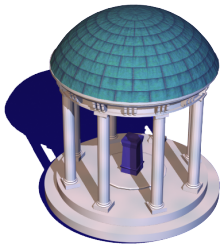
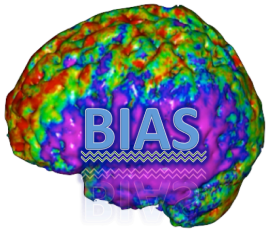
UNC SCHOOL OF PUBLIC HEALTH

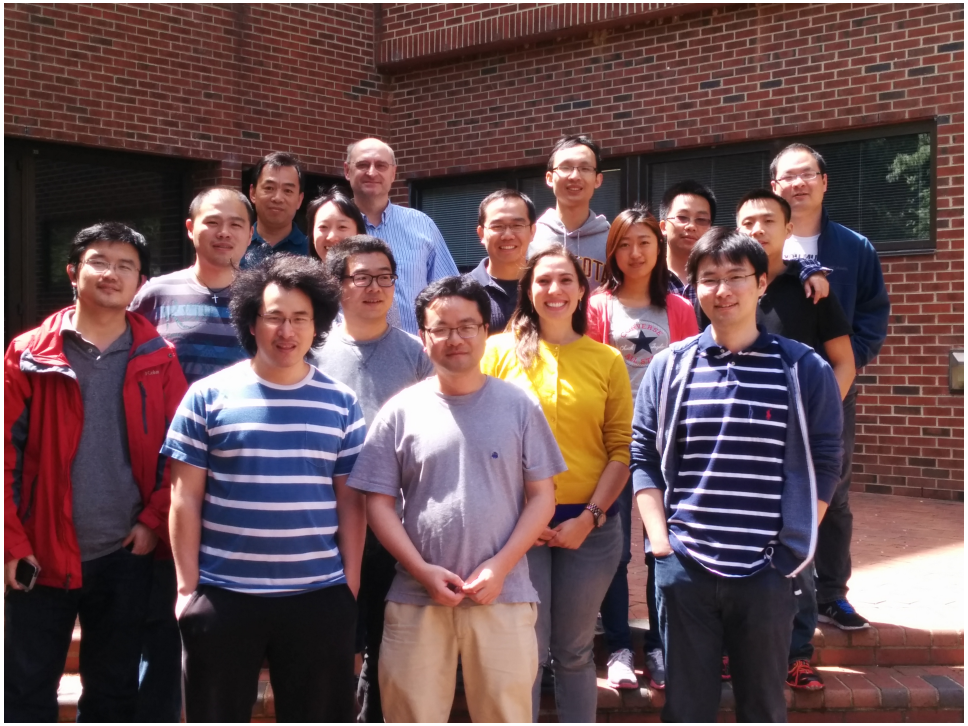


Functional Analysis of Large-scale Neuroimaging and Genetic Data

Hongtu Zhu, Ph.D
Department of Biostatistics† and Biomedical Research Imaging Center‡
The University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599, USA

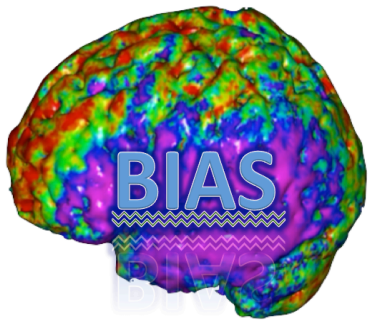


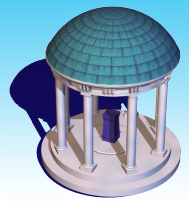
BIAS: Biostatistics and Imaging Analysis Lab



http://www.bios.unc.edu/research/bias/

Statistical Challenges in Neuroimaging Data Analysis

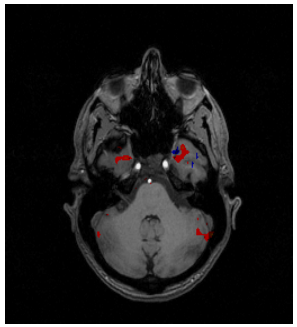




Reading Materials

1. Zhu, H. T., Chen, K. H., Yuan, Y. and Wang, J. L. (2015). Functional Mixed Processes Models for Repeated Functional Data. In submission.
2. Zhu, HT., Fan, J., and Kong, L. (2014). Spatial varying coefficient model and its applications in neuroimaging data with jump discontinuity. *JASA,* 109, 977-990, 2014.
3. J. W. Hyun, Li, Y. M., Gilmore, J., Lu, Z.H., Styner, M., and *Zhu, H.T.* SGPP: Spatial Gaussian Predictive Process Models for Neuroimaging Data. *NeuroImage*, 89, 70–80, 2014.
4. Yuan, Y., Gilmore, J., Geng, X. J., Styner, M., Chen, K. H., Wang, J. L., and *Zhu, H.T.* (2014). Fmem: Functional mixed effects modeling for the analysis of longitudinal white matter tract data. *NeuroImage* 84, 753–764.
7. Yuan, Y., Gilmore, J., Geng, X. J., Styner, M., Chen, K. H., Wang, J. L., and *Zhu, H.T.* (2013). A longitudinal functional analysis framework for analysis of white matter tract statistics. *NeuroImage*, 23:220-31, 2013.
8. Yuan, Y., Zhu, H.T., Styner, M., J. H. Gilmore., and Marron, J. S. (2013). Varying coefficient model for modeling diffusion tensors along white matter bundles. *Annals of Applied Statistics.* 7(1):102-125..
9. Zhu, H.T., Li, R. Z., Kong, L.L. (2012). Multivariate varying coefficient models for functional responses. *Ann. Stat.* 40, 2634-2666.
10. Li, YM, *Zhu HT*, Shen DG, Lin WL, Gilmore J, and Ibrahim JG. (2011). Multiscale adaptive regression models for neuroimaging data. *JRSS, Series B*, 73, 559-578.

Data Analysis



Raw Images

Image Reconstruction

Image Registration

Image Smoothing

Multiple Comparisons

Statistical Modelling

Statistical Analysis

Individual Imaging Analysis

Imaging Construction

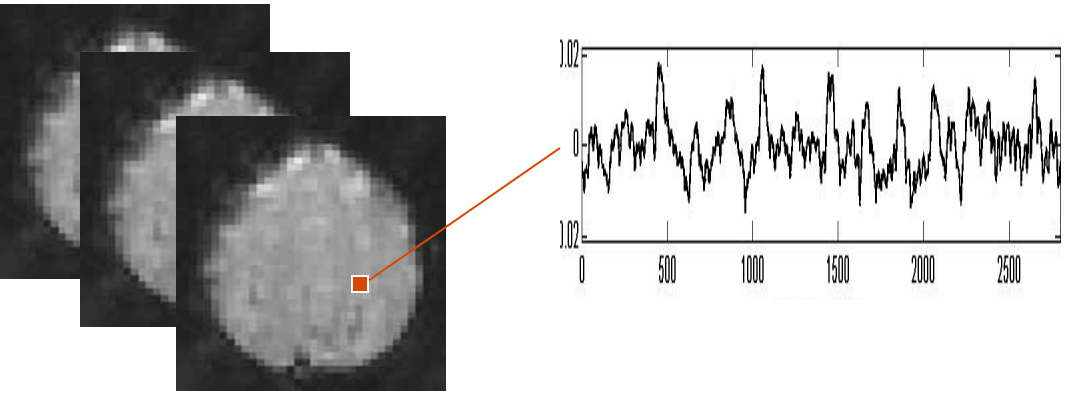
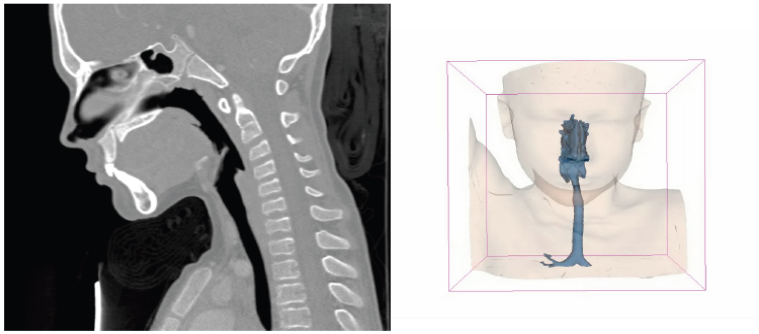


Image Segmentation

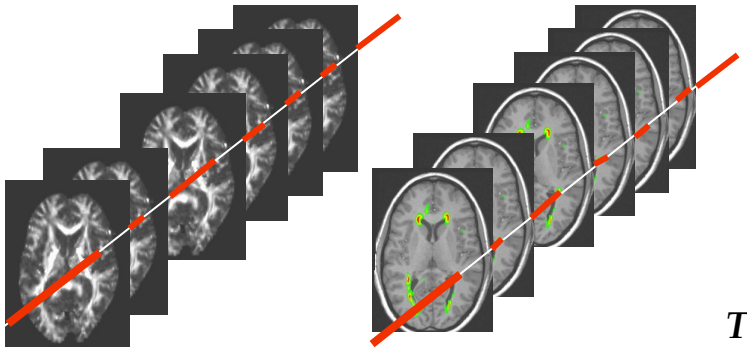
Example: Airway Segmentation from CT

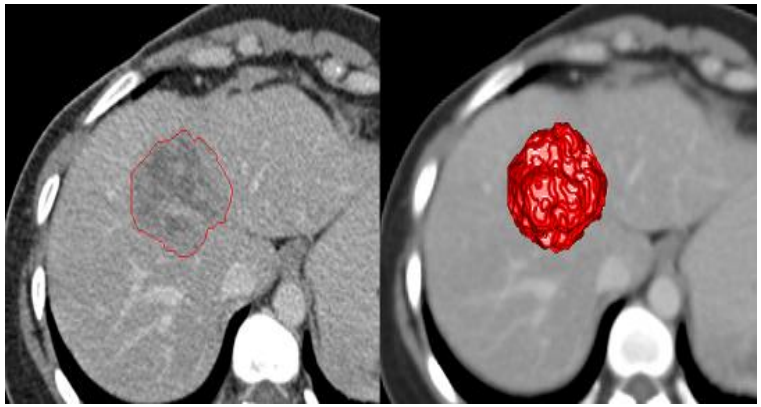


Multimodal Analysis

DTI

FLAIR

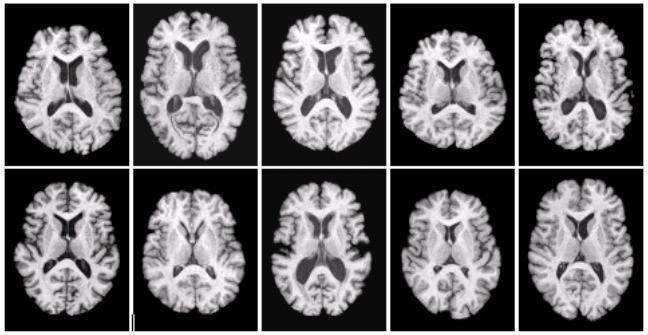


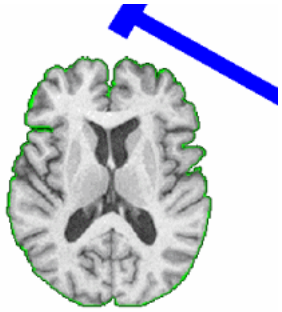


Marc

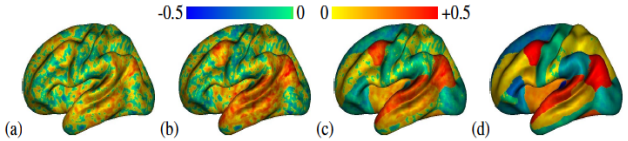
Group Imaging Analysis

Registration



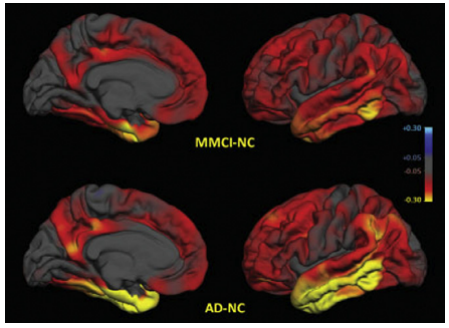


Prediction



NC/Diseased

Group Differences

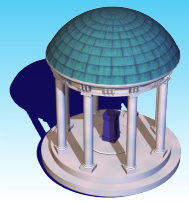


Longitudinal/Family Brain

Hibar, Dinggang, Martin

Imaging Genetics

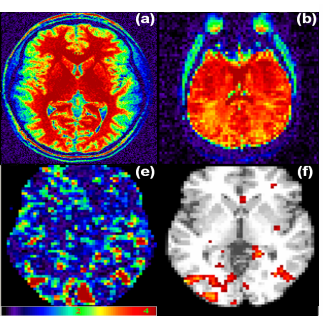
Genetics \ Imaging	Candidate ROI	Many ROI	Voxelwise
Candidate SNP	Imager	Imager	Imager
Candidate Gene	Geneticist		
Genome-wide SNP	Geneticist		
Genome-wide Gene	Geneticist		

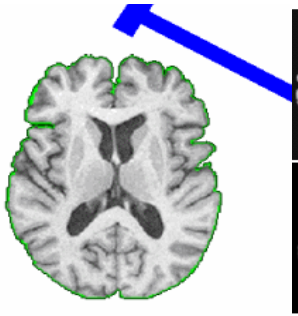
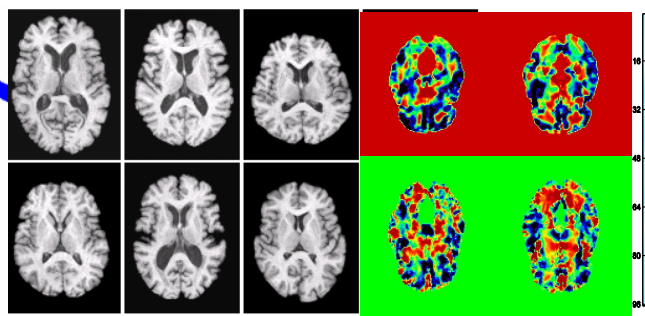
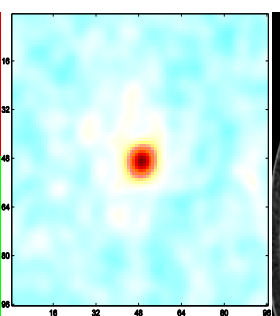
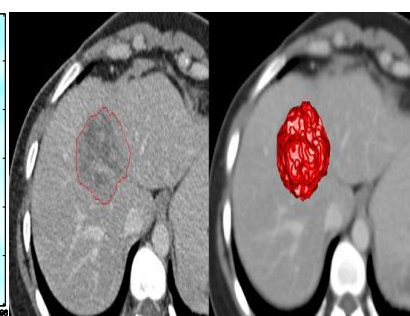


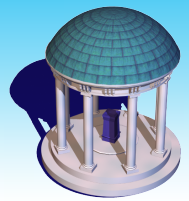
Noisy Imaging Data

Key Features

- Infinite Dimension
- Spatial Smoothness
- Spatial Correlation
- Spatial Heterogeneity





Infinite Dimensional Image

Mathematics.

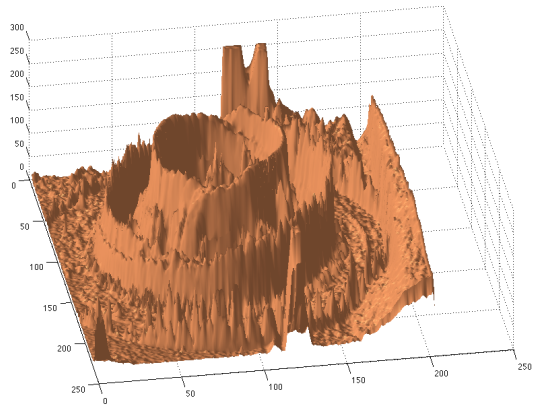
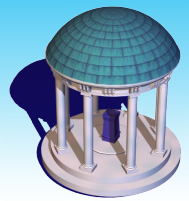


Image is the point or set of points in the range corresponding to a designated point in the domain of a given function.

- Ω is a compact set. $\tilde{x} \in \Omega \subseteq R^k$

$$\Rightarrow \quad f(\tilde{x}) \in M \subseteq R^m \qquad f : \Omega \rightarrow M \subseteq R^m$$

$$\bigstar \quad \int_{\Omega} \| f(\tilde{x}) \|^k \, d\tilde{x} < \infty \quad \text{for some } k>0$$

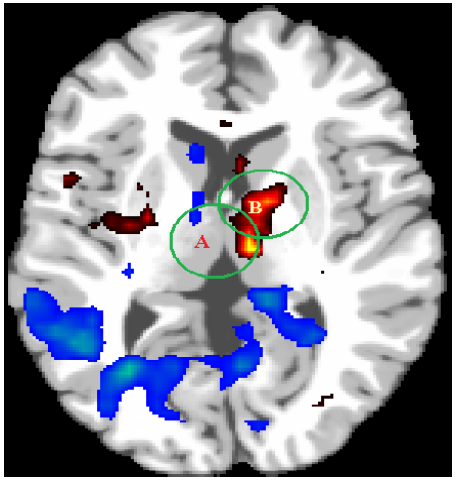


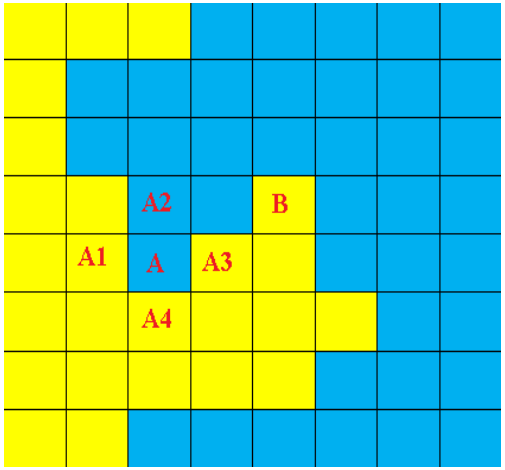
Spatial Smoothness

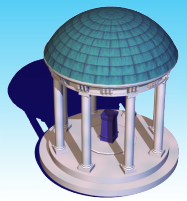
Cartoon Model

$\theta(d) \in R^K$ $\theta_k(d)$

- **Disjoint Partition** $D = \cup_{l=1}^{L} D_l$ and $D_l \cap D_{l'} = \phi$
- **Piecewise Smoothness: Lipschitz condition**
- **Smoothed Boundary**
- **Local Patch**
- **Degree of Jumps**





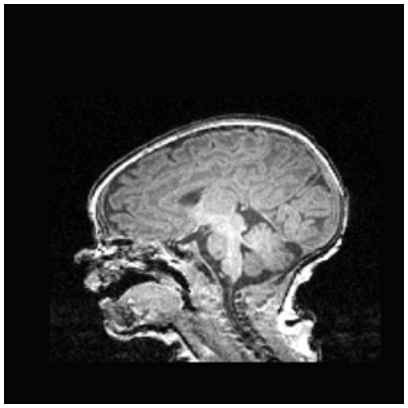


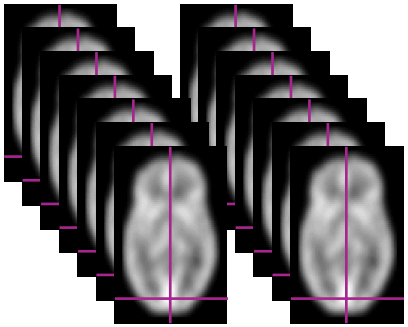
FDA: Functional Data Analysis

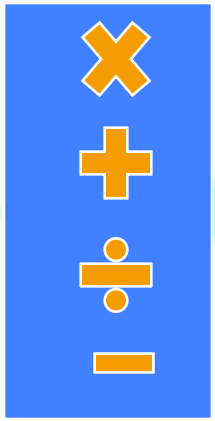
f

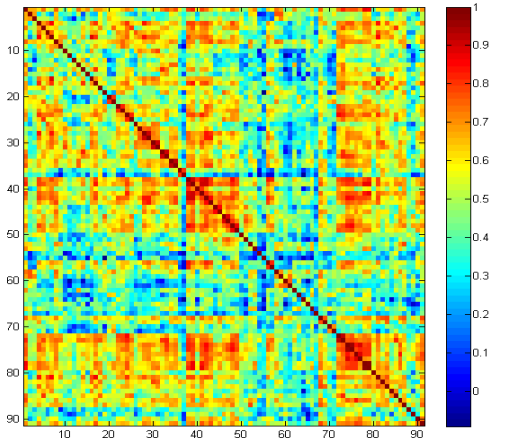
T

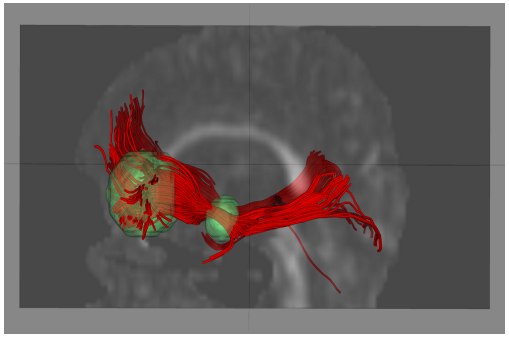
$\hat{F} = T[f]$

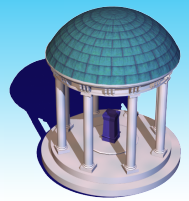












Statistical Analysis

Registration

Images

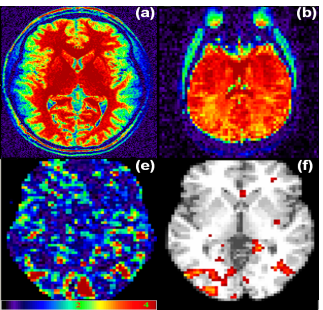
Estimation

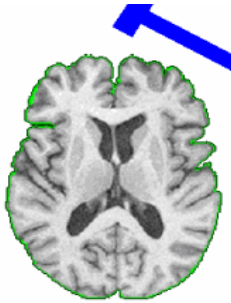
Prediction

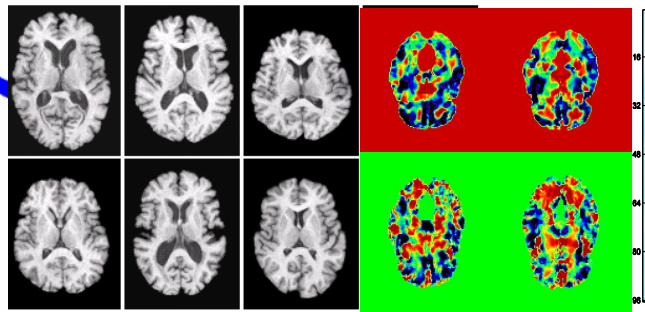
Smoothing

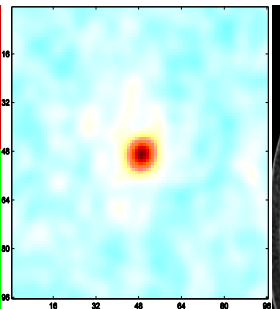
Voxel-wise Statistical Models

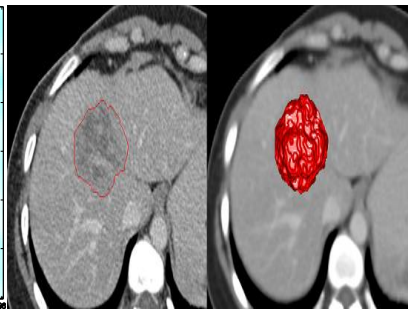
Multiple Comparisons





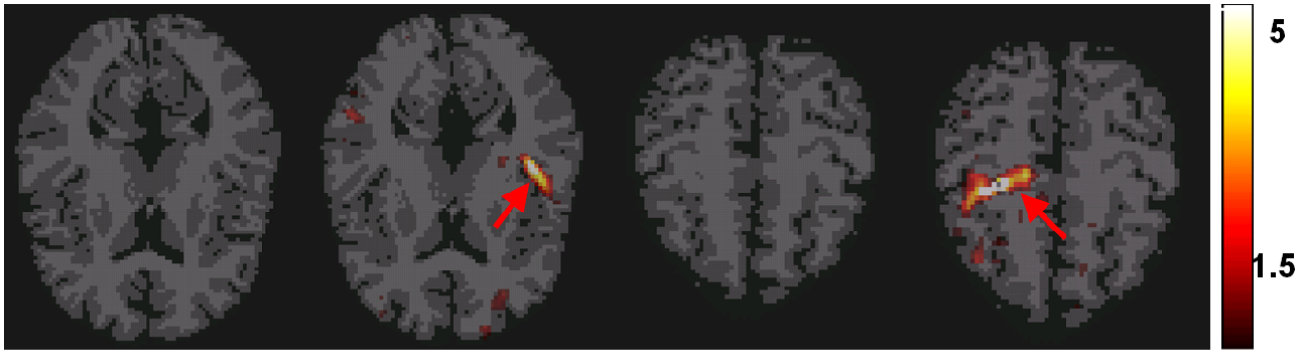




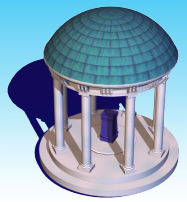


Smoothing Effect

- Smoothing method is independent of data
- Degree of smoothness is arbitrary
- Effect of smoothness is profound
- The relationship between smoothing method and study design is unknown



