure than disease, imaging genetics may be better powered to detect effects.

Slide courtesy of D. Hibar

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

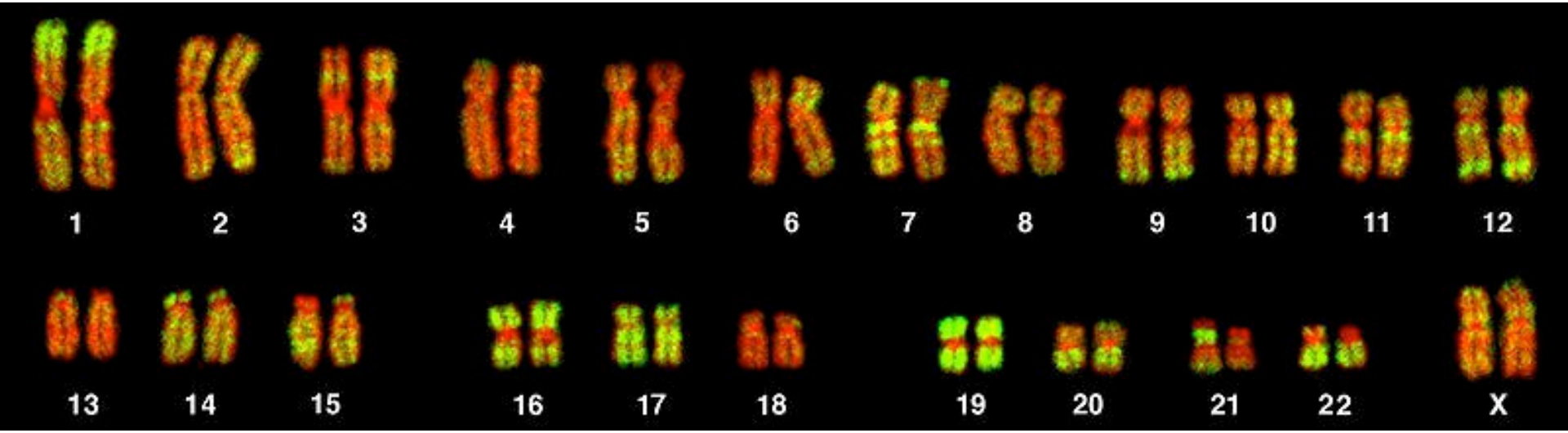
Genetics Terminology 1

- Quantitative Trait Locus (QTL), a location, usually a chromosomal region, implicated as containing one or more genes that influence a phenotype of interest
- Quantitative Trait Nucleotide (QTN), a specific sequence variation that has been implicated as having a functional effect on a phenotype of interest
- A QTL may contain multiple QTNs

Genetics Terminology 2

- Chromosome: a single long string of DNA, humans have two copies of each chromosome (diploid), one from mom, one from dad
- Autosome: chromosome not involved in sex determination (22 in humans)
- Sex chromosome: X or Y chromosome, females have 2X, males have 1X and 1Y
- Mitochondrial DNA: non-nuclear DNA, inherited only from the mother

Human Genome



23 Chromosomes
~25,000 genes, about half expressed in brain
~3 billion base pairs

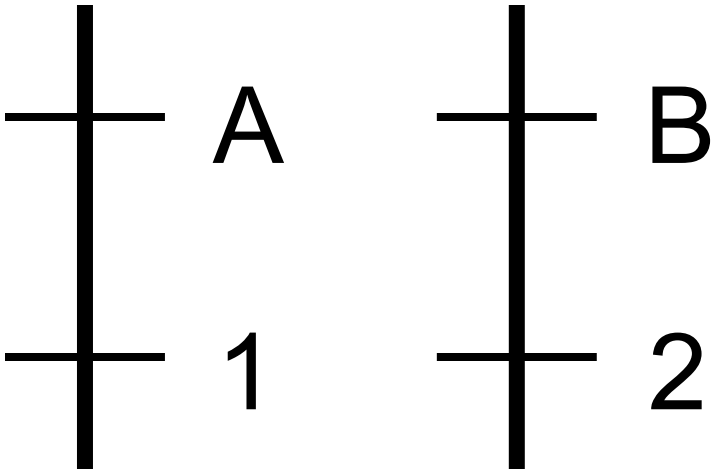
Genetics Terminology 3

- Gene: a unit of DNA that codes for a protein (but the term may be used to include both coding and non-coding elements)
- Locus: location, sometimes used interchangably with gene
- Allele: the specific variant you have at a particular site in the genome

Genetics Terminology 4

- Genotype: the combination of alleles on the two chromosomes of an individual
- Haplotype: the alleles at different loci being carried together on the same chromosome
- Phenotype: the trait of interest, some measurable property of the individual
  - Examples: neuroanatomy, neurophysiology, schizophrenia, bipolar disorder, Alzheimer's

Loci, Alleles, Genotypes, Haplotypes



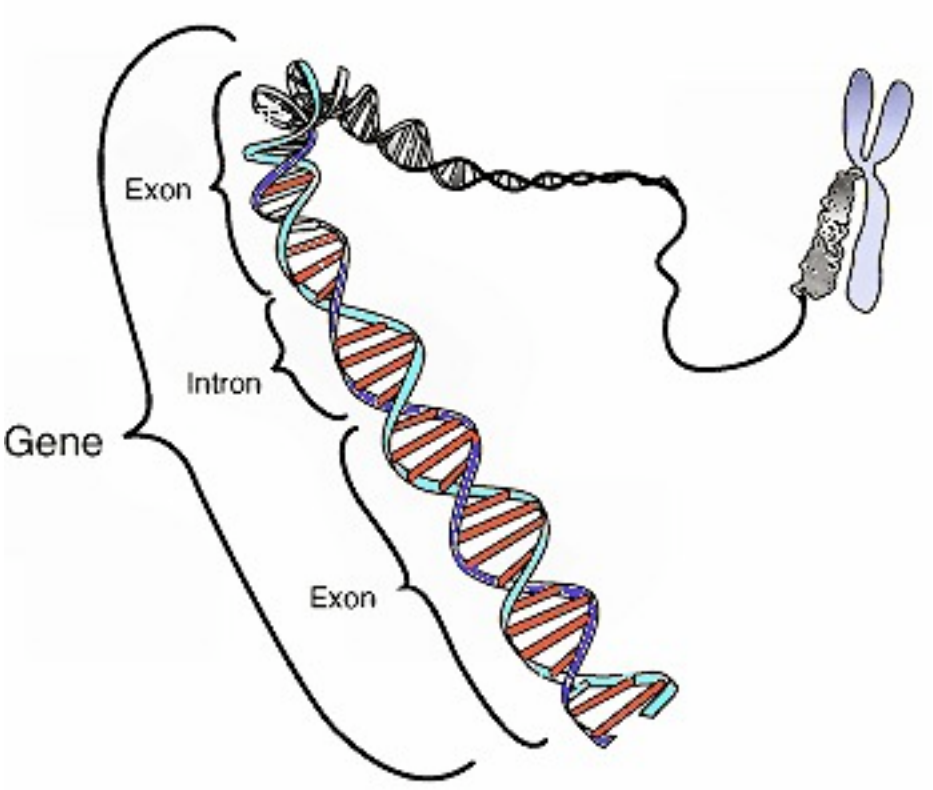
Two loci: letter locus & number locus
Each locus has two alleles: A or B, 1 or 2
The genotype at the letter locus is AB
The haplotype on the first chromosome shown is A1

Within A Gene

- Exon: specifies the mRNA, which is translated into the series of amino acids in the protein, coding sequence
- Intron: non-coding, intravening sequence
- Splice site: the juncture between an exon & an intron
- Promoter, enhancer: regulatory element (usually non-coding), controls time, place, amount of transcription

The Gene

- Functional & physical unit of heredity passed from parent to offspring (pieces of DNA)
- Typically contain information to make a specific protein
- Composed of nucleotides, sequence of four organic bases (Adenine, Guanine, Cytocine, and Thyamine)
- Matching nucleotides on the complimentary DNA strands form a base-pair



Genetics Analysis

Outline: 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?

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.

Types of Samples for Genetics

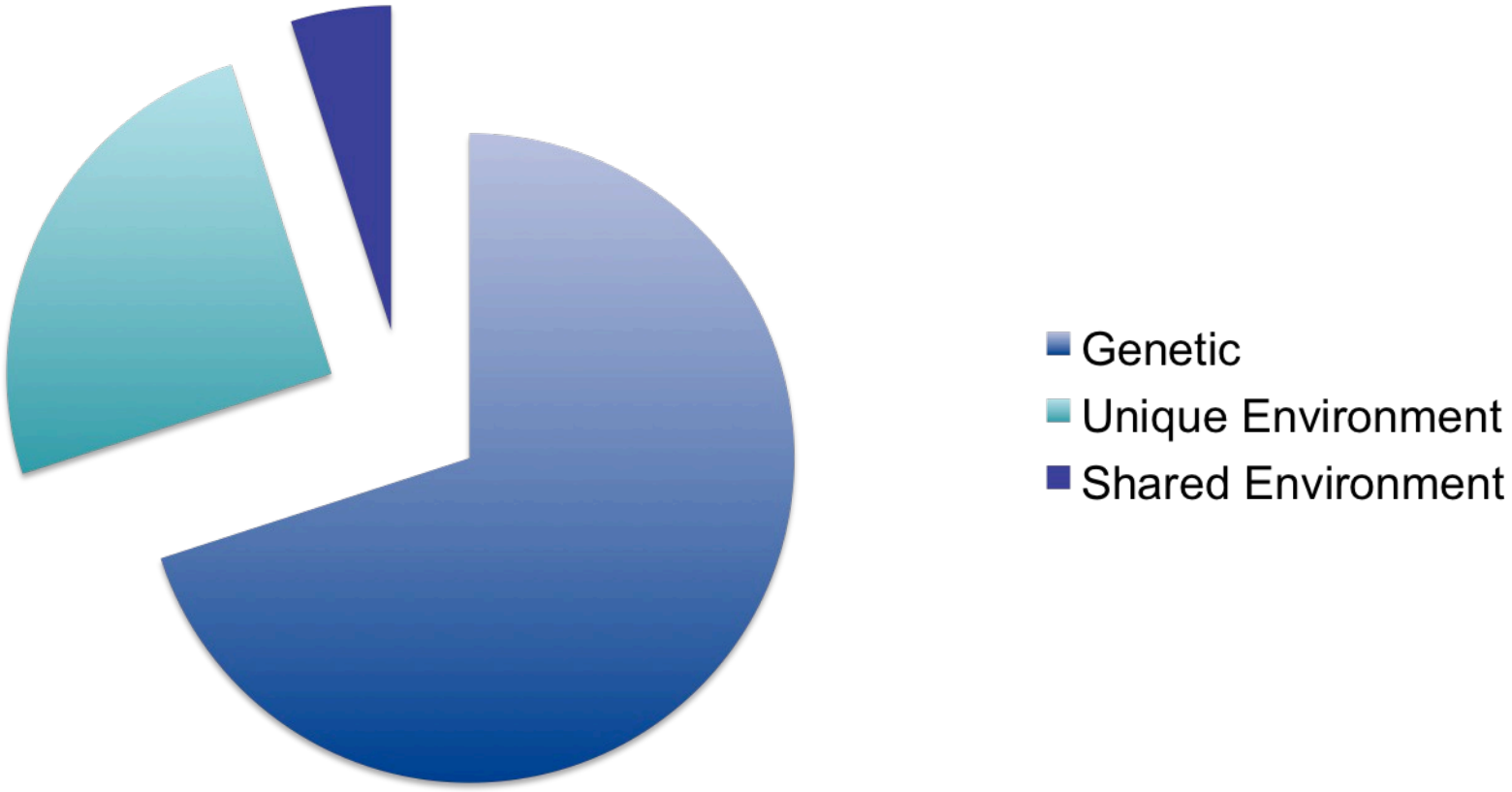
1. Adoptees: separating the effects of genes and family environment
2. Unrelated individuals: association only, estimation of effect size after variants are identified
3. Parent-child triads: association in the presence of linkage (transmission disequilibrium test), heritability/relative risk
4. Twins: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

Question 1: Heritability

Is this trait influenced by genetic factors?

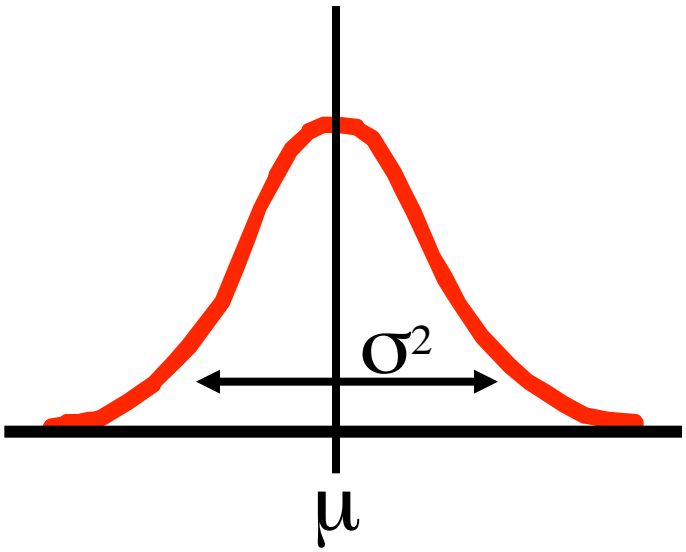
How strong are these genetic influences?

Defining Heritability



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

Variance Decomposition



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

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

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

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

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

σ_p^2 = total phenotypic

σ_g^2 = genetic

σ_e^2 = environmental

σ_a^2 = additive genetic

σ_d^2 = dominance

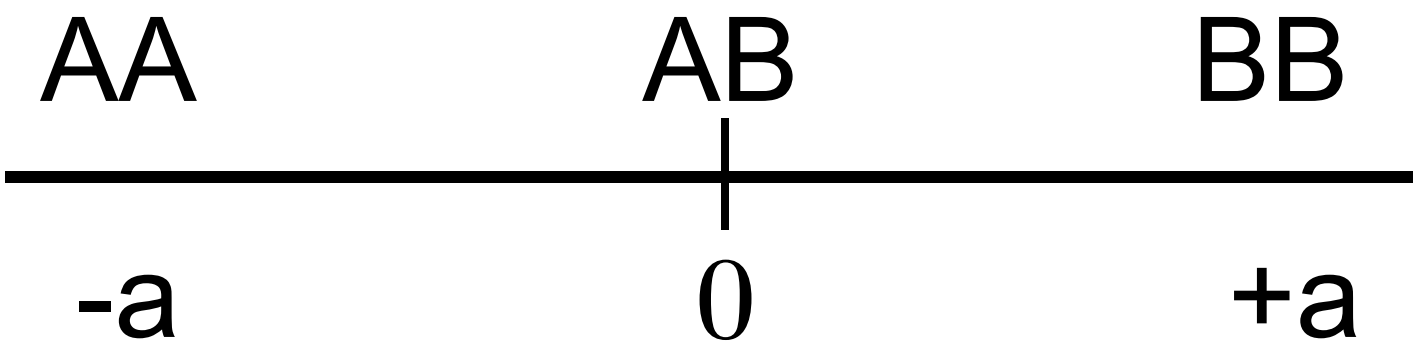
Almasy & Blangero, *Am J Hum Genet*, 1998

Broad-Sense Heritability (H^2)

- Proportion of total variance in a population, taken at a particular time or age, attributable to genetic variation
- All possible genetic contributions
  - allelic variation (additive variance)
  - dominance variation,
  - epistatic (multi-genic)
  - interactions
  - maternal and paternal effects

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2}$$

Defining Dominance



If the heterozygote is half way between the two homozygotes, there's a "dose-response" effect, d is zero, and there is no 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

