Adaptive Mantel Test for Penalized Inference, with Applications to Imaging Genetics

Dustin Pluta, Hernando Ombao, Zhaoxia Yu

01 AUG 2018

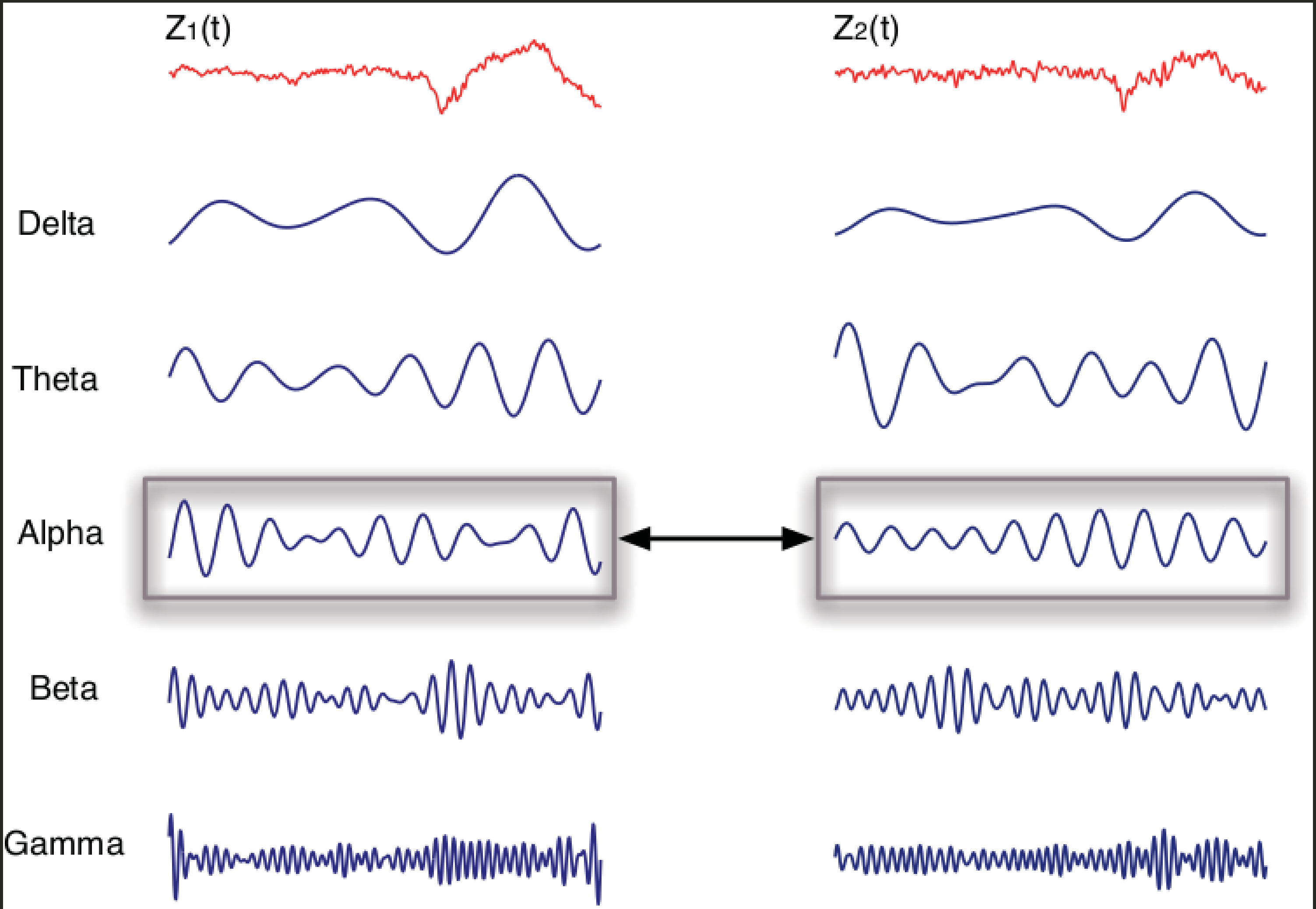
Acknowledgements

- **Gui Xue**, PI and data provider, Beijing Normal University, Center for Brain and Learning Sciences
- **Chuansheng Chen**, UCI, Dept. of Psychology and Social Behavior
- **Hernando Ombao**, KAUST
- **Zhaoxia Yu**, UCI, Dept. of Statistics
- **Tong Shen**, UCI, Dept. of Statistics (PhD Student)

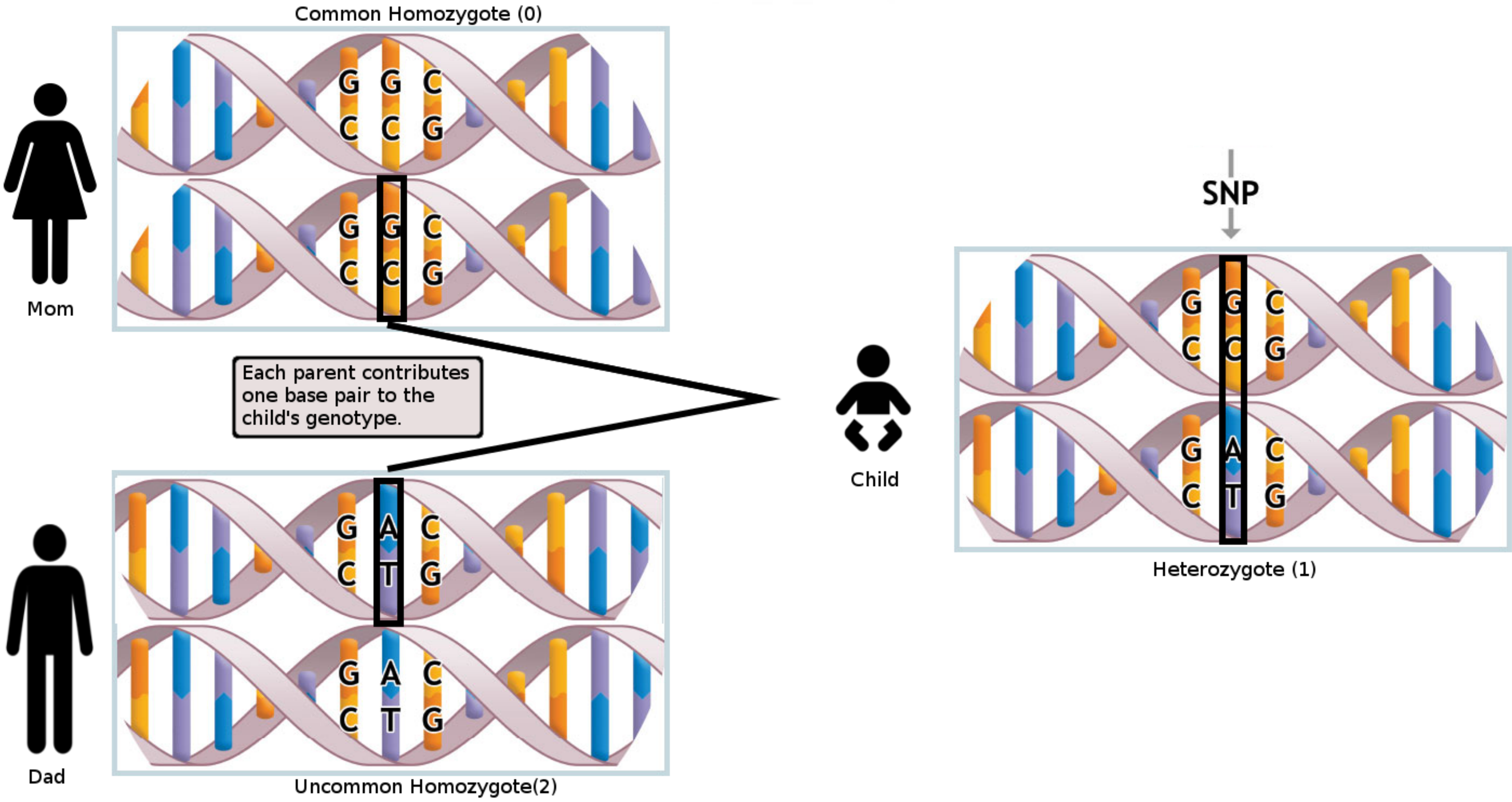
Overview of Talk

1. Scientific background of **connectome genetics**.
2. **Adaptive Mantel test** and metric-based association testing.
3. Heritability analysis of EEG coherence during a visual working memory task.

Functional Connectivity: Coherence



SNPs


Common Homozygote (0)
Mom
G G C
C C G
G G C
C C G
Each parent contributes one base pair to the child's genotype.
Dad
G A C
C T G
G A C
C T G
Uncommon Homozygote(2)
Child
SNP
G G C
C C G
G A C
C T G
Heterozygote (1)

Scientific Background

Estimating Heritability with Variance Components Model

- A popular approach to estimate narrow-sense heritability of a phenotype is the *variance components* (*random effects*) model.

$$Y = Xb + \varepsilon,$$

with

- $\mathrm{Var}(Y) = \sigma_b^2 XX^T + \sigma_\varepsilon^2 I_n,$
- $b \sim N(0, \sigma_b^2 I_p)$ is a random vector of SNP effects,
- $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$ residual vector,
- X is the SNP data matrix (column centered and scaled). Number of features in X may be on the order of 10^5.

Scientific Background

Estimating Heritability with Variance Components Model

- The above random effects model can be rewritten as

$$Y = g + \varepsilon,$$

where $g \sim N(0, \sigma_g^2 G)$, for $G = XX^T/p$ and $\sigma_g^2 = p\sigma_b^2$.

