

Statistical Methods for Neuroimaging Data Analysis

ICSA Canada Chapter 2015 Lecture 4

Imaging Genetics

Linglong Kong
University of Alberta

ICSA Canada Chapter August 4, 2015 @ Calgary, AB

Special Thanks to Professor Hongtu Zhu at UNC



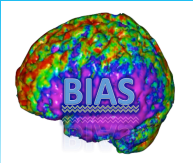
Reading materials (incomplete list)

1. H. Ganjgahi, A.M. Winkler, D.C. Glahn, J. Blangero, P. Kochunov, T.E. Nichol (2015) Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage*, in press.
2. P. Hibar, J. L. Stein, M. Renteria, A. Arias-Vasquez, S. Desrivieres, N. Jahanshad, et al., "Common genetic variants influence human subcortical brain structures," *Nature* (in press), 2014.
3. Y. Zhang, Z. Xu, X. Shen, W. Pan, for the ADNI. (2014). Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*. 96, 309-325.
4. J. L. Stein, S. E. Medland, A. A. Vasquez, D. P. Hibar, R. E. Senstad, A. M. Winkler, et al., "Identification of common variants associated with human hippocampal and intracranial volumes," *Nature Genetics*, vol. 44, pp. 552-561, 2012.
5. Ge, Tian, Nichols, Thomas E., Lee, Phil H., Holmes, Avram J., Roffman, Joshua L., Buckner, Randy L., Sabuncu, Mert R. and Smoller, Jordan W.. (2015) *Massively expedited genome-wide heritability analysis (MEGHA)*. Proceedings of the National Academy of Sciences of the United States of America, Volume 112 (Number 8). pp. 2479-2484.
6. D. Lin, H. Cao, V. D. Calhoun, e, Yu-Ping Wang, Y. P. Sparse models for correlative and integrative analysis of imaging and genetic data. *Journal of Neuroscience Methods* 237 (2014) 69–78.
7. H. Cao, J. Duan, D. Lin, Y. Y. Shugart, V. Calhoun, Y.-P. Wang. (2014). Sparse representation based biomarker selection for schizophrenia with integrated analysis of fMRI and SNPs. 102, 220–228.
8. Thompson, P. M., Ge, T., Glahn, D. C., Jahanshad, N., and Nichols, T. E. (2013) Genetics of the connectome. *NeuroImage*.
9. Medlan, S. E., Jahanshad, N., Neale, B. M., and Thompson, P. M. (2014). Whole-brain analyses of whole-brain data: working within an expanded search space. *Nature Neuroscience*.
10. Shen L, Thompson PM, Potkin SG, Bertram L, Farrer LA, Foroud TM, Green RC, Hu X, Huentelman MJ, Kim S, Kauwe JSK, Li Q, Liu E, Macciardi F, Moore JH, Munsie L, Nho K, Ramanan VK, Risacher SL, Stone DJ, Swaminathan S, Toga AW, Weiner MW, Saykin AJ, for the Alzheimer's Disease Neuroimaging Initiative (2014) Genetic analysis of quantitative phenotypes in AD and MCI: Imaging, cognition and biomarkers. *Brain Imaging and Behavior*, 8:183-207.
11. Vounou, M., Nichols, T. E., Montana, G., and ADNI (2010) Discovering genetic associations with high- dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *NeuroImage*, 53, 1147–1159.
12. Vounou, M., Janousova, E., Wolz, R., Stein, J., Thompson, P., Rueckert, D., Montana, G., and ADNI (2012) Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in alzheimer's disease. *Neuroimage*, 60, 700–716.
13. Ge, T., Feng, J., Hibar, D. P., Thompson, P. M., and Nichols, T. E. (2012) Increasing power for voxel- wise genome-wide association studies: The random field theory, least square kernel machines and fast permutation procedures. *NeuroImage*, 63, 858–873.
14. Wang, H., Nie, F., Huang, H., Kim, S., Nho, K., Risacher, S., Saykin, A., Shen, L., and ADNI (2012) Identifying quantitative trait loci via group-sparse multi-task regression and feature selection: An imaging genetics study of the adni cohort. *Bioinformatics*, 28, 229–237.
15. Silver M., Janousova E., Hue X., Thompson P. and Montana G. (2012) Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression. *Neuroimage*, 63(3), 1681-1694.



References

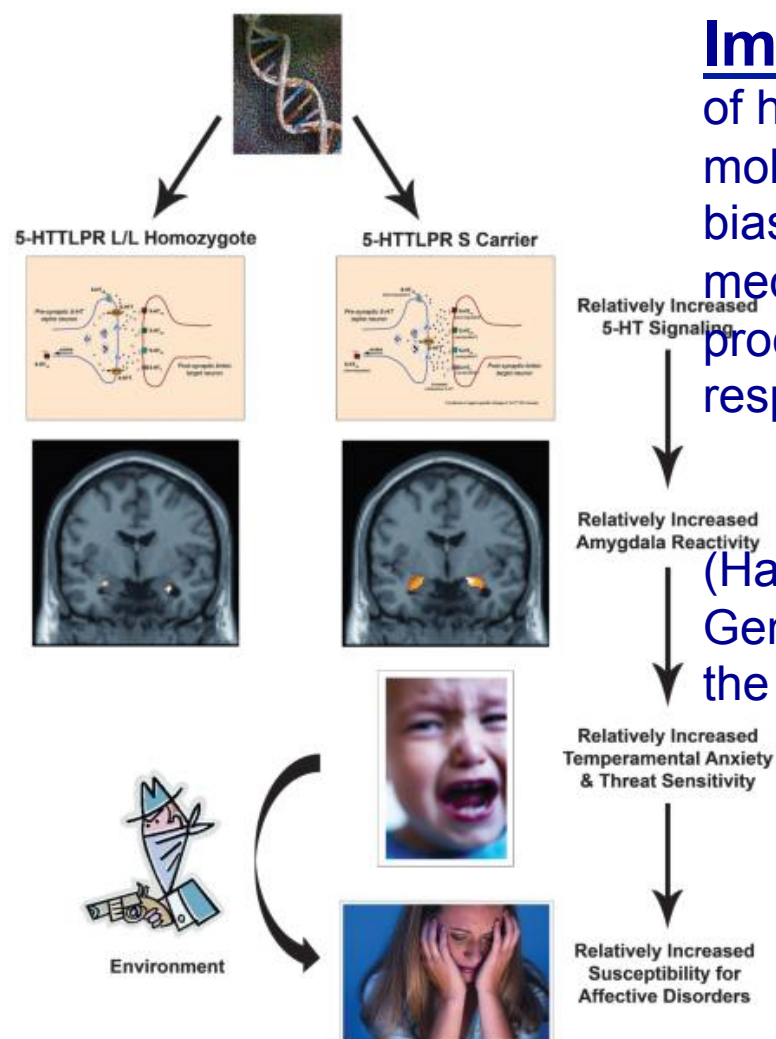
- Zhao, Y.Z., **Zhu, H.T.**, Lu, Z., Knickmeyer, R. C. and Zou, F. Bayesian Hierarchical Variable Selection for Genome-wide Association Studies. In submission, 2015.
- Zhu, H.T., Khondker, Z. S., Lu, Z.H.**, and Ibrahim, J. G. Bayesian Longitudinal Low Rank Regression Models for Longitudinal Imaging Genetic Data. *NeuroImage*, in revision, 2015.
- Huang, M., Nichols, T., Huang, C., Yu, Y., Lu, Z., Knickmeyer, R. C., Feng, Q., and **Zhu, H. T.** (2015). FVGWAS: Fast Voxelwise Genome Wide Association Analysis of Large-scale Imaging Genetic Data, *NeuroImage*, In revision.
- Sun, Q., Zhu, H.T.**, Liu, Y. F., and Ibrahim, J.G. SPReM: Sparse Projection Regression Model for High-dimensional Linear Regression. *Journal of American Statistical Association*, in press, 2014.
- S. J. Lee, R.J. Steiner; Shikai Luo; Michael C Neale; Martin Styner; **Hongtu Zhu**; John H. Gilmore. (2015). Quantitative tract-based white matter heritability in twin neonates. *NeuroImage*, 111:123-135.
- Lin, J., Zhu, H.T.**, Mihye, A., and Ibrahim, J.G. Functional Mixed Effects Models for Candidate Genetic Mapping in Imaging Genetic Studies. *Genetic Epidemiology*, 38(8):680-91, 2014.
- Zhu, H.T., Khondker, Z. S., Lu, Z.H.**, and Ibrahim, J. G. Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *Journal of American Statistical Association*, 109, 1084-1098, 2014.
- Wei Gao, Amanda Elton, **H. Zhu**, Sarael Alcauter, J. Smith, John H Gilmore, and Weili Lin. (2014). Inter-subject Variability of and Genetic Effects on the Brain's Functional Connectivity during Infancy. *Journal of Neuroscience*, 34: 11288-11296.
- Kai Xia, Yang Yu, Mihye Ahn, **H. Zhu**, Fei Zou, John Gilmore, Rebecca Christine Knickmeyer. Environmental and genetic contributors to salivary testosterone levels in infants. *Frontiers in Endocrinology*. 2014.
- Bryant, C., Giovanello, K. S., Ibrahim, J. G., Shen, D. G., Peterson, B. S., and **Zhu, H.T.** (2013) Mapping the heritability of regional brain volumes explained by all common SNPs from the ADNI study. *PLOS ONE*.
- Lin, J., Zhu, H.T.**, Knickmeyer, R., Styner, M., Gilmore, J. and Ibrahim, J.G. Projection Regression Models for Multivariate Imaging Phenotype. *Genetic Epidemiology*, 36, 631-641, 2012.



Motivation



Motivation

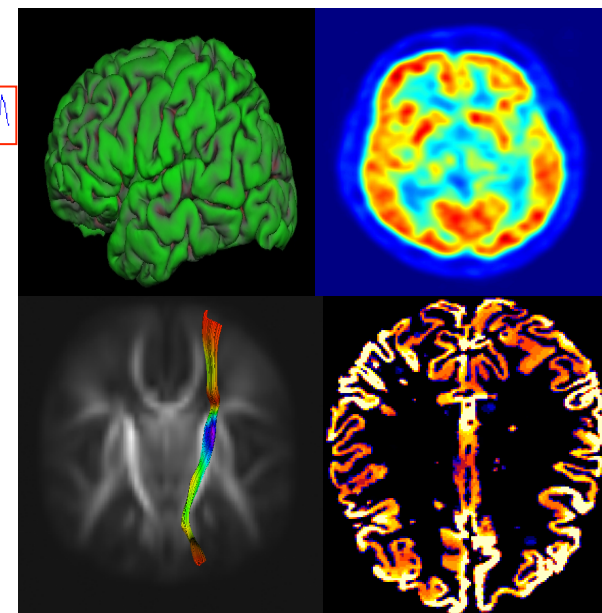
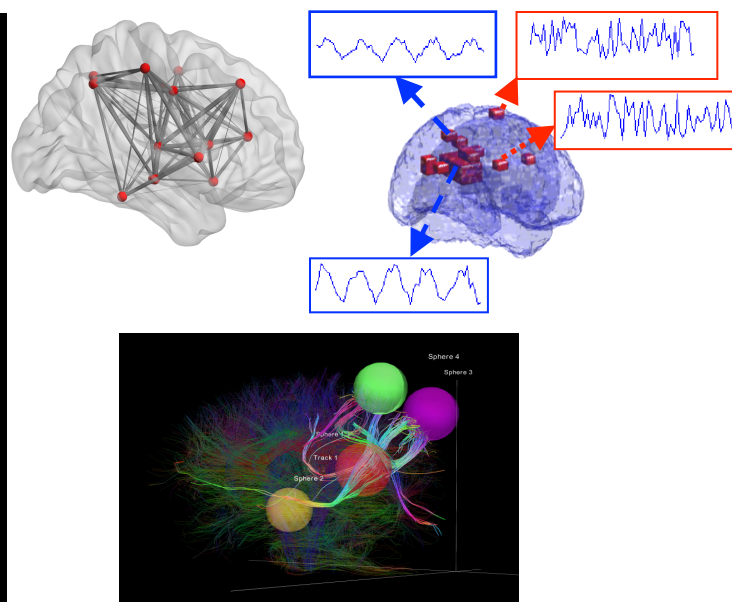
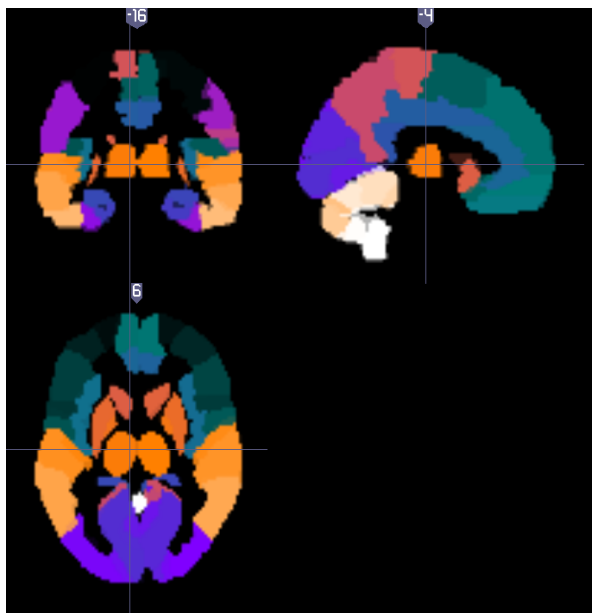


Imaging genetics allows for the identification of how common/rare genetic polymorphisms influencing molecular processes (e.g., serotonin signaling), bias neural pathways (e.g., amygdala reactivity), mediating individual differences in complex behavioral processes (e.g., trait anxiety) related to disease risk in response to environmental adversity.

(Hariri AR, Holmes A.
Genetics of emotional regulation:
the role of the serotonin transporter in neural function.
Trends Cogn Sci. [10:182–191])



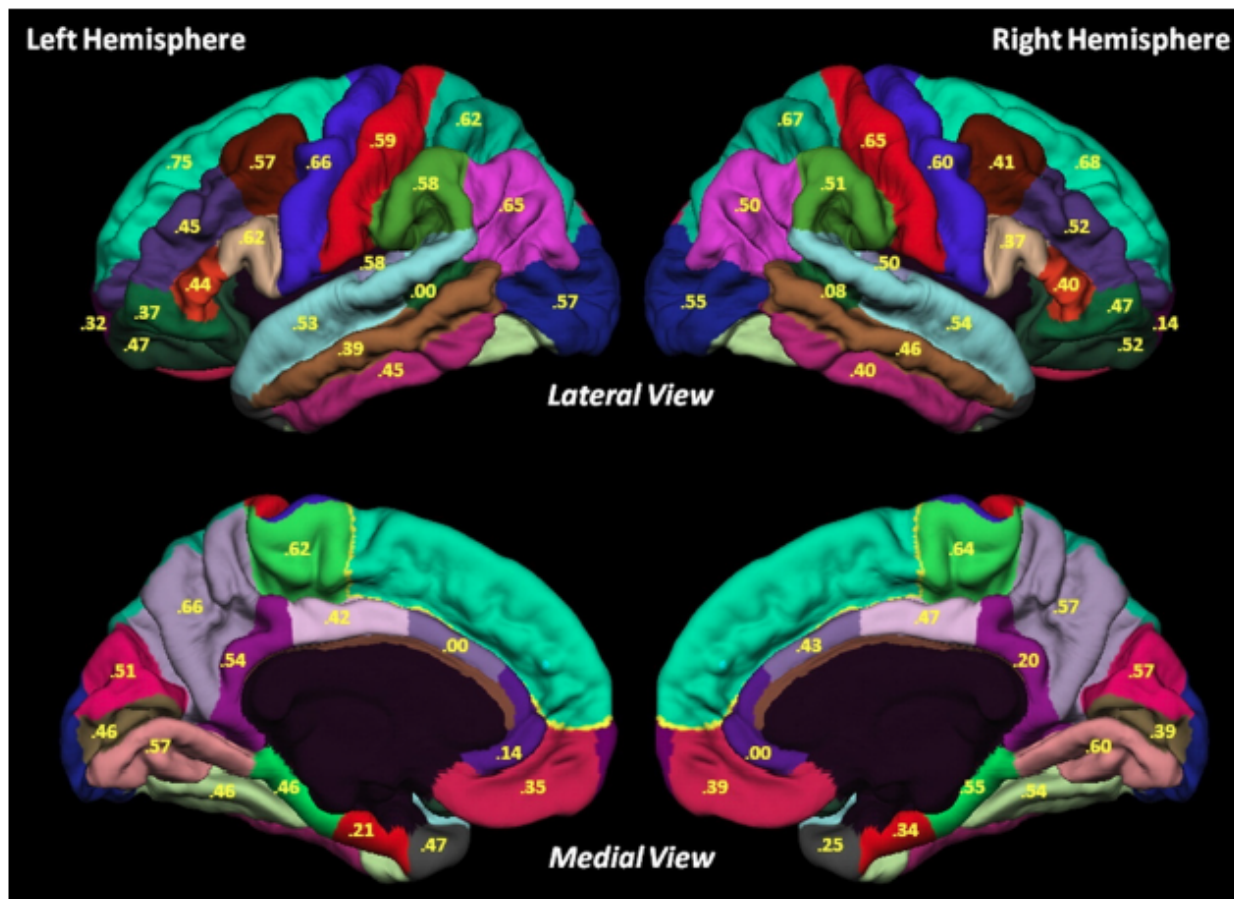
Neuroimaging Phenotype



Multivariate, smoothed functions, and piecewisely smoothed functions
Dimension varies from 100~500,000.



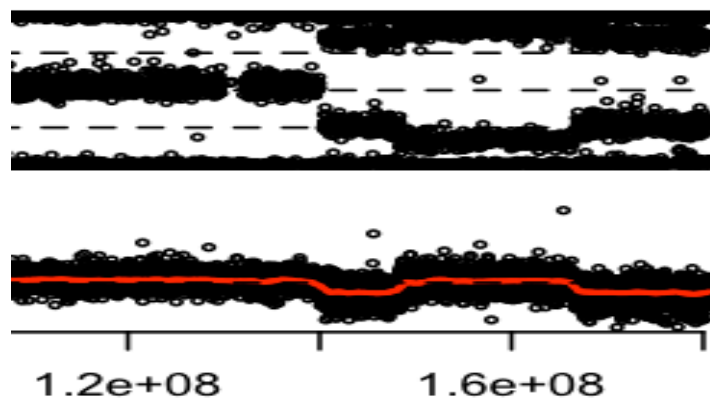
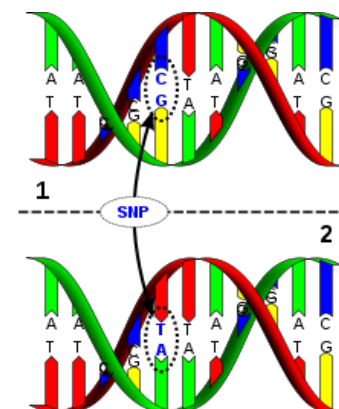
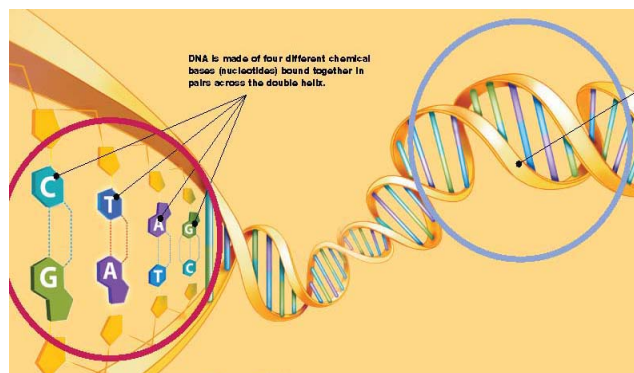
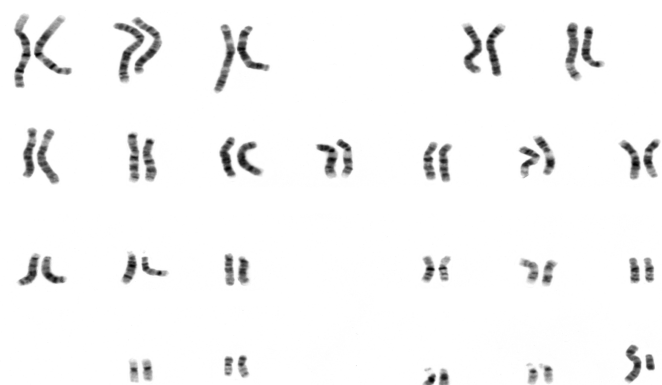
Brain Structure is highly heritable



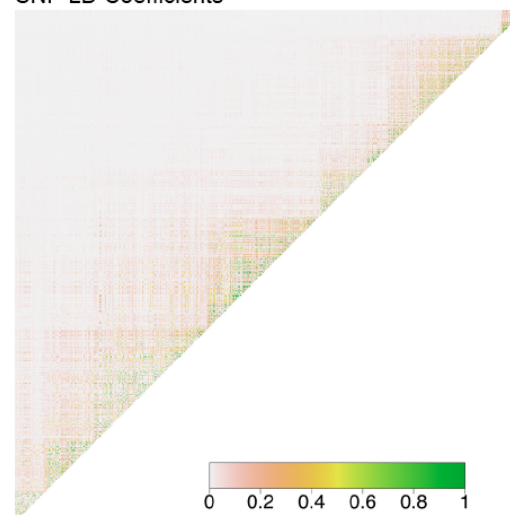
Must be specific genetic variants explaining the high heritability (Kremen et al.2010).



Genetic Data

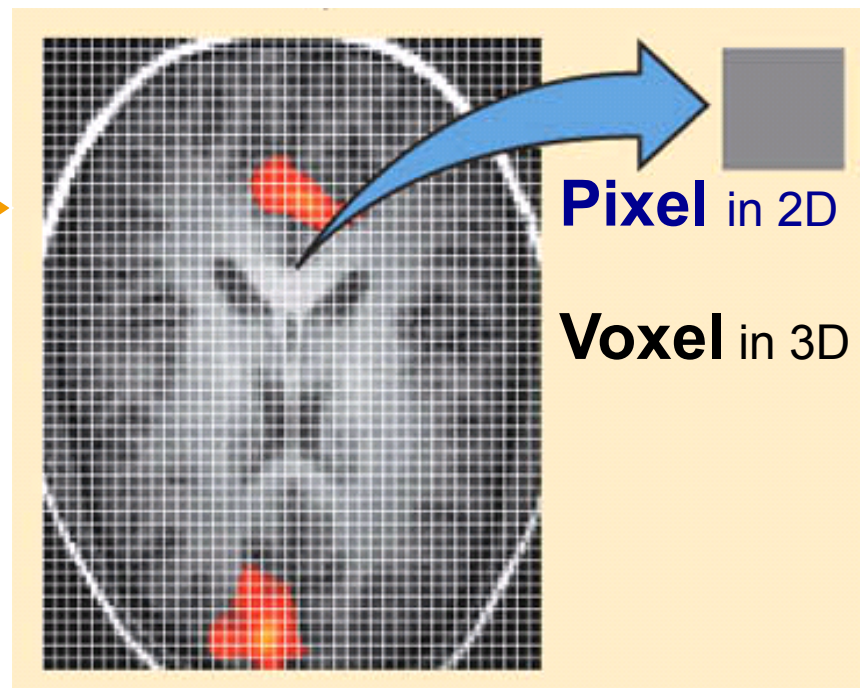
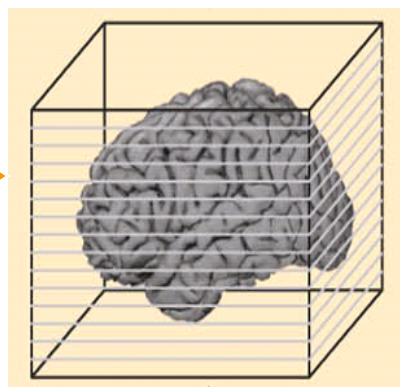


SNP LD Coefficients





Imaging Phenotype

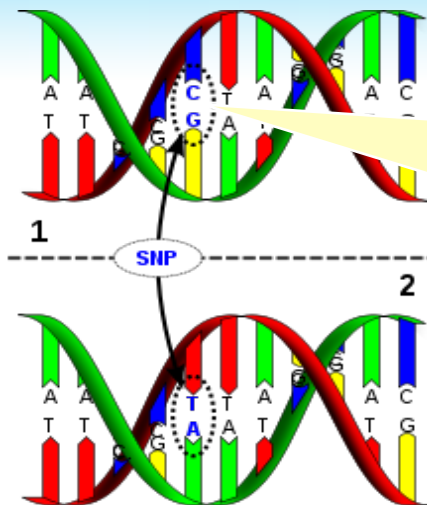
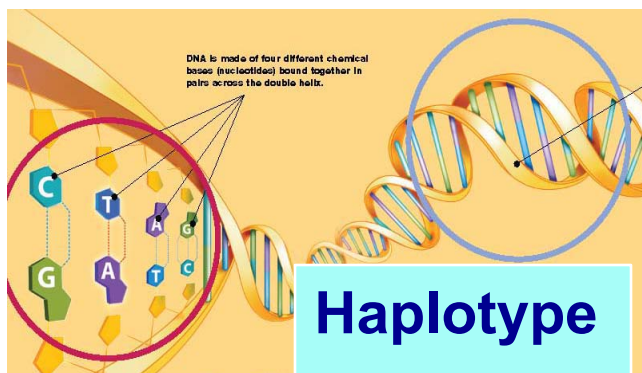
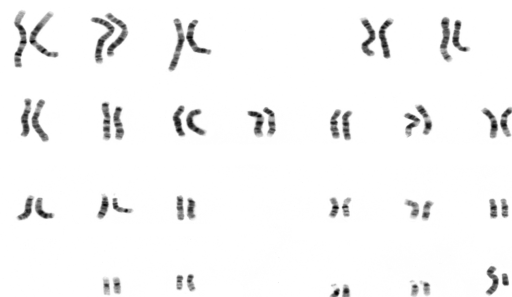


Features:

- Spatially correlated functional data
- High dimension ($\sim 2^{21} = 2,097,152$ voxels)
- Voxels with similar color form regions

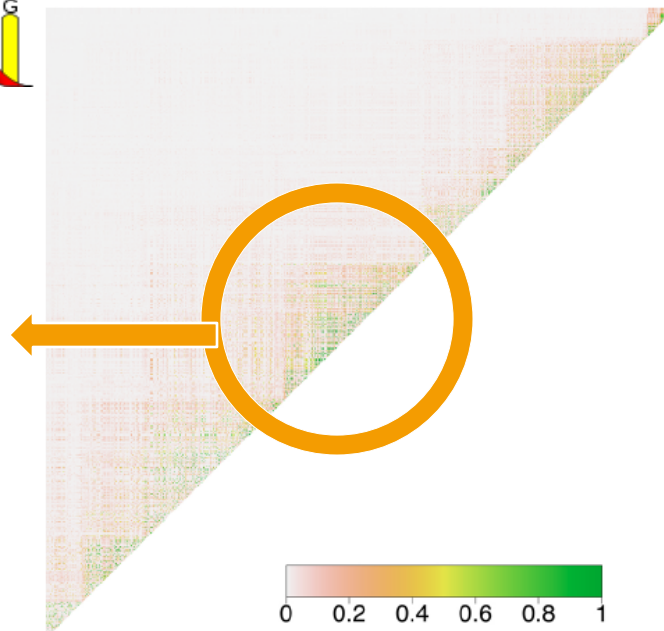


Genetic Variation Data



SNP
(Single Nucleotide Polymorphism)

SNP LD Coefficients



- **Correlated**
- **High Dimension**



Imaging Genetic Studies -- Brain Development

ALLEN BRAIN ATLAS
DATA PORTAL

Search

[Publications](#) [Contact Us](#)

[Home](#) [Mouse Brain](#) [Developing Mouse Brain](#) [Human Brain](#) [Mouse Connectivity](#) [More ▾](#)

[Welcome](#) [Announcements](#) [Highlights](#) [Tutorials](#) [Search](#) [API](#) [Help](#)

A growing collection of online public resources integrating extensive gene expression and neuroanatomical data, complete with a novel suite of search and viewing tools.

