Multivariate Phenotypes for association and linkage

Kochunov Peter, PhD, DABMP

Maryland Psychiatric Research Center

University of Maryland, Baltimore

And

Southwest Foundation for Biomedical Research, San Antonio

Introduction

- Types of genetic data
  - Commonly available
  - Potentially collectable
- Different genetic study design
  - Ranking based the power of genetic discovery
- Example: Genetics of cerebral aging analyses
  - Univariate analyses of imaging-based traits
  - Multivariate analyses of imaging-based traits
  - Identification of individual genes using SNP and transcript correlation analyses
  - Concordance/Discordance in findings when using different types of genetic information

Commonly collected genetic data

- Family information: twins/siblings/pedigree
  - Kinship matrix: degree of shared genetic variance
- Single-nucleotide polymorphism (SNP)
  - Single nucleotide in a polymorphic DNA region
- Quantitative trait locus (QTL) markers
  - Stretches of identifiable DNA 10-100kbp
  - Chromosomal markers
    - Selected to be proximal (linked) to genes during recombination
    - Tracking DNA inherited from each parents
    - 10-100 markers per chromosome
- Transcript data
  - Levels of transcribed mRNA measured from leukocytes: snapshot of gene expression in blood

Less commonly collected genetic data

- Deep sequencing data
  - DNA regions sequenced at fine intervals (1kbp)
  - Polymorphism of specific genes
- Copy-number variations
  - Information on deleted/repeated regions
  - Determined by hi-res kariotyping
- Methylation
  - Nature's way of regulating of gene transcriptions

My ranking of genetic studies by power of discovery

- Micro-deletion syndromes: 18q-, 7q-, etc
- Advantages:
  - Variable deletion of random size
    - Usually important genes
  - Variable symptoms
  - Variable imaging findings
  - Deep sequencing/CNV/Expression data
  - Potential treatment strategies: HGH in 18q- improves myelination
- Disadvantages
  - Rare disorders (~1 in 1000 to 10000)
  - Difficult recruitment
  - Difficult to image

Family/Twin study

- Advantages
  - Power of genetic discovery can be directly quantified
  - Familial relationship can be used to:
    - Calculate heritability
    - QTL analysis
    - Simplifies multivariate genetic analyses
- Disadvantages
  - Difficult to recruitment
    - Especially for a specific disorder
  - Need two-to-three generations for improved power

Genome-wide association studies

- Advantages
  - Can be used to study a particular disorder
  - Simplified recruitment
    - Subjects vs. controls
  - Many publically available datasets
- Disadvantages
  - Limits genetic analyses to GWAS
  - Cannot account for familial variability
  - Multiple testing makes it difficult to achieve statistical significance
  - Potential for high false positive results

Attempting genetics discovery in GOBS dataset

- Genetics of Brain Structure and Function
  - Funded by NIMH: PIs: David Glahn and John Blangero
  - A progeny of San Antonio Heart Foundation Study
  - Multi-family, three-to-four generation pedigree
- Subjects
  - 900 individuals with imaging data
  - SA area Hispanics, average family size ~ 11 individuals
  - Probands, ages 30-60 and their relatives
  - Wealth of longitudinal measurements (BP, clinical chemistry, etc)
- Genetic data
  - Kinship, QTL, 550k SNPs, expression levels, methylation data and deep sequencing data

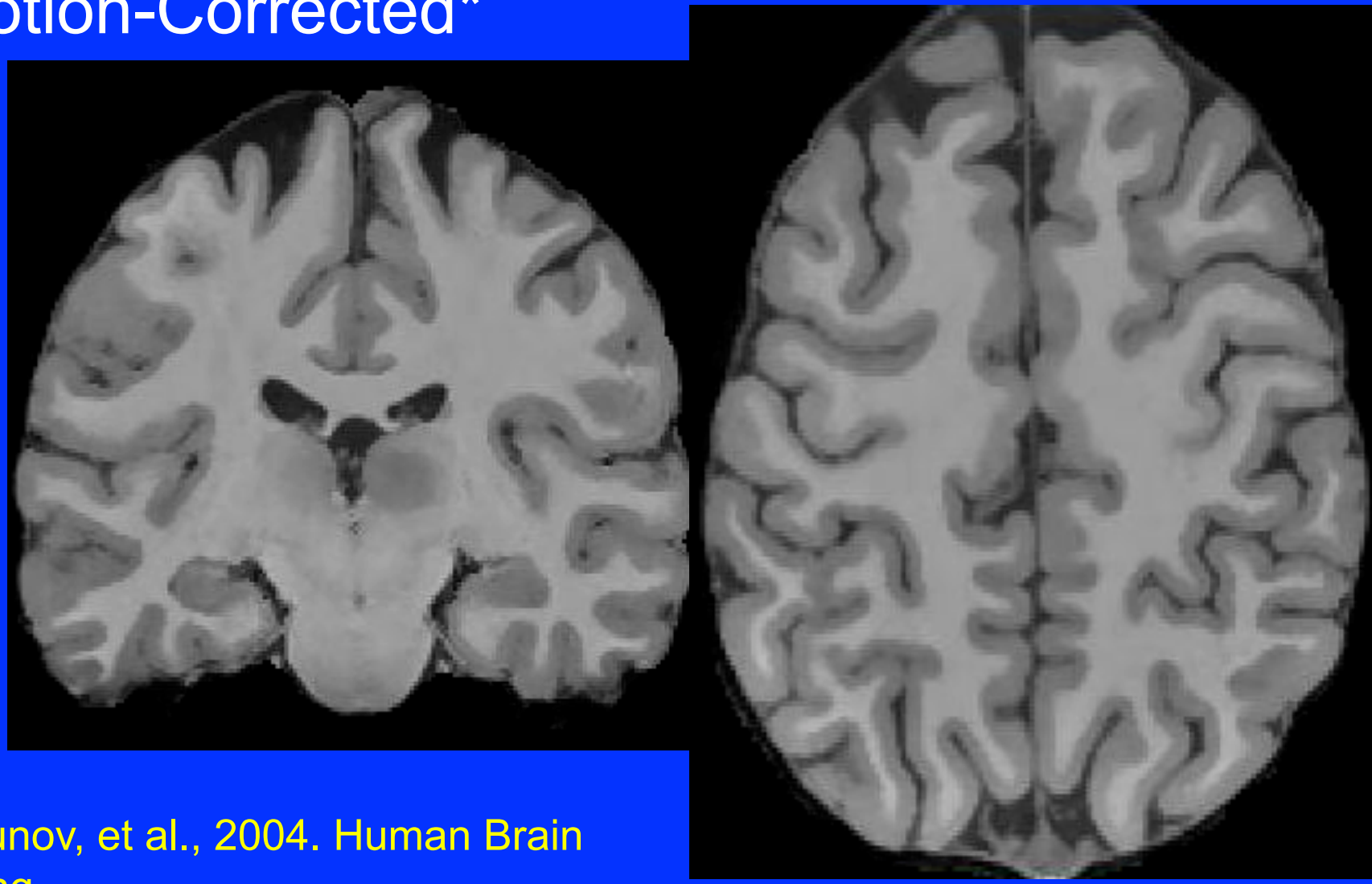
GOBS Imaging Protocol*

- An hour long imaging session
- Implemented on 3T Tim Trio Scanner
- Structural Part takes 50 min.
  - High-resolution T1w (800 μm isotropic)
  - HARDTI
  - 3D FLAIR
- rsFMRI takes 8 min
- Available at
  - http://ric.uthscsa.edu/personalpages/petr

*Kochunov (2009) Methods.

