Spectral Dynamic Causal Modelling of Resting-State fMRI: Relating Effective Brain Connectivity in the Default Mode Network to Genetics

Yunlong Nie[1], Laila Yasmin[2], Yin Song[2], Vanessa Scarapicchia[3]
Jodie Gawryluk[3], Liangliang Wang[1], Jiguo Cao[1], Farouk Nathoo[2]

[1]Department of Statistics and Actuarial Science, Simon Fraser University

[2]Department of Mathematics and Statistics, University of Victoria

[3]Department of Psychology, University of Victoria

Table of Contents

- Imaging genetics study of 112 subjects from the Alzheimer's Disease Neuroimaging Initiative.

- Imaging genetics study of 112 subjects from the Alzheimer's Disease Neuroimaging Initiative.
- Examine resting-state fMRI (rs-fMRI) and genetic data obtained from these subjects, where each subject is classified as either cognitively normal (CN), as having mild cognitive impairment (MCI), or as having Alzheimer's Disease (AD).

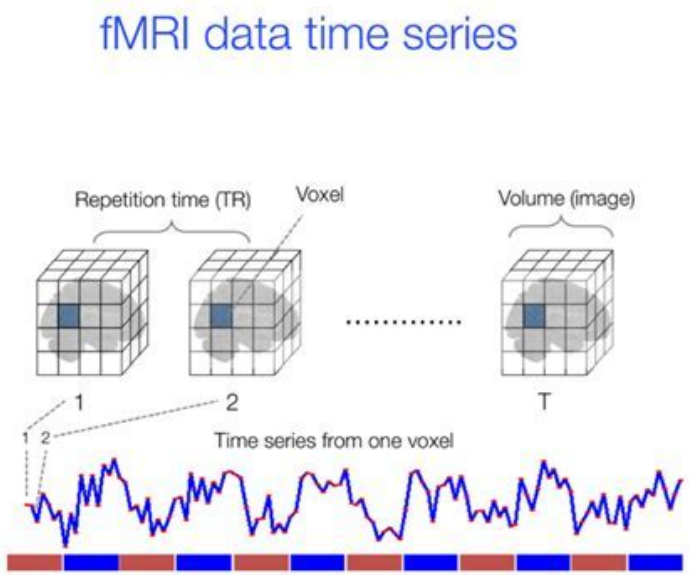
- Imaging genetics study of 112 subjects from the Alzheimer's Disease Neuroimaging Initiative.
- Examine resting-state fMRI (rs-fMRI) and genetic data obtained from these subjects, where each subject is classified as either cognitively normal (CN), as having mild cognitive impairment (MCI), or as having Alzheimer's Disease (AD).
- Our goal is to examine the relationship between effective brain connectivity and genetics in relation to disease.

- fMRI measures the amount of oxygenated versus deoxygenated blood in different parts of the brain

- fMRI measures the amount of oxygenated versus deoxygenated blood in different parts of the brain
- Since neural activity requires oxygen to be delivered to active parts of the brain through blood flow, this measurement is a proxy or indirect measurement of neural activity

- fMRI measures the amount of oxygenated versus deoxygenated blood in different parts of the brain
- Since neural activity requires oxygen to be delivered to active parts of the brain through blood flow, this measurement is a proxy or indirect measurement of neural activity
- We are measuring the metabolic demands of active neurons as a proxy for the actual electrical neural activity

- This gives relatively low temporal resolution (TR $\sim$ 2s) but high spatial resolution (1 cubic millimeter voxels/ 1 million voxels in a high resolution brain scan)



- Task-based fMRI: the subject is stimulated by some paradigm (e.g. visual or auditory stimulus) and fMRI can be used to examine active brain regions used to process stimuli and plan a response

- Task-based fMRI: the subject is stimulated by some paradigm (e.g. visual or auditory stimulus) and fMRI can be used to examine active brain regions used to process stimuli and plan a response
- Resting-state fMRI: no stimulus and therefore likely no overwhelmingly strong signal in localized regions; fMRI measurements in this case typically used to characterize regional interactions - connectivity

- We use Dynamic Causal Modeling (DCM; Li et al., 2011; Friston et al., 2003; Friston et al., 2014; Razi et al., 2015; Friston et al., 2017) applied to rs-fMRI time series in order to estimate a directed network representing effective brain connectivity within the default mode network (DMN).

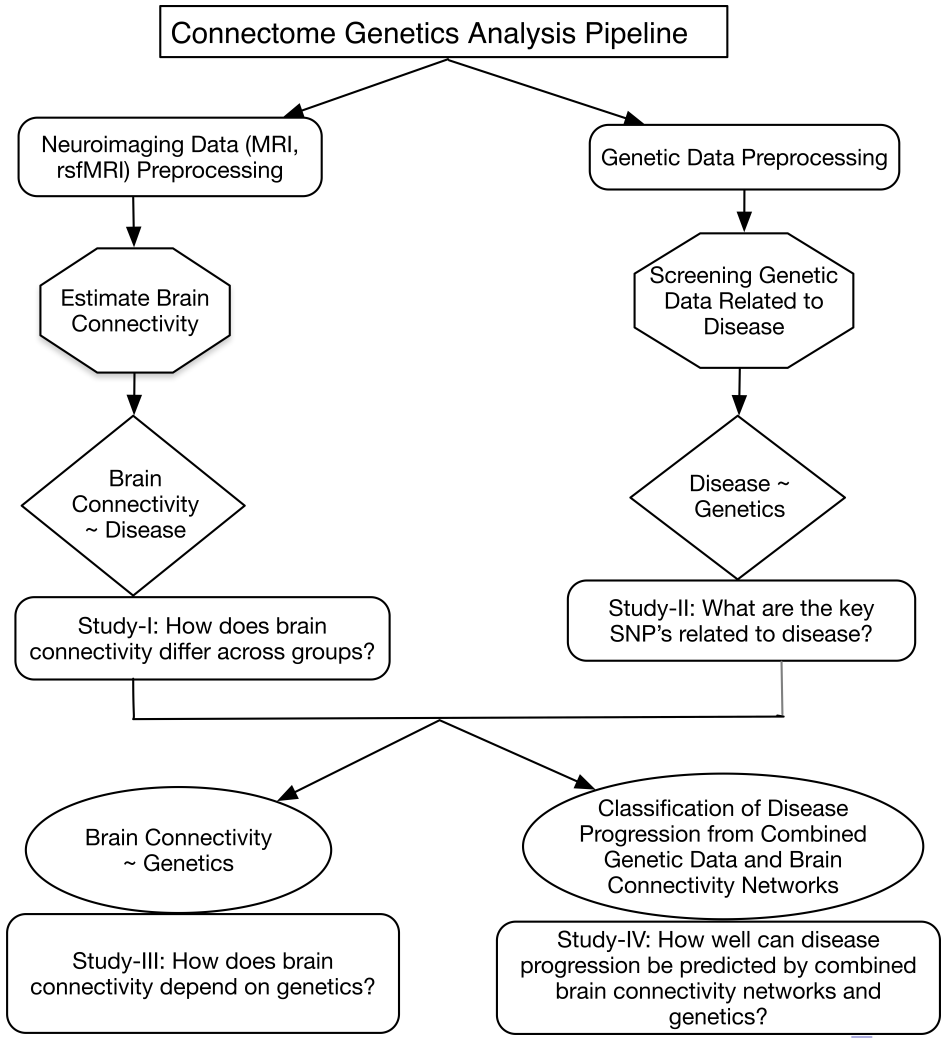
- We use Dynamic Causal Modeling (DCM; Li et al., 2011; Friston et al., 2003; Friston et al., 2014; Razi et al., 2015; Friston et al., 2017) applied to rs-fMRI time series in order to estimate a directed network representing effective brain connectivity within the default mode network (DMN).
- The DMN is a key a network commonly known to be active when the brain is at rest.