Narrow-sense heritability of the phenotype measured by Y can then be estimated as

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_\varepsilon^2}.$$

Scientific Background

Multivariate Heritability

Following Ge et al. (2016), a multivariate version of h^2 can be defined for when Y is a phenotype vector (or matrix) for each subject, such as a connectivity matrix.

$$h^2 = \frac{\mathrm{tr}(\Sigma_g)}{\mathrm{tr}(\Sigma_g) + \mathrm{tr}(\Sigma_\varepsilon)}.$$

When both X and Y are standardized and scaled so that $tr(XX^T) = tr(YY^T) = n$, a method of moments estimator for h^2 is

$$\hat{h}^2_{MOM} = \frac{\mathrm{tr}(YY^T XX^T) - n}{\mathrm{tr}(XX^T XX^T) - n}.$$

Metric-based Association Testing

The Inference Goal

Given observations of n subjects across two data modalities $\mathbf{X}$ and $\mathbf{Y}$, is similarity in $\mathbf{X}$ significantly associated with similarity in $\mathbf{Y}$?

Setup

- In our application, $X \in \mathrm{X}^n$ is an $n \times p$ matrix of SNP measurements, and $Y \in \mathrm{Y}$ is an $n \times 1$ vector of scalar phenotype measurements.
- Assume X and Y have been column centered and scaled.
- Let $\mathcal{K}$ be a bounded, symmetric, positive semi-definite similarity function, e.g. $\mathcal{K}(u, v) = u^T v$.

Mantel Test

- Given **similarity functions** $\mathcal{K}_X : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R}$ and $\mathcal{K}_Y : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we can form two $n \times n$ **Gram matrices** K and H, where

$$K_{ij} = \mathcal{K}_X(X_i, X_j)$$

$$H_{ij} = \mathcal{K}_Y(Y_i, Y_j).$$

- The **correlation** of these distance matrices is

$$r(H, K) := \frac{\langle K, H \rangle}{\|K\| \cdot \|H\|},$$

- In our application, X_i is the vector of SNPs for subject i, and Y_i is an observed phenotype, which could be a scalar, vector, matrix, graph, or function (or anything else that admits a sensible similarity measure).

- We will focus on using similarity measures based on the Euclidean inner product, but any similarity function can be used for $\mathcal{K}$.

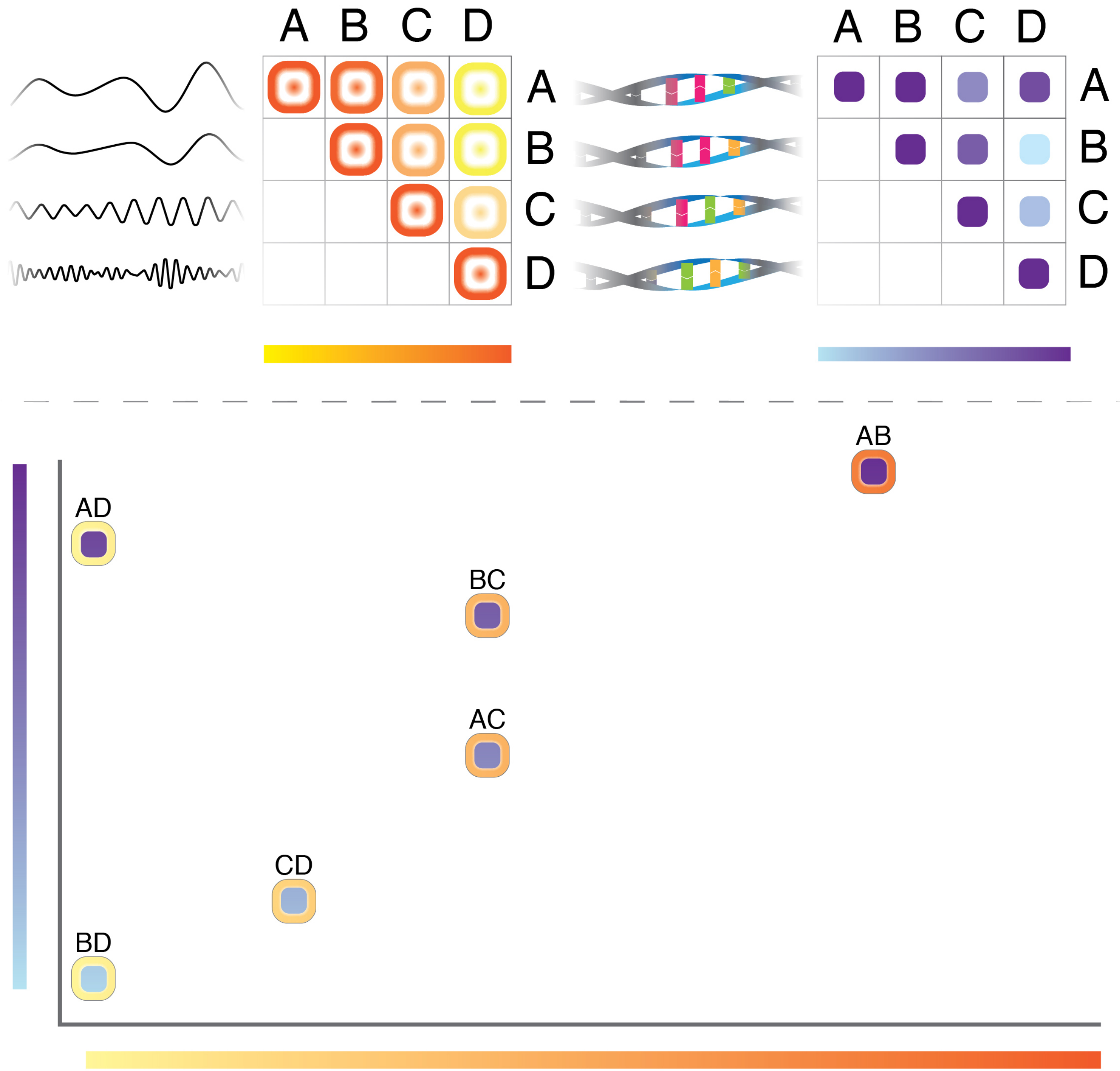


Image: Heno Wang, Scientific Illustrator, KAUST

GRM

$$\begin{matrix} K_{11} & K_{12} & \dots & K_{1N} \\ K_{21} & K_{22} & \dots & K_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ K_{N1} & K_{N2} & \dots & K_{NN} \end{matrix}$$

$$\begin{matrix} K_{11} \\ K_{12} \\ \vdots \\ K_{N(N-1)} \\ K_{NN} \end{matrix}$$

Connectivity Similarity

$$\begin{matrix} H_{11} & H_{12} & \dots & H_{1N} \\ H_{21} & H_{22} & \dots & H_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ H_{N1} & H_{N2} & \dots & H_{NN} \end{matrix}$$

$$\begin{matrix} H_{11} \\ H_{12} \\ \vdots \\ H_{N(N-1)} \\ H_{NN} \end{matrix}$$

$$Z^* = \langle K, H \rangle \equiv \sum_i \sum_j K_{ij} H_{ij}$$

Kernel Mantel Test

Similarity with Weighted Inner Products

For two vectors $u, v \in \mathbb{R}^p$, the **weighted inner product** $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ for some positive semi-definite matrix $\mathcal{W}$, is defined as

$$\langle u, v \rangle_{\mathcal{W}} = u^T \mathcal{W} v.$$

The **Mantel Test Statistic** for similarity $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ is

$$T_{\mathcal{W}} = \mathrm{tr}(X \mathcal{W} X^T Y Y^T) = Y^T X \mathcal{W} X^T Y.$$

Kernel Mantel Test

Similarity with Weighted Inner Products

For two vectors $u, v \in \mathbb{R}^p$, the **weighted inner product** $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ for some positive semi-definite matrix $\mathcal{W}$, is defined as

$$\langle u, v \rangle_{\mathcal{W}} = u^T \mathcal{W} v.$$

The **Mantel Test Statistic** for similarity $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ is

$$T_{\mathcal{W}} = \mathrm{tr}(X \mathcal{W} X^T Y Y^T) = Y^T X \mathcal{W} X^T Y.$$

Questions

How does the choice of weight matrix affect the test characteristics?

How should the weight matrix be chosen?