3D, T1w Structural Imaging

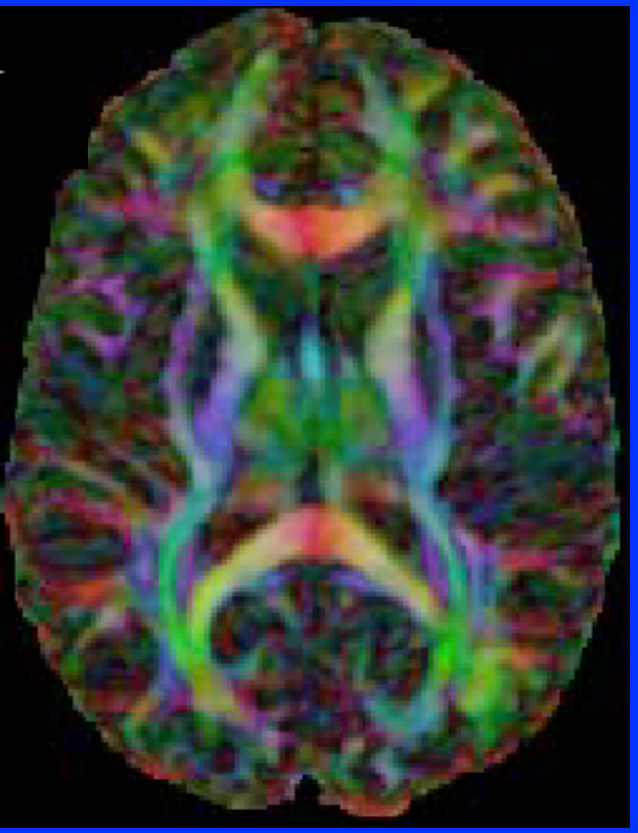
- High-contrast/resolution (25/Iso 800μm) Motion-Corrected*

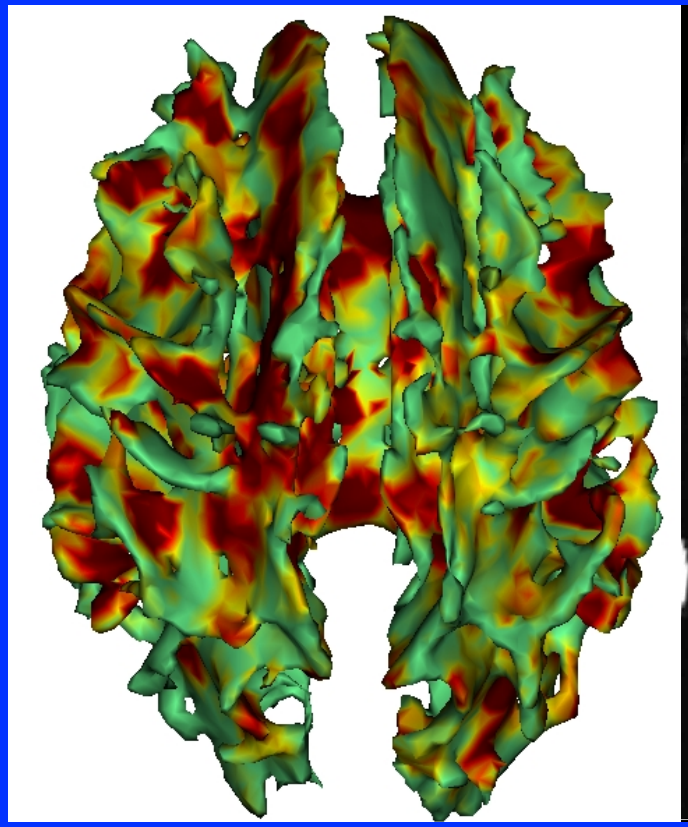


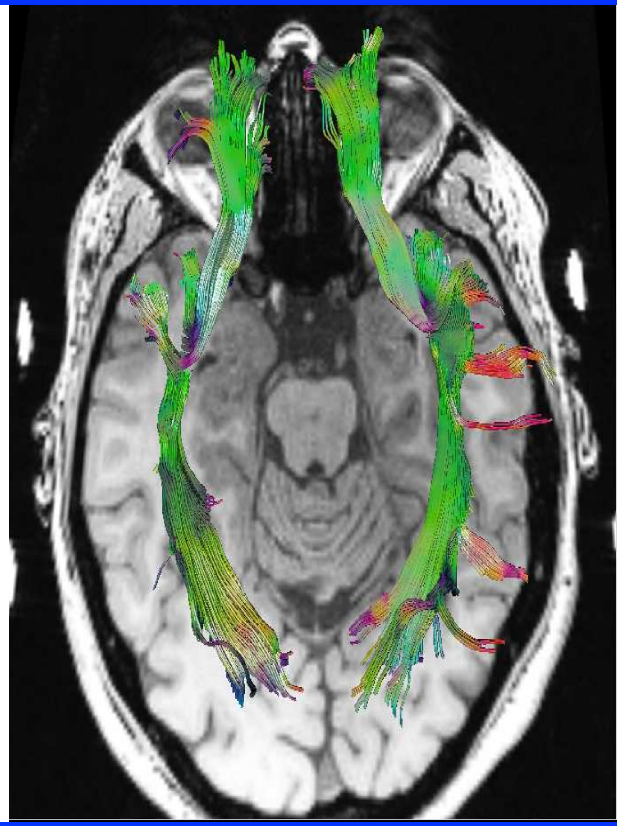
*Kochunov, et al., 2004. Human Brain Mapping

Diffusion Tensor Imaging

- MGH sequence. (1.7x1.7x3mm), 56 direction. Optimized for FA measurements (b=0, 700 s/mm^2)