Get started with tutorials offering introductory overviews and guided tours.

Mouse Brain
A genome-wide, high-resolution atlas of gene expression throughout the adult mouse brain

Human Brain
A multi-modal, multi-resolution atlas detailing gene expression across the adult human brain

Developing Mouse Brain
A detailed atlas of gene expression across mouse brain development

Developing Human Brain
A detailed atlas of gene expression across human brain development

Mouse Connectivity
A high-resolution map of neural connections in the mouse brain

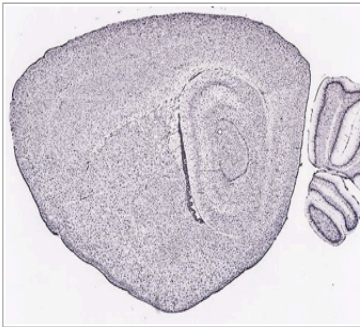
Non-Human Primate
A detailed atlas of gene expression across postnatal primate brain development

Mouse Spinal Cord
A genome-wide, high-resolution atlas of gene expression throughout the mouse spinal cord

Glioblastoma
A unique platform for exploring human glioblastoma at the cellular and molecular levels

Mouse Diversity
Gene expression data in the mouse brain across genetic backgrounds and sex

Sleep
Gene expression data in the mouse brain for five conditions of sleeping and waking

HIGHLIGHTS


Apoe is expressed in astroglia and microglia in the brain of the postnatal day 56 mouse.

[More highlights](#)

ANNOUNCEMENTS - Updated April 15, 2013
OHMB 2013
What's New - Latest Data Release March 7, 2013
[Educational and Training Resources](#)
[Connect with Us](#)

©2004 - 2013 Allen Institute for Brain Science. All Rights Reserved
[Privacy Policy](#) | [Terms of Use](#) | [Citation Policy](#) | [About the Allen Institute](#)

 **ALLEN INSTITUTE
for BRAIN SCIENCE**
Fueling Discovery

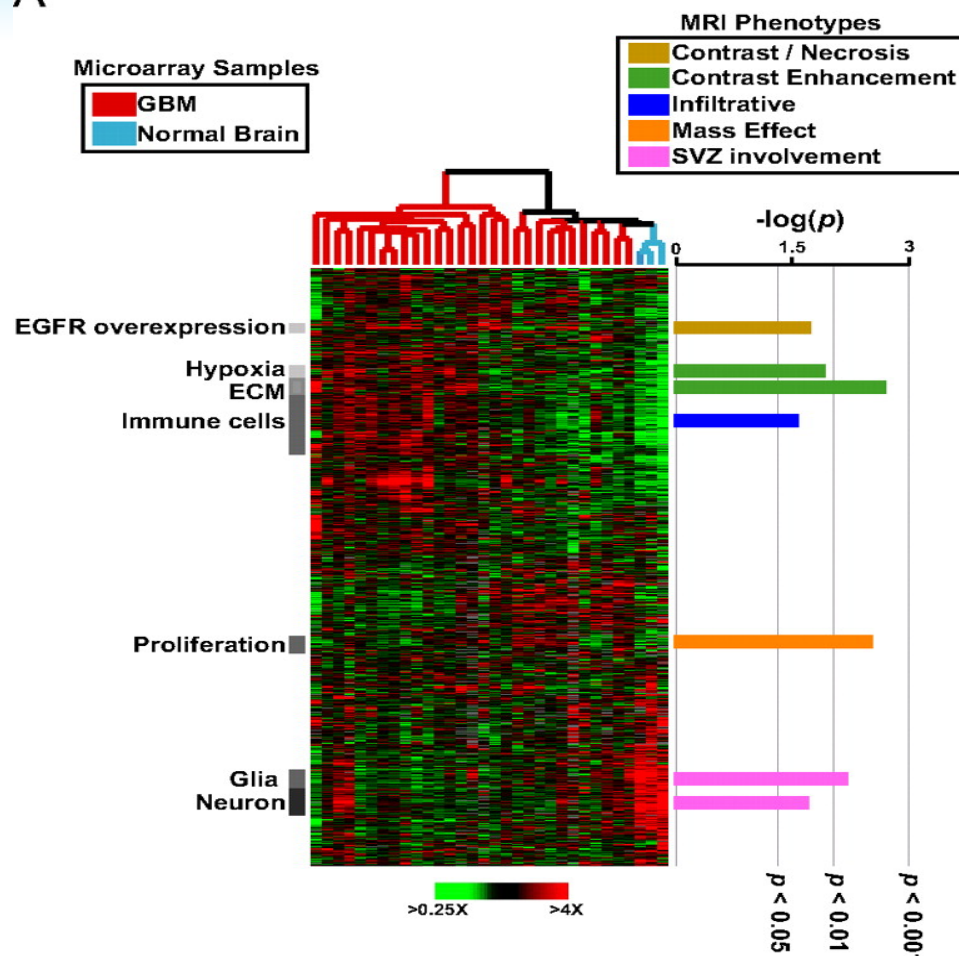
<http://www.brain-map.org/>

UNIVERSITY of ALBERTA



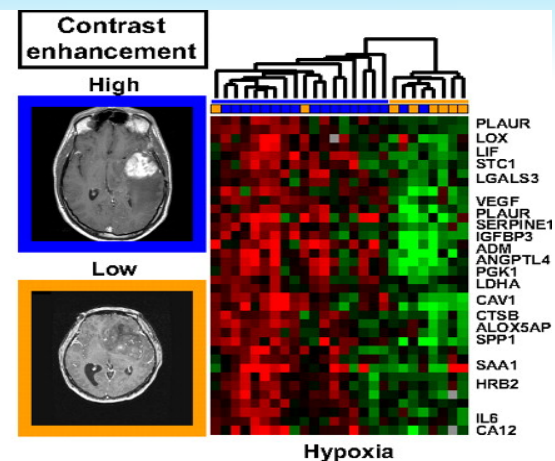
Imaging Genetic Studies -- Brain Tumor

A

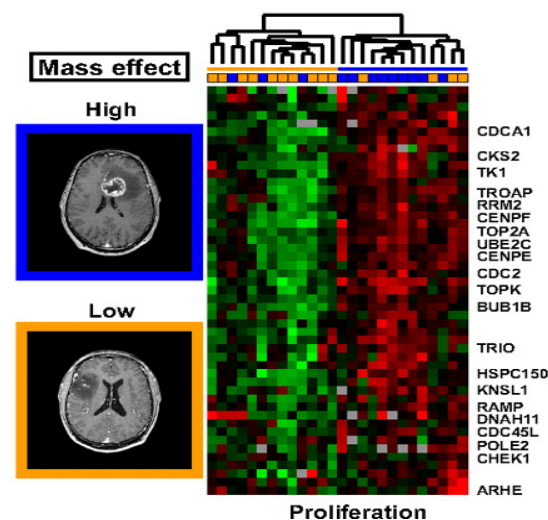


<http://www.pnas.org/content/105/13/5213/F1.expansion.html>

B

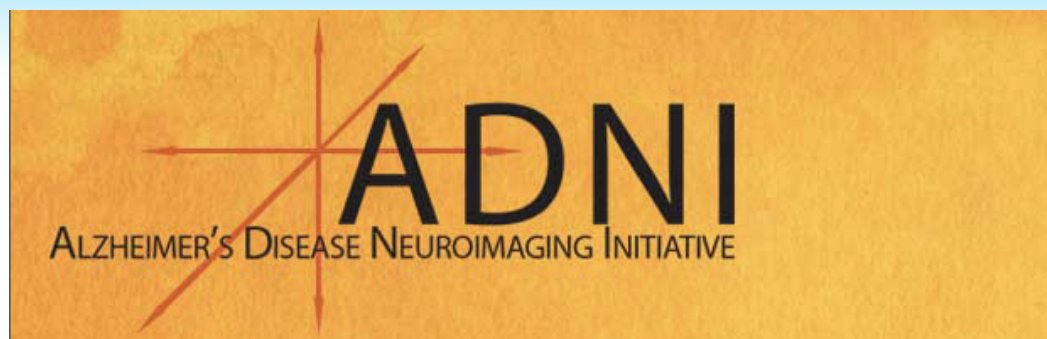


C





Imaging Genetic Studies -- More Studies



NDAR National Database for Autism Research
Serving the autism research community

09/28/2012 - First Publication Using NDAR Data Now Available
09/28/2012 - NDAR Announces the Autism Genome Project (AGP) Data Release
05/24/2012 - NDAR Publication on Sharing Heterogeneous Data

[Home](#) [Query](#) [Harmonization Tools](#) [Contribute](#) [Request Access](#) [Policy](#) [Tutorials](#) [About](#) [FAQ](#)

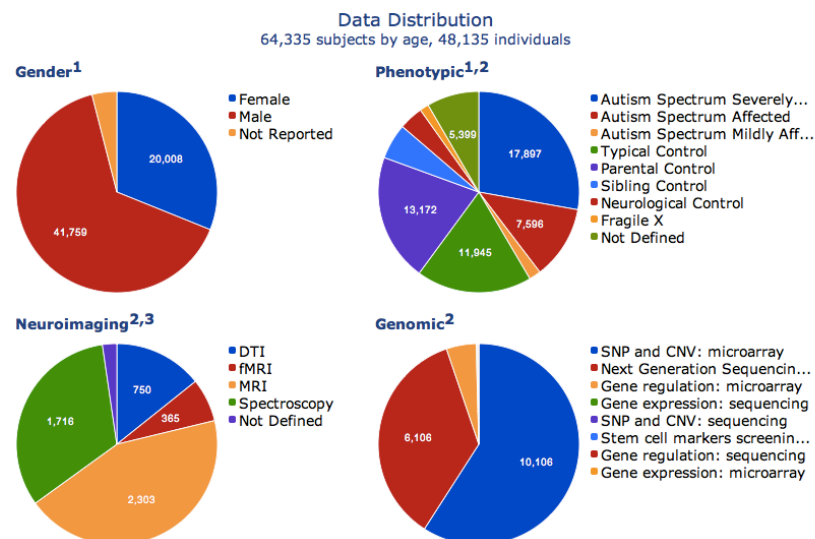
- Quick Navigation**
- NDAR Data
 - Query
 - Data from Labs
 - Data from Papers
 - Submission
 - Harmonization Standards
 - Prepare
 - Validation Tool
 - Submit
 - Data Dictionary
 - Contact Us

Learn About NDAR

National Database for ...

YouTube

The National Database for Autism Research (NDAR) is an NIH-funded research data repository that aims to accelerate progress in autism spectrum disorders (ASD) research through data sharing, data harmonization, and the reporting of research results. NDAR also serves as a scientific community platform and portal to multiple other research repositories, allowing for aggregation and secondary analysis of data.



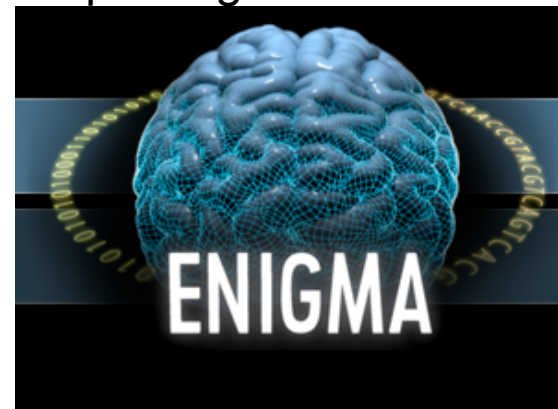


Genetic variation through collaboration

- ❑ Common genetic variation has only a **small effect** on brain structure.
- ❑ In a GWAS, many statistical tests are conducted.
- ❑ **Huge sample sizes** (>10,000 individuals) are needed in order to find and replicate association of individual genetic variants.
- ❑ As both imaging and genotyping are expensive, a **consortium is needed**.

Dr. Hibar

<http://enigma.ini.usc.edu>



> 185 institutions, 300+ co-authors,
world-wide consortium
countries resulting in **~30,000+** subjects

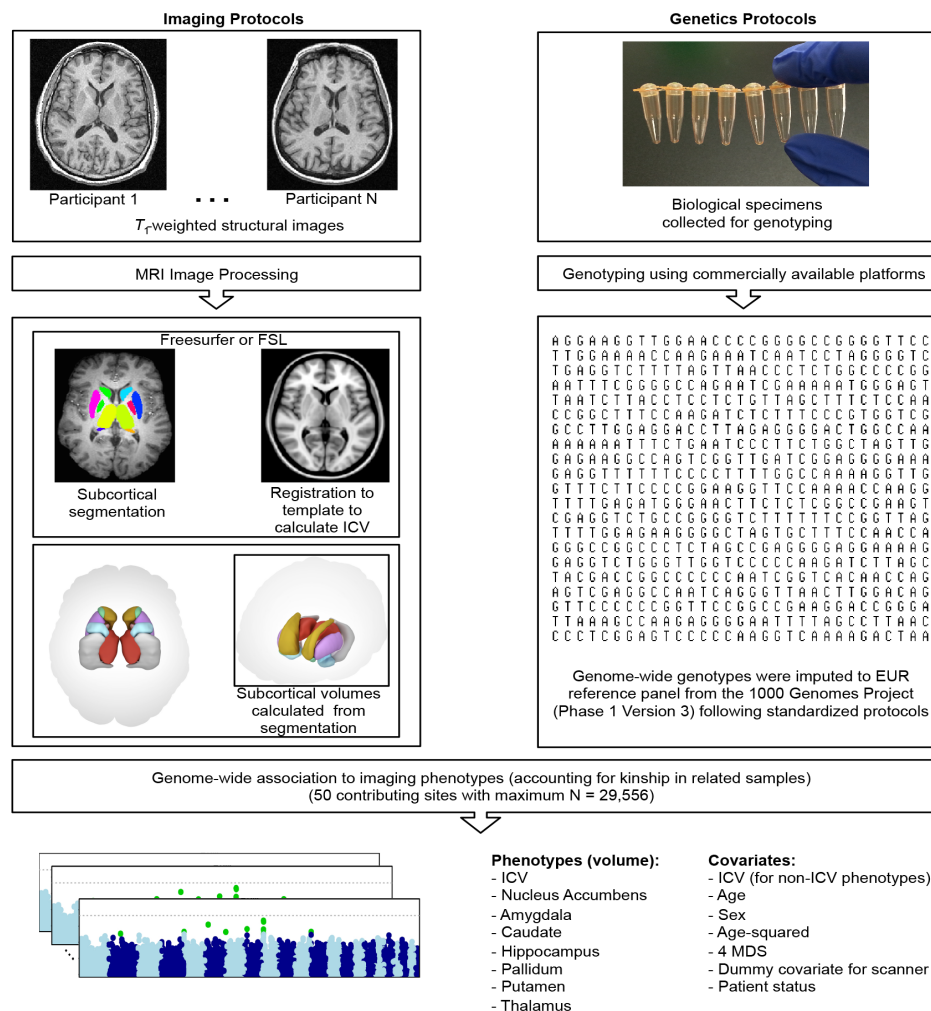
*Now expanded to multiple multi-national
meta-analytic projects:*

Genetics of brain structure
Genetics of brain function
Meta-analytic effects on disease



Procedure for meta-analysis (Part 1)

Completed at the level of the individual site

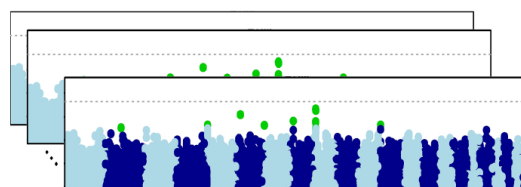


All protocols available at
<http://enigma.ini.usc.edu>



Procedure for meta-analysis (Part 2)

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

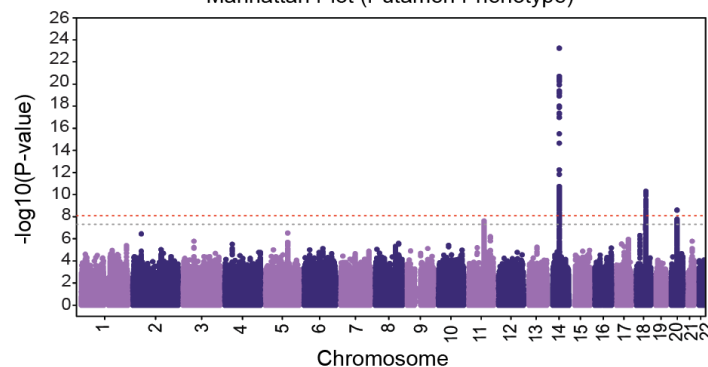
Covariates:

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

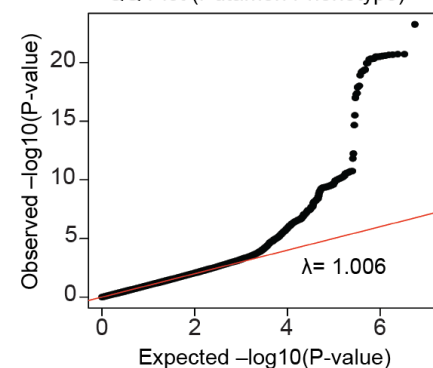
Quality Checking and Filtering ($MAC < 10$, $R^2 < 0.5$)

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

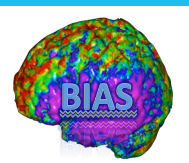
Manhattan Plot (Putamen Phenotype)



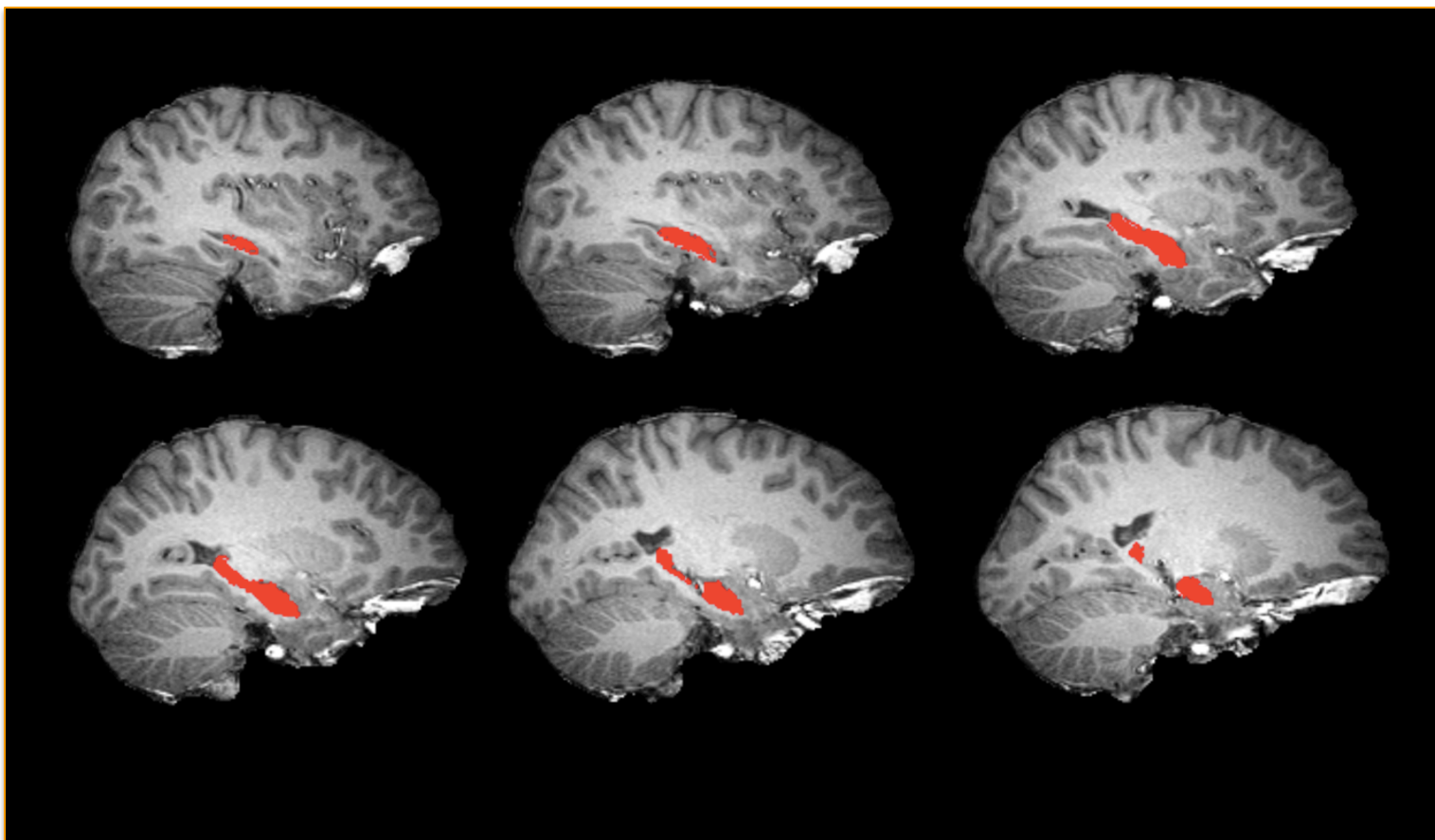
QQ Plot (Putamen Phenotype)



Uploaded to ENIGMA server for analysis at central site



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



21,151 individuals in discovery + replication



ENIGMA2: Genetics of subcortical structures

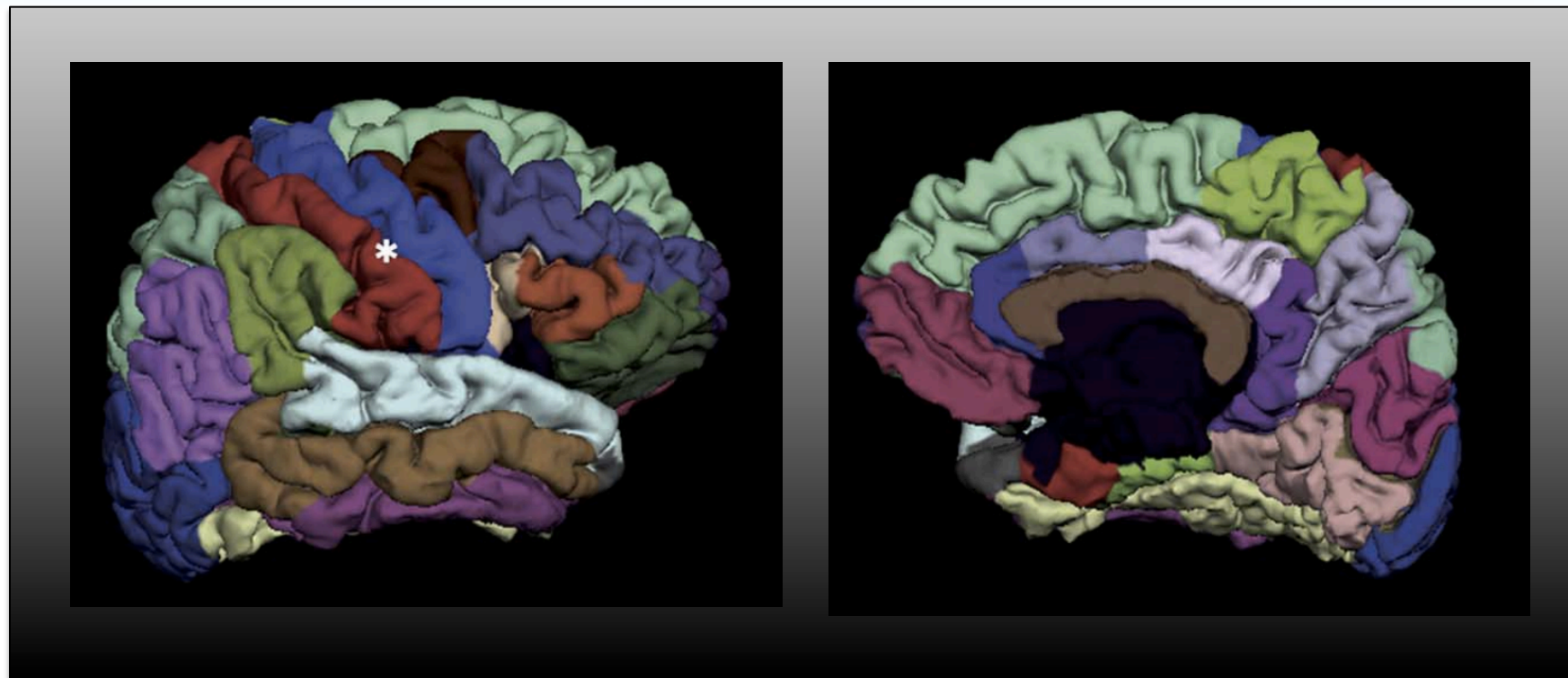


Discovery + Replication = 30,717 individuals
50 contributing cohorts around the world

(Hibar et al., 2015. *Nature*)



What's next for ENIGMA?