Similarity Measures

Euclidean Inner Product

- Choosing $\mathcal{W} = I_p$ gives

$$K_E = XX^T,$$

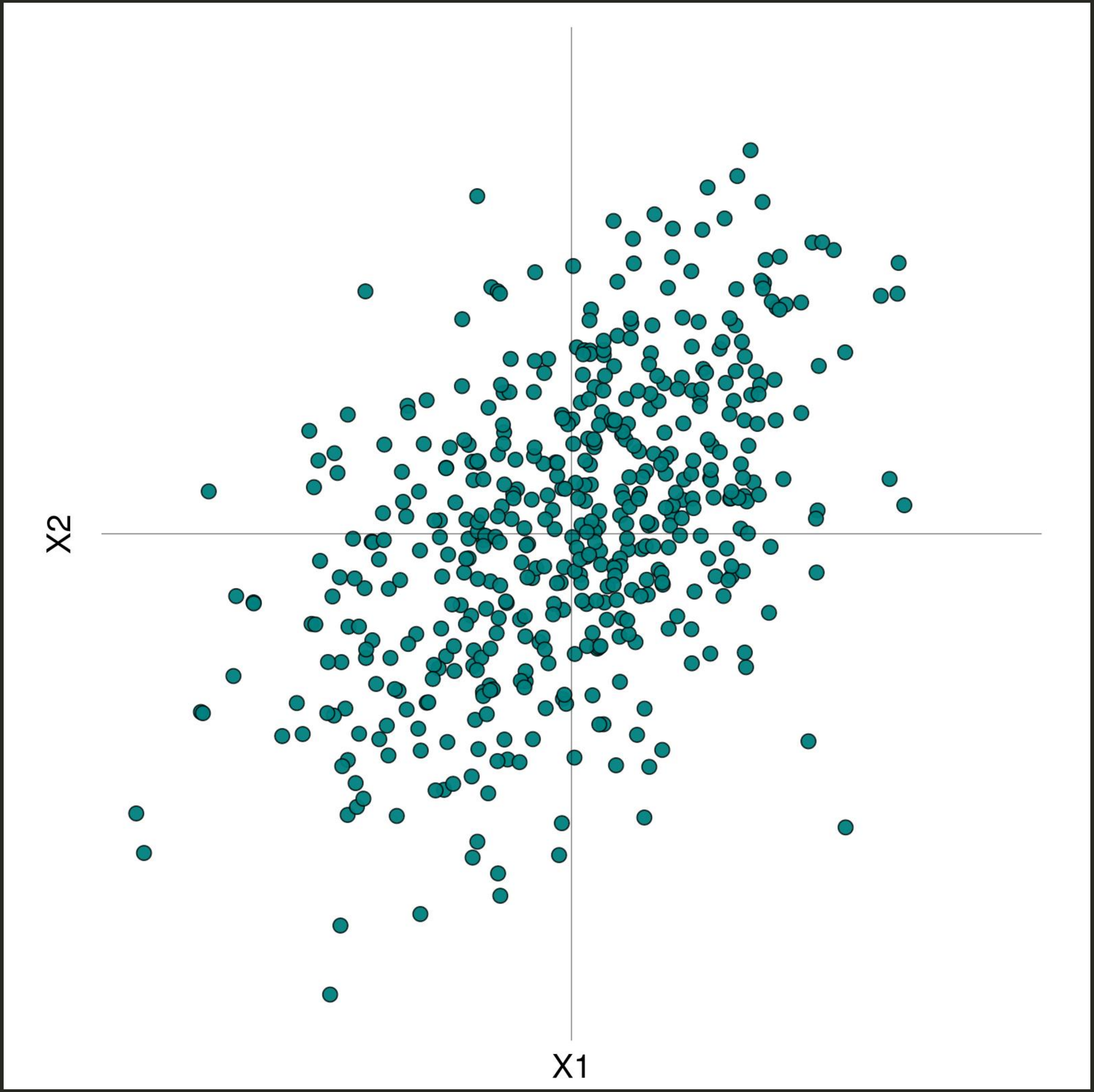
which is the Gram matrix for the standard Euclidean inner product.

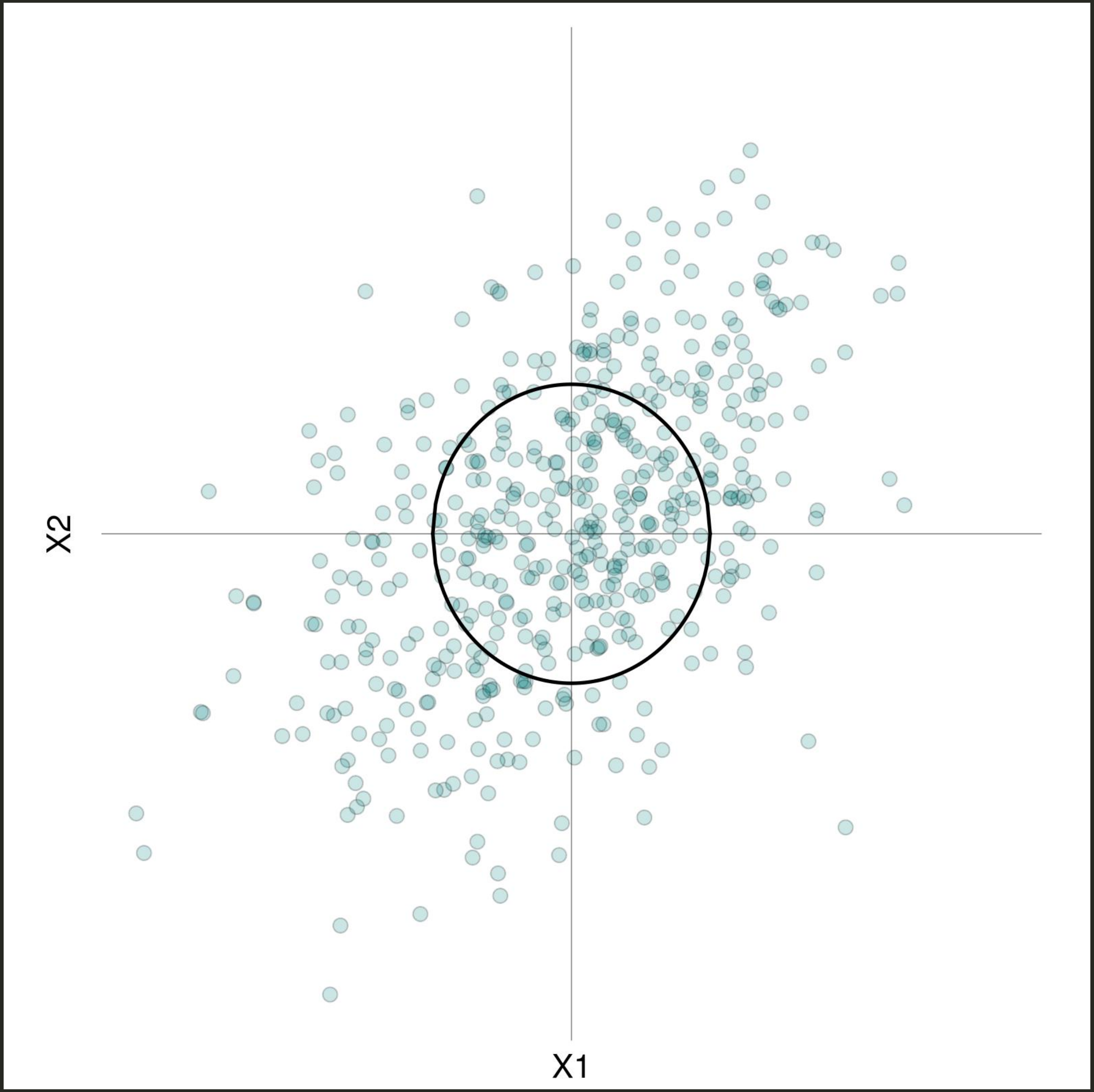
Mahalanobis Inner Product

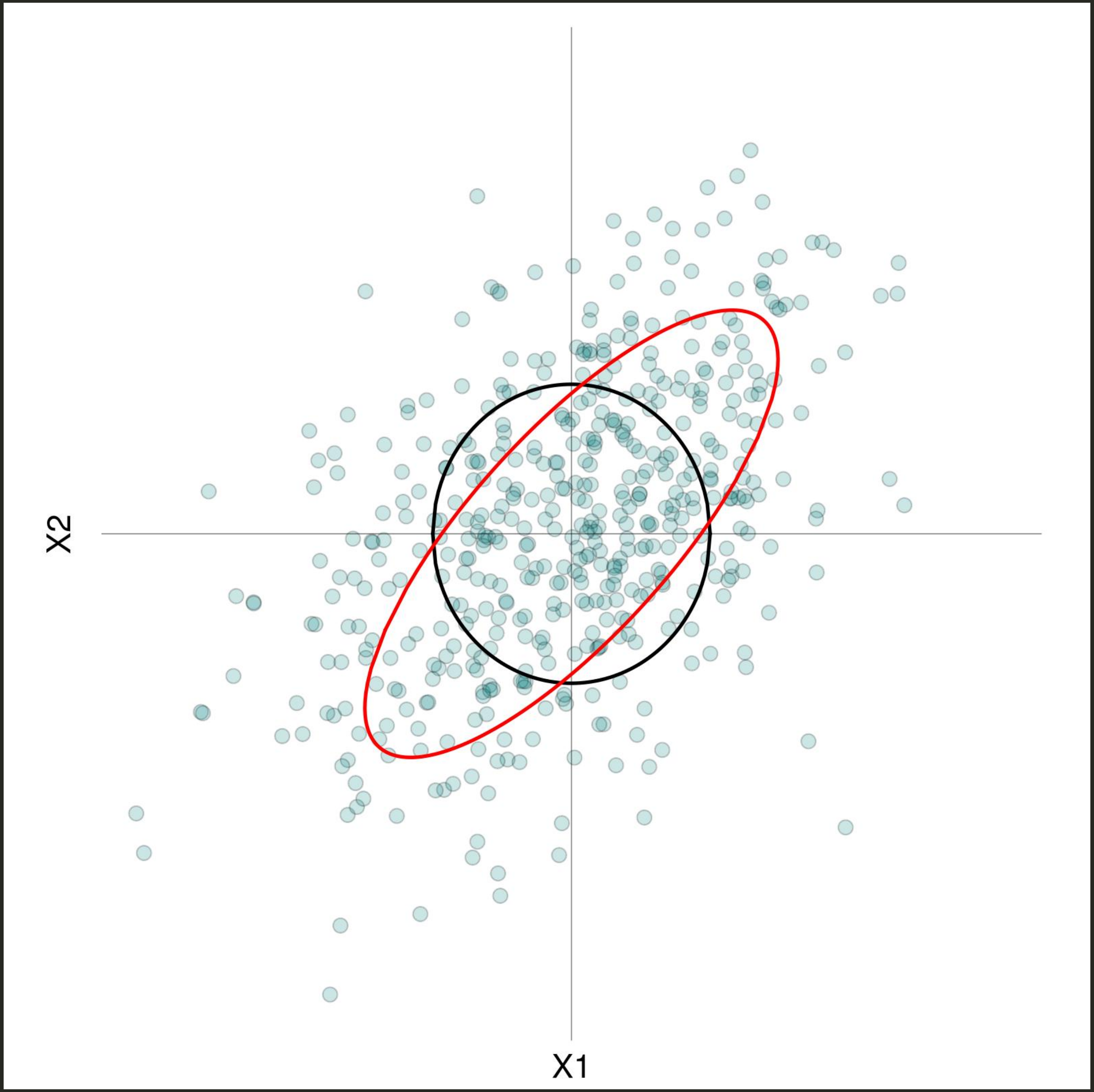
- Choosing $\mathcal{W} = (X^T X)^{-1}$ gives