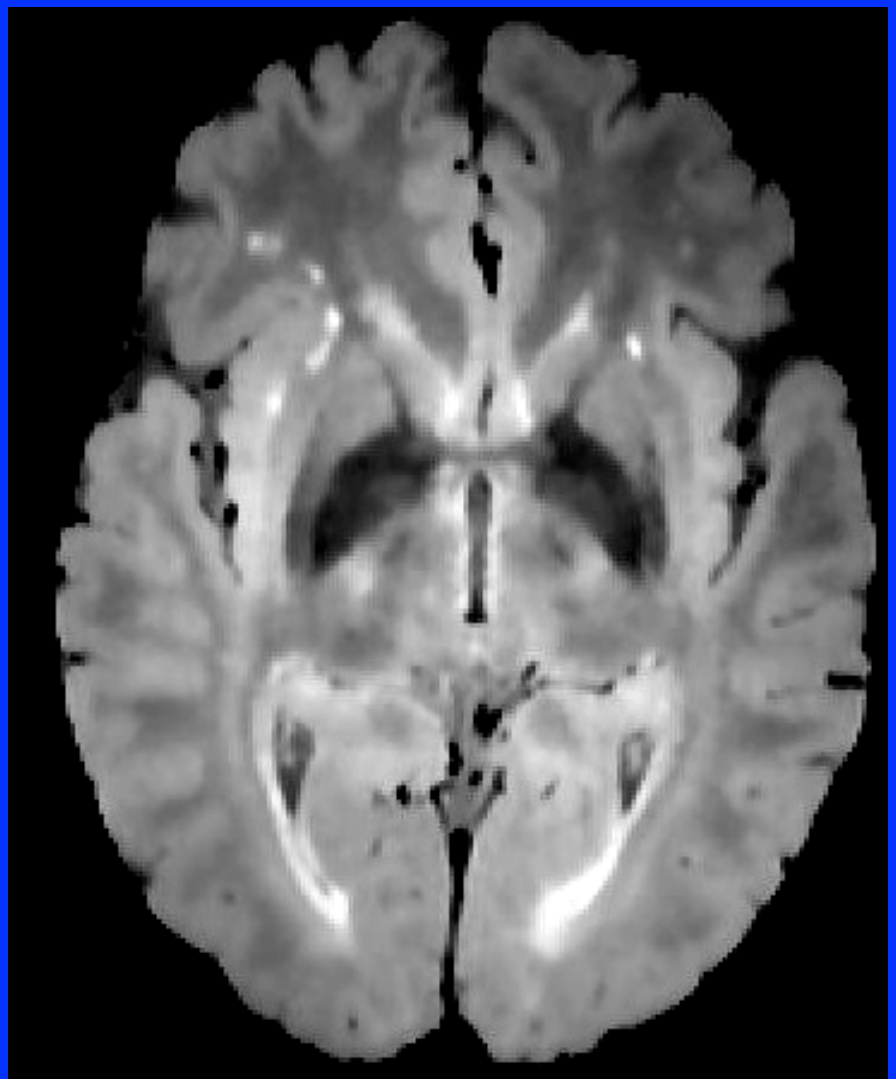


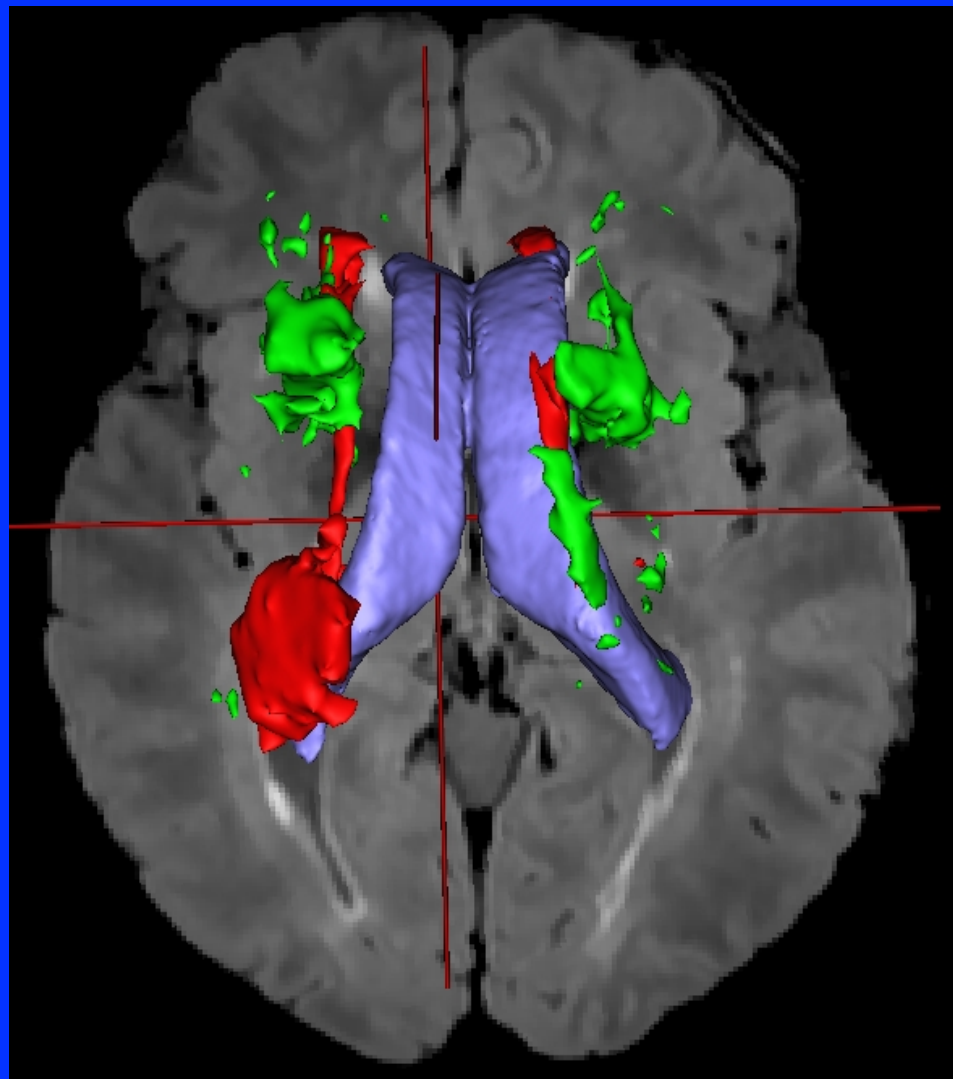




FLAIR

- 3D, Iso 1mm^3, Non-Selective IR. Optimized for lesion contrast





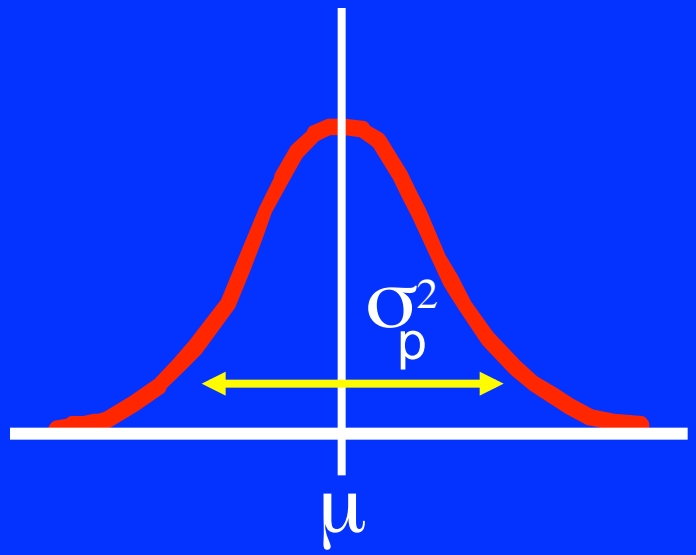
Quantification of Cerebral Decline in normal aging

- Discover genetic risks of accelerated aging
- WM health is quantified using
  - DTI measurements of water anisotropy
  - Volume, number and locations of FLAIR lesions
- GM health is quantified using
  - GM thickness
- Analysis tools used:
  - Tract-Based Spatial Statistics (DTI)
  - Manual tracing and labeling of FLAIR lesions
  - GM thickness calculations using BrainVisa