Jones et al. (2006),
Yue et al. (2010)



Prediction Accuracy: ADNI PET Data

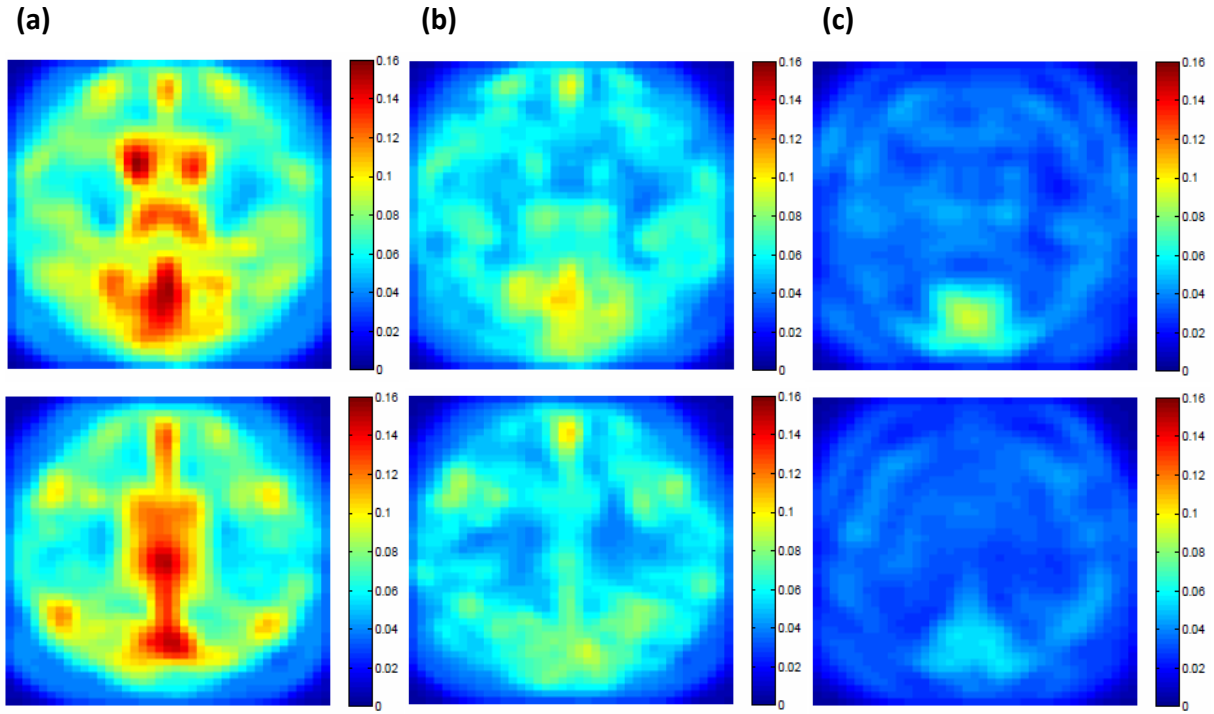
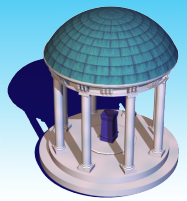


Figure : rtMSPE maps for prediction of ADNI PET images at month 12 for 79 test subjects. Selected slices are shown for (a) Semi-parametric model; (b) Semi-parametric model+FPCA; (c) Semi-parametric model+FPCA+Spatial-temporal model.



FSEM: Functional Structural Equation Models for Twin Functional Data

Reading Materials

Li, YM, John Gilmore, JP Wang, M. Styner, Weili Lin, and *Zhu, HT.* (2012). Two-stage spatial adaptive analysis of twin neuroimaging data. *IEEE Transactions on Medical Imaging.* 31, 1100-12.

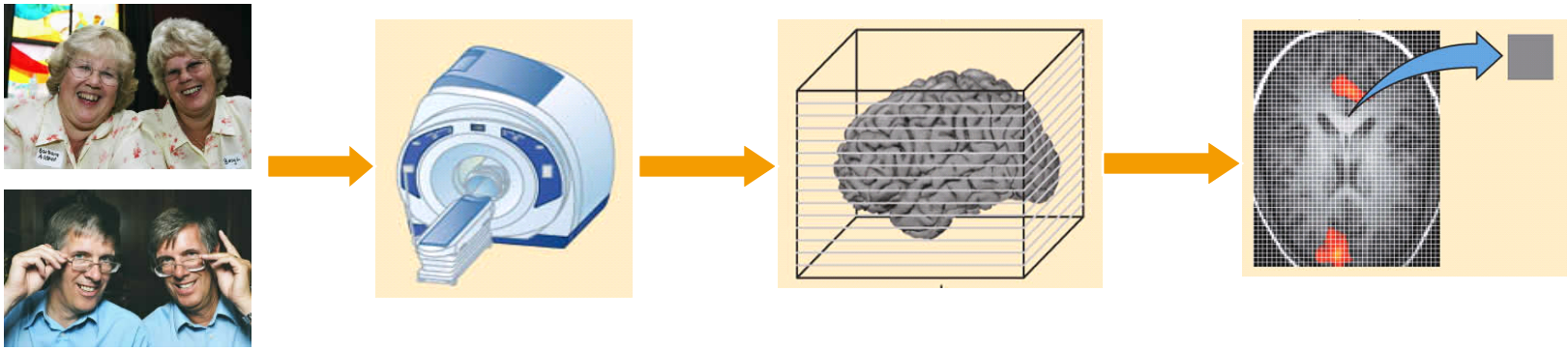
Li, YM, *Zhu HT*, Shen DG, Lin WL, Gilmore J, and Ibrahim JG. (2011). Multiscale adaptive regression models for neuroimaging data. *JRSS, Series B*, 73, 559-578.

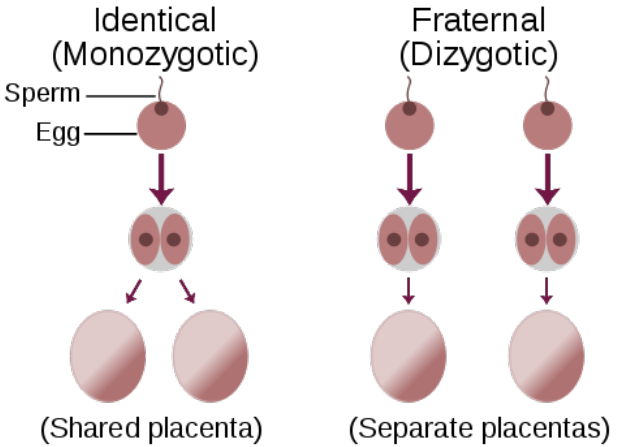
Luo, S., R. Song.,John Gilmore, M. Styner, and *Zhu, HT.* (2015). Functional Structural Equation Models for twin functional data. To be submitted.

S.J. Lee, R. J. Steiner, S. Luo, M. C. Neale, M. Styner, H. Zhu, J.H. Gilmore. (2015). Quantitative tract-based white matter heritability in twin neonates. *NeuroImage*, 111, 123–135.

Video:
http://www.birs.ca/events/2015/5-day-workshops/15w5096/videos/watch/201506301556-Zhu.html

Twin Neuroimaging Data





MZ twins share same genetic material

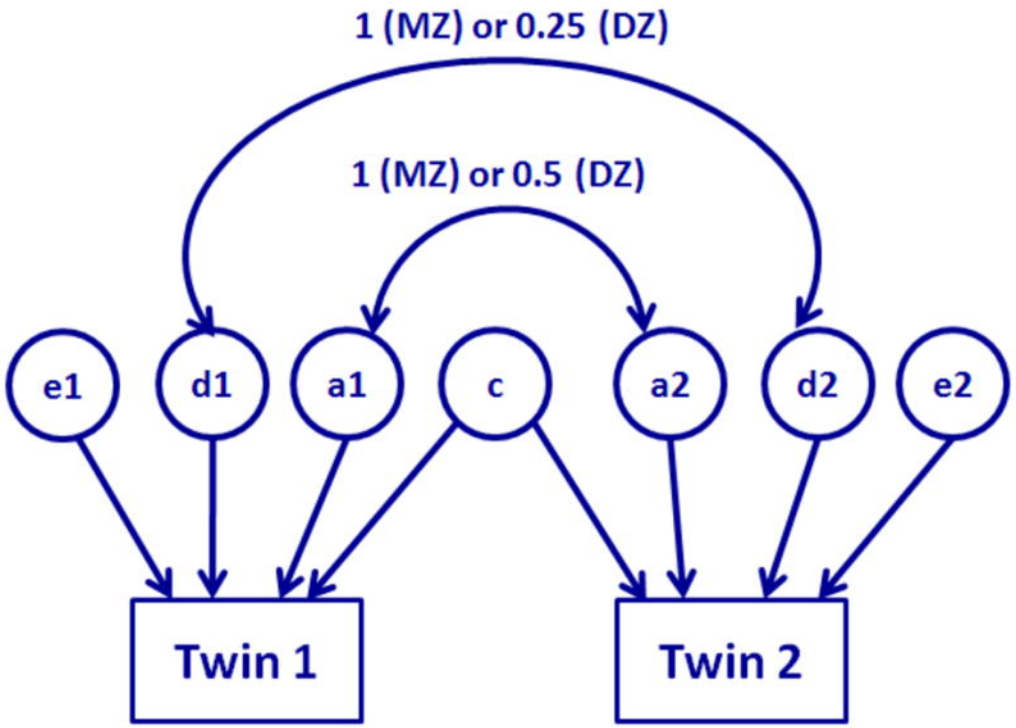
DZ twins share average 50% of their genes

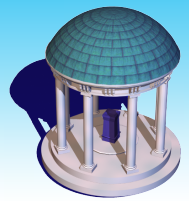
FSEM (I)

$$y_{ij}(d) = x_{ij}^T \beta(d) + a_{ij}(d) + d_{ij}(d) + c_i(d) + e_{ij}(d)$$

Additive and Dominance Genetic

Common Environmental





Two Strategies

Data $\{(Y_i, x_i) : i = 1, \cdots, n\}$ $Y_i = \{Y_i(d) : d \in D\}$

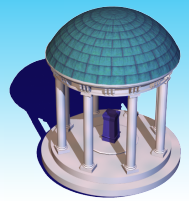
Strategy 1: Individual Approach

$$Y_i = \{Y_i(d) : d \in D\} \xrightarrow{W_i(.)} Y_{wi} = \{Y_{wi}(d) : d \in D\} = W_i(Y_i)$$

$$\prod_i p(Y_i \mid \theta = \{\theta(d) : d \in D\}) \qquad \prod_i p(Y_{wi} \mid \theta = \{\theta(d) : d \in D\})$$

Strategy 2: Global Approach

$$\prod_i p(Y_i \mid \theta = \{\theta(d) : d \in D\}) \xrightarrow{W(.)} W\left(\prod_i p(Y_i \mid \theta = \{\theta(d) : d \in D\})\right)$$



Hierarchical Smoothing Model

Strategy 1: Individual Approach (Hierarchical Smoothing Model)

- $Y_i \sim p(Y_i \mid \theta = \{\theta(d) : d \in D\})$

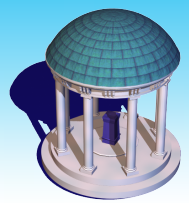
- $Y_i(d) = Y_{wi}(d) + \varepsilon_{wi}(d),$

  $\varepsilon_{wi} \sim p(\varepsilon_{wi} \mid 0, \sigma^2 = \{\sigma^2(d) : d \in D\})$

  $Y_{wi} \sim p(Y_{wi} \mid \tilde{\theta} = \{\tilde{\theta}(d) : d \in D\})$

Key Conditions:

- **Relative high SNR in individual image**
- **Consistency:** $\theta(d) = \tilde{\theta}(d) \text{ for } d \in D$



Problem 1: Smoothing

$$Y_i \sim p(Y_i \mid \theta = \{\theta(d) : d \in D\})$$

- **Parameters are not associated with the mean structure**

Twin Models

$$Y_i(d) = x_i^T \beta(d) + a_i(d) + c_i(d) + e_i(d),$$

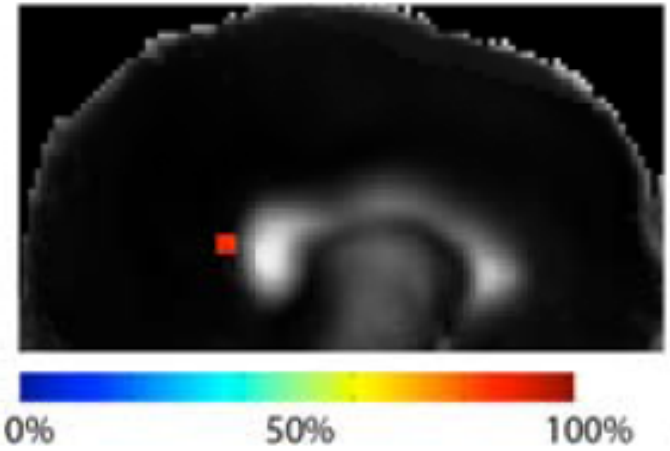
$$a_i(d) \sim (0, \sigma_a^2(d)),\ c_i(d) \sim (0, \sigma_c^2(d)),\ e_i(d) \sim (0, \sigma_e^2(d))$$

Consequence

Table 1. The effect of FA smoothing on the detection of FA-BDNF linkage

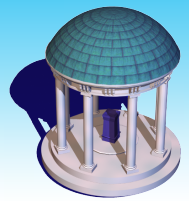
	FWHM (mm)				
	0	3	6	9	12
Number of FDR-significant voxels	142	68	44	0	0
FDR-significant clusters*	1(18), 2(9), 3(5), 4(4), 5(1), 10(1), 11(1), 13(1), 16(1), 20(1)	1(1), 2(1), 7(1), 58(1)	44(1)		

*Listed based on their size in voxels (the number of clusters of each size is in parentheses).



- **Moderate Smoothing: Single large cluster**
- **Excessive Smoothing: Effects disappeared**
- **No smooth: Small clusters**

Chiang et al. (2009)



Twin-MARM

There are two sets of parameters:
mean structure
variance structure

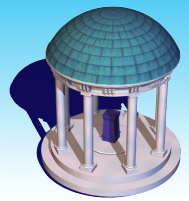
$$\{\beta(d) : d \in D\}$$

$$\{(\sigma_a^2(d), \sigma_c^2(d), \sigma_e^2(d)) : d \in D\}$$

Cartoon Model

Questions of interest:

- **Mean and variance images may have different patterns.**
- **Problematic Practice: Directly smooth imaging data**



Multiscale Adaptive Models

Strategy 2: Global Approach

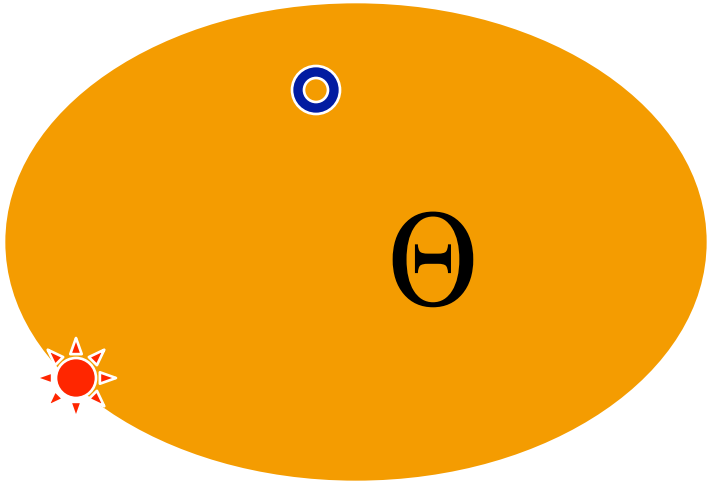
- $Y_i \sim \ell(\theta = \{\theta(d) : d \in D\} \mid Y_i)$

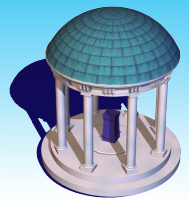
  $= \log p(Y_i \mid \theta = \{\theta(d) : d \in D\})$

- $W(\ell(\theta \mid Y_i))$

$\theta(d) \in \Theta$

Parametric Space





Twin-MARM

Two-stage Approach

- **Mean structure**

$$Y_{ij}(d) = x_{ij}^T \beta(d) + \varepsilon_{ij}(d) \Rightarrow \{\hat{\beta}(d;h) : d \in D\}$$

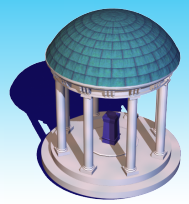
- **Variance structure**

$$\{Y_{ij}(d) - x_{ij}^T \hat{\beta}(d;h)\}^2 = z_{ij}^T \rho(d) + \delta_{ij}(d) \Rightarrow \{\hat{\rho}(d;h) : d \in D\}$$

Theorem 2: Substituting $\{\hat{\beta}(d;h) : d \in D\}$ into the second stage has negligible effect.

$$\omega_1(d,d';h) = K_{loc}(\| d - d' \|_2 / h) K_{st}(D_\beta(d,d';h)/C_n) \Rightarrow \{\hat{\beta}(d;h) : d \in D\}$$

$$\omega_2(d,d';h) = K_{loc}(\| d - d' \|_2 / h) K_{st}(D_\rho(d,d';h)/C_n) \Rightarrow \{\hat{\rho}(d;h) : d \in D\}$$



Simulation Study

ACE Model: $y_{ij}(v) = \mathbf{x}_{ij}^T \beta(v) + a_{ij}(v) + c_i(v) + e_{ij}(v)$

We set: $(\beta_2(v), \beta_3(v), \sigma_d(v)^2, \sigma_e(v)^2)^T = (1,1,1,1)^T$ across all voxels v

$(\beta_1(v), \sigma_a(v)^2)$ as $(0,0), (0.3,0.5), (0.6,1), (0.9,1.5)$ and $(1.2,2.0)$

across 5 regions of interest

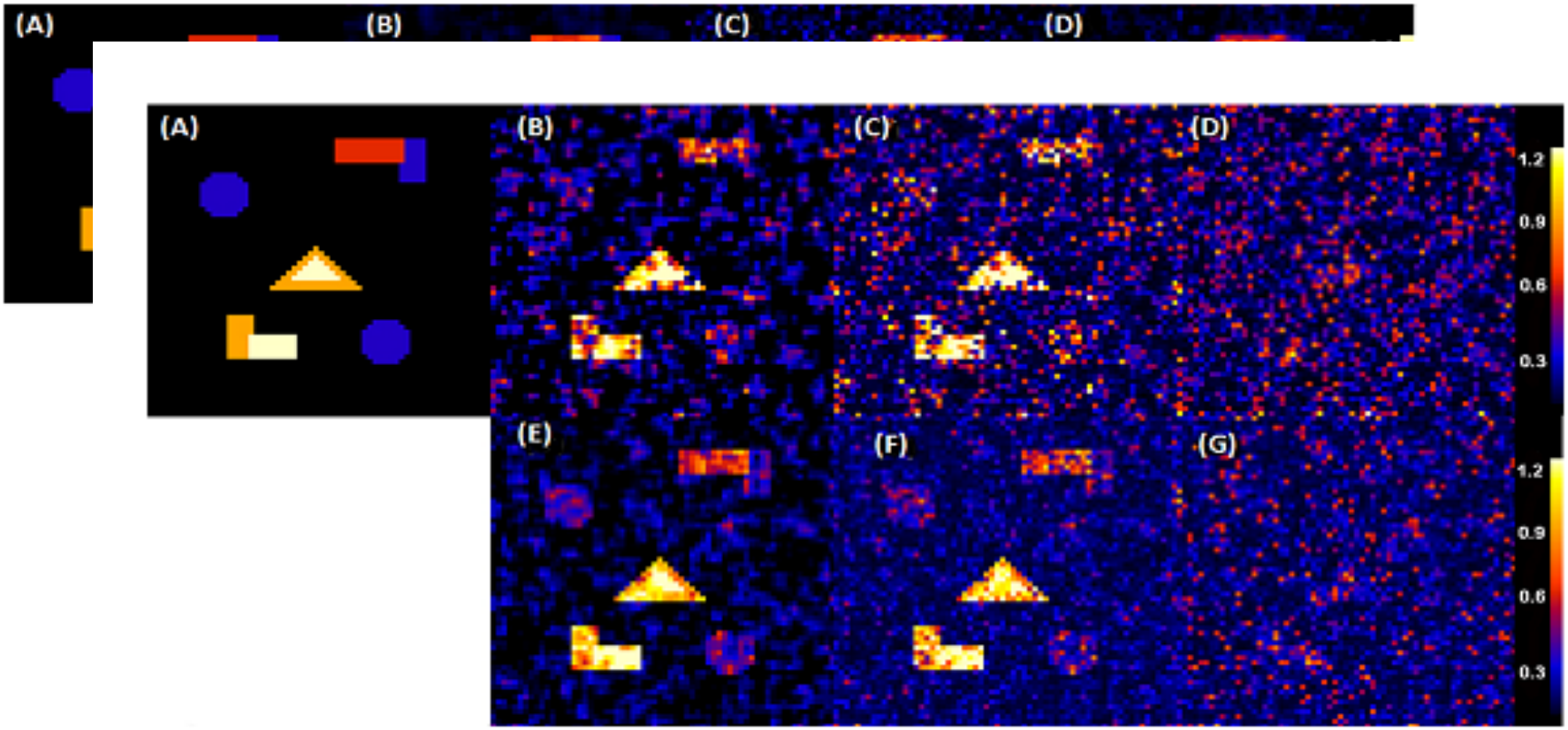
$a_{ij}(v)$, $d_{ij}(v)$, $c_i(v)$ and $e_{ij}(v)$ **: are independently normally distributed**

with mean 0 and variance: $\sigma_a(v)^2$, $\sigma_d(v)^2$, $\sigma_c(v)^2$, and $\sigma_e(v)^2$

$\mathrm{Cov}(a_{i1}(v), a_{i2}(v))$ **equals** $\sigma_a(v)^2$ **for MZ twins** $\sigma_a(v)^2/2$

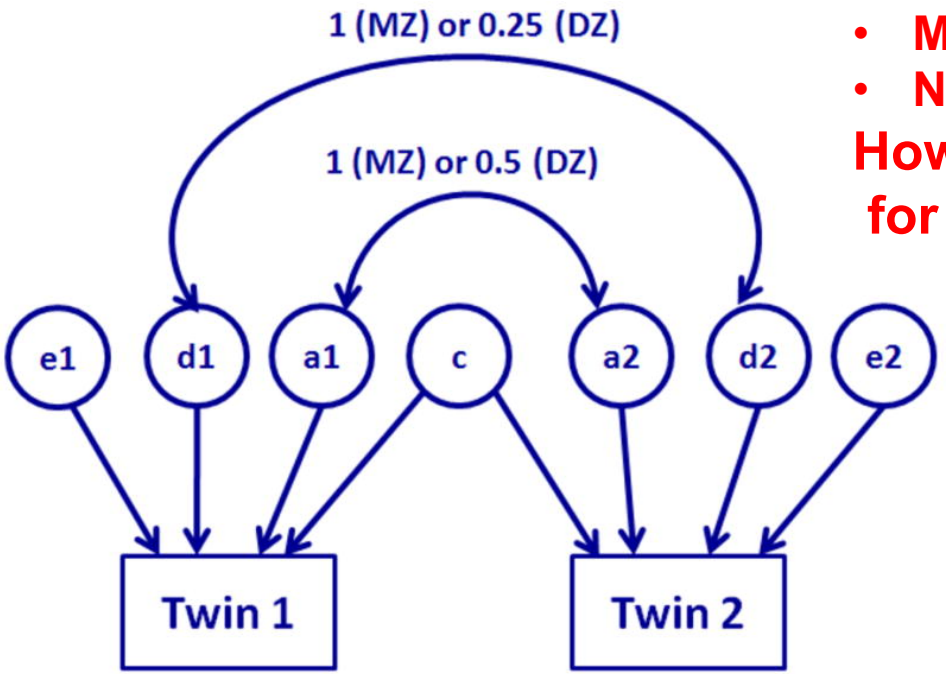
Simulation Study

It is dangerous to use Gaussian-kernel to smooth imaging data and then carry out twin analysis.