- We use Dynamic Causal Modeling (DCM; Li et al., 2011; Friston et al., 2003; Friston et al., 2014; Razi et al., 2015; Friston et al., 2017) applied to rs-fMRI time series in order to estimate a directed network representing effective brain connectivity within the default mode network (DMN).
- The DMN is a key a network commonly known to be active when the brain is at rest.
- These networks are then related to genetic data and Alzheimer's disease.

- We use Dynamic Causal Modeling (DCM; Li et al., 2011; Friston et al., 2003; Friston et al., 2014; Razi et al., 2015; Friston et al., 2017) applied to rs-fMRI time series in order to estimate a directed network representing effective brain connectivity within the default mode network (DMN).
- The DMN is a key a network commonly known to be active when the brain is at rest.
- These networks are then related to genetic data and Alzheimer's disease.
- Ours is the first imaging genetics study to use DCM as a neuroimaging phenotype and this is one novel aspect of the study.

- We use Dynamic Causal Modeling (DCM; Li et al., 2011; Friston et al., 2003; Friston et al., 2014; Razi et al., 2015; Friston et al., 2017) applied to rs-fMRI time series in order to estimate a directed network representing effective brain connectivity within the default mode network (DMN).
- The DMN is a key a network commonly known to be active when the brain is at rest.
- These networks are then related to genetic data and Alzheimer's disease.
- Ours is the first imaging genetics study to use DCM as a neuroimaging phenotype and this is one novel aspect of the study.
- Ours is also the first imaging genetics study to use the Bayes factor for examining evidence for genetic effects.

- In putting together the analysis we have developed a new pipeline for connectome genetics which can be applied generally to examine the relationship between brain connectivity and genetics in relation to disease.



- The selection criteria for our study: We first begin with ADNI2 subjects (1437 at this stage) and considered those subjects also having genome-wide data (774 left at this stage) and also with at least one resting-state fMRI scan at 3T (112 subjects).
- The 112 subjects comprise 37 CN, 63 MCI and 12 AD subjects, with these subjects having a mean age of 73.8 years with the range being 56.3-95.6 years, 46 of these subjects being male, 5 being left-handed, and with education measured in years ranging from 11 to 20.

Table: Distribution of Demographic Variables

		AD	MCI	NL	p-value
n		12	63	37	
PTGENDER (%)	Female	8 (66.7)	34 (54.0)	23 (62.2)	0.640
	Male	4 (33.3)	29 (46.0)	14 (37.8)	
PTHAND (%)	Left	1 (8.3)	2 (3.2)	2 (5.4)	0.507
	Right	11 (91.7)	61 (96.8)	35 (94.6)	
Age (mean (sd))		75.82 (7.91)	72.70 (7.66)	75.22 (6.68)	0.165
PTEDUCAT (mean (sd))		16.33 (2.53)	16.06 (2.66)	16.30 (2.21)	0.874
APOE Gene No.e4 (%)	Zero	1 (8.3)	35 (55.6)	24 (64.9)	0.0052
	One	9 (75.0)	22 (34.9)	12 (32.4)	
	Two	2 (16.7)	6 (9.5)	1 (2.7)	

- For these subjects we examine effective connectivity based on their baseline rs-fMRI scan.

- For these subjects we examine effective connectivity based on their baseline rs-fMRI scan.
- Functional Connectivity: refers simply to the correlation between measured time series over different locations.

- For these subjects we examine effective connectivity based on their baseline rs-fMRI scan.
- Functional Connectivity: refers simply to the correlation between measured time series over different locations.
- Effective Connectivity: directional time-lagged connectivity or the influence one neuronal system exerts over another.

Our study examines effective connectivity across four DMN regions.

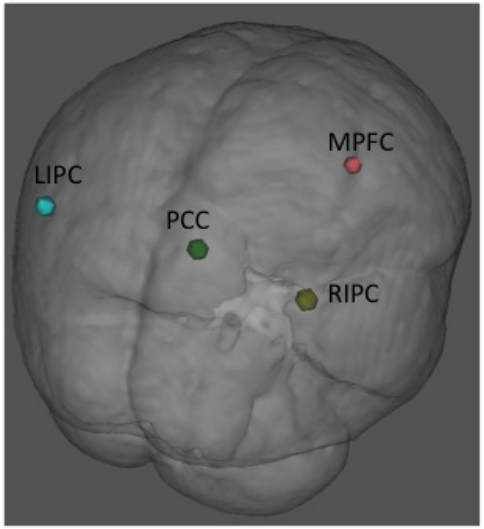


Figure: The locations of the four regions within the default mode network (DMN) examined in our studies: the medial prefrontal cortex (mPFC), the posterior cingulate cortex (PCC), the left and right intraparietal cortex (LIPC and RIPC).

- We estimate resting-state effective connectivity for fMRI using a variant of DCM known as spectral DCM.

- We estimate resting-state effective connectivity for fMRI using a variant of DCM known as spectral DCM.
- Use of DCM as an endophenotype is justified by the work of Xu et al. (*Cerebral Cortex*, 2017) who use DCM applied to rs-fMRI in a twin study in order to identify the genetic and environmental contribution to resting-state effective connectivity in the DMN.

- We estimate resting-state effective connectivity for fMRI using a variant of DCM known as spectral DCM.
- Use of DCM as an endophenotype is justified by the work of Xu et al. (*Cerebral Cortex*, 2017) who use DCM applied to rs-fMRI in a twin study in order to identify the genetic and environmental contribution to resting-state effective connectivity in the DMN.
- These authors estimate the heritability of effective resting-state connectivity within the DMN (same four regions) as 0.54 and thus suggest that the genes have an influence on effective resting-state connectivity.