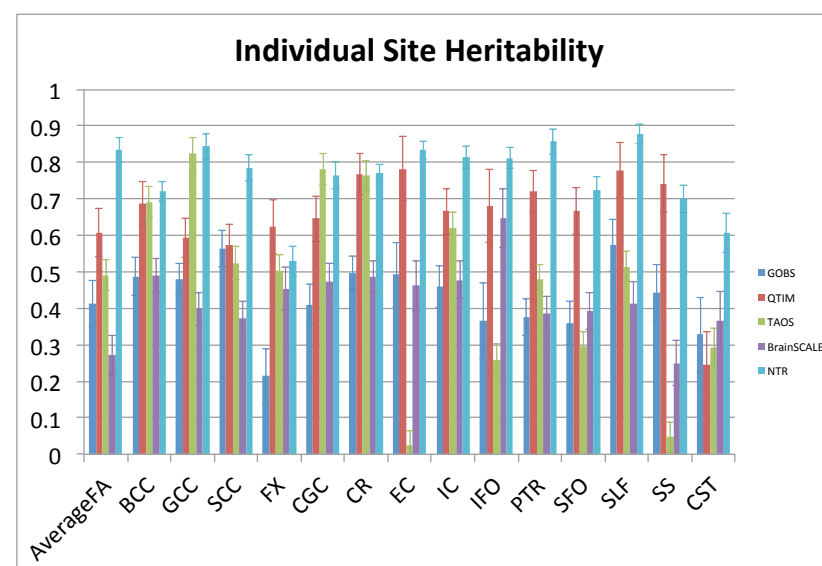
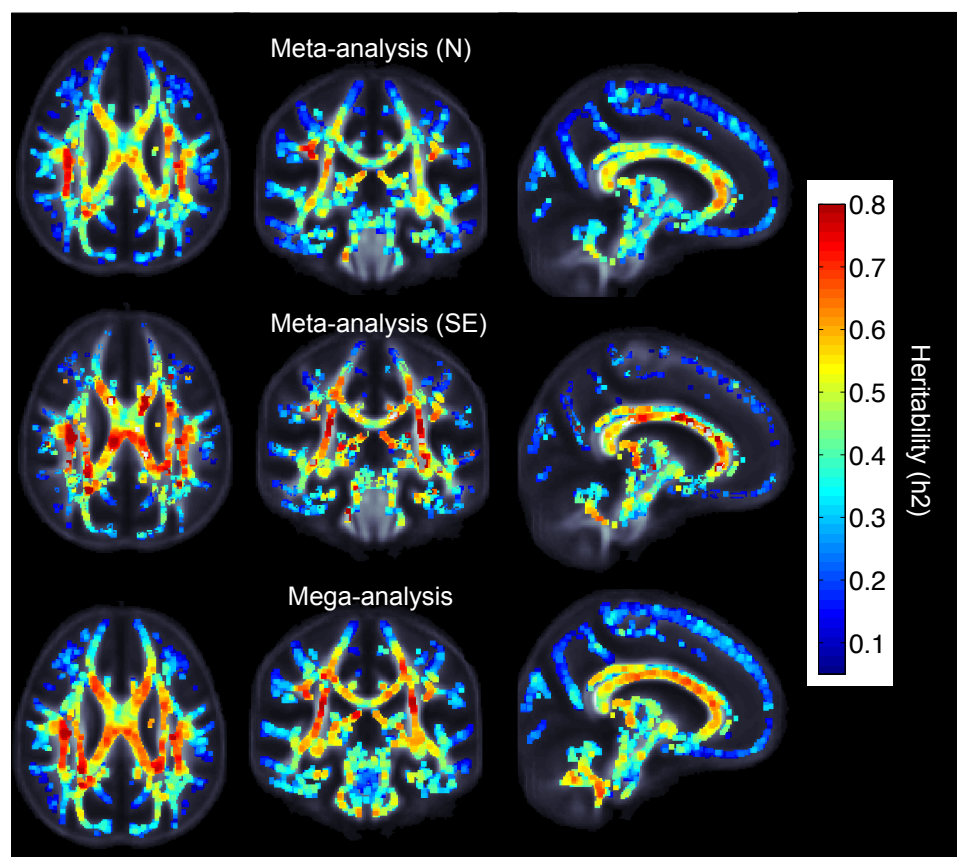


ENIGMA3: GWAS of cortical area and thickness
Projected N=30,000

Interested in joining? Sign up at <http://enigma.ini.usc.edu/>



ENIGMA-DTI





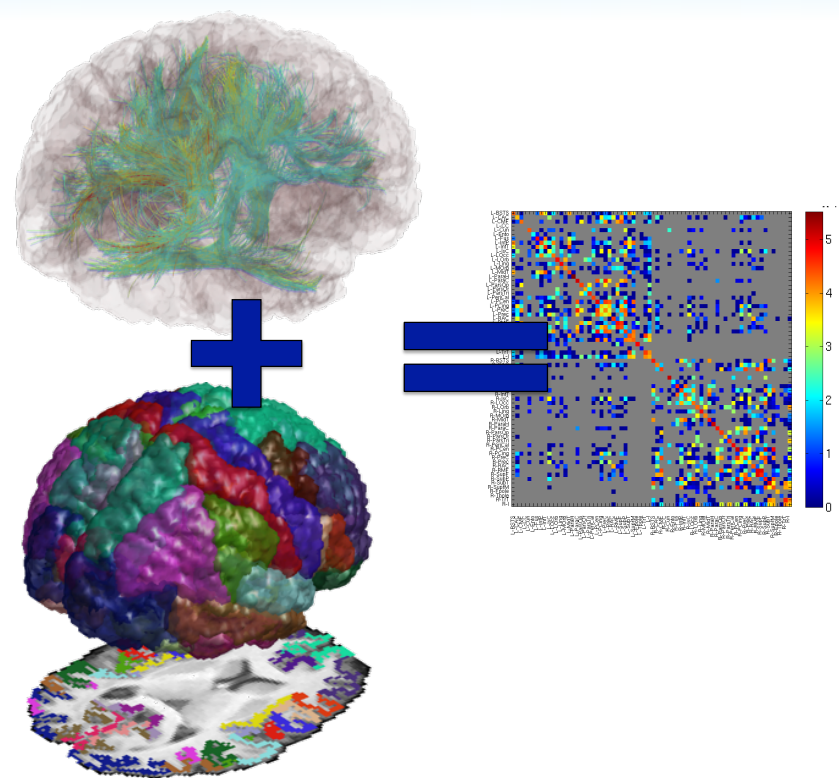
Mapping Connectivity

- Structural magnetic resonance imaging (MRI)
 - ✧ Anatomical analysis
 - ✧ Mapping Tissue Loss



Diffusion weighted

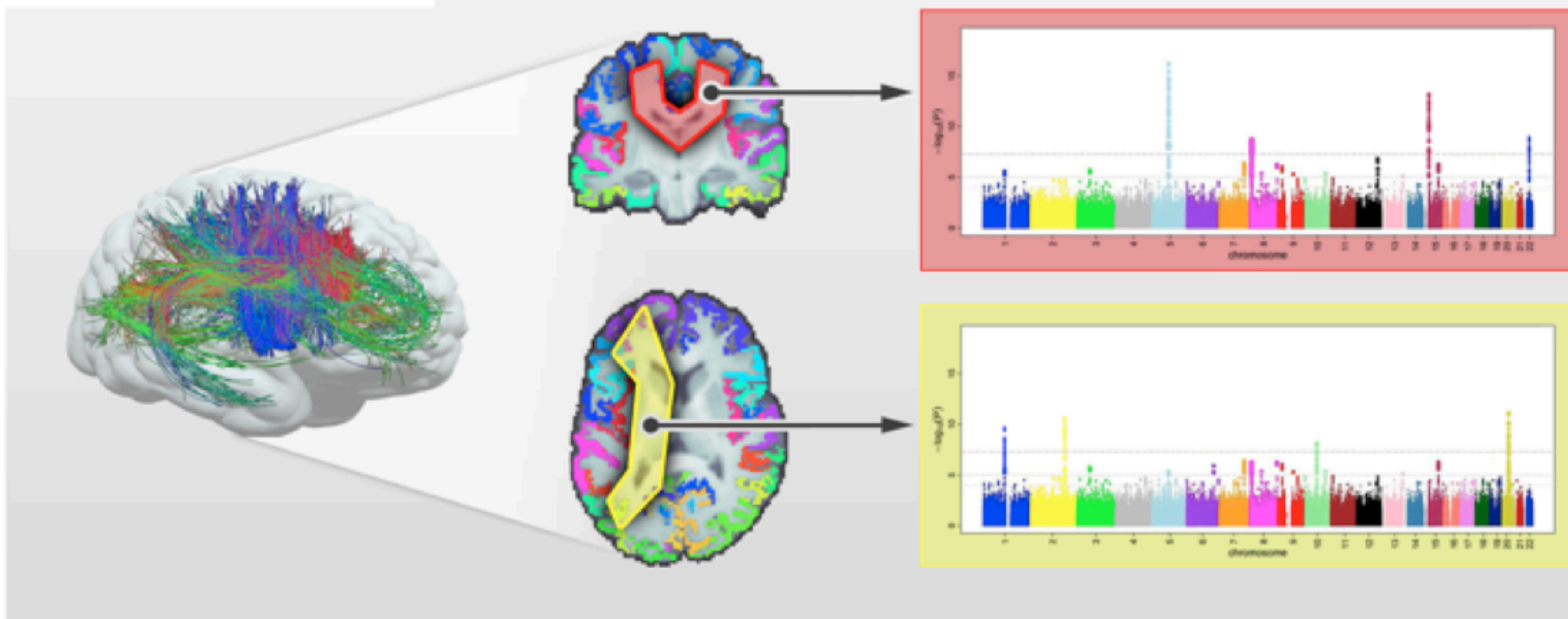
- ✧ Microstructure
- ✧ Connectivity





Connectome-Wide Genome-Wide Screen Alzheimer risk gene

Connectome-wide GWAS



Discovery sample – Young Adults

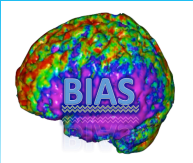
Effect in ADNI

Within 2 weeks Sherva et al. published *SPON1*

Found in a cognitive GWAS in AD



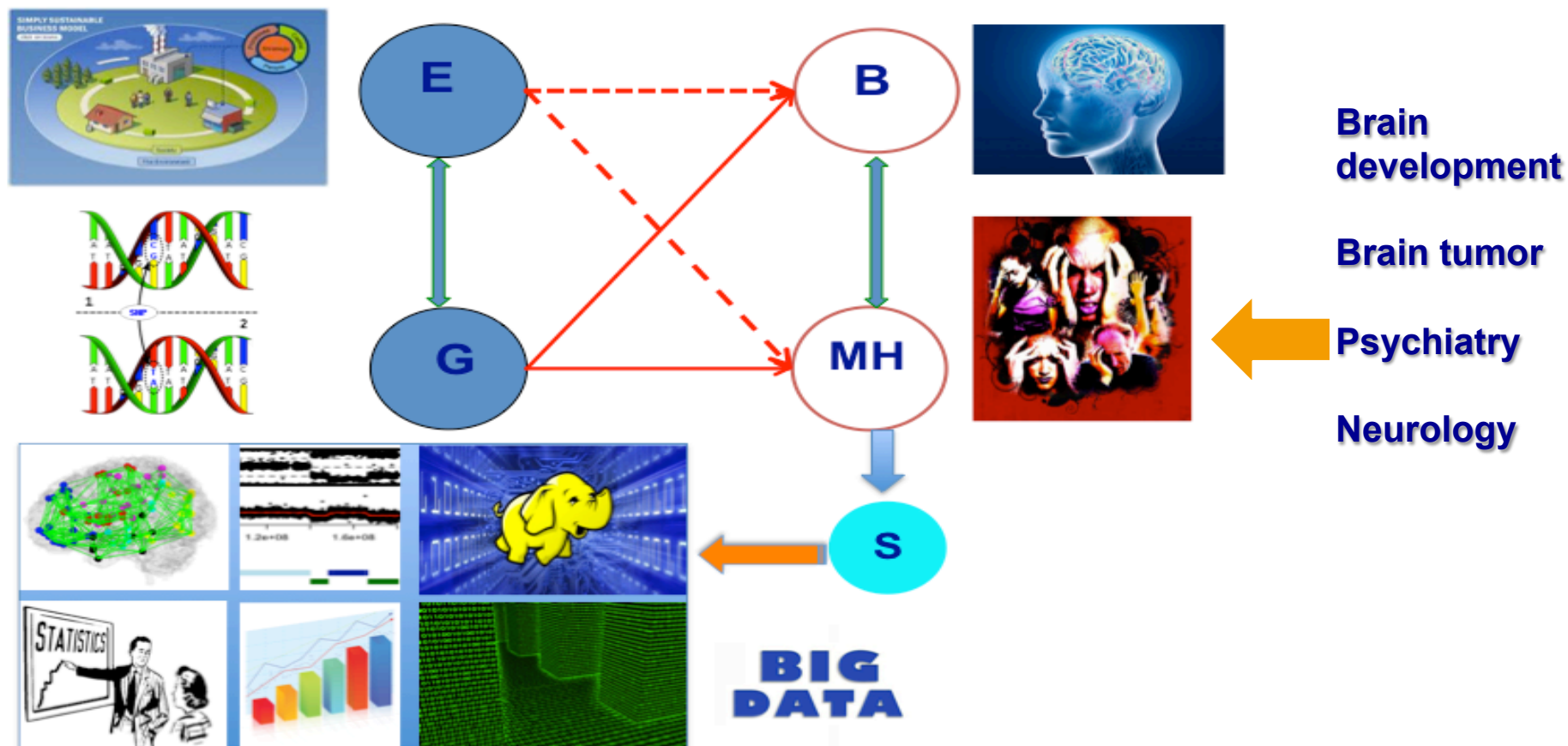
Jahanshad et al., PNAS 2013



Overview



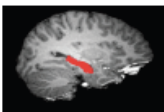
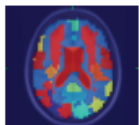
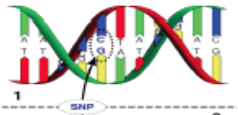
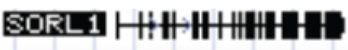
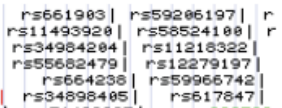
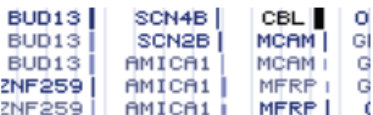
Imaging Genetic Studies



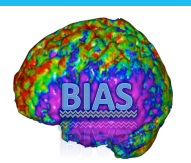
Common Goal: Detect potential genes for inherited phenotypes



Statistical Methods

Imaging Genetics	Candidate ROI 	Many ROI 	Voxelwise 
Candidate SNP 	Imager	Imager	Imager
Candidate Gene 	Geneticist	↑	↑
Genome-wide SNP 	Geneticist	↑	↑
Genome-wide Gene 	Geneticist		

Hibar, et al. HBM 2012



High Dimensional Regression Model

Data $\{(Y_i, X_i) : i = 1, \dots, n\}$

$$Y_i = \{y_i(v) : v \in V\}$$

$$\{X_i(g) : g \in G_0\}$$

Phenotype

Y

$$n \times p_y$$

Genotype

X

$$n \times p_x$$

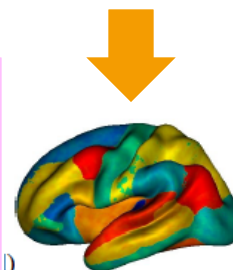
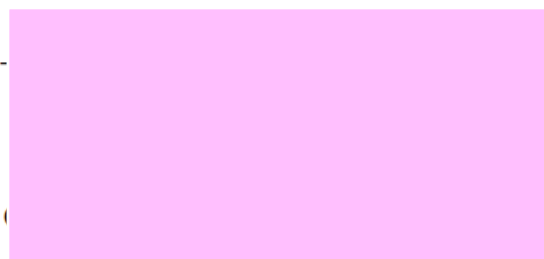
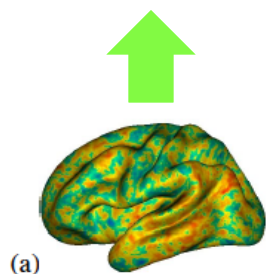
B

$$p_x \times p_y$$

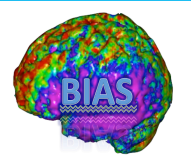
Error

E

$$n \times p_y$$



$$(p_x, n, p_y)$$



Current Approaches and Limitations

Candidate Gene
And / Or
Candidate Imaging Region



Only **part** of the picture



<http://www.tnooz.com/>

Genomic-wide Search
+
Voxel-based Approach

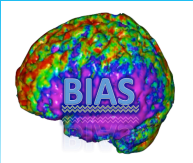


~~correlation (LD) structure~~
~~spatial correlation in voxels~~
Low statistical power

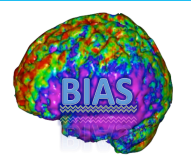
Everything in a big model
+
Penalized Method



Unstable variable selection
+
~~Statistical inference~~



Models for Candidate ROI Candidate Gene-set Analysis



High Dimensional Regression Model

Data $\{(Y_i, X_i) : i = 1, \dots, n\}$

$$Y_i = \{y_i(v) : v \in V_0\} \quad X_i = \{X_i(g) : g \in G_0\}$$

Phenotype

Y

Genotype

X

B

Error

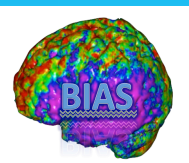
E

$$\begin{array}{c} n \times p_y \end{array} = \begin{array}{c} n \times p_x \end{array} \begin{array}{c} p_x \times p_y \end{array} + \begin{array}{c} n \times p_y \end{array}$$

Key Conditions:

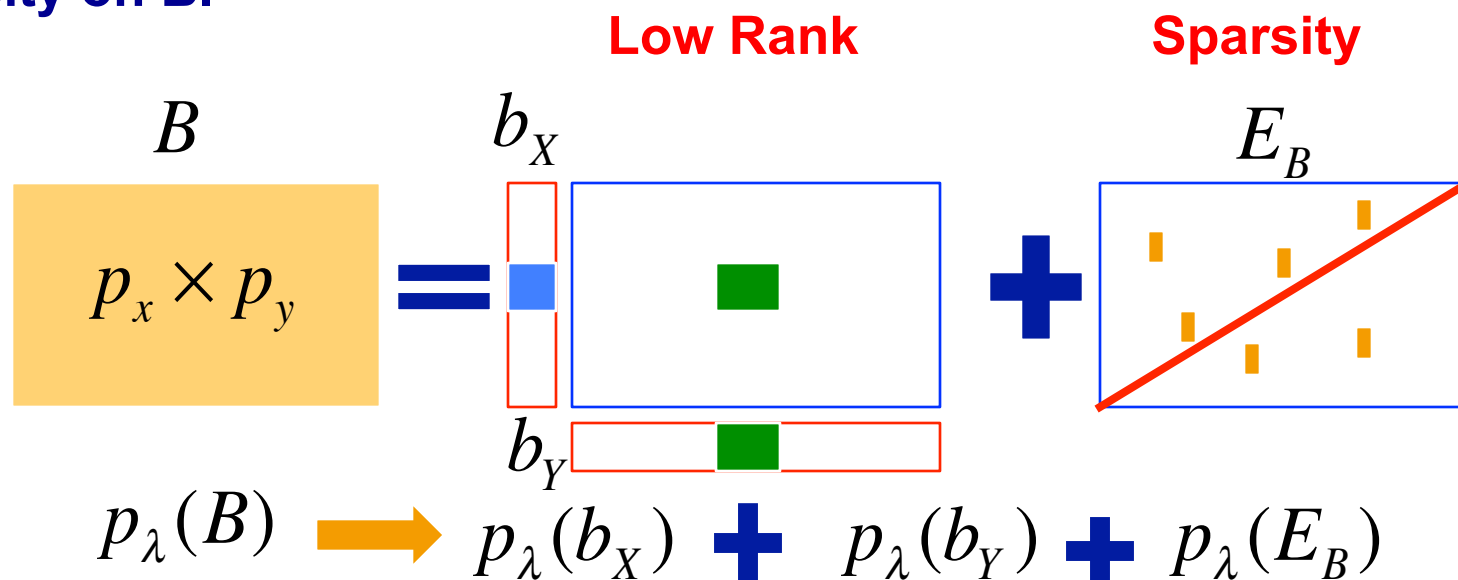
$$\max(p_x, p_y) \sim n$$

- Sparsity of B
- Restricted null-space property for design matrix X



Sparse and Low-rank Representation

Sparsity on B.



Regularization Methods

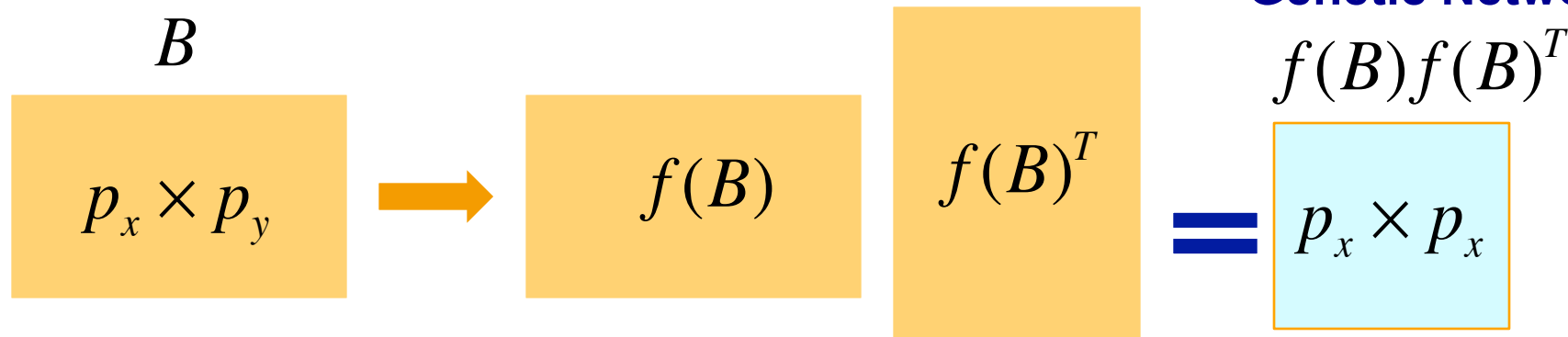
- Lasso 1, 2, 3,
- SCAD, MCP,

$$\hat{\theta} \in \arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n (y_i - x_i^T \theta)^2 + \lambda_n \sum_{j=1}^p |\theta_j|$$