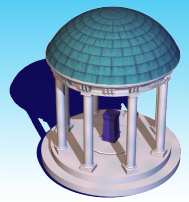
FSEM (I)

$$y_{ij}(d) = x_{ij}^T \beta(d) + a_{ij}(d) + d_{ij}(d) + c_i(d) + e_{ij}(d)$$



- **Marginal Modeling**
- **Not capture covariance structure**

How to define stochastic processes for DZ?



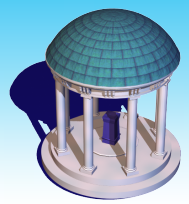
FSEM (II)

$$y_{ij}(d) = x_{ij}^T \beta(d) + \sqrt{0.5}1(\mathrm{DZ})a_{ij}(d) + [1(\mathrm{MZ}) + \sqrt{0.5}1(\mathrm{DZ})]a_i(d) + c_i(d) + e_{ij}(d)$$

$$a_{ij}(d) \sim GP(0, \Sigma_a)$$
$$a_i(d) \sim GP(0, \Sigma_a)$$
$$c_i(d) \sim GP(0, \Sigma_c)$$
$$e_{ij}(d) \sim GP(0, \Sigma_e)$$

$$\Sigma_Y(d,d') = \begin{bmatrix} \Sigma_a + \Sigma_c + \Sigma_e & \Sigma_a + \Sigma_c \\ \Sigma_a + \Sigma_c & \Sigma_a + \Sigma_c + \Sigma_e \end{bmatrix}(d,d')$$

$$\Sigma_Y(d,d') = \begin{bmatrix} \Sigma_a + \Sigma_c + \Sigma_e & 0.5\Sigma_a + \Sigma_c \\ 0.5\Sigma_a + \Sigma_c & \Sigma_a + \Sigma_c + \Sigma_e \end{bmatrix}(d,d')$$



FSEM (II)

Three-stage Approach

- **Mean structure**

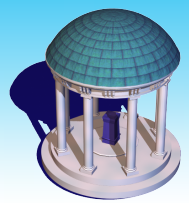
$$Y_{ij}(d) = x_{ij}^T \beta(d) + \varepsilon_{ij}(d) \Rightarrow \{\hat{\beta}(d;h) : d \in D\}$$

- **Variance structure (Weighted likelihood)**

$$\{Y_{ij}(d) - x_{ij}^T \hat{\beta}(d;h)\}^2 = z_{ij}^T \rho(d) + \delta_{ij}(d) \Rightarrow \{\hat{\rho}(d;h) : d \in D\}$$

- **Estimate covariance operators**

$$\Sigma_a(d,d') \text{ and } \Sigma_c(d,d')$$



FSEM (II)

Two Key Test Procedures

- **Test marginal genetic and environmental effects**

$$H_{0A}(d):\Sigma_a(d,d)=0 \text{ versus } H_{1A}(d):\Sigma_a(d,d)>0$$

$$H_{0C}(d):\Sigma_c(d,d)=0 \text{ versus } H_{1C}(d):\Sigma_c(d,d)>0$$

- **Test global genetic and environmental effects**

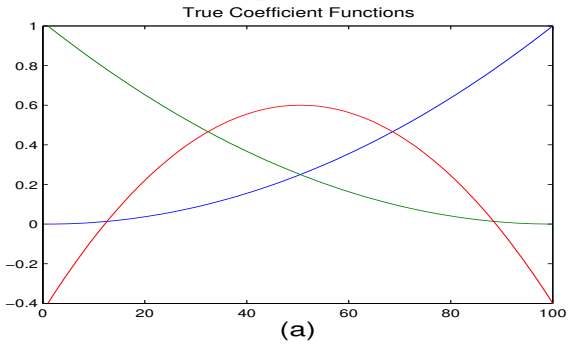
$$H_{0A}:\int\Sigma_a(d,d)m(d)=0 \text{ versus } H_{1A}(d):\int\Sigma_a(d,d)m(d)>0$$

$$H_{0C}:\int\Sigma_c(d,d)m(d)=0 \text{ versus } H_{1C}(d):\int\Sigma_c(d,d)m(d)>0$$

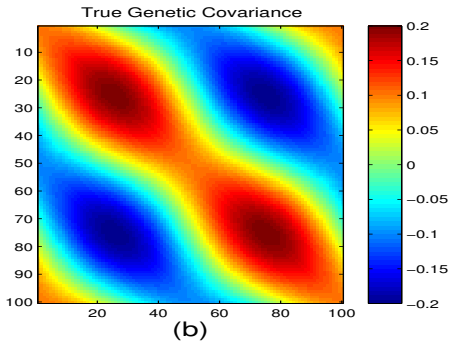
Simulations

True

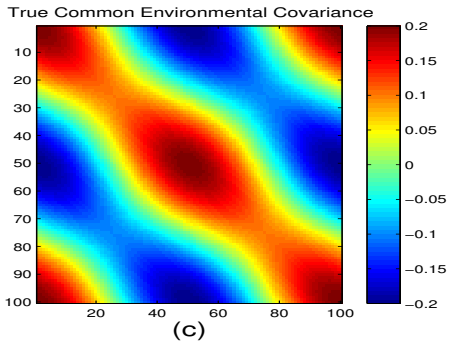
$\beta(d)$

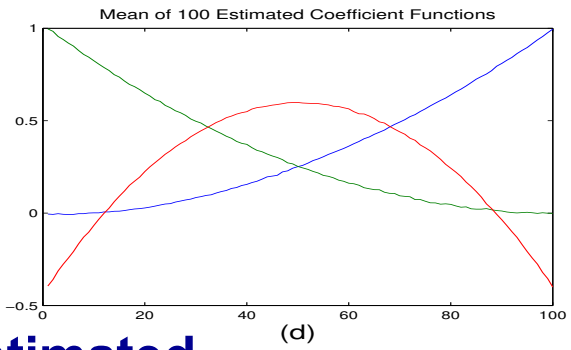


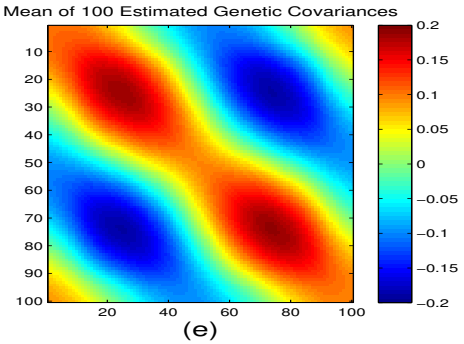
$\Sigma_a(d,d')$

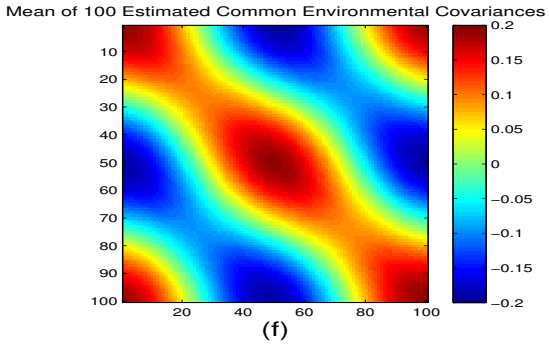


$\Sigma_c(d,d')$



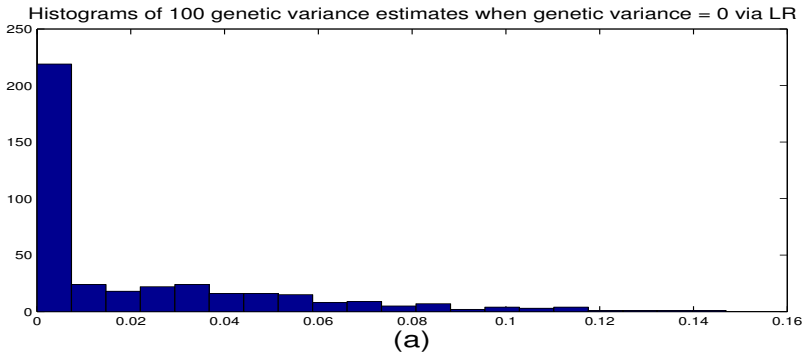


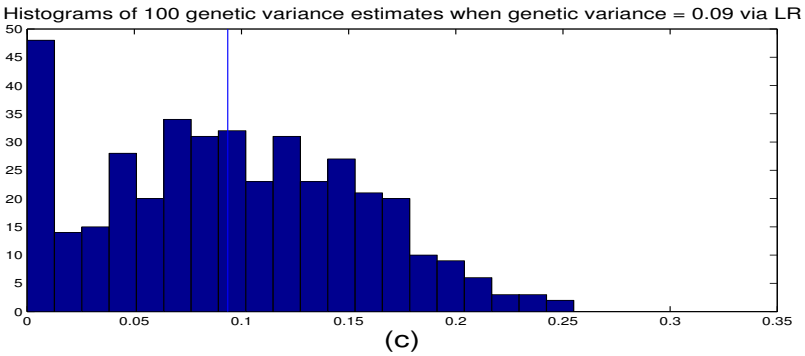


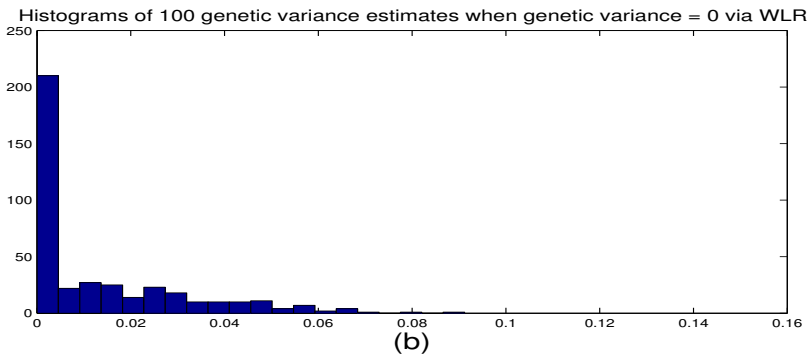


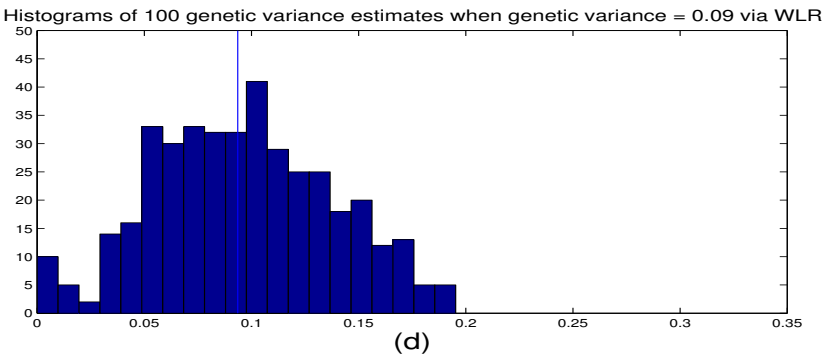
Estimated

Simulations





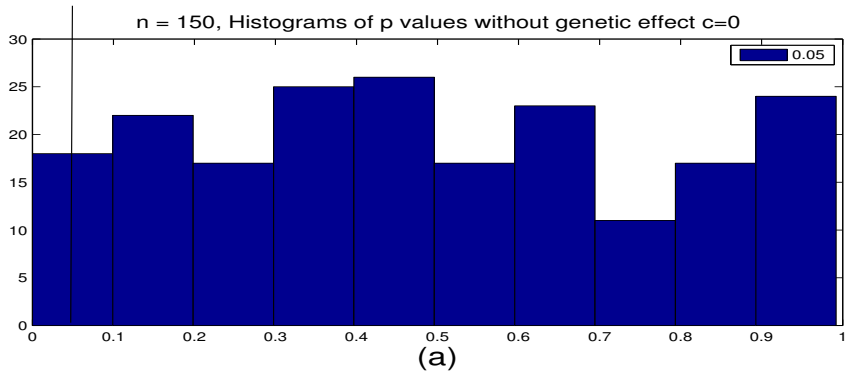


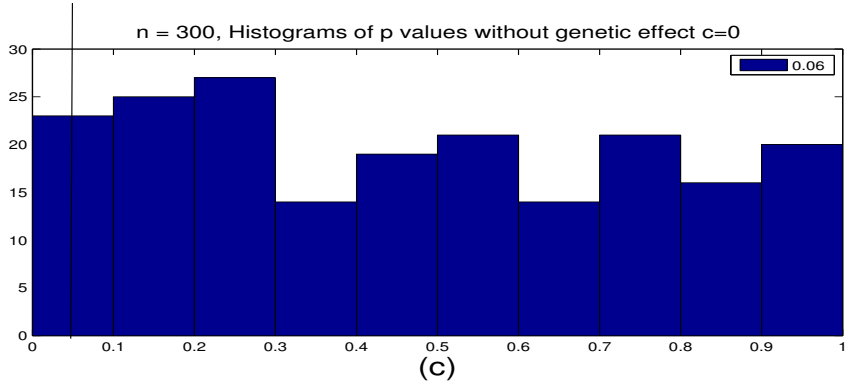


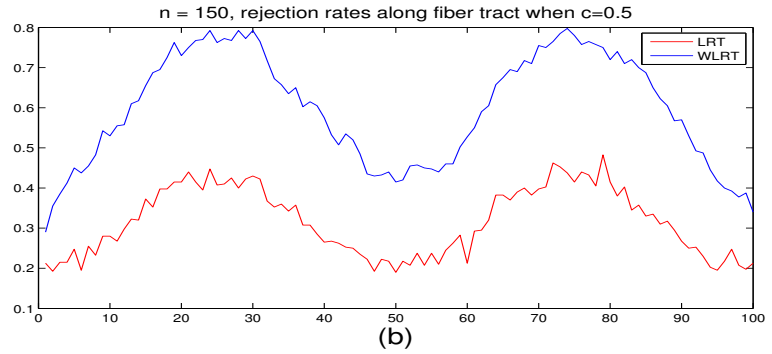
No Genetic Effect

Genetic Variance=0.09

Simulations

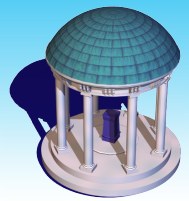






Type I

Power



UNC Early Brain Development Studies

PIs: Drs. John H. Gilmore and Weili Lin

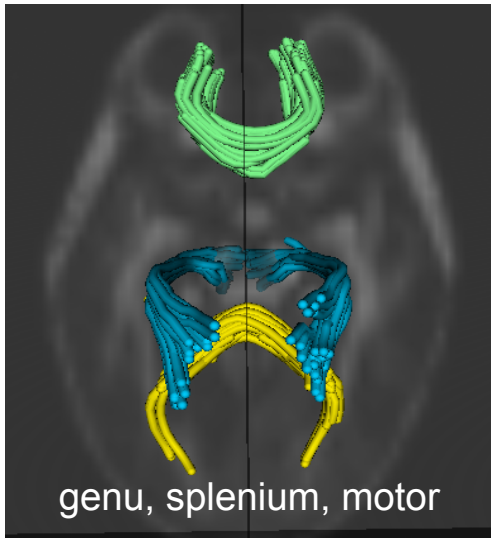
To track changes in behavior with brain structure, connectivity, and function, in order to characterize the progression from primary changes to subsequent clinical presentation, and to identify predictors of divergence from the typical trajectory.

- Singletons, twins, high risk
- A longitudinal prospective study
- 900 young children aged 0 to 6 years
- Recruited prenatally
  - Exclusion: ultrasound abnormality, significant fetal/ maternal medical problem, substance abuse
- 3TMRI (Seimens Allegra)
  - T1, T2, DTI, resting state fMRI
- Scanned during normal sleep(no meds)
- Ear protection, head in vac-fix device
- Success rate: 87% @ 2 weeks, 71% @ 1 year, 62% at 2 years

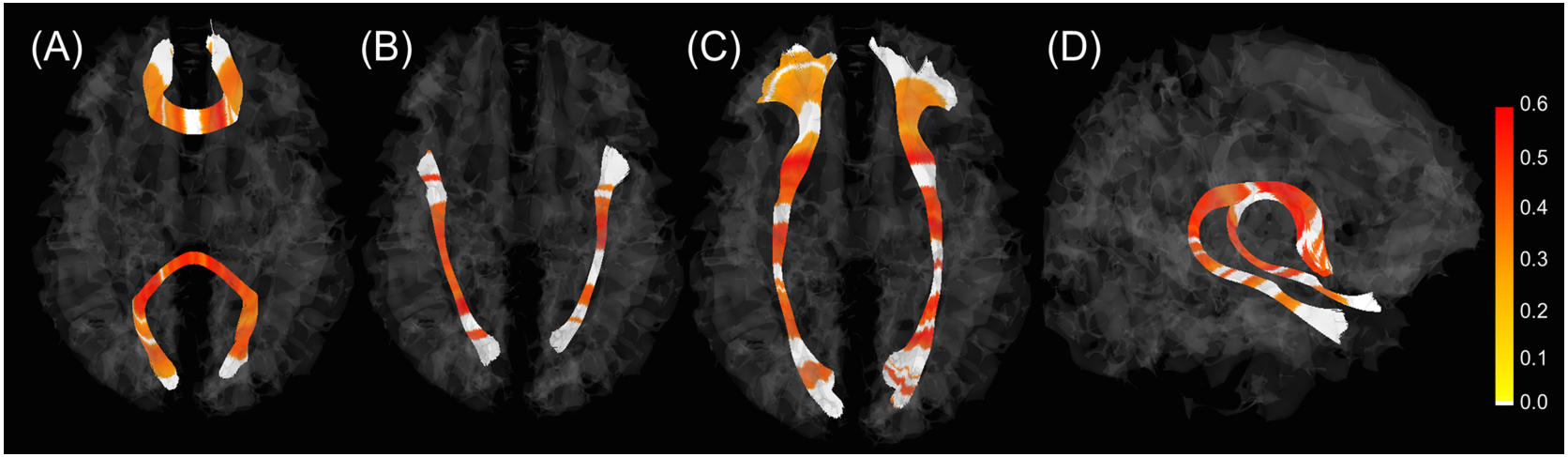
Quantitative tract-based white matter heritability in twin neonates

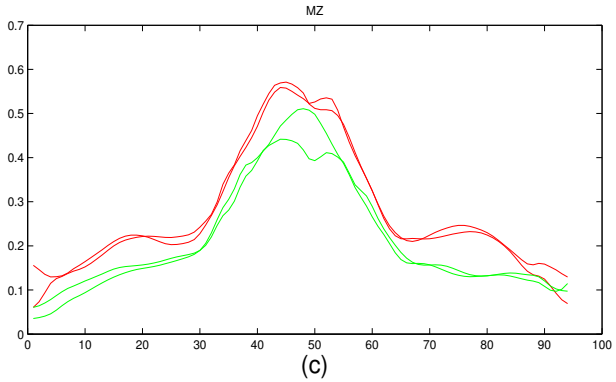
- The data set consists of **356** healthy twin neonates with **190 males and 166 females** from the neonatal project as part of the UNC Early Brain Development Studies.
- There are **129** twin pairs (**48 MZ** twin pairs and **81 DZ** twin pairs) and **98 unrelated "singleton" twins** - a single unpaired twin subject in which a usable scan was not obtained from the co-twin.
- The gestational ages of these infants range from 257 to 401 days, and their mean gestational age is 289 days with standard deviation 18 days.

Question of interest:
comprehensive heritability data on white matter microstructure fractional anisotropy (FA), radial diffusion (RD), and axial diffusion (AD) along 47 fiber tracts.