- The fMRI and anatomical data are pre-processed using a combination of the FSL and SPM12 software.
- As part of this each subject's anatomy is normalized to a standardized space defined by a Montreal Neurological Institute (MNI) template brain.
- BOLD time series from the DMN regions of interest are obtained by extracting time series from all voxels within an 8mm radius of the associated MNI coordinate, and then applying a principle component analysis and extracting the first eigenvariate.

fMRI Preprocessing

- Slice timing correction: the entire brain volume is not sampled at the same instant in time; different layers of the volume are sampled at different time points during TR

fMRI Preprocessing

- Slice timing correction: the entire brain volume is not sampled at the same instant in time; different layers of the volume are sampled at different time points during TR
- Slice timing correction: shifts the measured time course at each voxel by a small amount so that after correction we can assume all voxels are sampled at the same time

fMRI Preprocessing

- Slice timing correction: the entire brain volume is not sampled at the same instant in time; different layers of the volume are sampled at different time points during TR
- Slice timing correction: shifts the measured time course at each voxel by a small amount so that after correction we can assume all voxels are sampled at the same time
- Realignment: realign images to account for motion correction during the scan

fMRI Preprocessing

- Slice timing correction: the entire brain volume is not sampled at the same instant in time; different layers of the volume are sampled at different time points during TR
- Slice timing correction: shifts the measured time course at each voxel by a small amount so that after correction we can assume all voxels are sampled at the same time
- Realignment: realign images to account for motion correction during the scan
- Normalization: refers to the coregistration of a subject's image to a standard template space (shape normalization across subjects)

fMRI Preprocessing

- Slice timing correction: the entire brain volume is not sampled at the same instant in time; different layers of the volume are sampled at different time points during TR
- Slice timing correction: shifts the measured time course at each voxel by a small amount so that after correction we can assume all voxels are sampled at the same time
- Realignment: realign images to account for motion correction during the scan
- Normalization: refers to the coregistration of a subject's image to a standard template space (shape normalization across subjects)
- Coregistration: alignment of functional and structural images from the same subject to map functional (brain activity) information into anatomical space

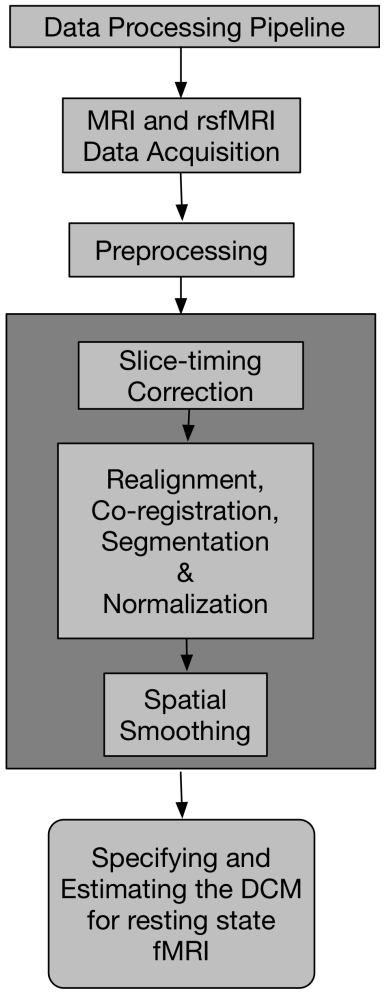


Figure: The steps involved in preprocessing the fMRI data and obtaining an effective brain connectivity network.

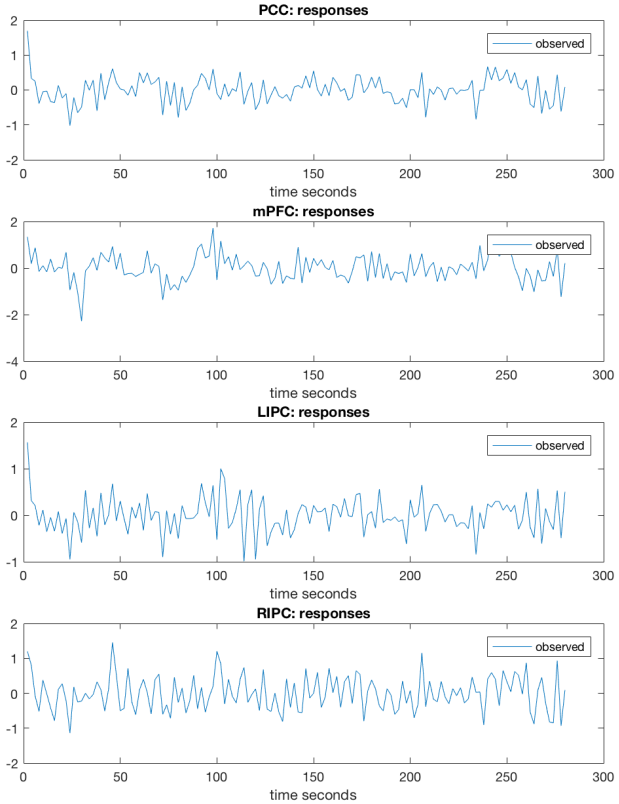


Figure: An example of the rs-fMRI data obtained from a single subject from the four regions of interest (PCC, mPFC, LIPC, RIPC).

- Given observed time series at R regions $\boldsymbol{y}(t) = (y_1(t), \ldots, y_R(t))'$, DCM is a state space model specified as follows:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{v}(t)$$

$$\boldsymbol{y}(t) = h(\boldsymbol{x}(t), \boldsymbol{\theta}) + \boldsymbol{e}(t)$$

- $\boldsymbol{x}(t) = (x_1(t), \ldots, x_R(t))'$ are latent variables used to represent the states of neuronal populations at some time t and $\dot{\boldsymbol{x}}(t)$ is a time derivative defining a differential equation approximating neuronal dynamics

- Given observed time series at R regions $\boldsymbol{y}(t) = (y_1(t), \ldots, y_R(t))'$, DCM is a state space model specified as follows:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{v}(t)$$

$$\boldsymbol{y}(t) = h(\boldsymbol{x}(t), \boldsymbol{\theta}) + \boldsymbol{e}(t)$$

- The differential equation approximating the dynamics for neuronal states represents a first-order Taylor approximation for an underlying non-linear biophysical model; $\boldsymbol{A}$ represents the Jacobian matrix for the function defining the nonlinear dynamics; its estimation allows us to characterize (to first-order) the effective connectivity to first-order $\rightarrow$ this gives us the network

- Given observed time series at R regions $\boldsymbol{y}(t) = (y_1(t), \ldots, y_R(t))'$, DCM is a state space model specified as follows:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{v}(t)$$

$$\boldsymbol{y}(t) = h(\boldsymbol{x}(t), \boldsymbol{\theta}) + \boldsymbol{e}(t)$$

- $h(\boldsymbol{x}(t), \boldsymbol{\theta})$ is a nonlinear mapping from neuronal states to the haemodynamic response also depending on parameters θ (see, e.g., Friston, 2007, *Neuroimage*)
- $\boldsymbol{v}(t)$ and $\boldsymbol{e}(t)$ represent neuronal and measurement noise respectively

- For resting-state fMRI this stochastic DCM model can be fit in the time-domain using a mean-field variational Bayes approach (see, e.g., Li et al., 2011, *Neuroimage*) which involves inference on both model parameters and latent states.

- For resting-state fMRI this stochastic DCM model can be fit in the time-domain using a mean-field variational Bayes approach (see, e.g., Li et al., 2011, *Neuroimage*) which involves inference on both model parameters and latent states.
- The model can also be fit in the spectral domain using an approach known as spectral DCM (Friston et al., 2014, *Neuroimage*).

- For resting-state fMRI this stochastic DCM model can be fit in the time-domain using a mean-field variational Bayes approach (see, e.g., Li et al., 2011, *Neuroimage*) which involves inference on both model parameters and latent states.
- The model can also be fit in the spectral domain using an approach known as spectral DCM (Friston et al., 2014, *Neuroimage*).
- The latter approach is somewhat akin to a method of moments approach where the theoretical cross spectra associated with the dynamic model is equated with the sample cross spectra.