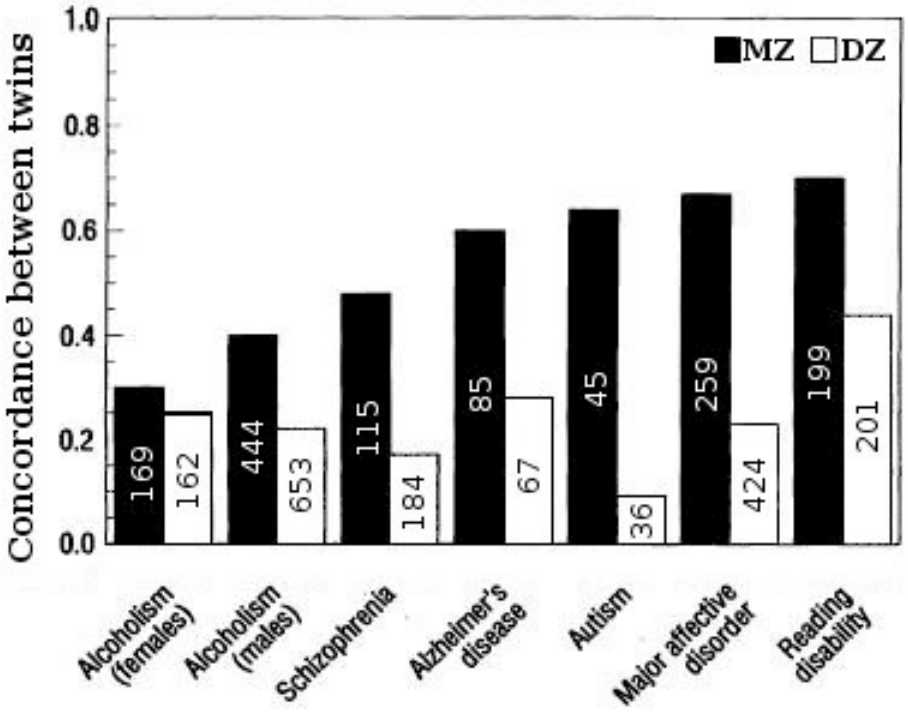
Falconer's Method

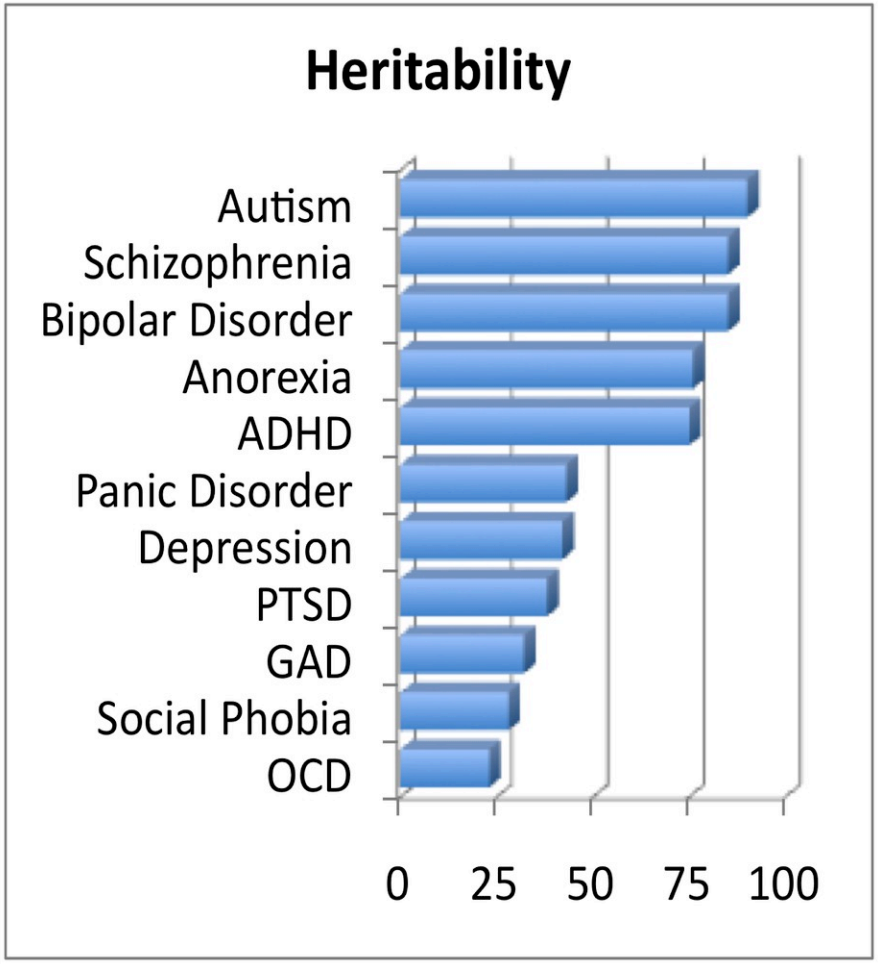
$$h^2 = 2*(r_{MZ} - r_{DZ})$$

r_{MZ}= correlation between monozygotic co-twins

r_{DZ}= correlation between dizygotic co-twins

Twin Concordances

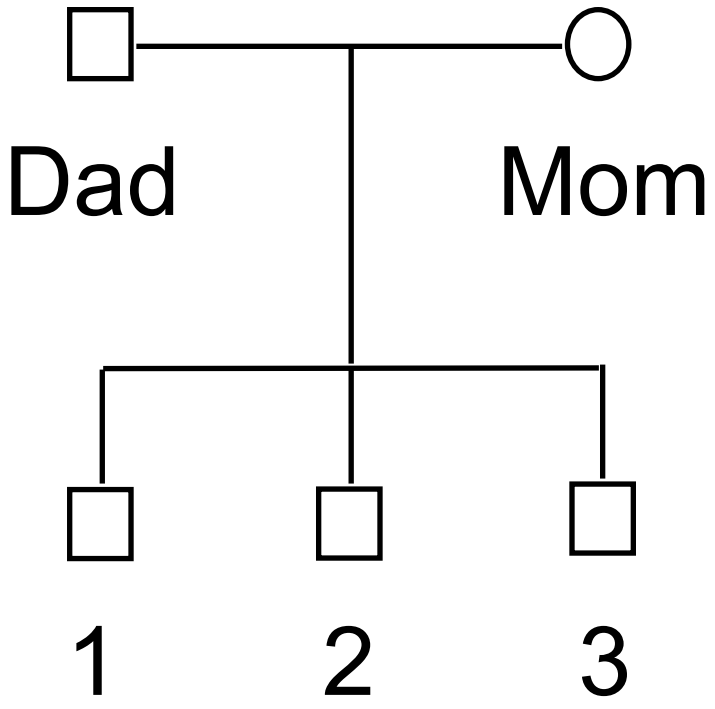




Limitations of Twins

- Common Environment Unless Raised Apart
- Twins reared apart are difficult to find, non-representative
- Common prenatal environment;
  - intrauterine competition
- Mother may be physically stressed

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

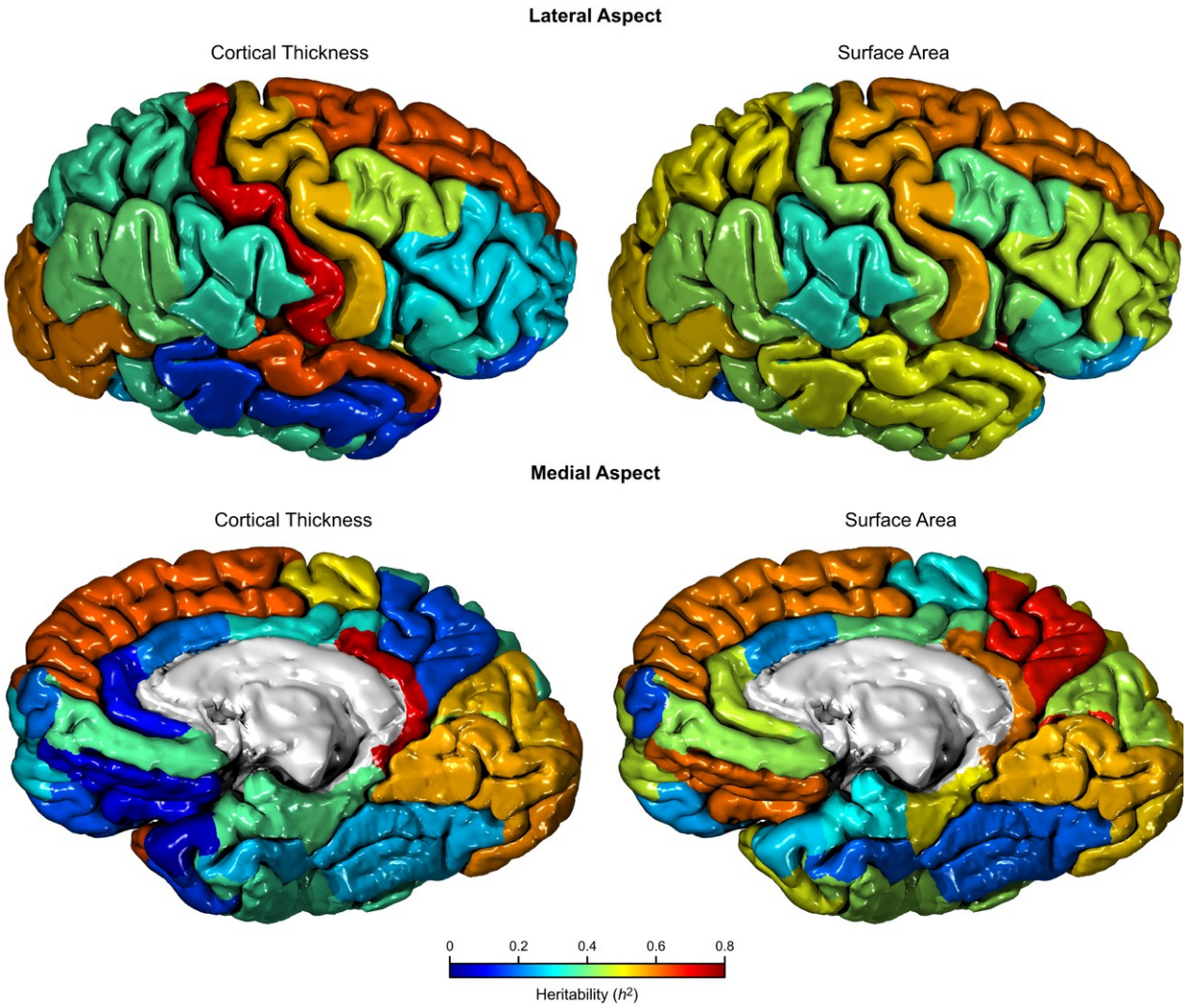
Limitations of Heritability Estimates

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 more genetically tractable if it is due to many genes.