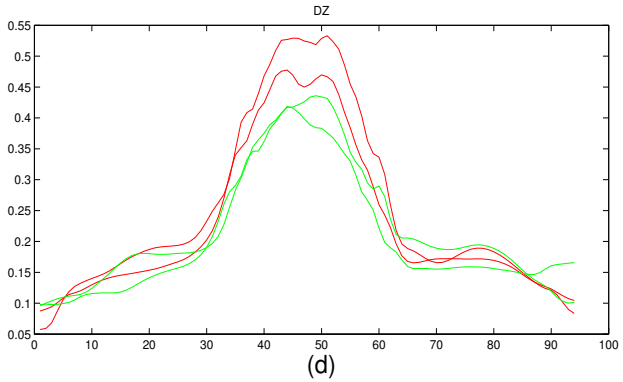


Twin Functional Data

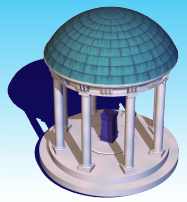




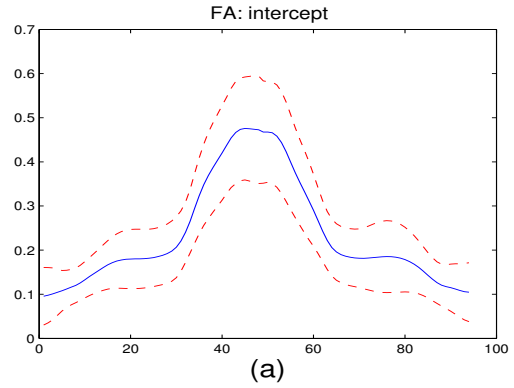
MZ

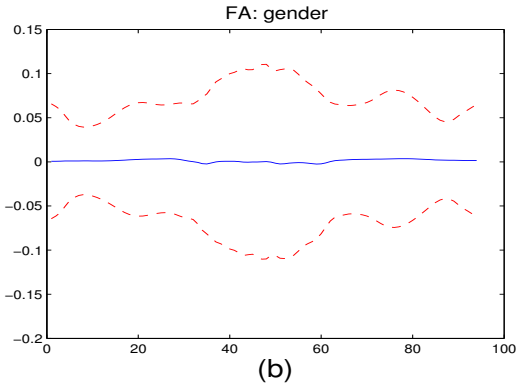


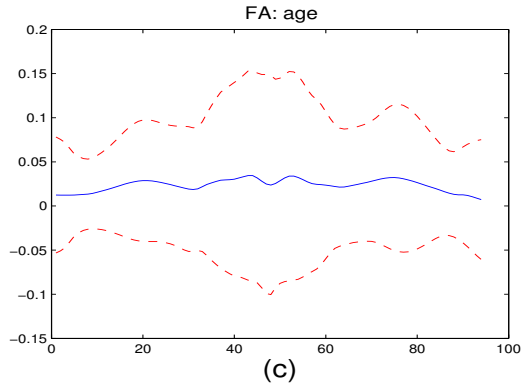
DZ

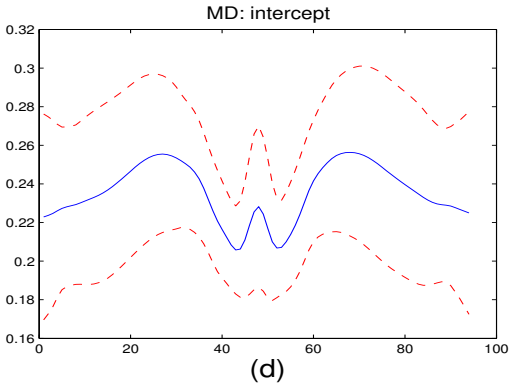


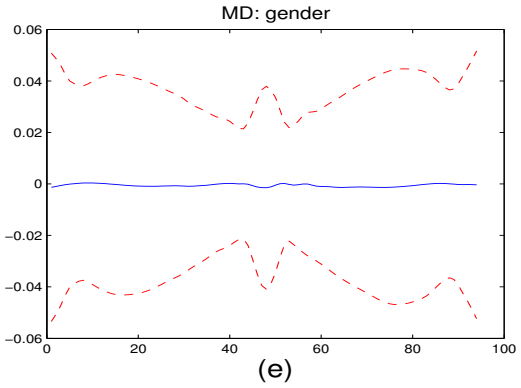
Coefficient Functions

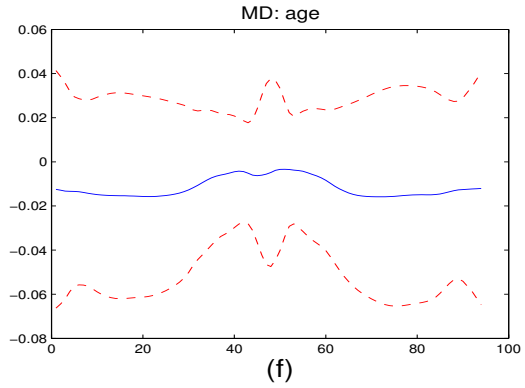




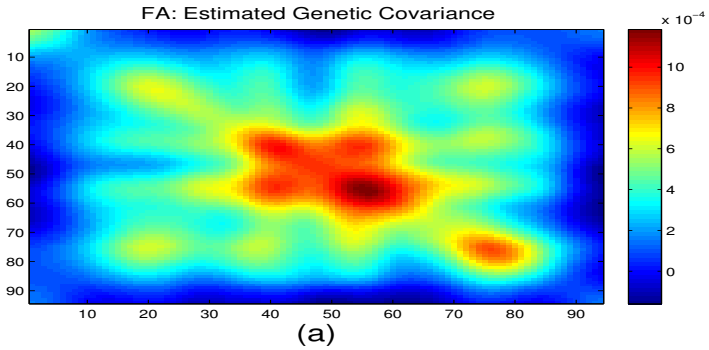


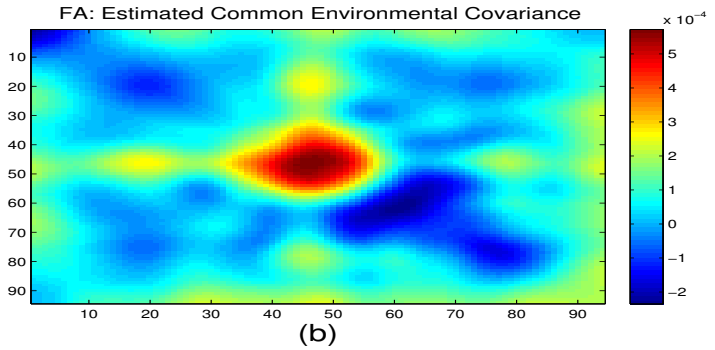


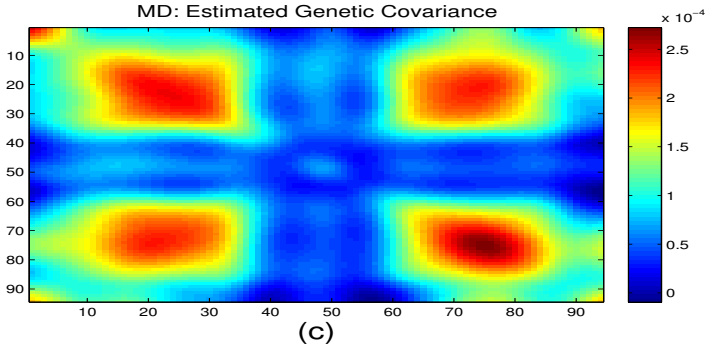


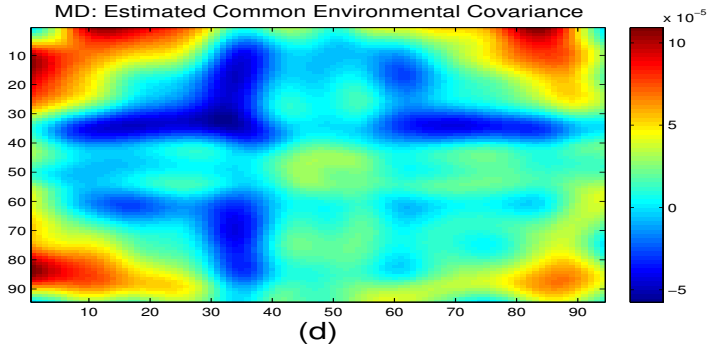


Genetic and Environmental COs

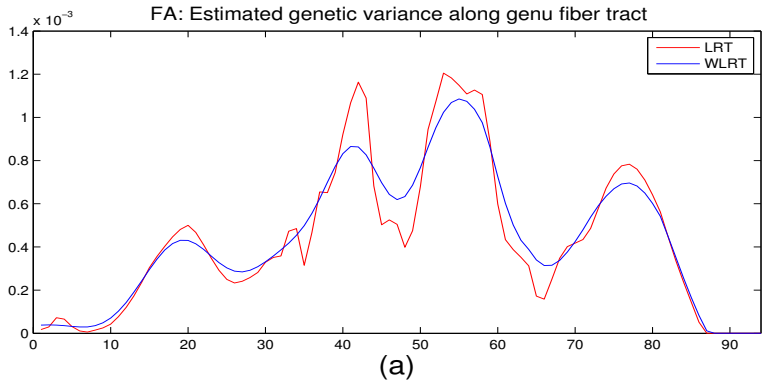


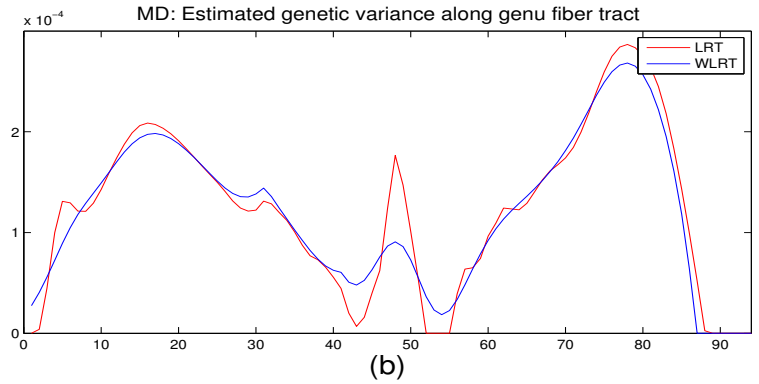


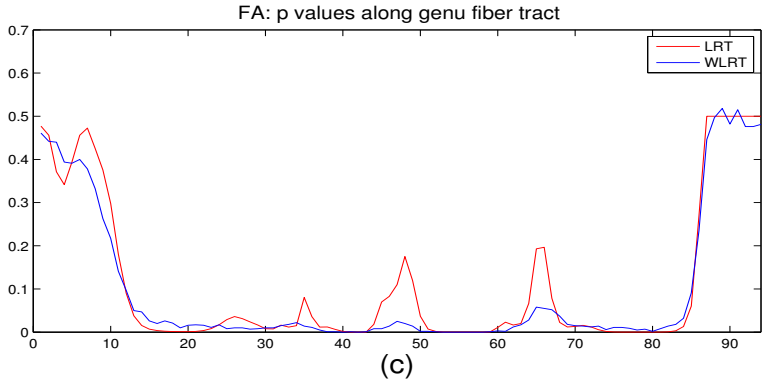


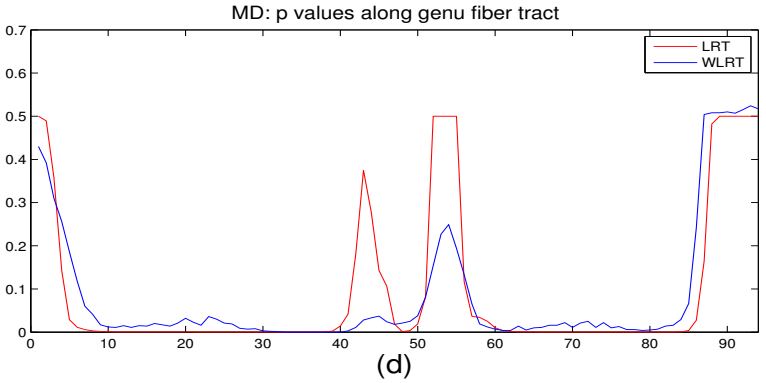


Genetic Effects

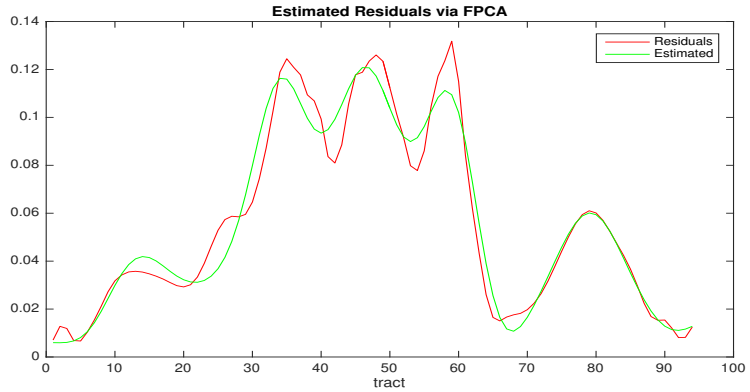


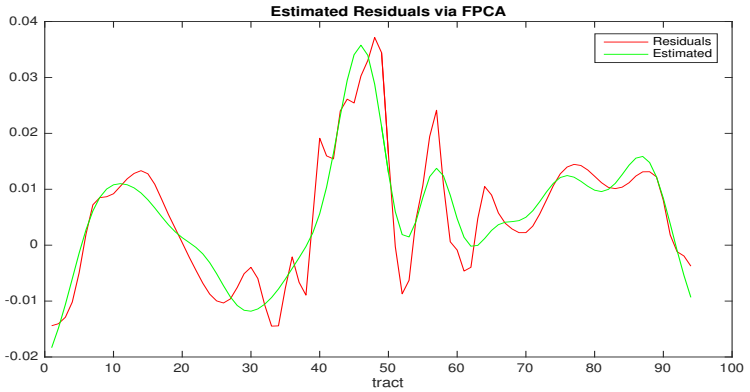


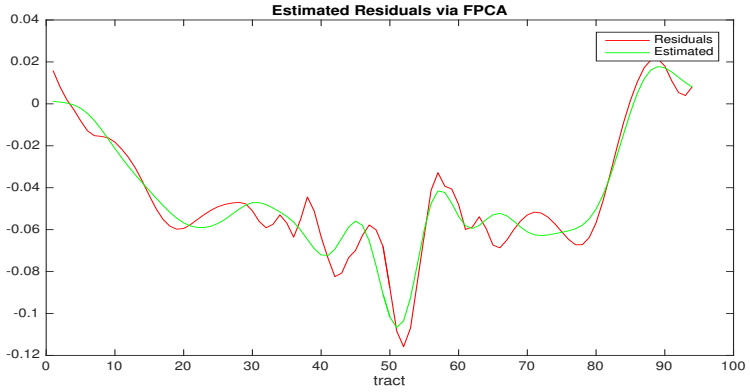


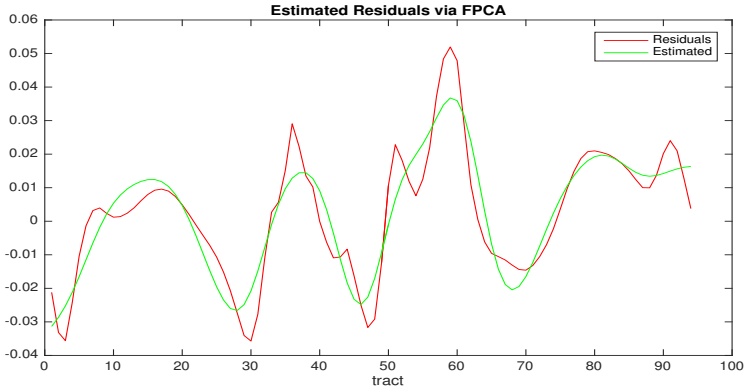


Prediction Accuracy

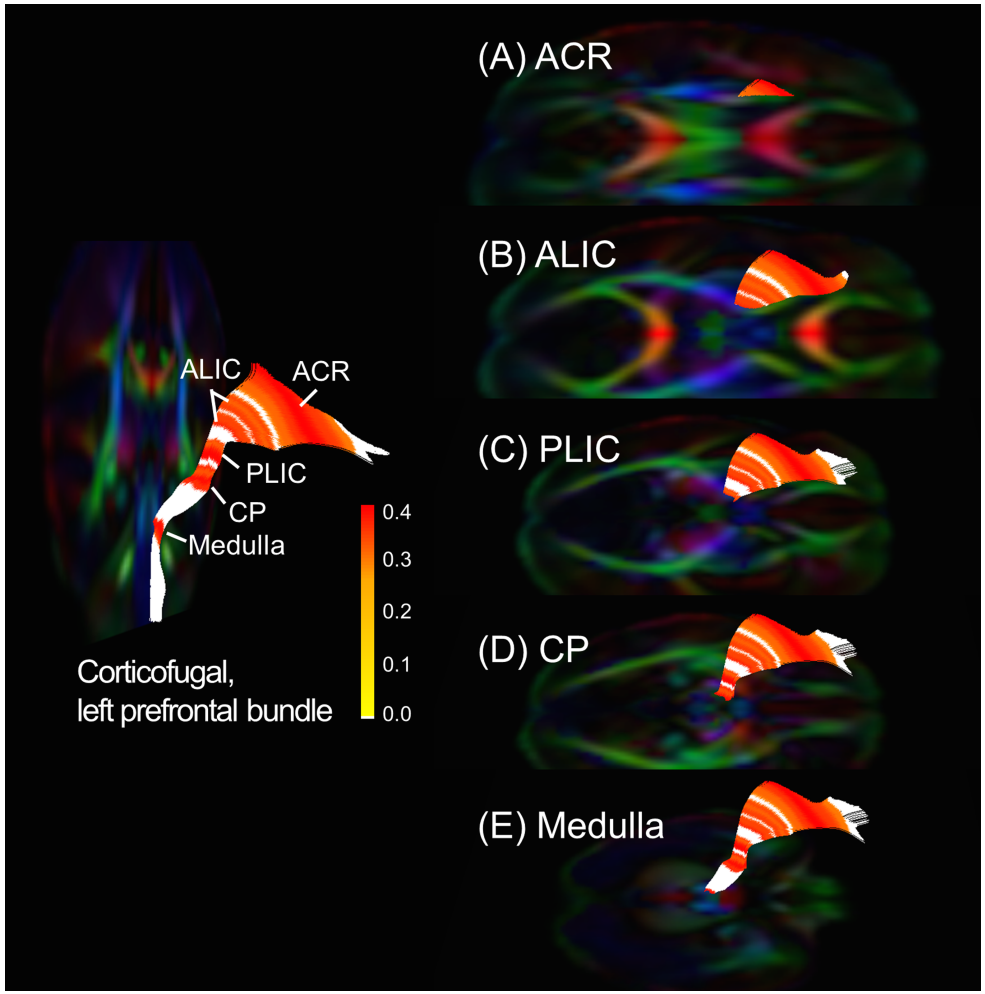


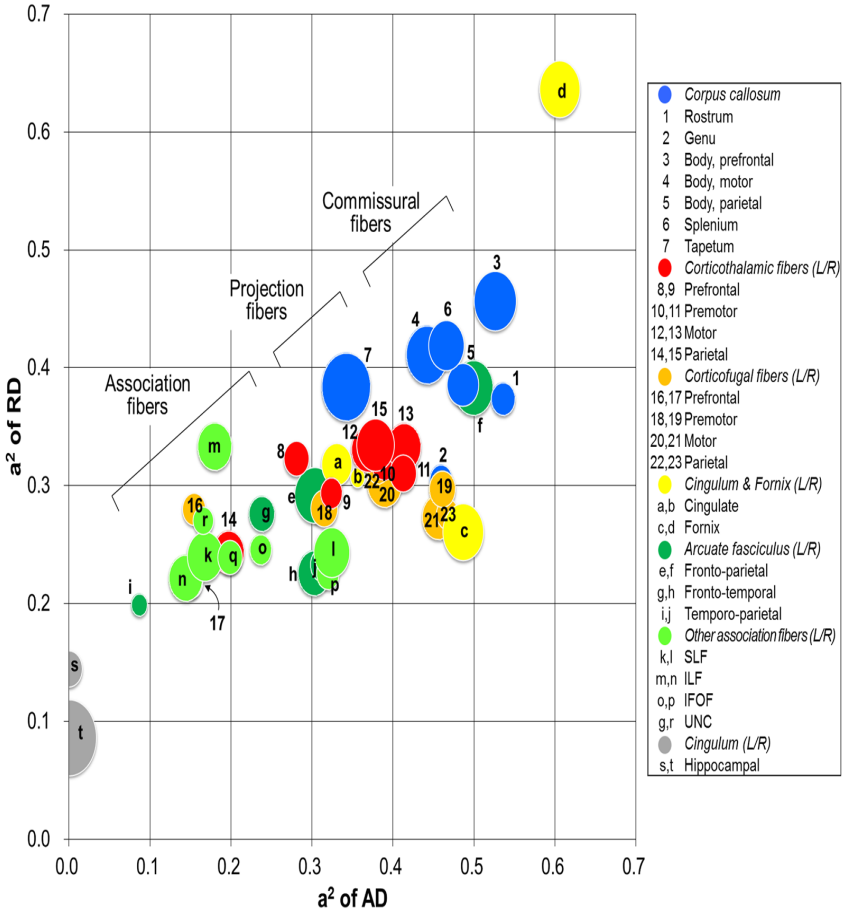


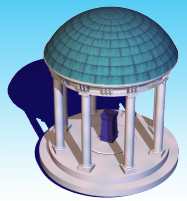




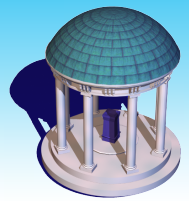
Heritability







HPRM: Hierarchical Principal Regression Model of Diffusion Tensor Bundle Statistics



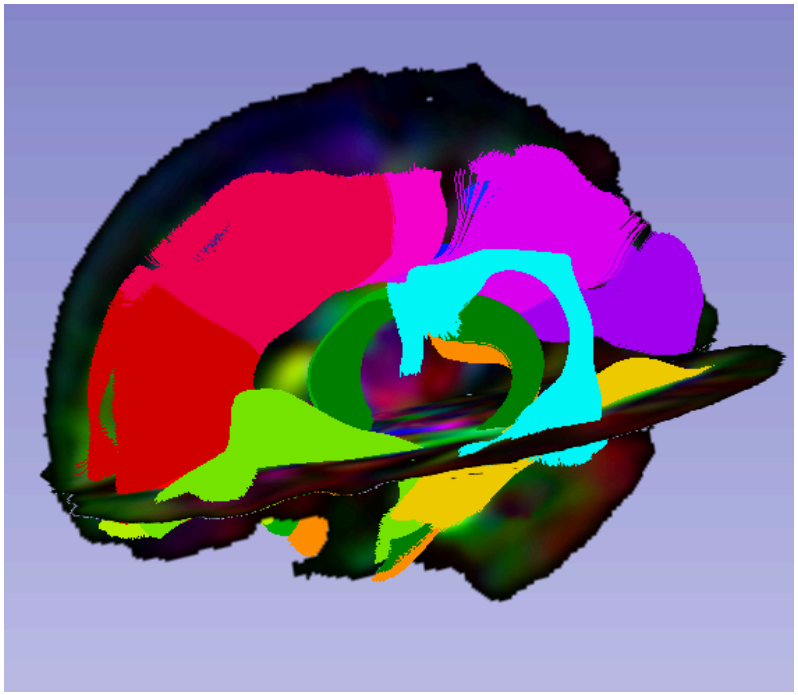
Reading Materials

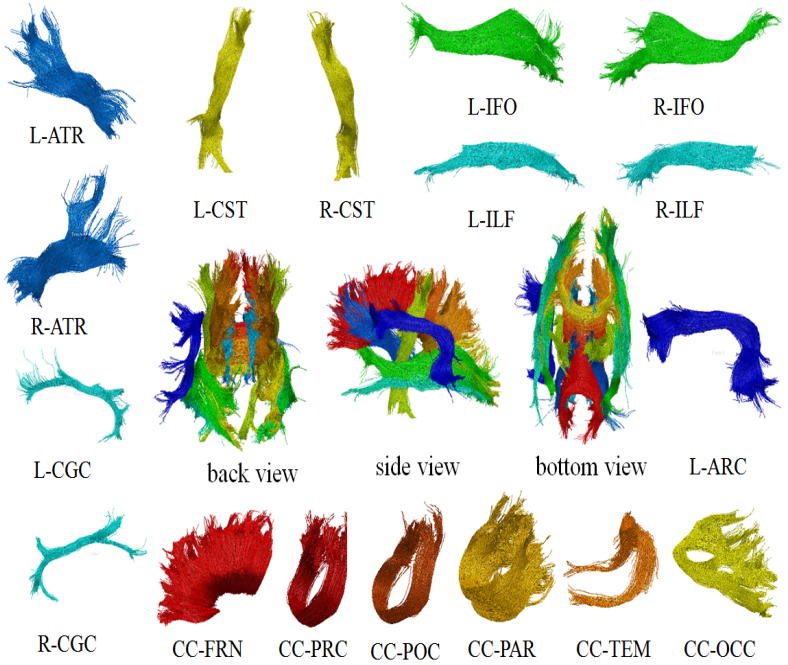
1. Zhang, J. W., Ibrahim, J. G., J. Gilmore., M. Styner and Zhu, H.T. HPRM: Hierarchical Principal Regression Model of Diffusion Tensor Bundle Statistics. 2016. In Submission.
2. Luo, X. C., Zhu, L. X., Kong, L., ***Zhu, H.T.*** Functional Nonlinear Mixed Effects Models For Longitudinal Image Data. ***Information Processing in Medical Imaging (IPMI)*** 2015.
3. Luo. X. C., Zhu, L.X., and Zhu, H.T. (2016). Single-index Varying Coefficient Model for Functional Responses. ***Biometrics***, in revision.
4. Liang, J. L., Huang, C., and Zhu, H.T. (2014). Functional single-index varying coefficient models. In submission.
5. Hua, Z.W., Dunson, D., Gilmore, J.H., Styner, M., and ***Zhu, HT.*** (2012). Semiparametric Bayesian local functional models for diffusion tensor tract statistics. ***NeuroImage,*** 63, 460-674.
6. Zhu, HT., Kong, L., Li, R., Styner, M., Gerig, G., Lin, W. and Gilmore, J. H. (2011). FADTTS: Functional Analysis of Diffusion Tensor Tract Statistics, ***NeuroImage***, 56, 1412-1425.
7. Zhu, H.T., Styner, M., Tang, N.S., Liu, Z.X., Lin, W.L., Gilmore, J.H. (2010). FRATS: functional regression analysis of DTI tract statistics. ***IEEE Transactions on Medical Imaging***, 29, 1039-1049.