- For resting-state fMRI this stochastic DCM model can be fit in the time-domain using a mean-field variational Bayes approach (see, e.g., Li et al., 2011, *Neuroimage*) which involves inference on both model parameters and latent states.
- The model can also be fit in the spectral domain using an approach known as spectral DCM (Friston et al., 2014, *Neuroimage*).
- The latter approach is somewhat akin to a method of moments approach where the theoretical cross spectra associated with the dynamic model is equated with the sample cross spectra.
- Developed in Friston et al. (2014), it assumes a parameterized power law form for the spectral densities of the noise terms in the state-space model and then express the empirical cross spectra as the sum of the theoretical cross spectra and Gaussian error.

- This then yields a Gaussian likelihood for the empirical cross spectra depending on the time-invariant parameters but not depending on the latent variables $\boldsymbol{x}(t)$. This likelihood is then combined with a prior distribution for these parameters and an approximation to a posterior distribution for these parameters is obtained using variational Bayes.

- This then yields a Gaussian likelihood for the empirical cross spectra depending on the time-invariant parameters but not depending on the latent variables $\boldsymbol{x}(t)$. This likelihood is then combined with a prior distribution for these parameters and an approximation to a posterior distribution for these parameters is obtained using variational Bayes.
- Interestingly, Razi et al. (2015, *Neuroimage*) report simulation results that demonstrate estimators obtained from spectral DCM having higher accuracy (in the sense of mean-squared error) than those obtained from stochastic DCM. In addition the former has a higher computational efficiency since estimation of the latent states is not required.

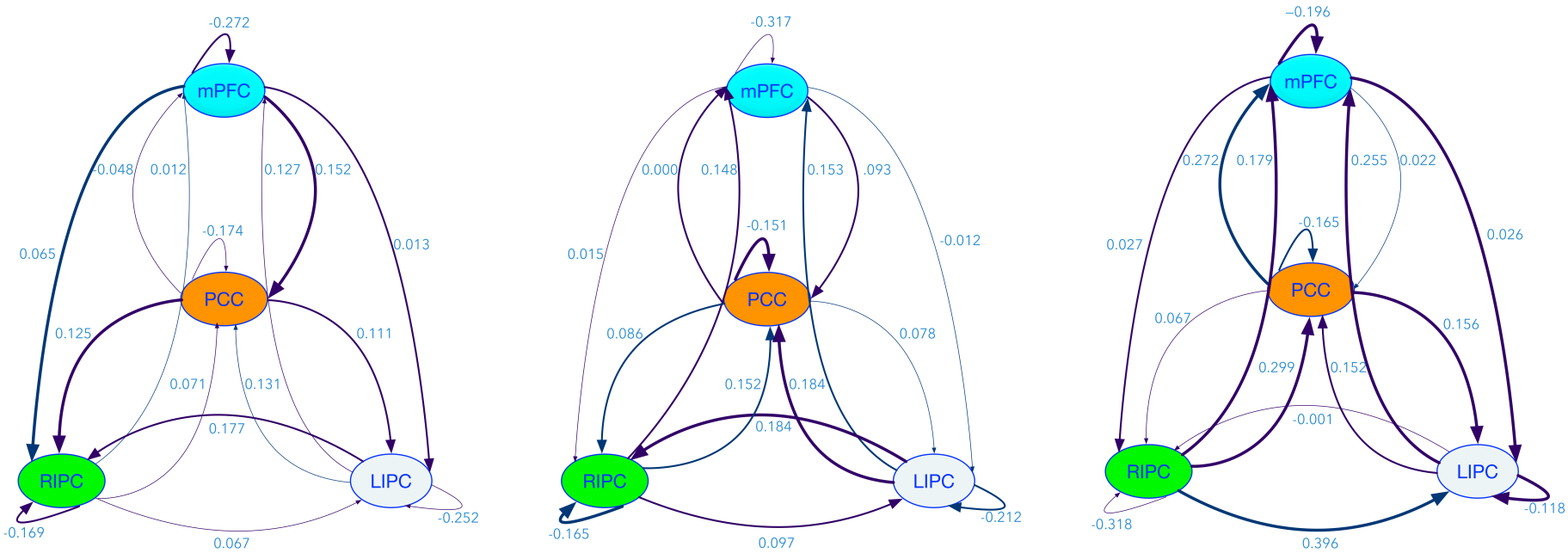


Figure: Average connectivity weights, left to right: normal controls, MCI, AD.

1. Average estimated connection from RIPC to LIPC is higher for subjects with AD (0.396) when compared with subjects having MCI (0.097) and NC subjects (0.067).
2. Average connectivity from PCC to MPFC is higher for AD subjects (0.272) than MCI (0.00) and NC (-0.048) subjects.

- Next use regression to examine association between disease status and connectivity at each edge of the network.
- Fit a set of 16 linear models where each model corresponds to one edge in the network; the response for each regression model is the value associated with that edge as estimated by spectral DCM.
- Here we have 112 networks, 16 edges; disease group (CN/MCI/AD) is included in each model as a categorical variable
- Additional variables representing age, gender, right/left handedness, and education are included to adjust for potential confounding.
- An F-test is conducted at each edge to assess the significance of disease on effective connectivity at that given edge.

- In our case the smallest p-values across the network correspond to $p = 0.081$ for the connection from RIPC to LIPC, and $p = 0.087$ for the connection from PCC to mPFC.

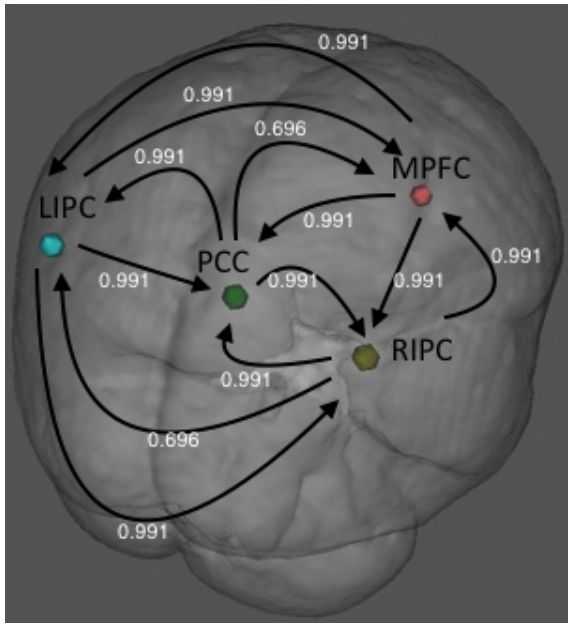


Figure: Network of FDR adjusted p-values from an analysis of covariance testing for the effect of disease group on edge weight while adjusting for age, gender, right/left hand, and education. The self-connections have FDR adjusted p-values 0.991 (LIPC), 0.991 (RIPC), 0.991 (PCC) and 0.991 (mPFC).