$$K_M = X(X^T X)^{-1} X^T,$$

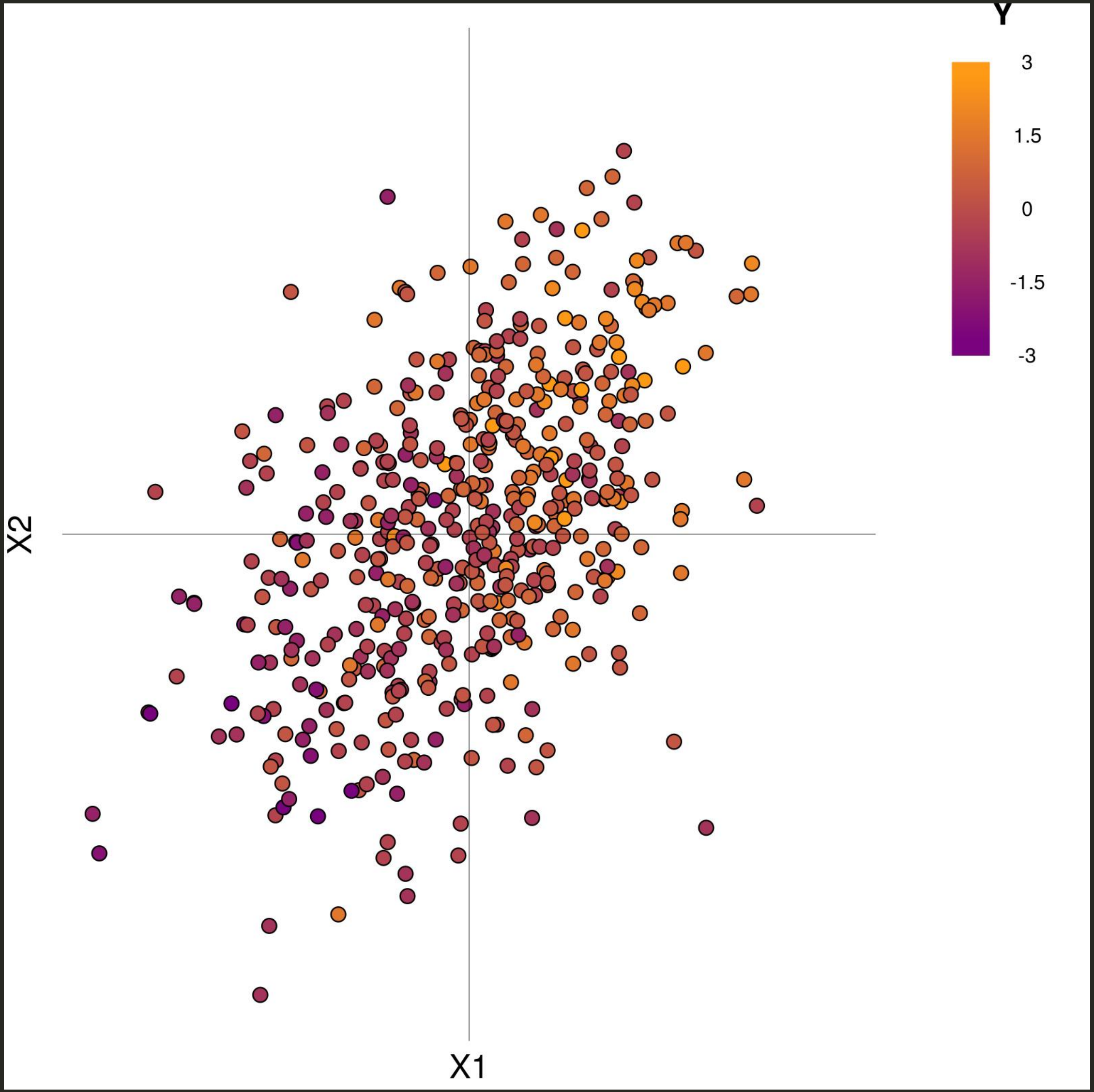
which is a similarity matrix related to the Mahalanobis distance, and is the projection matrix for the column space of X.

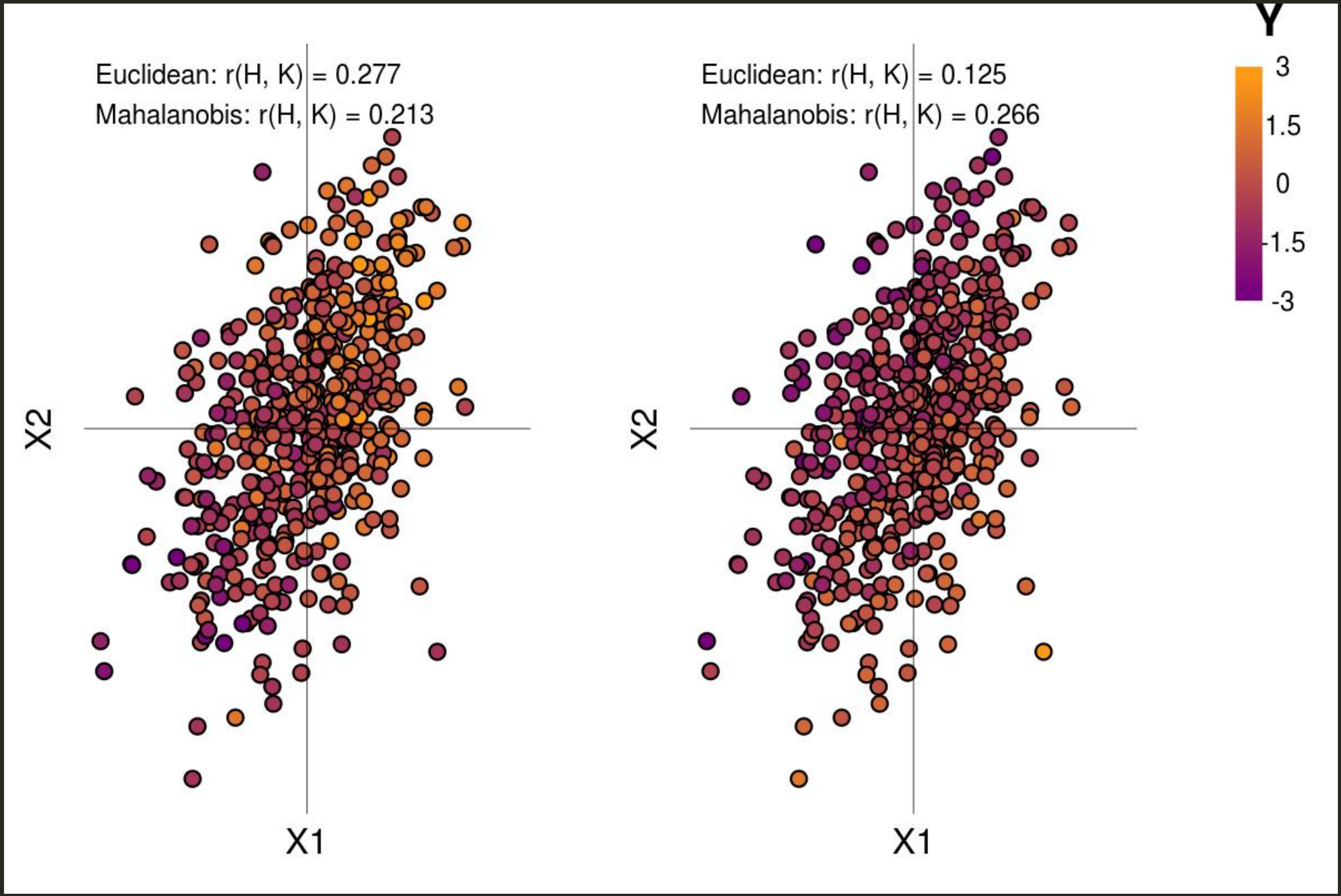

X2
X1


X2
X1


X2
X1

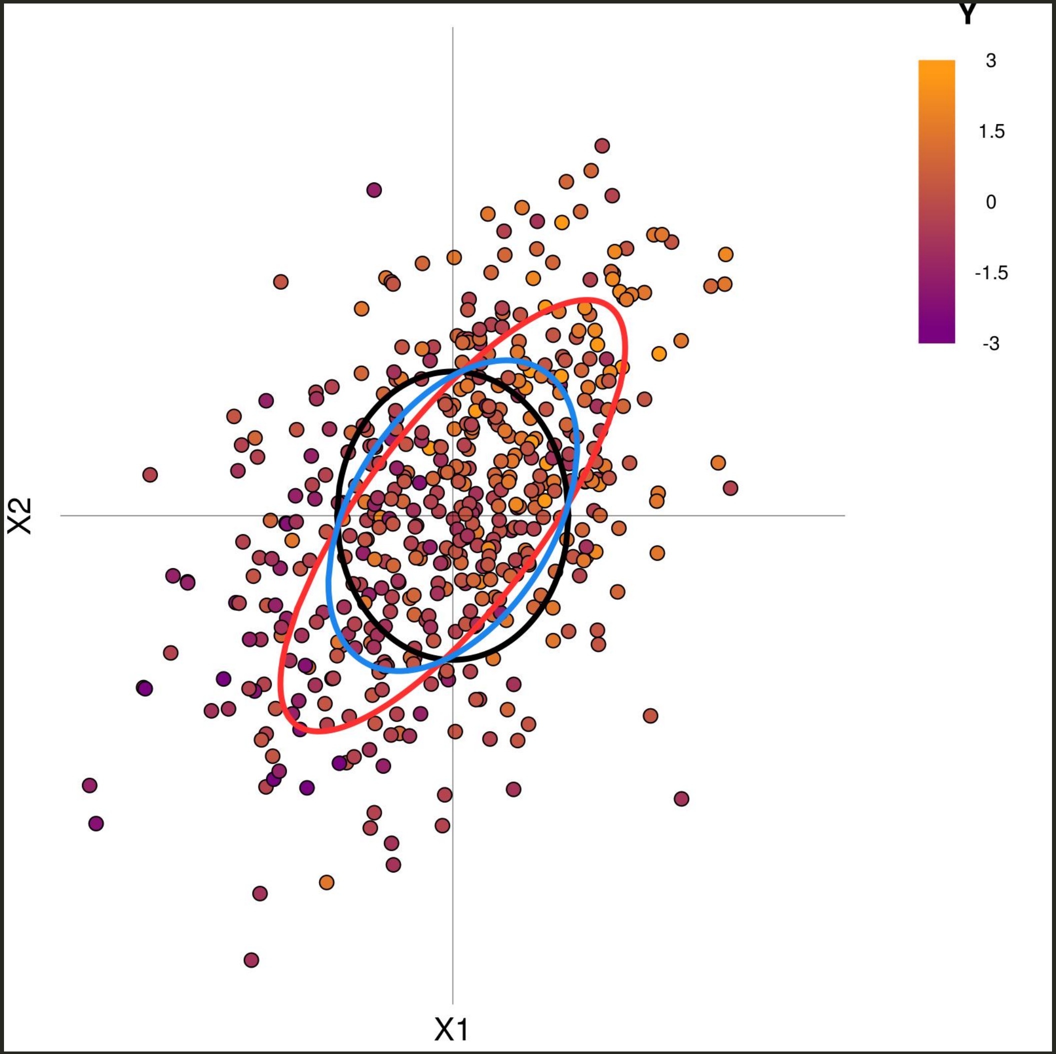
Data generated from variance components model with $\sigma_b^2 = 1$