Video

http://www.birs.ca/events/2016/5-day-workshops/16w5036/videos/watch/201602021312-Zhang.html

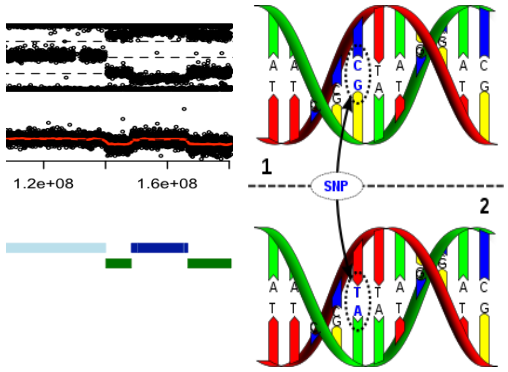
White Matter Fiber Bundles



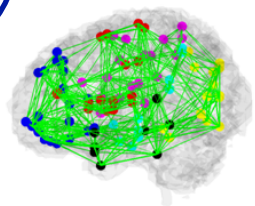


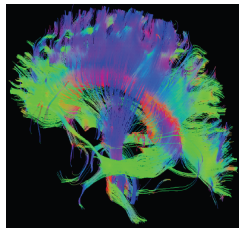
Scalar-on-Functional Models

E: environmental factors



G: genetic/genomics



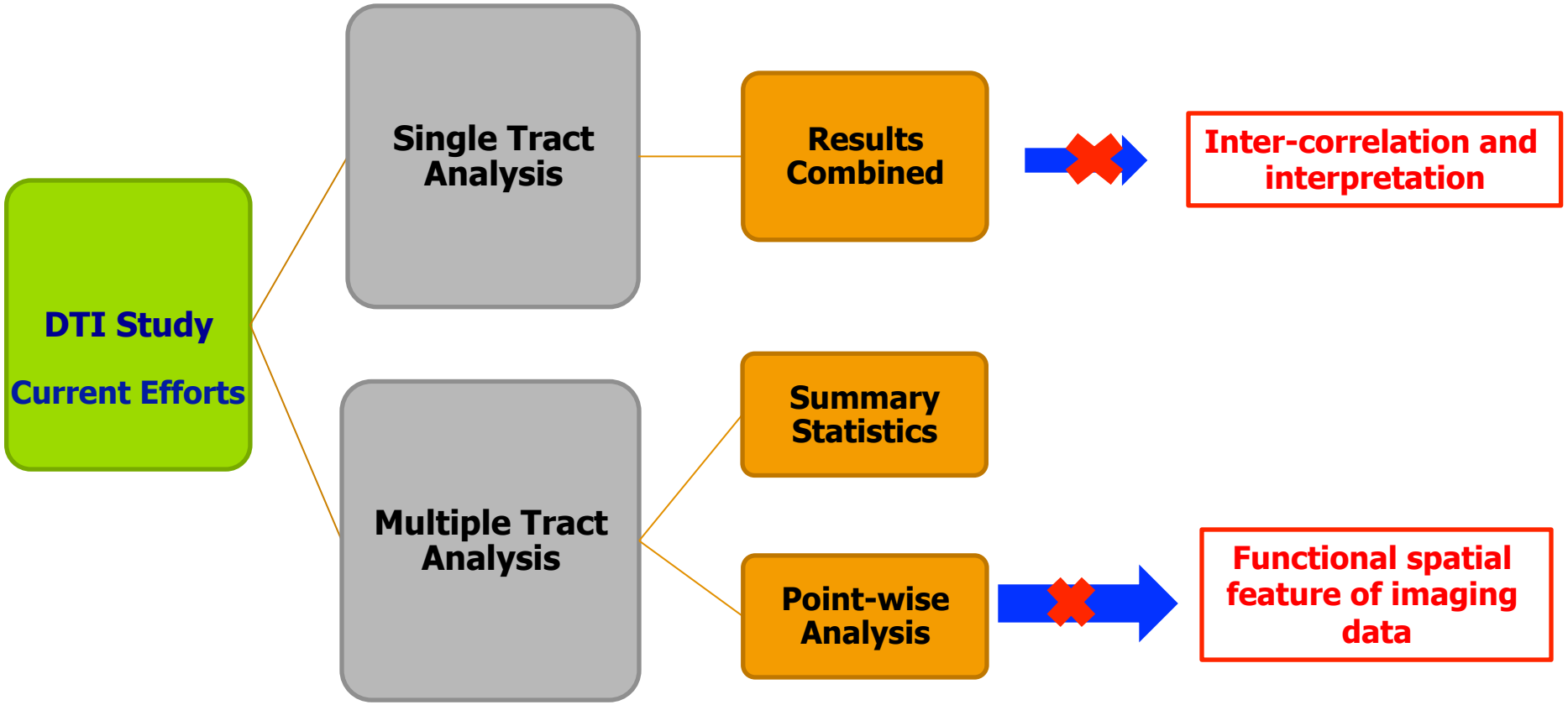


I: imaging/device

D: disease

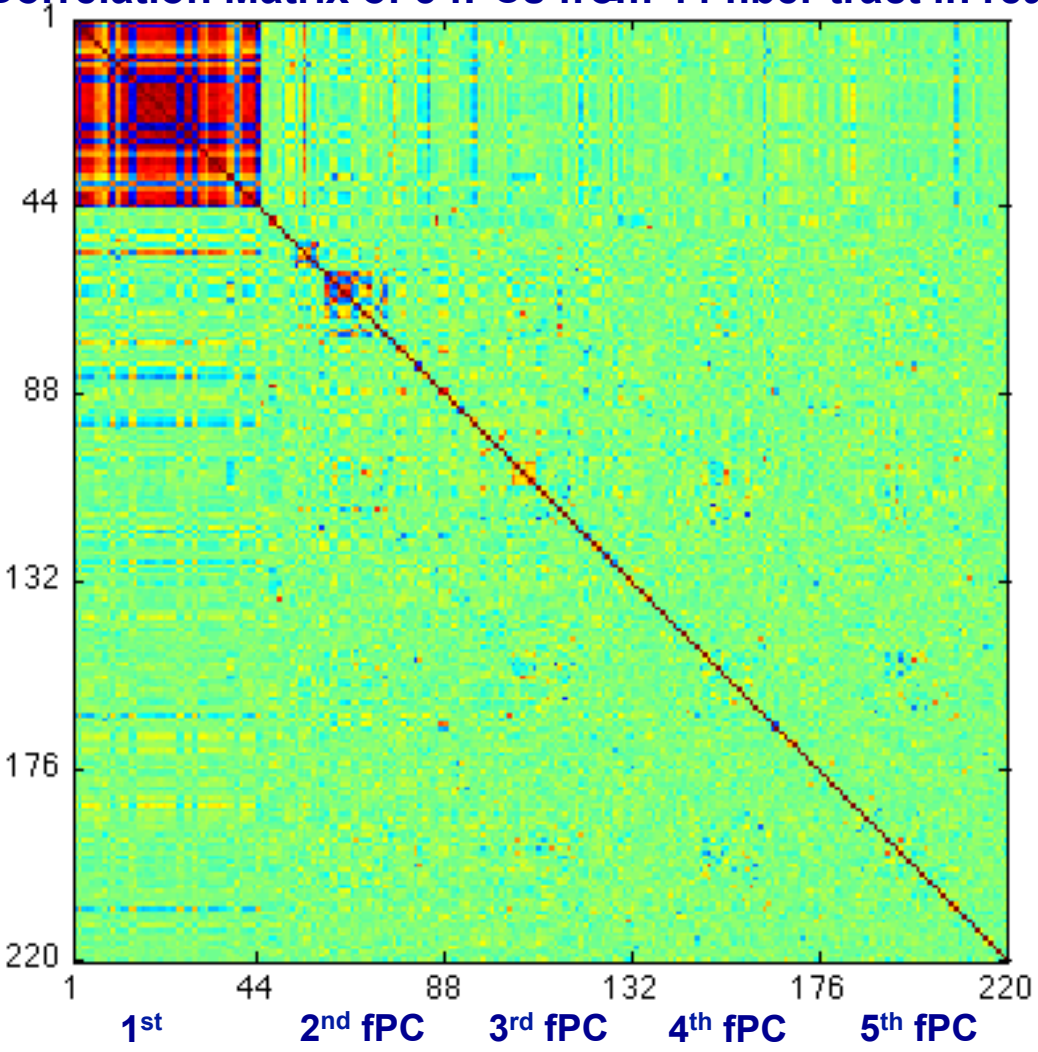
http://en.wikipedia.org/wiki/DNA_sequence

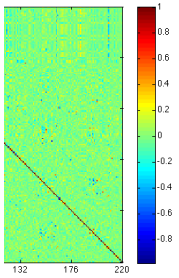
Existing Methods



Path Diagram

Correlation Matrix of 5 fPCs from 44 fiber tract in real data





Gau

$y_{i,m}(s) =$

Mean profile

Est
Lin

Tes
Sui

Factors

$B_\eta \mathbf{x}_i + \boldsymbol{e}_{\eta,i}$

residual

$\beta_{\delta m} \mathbf{x}_i + \boldsymbol{e}_{\delta_m,i}$

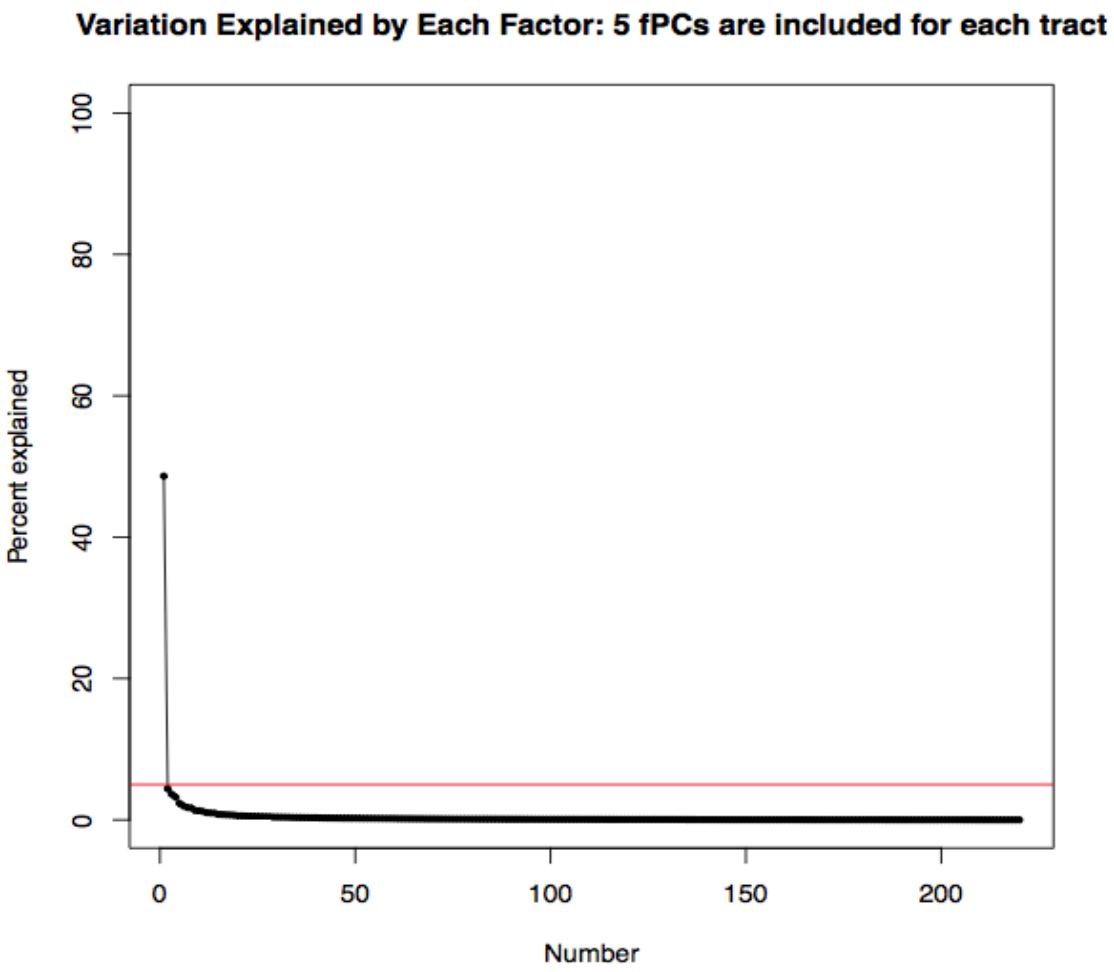
GWAS of early brain development: DTI Tracts

- 472 twin subjects
  - 236 DZ pairs, 32 MZ pairs and 260 Singletons
- Neonatal MRI (around one month old)
  - 3T Siemens Allegra head-only scanner or 3T Siemens TIM Trio
  - DTIPrep (Quality Control), Slicer[1] (Visual QC, DTI atlas creation, Fiber tract segmentation, Registration)
  - DTI Data: FA measure of 44 Fiber Tracts
  - TBSS Data: FA measure of 21 bundles
- Genetic markers
  - ~ 800k genetic marker
  - Imputation with MACH-Admix, template 1000G Phase I v3
  - ~ 6 million SNPs and indels with MAF>0.05
- Fit ACE model in regression
- Covariates
  - Gestational age at birth, family income, DTI direction, Scanner Type, 3 genetic PC scores

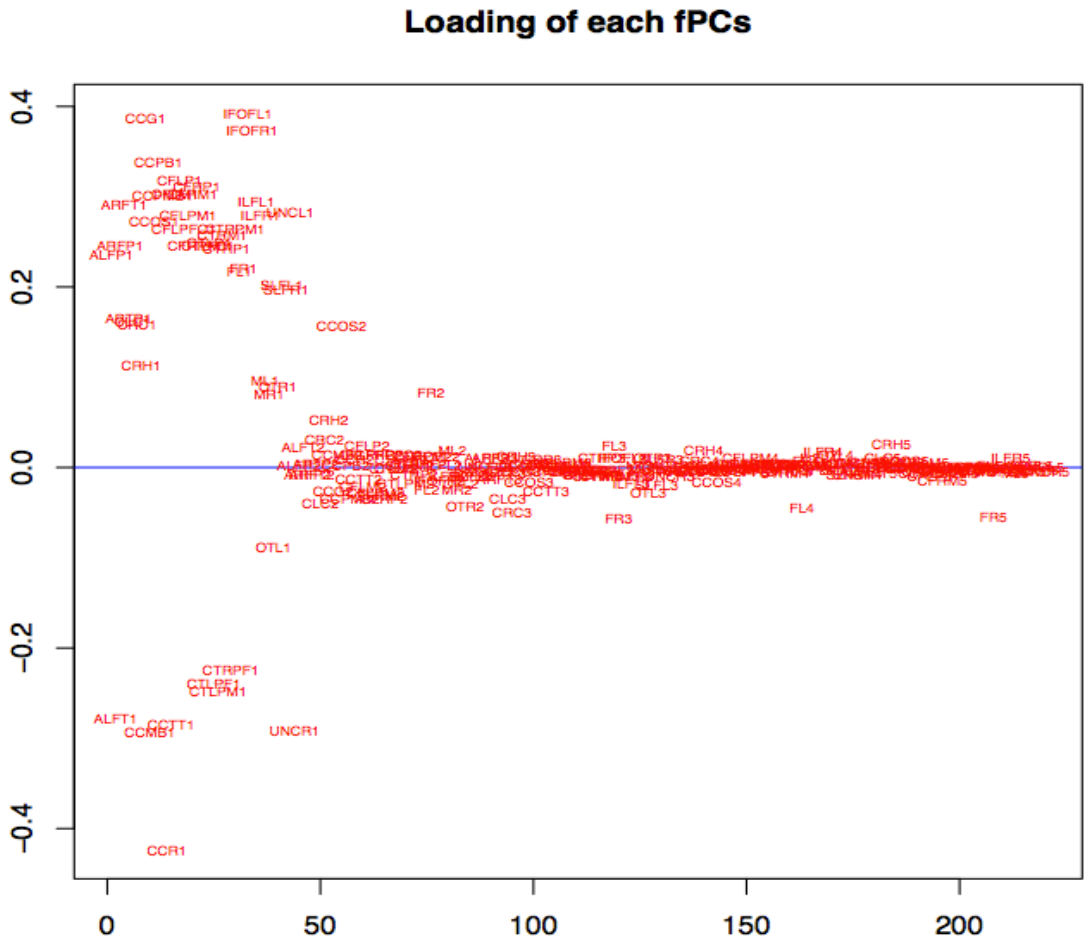
GWAS of early brain development: DTI Tracts