3. Heritability vs. Familiality- A trait can be familial without being heritable.

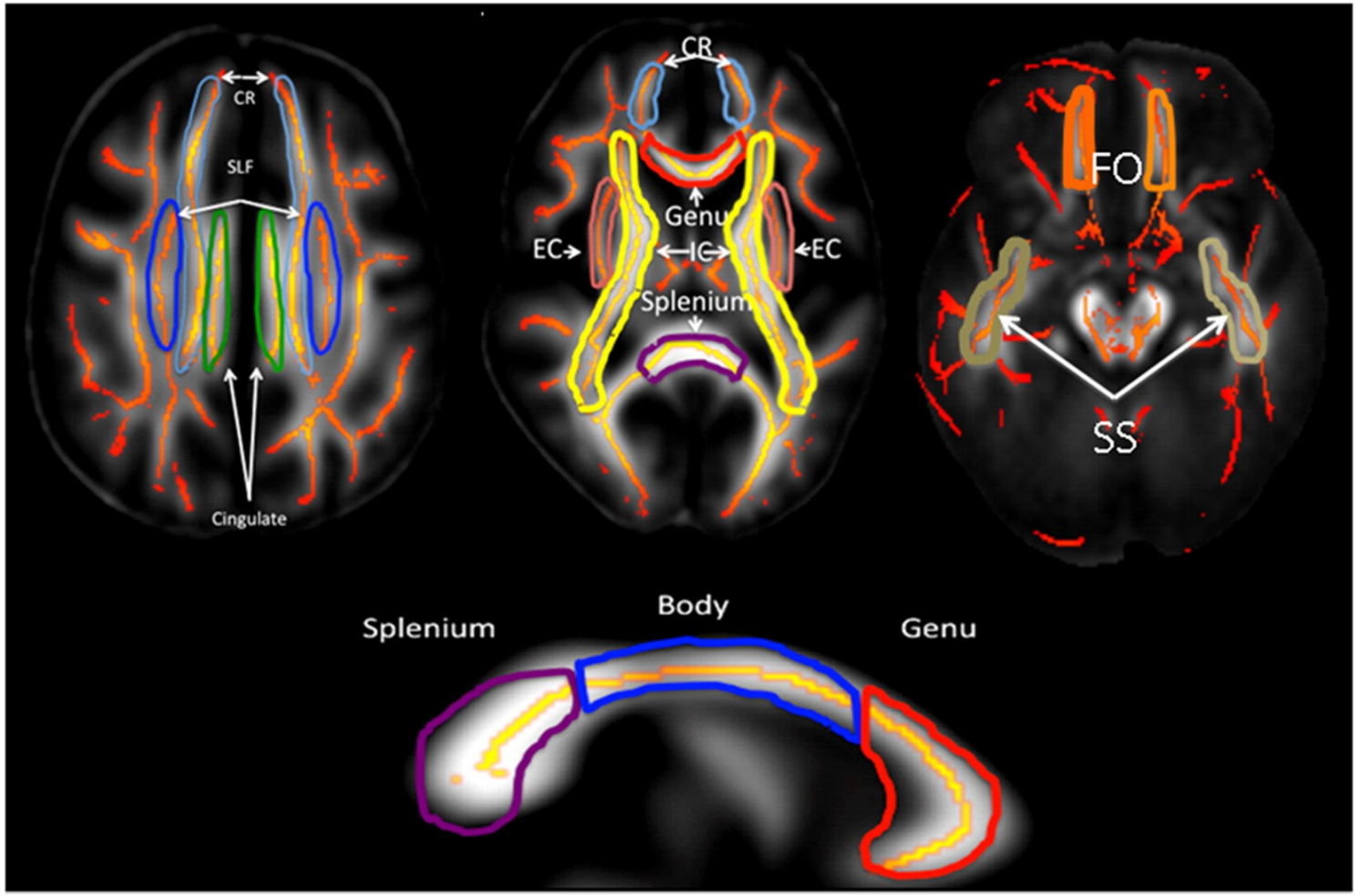
Cortical Thickness & Surface Area



Winkler et al., *NeuroImage*, 2010

486 family members

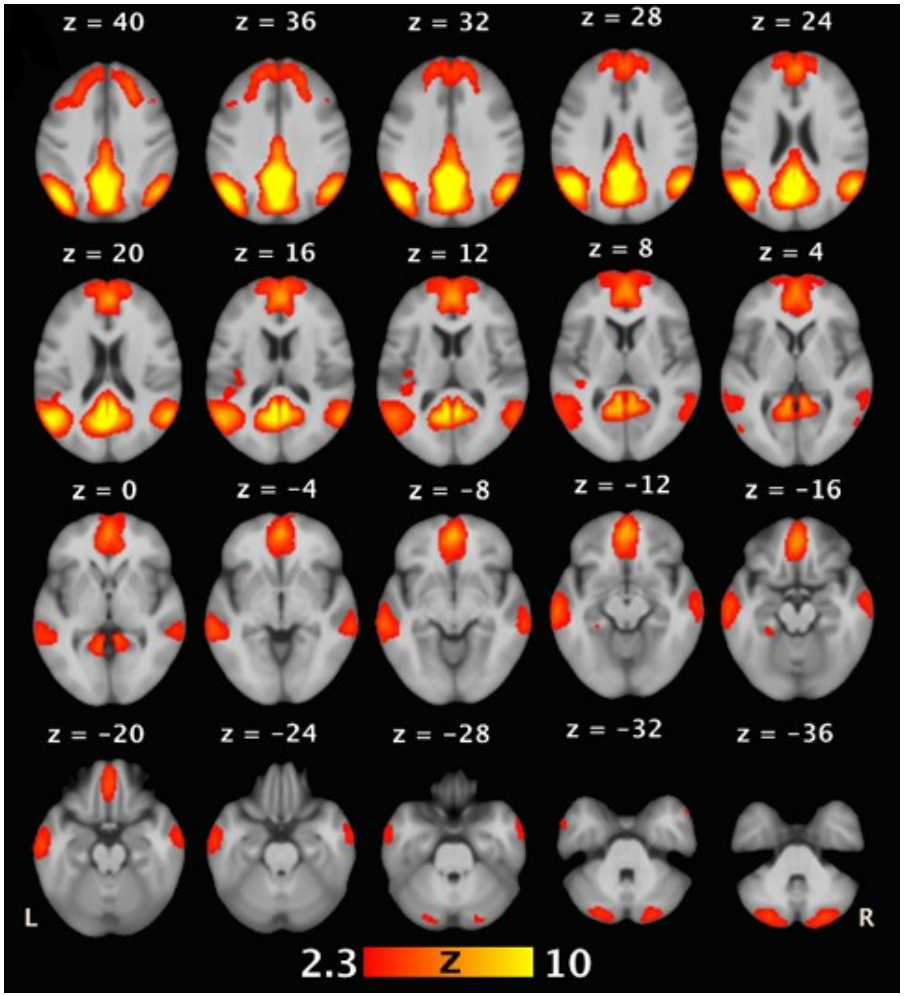
White-Matter Tracts (DTI) h^2



Kochunov et al., *NeuroImage*, 2010

467 family members

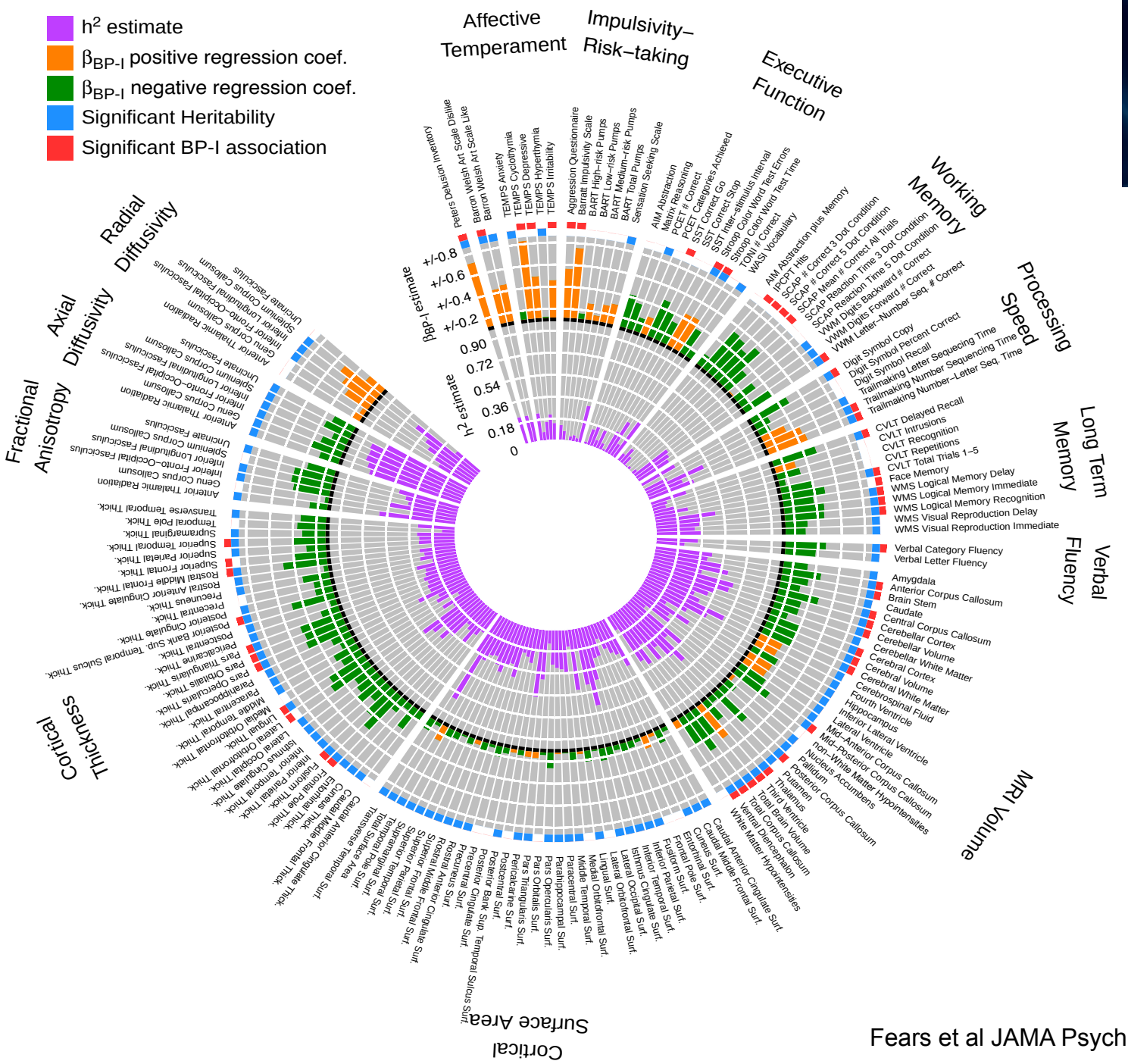
Resting State fMRI Heritability



h^2= 0.424

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

333 family members



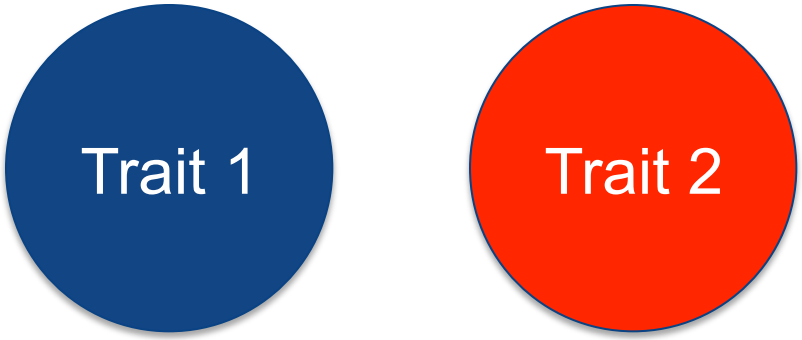
Fears et al JAMA Psych 2014

Question 2: Pleiotropy

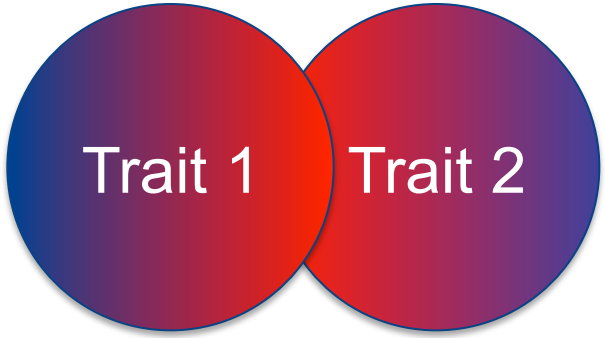
Which traits are influenced by the same genes?

Levels of Pleiotropy

No Pleiotropy



Partial Pleiotropy



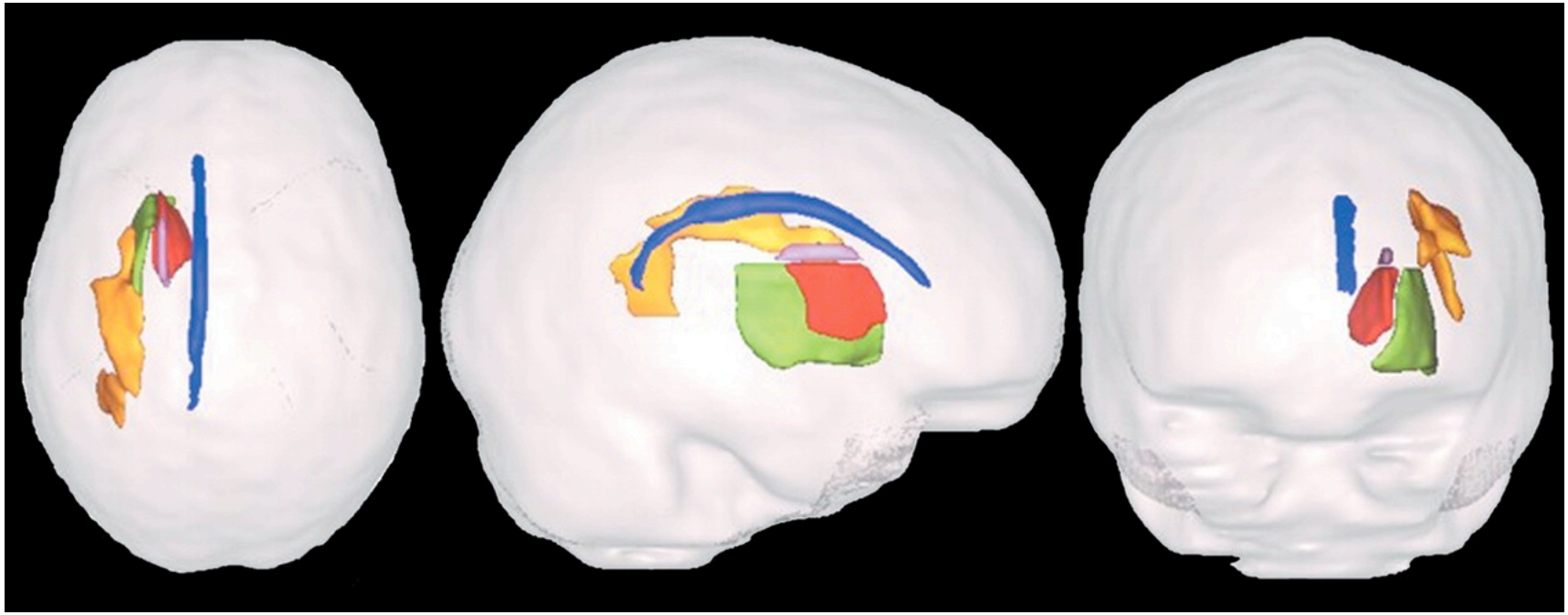
Full Pleiotropy

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

White Matter Tracts & Working Memory

All cognitive & imaging measures were heritable, but only WM performance and SLF integrity shared genetic factors

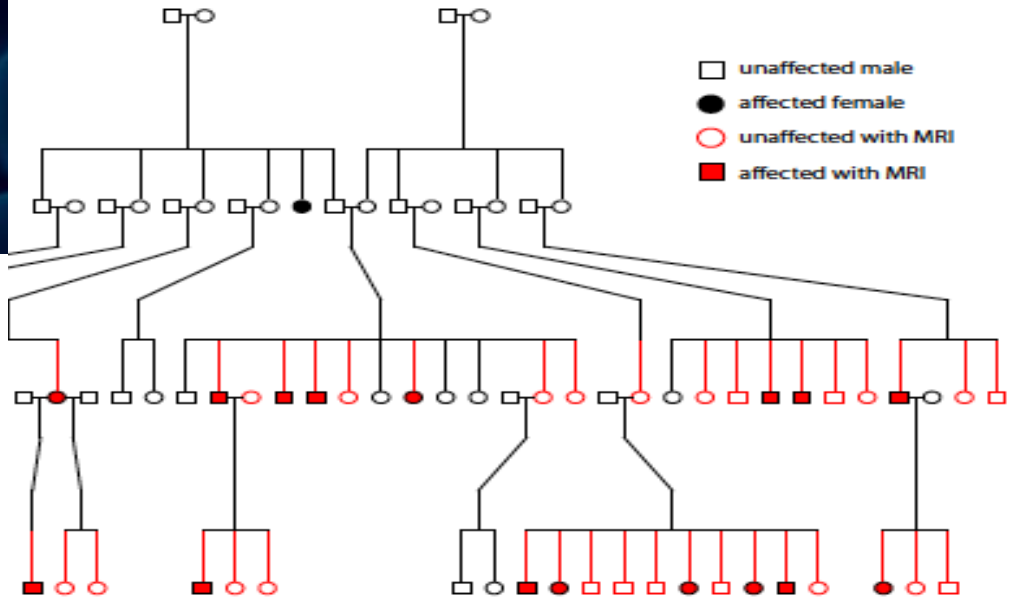


Superior longitudinal fasciculus – Spatial DRT: $\rho_g = 0.593$

Karlsgodt et al., *J Neurosci*, 2010

467 family members

Neuroimaging Phenotypes in Bipolar Pedigrees



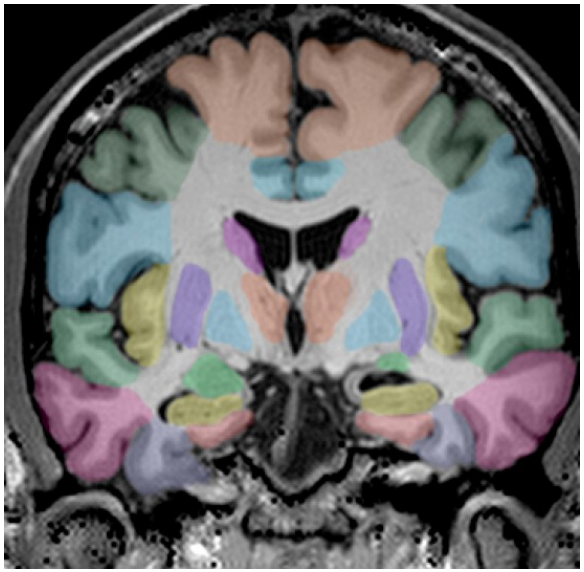
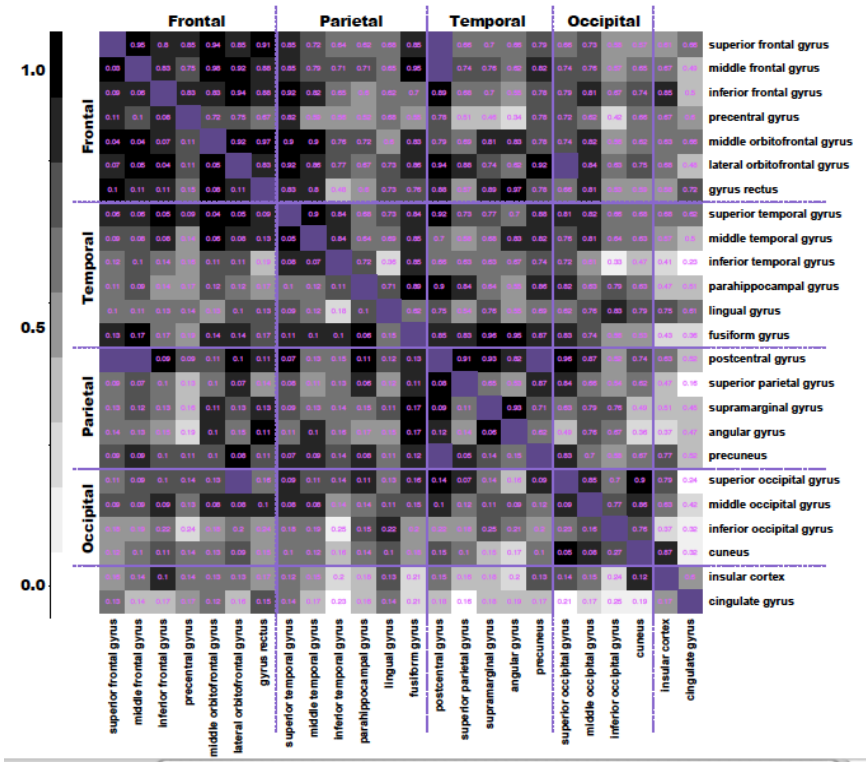


FIGURE 5: Genetic Correlation (ρ_G) for Cortical Volume Phenotypes



Fears et al JAMA Psychiatry 2014

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

What is Association?