Left Data generated from variance components model with $\sigma_b^2 = 1$.
Right Data generated from fixed effects model with $\beta = (0.75, -0.75)$.

Can we compromise between the Mahalanobis and Euclidean metrics?



Weight Matrices

Goal

We want a weight matrix that adjusts for the correlation structure in the data like the Mahalanobis metric, but which is closer to the Euclidean metric.

Weight Matrices

Goal

We want a weight matrix that adjusts for the correlation structure in the data like the Mahalanobis metric, but which is closer to the Euclidean metric.

Solution: Penalize the Mahalanobis weight matrix.

- Let $\lambda \geq 0$. Consider the penalized weight matrix:

$$\mathcal{W}_\lambda = (X^T X + \lambda I_p)^{-1}.$$

- As $\lambda \to \infty$, the penalty term λI_p dominates $X^T X$, and so W_λ tends to a constant diagonal matrix.
- We call $\mathcal{K}(u, v) = u^T \mathcal{W}_\lambda v$ the **ridge kernel**.

Kernel Mantel Test

Correlation of Similarities

Assume $\operatorname{rank}(X) = r$ with singular value decomposition $X = U_{n\times r} D_{r\times r} V_{p\times r}^T$, where $\eta_j, j = 1, \cdots, r$ are the squared singular values. Let $H = YY^T$ and $Z = U^T Y$.

Mahalanobis Metric $\quad r(H, K_M) \asymp \sum_{j=1}^r z_j^2 \quad = \operatorname{tr}(HK_M) =: T_M,$

Euclidean Metric $\quad r(H, K_E) \asymp \sum_{j=1}^r \eta_j z_j^2 \quad = \operatorname{tr}(HK_E) =: T_E,$

Ridge Similarity $\quad r(H, K_\lambda) \asymp \sum_{j=1}^r \frac{\eta_j}{\lambda+\eta_j} z_j^2 \quad = \operatorname{tr}(HK_\lambda) =: T_\lambda.$

Kernel Mantel Test

Limiting Relationship

From the previous results, we get the following limiting relationships between the ridge test, and tests for the fixed effects and random effects models.

$$T_{\lambda=0} = T_M$$

$$T_\lambda \asymp \left\{ \lambda \sum_{j=1}^{r} \frac{\eta_j}{\lambda + \eta_j} z_j^2 \right\} \quad \overset{\lambda \to \infty}{\to} \quad T_E$$

Similarly, for the matrix correlations

$$r(H, K_{\lambda=0}) = r(H, K_M)$$

$$\lim_{\lambda \to \infty} r(H, K_\lambda) = r(H, K_E)$$

Kernel Mantel Test

Linear Model Definitions

Model Name	Definition	
Fixed	$Y \sim N(X\beta, \sigma_\varepsilon^2 I_N)$	
Random	$Y \sim N(0, \sigma_b^2 G + \sigma_\varepsilon^2 I_N)$,	$G = XX^T/p$
Ridge	$Y \sim N(X\beta, \sigma_\varepsilon^2 I_N)$,	$\|\hat{\beta}\|_2^2 < c(\lambda)$

Linear Model Score Tests

Model	Score Stat.	Equivalent Stat.	Null Distribution
Fixed	$S = Z^T D(D^T D)^{-1} D^T Z$	$T_M = \mathrm{tr}(K_M H)$	$c_1 \chi_p^2$
Random	$S = Z^T D D^T Z$	$T_E = \mathrm{tr}(K_E H)$	$c_2 \sum_{j=1}^{r} \eta_j \chi_1^2$
Ridge	$S = Z^T D(D^T D + \lambda I_p)^{-1} D^T Z$	$T_\lambda = \mathrm{tr}(K_\lambda H)$	$c_3 \sum_{j=1}^{r} \frac{\eta_j}{\lambda + \eta_j} \chi_1^2$

Kernel Mantel Test

Proportion of Variance Explained

$$R^2(X,Y) = \sqrt{p} \cdot r(H, K_M)$$

Heritability

For large n and assuming that $\mathrm{rank}(X) = p$, and letting $G = K_E/p$

$$\hat{h}^2_{MOM} = \frac{tr(HG) - n}{tr(G^2) - n} \approx p\sqrt{\frac{tr(H^2)}{\mathrm{tr}(K_E^2)}} r(H, K_E) \in [r(H, K_E), \sqrt{p} \cdot r(H, K_E)]$$

Adaptive Mantel Test

Goal

We want to **use the ridge kernel in the Mantel test** to improve power in high-dimensional settings.

Adaptive Mantel Test

Goal

We want to **use the ridge kernel in the Mantel test** to improve power in high-dimensional settings.

Problem

However, **choosing a good penalty term for inference can be difficult,** since we must control the type I error.

Adaptive Mantel Test

Goal

We want to **use the ridge kernel in the Mantel test** to improve power in high-dimensional settings.

Problem

However, **choosing a good penalty term for inference can be difficult,** since we must control the type I error.

Idea

To simultaneously test a set of tuning parameters, use the **minimum *P*-value** across all parameters as the test statistic.

Adaptive Mantel Test

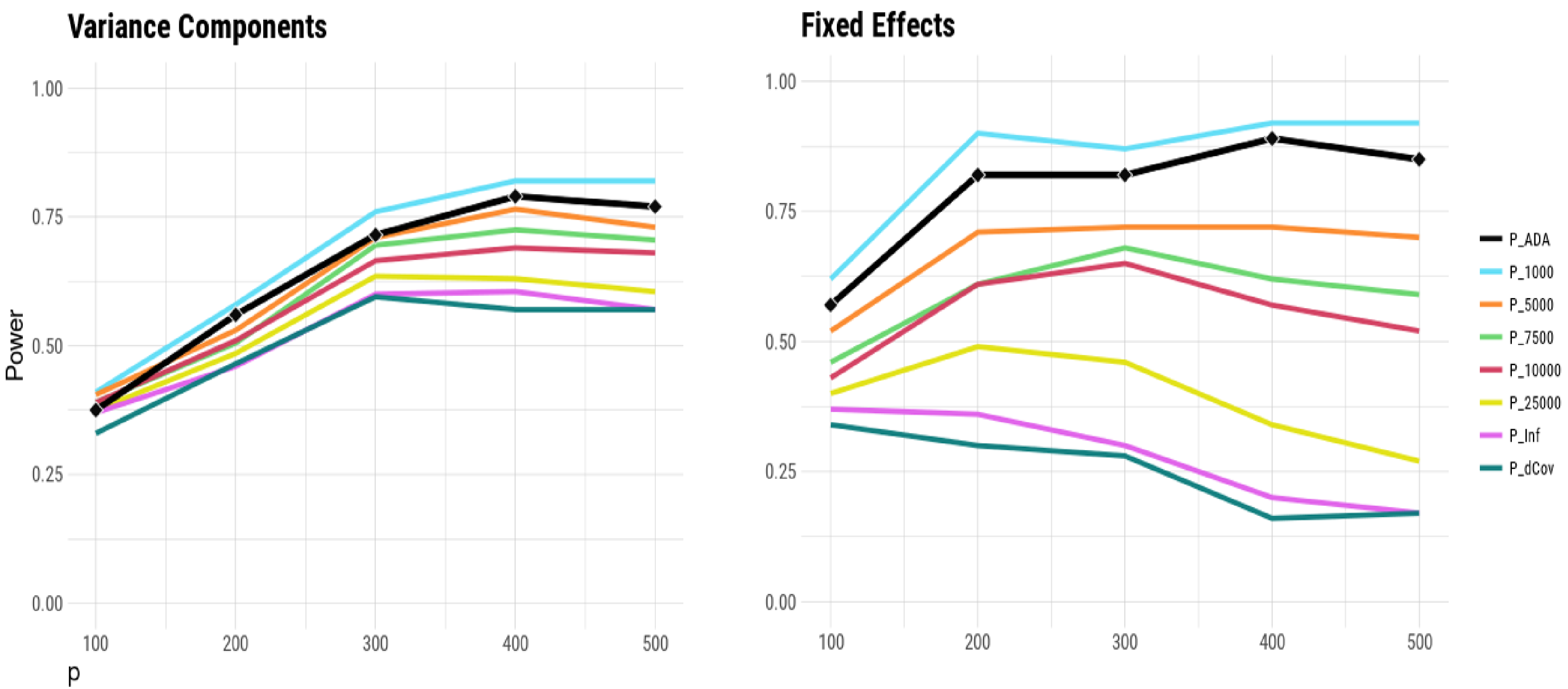
Idea

To simultaneously test a set of tuning parameters, use the **minimum** ***P*****-value** across all parameters as the test statistic.