5 fPcs to include >70%

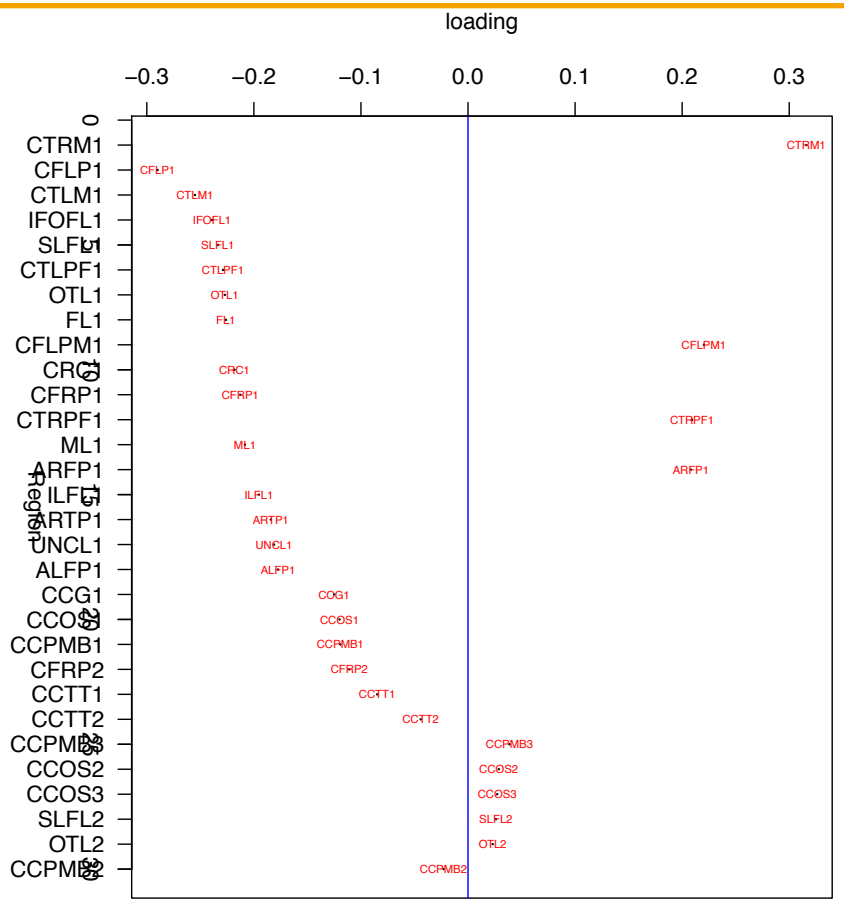
Factor Analysis

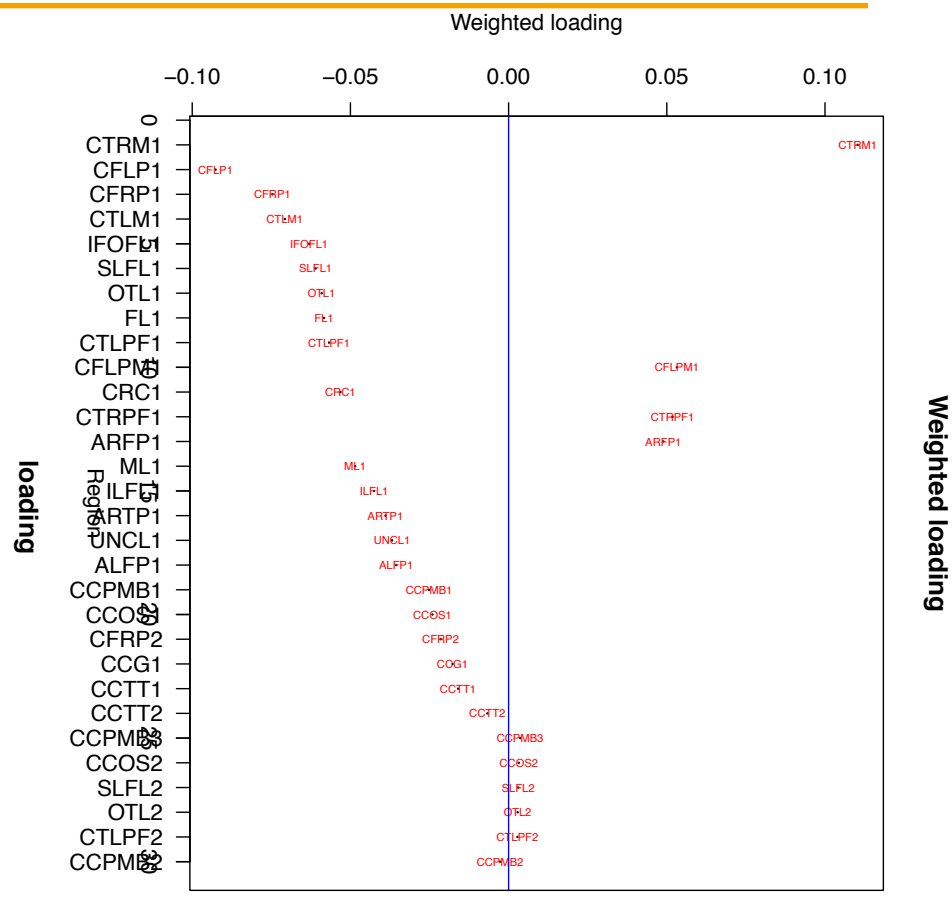


Plot of Loading

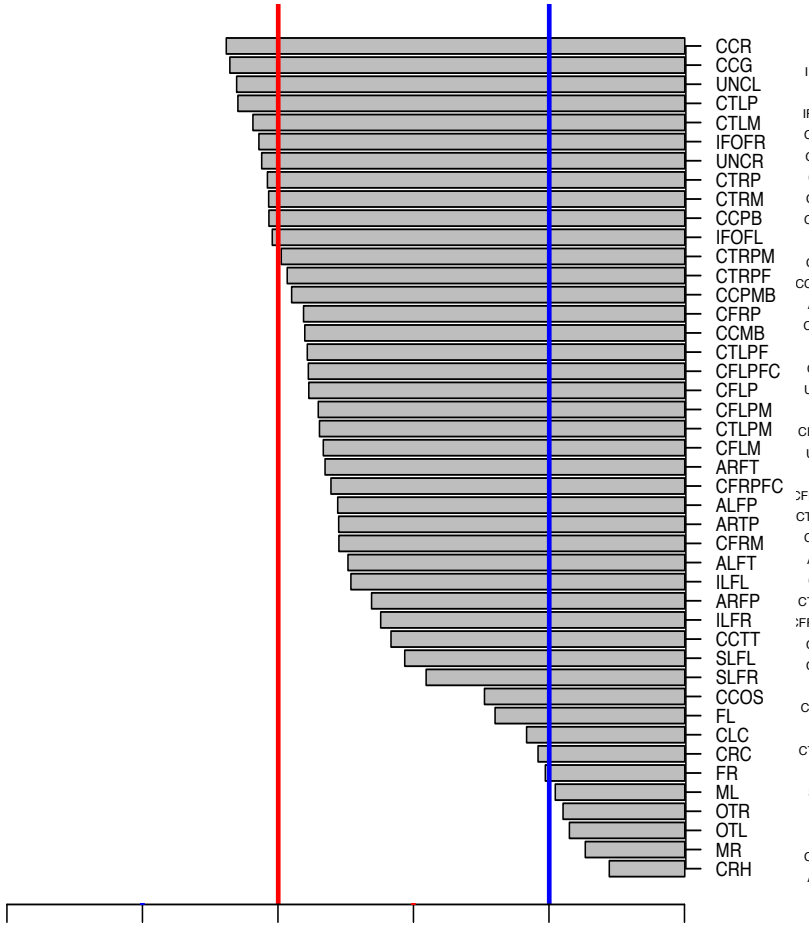


Loading Zoom

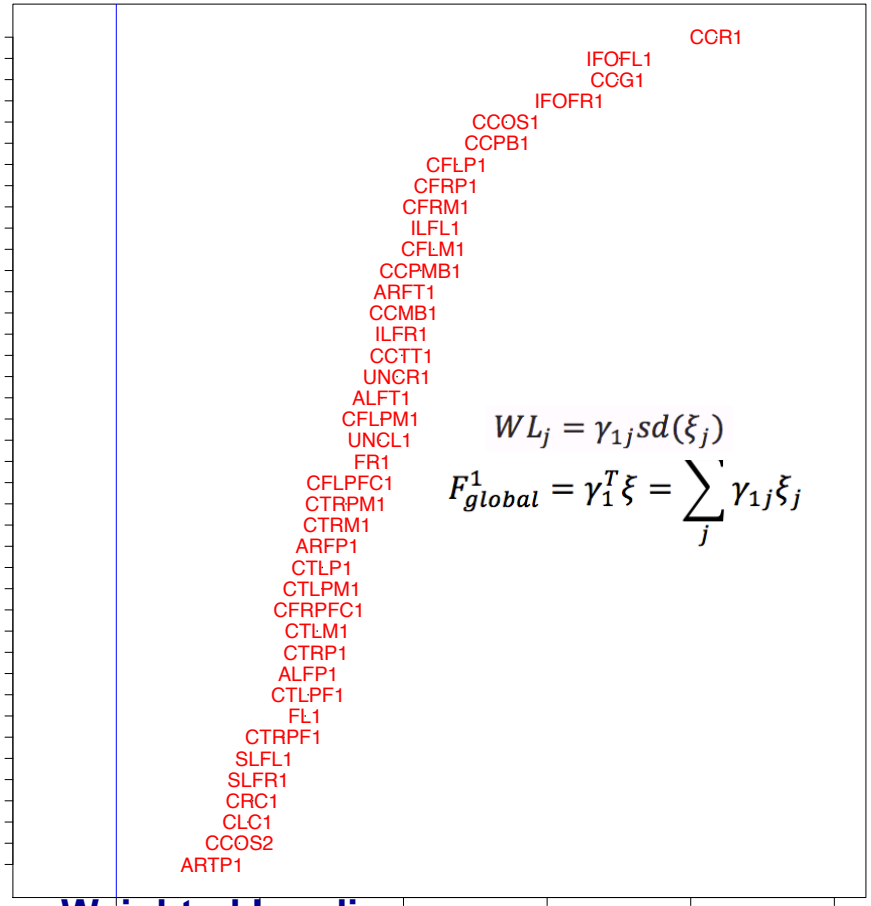




Global Factor in Real Data Analysis



Percent of variation explained by global factor

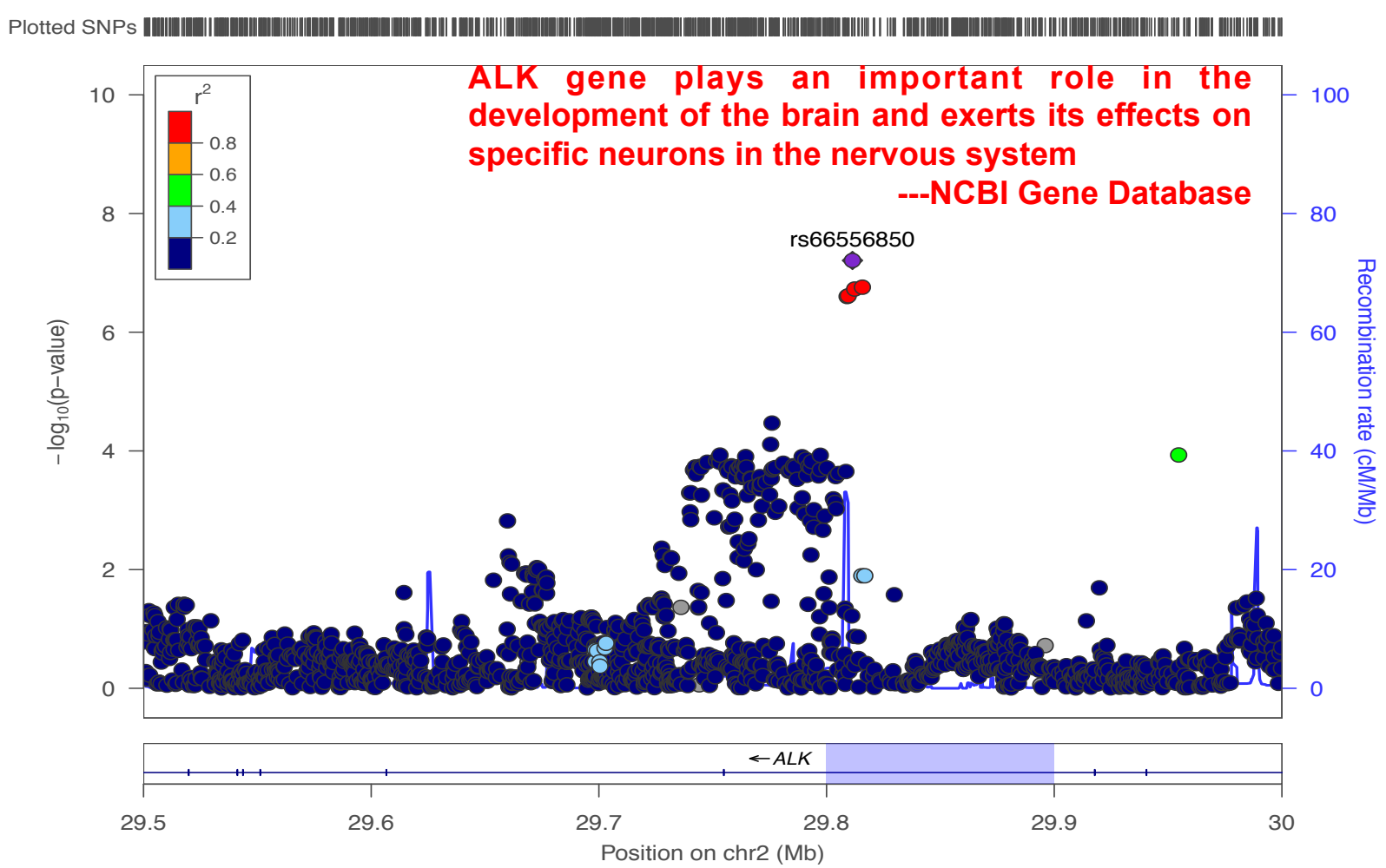


$$WL_j = \gamma_{1j} sd(\xi_j)$$

$$F^1_{global} = \gamma_1^T \xi = \sum_j \gamma_{1j} \xi_j$$

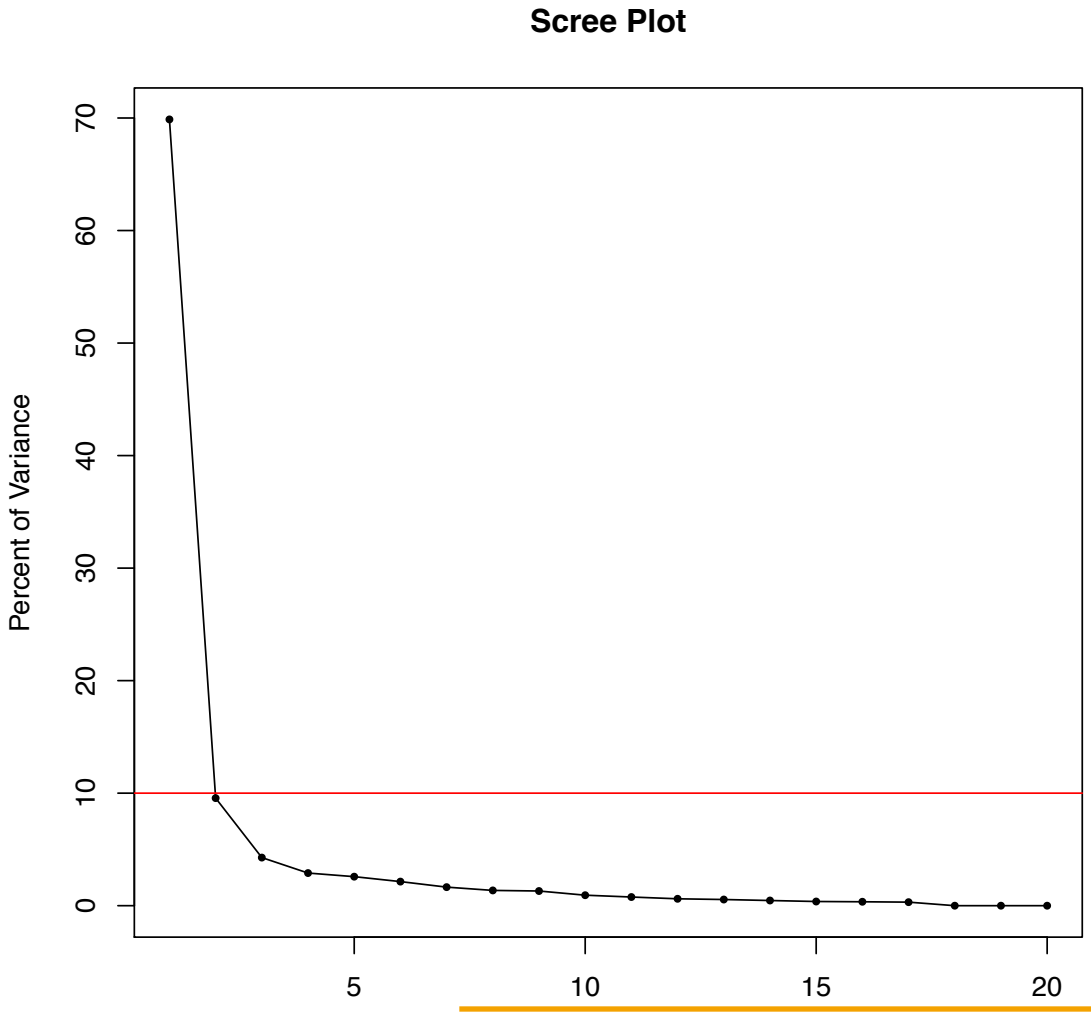
Weighted Loading

GWAS Result of global factor



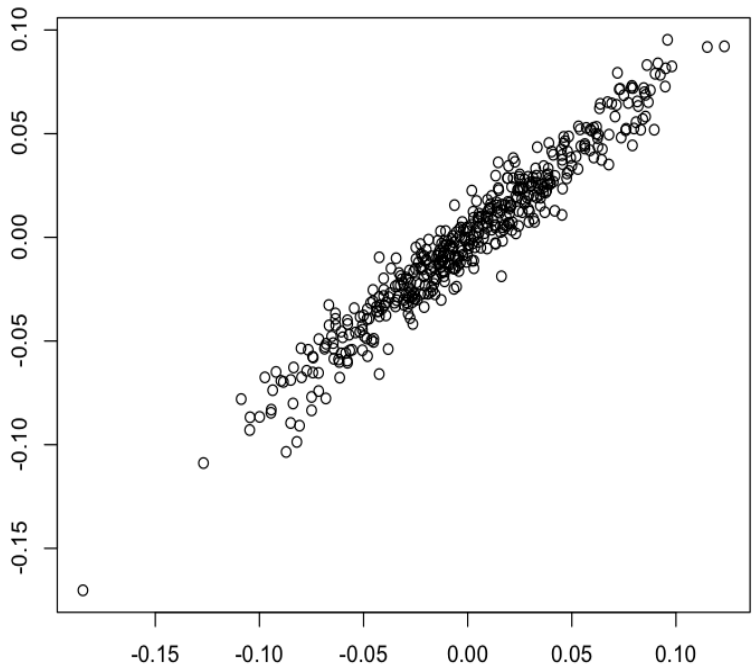
GWAS Result of global factor: TBSS

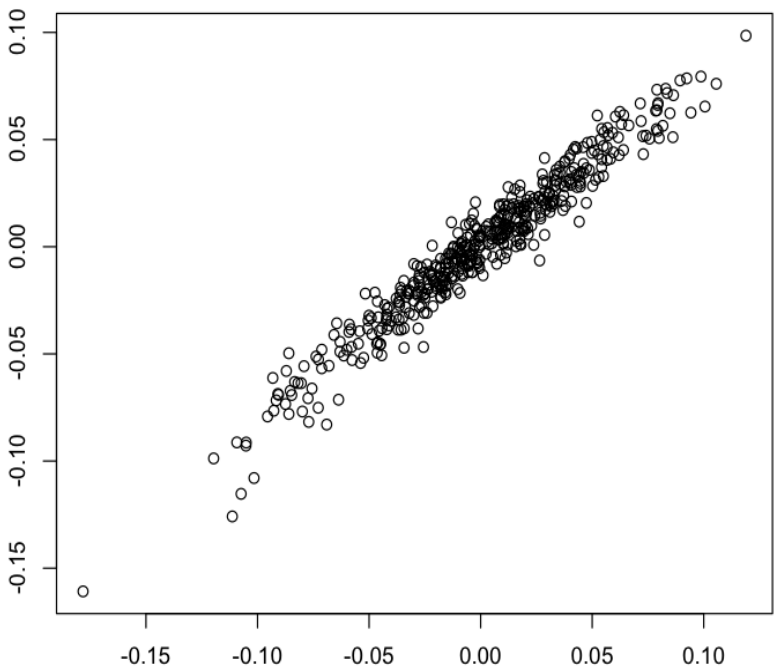
Factor Analysis



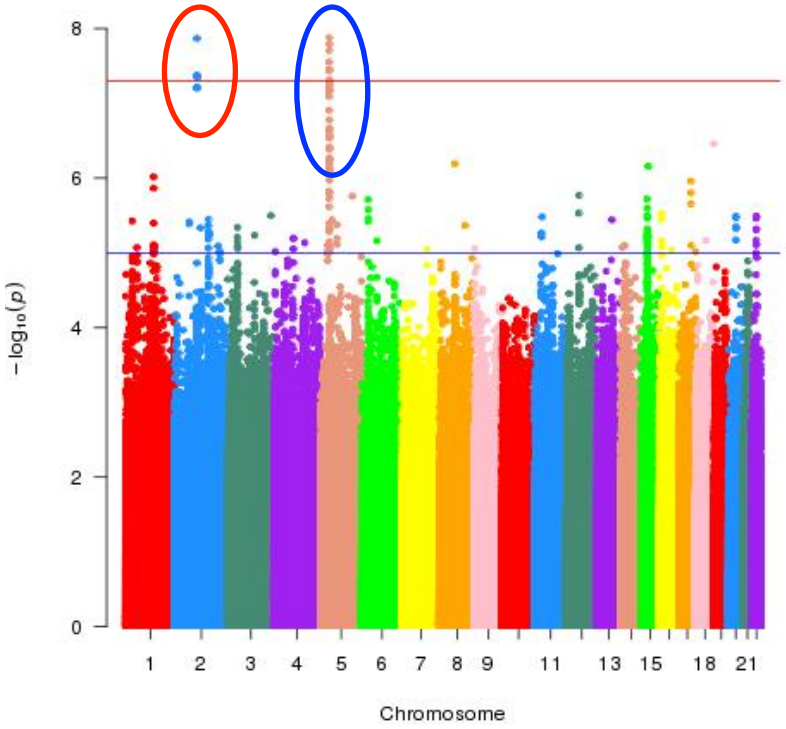
Global factors: Tracts versus TBSS

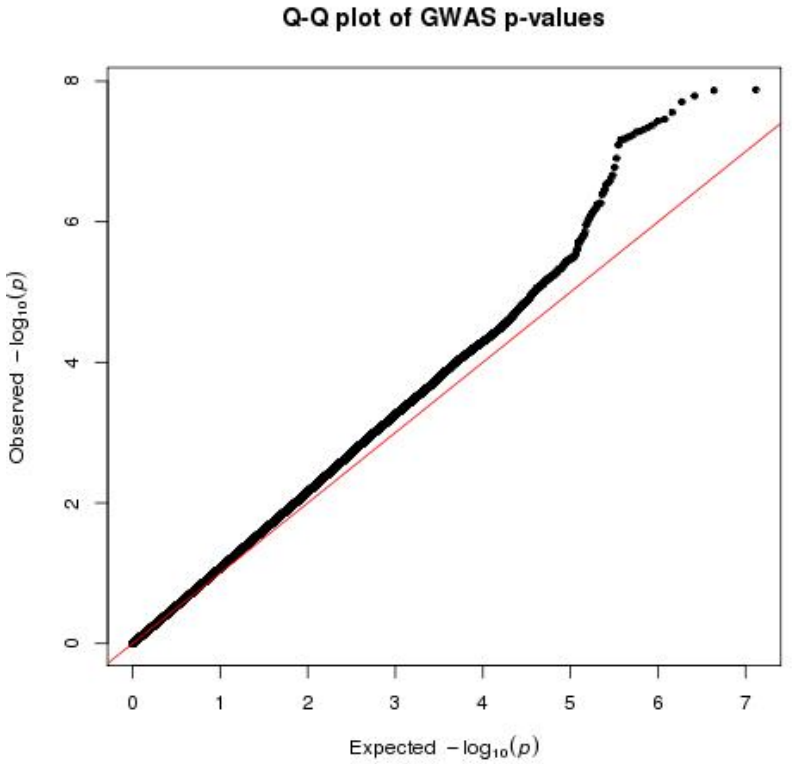
The global factors of DTI tracts and TBSS are highly correlated !