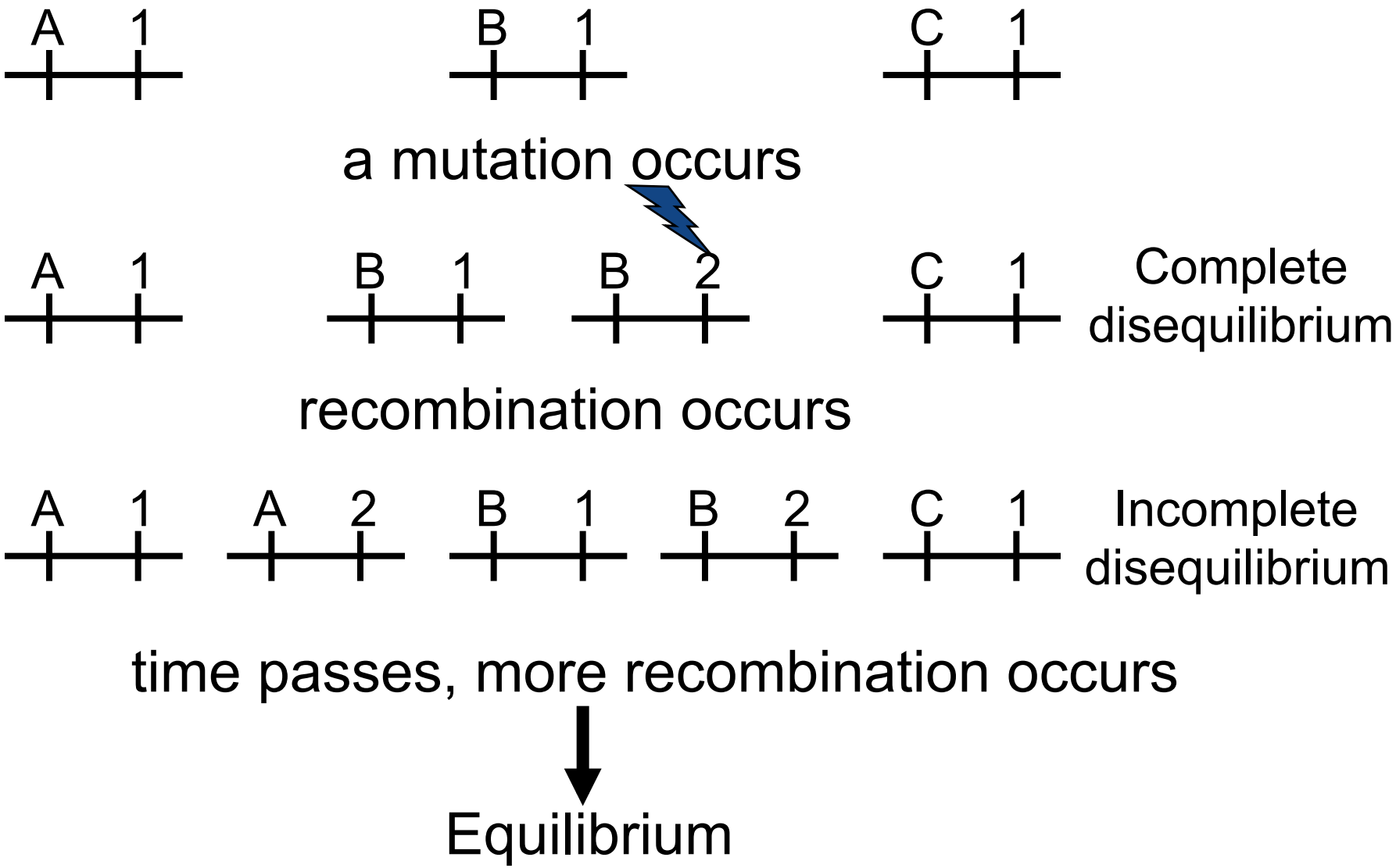
- Tests for correlation between genotype and phenotype

- Association analyses work when:
  - 1) your genotyped marker is a functional polymorphism
  - 2) your genotyped marker is in linkage disequilibrium with a functional polymorphism

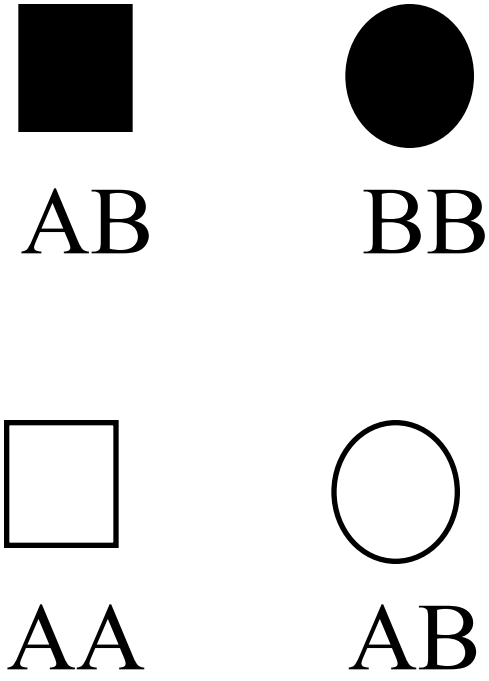
Linkage Disequilibrium (LD)

- Linkage disequilibrium is the non-random association of alleles at two or more loci
- LD = presence of statistical associations between alleles at different loci that differ from what would be expected if alleles were independently, randomly sampled based on individual allele frequencies
- Level of LD is influenced by many factors-genetic linkage, selection, rate of recombination, rate of mutation, genetic drift, non-random mating, and population structure.
- LD is unpredictable

How do we get LD?



Association test for unrelated individuals: Discrete traits



	Aff	Unaff
A	a	b
B	c	d

$$\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Association test for unrelated individuals: Quantitative traits

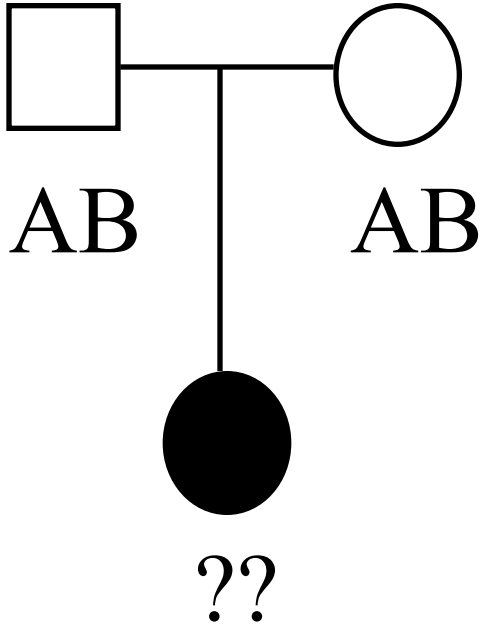
Genotype	N	mean	variance
AA	n_{AA}	μ_{AA}	σ^2_{AA}
AB	n_{AB}	μ_{AB}	σ^2_{AB}
BB	n_{BB}	μ_{BB}	σ^2_{BB}

Assuming the trait values are normally distributed

$$d = \frac{\mu_{AA} - \mu_{BB}}{\sqrt{\frac{\sigma^2_{AA}}{n_{AA}} + \frac{\sigma^2_{BB}}{n_{BB}}}}$$

Transmission disequilibrium test (TDT): association

-Family-based association test for presence of linkage between a genetic marker and a trait (only will detect linkage in the presence of association)



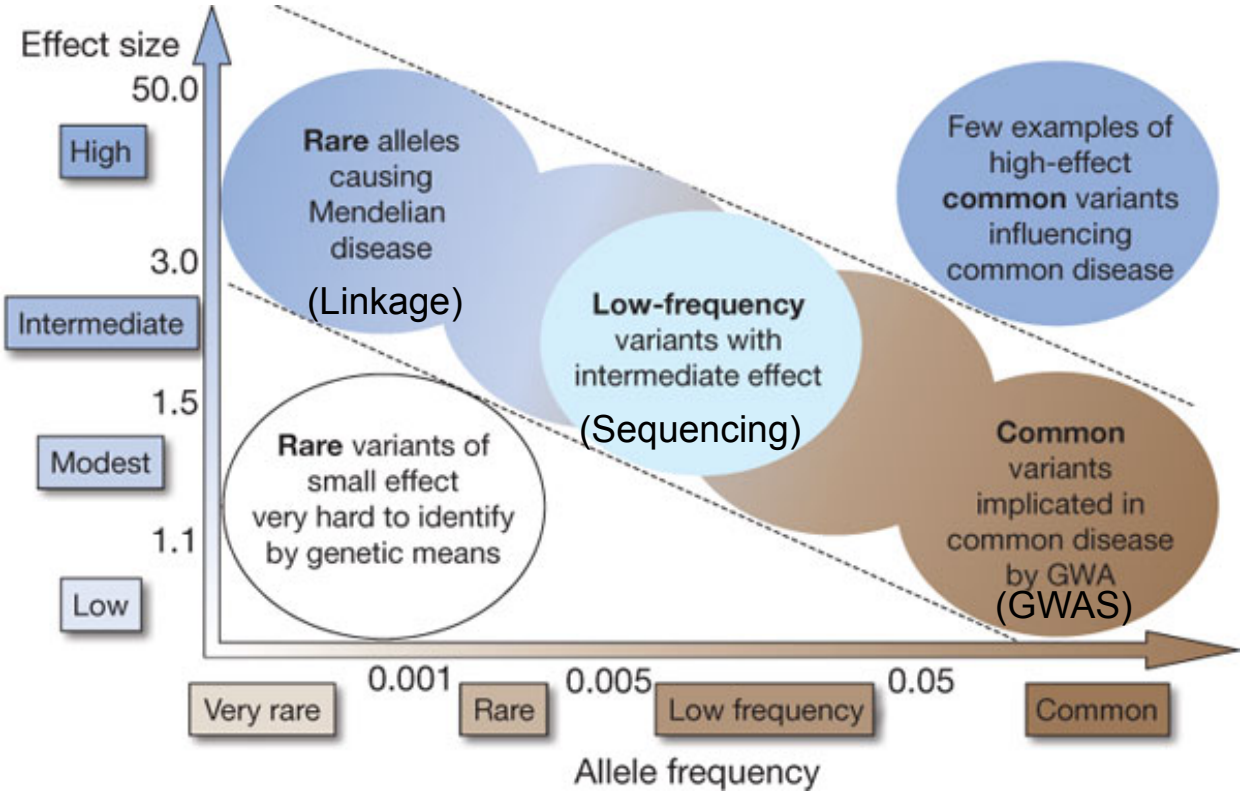
Transmitted allele \ Nontransmitted allele	A	B
A	a=0	b
B	c	d=0

$$\chi^2 = (b\text{-}c)^2/(b+c)$$

Candidate genes: Do you feel lucky?

~3 Billion base pairs

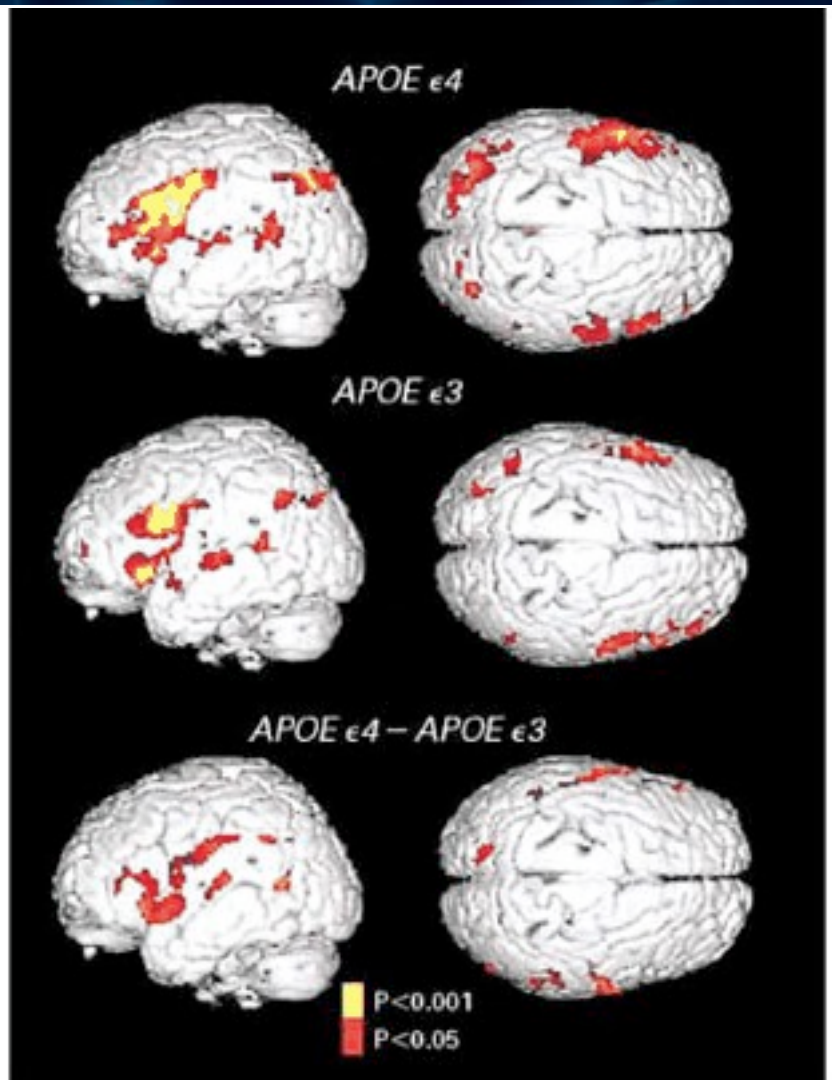
Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).



Manolio *et al. Nature* **461**, 747-753 (2009) doi:10.1038/nature08494

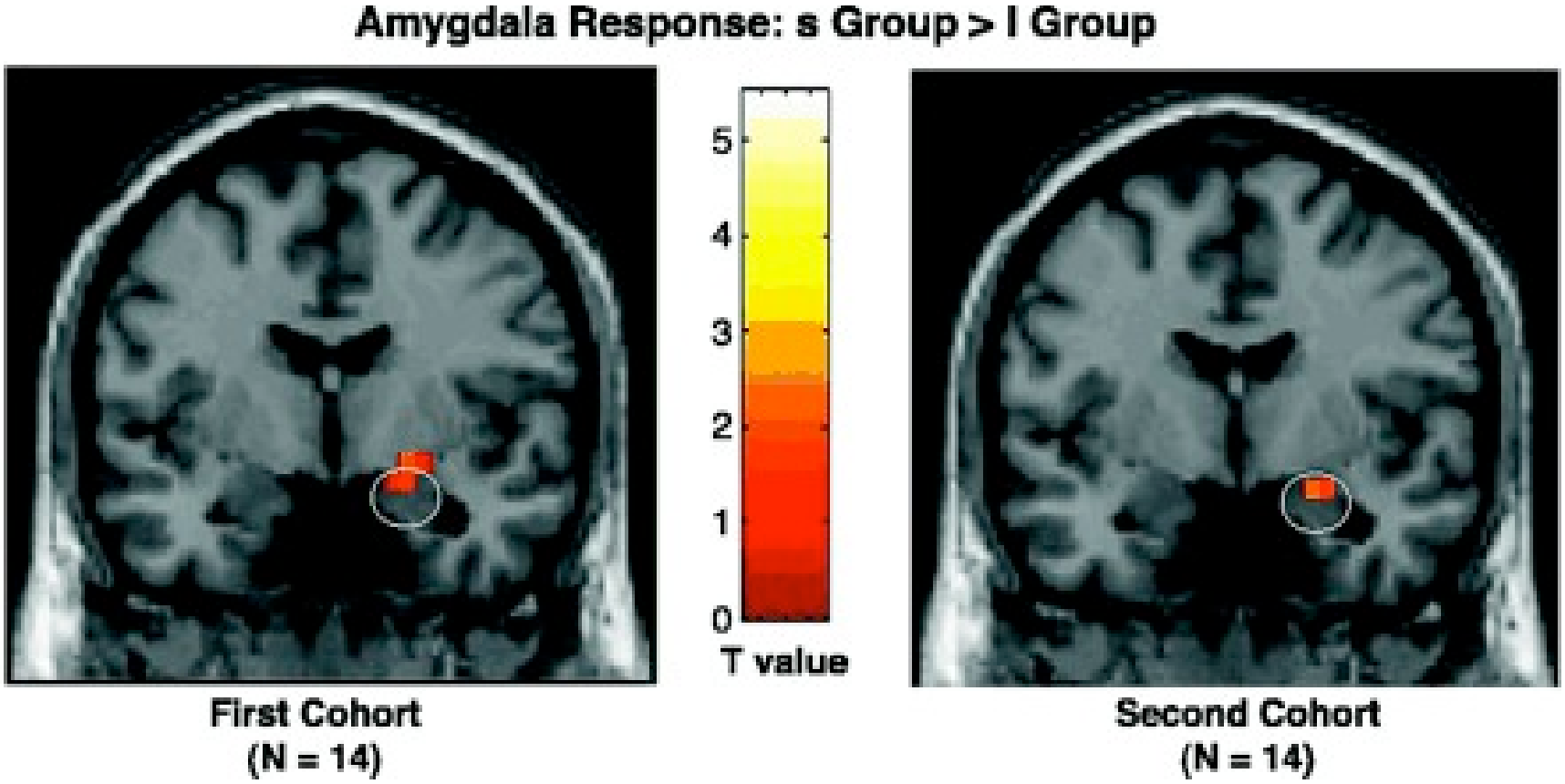
Memory Activation & APOE ε4

- Major Risk gene for Alzheimer dx
- 16 APOE ε4
- 14 APOE ε3
- APOE ε4 allele carriers had increased neural activity during memory task; predicted cognitive decline



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

Fear Response & Serotonin Transporter Gene (SLC6A4): Short allele hetero/homozygotes show greater amygdala reactivity to fearful stimuli



Hariri et al., *Science*, 2002

To obtain this degree of significance- locus must explain ~28% of phenotypic variance

"*Excess significance bias in the literature on brain volume abnormalities*"

Table 2. Observed and Expected Number of "Positive" Study Data Sets Across All Meta-analyses for Each Condition and for Each Brain Structure

	Study Data Sets, No.	Observed Positive Data Sets, No.	Expected Positive Data Sets,[a] No.	Expected Postive Data Sets Under Half-Effect Assumption,[b] No.
According to condition				
Major depressive disorder	109	42	26.8[c]	11.4[d]
Bipolar disorder	191	36	15.6[c]	11.2[c]
Obsessive-compulsive disorder	16	3	0.8[d]	0.8[d]
Posttraumatic stress disorder	44	17	4.8[c]	2.8[c]
Autism	28	12	3.1[c]	1.7[c]
First-episode schizophrenia	48	26	22.3	7.2[c]
Relative of patient with schizophrenia	25	6	4.9	2.0[d]
According to brain structure				
Anterior cingulate cortex	8	4	4.9	1.7[c]
Orbitofrontal cortex	7	4	2.9	1.2[d]
Prefrontal cortex	7	1	1.7	0.9
Hippocampus	138	44	28.5[c]	11.9[c]
Putamen	20	3	3.4	1.7
Caudate nucleus	48	5	2.0[d]	2.0[d]
Amygdala	82	32	4.7[c]	4.3[c]
Lateral ventricles	35	15	14	3.5[c]
Third ventricle	27	9	6.8	2.5[c]
Gray matter	21	5	2.0[d]	1.2[c]
White matter	14	1	0.8	0.7
Globus pallidus	6	2	0.5[d]	0.4[d]
Thalamus	10	3	0.5[c]	0.5[c]
Temporal lobe	24	6	1.6[c]	1.2[c]
Vermal lobules	22	8	2.6[c]	1.4[c]

[a] Based on the assumption that the plausible effect is the one seen in each meta-analysis of each particular brain structure and condition.
[b] Based on the assumption that the plausible effect is half of what is seen in each meta-analysis.
[c] Statistically significant difference between expected and observed even after Bonferroni adjustment for the total number of tested conditions or brain structures.
[d] Nominally statistically significant difference between expected and observed.