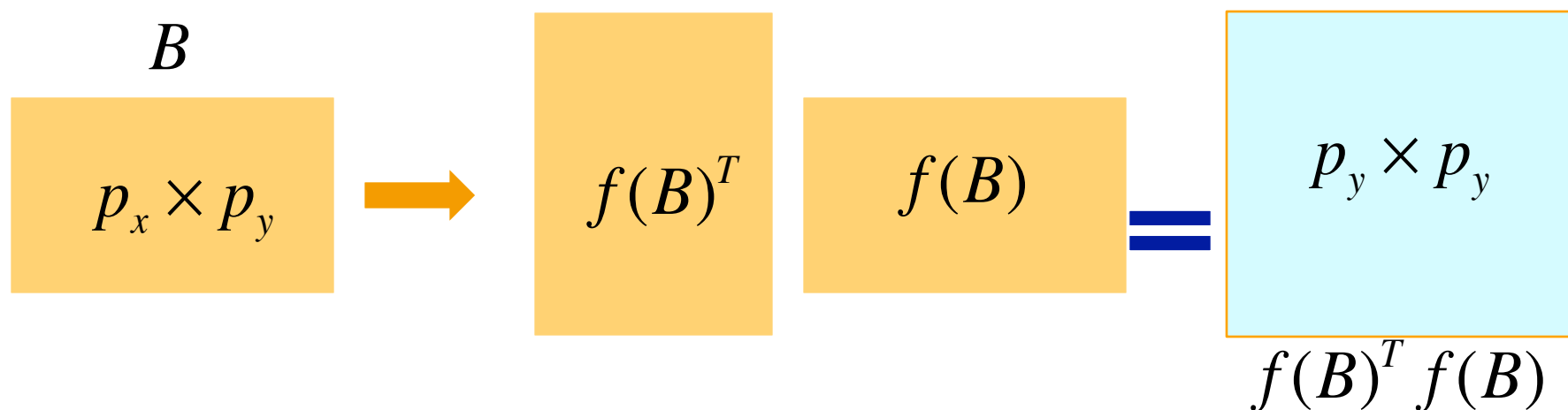


Genetic and Imaging Networks

Genetic Network

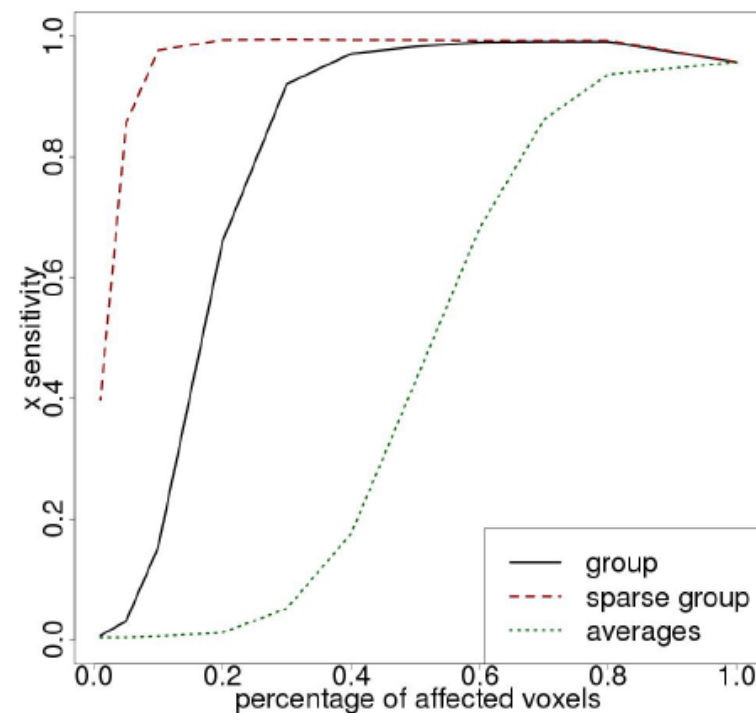
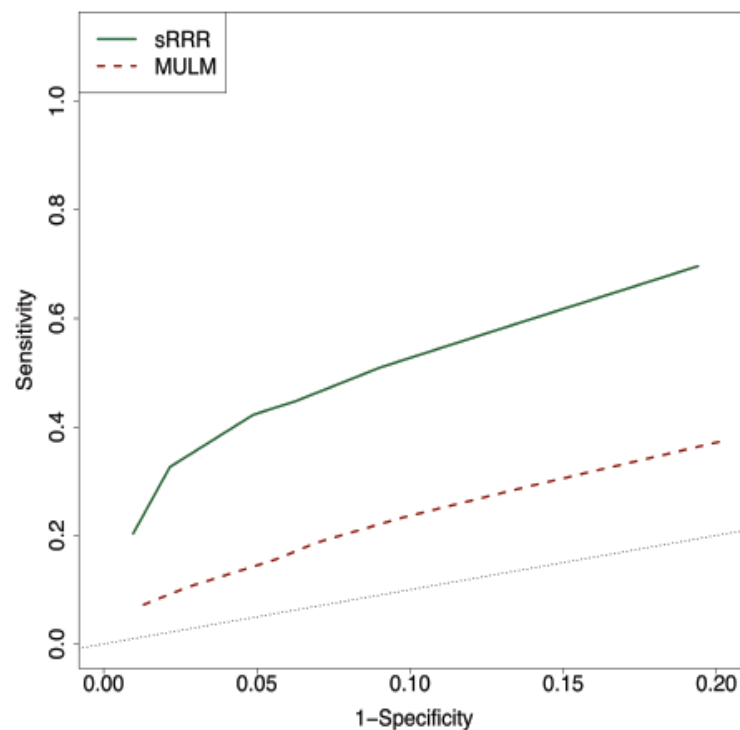
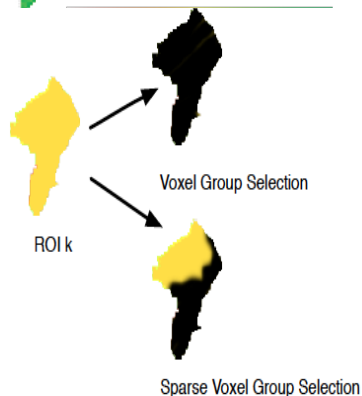
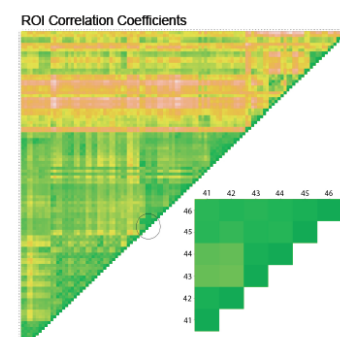
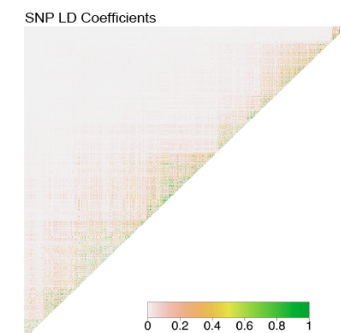


Imaging Network





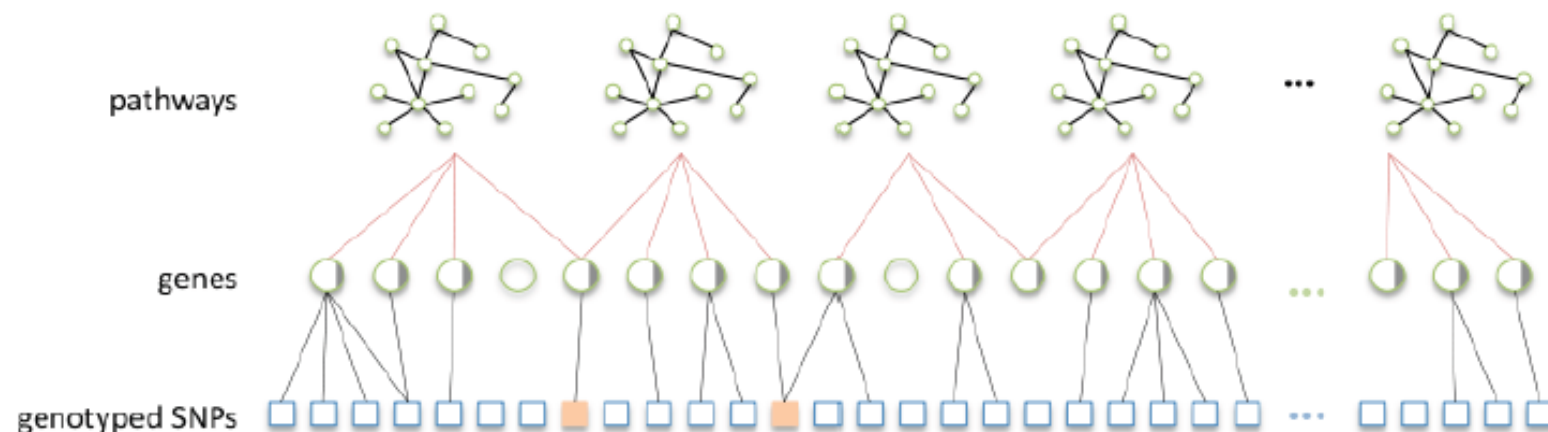
Simulation



Vounou et al. (2010, 2012) NeuroImage



SNPs to Genes to Pathways



- ▶ Group all available SNPs into L pathways $\mathcal{G}_1, \dots, \mathcal{G}_L$
- ▶ Adopt a *group lasso* penalty to force group selection,

$$P_b(\mathbf{b}) = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda \sum_{l=1}^L w_l \|\mathbf{b}_l\|_2$$

to select only most predictive pathways,

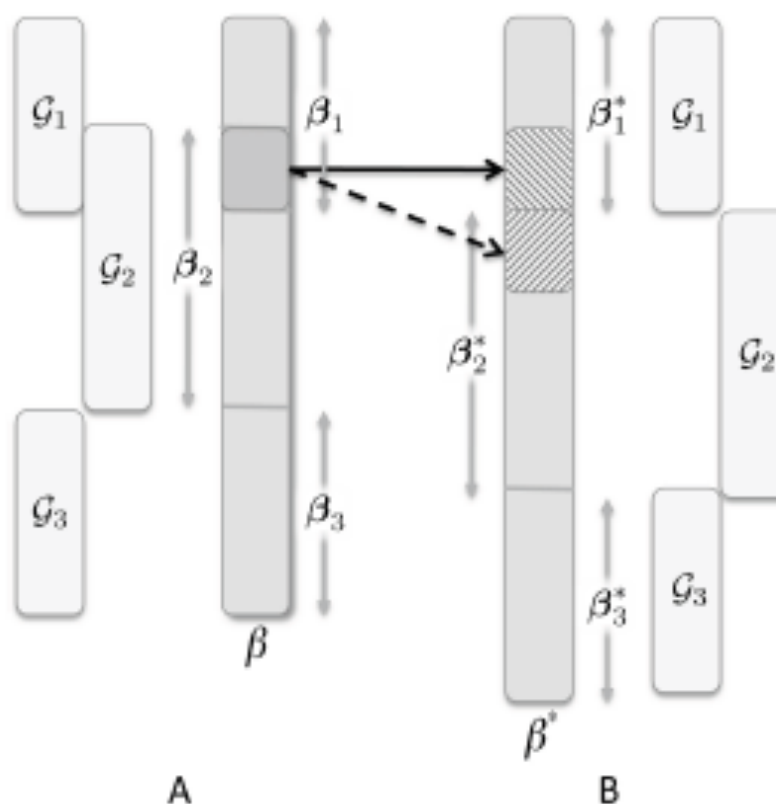
$$\mathbf{b} = \left\{ \underbrace{(0, \dots, 0)}_{\mathcal{G}_1}, \dots, \underbrace{(0, \dots, \mathbf{b}_{l_a}, 0, \dots, \mathbf{b}_{l_b}, 0, \dots, 0)}_{\mathcal{G}_l}, \dots, \underbrace{(0, \dots, 0)}_{\mathcal{G}_L} \right\}$$

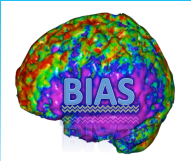


Overlapping Pathways

SNPs Duplication in Overlapping Pathways

Three pathways $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ and grouped regression coefficients $\beta_1, \beta_2, \beta_3$





ANDI

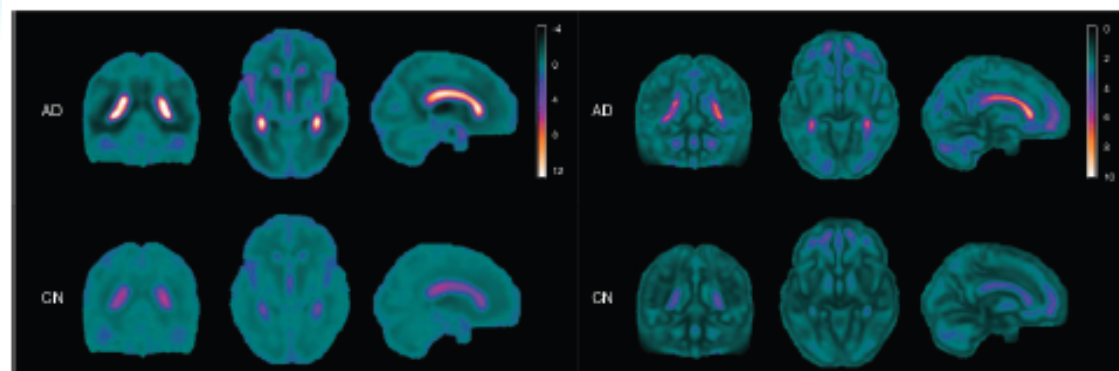
- Serial brain MRI scans were analyzed from 200 probable AD patients and 232 healthy elderly controls (CN)
- Longitudinal scans available at three time points

	Screening	6Mo	12Mo	24Mo
AD	200	165	144	111
CN	232	214	202	178
Total	432	379	346	289

At screening:			
Group	age (years)	N male	N female
AD	75.7±7.7	103	97
CN	76.0±5.0	120	112

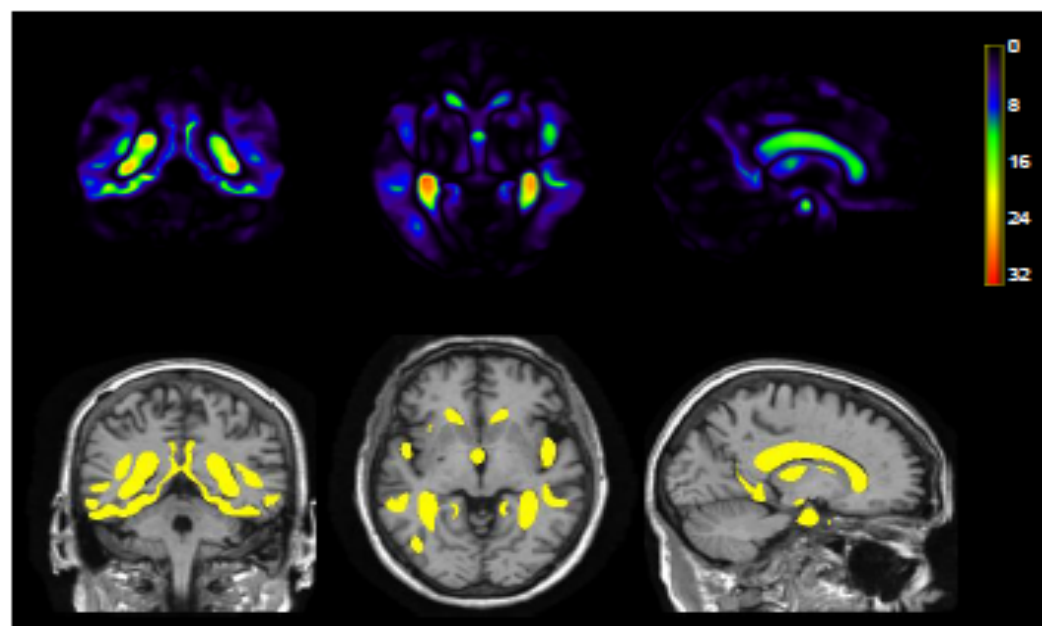


AD Imaging Signature



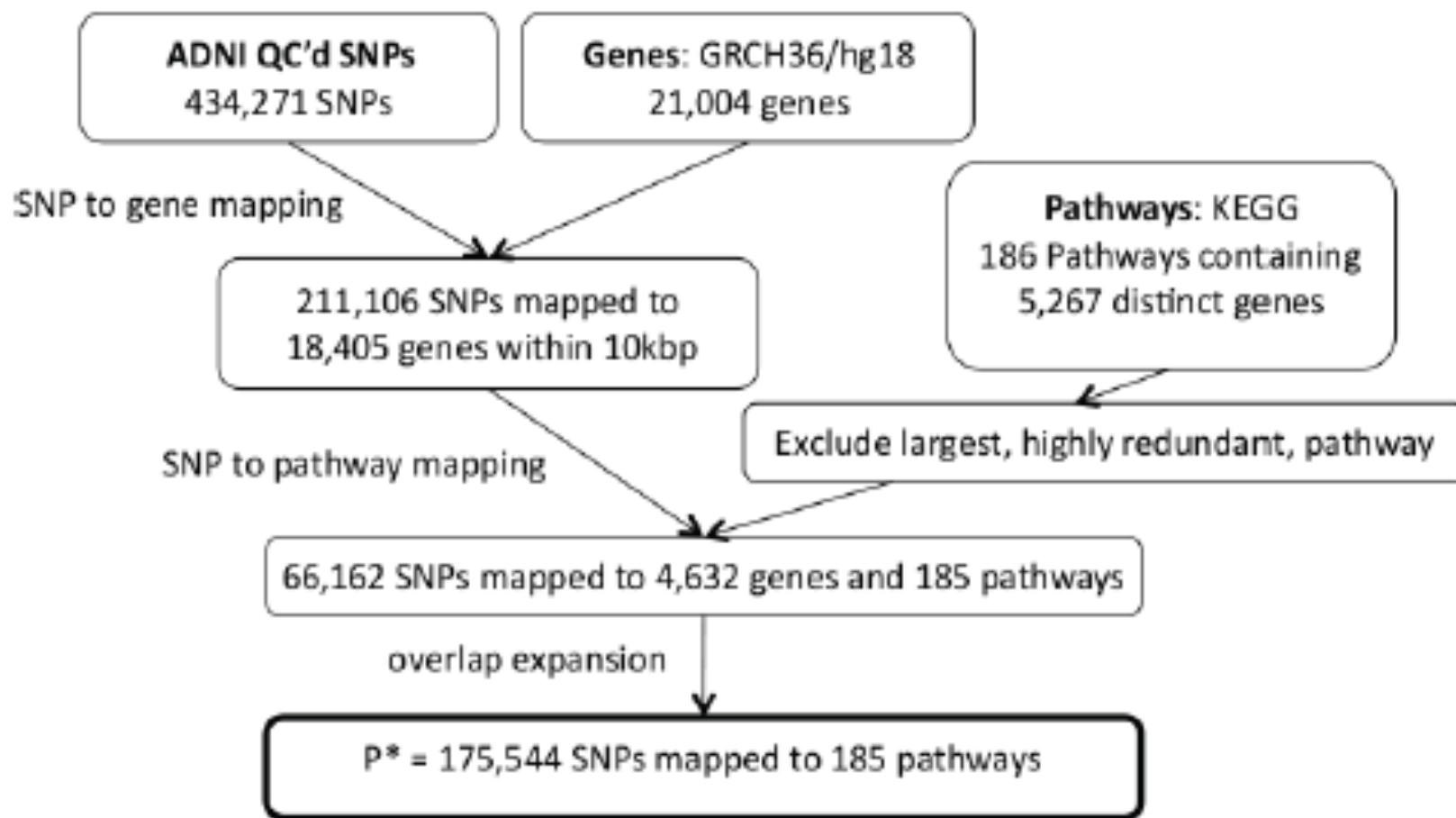
Mean and STD of slopes

$-\log_{10}(p)$ value map





Mapping SNPs to Pathway





Top 15 SNPs and Genes Ranked by SRRR

Rank	SNP	SNP RANKING		GENE RANKING		
		π^{SNP}	Mapped gene(s)	Gene	π^{gene}	# SNPs
1	rs4788426	0.451	<i>PRKCB</i>	PRKCB	0.451	73
2	rs11074601	0.429	<i>PRKCB</i>	ADCY8	0.411	69
3	rs263264	0.411	<i>ADCY8</i>	ADCY2	0.392	106
4	rs13189711	0.392	<i>ADCY2</i>	HK2	0.302	28
5	rs680545	0.302	<i>HK2</i>	PRKCA	0.290	99
6	rs4622543	0.290	<i>PRKCA</i>	PIK3R3	0.267	9
7	rs9896483	0.274	<i>PRKCA</i>	MYLK	0.234	24
8	rs1052610	0.267	<i>PIK3R3</i>	PIK3CG	0.207	9
9	APO ϵ 4	0.251	<i>TOMM40 APOE</i>	COL5A3	0.174	14
10	rs1254403	0.234	<i>MYLK</i>	GNAI1	0.167	22
11	rs4730205	0.207	<i>PIK3CG</i>	ACACA	0.164	23
12	rs889130	0.174	<i>COL5A3</i>	G6PC	0.163	6
13	rs6973616	0.167	<i>GNAI1</i>	DGKA	0.160	3
14	rs9906543	0.164	<i>ACACA</i>	CR1	0.154	21
15	rs2229611	0.163	<i>G6PC</i>	TOMM40	0.152	6



Factor Model

$$\begin{array}{c}
 E \\
 n \times p_y
 \end{array}
 =
 \begin{array}{c}
 \text{Long-range} \\
 \text{Correlation}
 \end{array}
 \begin{array}{c}
 \Lambda \\
 p_y \times q
 \end{array}
 \begin{array}{c}
 \xi_i \\
 q \times 1
 \end{array}
 +
 \begin{array}{c}
 \text{Short-range} \\
 \text{Correlation}
 \end{array}
 \begin{array}{c}
 \eta_i \\
 p_y \times 1
 \end{array}$$

$$\Sigma_E = \begin{array}{c} \Lambda \\ \Lambda^T \end{array} + \Sigma_\eta$$



Simulation

Patterns

Plus

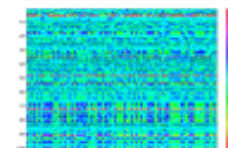
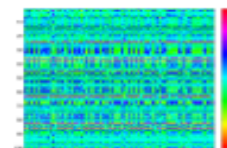
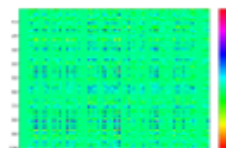
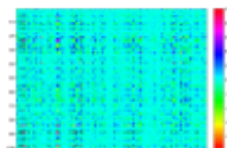
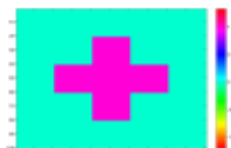
SVD

SVD

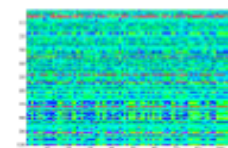
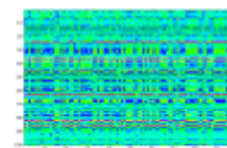
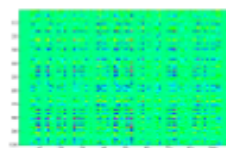
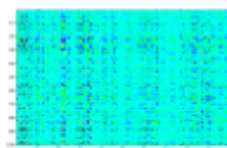
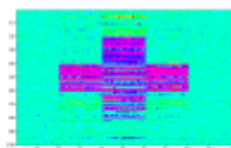
UN

UN

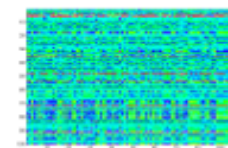
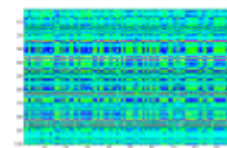
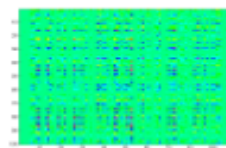
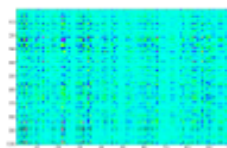
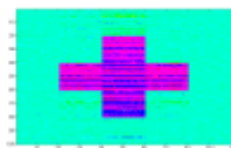
True B



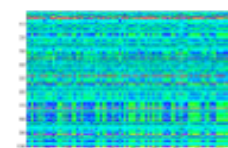
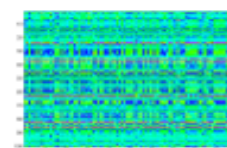
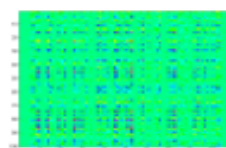
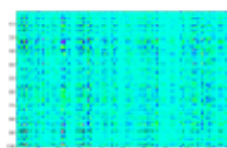
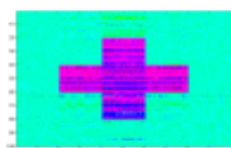
LASSO



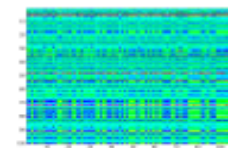
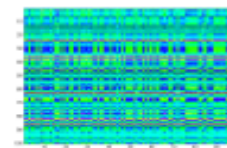
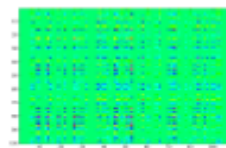
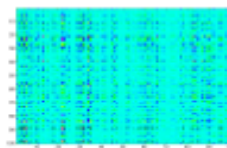
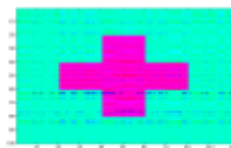
BLASSO



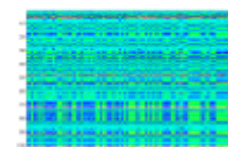
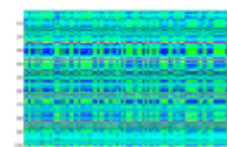
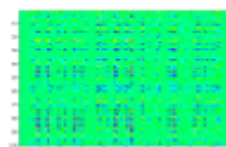
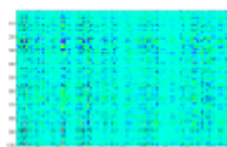
G-SMuRFS



GLRR3



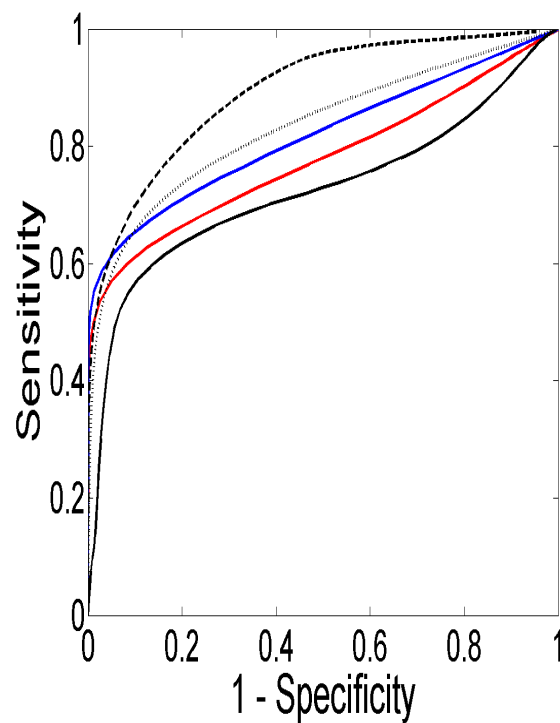
GLRR5



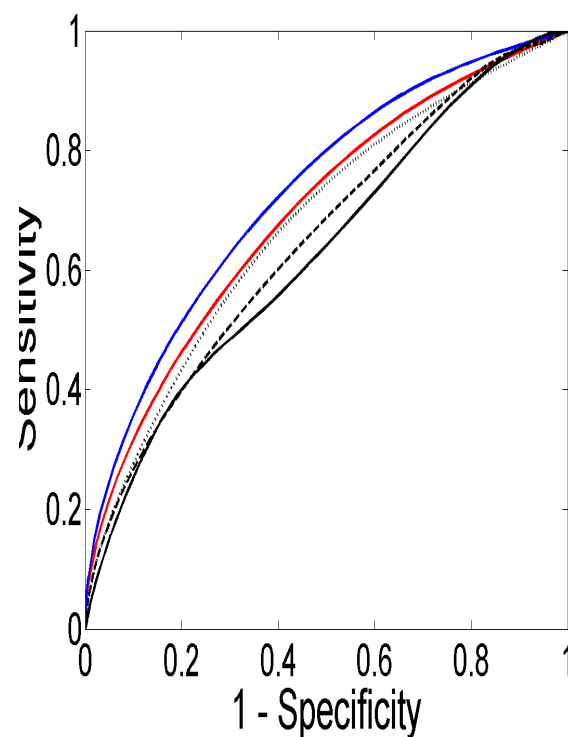


Simulation

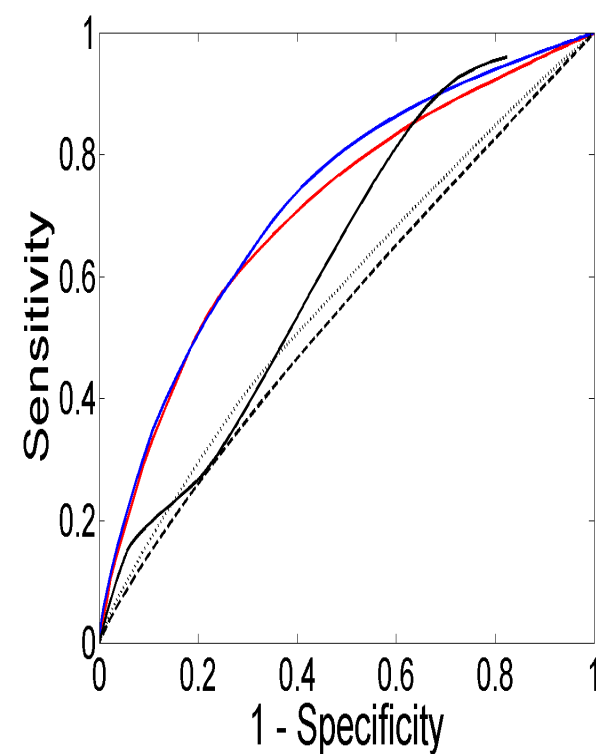
B=Plus X=Cont.