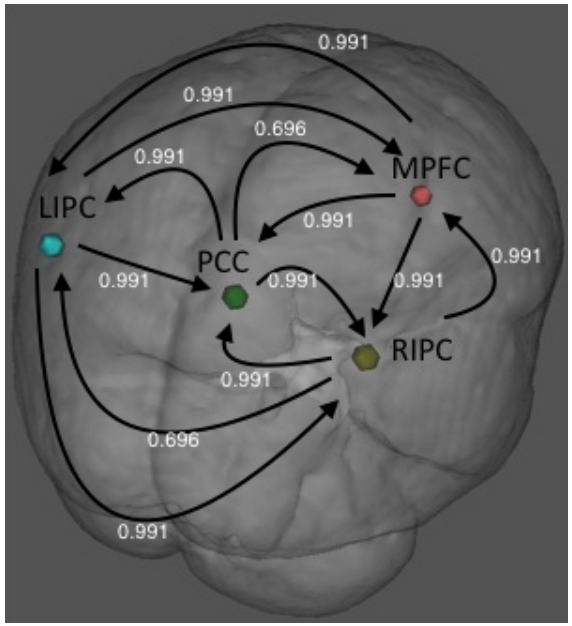


Figure: Network of FDR adjusted p-values from an analysis of covariance testing for the effect of disease group on edge weight while adjusting for age, gender, right/left hand, and education.

- The point estimates suggest the connections from RIPC to LIPC and PCC to mPFC may be higher in those subjects with AD.
- After correcting for multiple comparisons the FDR adjusted p-values do not indicate an association between connection strength and disease at any connection.

- **Study II:** We next examine the relationship between the probability of disease and genetics.

- **Study II:** We next examine the relationship between the probability of disease and genetics.
- The freely available software package PLINK (Purcell et.al., 2007) was used for genomic quality control.

- **Study II:** We next examine the relationship between the probability of disease and genetics.
- The freely available software package PLINK (Purcell et.al., 2007) was used for genomic quality control.
- The genetic data are SNPs from non-sex chromosomes, i.e., chromosome 1 to chromosome 22.

- **Study II:** We next examine the relationship between the probability of disease and genetics.
- The freely available software package PLINK (Purcell et.al., 2007) was used for genomic quality control.
- The genetic data are SNPs from non-sex chromosomes, i.e., chromosome 1 to chromosome 22.
- SNPs with minor allele frequency less than 5% are removed as are SNPs with a Hardy-Weinberg equilibrium p-value lower than 10^{-6} or a missing rate greater than 5%.

- **Study II:** We next examine the relationship between the probability of disease and genetics.
- The freely available software package PLINK (Purcell et.al., 2007) was used for genomic quality control.
- The genetic data are SNPs from non-sex chromosomes, i.e., chromosome 1 to chromosome 22.
- SNPs with minor allele frequency less than 5% are removed as are SNPs with a Hardy-Weinberg equilibrium p-value lower than 10^{-6} or a missing rate greater than 5%.
- After preprocessing we are left with 1,220,955 SNPs for each of the 112 subjects.

- We conduct a genome-wide association study (GWAS) with the goal of identifying a smaller subset of SNPs that are related to disease (CN/MCI/AD).

- We conduct a genome-wide association study (GWAS) with the goal of identifying a smaller subset of SNPs that are related to disease (CN/MCI/AD).
- A multinomial logistic regression with disease category as the response is fit for each SNP to assess that SNPs marginal association with disease.

- We conduct a genome-wide association study (GWAS) with the goal of identifying a smaller subset of SNPs that are related to disease (CN/MCI/AD).
- A multinomial logistic regression with disease category as the response is fit for each SNP to assess that SNPs marginal association with disease.
- We adjust adjusting for covariates representing age, sex, right hand or left hand, and education.

- We conduct a genome-wide association study (GWAS) with the goal of identifying a smaller subset of SNPs that are related to disease (CN/MCI/AD).
- A multinomial logistic regression with disease category as the response is fit for each SNP to assess that SNPs marginal association with disease.
- We adjust adjusting for covariates representing age, sex, right hand or left hand, and education.
- We sort the SNPs by the resulting p-values from a likelihood ratio test, where the null hypothesis corresponds to the case where the probability distribution of disease does not depend on the given SNP.

- We conduct a genome-wide association study (GWAS) with the goal of identifying a smaller subset of SNPs that are related to disease (CN/MCI/AD).
- A multinomial logistic regression with disease category as the response is fit for each SNP to assess that SNPs marginal association with disease.
- We adjust adjusting for covariates representing age, sex, right hand or left hand, and education.
- We sort the SNPs by the resulting p-values from a likelihood ratio test, where the null hypothesis corresponds to the case where the probability distribution of disease does not depend on the given SNP.
- A subset of the top 100 SNPs is selected based on this ranking.

The distribution of p-values by chromosome and the cut-off for selecting the best subset of 100.

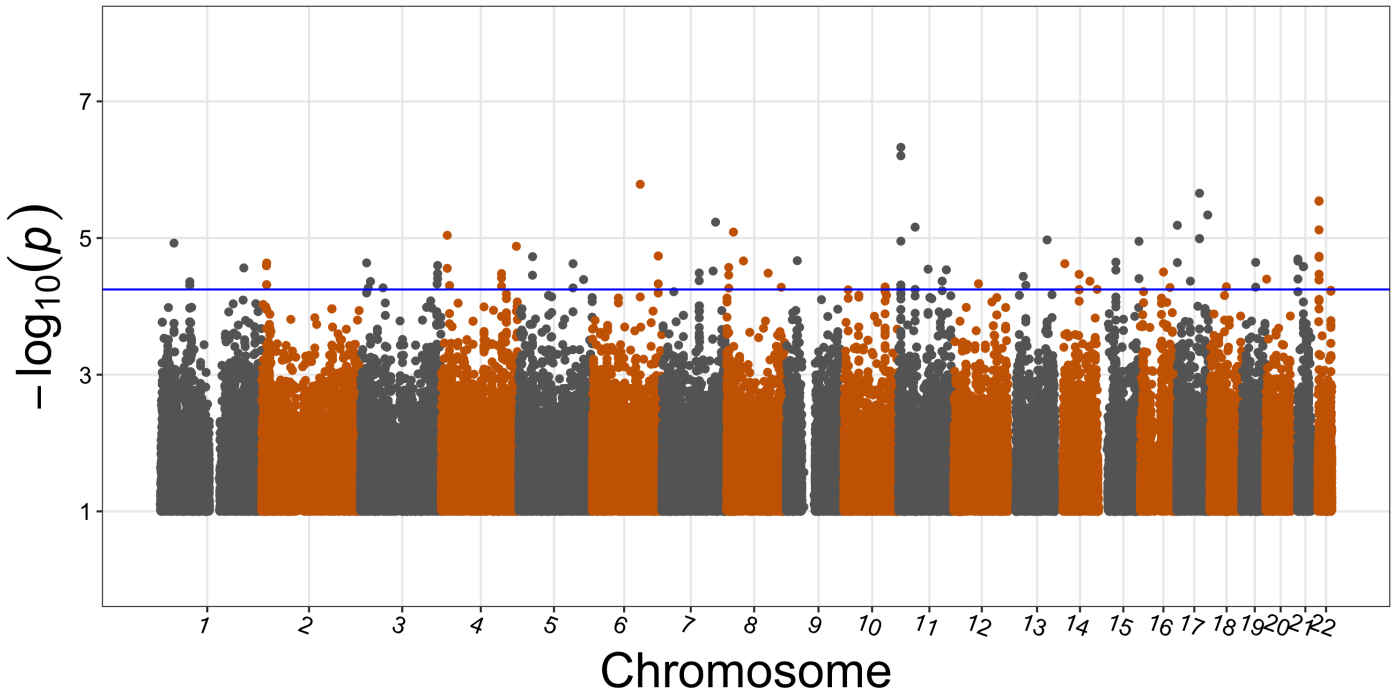


Figure: The p-values associating disease status with SNPs adjusting for age, sex, right or left hand and education. The blue line represents the cutoff used to obtain the top 100 SNPs.