21/41 meta-analyses found statistically significant associations; 142/461 (31%) data sets had positive results. Even if the summary effect sizes of the meta-analyses were unbiased, the expected number of positive results would have been only 78.5 compared with the observed number (142; $P < .001$).

Ioaniddis Arch Gen Psych 2012

Enhancing the Informativeness and Replicability of Imaging Genomics Studies

Carter CS, Bearden CE, Bullmore ET, Geschwind DH, Glahn DC, Gur RE, Meyer-Lindenberg A, Weinberger DR

-False positive problem
-How can we appropriately minimize search space?
-Target genes with association to disease-related quantitative trait (not necessarily disease itself)
-Genes with known syndromic associations with high risk
-Genes with known anatomic expression sites (e.g., fetal expression studies, knock-out models)

Approaches to Genotyping

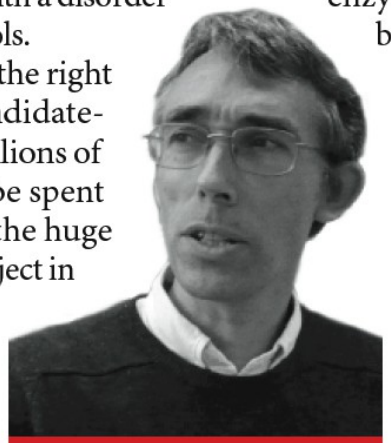
Candidate genes: genotype only markers in genes potentially related to the trait.

- 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 novel finding

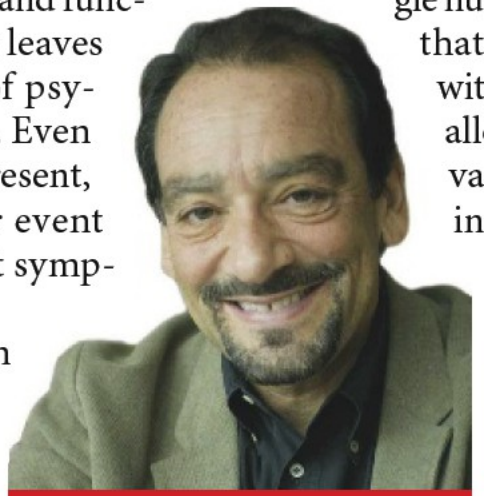
Genome screen: genotype anonymous markers spanning the genome at regular intervals

- 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

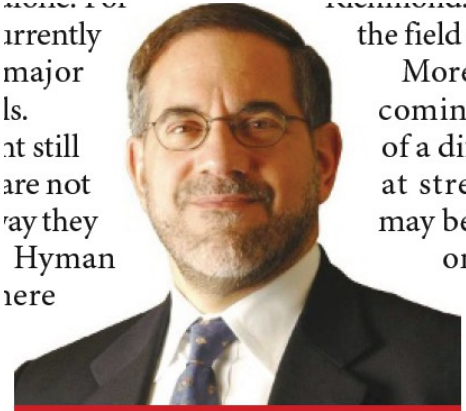
Candidate Gene Controversy



"The studies so far are statistically underpowered. We need bigger studies."
— Jonathan Flint



"Geneticists know nothing about psychiatric disease."
— Daniel Weinberger

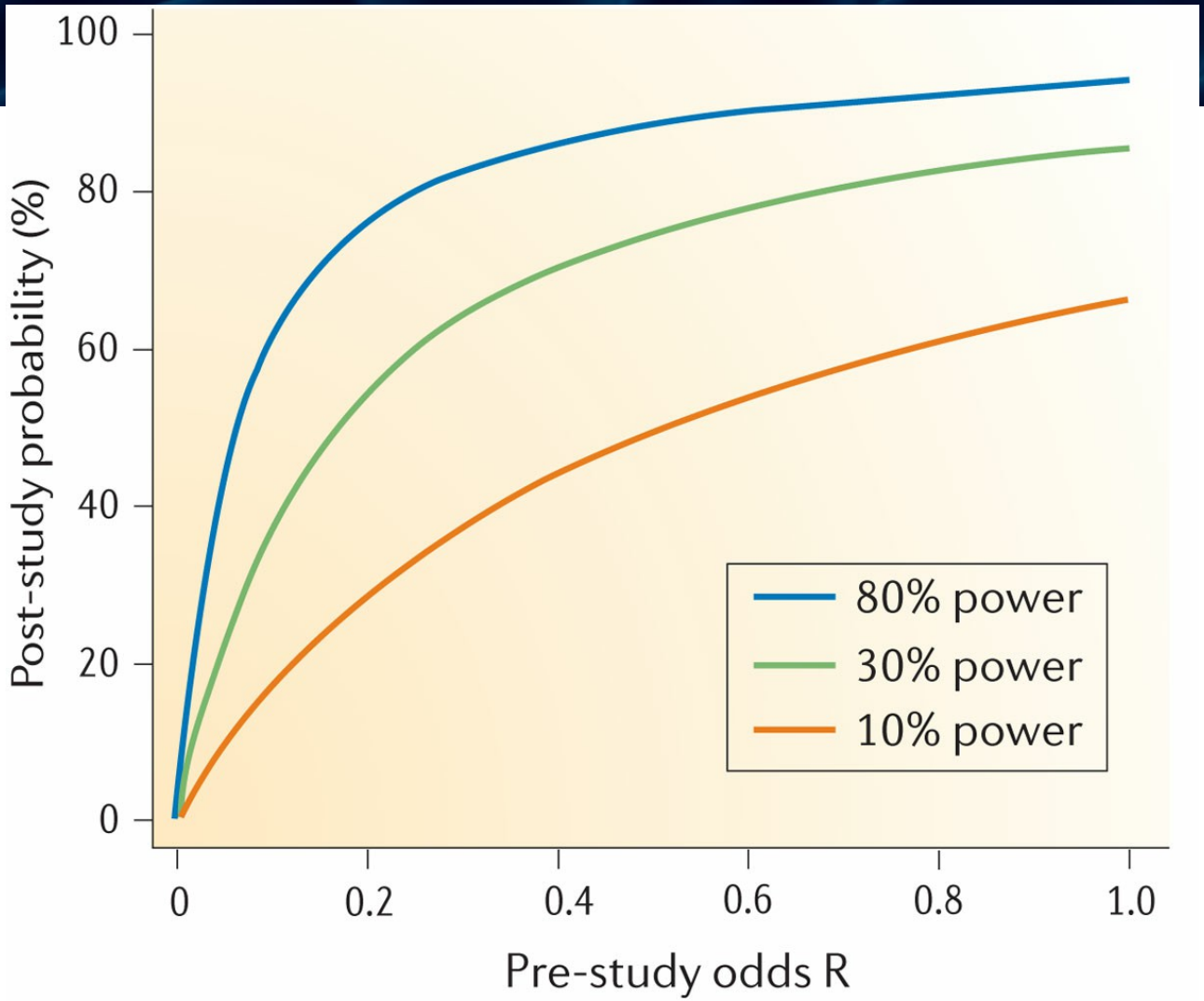


"We are just too ignorant of the underlying neurobiology to make guesses about candidate genes." —Steven Hyman

But Flint argues that no one has yet shown that intermediate phenotypes have a tighter link to the genes responsible than the disease itself. "I just don't buy that brain size or whatever will work better," he says.

"Candidate genes are like packing your own lunch box and then looking in the box to see what's in it."

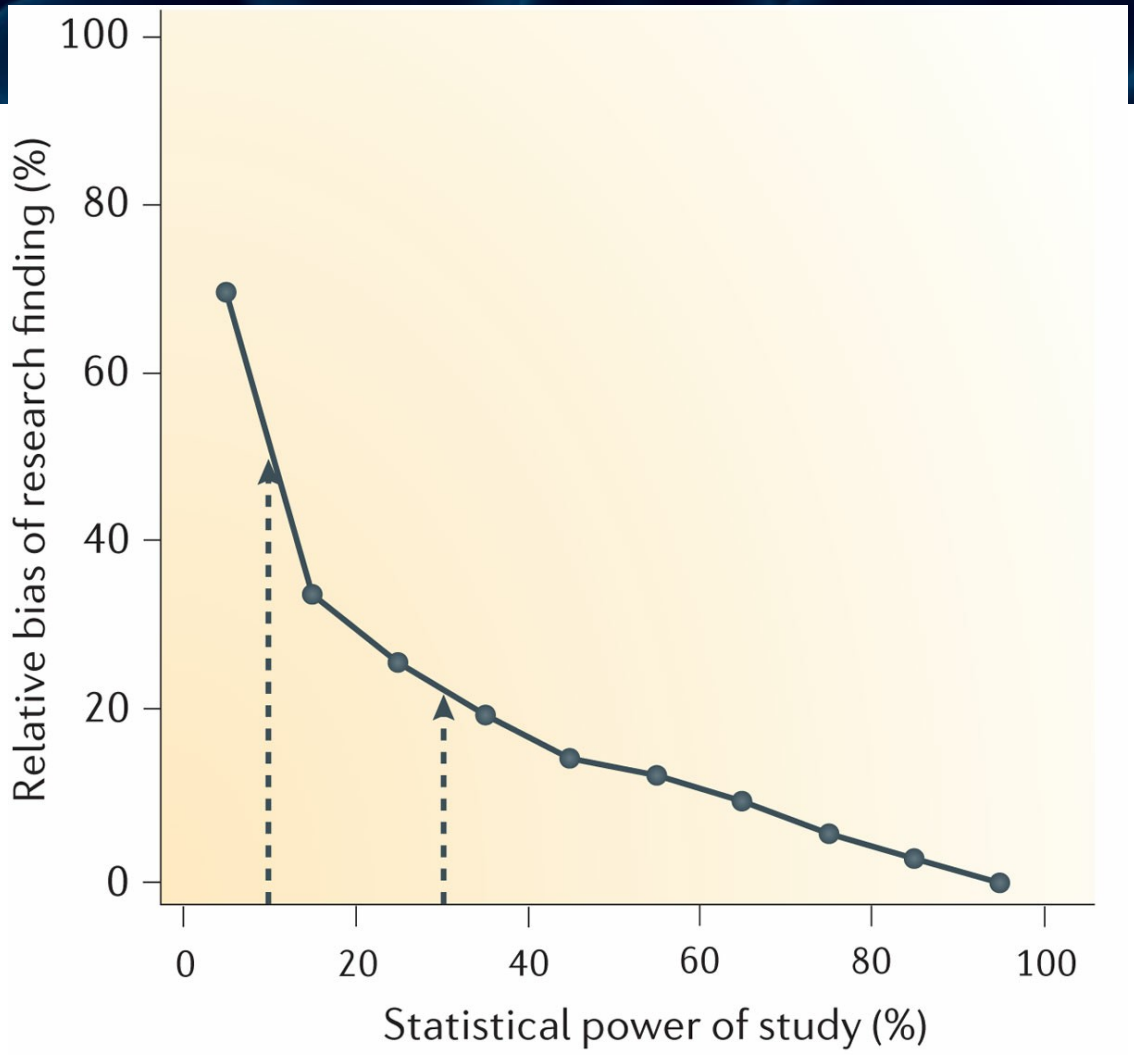
Positive predictive value as a function of the pre-study odds of association for different levels of statistical power.



Button et al 2013

Nature Reviews | Neuroscience

Winner's curse: effect size inflation as a function of statistical power.



Button et al 2013

Nature Reviews | Neuroscience

What if you don't know the allele?

A genome-wide association (GWA) study is an approach that involves rapidly scanning markers across the complete sets of DNA, or genomes, of many people to find genetic variations associated with a trait.

GWA studies identify SNPs and other variants in DNA associated with a disease, but cannot on their own specify which genes are causal

Currently: ~1.5Mil SNPs

Multiple testing: p-values

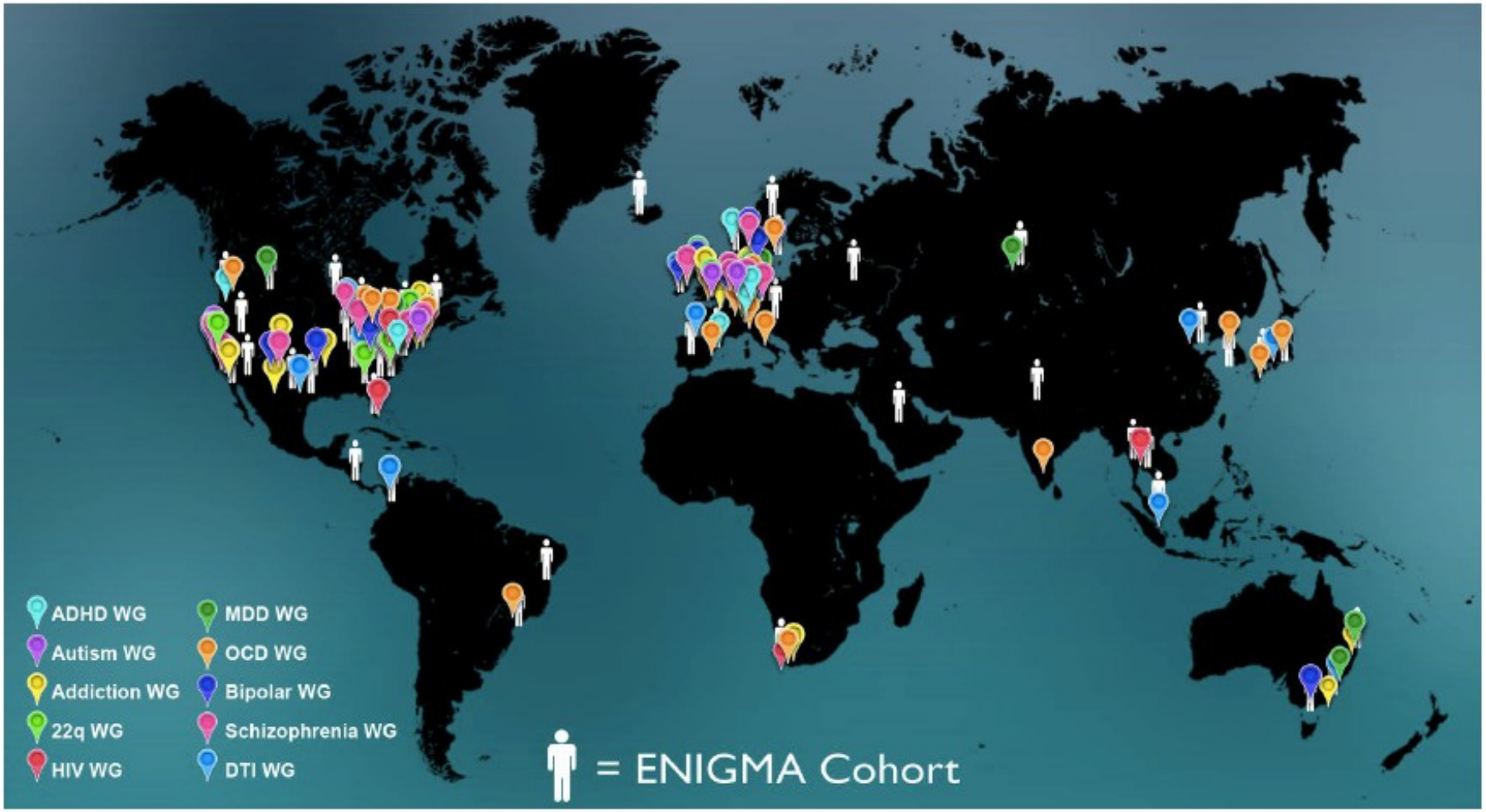
- A p-value of 0.05 implies that 5% of the time we will reject the null hypothesis (i.e. conclude that we have an association) when the null hypothesis is actually correct