B=SVD X=Cont.



B=SVD X=SNP

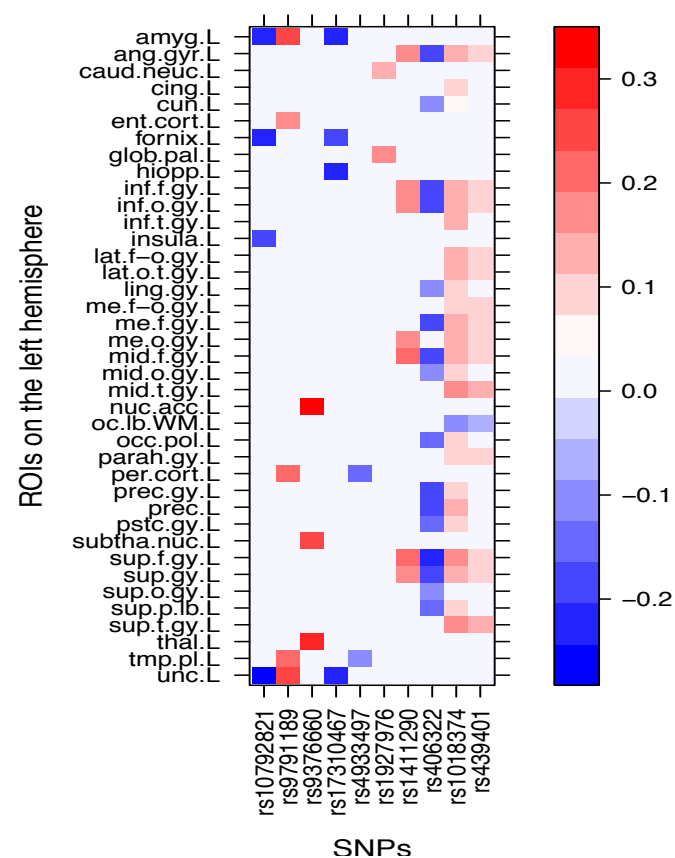
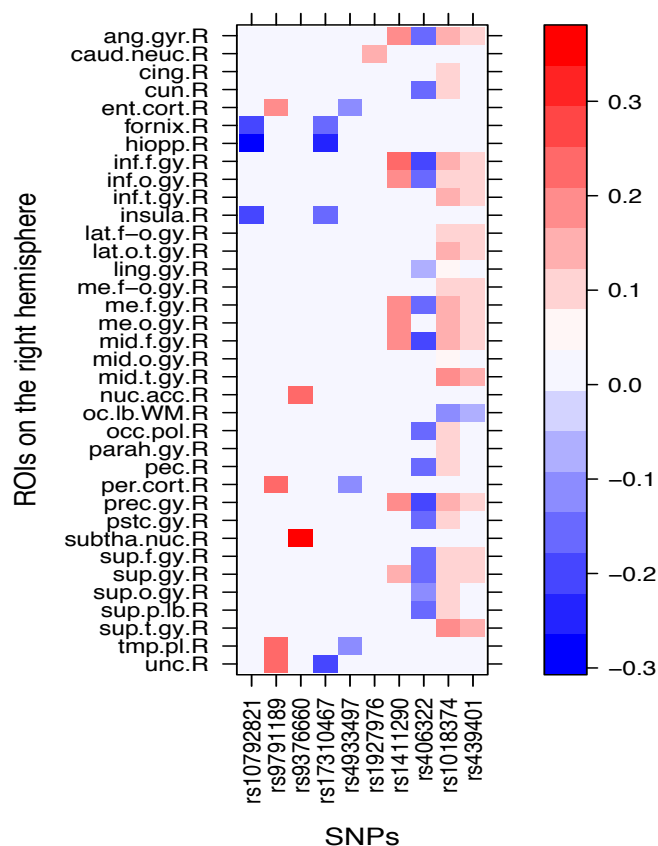
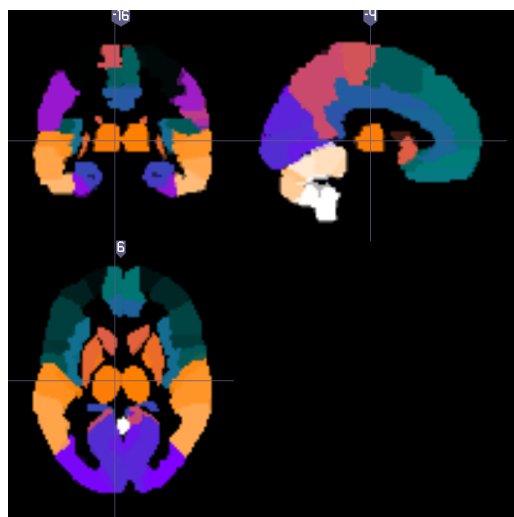




ADNI

749 AD/MCI/NC subjects, 93 ROIs

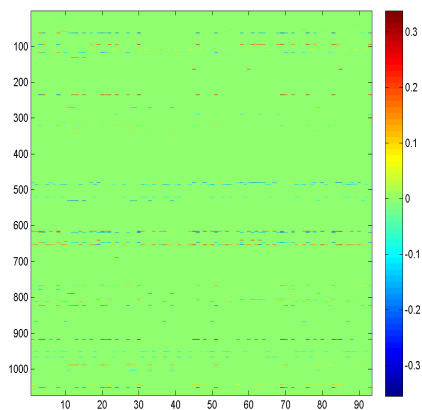
40 AD candidate genes on the AlzGene web



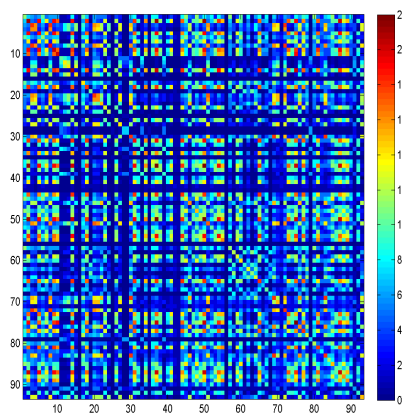


ADNI

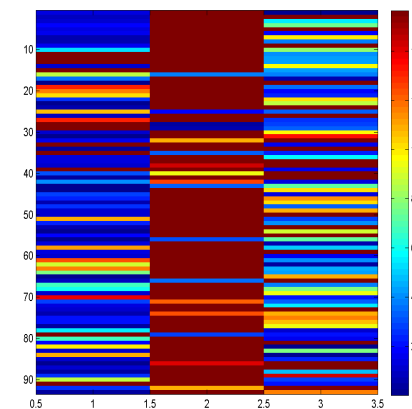
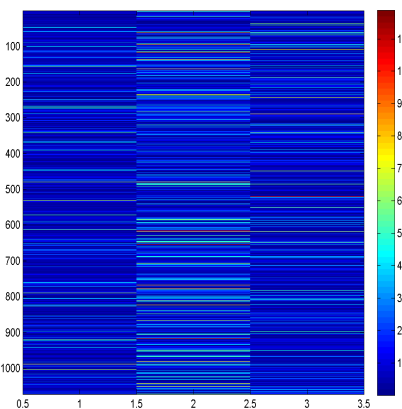
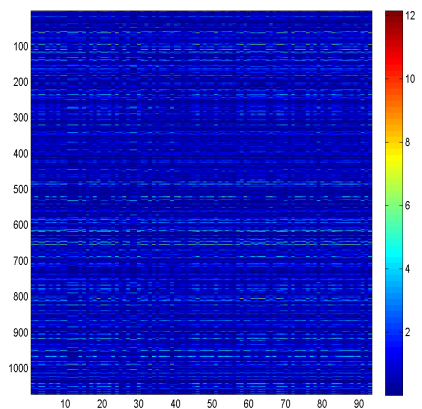
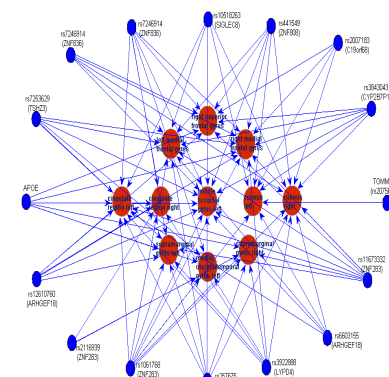
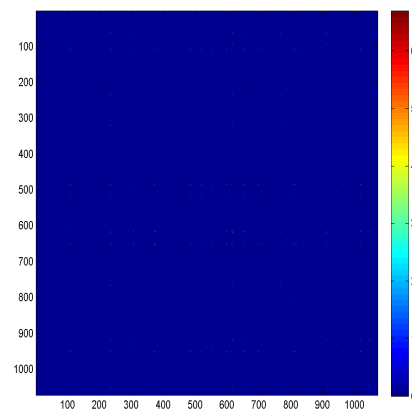
$\hat{B}$



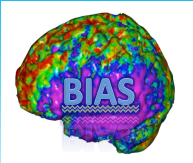
ROI network



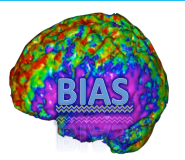
Genetic network



$-\log_{10}(p)$ for $\hat{B}$

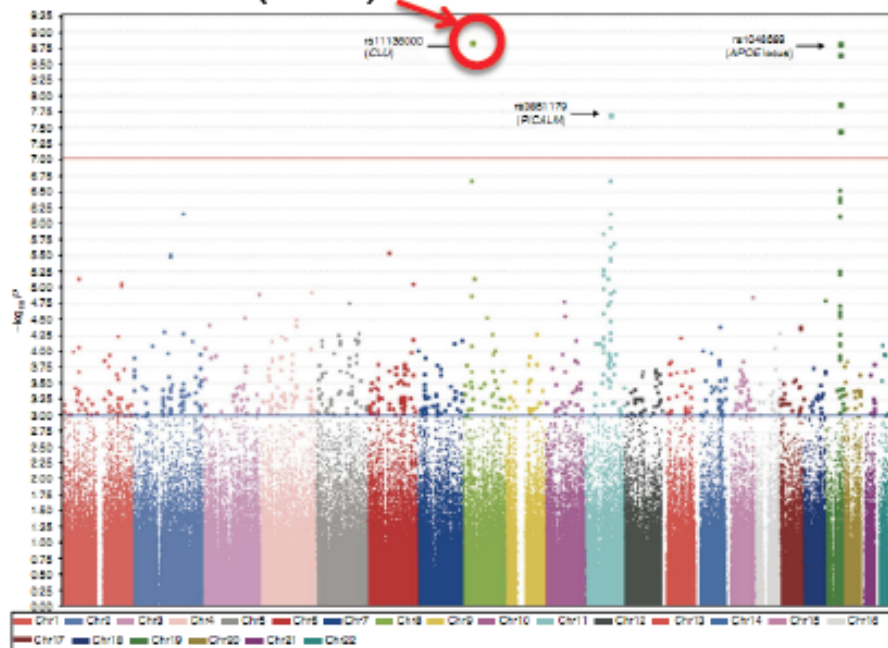


Models for Candidate SNP/Gene Voxel-wise Analysis



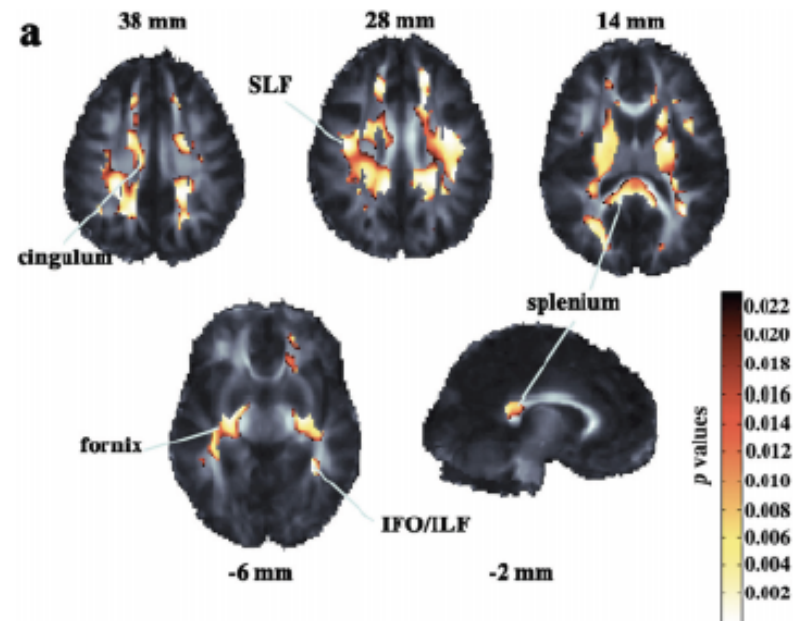
Candidate SNP

rs11136000 (CLU)



Genome-wide association identifies variant within the CLU gene in ~4000 Alzheimer's patients and ~8000 controls – but what does it do?

(Harold et al., 2009)



The Alzheimer's associated variant broadly affects white matter integrity in a young cohort – may create early predisposition for disease

(Braskie et al., 2009)



High Dimensional Regression Model

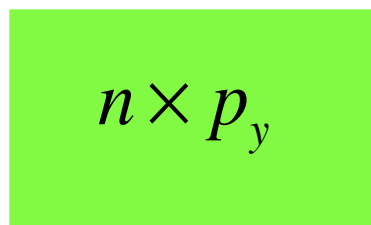
Data $\{(Y_i, X_i) : i = 1, \dots, n\}$

$$Y_i = \{y_i(v) : v \in V\}$$

$$\{X_i(g) : g \in G_0\}$$

Phenotype

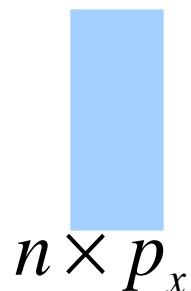
Y



=

Genotype

X



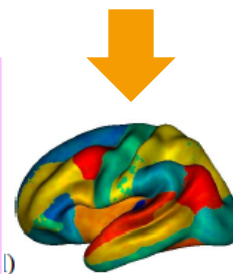
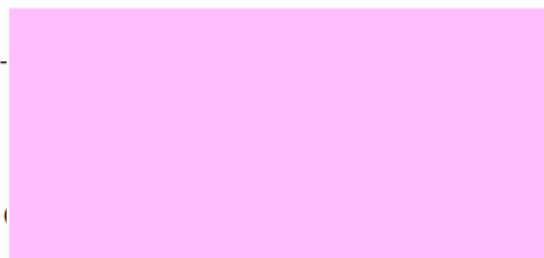
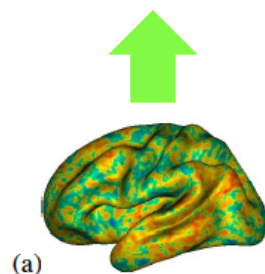
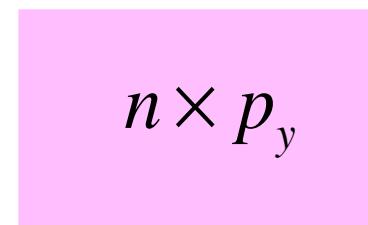
B



+

Error

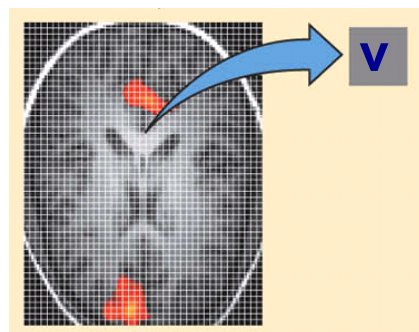
E



$$p_x \ll n \ll p_y$$



Functional Mixed Effects Model



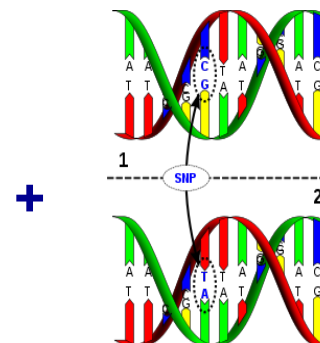
BEHAVIOR GENDER



=



Age



+

+

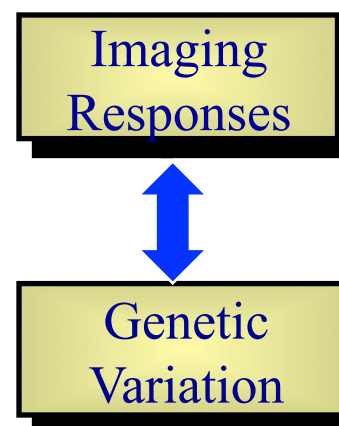
Error (v)

$$y_i(v) = x_i' \beta(v) + z_i' \gamma(v) + e_i(v)$$

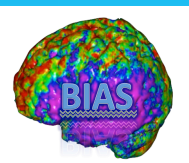
$$\gamma(v) \sim N(0, \sigma_\gamma^2(v) I_L)$$

$$e_i(v) \sim N(0, \sigma_e^2(v))$$

$$\dim(y_i(v)) = 1$$

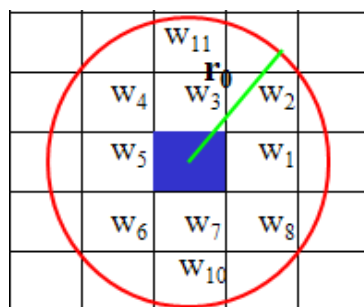


$$\sigma_\gamma^2(v) \stackrel{?}{=} 0$$

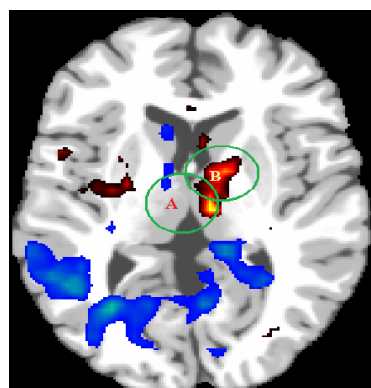


Multiscale Adaptive Smoothing

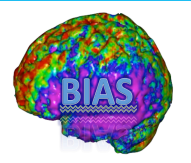
Estimate $\sigma_{\gamma}^2(v)$, $\sigma_e^2(v)$ marginally, using **regular** REML



Estimate $\sigma_{\gamma}^2(v)$ adaptively, using **weighted** REML



Hypothesis Testing for $H_0(v)$: $\sigma_{\gamma}^2(v) = 0$



Weighted REML

$$Y_{nx1}(v) = X_{n \times p} \beta(v)_{p \times 1} + Z_{n \times L} \gamma(v)_{L \times 1} + E(v)_{nx1}$$

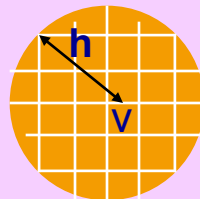
$$\downarrow K_{(n-p) \times n} X_{n \times p} = 0$$

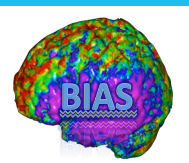
$$Y^*(v) = KY(v) = K Z \gamma(v) + K E(v) = Z^* \gamma(v) + E^*(v)$$

$$\downarrow \text{Regular REML}$$

$$\downarrow \text{Weighted REML}$$

$B(v, h) :$





Hypothesis Testing

$$H_0 : \sigma_v^2(v) = 0 \text{ vs } H_1 : \sigma_v^2(v) \neq 0$$

Test statistic $T(v)$:

$$2\{L_{REML}(Y^*(v) | Z, \hat{\sigma}_\gamma^2(v), \hat{\sigma}_e^2(v)) - L_{REML}(Y^*(v) | Z, 0, \hat{\sigma}_e^2(v))\}$$

Traditional approach: $T \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$

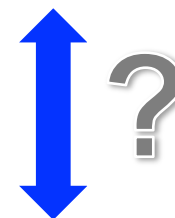
Our approach : **Exact Null Distribution**

$$T_n(v) \sim \sup_{\sigma^2(v) \geq 0} \sum_{B(v,h)} \omega(v, v', h) \sum_{i=1}^L \frac{\eta_i^2(v) \sigma_\gamma^2(v) d_i}{\hat{\sigma}_e^2(v') + \sigma_\gamma^2(v) d_i} - \log(\hat{\sigma}_e^2(v') + \sigma_\gamma^2(v) d_i)$$

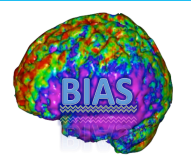
d_i : i^{th} eigenvalue of $KZ^T ZK^T$

$$\eta_i(v) \sim N(0, 1)$$

Imaging
Responses



Genetic Variation



Score Test

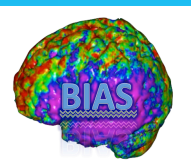
Likelihood Ratio Test:

- Smoothing estimator then computing LRT
- Solve it for **EVERY** random sample and **EVERY** voxels

Regular Score Test : $SC(v) = \frac{1}{2} \sigma_{ev}^{-4} Y_v^T \Omega Y_v$ (Tzeng2007)

- **No need to solve** maximizer every time
- **E(SC) and E(SC²)** feasible to calculate

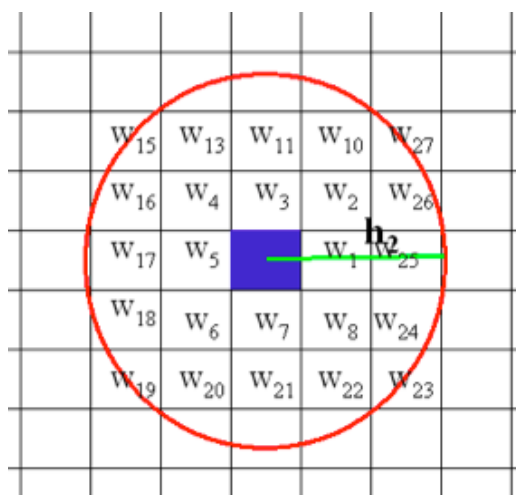
But statistical power is not good enough...



Weighted Score Test

Regular Score Test : $SC_{\sigma_{\gamma}^2}(v) = \frac{1}{2} \sigma_{ev}^{-4} Y_v^T \Omega Y_v \sim \text{Gamma}(\alpha, \beta)$

(Tzeng2007)



Weighted Score Test

$$WSC_{\sigma_{\gamma}^2}(v) = \frac{1}{2} \sum_{v' \in B(v, h)} \omega_{v'} \sigma_{ev'}^{-4} Y_{v'}^T \Omega Y_{v'}$$

$$\sim \text{Gamma}(\alpha', \beta')$$

No tricks : 20 Mins!