Genetics of Cerebral Aging

- Part I: Univariate Genetic Analysis
  - Measure heritability ($h^{2)}$
  - Perform QTL analysis
- Part II: Multivariate Genetic Analysis
  - Improve power of genetic discovery by
    - Use shared genetic variability from multiple traits*
  - Quantify shared genetic variability
    - Genetic correlation analysis
  - Localize DNA regions using
    - QTL
    - GWAS
    - Transcripts

Univariate Genetic Analyses: Variance Decomposition



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

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

$$\sigma^2_p = \sigma^2_g + \sigma^2_e$$

$$\sigma^2_g = \sigma^2_a + \sigma^2_d$$

σ^2_p = total phenotypic

σ^2_g = genetic

σ^2_e = environmental

σ^2_a = additive genetic

σ^2_d = dominance

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

Defining Heritability (h2)

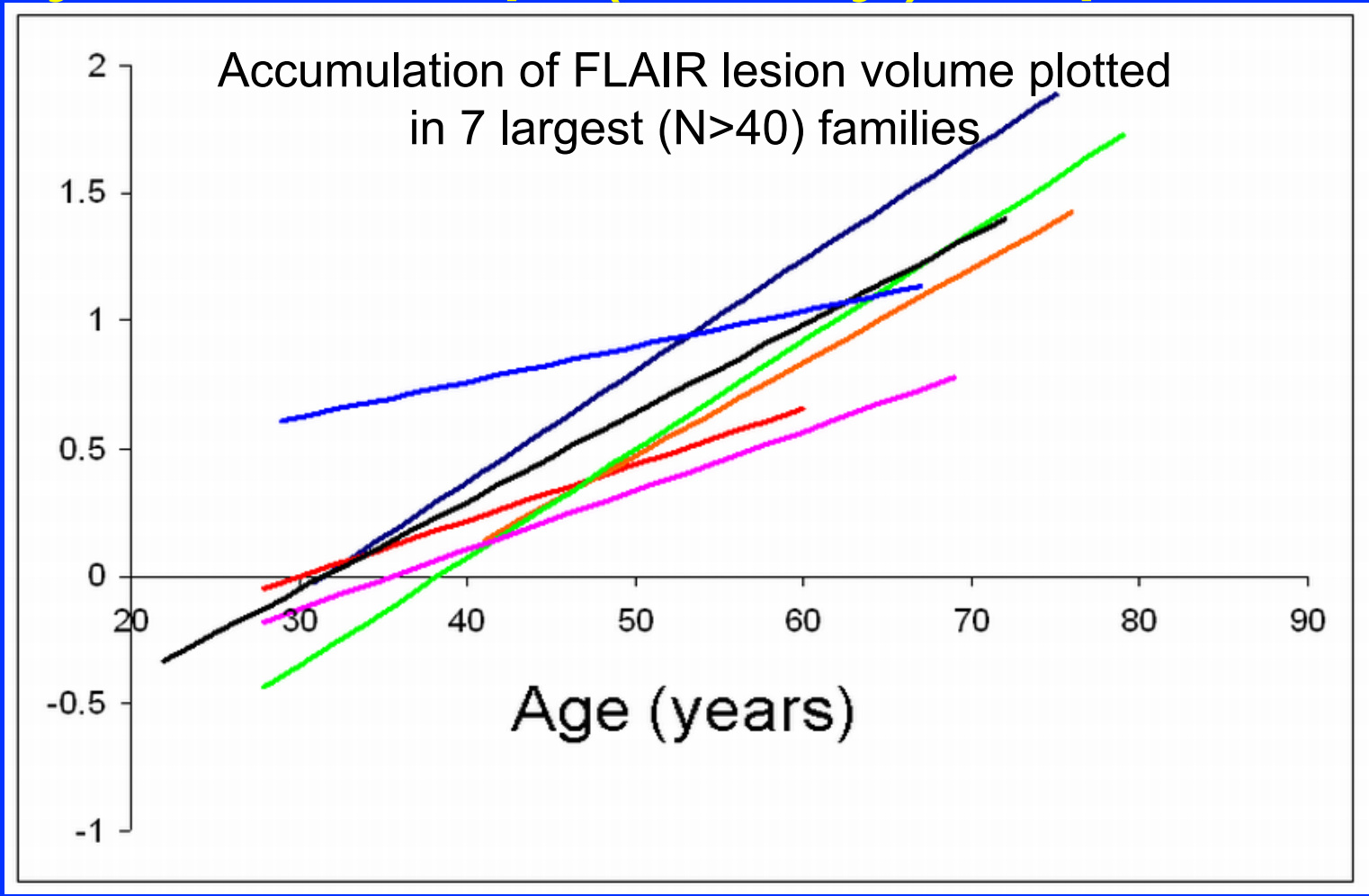
- Heritability (h2): the proportion of the phenotypic variance explained by the additive genetic effects.

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

Using kinship information to estimate heritability

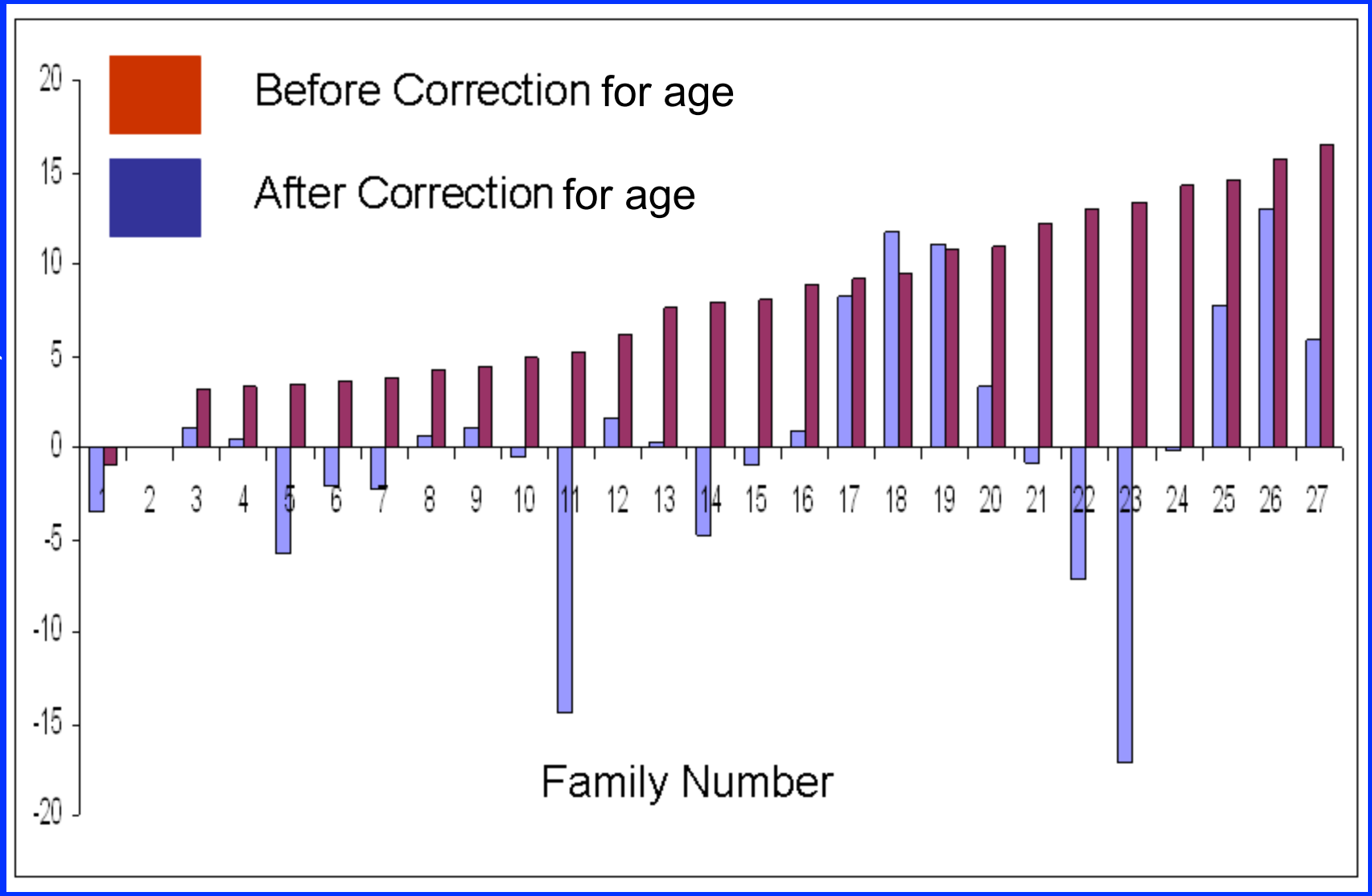
Relatives	Covariance	Heritability
Parent-child	$1/2\sigma^2_a$	$r = 1/2\ h^2$
Half siblings	$1/4\sigma^2_a$	$r = 1/4\ h^2$
Full siblings	$1/2\sigma^2_a + 1/4\sigma^2_d$	$r \geq 1/2\ h^2$
Cousins	$1/8\sigma^2_a$	$r = 1/8\ h^2$