- 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

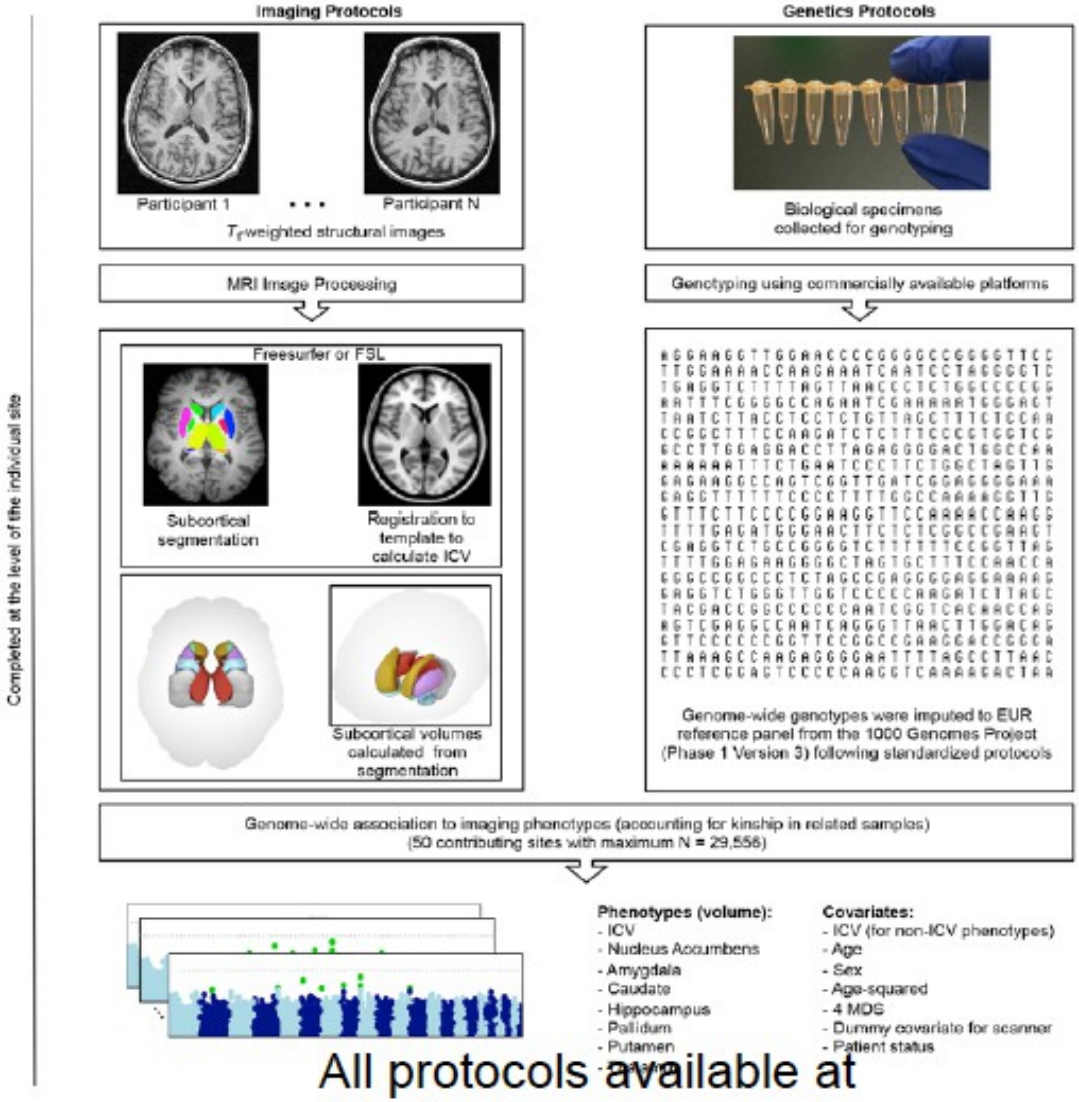
- The simplest correction is the Bonferroni: multiply each p-value by the total number of tests, or divide the significance threshold required by the number of tests (0.05 / #).
- Genome-wide significance (~1.5 Mil SNPs) requires $p=5x10^{-8}$
- This maintains an experiment-wide significance threshold, but may be too conservative when the tests are correlated, e.g. if some markers tested are in LD with each other.

A world-wide collaborative initiative to find replicable genetic influences on brain structure



> 185 institutions, 300+ co-authors, world-wide consortium

Procedure for meta-analysis (Part 1)

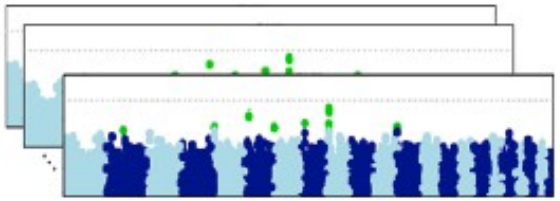


All protocols available at http://enigma.ini.usc.edu

Procedure for meta-analysis (Part 2)

Uploaded to ENIGMA server for analysis at central site

Genome-wide association to imaging phenotypes (accounting for kinship in related samples)
(50 contributing sites with maximum N = 29,556)

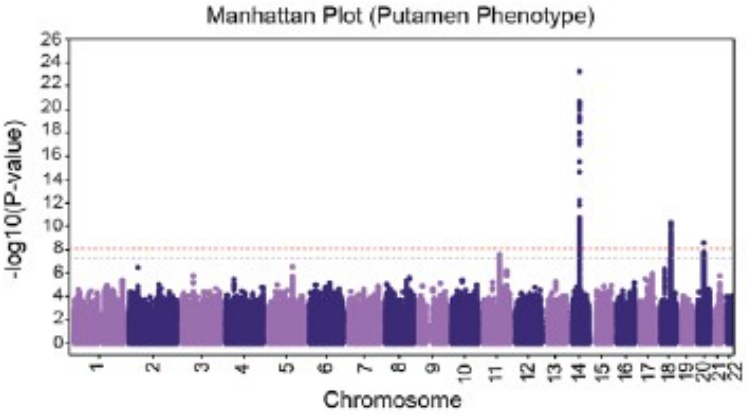


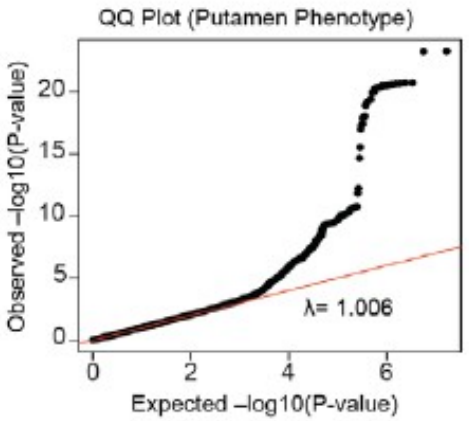
Phenotypes (volume):
- ICV
- Nucleus Accumbens
- Amygdala
- Caudate
- Hippocampus
- Pallidum
- Putamen
- Thalamus

Covariates:
- ICV (for non-ICV phenotypes)
- Age
- Sex
- Age-squared
- 4 MDS
- Dummy covariate for scanner
- Patient status

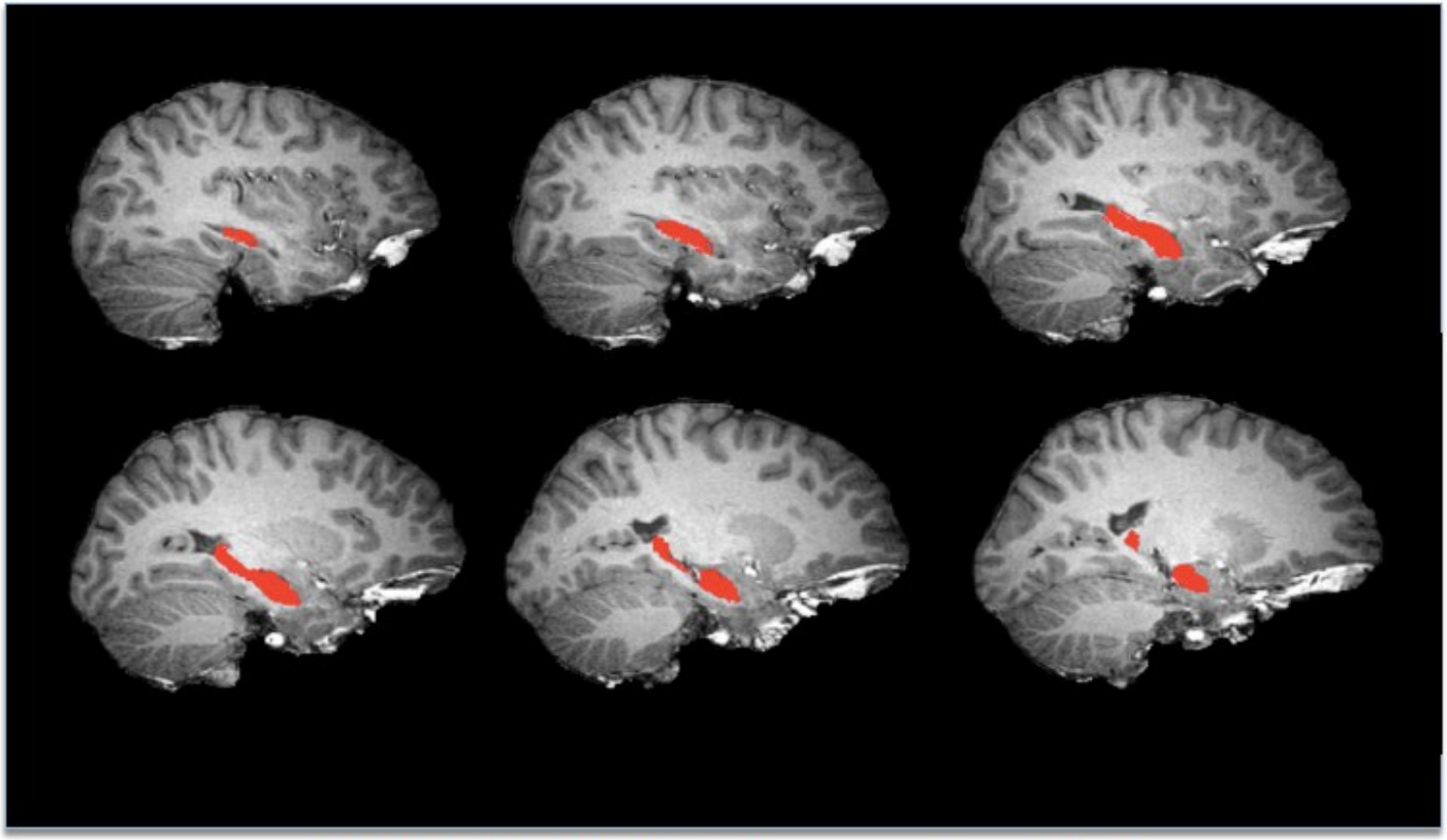
Quality Checking and Filtering (MAC < 10, R^2 < 0.5)

Meta-analyses
Fixed-effect, inverse variance-weighted model





ENIGMA1: Pilot Project Hippocampal and Intracranial Volume GWAS Meta-analysis



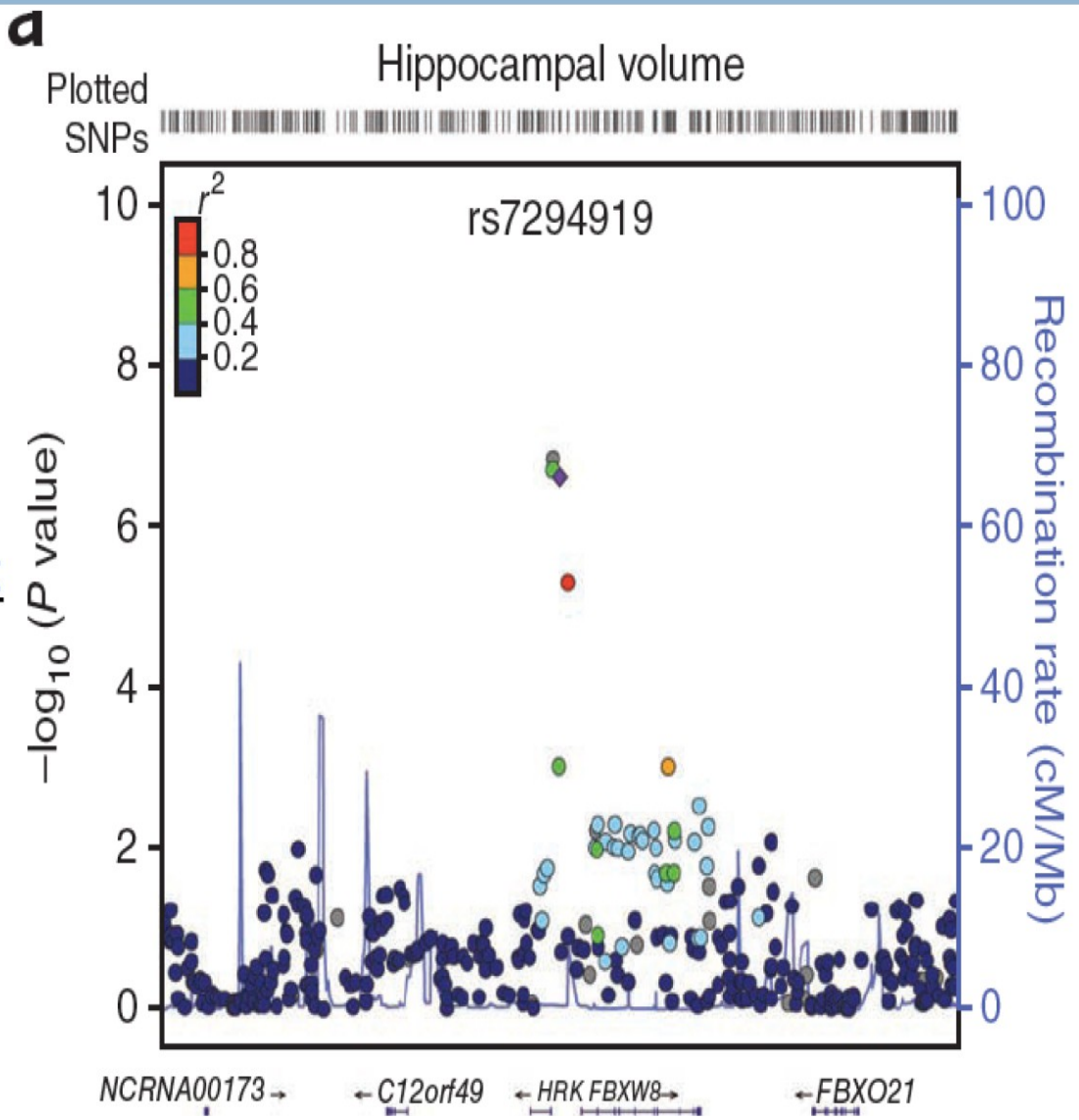
21,151 individuals in discovery + replication

(Stein et al., *Nature Genetics*, 2012)

After Replication:

N = 21,151;

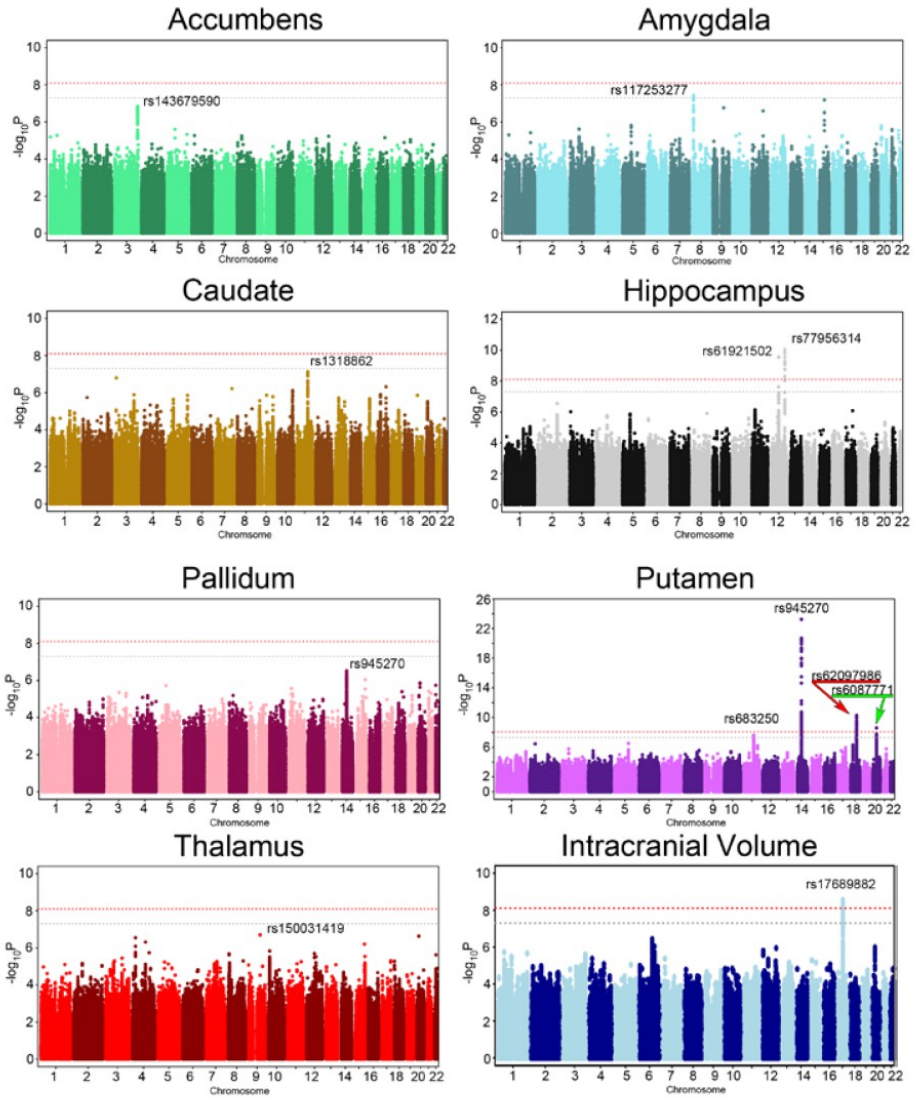
$P = 6.70 \times 10^{-16}$



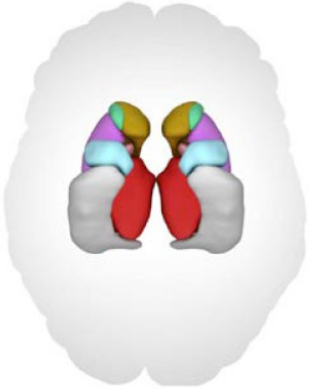
Associated with 1.2% decrease in the average hippocampal volume per risk allele

Stein et al 2012

Five novel genetic variants identified; hippocampal and ICV results replicated



P<7.1x10-9



Discovery sample: 13,171 individuals

Hibar et al Nature 2015

Conclusions..?