Results: Hypothesis Testing – Local view

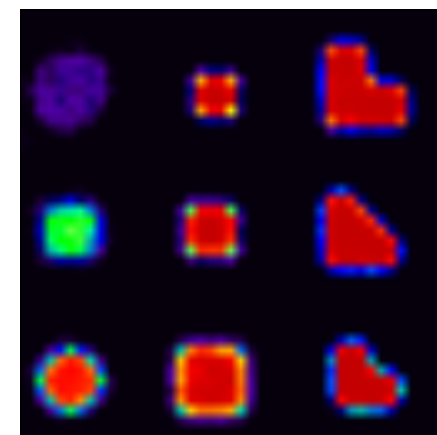
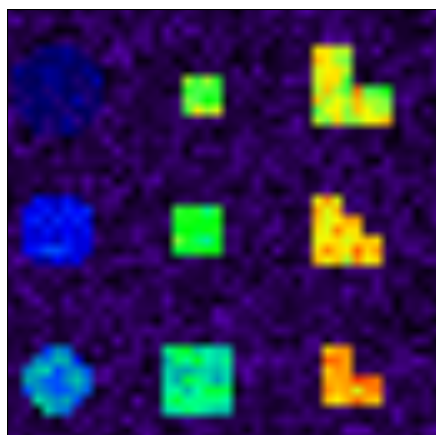
$$H_0 : \sigma^2_v(v) = 0 \text{ vs } H_1 : \sigma^2_v(v) \neq 0$$

Voxel Based

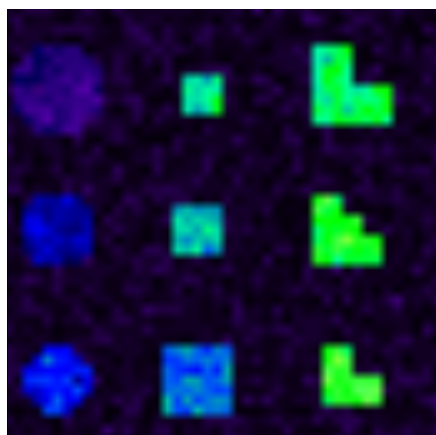
True Value

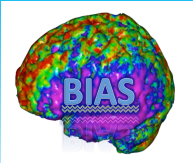
FMEM

LRT



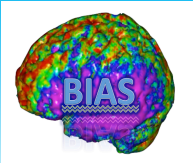
Score T





ADNI

- **Objective:** Identify brain regions affected by CR1
- **Subject:** 335 elders (174 Controls, 161 ADs)
- **Image:** 128 * 128 * 128 RAVEN maps from baseline T1-weighted images
- **Demographic Info:** Gender, Baseline age, Baseline age square, APOE Risk, Handedness, Education Level, Baseline intracranial volume
- **Genetic Variation:** CR1 (16 SNPs, MAFs from 13% ~ 41%)
- **Method:** Functional Mixed Effects Model (FEME)



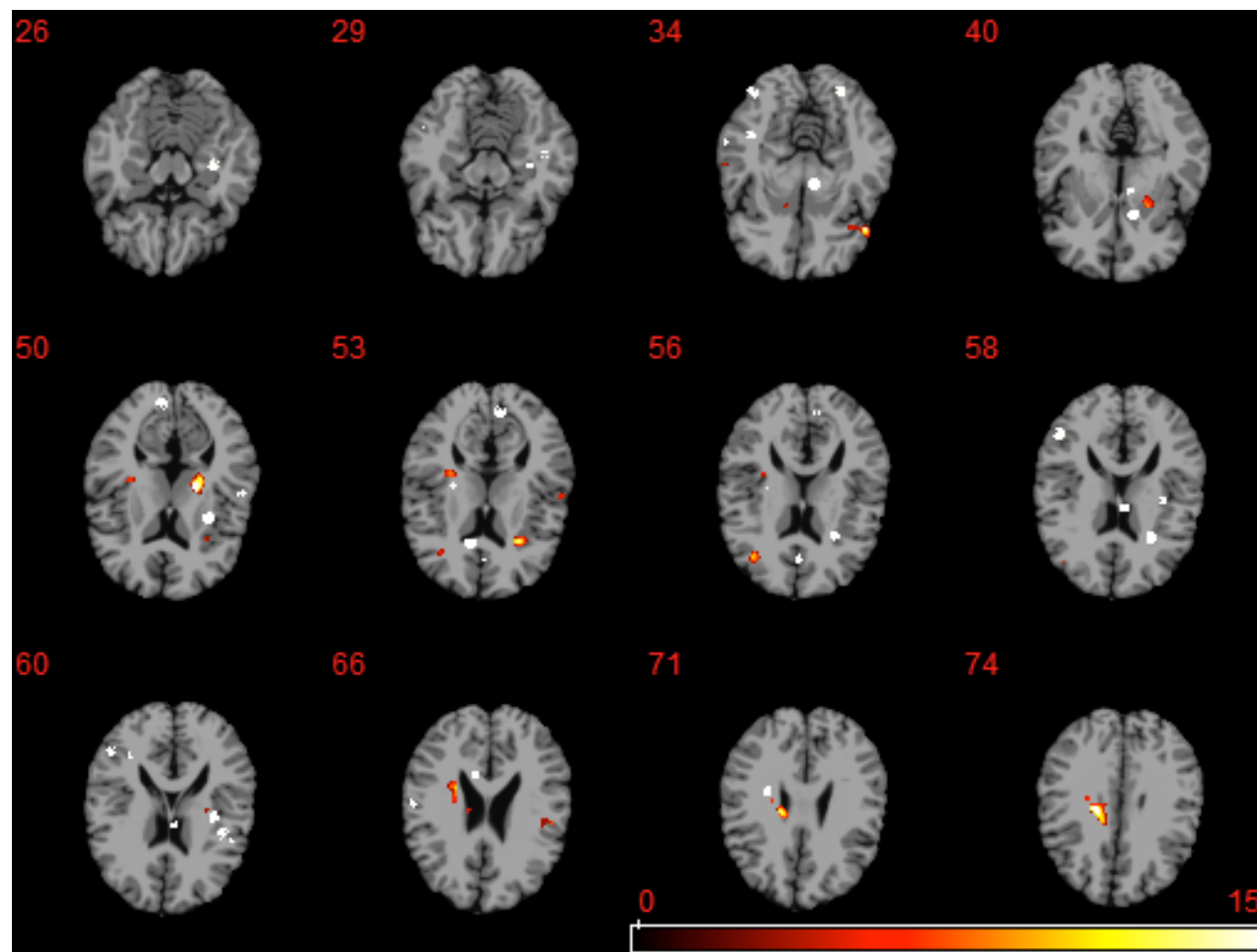
ADNI

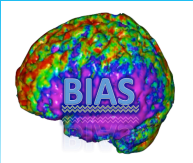
45 ROIs by FMEM, 5 ROIs by Voxel-based

Superior temporal gyrus (R167)	Median cingulate and paracingulate gyri (R209, L116)
Inferior temporal gyrus (R150, L378)	Calcarine fissure and surrounding cortex (R53, L266)
Precentral gyrus (R170, L*439)	Cuneus (R151, L345)
Middle frontal gyrus (R127, L295)	Superior occipital gyrus (L317)
Postcentral gyrus (R60, **L65)	Middle occipital gyrus (R144)
Insula (R53, L74)	Precuneus (R61)
Putamen (R200, L*123)	Paracentral Lobule (R86)
Pallidum (R58)	Caudate (R197)
Fusiform (R117, L306)	Lingual (L114)
Inferior temporal gyrus (R219)	Inferior frontal gyrus, opercular part (L78)
Inferior parietal but supramarginal and angular gyri (R87)	Middle temporal gyrus (L*394)
Angular (R433)	Inferior frontal gyrus, orbital part (L61)
Inferior frontal gyrus, triangular part (R180, L*79)	
Inferior occipital gyrus (R269, L133)	
Superior frontal gyrus (R71)	
Supplementary Motor Area (R117)	
Postcentral gyrus (R104, L183)	
Superior frontal gyrus, medial (R188, L159)	
Anterior cingulate and paracingulate gyri (R428, L88)	

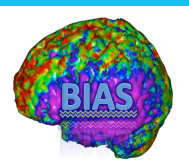


ADNI





Models for Candidate ROIs Whole Genome-wise Analysis



High Dimensional Regression Model

Data $\{(Y_i, X_i) : i = 1, \dots, n\}$

$$Y_i = \{y_i(v) : v \in V\}$$

$$\{X_i(g) : g \in G_0\}$$

Phenotype

Y

$$n \times p_y$$

Genotype

X

$$n \times p_x$$

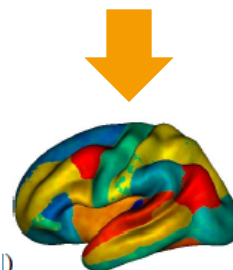
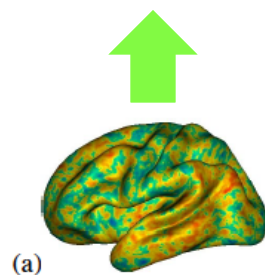
B

$$p_x \times p_y$$

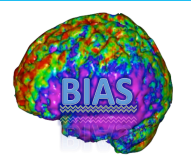
Error

E

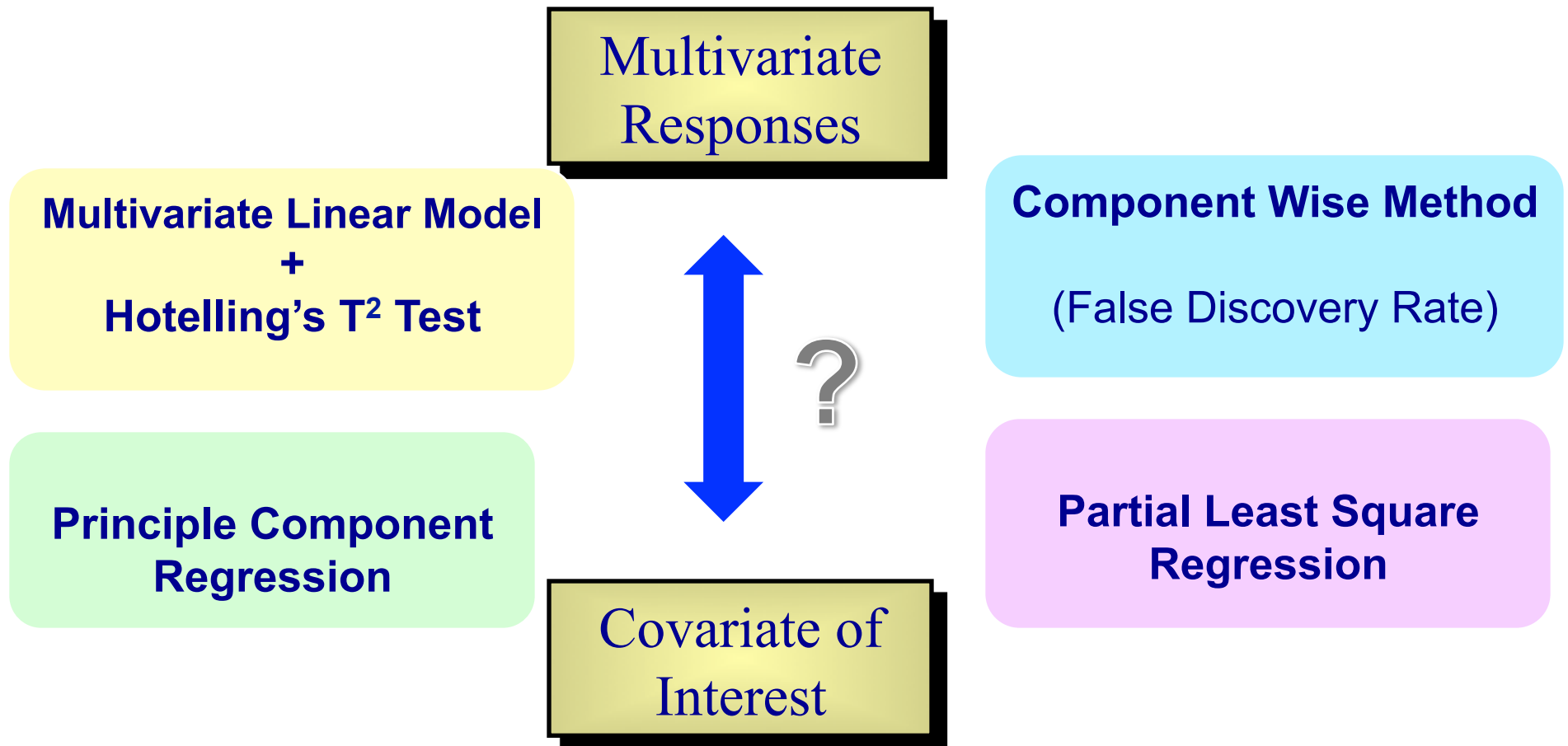
$$n \times p_y$$

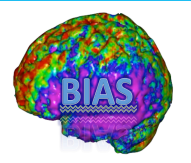


$$p_x \ll n < p_y$$



Current Approaches





Limitations

**Multivariate Linear Model
+
Hotelling's T^2 Test**



$N > p$
~~Hotelling T^2 Test~~

**Principle Component
Regression**



Low Statistical Power
↓
Larger N

Component Wise Method

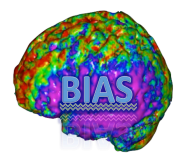


Inflated Type I Error Rate
↓
Wrong Conclusions

**Partial Least Square
Regression**



No p-values!!



Sparse Projection Regression Model

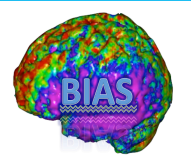
- Multivariate regression with a high-dimensional responses and a multivariate covariate of interest
- Consider a Multivariate Linear Model (MLM):

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \quad \text{or} \quad \mathbf{y}_i = \mathbf{B}^T \mathbf{x}_i + \mathbf{e}_i$$

- We are interested in the hypothesis testing problem:

$$H_0 : \mathbf{CB} = \mathbf{B}_0 \quad \text{v.s.} \quad H_1 : \mathbf{CB} \neq \mathbf{B}_0$$

- Diverging q , fixed p case
 - High-dimension two sample test
 - Imaging genetics association study



Sparse Projection Regression Model

- Let $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_k]$, then a projection regression model is given by:

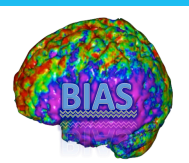
$$\mathbf{W}^T y_i = (\mathbf{B}\mathbf{W})^T \mathbf{x}_i + \mathbf{W}^T \mathbf{e}_i = \beta_{\mathbf{w}}^T \mathbf{x}_i + \epsilon_i$$

- Hypothesis problem reduces to:

$$H_{0W} : \mathbf{C}\beta_{\mathbf{w}} = \mathbf{b}_0 \quad \text{v.s.} \quad H_{1W} : \mathbf{C}\beta_{\mathbf{w}} \neq \mathbf{b}_0$$

$$\text{where } \mathbf{C}\beta_{\mathbf{w}} = \mathbf{C}\mathbf{B}\mathbf{W} \text{ and } \mathbf{b}_0 = \mathbf{B}_0\mathbf{W}$$

- How to determine an 'optimal' $\mathbf{W}$?



Sparse Projection Regression Model

- We show that this is achieved by optimizing the following generalized heritability ratio (GHR):

$$\text{GHR}(\mathbf{w}; \mathbf{C}) = \frac{\mathbf{w}^T (\tilde{\mathbf{B}}_1 - \mathbf{B}_0)^T S_{\tilde{\chi}_1} (\tilde{\mathbf{B}}_1 - \mathbf{B}_0) \mathbf{w}}{\mathbf{w}^T \Sigma_R \mathbf{w}} = \frac{\mathbf{w}^T \Sigma_C \mathbf{w}}{\mathbf{w}^T \Sigma_R \mathbf{w}}$$

- High Dimensional Setting
- noise accumulation
 - ill-conditioned sample covariance estimator: $\hat{\Sigma}_R$
- Sparse Projection Regression Model is proposed as following:

$$\text{argmax} \left\{ \frac{\mathbf{w}^T \hat{\Sigma}_C \mathbf{w}}{\mathbf{w}^T \hat{\Sigma}_R \mathbf{w}} \right\} \quad \text{s.t.} \quad \|\mathbf{w}\|_1 \leq t$$

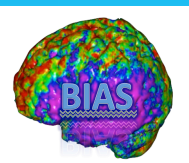


Sparse Projection Regression Model

- After estimating $\mathbf{W}$, we can calculate a $k \times k$ matrix as:

$$T_n = (\mathbf{C}\hat{\beta}_{\mathbf{w}} - \mathbf{b}_0)^T \Sigma_{\tilde{\Omega}}^{-1} (\mathbf{C}\hat{\beta}_{\mathbf{w}} - \mathbf{b}_0)$$

- Test statistic: $\text{Tr}_n = \text{trace}(T_n)$
- Wild bootstrap
 - Fit MLM under the null hypothesis to calculate the estimated multivariate regression coefficient, denoted by $\hat{\mathbf{B}}_0$, residuals $\hat{\mathbf{e}}_i = \mathbf{y}_i - \hat{\mathbf{B}}_0^T \mathbf{x}_i$.
 - Generate G bootstrap samples $\mathbf{z}_i^{(g)} = (\hat{\mathbf{B}}_0)^T \mathbf{x}_i + \eta_i^{(g)} \hat{\mathbf{e}}_i$.
 - Repeat the estimation procedure for estimating the optimal weights and the calculation of the test statistic $\text{Tr}_n^{(g)}$.
 - p -value of Tr_n is computed as $\frac{1}{G} \sum_{g=1}^G \mathbf{1}(\text{Tr}_n^{(g)} \geq \text{Tr}_n)$.



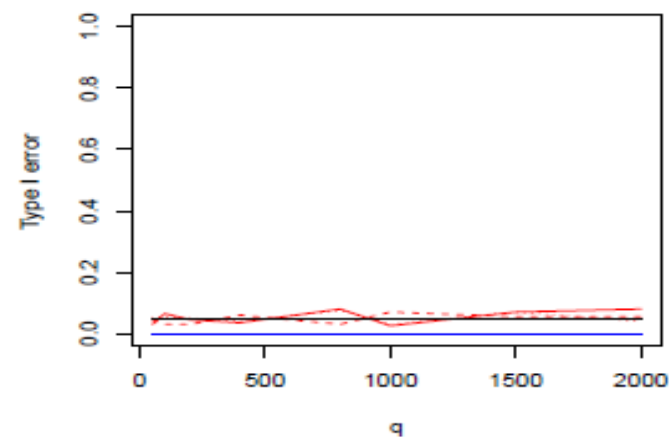
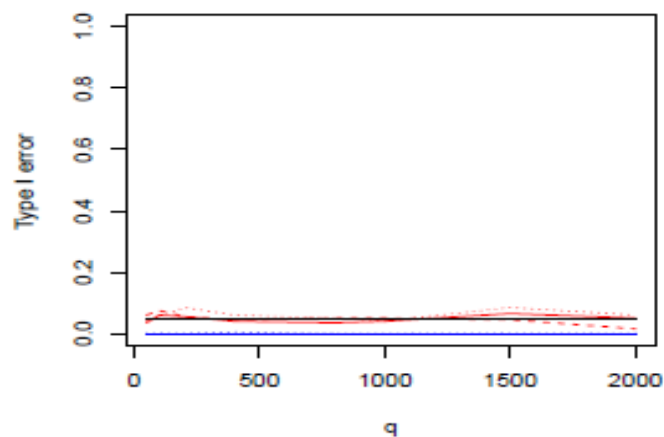
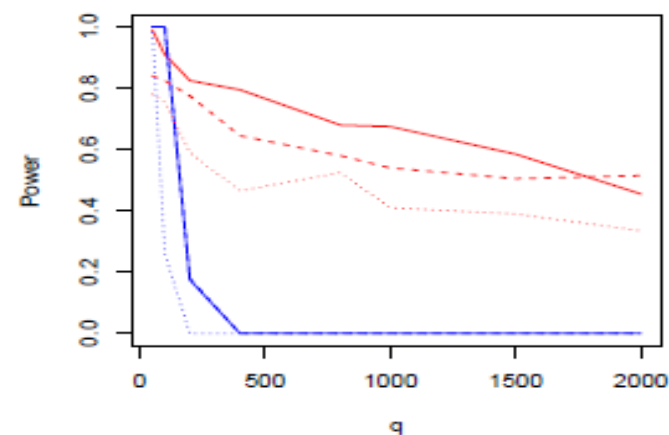
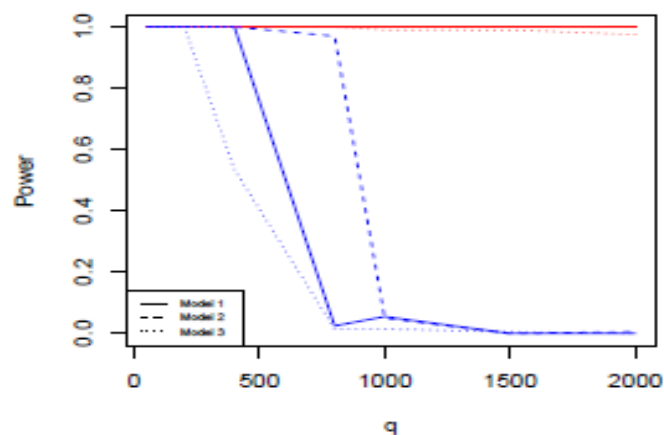
Simulation

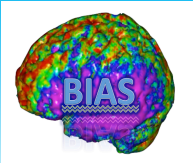
Numerical Example: High Dimensional Two Sample Test

- $\{\mathbf{y}_1, \dots, \mathbf{y}_{n_1}\}$ and $\{\mathbf{y}_{n_1+1}, \dots, \mathbf{y}_n\} \subset R^q$ from $N(\beta_1, \Sigma_R)$ and $N(\beta_2, \Sigma_R)$, respectively.
- We set: $n = 2n_1 = 100$ and q is 50, 100, 200, 400, 800, 1000, 1500, and 2000, respectively.
- $H_0 : \beta_1 = \beta_2$ against $H_1 : \beta_1 \neq \beta_2$
- Can be formulated by a regression model with $\mathbf{B}^T = [\beta_1, \beta_2]$ and $\mathbf{C} = (1, -1)$.
- Error covariance matrix $\Sigma_R = \sigma^2(\rho_{j,j'})$:
 - Model 1: is an independent covariance matrix with $(\rho_{jj'}) = \text{diag}(1, \dots, 1)$.
 - Model 2: is a weak correlation matrix with $\rho_{jj'} = \mathbf{1}(j' = j) + 0.3 \times \mathbf{1}(j' \neq j)$.
 - Model 3: is a strong correlation covariance matrix with $\rho_{jj'} = 0.8^{|j'-j|}$.



Simulation



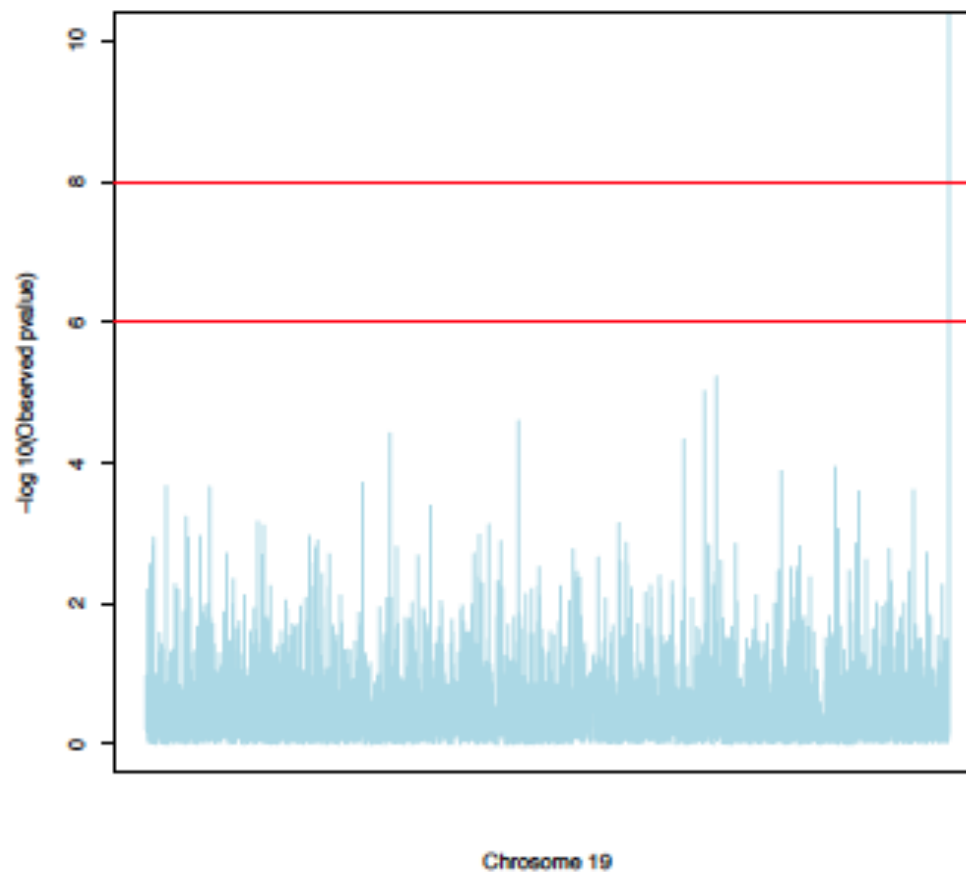


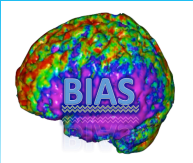
Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Analysis

- Problem of interest: to perform a genome-wide search for establishing the association between the 10,479 SNPs collected on chromosome 19 and the brain volume of 93 regions of interest (ROIs).
- 93 ROIs~intercept, a specific SNP, age, gender, whole brain volume, the top 5 principal components
- Several important findings:
 - The ApoE allele was identified as the top one significant covariate with $-\log_{10}(p) \sim 15$.
 - The rs207650 on the TOMM40 gene is found to be one of the top 3 significant SNPs with $-\log_{10}(p) \sim 5$, Vounou et al. (2012).
 - Able to detect some additional SNPs, such as rs11667587 on the NOVA2 gene.



ADNI



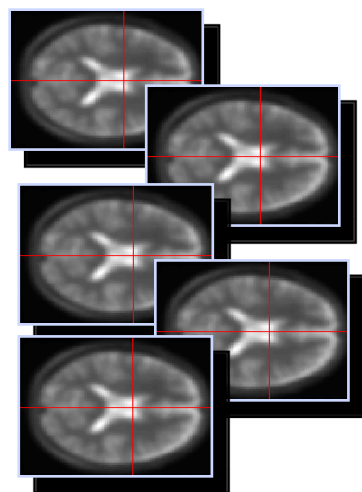


Fast Voxel-wise Genome-wise Analysis

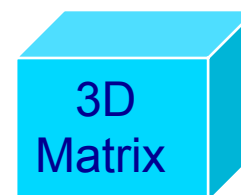


Data Structure

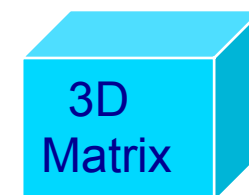
Imaging:



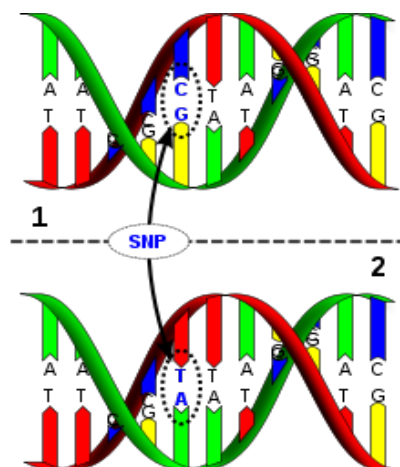
Person No.1 Person No.100



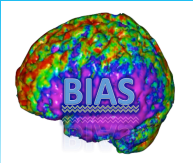
.....