Why is kinship (family) important?



Different families accumulate FLAIR lesions at different rates

Rate FLAIR volume increase with age



Family structure – explains a lot of variability

Results of heritability analysis

Heritability	Flair Volume*	GM thickness**	FA***
h^2	~80%	~60%	~60%

A large proportion of variability in anatomic traits is controlled by familial factors

*Kochunov et al 2009 Stroke

**Winkler et al 2010 NeuroImage

***Kochunov et al 2010 NeuroImage

Gene localization using univariate QTL

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

$$\Omega = 2\Phi\sigma_a^2 + \delta_7\sigma_d^2 + I\sigma_e^2$$

- μ Baseline mean
- β Regression coefficients
- x Scaled covariates
- a Additive genetic effects
- d Dominance genetic effects
- e Random environmental effects
- Φ Kinship matrix
- I Identity matrix

Co-Inheritance of chromosomal (QTL markers) regions and the trait

Result of QTL analyses

QTL	Flair Volume*	GM thickness**	FA***
Significance of QTL	suggestive	suggestive	suggestive

No chromosomal region was in significant control of the variability in these traits

The locations of suggestive QTL only partially replicated suggestive QTLs reported by others

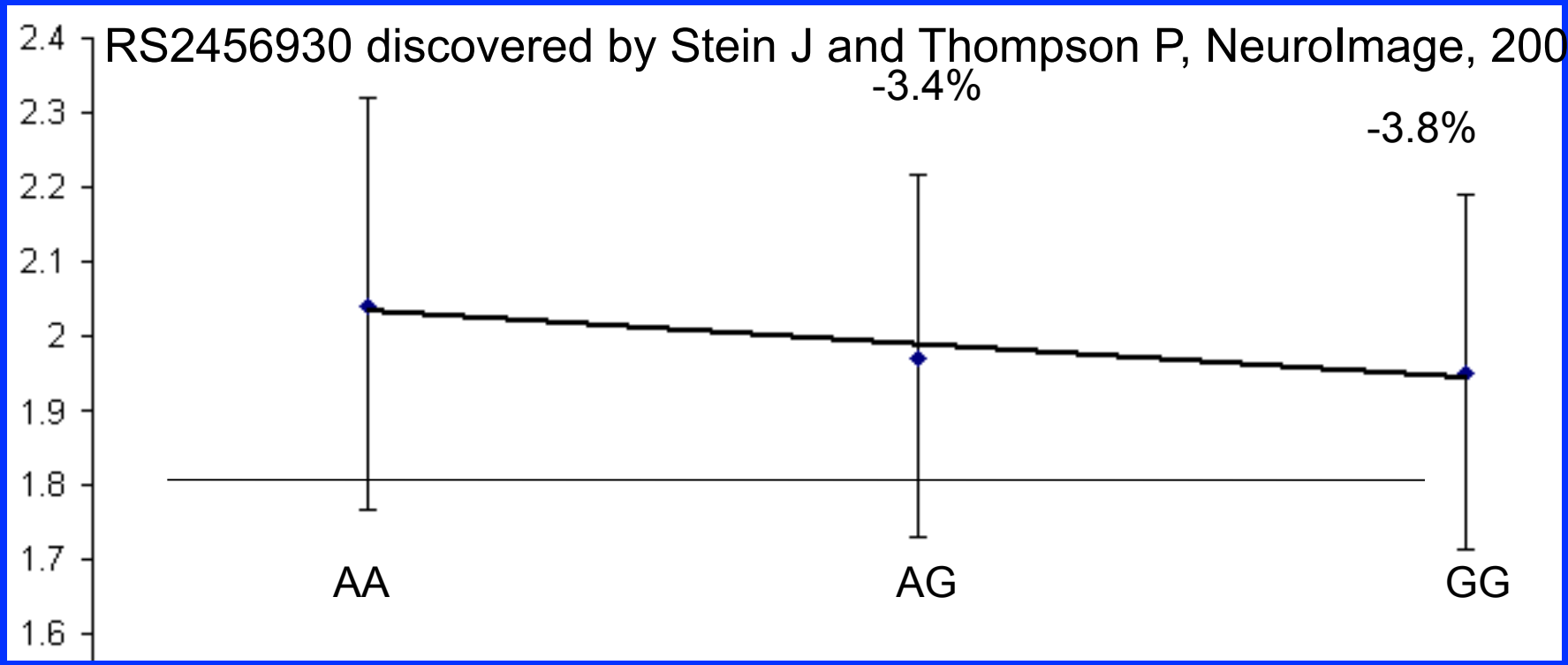
*Kochunov et al 2010 Stroke

**WIP

***Kochunov et al 2010 NeuroImage

Gene localization using GWAS

- Calculate the proportion of variability in trait that is explained by a single polymorphism



GWAS Results

- Not significant for continuous traits

Cont. Trait	SNP	p	Gene
GM thickness	RS6756736	10^{-6}	Thyroid adenoma associated
FLAIR vol.	RS3731213	10^{-7}	CDKN2A cyclin-dependent kinase inhibitor
FA	RS2803424	10^{-5}	Proprotein convertase subtilisin/kexin type 5
Binary Trait			
Stuttering	WIP	10^{-10}	Intergenic