- Strongest evidence for novel intergenic locus (rs945270; 14q22.3) near KTN1 gene which influences putamen volume; encodes kinectin (receptor involved in organelle transport)
- Pathway analysis- SNPs associated with putamen volume showed enrichment of genes involved in apoptosis and axon guidance pathways
- Functional validation study- looked for association with gene expression 1 Mb up/downstream. C allele associated with increased KTN1 expression in frontal cortex and in blood
- Kinectin only found in dendrites and soma, not cell bodies- volumetric effects may reflect genetic control of neuronal cell size and/or dendritic complexity

Determining Association Power

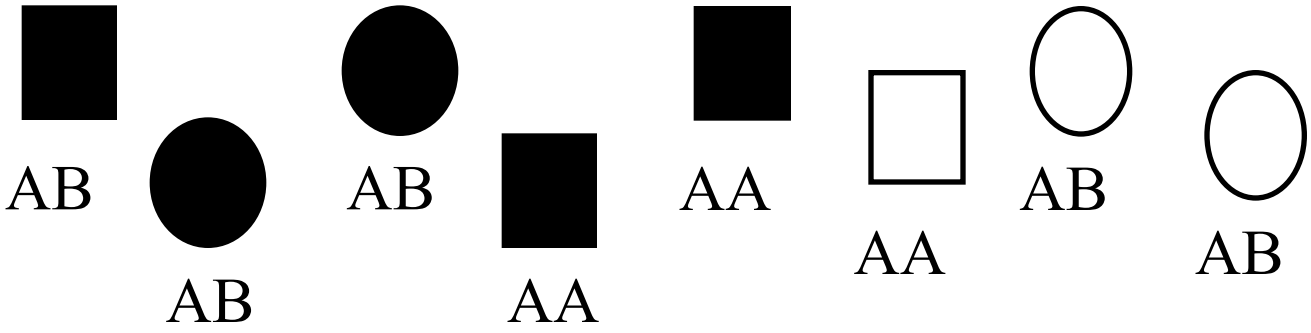
The power to find association is a function of:

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

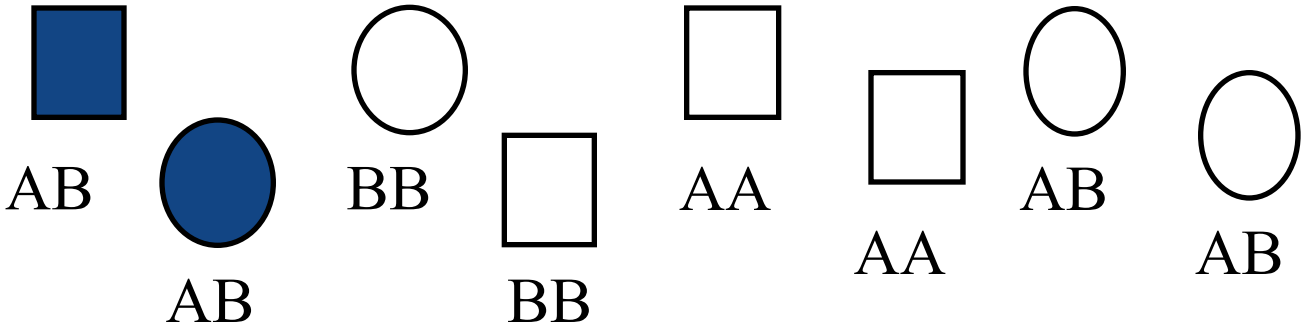
Limitations of Association

- 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

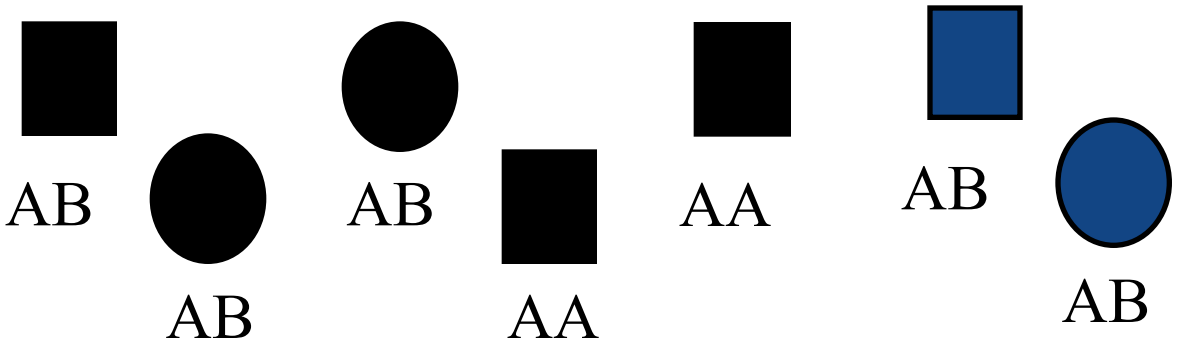


African Americans
70% A,
30% B

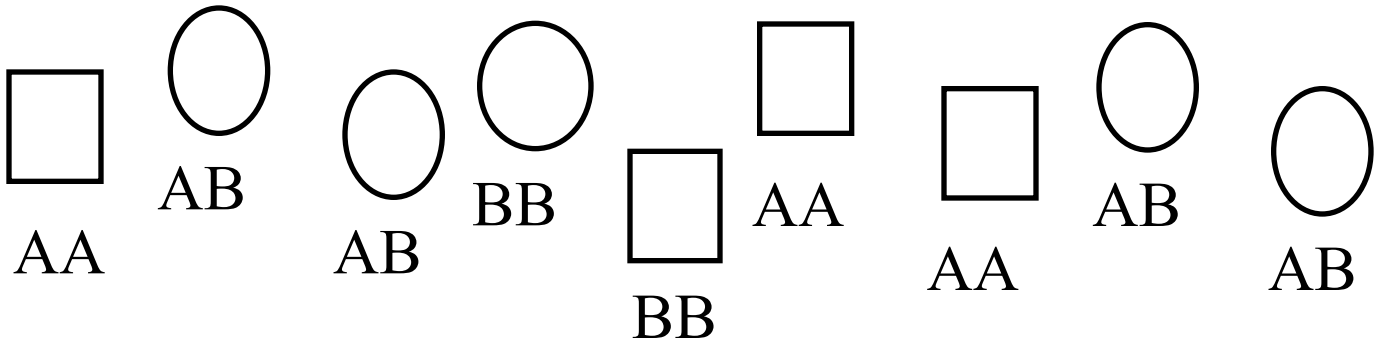


European Americans
50% A,
50% B

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.

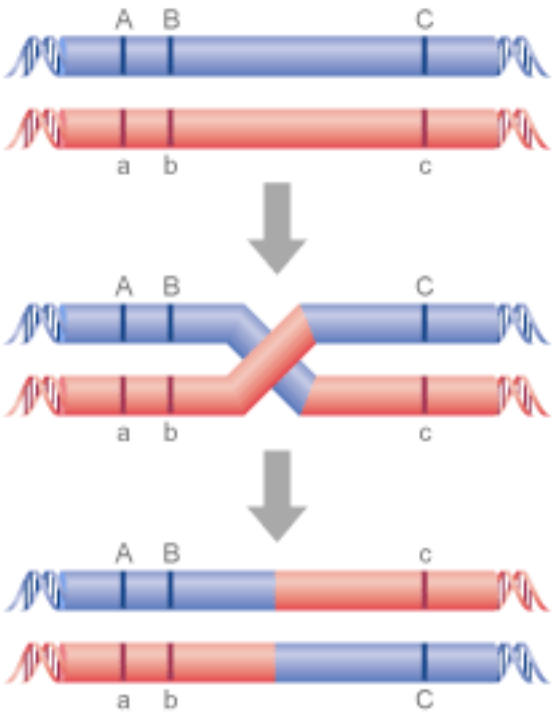
2. Use one of a variety of statistical approaches designed to deal with population stratification (e.g. TDT, genomic control)

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

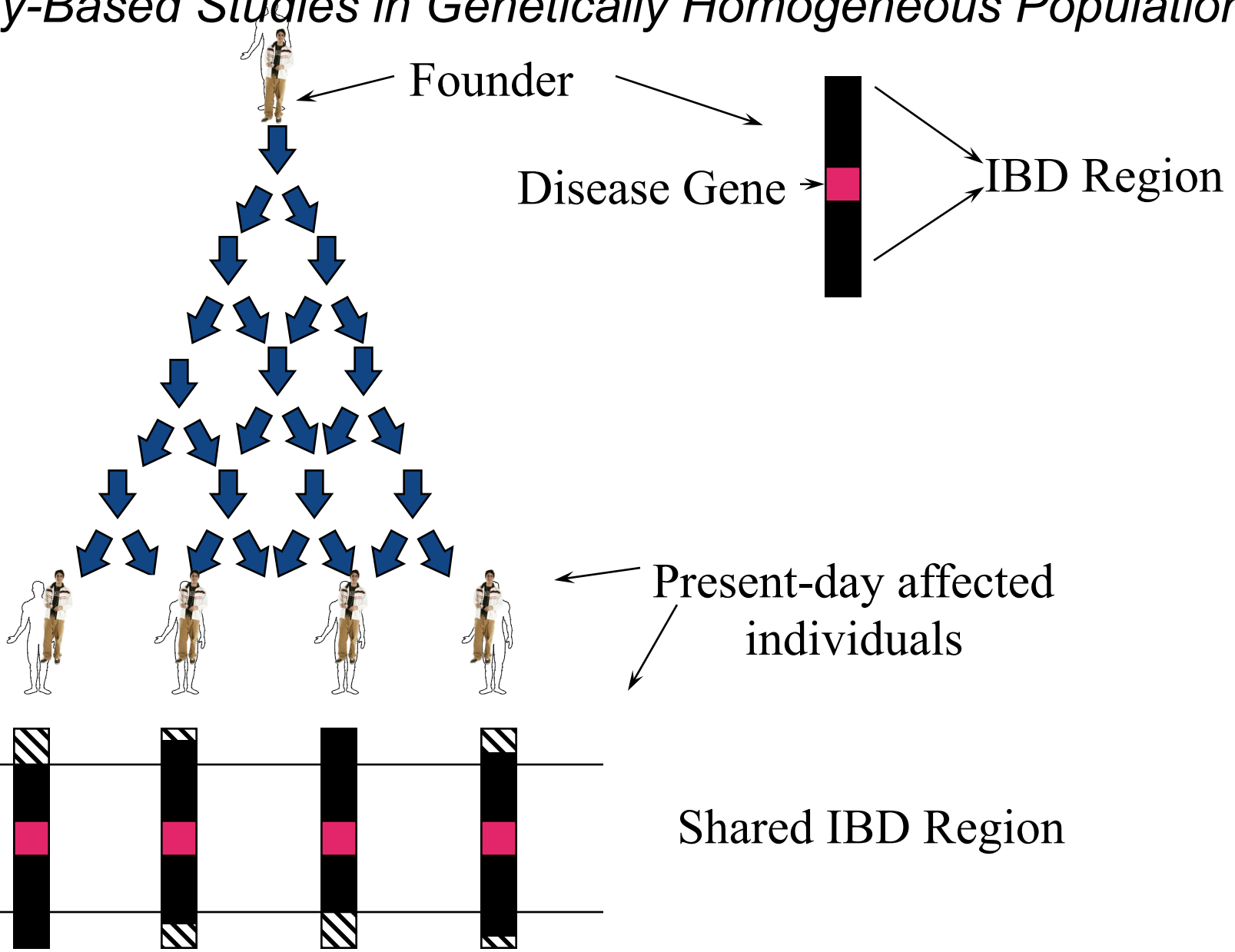
LOD = statistical estimate of whether 2 genes (a gene and a disease gene) are likely to be located near each other & thus inherited together

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

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

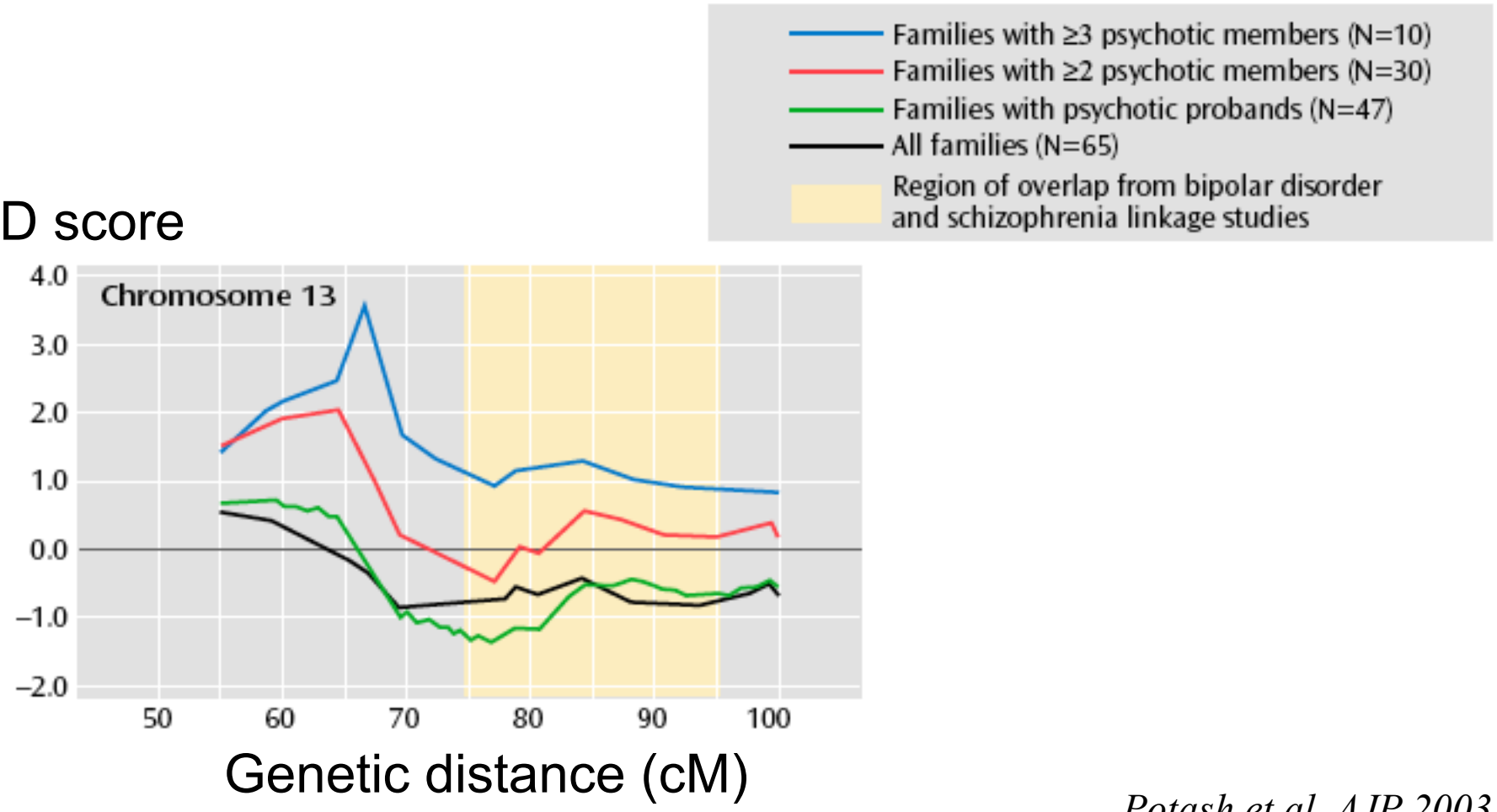
A LOD score >=3.0 is considered evidence for linkage
A LOD score of 3 indicates 1000 to 1 odds that the linkage being observed did not occur by chance
A LOD score <=-2.0 is considered evidence to exclude linkage