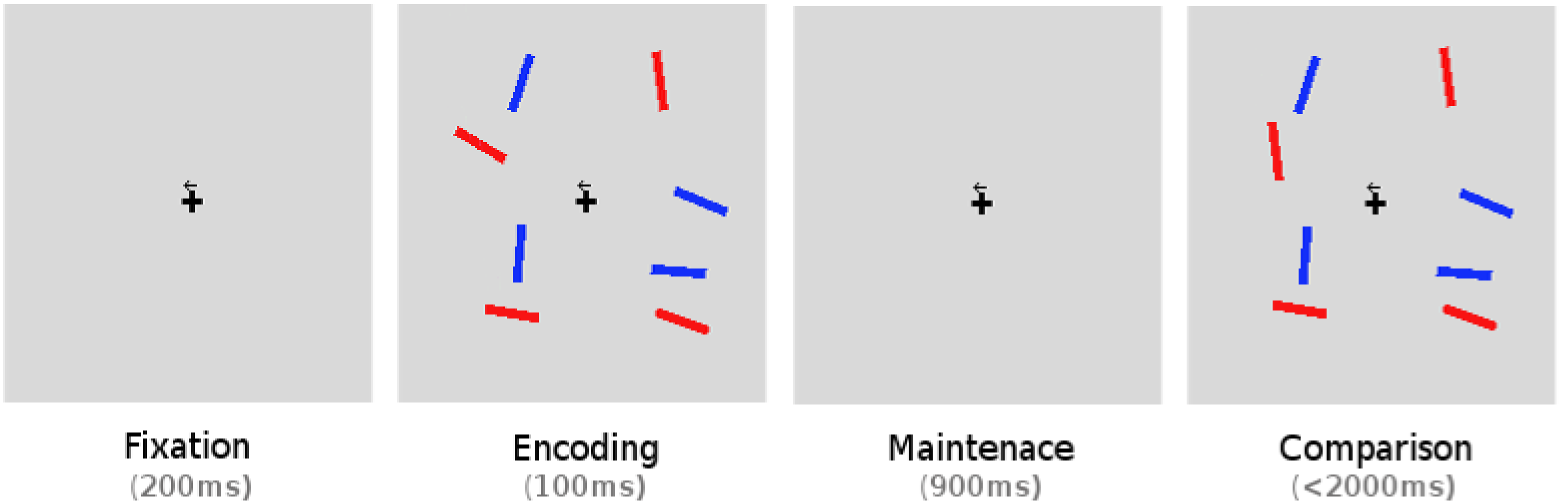
Algorithm

- **Input:**
  - $X, n \times p$ covariates, column centered and scaled
  - $Y, n \times 1$ response, centered and scaled
  - $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$ set of candidate penalty terms for the ridge kernel
- **Output:** P_{ADA} = adaptive Mantel *P*-value for global test of significant association

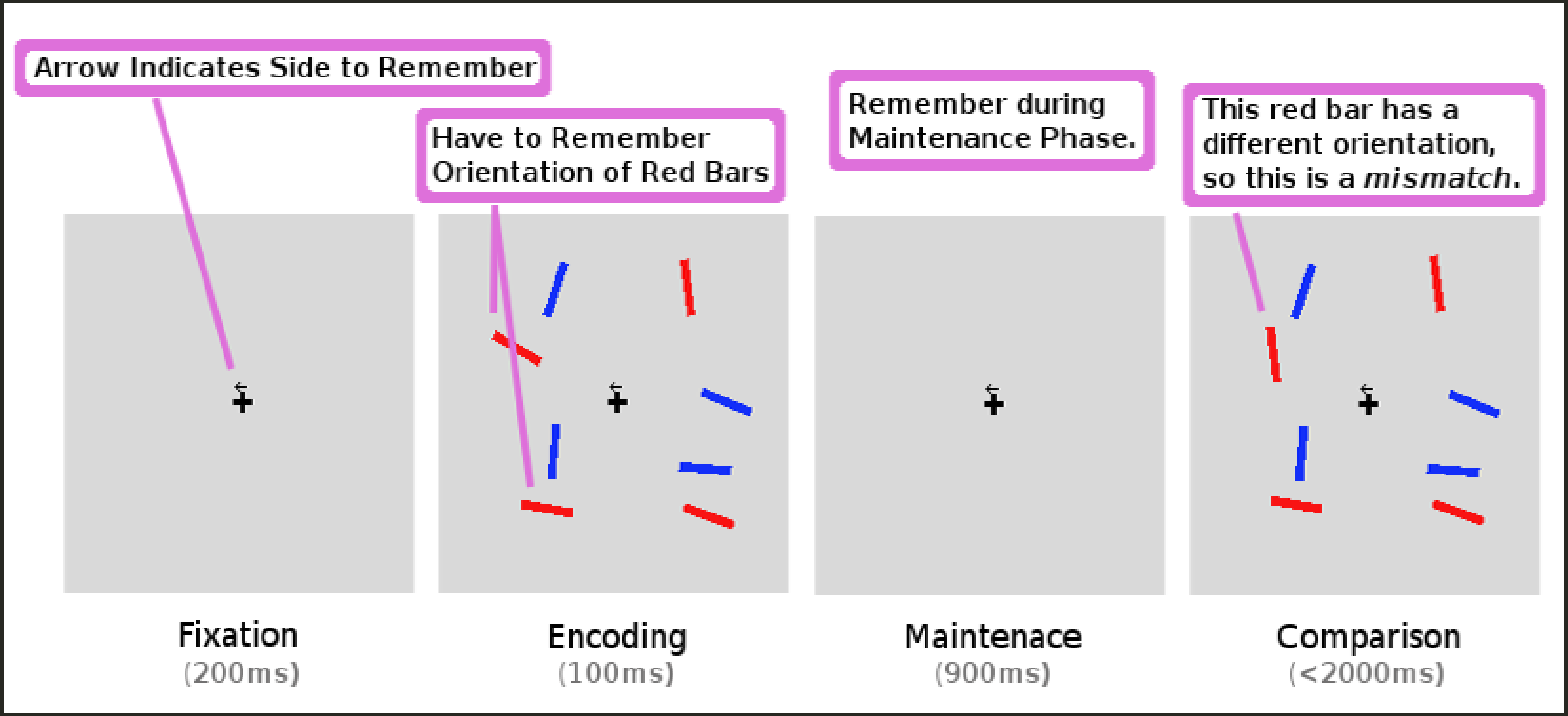
Simulations



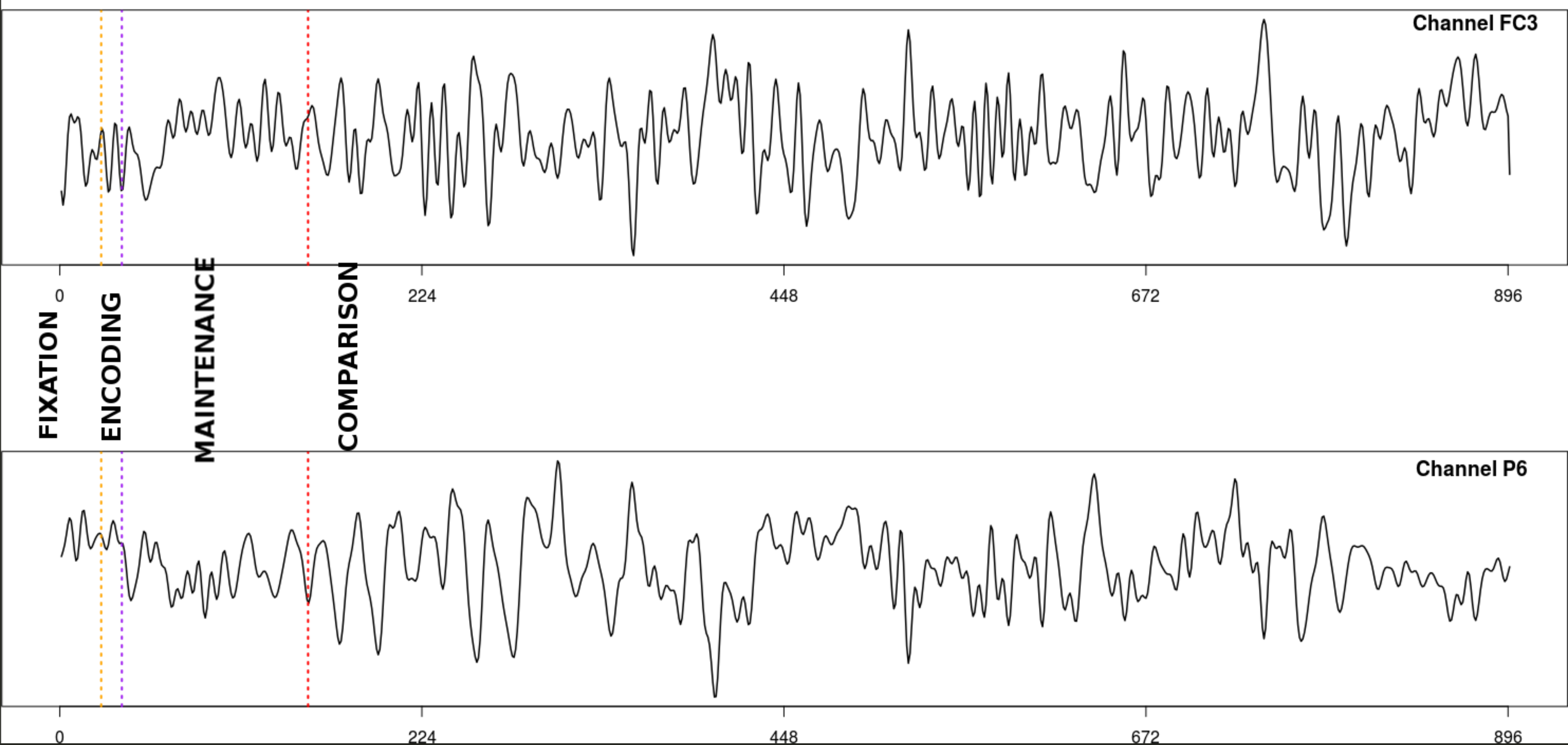
Visual Working Memory Experiment



Visual Working Memory Experiment



Visual Working Memory Experiment



Data Description

- 350 Subjects from the BNU data set
- ~10 minute 64 channel EEG recording during VWM task
  - Preprocessed according to standard pipeline
  - Coherence measures for each channel pair was calculated by the FFT, and grouped into five frequency bands (in Hz): $\delta\ (1-4), \theta\ (4-8), \alpha\ (8-16), \beta\ (16-32), \gamma\ (32+)$
- 13 SNPs selected for analysis, previously identified as potential factors for Alzheimer's disease risk
  - All 13 SNPs passed standard MAF and HWE quality control checks