To summarize univariate analyses

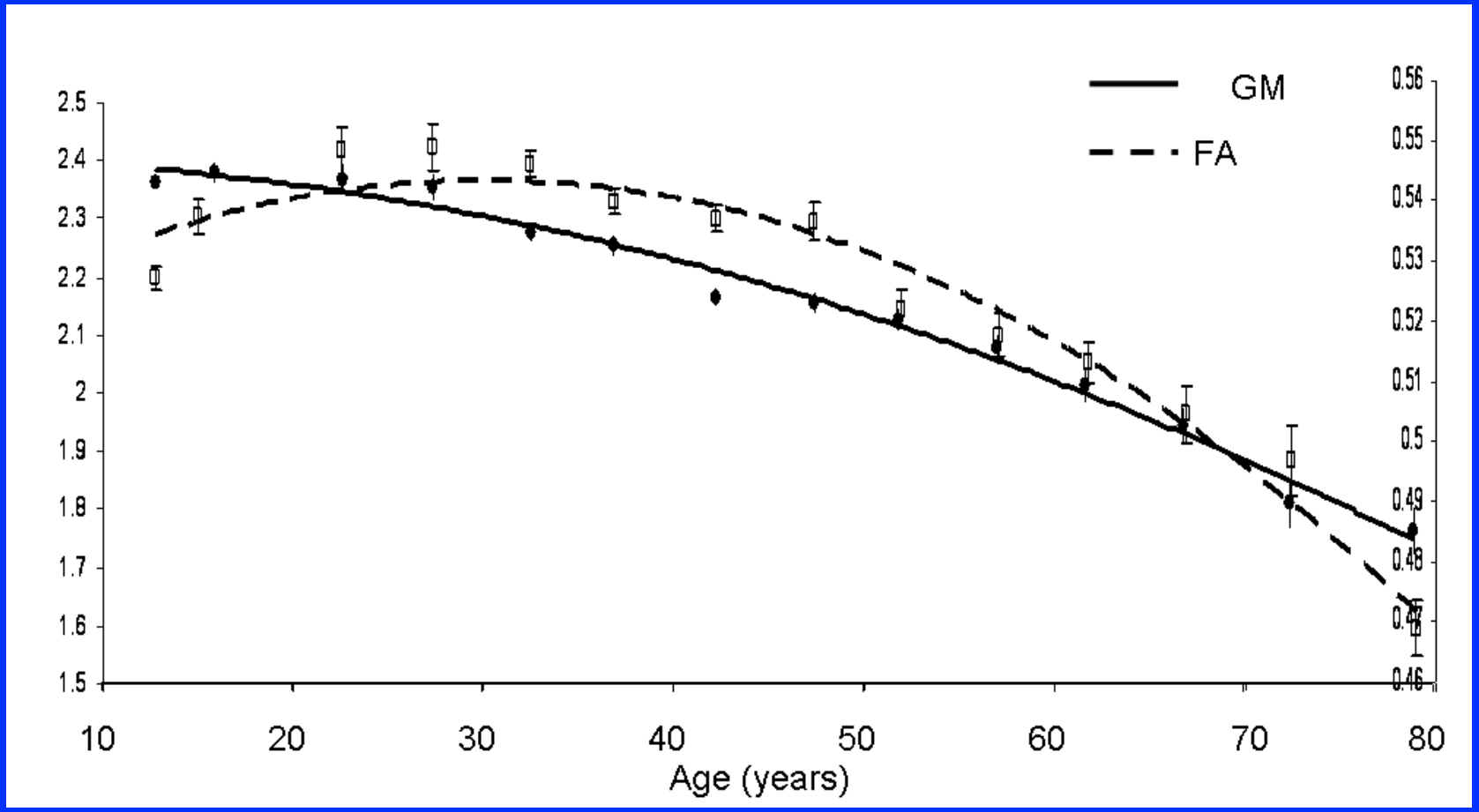
- The univariate genetic analysis
  - Demonstrated high fraction of variability is explained by additive genetic factors
  - Underpowered to identify genes for complex traits
    - Complex traits are controlled by pleotropically acting genes
    - Contribution from individual genes is difficult to separate
- How to improve the power of genetic discovery?
- What if two traits are correlated?
  - Pleiotropy?
  - Multivariate genetic analysis improve the statistical power by 2-100 times depending on degree of shared genetic variance

* Amos et al., 2001 Human Heredity

Multivariate genetic analysis

- Quantify the sources of genetic variability that are shared among multiple traits
  - Gets us closer to pleotropically acting genes
  - Helps to reduce the source of enviromental variability
  - Reduces gene x age and gene x environment interaction confound
    - If age trajectory is influenced by gene
- Two types of analyses
  - Genetic correlation
    - Calculated the fraction of shared genetic variance
  - Multivariate QTL/GWAS
    - Localize DNA regions in control of variability

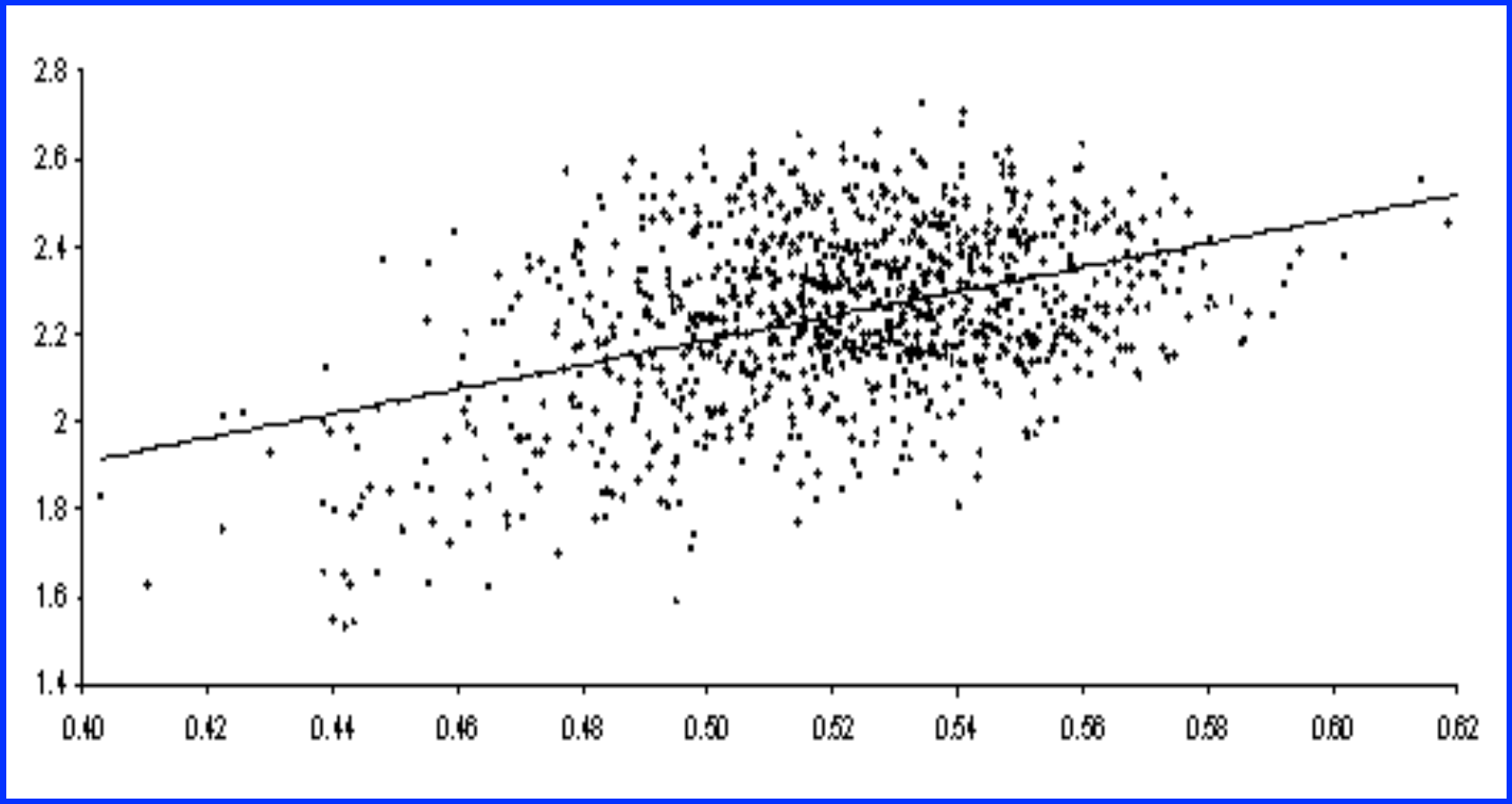
Genetics of FA of WM and Thickness of Cortical GM