Family-Based Studies in Genetically Homogeneous Populations



IBD= Identical By Descent

Example of Linkage Map



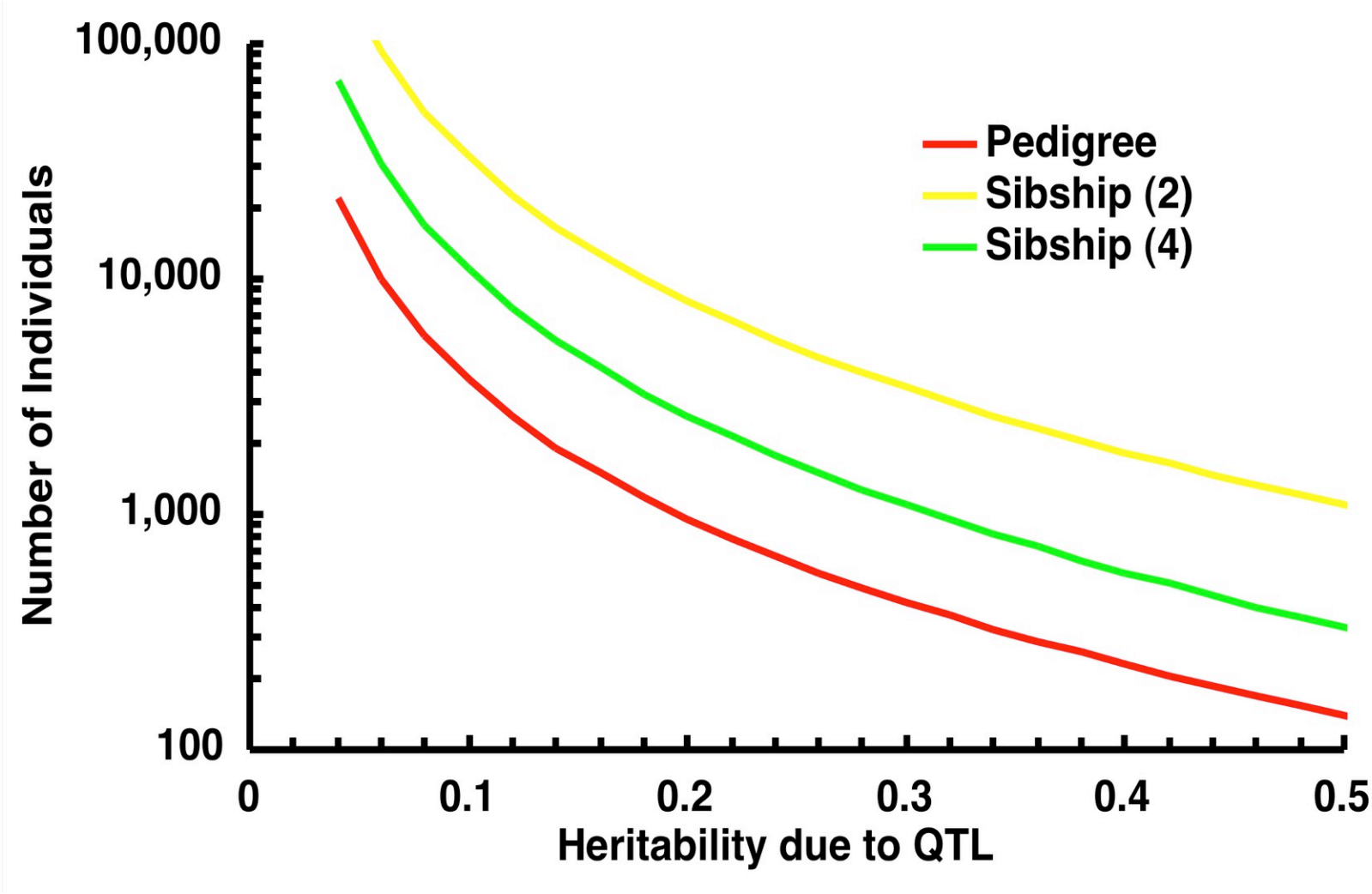
Potash et al. AJP 2003

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



Linkage vs. Association

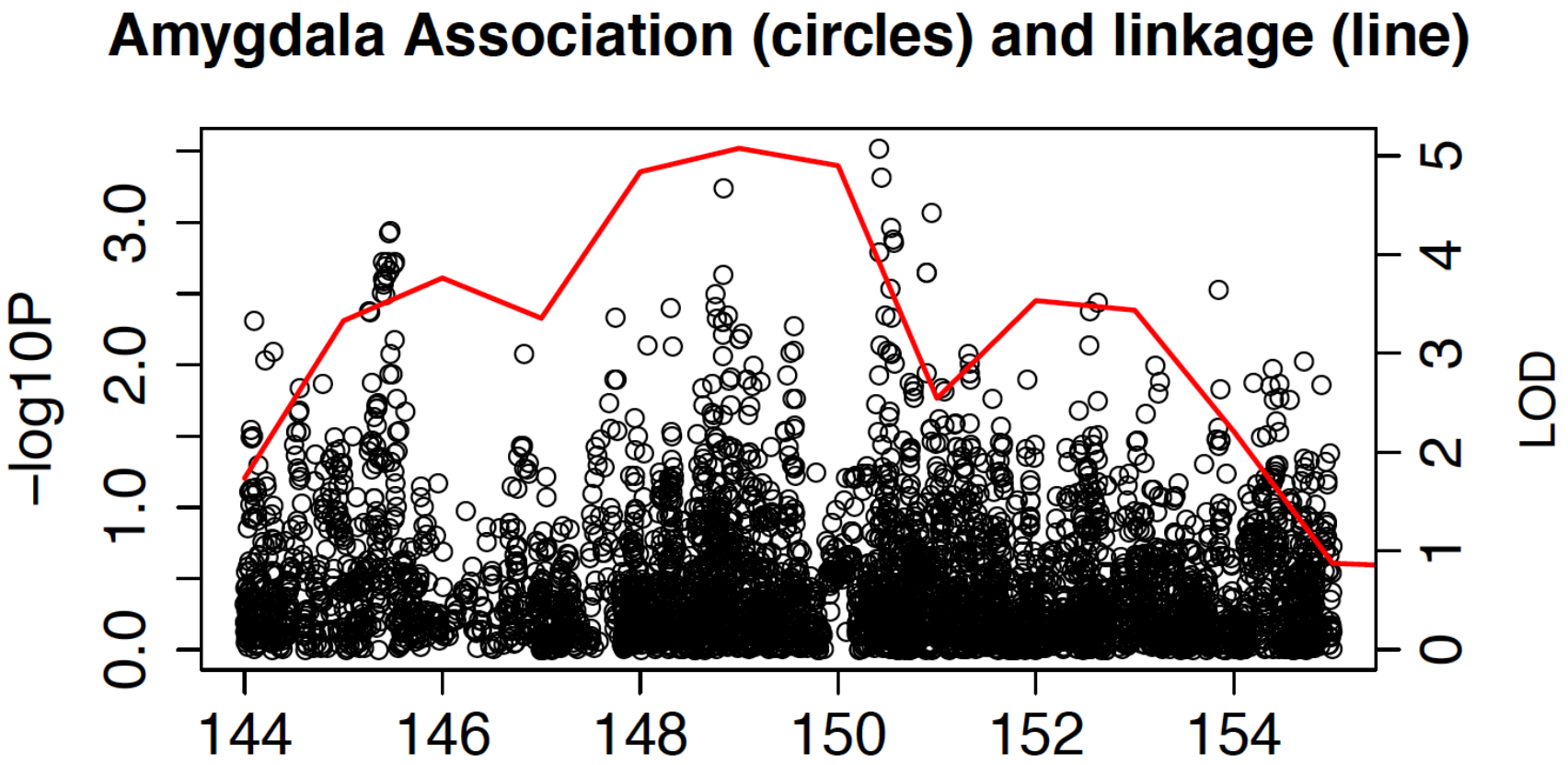
Association: testing for an excess of a specific combination of alleles at two loci. The same alleles must be traveling together at a population level. Detects effects of common variants.

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

Combined Linkage/Association Analysis

- Best of both worlds QTL localization approach
- Linkage can detect cumulative effect of multiple variants (including rare variants).
- Association detects effects of common variants.
- Joint test of linkage/association more powerful than association alone when there is linkage. Only minor loss of power in the absence of linkage.
- Implemented in SOLAR

Combined Linkage and Association Signal for Amygdala Volume in Latin American Pedigrees (n~580)



Fears et al in prep

Question 4: Identification

What specific genes influence the trait?

Identifying a Causal Gene

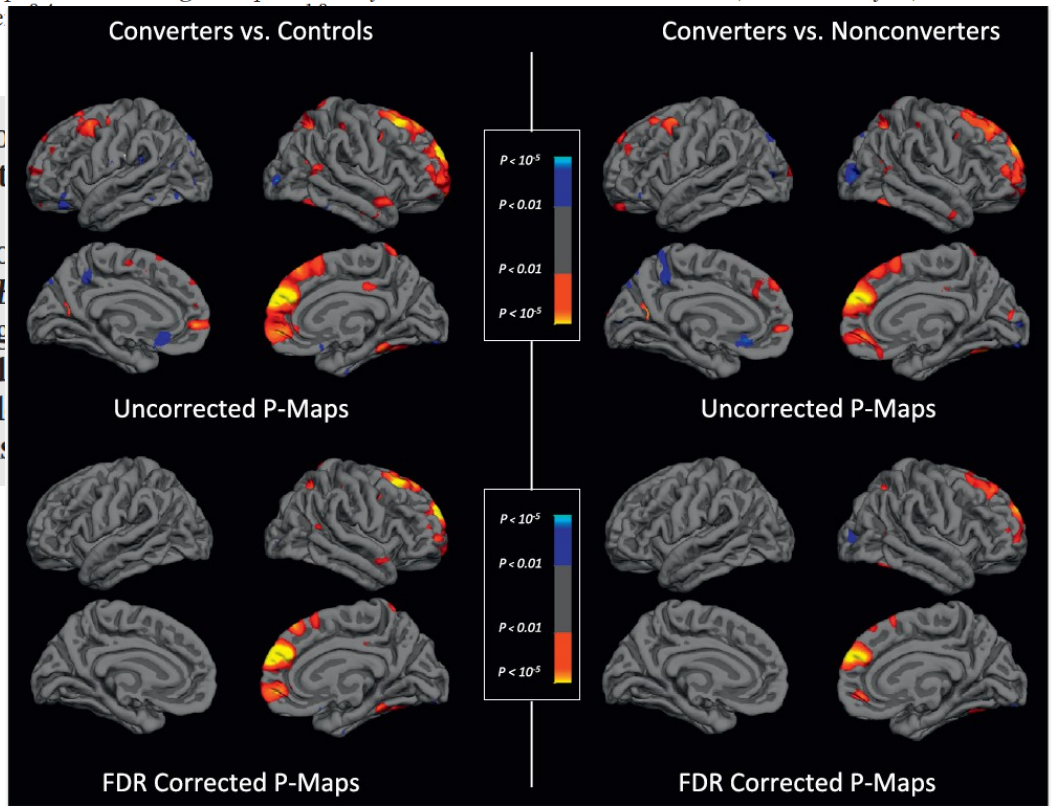
- Once a significant QTL is identified, additional genetic tests are needed to determine the exact identity of the gene
  - Association: identifies a genomic region of ~500kb (250kb to either side of the association) determined by the general extent of linkage disequilibrium
  - Linkage: detect the cumulative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

'Schizophrenia Gene' Discovery Sheds Light on Possible Cause

Schizophrenia risk from complex variation of complement component 4

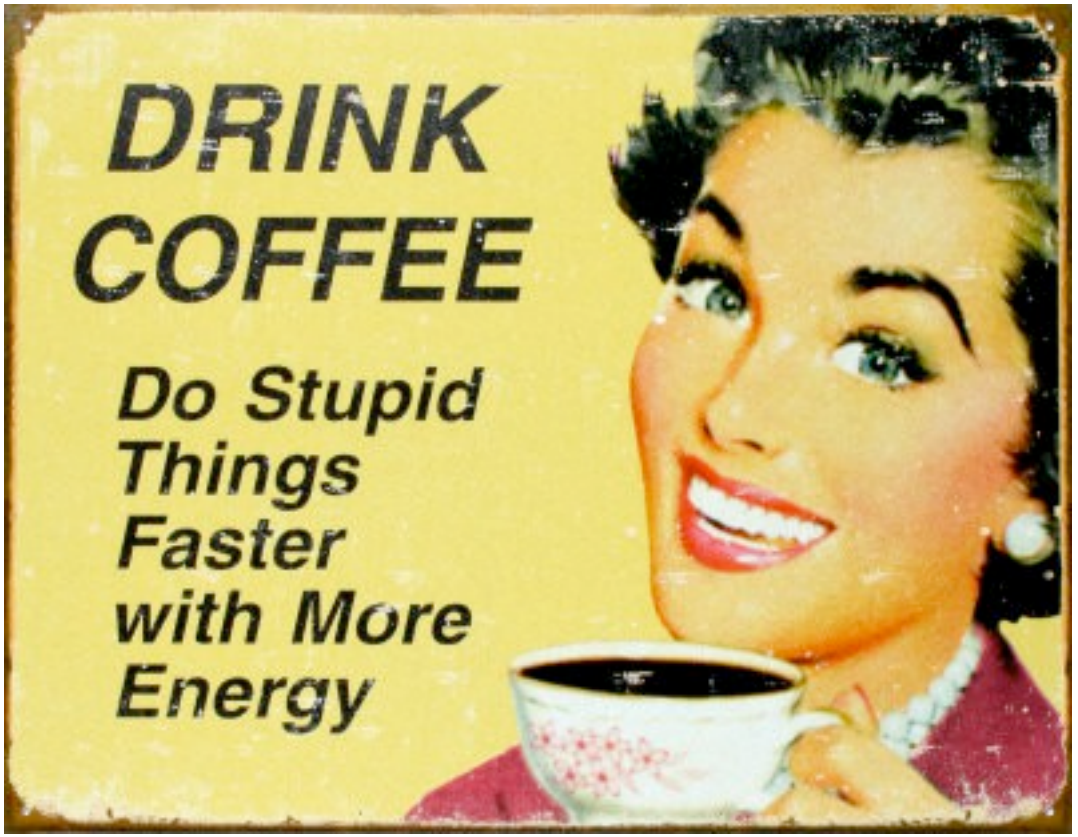
Aswin Sekar[1,2,3], Allison R. Bialas[4,5], Heather de Rivera[1,2], Avery Davis[1,2], Timothy R. Hammond[4], Nolan Kamitaki[1,2], Katherine Tooley[1,2], Jessy Presumey[5], Matthew Baum[1,2,3,4], Vanessa Van Doren[1], Giulio Genovese[1,2], Samuel A. Rose[2], Robert E. Handsaker[1,2], Schizophrenia Working Group of the Psychiatric Genomics Consortium*, Mark J. Daly[2,6], Michael C. Carroll[5], Beth Steve

Schizophrenia is a heritable brain illness with unkno association at a population level involves variation in t and molecular mechanisms accounting for this have arises in part from many structurally diverse alleles c alleles generated widely varying levels of *C4A* and *C4I* with schizophrenia in proportion to its tendency to g to neuronal synapses, dendrites, axons, and cell bod development. These results implicate excessive compl explain the reduced numbers of synapses in the brains



Cannon et al Biol Psych 2015

Coffee Break



Slide courtesy of D Glahn