- To jointly assess the effect of the 100 priority SNPs, we include these variables along with potential confounders as covariates and the disease status (CN/MCI/AD) as the response in a new symmetric multinomial logistic regression with LASSO penalty.

- To jointly assess the effect of the 100 priority SNPs, we include these variables along with potential confounders as covariates and the disease status (CN/MCI/AD) as the response in a new symmetric multinomial logistic regression with LASSO penalty.
- The goal is to jointly identify the SNPs that are most related to disease using LASSO.

- We use the glmnet software (Friedman et al., 2010) to fit the LASSO-penalized symmetric multinomial logistic regression model containing the top 100 SNPs and potential confounders.

- We use the glmnet software (Friedman et al., 2010) to fit the LASSO-penalized symmetric multinomial logistic regression model containing the top 100 SNPs and potential confounders.
- The model relates the probability of each class of disease to these variables (linear on log-odds scale).

- We use the glmnet software (Friedman et al., 2010) to fit the LASSO-penalized symmetric multinomial logistic regression model containing the top 100 SNPs and potential confounders.
- The model relates the probability of each class of disease to these variables (linear on log-odds scale).
- The Lasso penalty parameter is selected using 10-fold cross validation.

CN - Top SNPs		MCI - Top SNPs		AD - Top SNPs	
ID	$\hat{\beta}$	ID	$\hat{\beta}$	ID	$\hat{\beta}$
kgp324005_G	1.26	rs341080_T	-2.39	kgp5568290_G	1.88
rs175174_G	-0.93	rs442277_A	-1.62	rs7617199_A	1.36
rs9346859_C	0.90	kgp10950119_T	1.07	kgp239829_C	1.33
rs2759295_G	-0.86	rs4875544_T	0.96	rs10900494_G	-1.21
rs2929981_G	0.83	kgp8323161_C	0.92	rs7781597_C	1.00

Table: Table of estimates showing the top five SNPs for each disease category along with the estimated regression coefficient from multinomial logistic regression with Lasso penalty.

These findings suggest that SNPs kgp5568290, rs7617199 and **kgp239829** have the largest estimated effects on the log-odds of AD.

- Note: our top subset of 100 SNPs did not contain any from the APOE gene.

- Note: our top subset of 100 SNPs did not contain any from the APOE gene.
- Our data have two SNPs rs769449 and kgp2187574 that are located within the range of gene APOE (chr19:45409039, chr19:45412650).

- Note: our top subset of 100 SNPs did not contain any from the APOE gene.
- Our data have two SNPs rs769449 and kgp2187574 that are located within the range of gene APOE (chr19:45409039, chr19:45412650).
- The individual associations between these two SNPs with the disease are not found to be significant, where the specific results are p-value $= 0.07$ for rs769449 and p-value $= 0.57$ for kgp2187574.

- Note: our top subset of 100 SNPs did not contain any from the APOE gene.
- Our data have two SNPs rs769449 and kgp2187574 that are located within the range of gene APOE (chr19:45409039, chr19:45412650).
- The individual associations between these two SNPs with the disease are not found to be significant, where the specific results are p-value $= 0.07$ for rs769449 and p-value $= 0.57$ for kgp2187574.
- As a result, neither of these two APOE SNPs are within the top 100 SNPs that we have identified and included in the larger model as their rank is quite low.

- Note: our top subset of 100 SNPs did not contain any from the APOE gene.
- Our data have two SNPs rs769449 and kgp2187574 that are located within the range of gene APOE (chr19:45409039, chr19:45412650).
- The individual associations between these two SNPs with the disease are not found to be significant, where the specific results are p-value $= 0.07$ for rs769449 and p-value $= 0.57$ for kgp2187574.
- As a result, neither of these two APOE SNPs are within the top 100 SNPs that we have identified and included in the larger model as their rank is quite low.
- If we define APOE related SNPs as those within a 1 million base pair range of APOE, we find 503 SNPs with p-values between 0.0005 and 0.99. In this case none of these 503 SNPs fall within the top 100 SNPs, though the highest rank among these is a rank of 1002 obtained by SNP kgp3912453.

- Given this, we look for more SNPs that might not be in the neighbourhood of APOE, but may still be related to the APOE SNPs.

- Given this, we look for more SNPs that might not be in the neighbourhood of APOE, but may still be related to the APOE SNPs.
- To do this we start with the top 100 SNPs that are associated with disease status and we correlate them with the first principal component obtained from those 503 SNPs that are in the 1 million base pair window of APOE.

- Given this, we look for more SNPs that might not be in the neighbourhood of APOE, but may still be related to the APOE SNPs.
- To do this we start with the top 100 SNPs that are associated with disease status and we correlate them with the first principal component obtained from those 503 SNPs that are in the 1 million base pair window of APOE.
- In this case the largest correlation is only about 0.28.

- Given this, we look for more SNPs that might not be in the neighbourhood of APOE, but may still be related to the APOE SNPs.
- To do this we start with the top 100 SNPs that are associated with disease status and we correlate them with the first principal component obtained from those 503 SNPs that are in the 1 million base pair window of APOE.
- In this case the largest correlation is only about 0.28.
- Out of the 1,220,955 SNPs there are no APOE SNPs in our top 100.

- We do have the expected APOEe4 signal in our data but APOE SNPs do not make it into our top 100 SNPs.

Table: Distribution of Demographic Variables

		AD	MCI	NL	p-value
n		12	63	37	
APOE Gene No.e4 (%)	Zero	1 (8.3)	35 (55.6)	24 (64.9)	0.0052
	One	9 (75.0)	22 (34.9)	12 (32.4)	
	Two	2 (16.7)	6 (9.5)	1 (2.7)	

- **Brain Connectivity** $\sim$ **Genetics**: We relate the resting-state effective connectivity networks estimated using DCM to each of the 100 priority SNPs using the multivariate approach proposed in Marchini et al. (2007).

- **Brain Connectivity** $\sim$ **Genetics**: We relate the resting-state effective connectivity networks estimated using DCM to each of the 100 priority SNPs using the multivariate approach proposed in Marchini et al. (2007).
- This approach tests for SNP effects on a multi-dimensional phenotype using the Bayes factor. The test is based on the linear model

$$\mathbf{y}_i = x_i\boldsymbol{\beta} + \mathbf{e}_i, \;\; \mathbf{e}_i \overset{iid}{\sim} N_{16}(0, \Sigma)$$

where $\mathbf{y}_i = (y_{i_1}, \ldots, y_{i_{16}})^T$ denotes the vector of residual estimated brain connectivity edges for the ith individual.