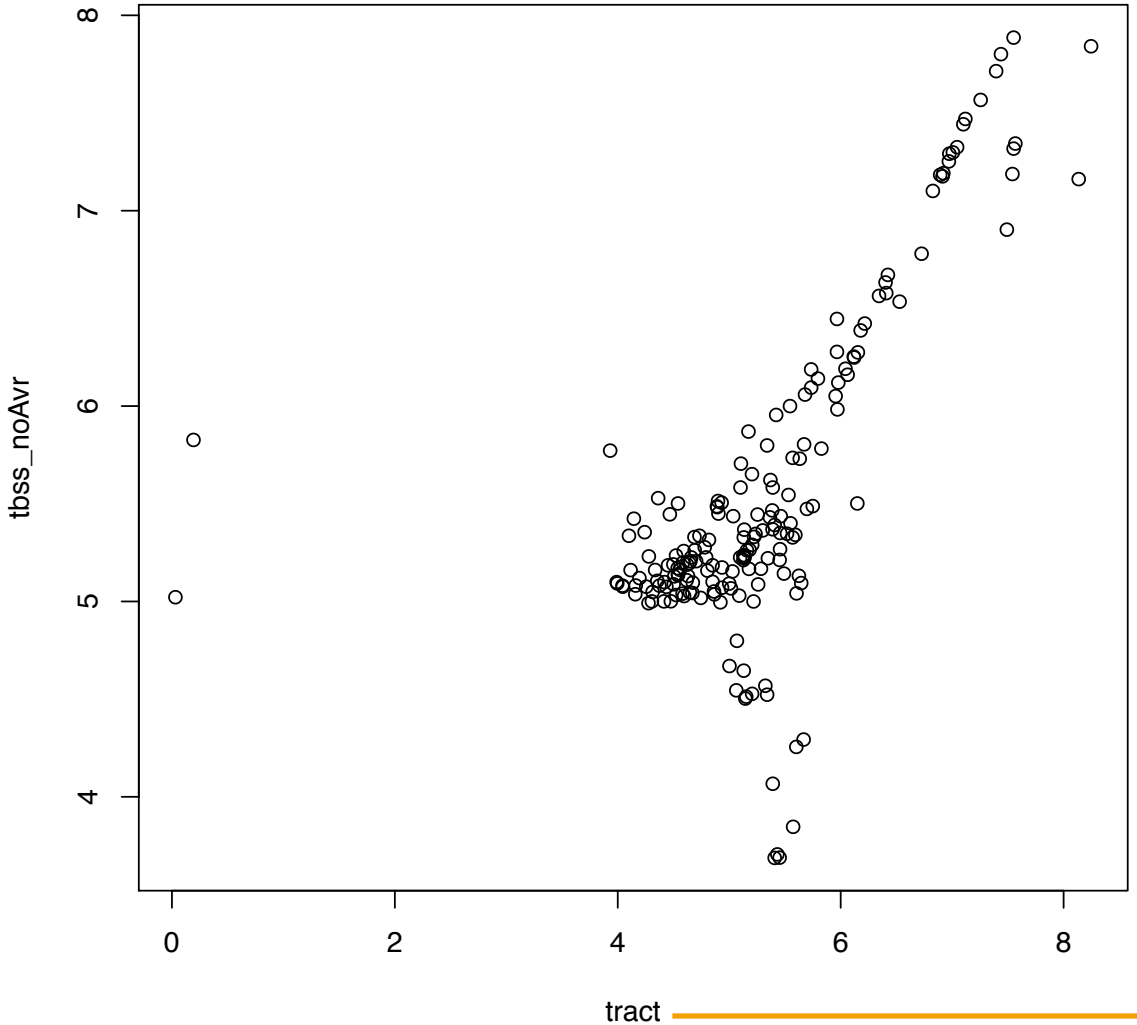


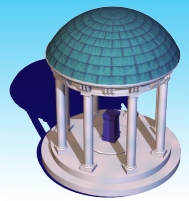
GWAS Result of global factor: TBSS





Comparison of Top Snps





GWAS Result of global factor

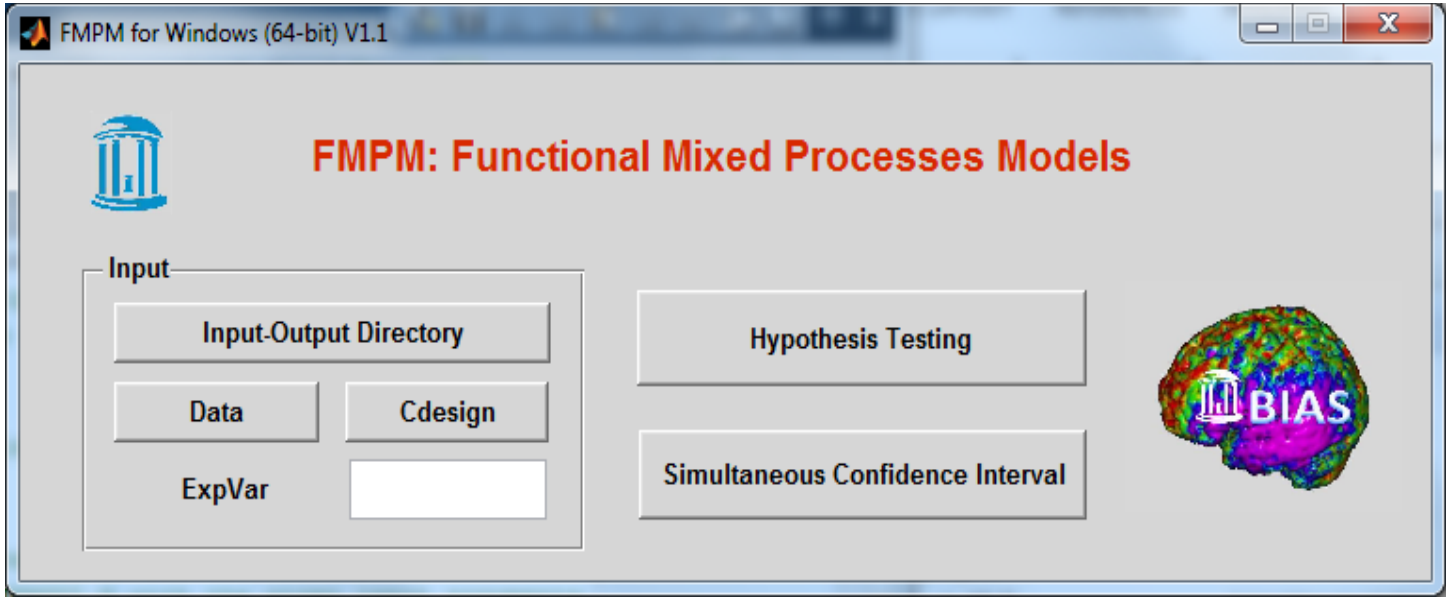
Top 20 SNPs and corresponding/proximal Genes							
snpname	chr	Rank in DTI	Pval in DTI	Rank in TBSS	Pval in TBSS	gene	Gene function
rs66556850	2	1	5.67E-09	2	1.44E-08		
rs62131138	2	2	7.32E-09	17	6.89E-08		
rs34328925	2	3	2.71E-08	8	4.53E-08	ALK	Brain development
rs34938026	2	4	2.81E-08	10	4.81E-08		
rs10167952	2	6	2.88E-08	15	6.49E-08		
rs6878826	5	5	2.81E-08	1	1.30E-08		
rs6866769	5	7	3.23E-08	18			
rs6878602	5	8	3.65E-08	3	1.58E-08		
rs6883230	5	9	4.03E-08	4	1.93E-08		
5:115008755:A_AGT	5	10	5.55E-08	5	2.71E-08	LOC102467217	
5:115008760:G_GTG	5	11	7.61E-08	6	3.39E-08	TMED7	
rs7705506	5	12	7.94E-08	7	3.61E-08	LOC10927100	
rs6594898	5	13	9.03E-08	9	4.73E-08	TICAM2	Progressive Multifocal
5:115009046:CA_C	5	14	9.85E-08	11	5.03E-08		Leukoencephalopathy
rs7732489	5	15	1.06E-07	12	5.11E-08		
rs7712289	5	16	1.07E-07	13	5.59E-08		
rs6594897	5	17	1.20E-07	14	6.43E-08		
rs6594896	5	18	1.22E-07	16	6.67E-08		
rs73116519	3	19	7.09E-07	55	3.15E-06		
rs72734794	15	20	8.66E-07	23	6.91E-07	UNC13C	infantile epileptic encephalopathy

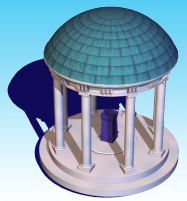
Risk Score

Common factor extracted from…	Tract	TBSS	TBSS - avrFA
P<1e-04	0.186767379	0.149947354	0.151419505
P<1e-03	0.223172509	0.183620164	0.18614227
P<1e-02	0.802213498	0.831979203	0.826575438
P<0.05	0.7030879	0.653264153	0.655184678
P<0.1	0.488391459	0.41296802	0.414686436
P<0.2	0.405096436	0.373412981	0.373995957
P<0.3	0.555212245	0.533781039	0.534005301
P<0.4	0.454465118	0.418048888	0.418313965
P<0.5	0.537275523	0.472857816	0.472769655
P<1	0.61317957	0.523945656	0.523566001

FADTTS

- http://www.nitrc.org/projects/fadtts
- FMPM GUI is a MATLAB graphical user interface

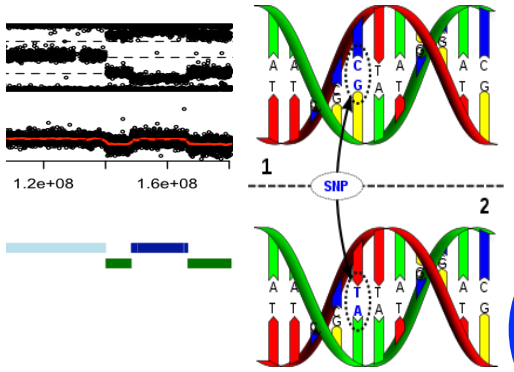




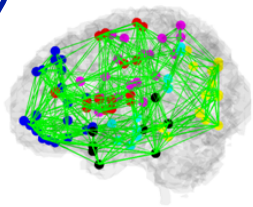
FFGWAS: Fast Functional Genome Wide Association AnalysiS of Surface-based Imaging Genetics

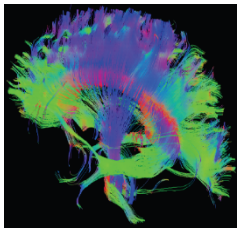
Imaging Genetics

E: environmental factors



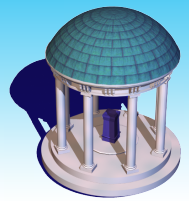
G: genetic markers





D: disease

http://en.wikipedia.org/wiki/DNA_sequence



Reading Materials

1. **Lin, J., Zhu, H.T.,** Knickmeyer, R., Styner, M., Gilmore, J. H. and Ibrahim, J.G. (2012). Projection Regression Models for Multivariate Imaging Phenotype. *Genetic Epidemiology*, 36, 631-641.
2. **Lin, J**., ***Zhu, H.T.,*** , Mihye, A., and Ibrahim, J.G. (2014). Functional Mixed Effects Models for Candidate Genetic Mapping in Imaging Genetic Studies. *Genetic Epidemiology,* 38(8):680-91.
3. **Zhu, H.T.,** Khondker, Z. S., Lu, Z.H., and Ibrahim, J. G. (2014). Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *Journal of American Statistical Association,* 507, 977-990.
4. **Zhu, HT,** Fan, J., and Kong, L. (2014). Spatial varying coefficient model and its applications in neuroimaging data with jump discontinuity. *Journal of American Statistical Association,* 109, 1084-1098.
5. **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, 2015.
6. 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 press.

Video

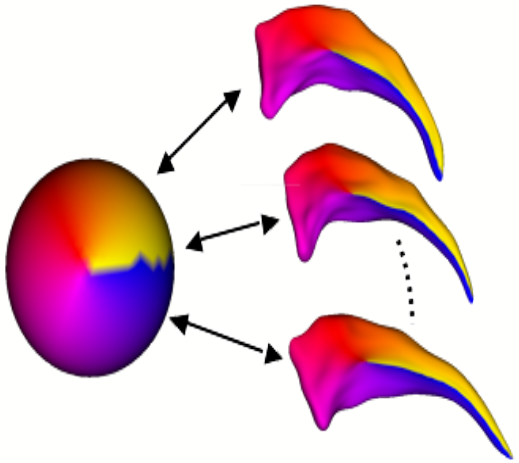
http://www.birs.ca/events/2016/5-day-workshops/16w5036/videos/watch/201602021521-Huang.html

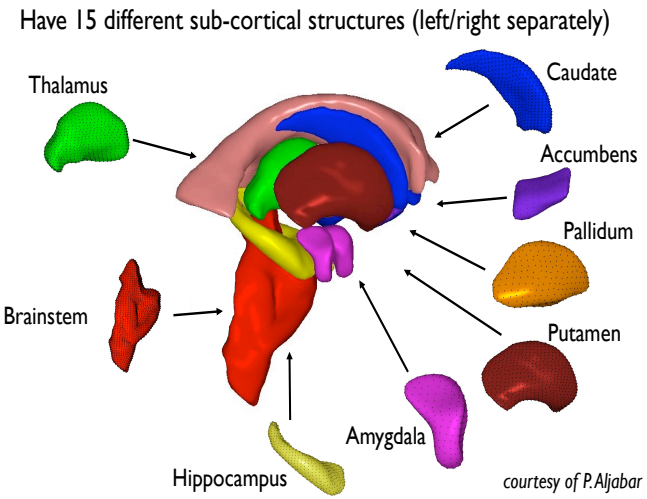
Statistical Methods

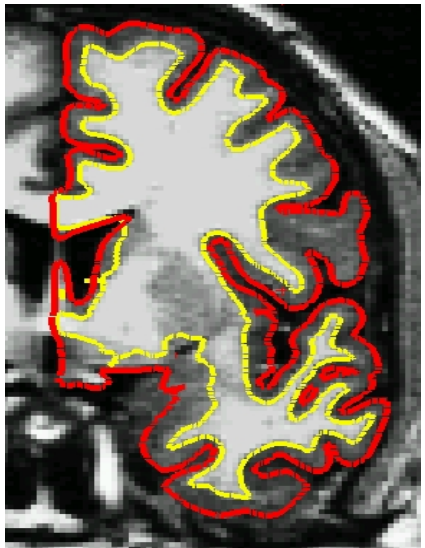
Genetics \ Imaging	Candidate ROI	Many ROI	Voxelwise
Candidate SNP	Imager	Imager	Imager
Candidate Gene	Geneticist	↑	↑
Genome-wide SNP	Geneticist	↑	↑
Genome-wide Gene	Geneticist		

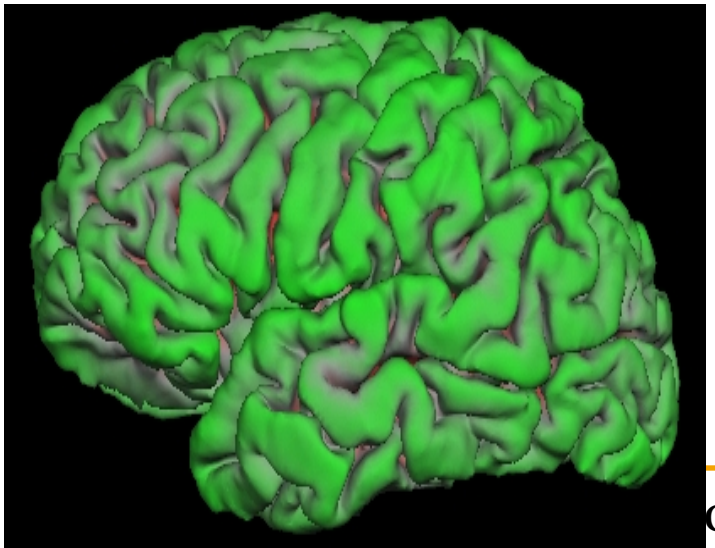
Hibar, et al. HBM 2012

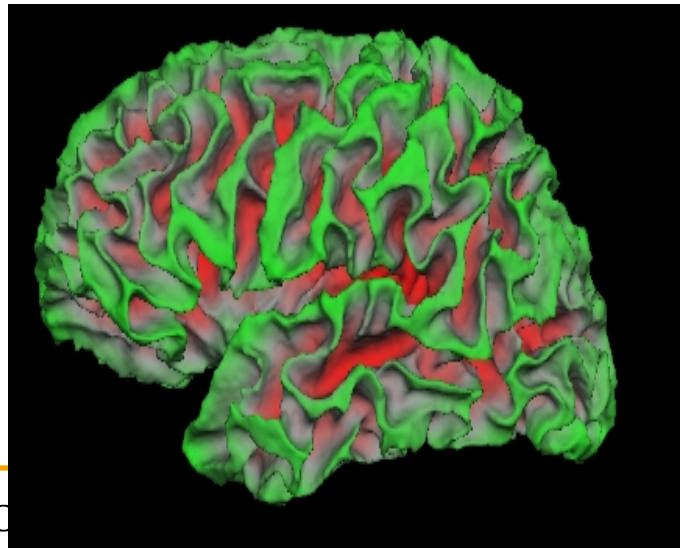
Data Structure









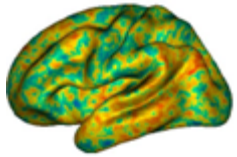


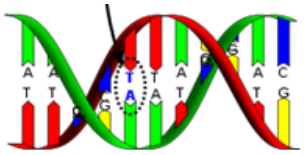
High Dimensional Regression Model

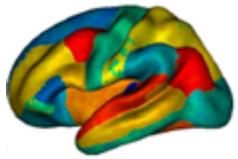
Data $\{(Y_i, X_i): i = 1, \cdots, n\}$

$$Y_i = \{y_i(v): v \in V\} \qquad X_i = \{X_i(g): g \in G_0\}$$

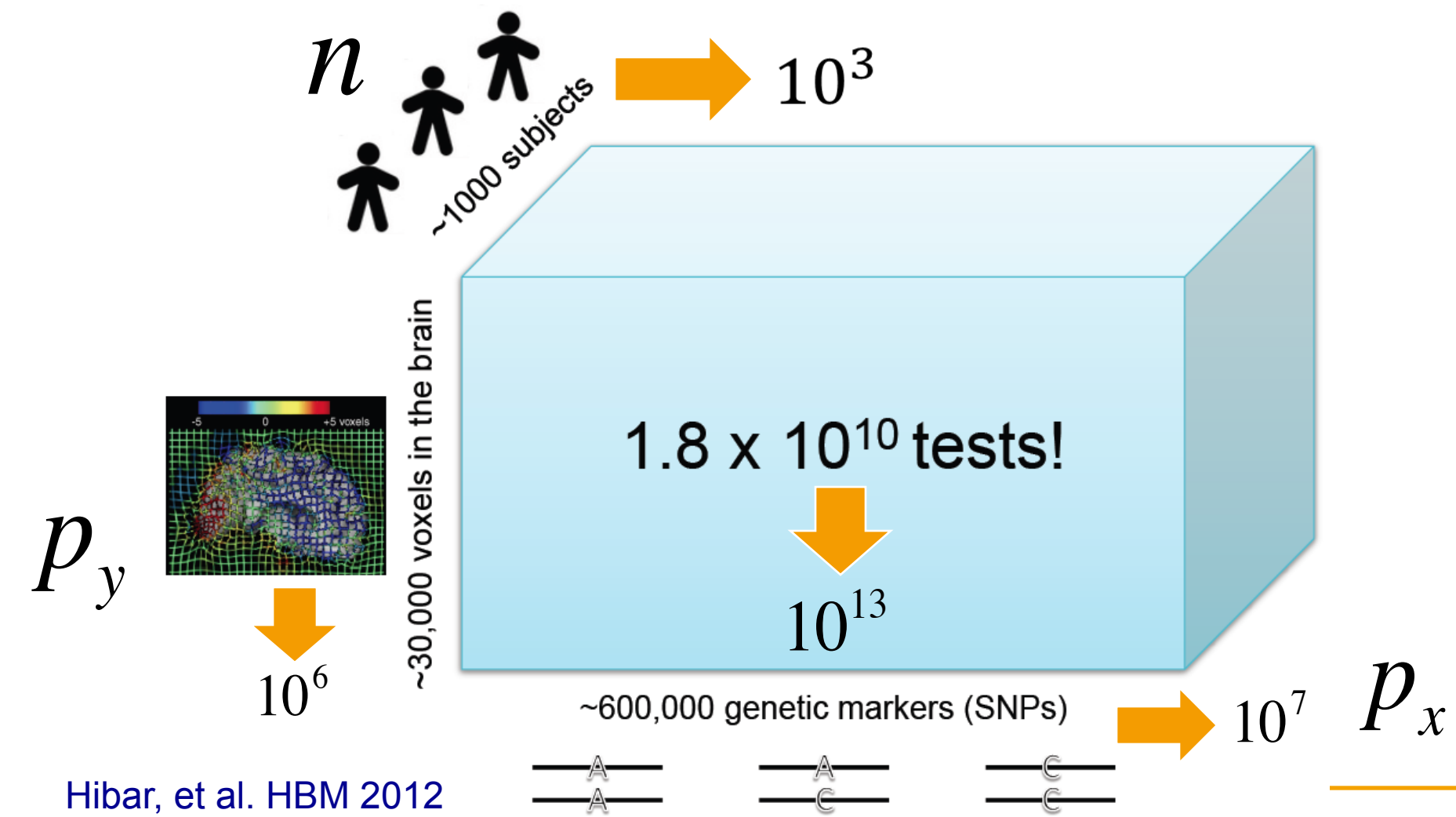
Phenotype		Genotype		Error
Y		X	B	E
$n \times p_y$	$=$	$n \times p_x$	$p_x \times p_y$	$+$ $n \times p_y$







Challenges



Hibar, et al. HBM 2012

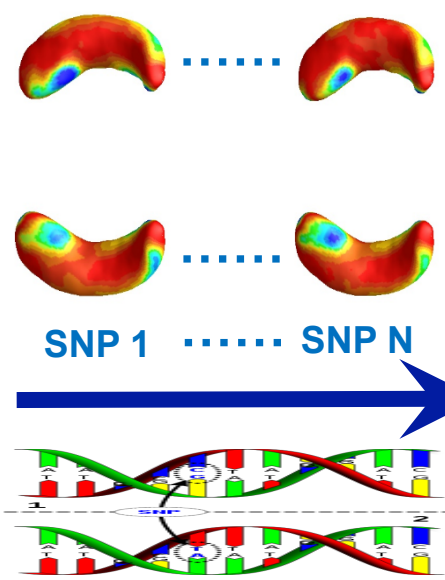
Fast Functional GWAS

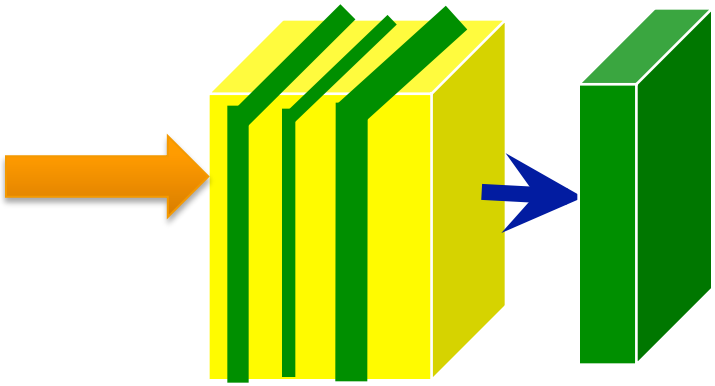
Functional Models

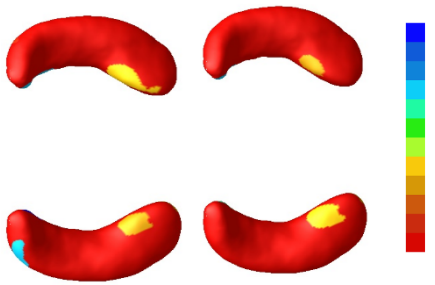
Smoothing

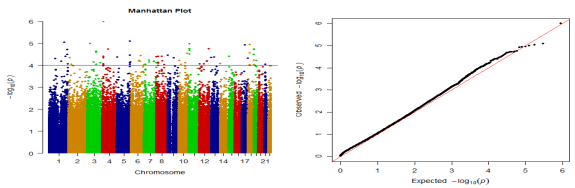
Global Sure Independence Screening Procedure

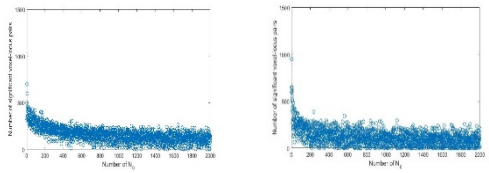
Significant Voxel-locus Cluster-locus Detection









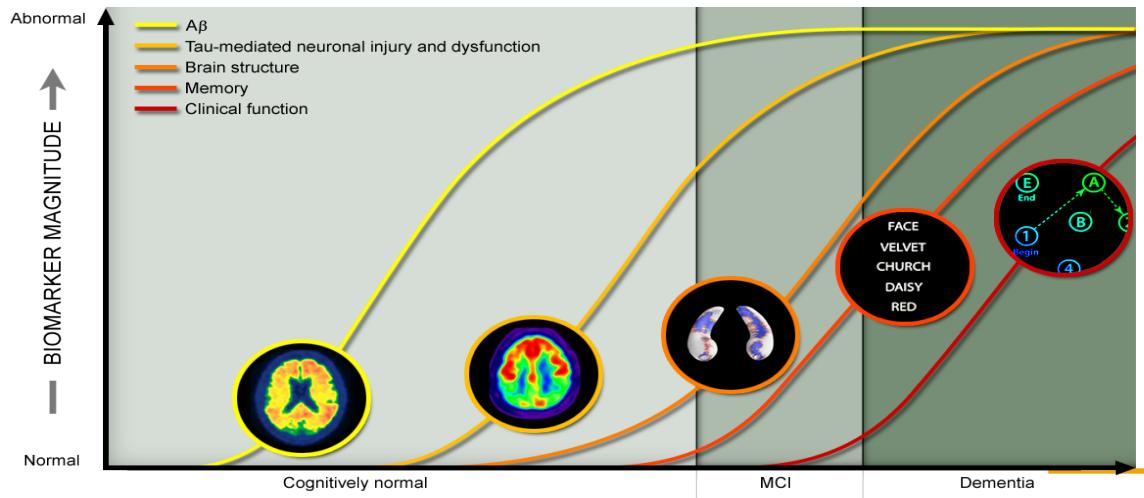


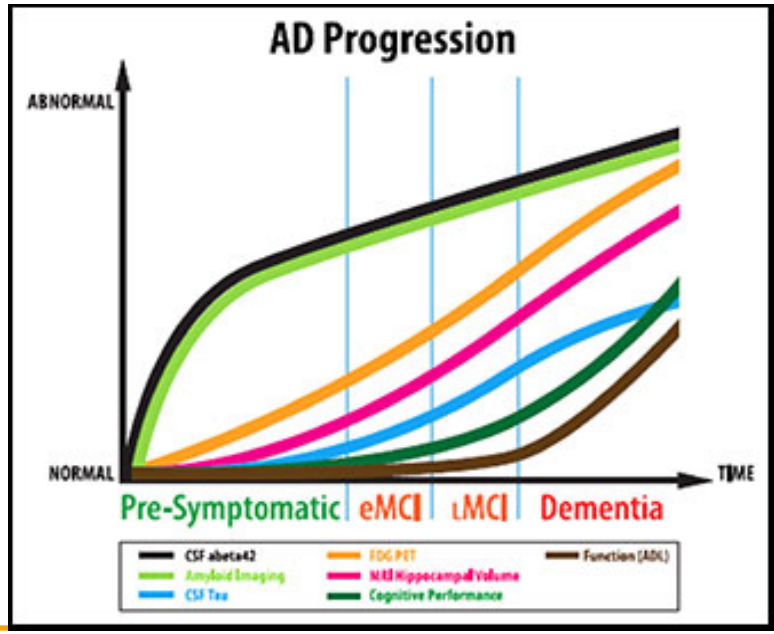
Imaging Genetics for ADNI

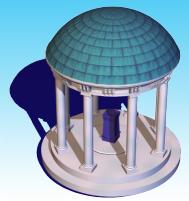
PI: Dr. Michael W. Weiner

- detecting AD at the earliest stage and marking its progress through biomarkers;
- developing new diagnostic methods for AD intervention, prevention, and treatment.