Both traits exhibit inverse U-trajectory with age

Kochunov, et al., 2011, NeuroImage

Linear Relationship between them



Putatively suggests a common biological mechanism

Kochunov, et al., 2011, NeuroImage

Genetic correlation analysis

- Calculation of the shared genetic variance
  - Correlation analysis between genetic portions of variability
- Use genetic correlation (ρ_G)

$$r = \sqrt{h_A^2}\sqrt{h_B^2} \cdot \rho_G + \sqrt{1 - h_A^2}\sqrt{1 - h_B^2} \cdot \rho_E$$

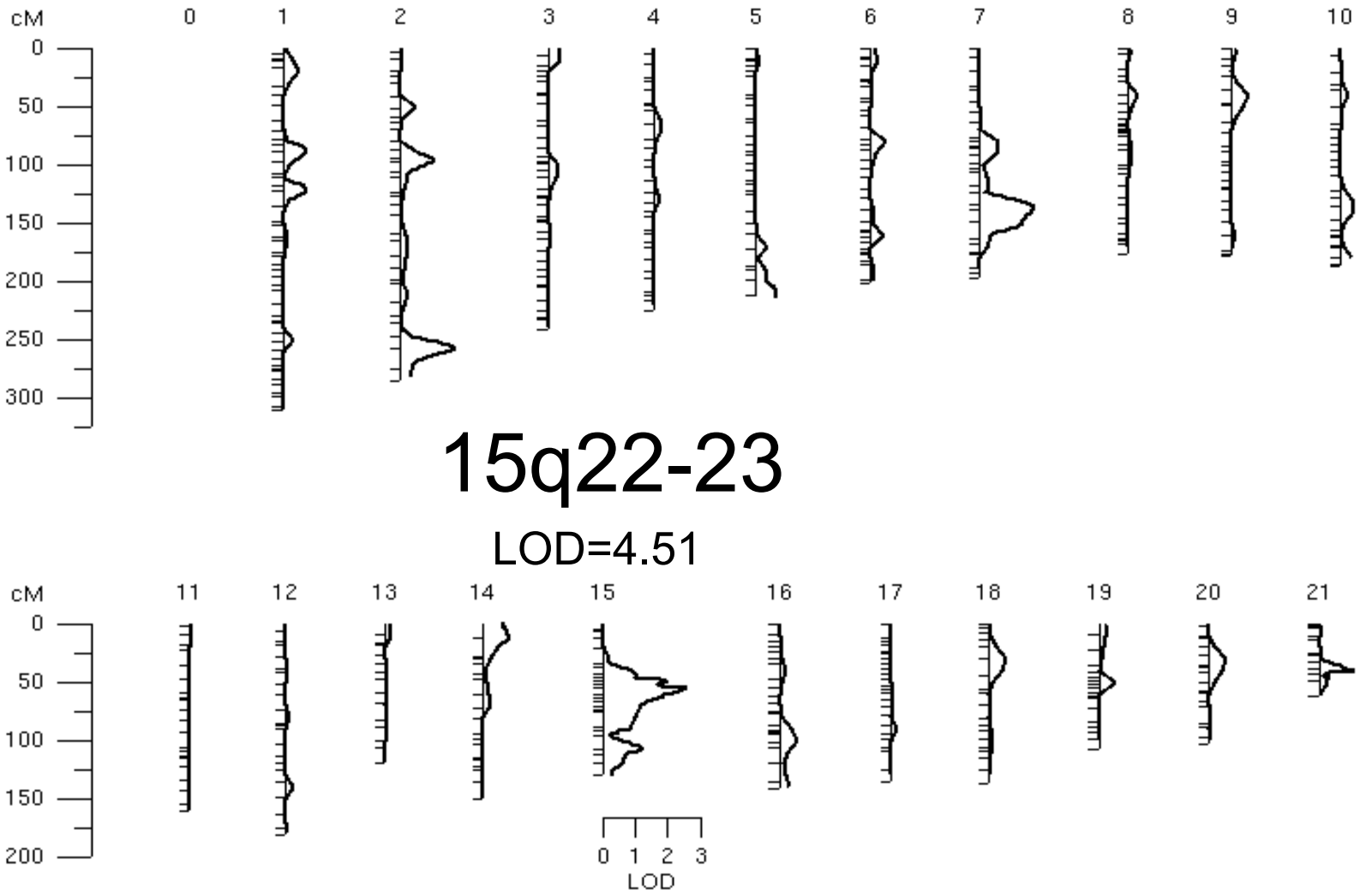
  - Pearson's r decomposed into ρ_G and ρ_E
  - ρ_G is the proportion of variability due to shared genetic effects
- Calculate degree of shared genetic variance
  - Regional GM thickness values (14 gyral regions)
  - Regional FA values (11 WM tracts)
    - 126 Trait pairs in total

Results of bivariate analysis

- Whole-brain average FA and GM
  - ρ_p=0.27; $p<10^{-7}$ and ρ_G=0.31; p=0.001
  - Suggestive QTL (2.54) at 15q22-23
- Regional analysis for 14 GM areas and 11 WM tracts (126 trait-pairs)
  - 101 showed significant ρ_G ($p<0.05$)
  - 51 suggestive QTL (LOD>2.0)
  - 13 significant QTL (LOD>3.0)