Adaptive Mantel Test Results

- Results of adaptive Mantel test for association of AD SNPs and EEG Coherence at particular frequency bands
- Used L_2 similarity for SNPs, and ridge kernel similarity for coherence, with penalty terms $\Lambda = \{0.5, 1, 5, 10, 100, 1000, \infty\}$

Band	Channels	$P-$value
α	All	0.075
α	Frontal	0.381
θ	All	0.416
θ	Frontal	0.081

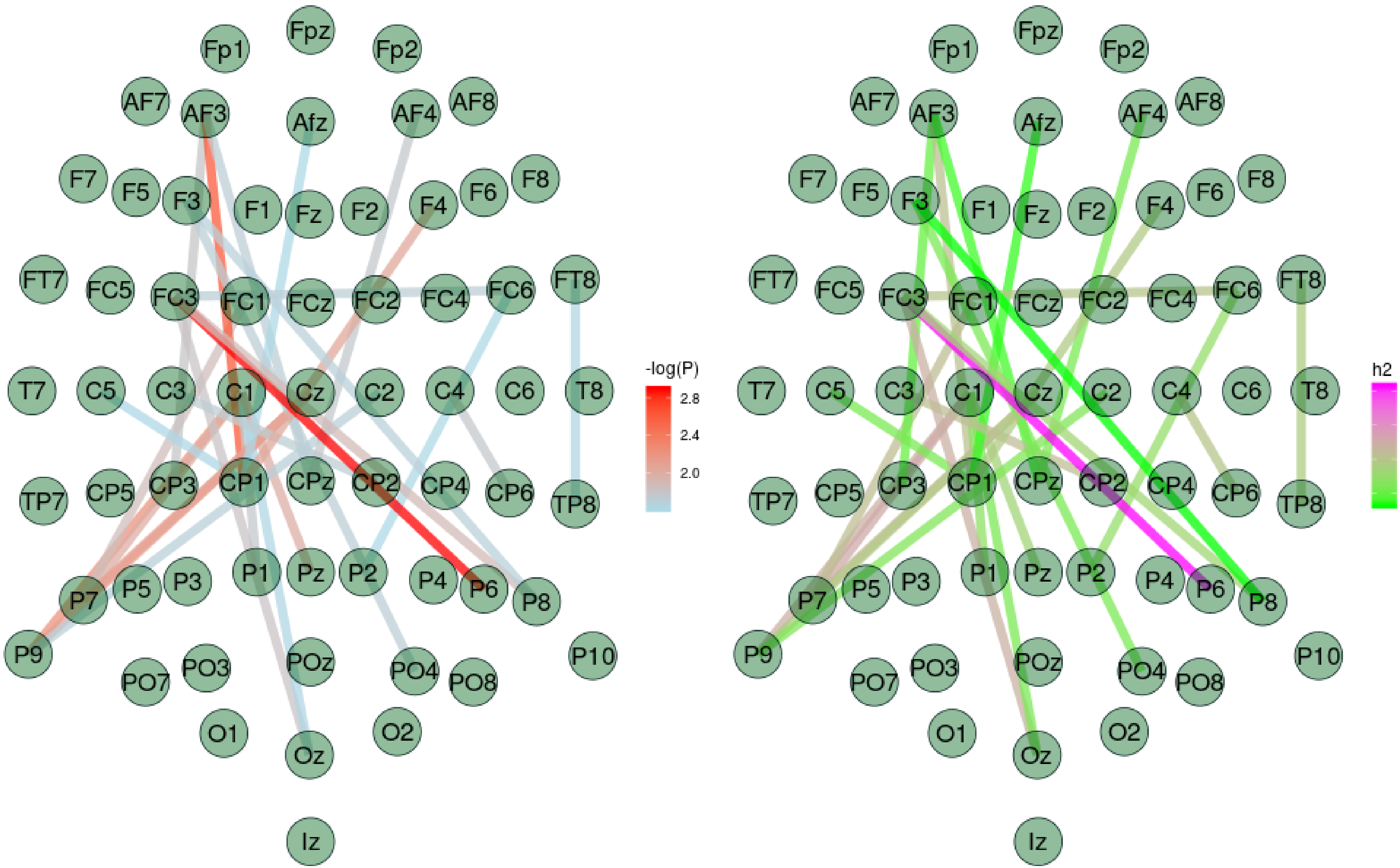
Adaptive Mantel Test Results for Genome-wide Association

- Results of adaptive Mantel test for **genome-wide heritability** of EEG Coherence at particular frequency bands
- Used L_2 similarity for SNPs, and ridge kernel similarity for coherence, with penalty terms $\Lambda = \{0.5, 1, 5, 10, 100, 1000, \infty\}$

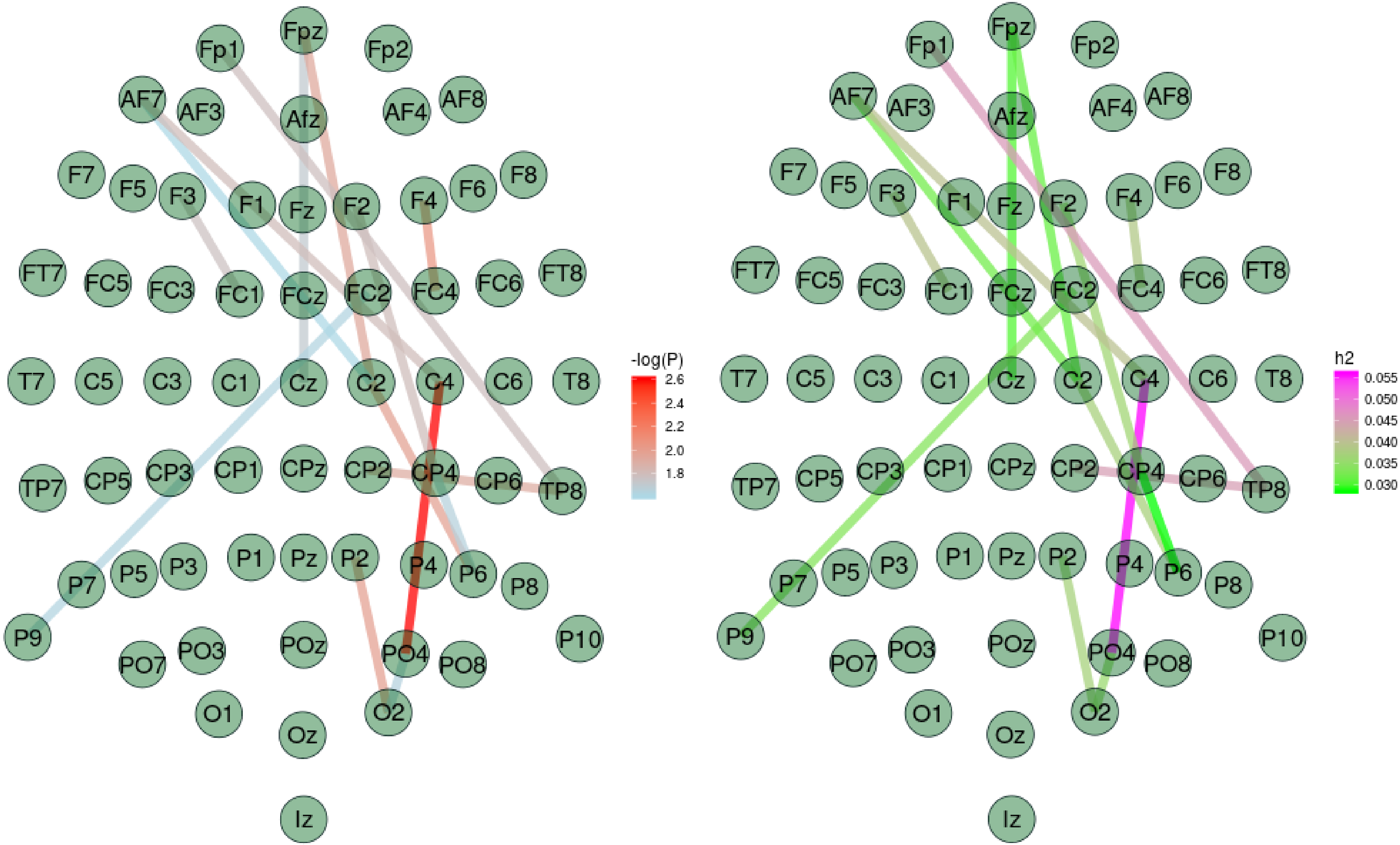
Band	Channels	$P-$ value
α	All	0.861
α	Frontal	0.410
θ	All	0.261
θ	Frontal	0.046