Genetic Variation :



	SNP1	SNP2		SNP2000
Person No. 1	1	2	...	0
.	0	.		1
.	:		.	:
.	.		.	:
Person No. 100	1	0	...	2

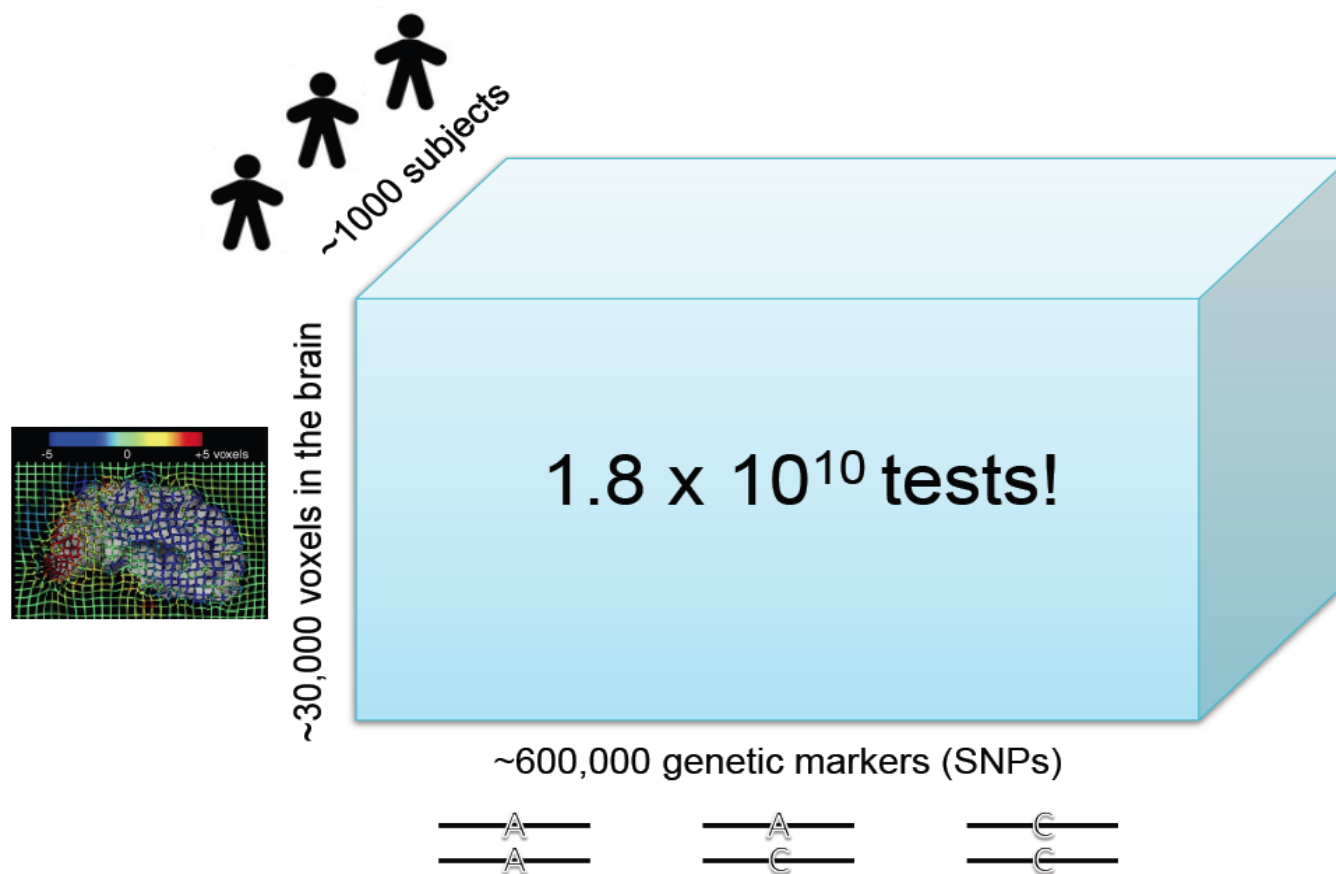


vGWAS

Advantages	Disadvantages
Candidate SNPs allow you to test a specific biological hypothesis	It is highly likely that we don't know the genetic underpinnings of a trait like brain structure so in general don't know the right SNP to pick
Strong hypothesis drives clearly interpretable results	In order to be widely accepted, the variant needs to have strong prior evidence (genome-wide significant in a meta-analysis or have clear function)
Multiple comparisons burden is reduced (one SNP – many voxels)	Unable to search the genome, only characterize the effect of a known variant
Quick way to provide functional relevance to unbiased genome-wide search results	



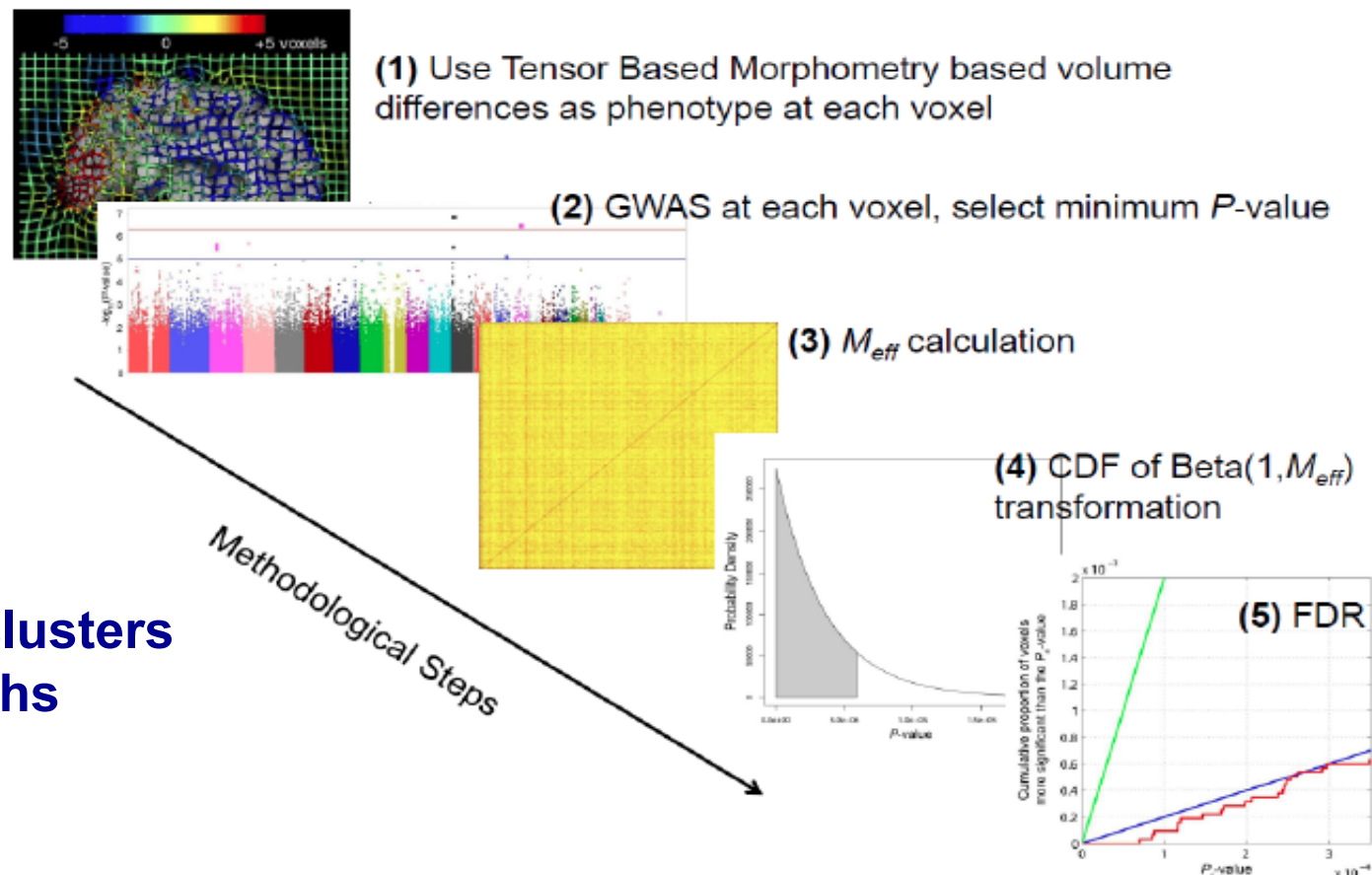
vGWAS



Hibar, et al. HBM 2012

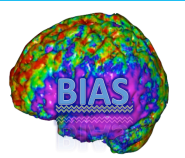


vGWAS

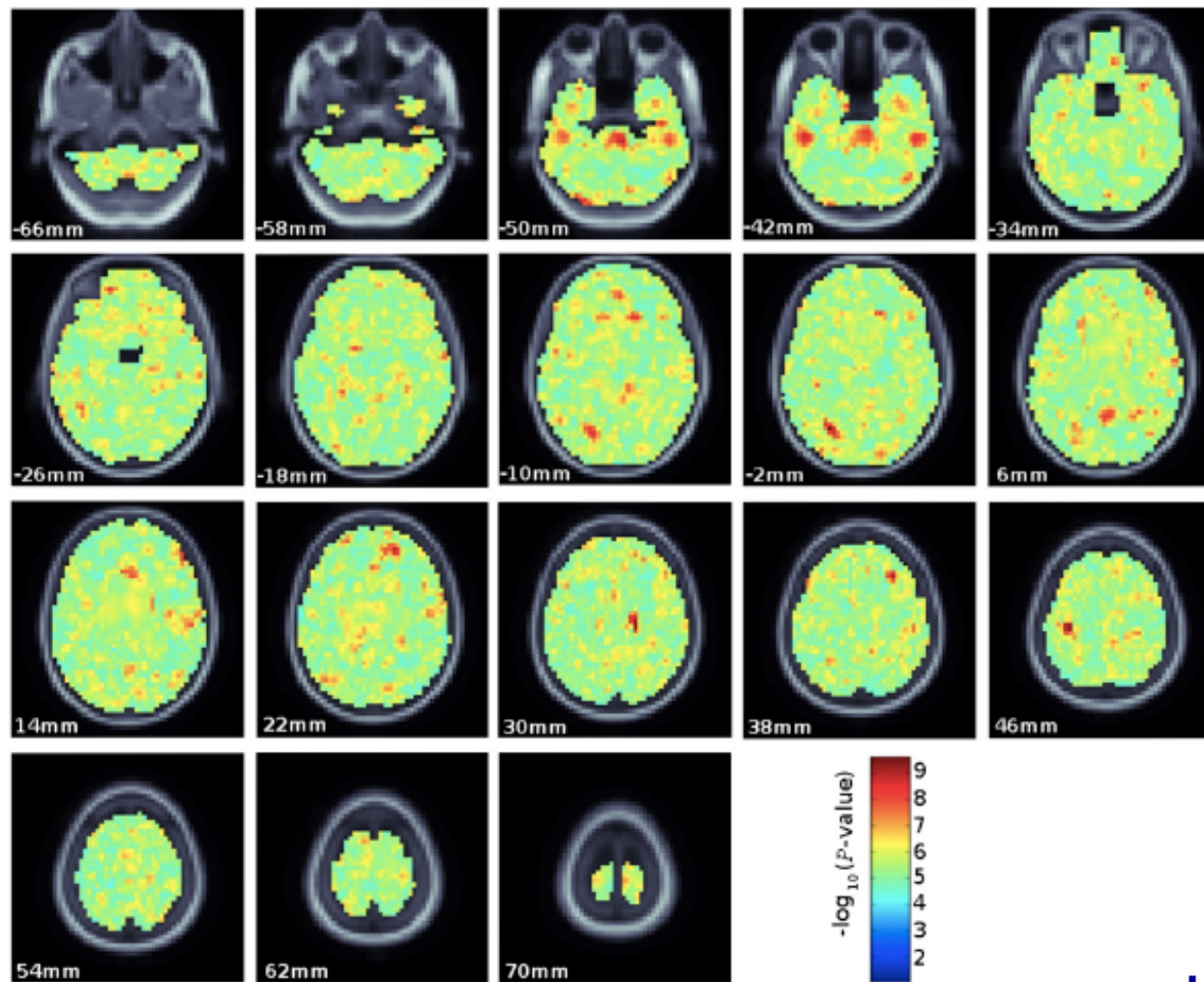


500 clusters
months

Figure 1: Summary of vGWAS method for conducting association and assessing statistical significance.



Raw Minimum p-value at each voxel



Hibar et al. 2011



Most associated genes

Chr	Gene	# of SNPs in gene	# of	Volume (mm ³)	Proportion of brain volume	Clustermax (mm ³)	# of clusters
11	GAB2	20		6336	0.0049	2688	9
11	LRDD	2		8128	0.0063	7872	4
12	PTPRB	17		3200	0.0024	3008	5
9	ZNF462	9		2688	0.0021	2688	1
21	IGSF5	27		16,384	0.013	9344	3
2	SLC25A12	10		1856	0.0014	1792	2
11	MRE11A	9		9344	0.0072	9216	3
19	SLC8A2	11		5632	0.0043	5376	3
15	CHRM5	3		1280	0.00099	1216	2
18	SPIRE1	19		6016	0.0046	3072	12
3	C3orf64	9		4352	0.0034	2112	4
21	S100B	1		9344	0.0072	6656	7
1	CRCT1	1		4096	0.0032	3456	4
19	ZNF626	6		2560	0.002	2112	3
1	ELK4	1		4032	0.0031	2688	4
11	RSF1	8		768	0.00059	768	1
20	WFDC11	2		1280	0.00099	512	5
6	SCML4	27		3328	0.0026	3328	1
12	ERP27	8		2624	0.002	2176	2

AD Risk Gene
(Reiman et al.,
2007)



A Heteroscedastic Linear Model

$$y_i(v) = \mathbf{x}_i^T \boldsymbol{\beta}(v) + \mathbf{z}_i(c)^T \boldsymbol{\gamma}(c, v) + e_i(v) \quad \text{for } i = 1, \dots, n$$

where $\boldsymbol{\beta}(v) = (\beta_1(v), \dots, \beta_K(v))^T$ is a $K \times 1$ vector associated with non-genetic predictors, and $\boldsymbol{\gamma}(c, v) = (\gamma_1(c, v), \dots, \gamma_L(c, v))^T$ is an $L \times 1$ vector of genetic fixed effects (e.g., additive or dominant).

Moreover, $e_i(v)$ are measurement errors with zero mean and $\mathbf{e}_i = \{e_i(v) : v \in V\}$ are independent across i .



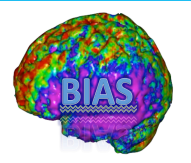
A Heteroscedastic Linear Model

We need to test:

$H_0(c, v) : \gamma(c, v) = 0$ versus $H_1(c, v) : \gamma(c, v) \neq 0$ for each (c, v)

We calculate a Wald-type statistic as:

$$\begin{aligned} W(c, v) &= \tilde{\gamma}(c, v)^T \left\{ \text{Cov}(\tilde{\gamma}(c, v)) \right\}^{-1} \tilde{\gamma}(c, v) \\ &= \text{tr} \left\{ \left\{ \mathbf{Z}_c^T (\mathbf{I}_n - \mathbf{P}_X) \mathbf{Z}_c \right\}^{-1} \mathbf{Z}_c^T (\mathbf{I}_n - \mathbf{P}_X) \sigma_e^{-2}(c, v) \mathbf{Y}(v) \mathbf{Y}(v)^T (\mathbf{I}_n - \mathbf{P}_X) \mathbf{Z}_c \right\} \end{aligned}$$



Big-Data Challenges

10^4

$$X : p_x \times n$$

10^6

10^6

$$B : p_x \times p_y$$

10^7

Memory:

$$O((p_x + p_y)n + p_x p_y)$$

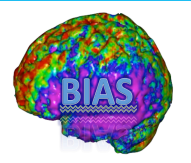
10^4

$$Y : n \times p_y$$

10^7

Computational time:

$$O(p_x p_y n) = O(10^{17})$$



Big-data Challenges

Several big-data challenges arise from the calculation of $W(c, v)$ as follows.

- Calculating $\sigma_e^2(c, v)$ across all (c, v) 's can be computationally intensive.
- Holding all $W(c, v)$ in the computer hard drive requires substantial computer resources.
- Speeding up the calculation of $W(c, v)$.

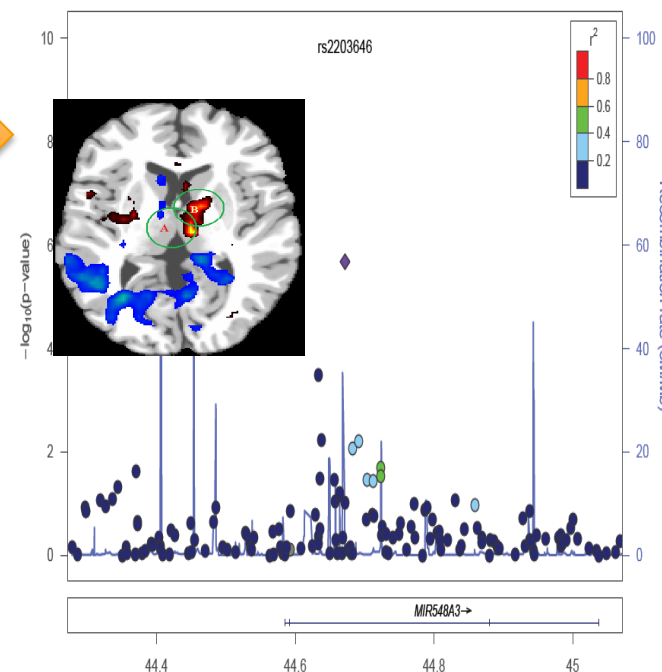
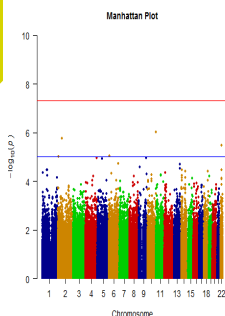
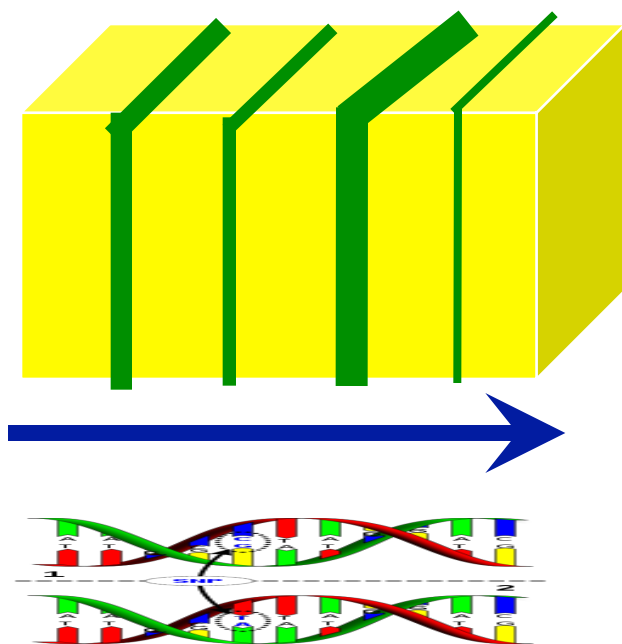


FVGWAS

(I) Spatially
Heteroscedastic
Linear Model

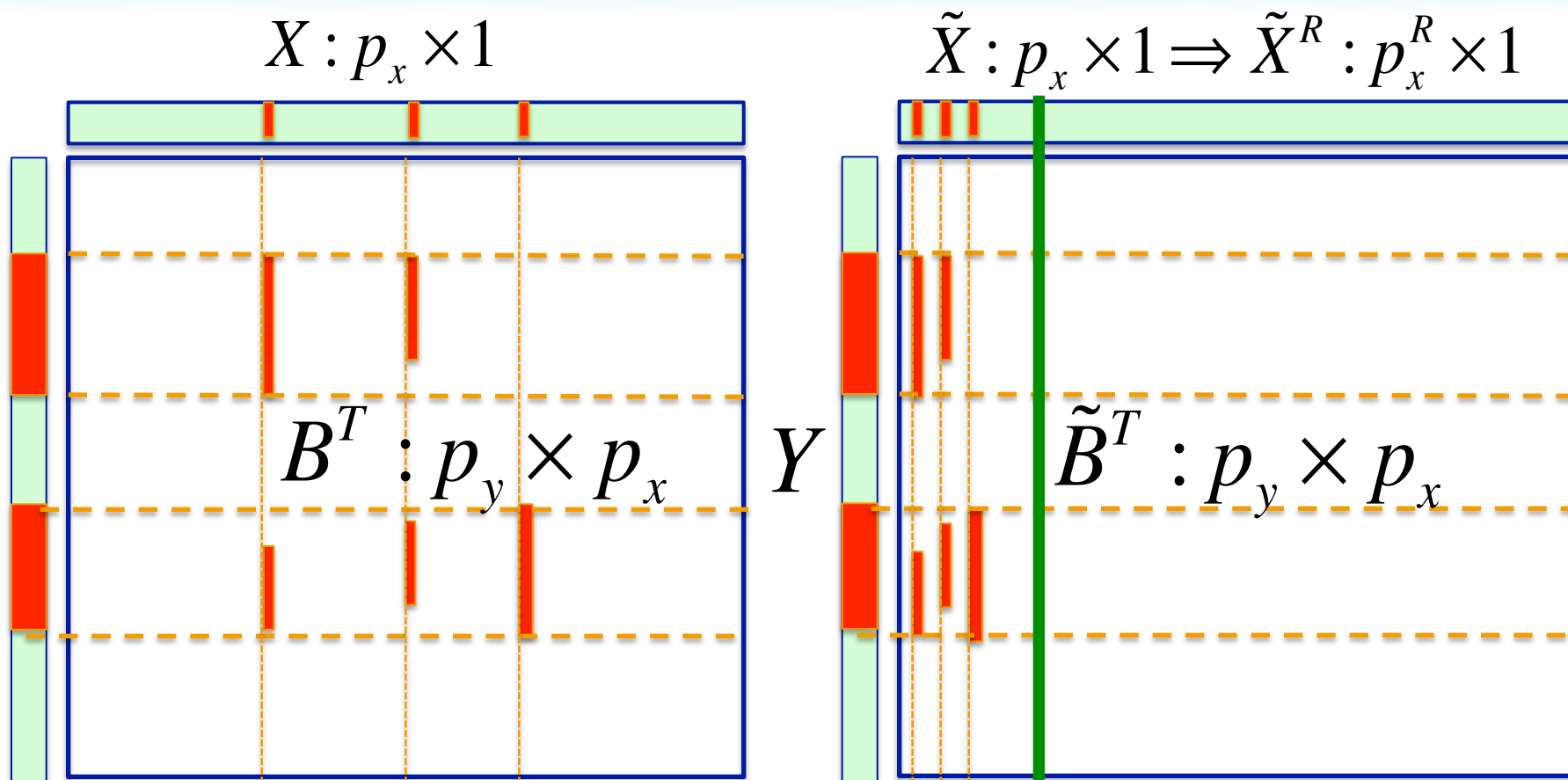
(II) Global Sure
Independence
Screening
Procedure

(III) Detection
Procedure





Key Features



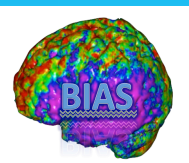
X: Sparsity; Y|X: Clustered ROIs



FVGAWS

To solve these computational bottlenecks, we propose two solutions as follows.

- Calculate $\sigma_e^2(c, v)$ under the null hypothesis $H_0(c, v)$ for each v and c .
- Develop a GSIS procedure to eliminate many ‘noisy’ loci based on a global Wald-type statistic.
- ❖ By using these two solutions, we are able to reduce the computational complexity from $O(p_x p_y n)$ to $O((p_x + p_y)n^2)$.