- **Brain Connectivity** $\sim$ **Genetics**: We relate the resting-state effective connectivity networks estimated using DCM to each of the 100 priority SNPs using the multivariate approach proposed in Marchini et al. (2007).
- This approach tests for SNP effects on a multi-dimensional phenotype using the Bayes factor. The test is based on the linear model

$$\mathbf{y}_i = x_i\boldsymbol{\beta} + \mathbf{e}_i, \;\; \mathbf{e}_i \overset{iid}{\sim} N_{16}(0, \Sigma)$$

  where $\mathbf{y}_i = (y_{i_1}, \ldots, y_{i_{16}})^T$ denotes the vector of residual estimated brain connectivity edges for the ith individual.
- Residual connectivity is calculated by subtracting a baseline term based on linear regression estimates of an overall mean and the effects of covariates sex, age, left or right hand and education

- **Brain Connectivity** $\sim$ **Genetics**: We relate the resting-state effective connectivity networks estimated using DCM to each of the 100 priority SNPs using the multivariate approach proposed in Marchini et al. (2007).
- This approach tests for SNP effects on a multi-dimensional phenotype using the Bayes factor. The test is based on the linear model

$$\mathbf{y}_i = x_i\boldsymbol{\beta} + \mathbf{e}_i, \;\; \mathbf{e}_i \overset{iid}{\sim} N_{16}(0, \Sigma)$$

  where $\mathbf{y}_i = (y_{i_1}, \ldots, y_{i_{16}})^T$ denotes the vector of residual estimated brain connectivity edges for the ith individual.
- Residual connectivity is calculated by subtracting a baseline term based on linear regression estimates of an overall mean and the effects of covariates sex, age, left or right hand and education
- x_i represents the number of a particular allele for a given SNP for the i^{th} individual; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{16})^T$ are the parameters relating this SNP to all of the edge values of the network.

- The null model corresponds to $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The alternative is $H_1 : \exists j, \;\; s.t. \; \beta_j \neq 0$.

- The null model corresponds to $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The alternative is $H_1 : \exists j, \;\; s.t. \; \beta_j \neq 0$.
- The Bayes factor (Kass and Raftery, 1995; Wagenmakers, 2007; Masson, 2011; Nathoo and Masson, 2015) is used to assess the evidence in favour of H_1.

- The null model corresponds to $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The alternative is $H_1 : \exists j, \;\; s.t. \; \beta_j \neq 0$.
- The Bayes factor (Kass and Raftery, 1995; Wagenmakers, 2007; Masson, 2011; Nathoo and Masson, 2015) is used to assess the evidence in favour of H_1.
- The Bayes factor is defined by the ratio of marginal likelihoods $BF = m(y|H_1)/m(y|H_0)$, and quantifies the strength of evidence in favour of H_1; that is, the hypothesis that at least one network edge depends on the included SNP.

- The null model corresponds to $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The alternative is $H_1 : \exists j, \;\; s.t. \; \beta_j \neq 0$.
- The Bayes factor (Kass and Raftery, 1995; Wagenmakers, 2007; Masson, 2011; Nathoo and Masson, 2015) is used to assess the evidence in favour of H_1.
- The Bayes factor is defined by the ratio of marginal likelihoods $BF = m(y|H_1)/m(y|H_0)$, and quantifies the strength of evidence in favour of H_1; that is, the hypothesis that at least one network edge depends on the included SNP.
- The model is fit and the Bayes factor computed separately for each of the 100 priority SNPs.

The top 10 SNPs with the corresponding values of their Bayes factor.

SNP id	Bayes Factor
rs11036476	69.18
kgp239829	54.95
rs4910743	14.13
rs7482144	11.75
rs2855039	10.96
kgp10652659	10.23
kgp3458705	9.55
kgp5413383	6.61
kgp4994042	6.61
kgp161879	6.46

Table: The top 10 SNPs (based on the value of the Bayes factor) related to effective brain connectivity within the four regions of the DMN considered. These SNPs are selected from the 100 SNPs identified as most related to disease in Study II.

- There are two distinct SNPs, rs11036476 (Bayes factor = 69.18) and kgp239829 (Bayes factor = 54.95), which give values of the Bayes factor indicating very strong evidence that effective connectivity in the DMN depends on these SNPs.

- There are two distinct SNPs, rs11036476 (Bayes factor $= 69.18$) and kgp239829 (Bayes factor $= 54.95$), which give values of the Bayes factor indicating very strong evidence that effective connectivity in the DMN depends on these SNPs.
- Among the top 10 SNPs these two SNPs have Bayes factors that are noticeably larger than the others.

- There are two distinct SNPs, rs11036476 (Bayes factor $= 69.18$) and kgp239829 (Bayes factor $= 54.95$), which give values of the Bayes factor indicating very strong evidence that effective connectivity in the DMN depends on these SNPs.
- Among the top 10 SNPs these two SNPs have Bayes factors that are noticeably larger than the others.
- As we shall see, they both correspond to the same genetic signal from chromosome 11 from two neighbouring SNPs in high linkage.

Importantly, SNP kgp239829 was also one of the top three SNPs found related to AD.

Table: Study III: Effective Connectivity $\sim$ Genetics

SNP id	Bayes Factor
rs11036476	69.18
kgp239829	**54.95**
rs4910743	14.13
rs7482144	11.75
rs2855039	10.96
kgp10652659	10.23
kgp3458705	9.55
kgp5413383	6.61
kgp4994042	6.61
kgp161879	6.46

Table: Study II: Disease $\sim$ Genetics

CN - Top SNPs		MCI - Top SNPs		AD - Top SNPs	
ID	$\hat{\beta}$	ID	$\hat{\beta}$	ID	$\hat{\beta}$
kgp324005_G	1.26	rs341080_T	-2.39	kgp5568290_G	1.88
rs175174_G	-0.93	rs442277_A	-1.62	rs7617199_A	1.36
rs9346859_C	0.90	kgp10950119_T	1.07	**kgp239829_C**	**1.33**
rs2759295_G	-0.86	rs4875544_T	0.96	rs10900494_G	-1.21
rs2929981_G	0.83	kgp8323161_C	0.92	rs7781597_C	1.00

- To further investigate what specific connection in the networks depend on each SNP, we conduct a posthoc analysis independently for each of the two top SNPs.

- To further investigate what specific connection in the networks depend on each SNP, we conduct a posthoc analysis independently for each of the two top SNPs.
- We associate each edge value to a given SNP by fitting a linear regression of the edge value onto the SNP while also including covariates representing sex, age, left or right hand and education.

- To further investigate what specific connection in the networks depend on each SNP, we conduct a posthoc analysis independently for each of the two top SNPs.
- We associate each edge value to a given SNP by fitting a linear regression of the edge value onto the SNP while also including covariates representing sex, age, left or right hand and education.
- The importance of the given SNP to each edge is then summarized through a p-value from an F-test.