- A longitudinal prospective study with 1700 aged between 55 to 90 years
- Clinical Data including Clinical and Cognitive Assessments
- Genetic Data including Ilumina SNP genotyping and WGS
- MRI (fMRI, DTI, T1, T2)
- PET (PIB, Florbetapir PET and FDG-PET)
- Chemical Biomarker



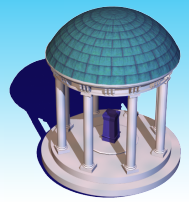




ADNI Data Analysis: Dataset Description

- 708 MRI scans of AD (186), MCI (388), and healthy controls (224) from ADNI-1.
- These scans on 462 males and 336 females are performed on a 1.5 T MRI scanners.
- The typical protocol includes the following parameters:
  (i) repetition time (TR) = 2400 ms;
  (ii) inversion time (TI) = 1000 ms;
  (iii) flip angle = 8^{o};
  (iv) field of view (FOV) = 24 cm with a 256 x 256 x 170 acquisition matrix in the x−, y−, and z−dimensions,
  (v) voxel size: 1.25 x 1.26 x 1.2 mm^{3}.

- Covariates: gender, age, APOE ε4, and the top 5 PC scores in SNPs



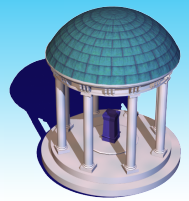
Imaging Data Preprocessing

Surface fluid registration based hippocampal sub-regional analysis package (Shi et al., Neuroimage, 2013)

- **Hippocampal surface registration**
  isothermal coordinates and fluid registration

- **Surface statistics computation**
  1. **multivariate tensor-based morphometry (mTBM) statistics**
  2. **radial distance**

Finally, we obtained left and right hippocampus shape representations as 100×150 matrices.



Our computational time
About 92,000 s

ADNI Data Analysis

Top 10 SNPs (Left Hippocampus)

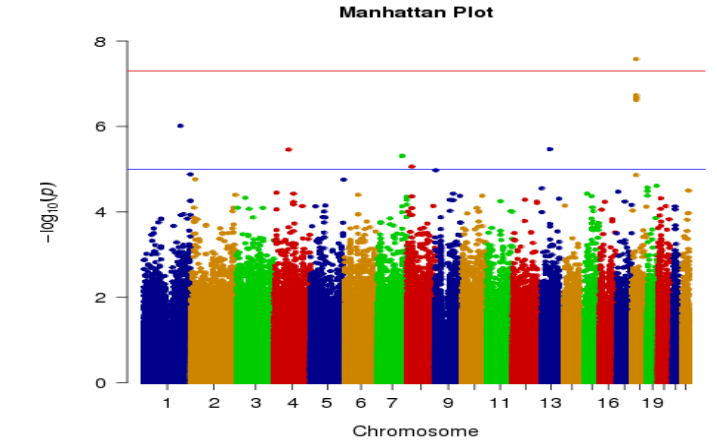
SNP	CHR	BP	-LOG 10(p)
rs657132	18	2.20533e+07	7.579767
rs604345	18	2.20033e+07	6.729377
rs582110	18	2.19954e+07	6.672876
rs546000	18	2.20031e+07	6.672876
rs489631	18	2.1989e+07	6.620395
rs16837577	1	1.94871e+08	6.016773
rs3812872	13	6.19869e+07	5.468391
rs6826085	4	7.68702e+07	5.459163
rs929714	7	1.3263e+08	5.314317
rs2042067	7	1.32651e+08	5.306583

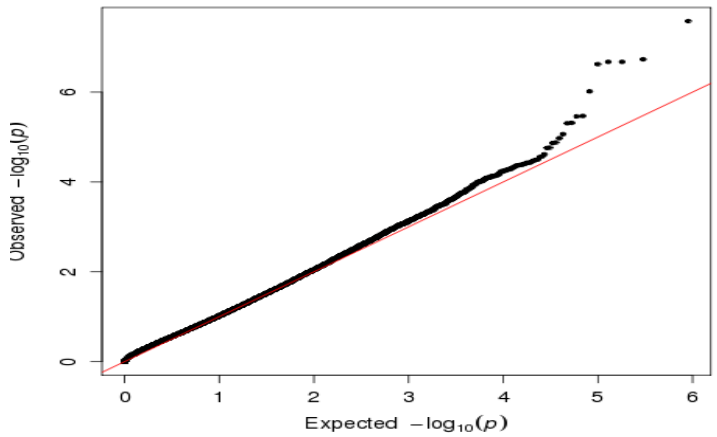
Top 10 SNPs (Right Hippocampus)

SNP	CHR	BP	-LOG 10(p)
rs4681527	3	1.44e+08	6.764886
rs3108514	2	1.51279e+08	6.274511
rs12264728	10	1.3214e+08	5.961976
rs652911	10	1.3214e+08	5.739661
rs10801705	1	8.95004e+07	5.622668
rs366346	10	1.32141e+08	5.617185
rs7312068	12	2.94352e+07	5.604041
rs7617465	3	1.43999e+08	5.522112
rs17605251	7	1.02746e+08	5.486603
rs749788	2	2.84618e+06	5.474675

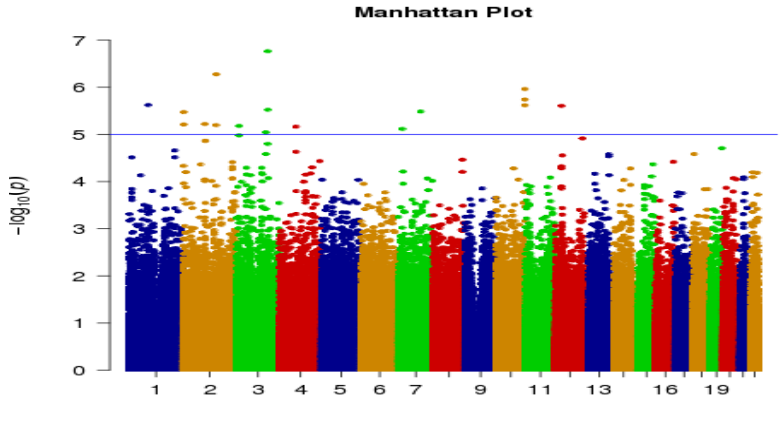
ADNI Data Analysis

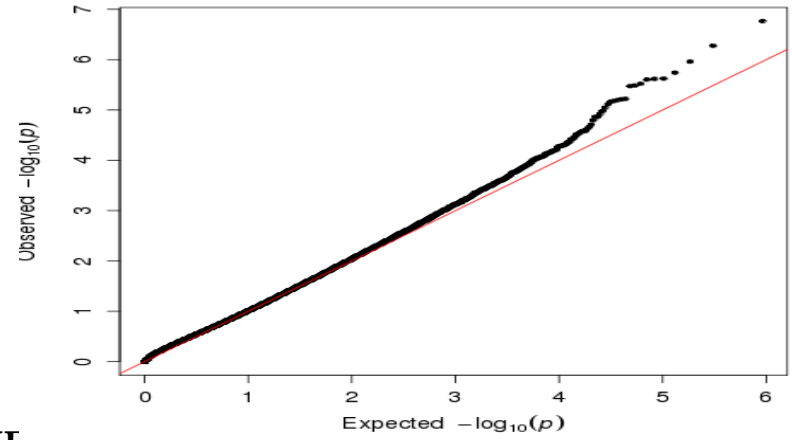
(Left Hippocampus)





(Right Hippocampus)

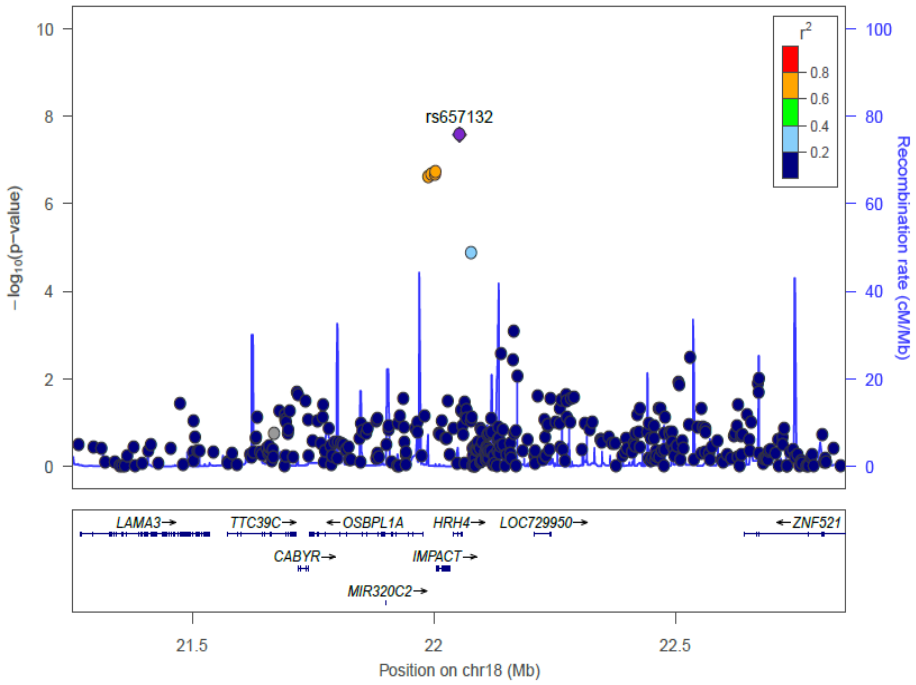




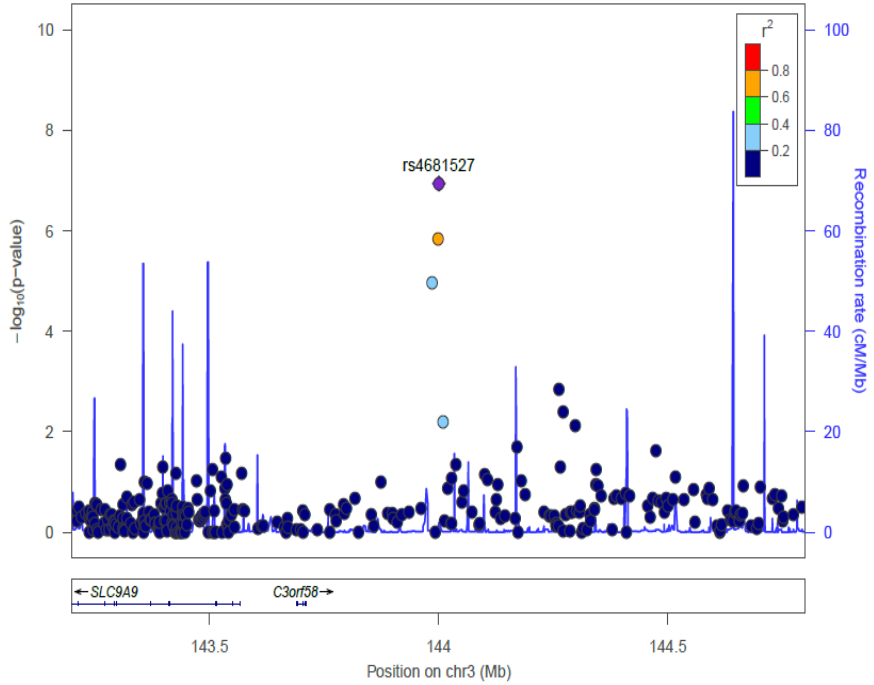
ADNI Data Analysis: Left Hippocampus

Significant Loci Zoom

(Left Hippocampus)



(Right Hippocampus)



ADNI Data Analysis: Left Hippocampus

(Left Hippocampus)

Top 1 SNP: rs657132
Closed Gene: HRH4

HRH4 (Histamine Receptor H4) is a Protein Coding gene.
Diseases associated with HRH4: cerebellar degeneration
An important paralog of this gene: CHRM4

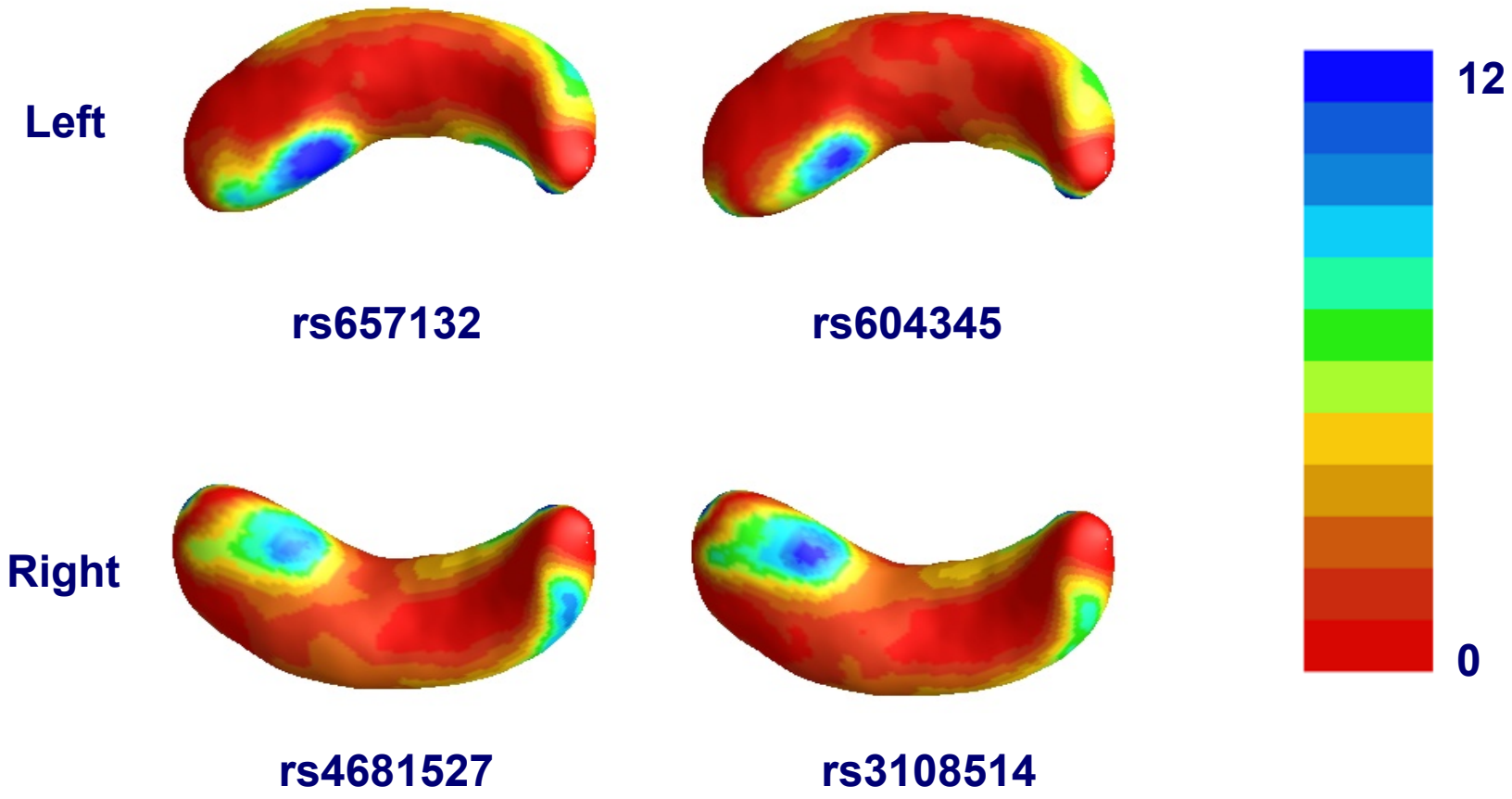
Mirshafiey & Naddafi, Am J Alzheimers Dis Other Demen. 2013

(Right Hippocampus)

Top 1 SNP: rs4681527
Closed Gene: C3orf58

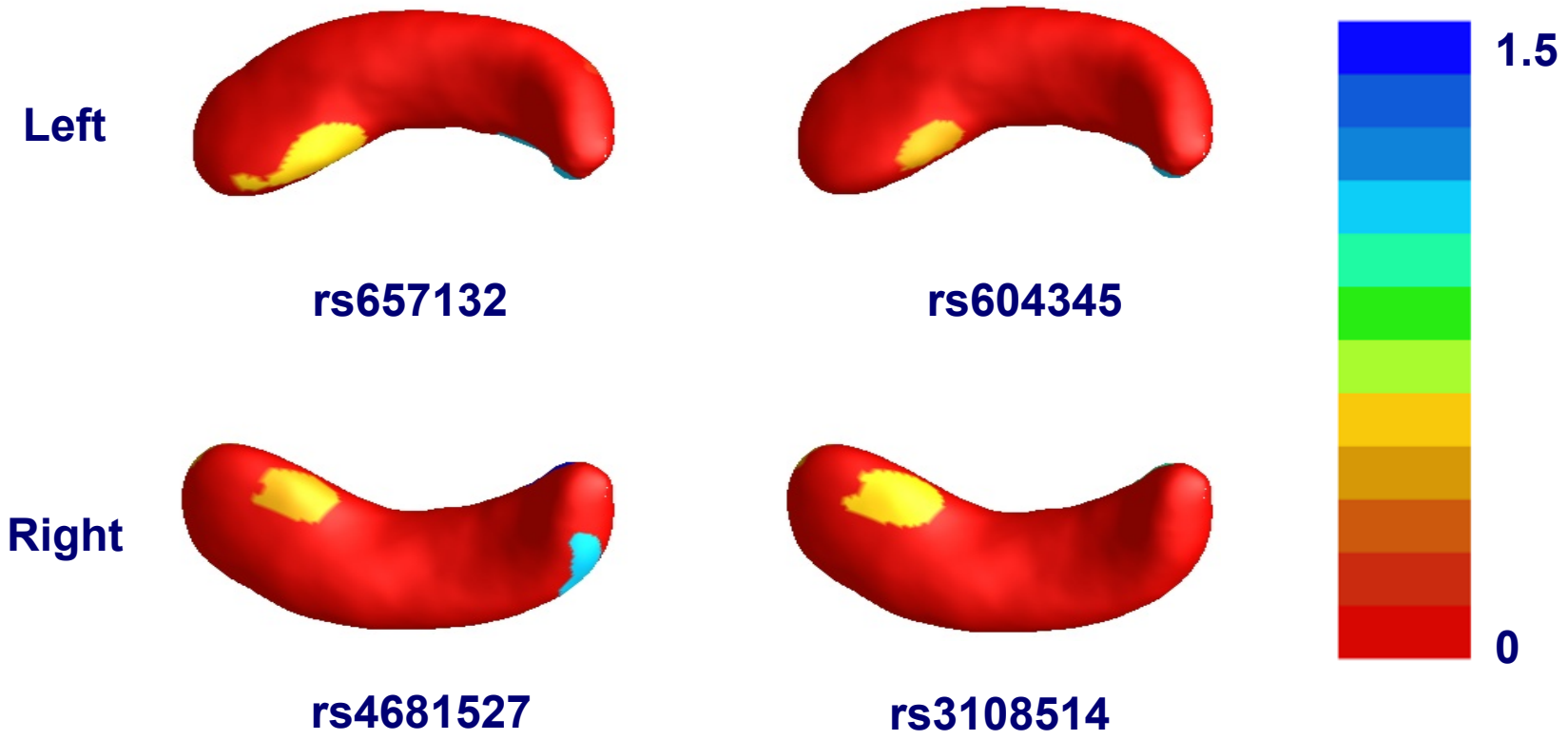
C3orf58 (Chromosome 3 Open Reading Frame 58) is a Protein Coding gene.
Diseases associated with C3orf58: hypoxia

ADNI Data Analysis



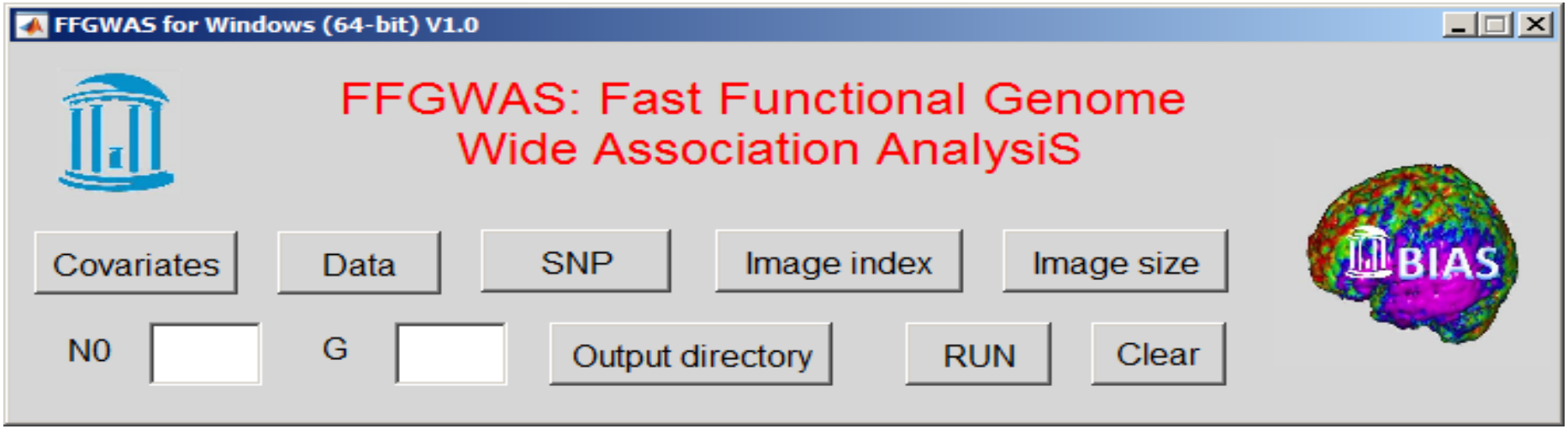
$-\log_{10}(p)$ values on Hippocampus (L & R) corresponding to Top 2 SNPs

ADNI Data Analysis



$-\log_{10}(p)$ values of significant clusters on Hippocampus (L & R) corresponding to Top 2 SNPs

A Software for FFGWAS



SAMSI

- **2013 Neuroimaging Data Analysis**
- **2015-2016 Challenges in Computational Neuroscience**

2016 Banff Birds Neuroimaging Data Analysis

Thank You!!