A Global Sure Independence Screening

The global Wald-type statistic at locus c is defined as

$$W(c) = N_V^{-1} \text{tr} \left\{ \left\{ \mathbf{Z}_c^T (\mathbf{I}_n - \mathbf{P}_X) \mathbf{Z}_c \right\}^{-1} \mathbf{Z}_c^T (\mathbf{I}_n - \mathbf{P}_X) \left\{ \sum_{v \in V} \hat{\sigma}_e(v)^{-2} \mathbf{Y}(v) \mathbf{Y}(v)^T \right\} (\mathbf{I}_n - \mathbf{P}_X) \mathbf{Z}_c \right\}$$

$$\left\{ \sum_{v \in V} \hat{\sigma}_e(v)^{-2} \mathbf{Y}(v) \mathbf{Y}(v)^T \right\}$$

is independent of c ,

the complexity of computing $\{W(c)\}$ is $O((p_x + p_y)n^2)$

the complexity of computing $\{W(c, v)\}$ is $O(p_x p_y n)$

$$\begin{aligned} p_x p_y / \left\{ (p_x + p_y) n \right\} &= 10^{6+7} / \left\{ (10^7 + 10^6) 10^4 \right\} = 10^2 \\ &= 10^{6+7} / \left\{ (10^7 + 10^6) 10^3 \right\} = 10^3 \end{aligned}$$



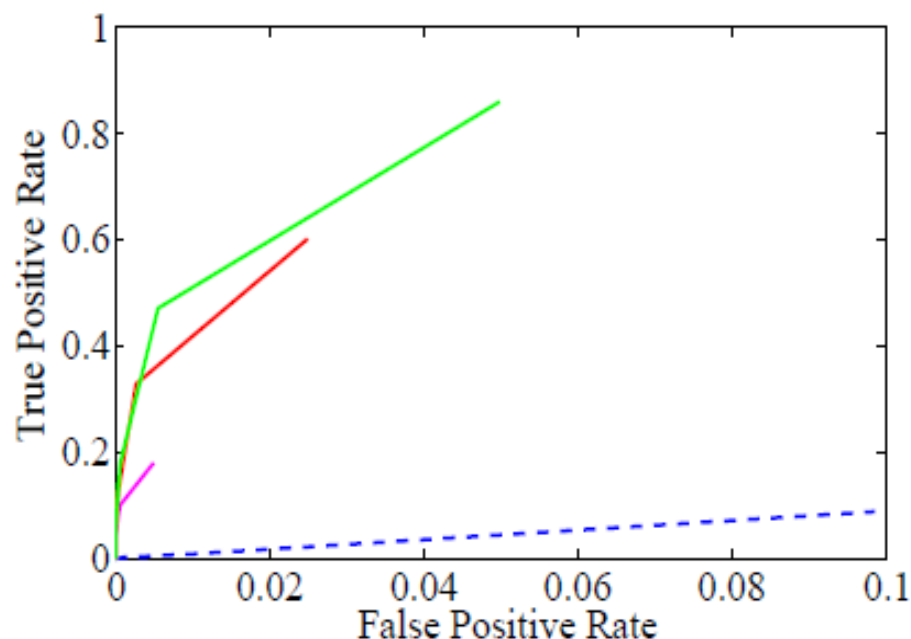
Simulation Studies

Simulation settings: the dark, gray, and white regions in the figure, respectively, represent background, brain region, and the effected ROI associated with the causal SNPs.

ROI: 10×10



$\gamma_* = 0.005$



$\gamma_* = 0.01$

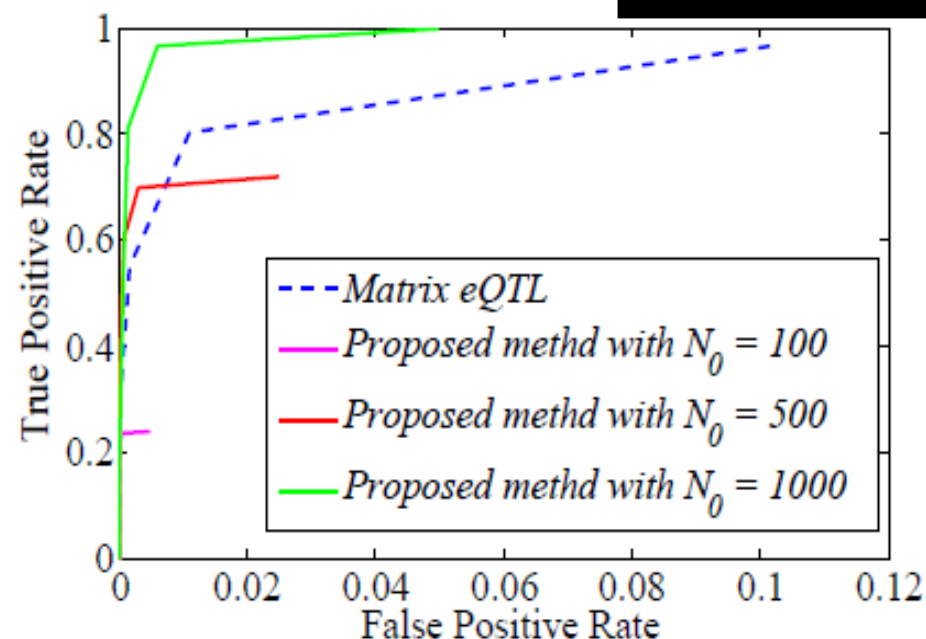


Fig. Simulation results for comparisons between FVGWAS and the Matrix eQTL in identifying significant voxel-SNP pairs.

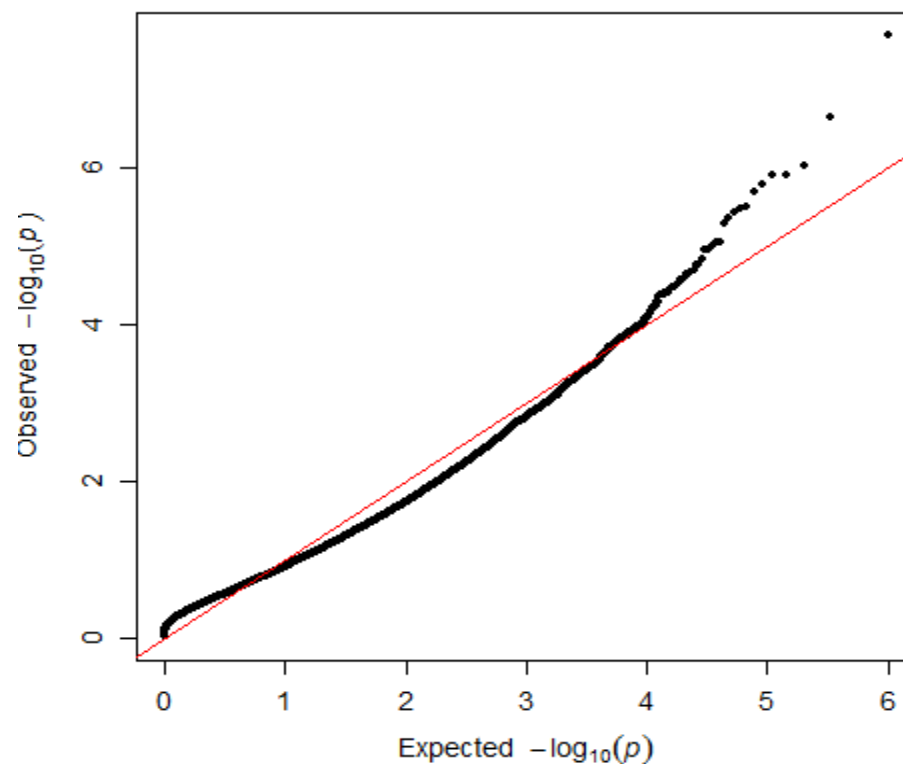
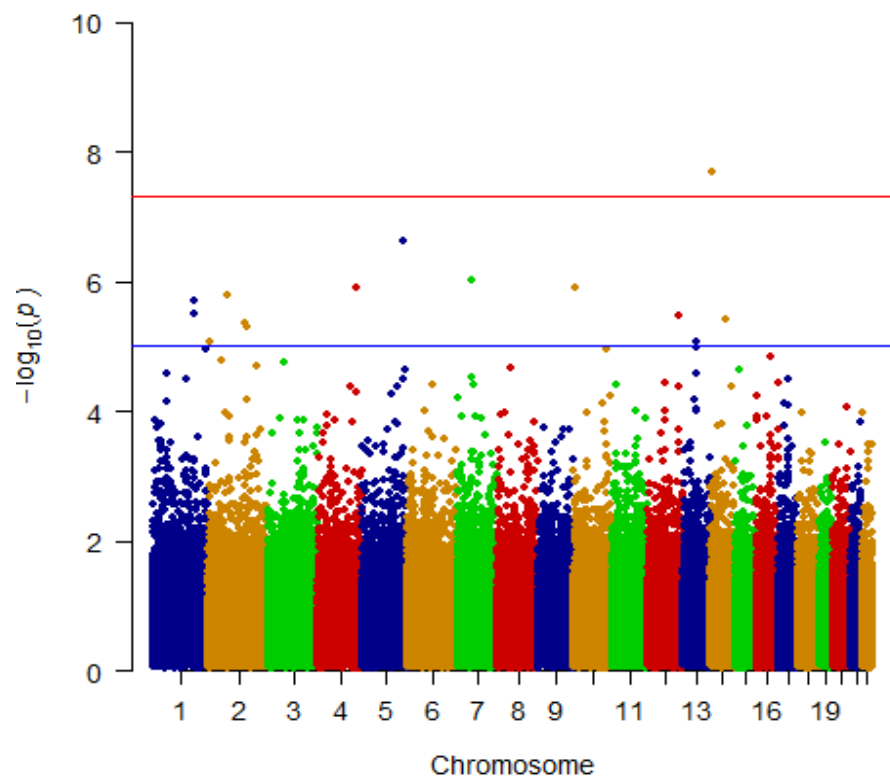


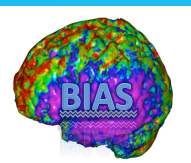
Results

Our computational time

About 33,800 s

Manhattan Plot





ADNI Data Analysis

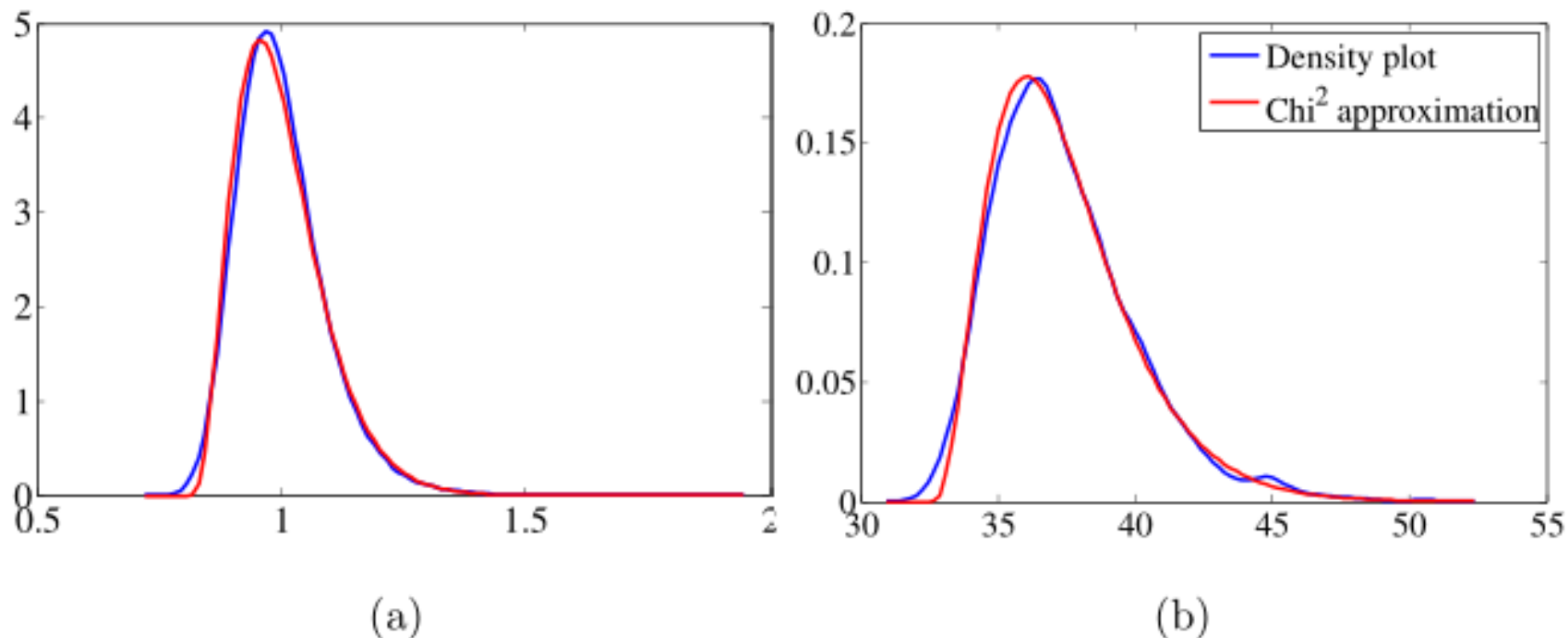
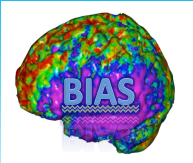


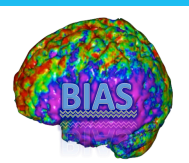
Fig. ADNI whole-brain GWAS: (a) the density plot of and its approximation; (b) the density plot of and its approximation



ADNI Data Analysis

Table. RAVEN map GWAS: significant voxel-SNP pairs at the 0.5 significance level (left) and significant cluster-SNP pairs at the 0.5 significance level (right)

SNP	Number of voxel-SNP pairs	SNP	Number of cluster-SNP pairs	Max cluster	<i>p</i> -value of the max cluster
rs2075650 (TOMM40)	23	rs11815438	1	7906	0.11
rs9490103	4	rs2480271	1	7365	0.23
rs2244634	2	rs7001339	1	6864	0.45



ADNI Data Analysis

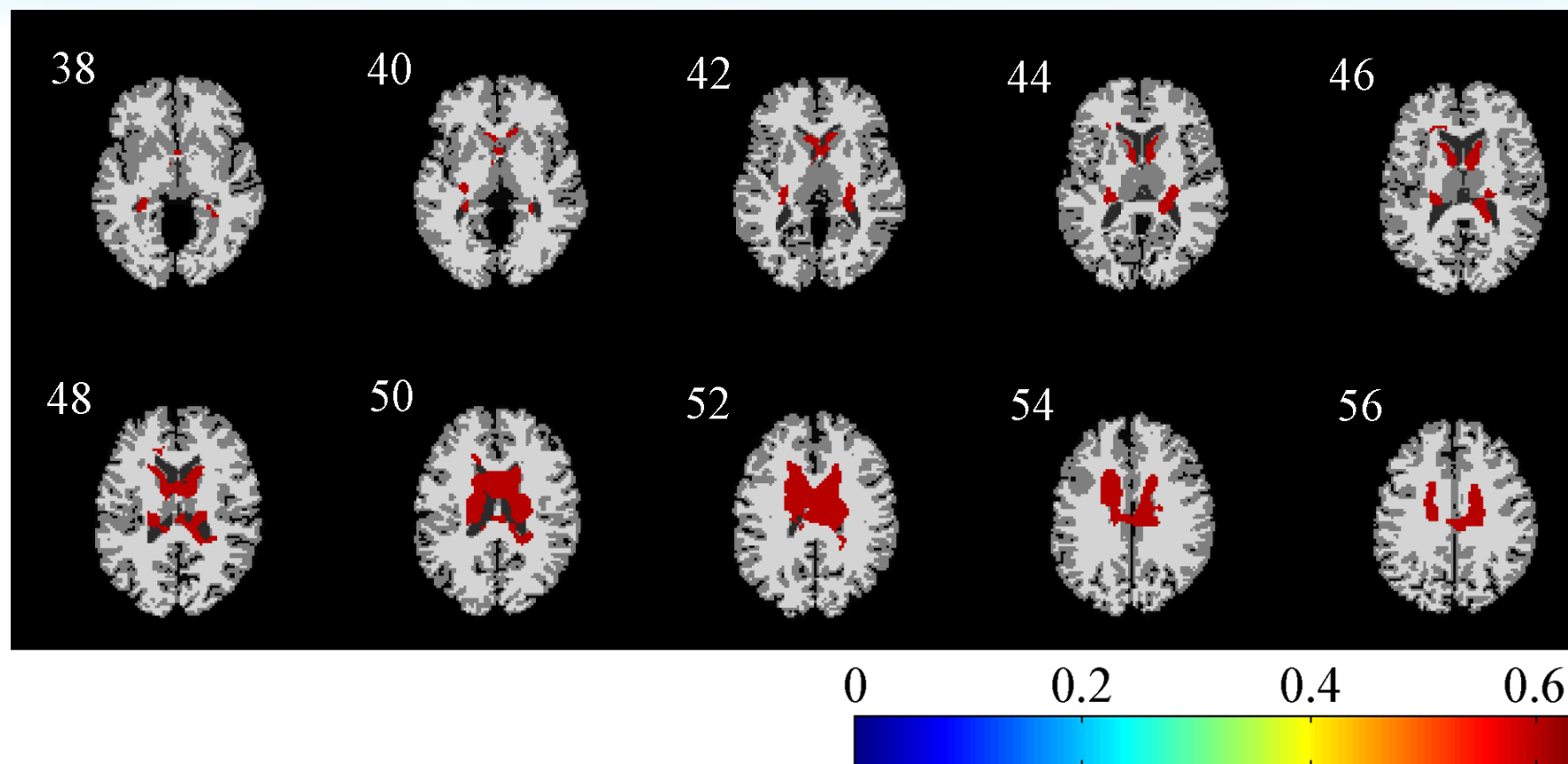
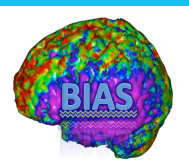


Fig. ADNI whole-brain GWAS: selected slices of $-\log_{10}(p)$ for significant clusters corresponding to a SNP (rs2480271).



Motivating Ex: DTI Fiber Tract Data

Data

- Diffusion properties (e.g., FA, RA)

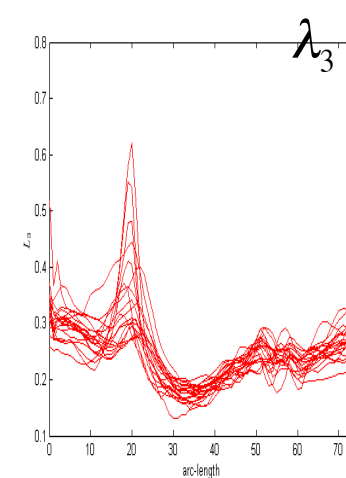
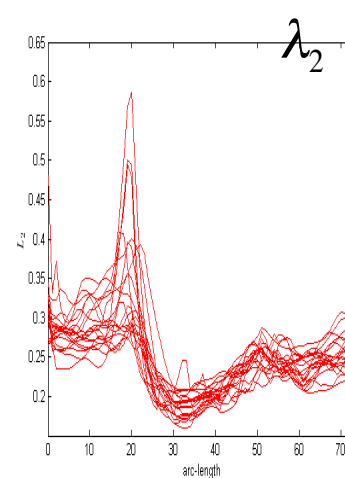
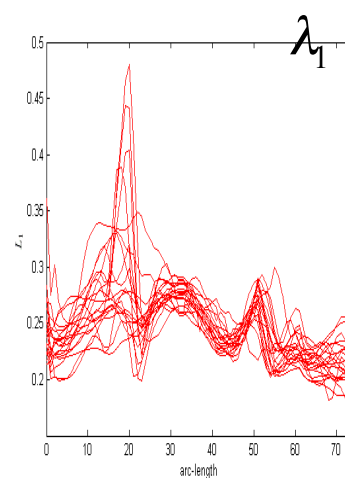
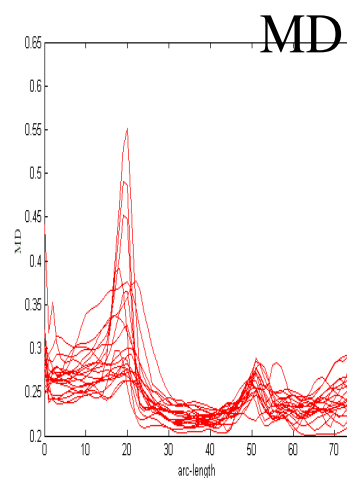
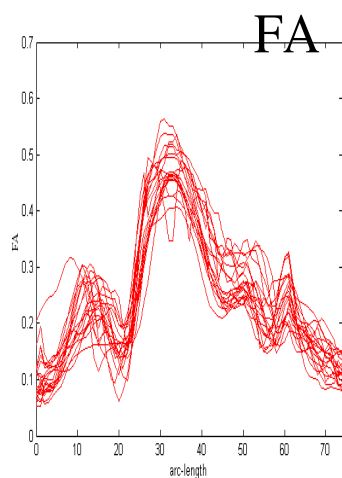
$$Y_i(s_j) = (y_{i,1}(s_j), \dots, y_{i,m}(s_j))^T$$

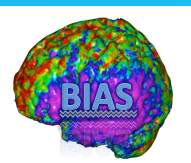
- $\{s_1, \dots, s_{n_G}\}$

Grids

- Covariates (e.g., age, gender, diagnostic)

$$x_1, \dots, x_n$$





MVCM

Decomposition:

$$y_{i,k}(s) = x_i^T B_k(s) + z_i(g)^T \beta_k(s, g) + \eta_{i,k}(s) + \varepsilon_{i,k}(s)$$

Coefficients Long-range Correlation Short-range Correlation

$x_1, \dots, x_n$ $\eta_{i,k}(\bullet) \sim SP(0, \Sigma_\eta)$ $\varepsilon_{i,k}(\bullet) \sim SP(0, \Sigma_\varepsilon),$

Covariance operator: $\Sigma_y(s, s') = \Sigma_\eta(s, s') + \Sigma_\varepsilon(s, s')$

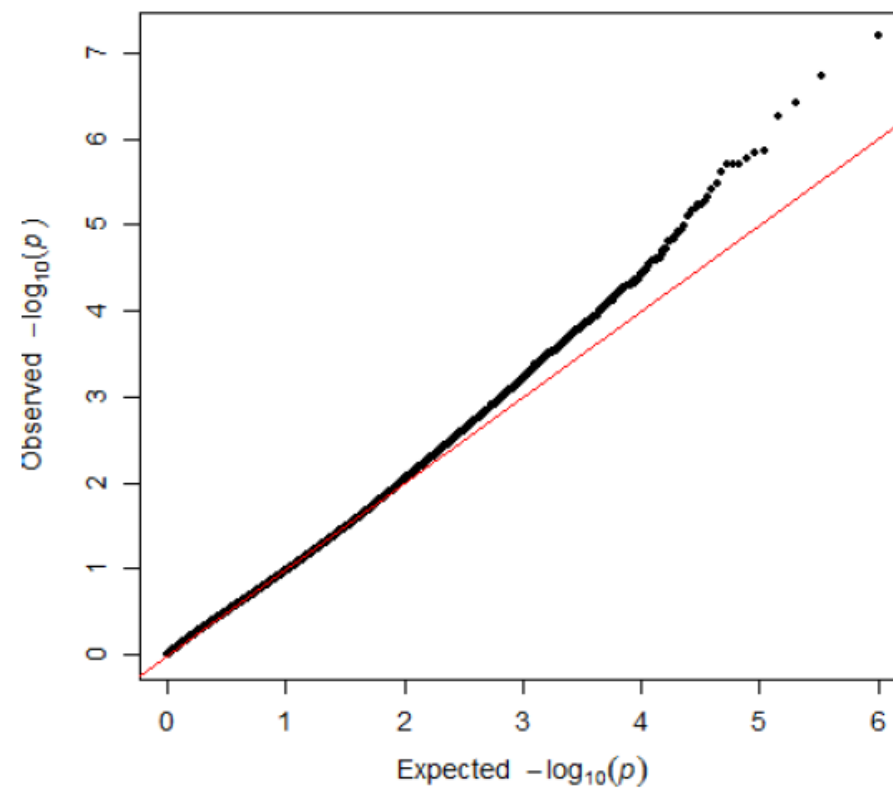
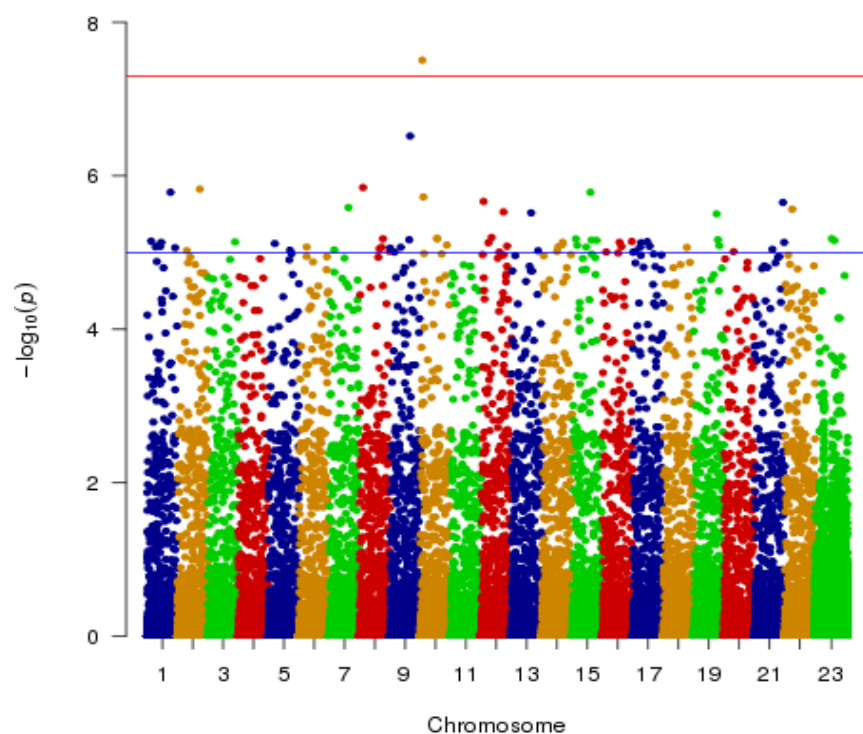
$$\sqrt{n} \{ \text{vec}(\hat{B}(d) - B(d) - 0.5O(H^2)) : d \in D \} \xrightarrow{L} G(0, \Sigma_B(d, d'))$$

Zhu, Li, and Kong (2012). AOS



Tract-GWAS

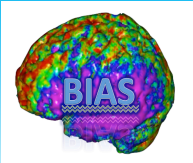
Manhattan Plot





Top SNPs

SNP ID	CHR ID	$-\log_{10}(\text{p-value})$
'rs1384333'	10	7.3057
'rs2805816'	9	6.6177
'rs34608777'	8	6.2466
'2:170144362:T_TTA'	2	6.1259
'rs11263882'	1	5.8854
'rs78723271'	15	5.7848
'rs11580558'	10	5.7236
'rs10841375'	12	5.6652
'rs11910115'	21	5.6525
'rs28839626'	7	5.5871



A Software for FVGWAS