15q22-23

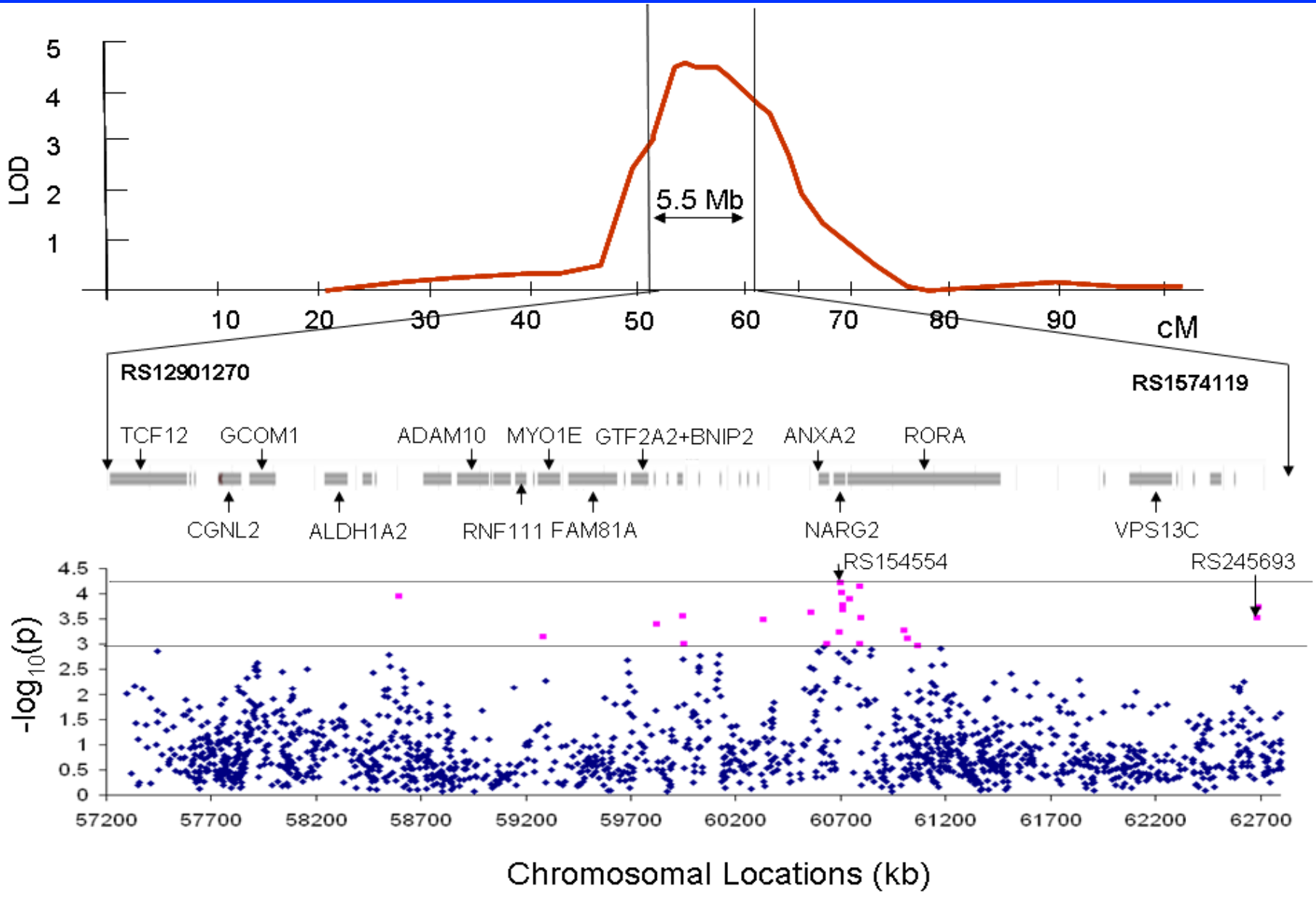
Highest LOD: FA of Body of CC and GM of Superior Parietal Lobule



Gene localization

- Use 13 pair traits that showed significant LOD to identify individual genes
- Use
  - Bivariate SNP association analysis
    - 1565 SNPs
    - Effective number = 985;
    - Criterion of significance = 5×10^{-5}
  - Transcript association analysis
    - 22 expression measurements for 20 genes

5.5Mbpairs and 20 genes



Results: SNP association

- No significant association
  - Highest for RS154554 (6×10^{-5}) – Intergenic
  - RS2456930 ($p=0.0001$)
    - Identified by Stein et al. 2010 as significant WGAS for medial temporal lobe volume
    - Located in the intergenic region
- Clustering analysis for 22 suggestively ($p<10^{-3}$) associated SNP
  - 40% (9) localized to NARG2 and RORA genes
  - The rest were intergenic

Kochunov et. al., In Submission

Results: Transcription correlations

Correlated transcript values with FA and GM thickness

- Transcript data available for 60% subjects
- Data collected 17 years ago
- RORA and ADAM10 gene transcripts
  - Significantly ($p \sim 0.01$) correlated with both FA and GM
- NARG2 transcripts
  - Significantly correlated with GM ($p=0.01$)
  - Suggestively correlated with FA ($p=0.09$)
- NARG2 and ADAM10 are significantly correlated
  - $R=0.44$; $p=1\times10^{-5}$
- NARG2 and RORA are uncorrelated
- No other significant correlation were observed

Three genes: RORA, NARG2 and ADAM10 emerged

- RORA: SNP and Transcripts
  - Discovered in *staggerer* mice mutation
    - Ataxia and neurodegeneration
  - Encodes an activator of transcription and a receptor for glucocorticoids
    - Neuroprotective and anti-inflammatory action
  - Loss of function
    - Activation of apoptosis
    - Activation of reactive atrocities
  - Candidate gene for
    - ADHD
    - Depression

NARG2

Identified by SNP association and transcript analysis

- Small Gene (59kb)
- Proximal to RORA
- Codes the *N*-methyl-d-aspartate (NMDA) receptor
- Highly expressed in developing brain
- Expressions are correlated with expressions of ADAM10

ADAM10

Identified by Transcript correlation analysis

- Its expressions were positively correlated with GM and FA
- Codes an α-secretase
  - anti-amyloidogenic proteolysis enzyme
  - Lyses amyloid precursor protein
- Blockage of ADAM10 expressions
  - rapid increase in the concentration of Aβ plaque and increased brain atrophy
- Two of its polymorphisms are associated with increased risks of late-onset AD (Kim et al., 2009)
- Activation of its transcription levels are considered a potential treatment strategy

Conclusions

- Univariate genetic analysis are underpowered
  - Imaging traits are physiologically removed from individual genes
  - Dependent on gene x age and gene x phenotype interactions
- Multivariate genetic analysis increases the power of genetic discovery
  - Requires related individuals
- Multimodal genetic information is necessary to understand findings
  - Different genetic information can lead to divergent findings
  - Eliminate the false positives

Words of wisdom*

- Symptoms in most genetic disorders are not caused by inherited mutations in one of the genes
- Instead, the majority of conditions are due to copy number changes.
  - E.g. deletions or duplications of otherwise normal genes.
- Extra and missing copies of the genes lead to changes in the RNA amount
- Thus altering a critical stoichiometry and leading to changes in critical biological processes.
- Only 5% of genes cause a problem when present in an abnormal copy number.
- Determining the function of every single gene on the chromosome may be interesting, but unnecessary
- Instead: We need to find “dosage sensitive” gene, e.g. these that lead to medical or developmental problems when deleted or duplicated.

Janine Cody, Head of the 18q- project

Acknowledgment

- John Blangero and David Glahn
- Thomas Nichols
- NIH
  - K01 EB006395
    - to P.K.,
  - RO1s MH078111, MH0708143 and MH083824
    - to J.B. and D.G..