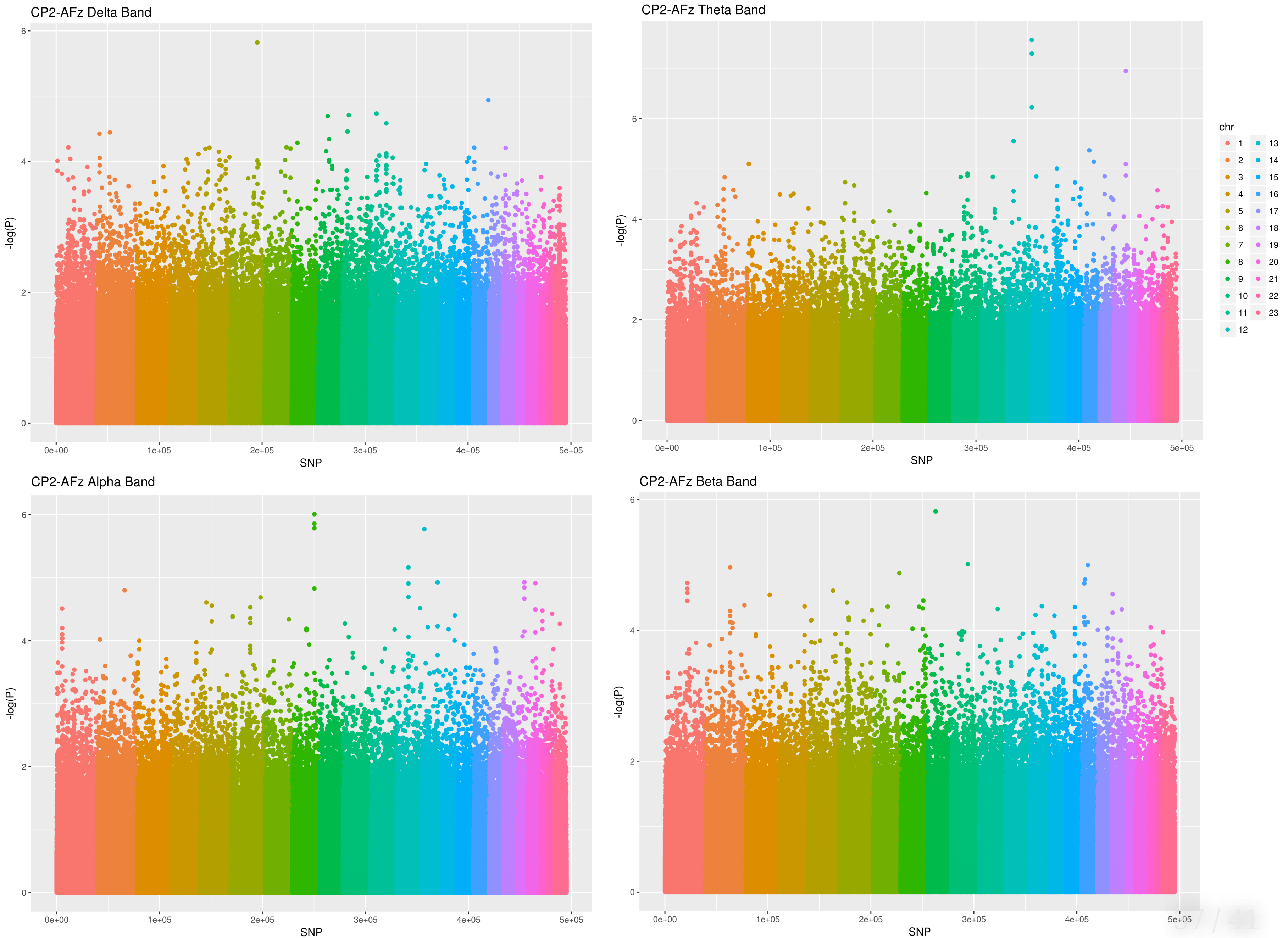
For θ Frontal channels, $\hat{h}^2 = 0.165$; best $\lambda = \infty$.

Theta Coherence



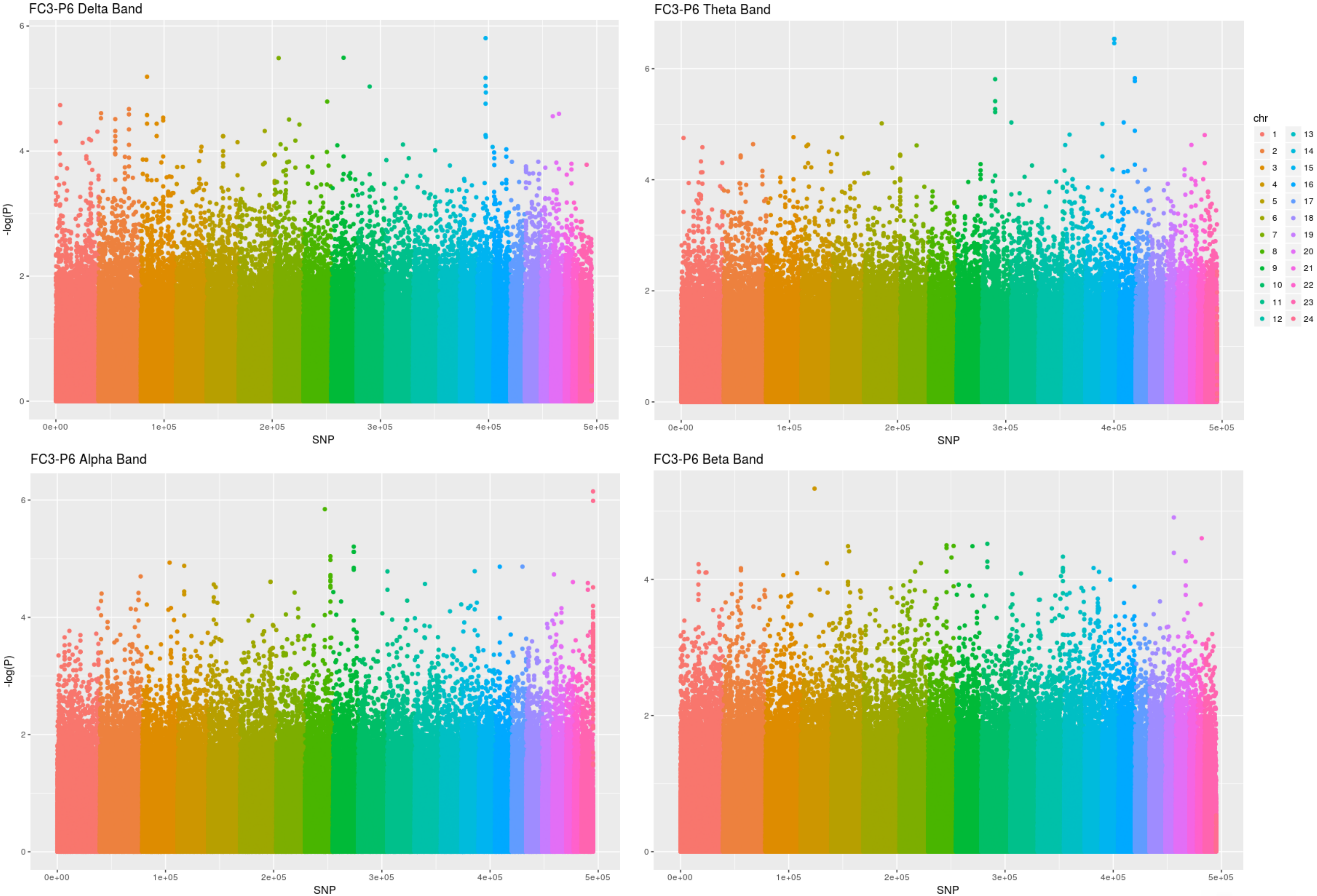
Alpha Coherence




CP2-AFz Delta Band
CP2-AFz Theta Band
CP2-AFz Alpha Band
CP2-AFz Beta Band
-log(P)
SNP
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
chr
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Most Significant SNPs for CP2-AFz Coherence

SNP	Chr	$P - value$	$Band$
$rs11061121$	12	2.7×10^{-8}	Theta
$rs7137937$	12	5.1×10^{-8}	Theta
$rs4759774$	12	5.1×10^{-8}	Theta
$rs7954170$	12	5.9×10^{-7}	Theta
$rs4891649$	18	1.13×10^{-7}	Theta
$rs16886$	8	9.8×10^{-7}	Alpha


FC3-P6 Delta Band
FC3-P6 Theta Band
FC3-P6 Alpha Band
FC3-P6 Beta Band
-log(P)
SNP
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
chr
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Most Significant SNPs for FC3-PO4 Coherence

SNP	Chr	$P - value$
$rs7182385$	15	2.9×10^{-7}
$rs3825876$	15	2.9×10^{-7}
$rs7342686$	15	3.5×10^{-7}
$rs2075181$	Y	7.11×10^{-7}

- No known clinical relevance for these Chr 15 SNPs, although this chromosome contains many genes that are expressed in the brain and nervous system, and which are related to developmental and mental disorders.
- The haplotype marker on the Y chromosome may indicate that gender differences are significant

Links

- **Statistical methods and challenges in connectome genetics**, *Stat. \& Probability Letters* 136 (2018), P. et al.
- **Adaptive Mantel Test Paper:** arxiv.org/pdf/1712.07270.pdf
- **Adaptive Mantel R Package:** github.com/dspluta/adamant
- **Presentation Slides:** github.com/dspluta/Presentations

References

- Wu M., et al. Kernel Machine SNP-set Testing under Multiple Candidate Kernels. Genetic Epideomiology. 2013. 37(3): 267-275.
- Cai T., et al., Kernel Machine Approach to Testing the Significance of Multiple Genetic Markers for Risk Prediction. Biometrics. 2011. 67(3): 975-986.
- Ge T, et al. Massively Expedited Genome-Wide Heritability Analysis (MEGHA). PNAS. 2015. 112, 2479-2484.
- Xue G, et al. Functional Dissociations of Risk and Reward Processing in the Medial Prefrontal Cortex. Cerebral Cortex. 2009. 19, 1019-1027.
- Yang J, et al. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics. (2011) 88, 76-82.
- Tzeng et al. (2009) Biometrics 65, 822.
- Visscher et al. (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data. PLoS Genetics.
- GCTA Power, http://cnsgenomics.com/shiny/gctaPower/