- Examining the networks for both SNPs and the p-values arising from the posthoc analysis relating each SNP to the 16 specific connections of interest, we find looking at both networks that these two SNPs are both strongly related to the same connections
- MPFC $\rightarrow$ LIPC (rs11036476, p-value $= 0.0055$; kgp239829, p-value $= 0.0052$) and LIPC $\rightarrow$ RIPC (rs11036476, p-value $= 0.0058$; kgp239829, p-value $= 0.0063$)

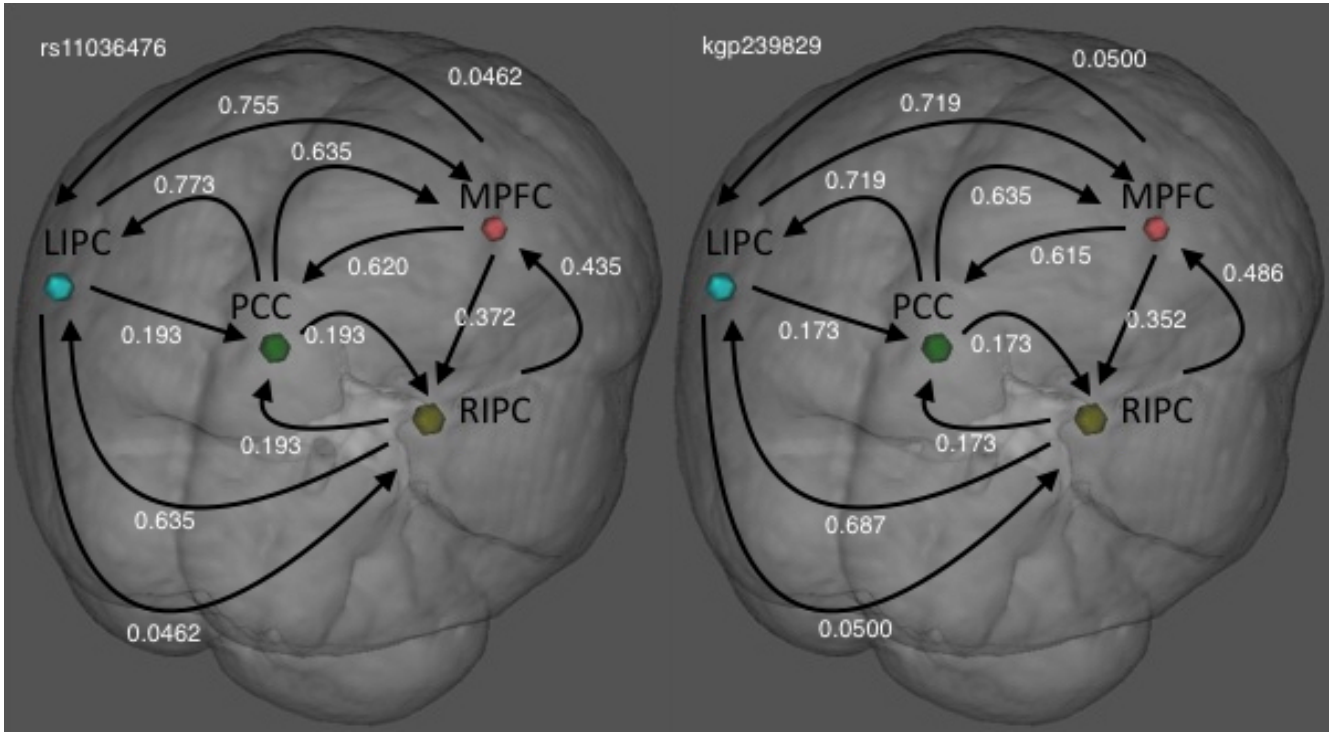


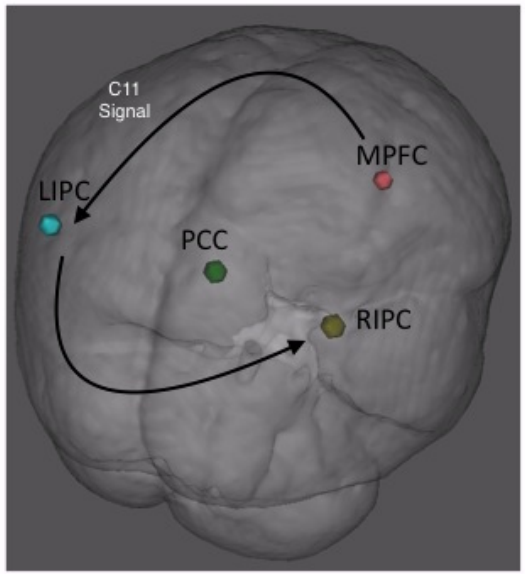
Figure: The FDR adjusted p-values obtained from F-tests associating effective connectivity between the brain regions of interest with each of the top two SNPs, rs11036476 (Bayes factor = 69.18) and kgp239829 (Bayes factor = 54.95), where the top 2 SNPs are identified by applying a multivariate linear model and the Bayes factor.

- Investigating further, we find that both rs11036476 and kgp239829 are very close to each other on chromosome 11 and their sample correlation is close to $\hat{\rho} = 1$, so that both SNPs represent the same genetic signal from chromosome 11 onto DMN connections MPFC $\rightarrow$ LIPC and LIPC $\rightarrow$ RIPC.

- Investigating further, we find that both rs11036476 and kgp239829 are very close to each other on chromosome 11 and their sample correlation is close to $\hat{\rho} = 1$, so that both SNPs represent the same genetic signal from chromosome 11 onto DMN connections MPFC $\rightarrow$ LIPC and LIPC $\rightarrow$ RIPC.
- As noted earlier, this same signal is also related to AD and thus suggests a potential a biomarker for AD that may have a relationship to DMN connections MPFC $\rightarrow$ LIPC and LIPC $\rightarrow$ RIPC.

- Investigating further, we find that both rs11036476 and kgp239829 are very close to each other on chromosome 11 and their sample correlation is close to $\hat{\rho} = 1$, so that both SNPs represent the same genetic signal from chromosome 11 onto DMN connections MPFC $\rightarrow$ LIPC and LIPC $\rightarrow$ RIPC.
- As noted earlier, this same signal is also related to AD and thus suggests a potential a biomarker for AD that may have a relationship to DMN connections MPFC $\rightarrow$ LIPC and LIPC $\rightarrow$ RIPC.
- Worth noting that the SNP with the third highest value of the Bayes factor rs4910743 (Bayes factor $= 14.13$) also exhibits a high correlation of $\hat{\rho} \approx 0.93$ with the top two SNPs and its network obtained from a posthoc analysis exhibits a very similar pattern.

- This further suggests the same genetic signal potentially related to AD and potentially modulating DMN connections MPFC $\rightarrow$ LIPC and LIPC $\rightarrow$ RIPC. **Study III**: *Potential path of information flow within the brain related to the identified c11 signal.* **Study II**: *The same genetic signal is identified by LASSO as potentially associated with Alzheimer's disease.*



- We also note that while only one (kgp239829) of the top three SNPs with highest Bayes factors is also listed as having a large estimated effect on the probability of AD, this is likely the result of a well known property of LASSO.

- We also note that while only one (kgp239829) of the top three SNPs with highest Bayes factors is also listed as having a large estimated effect on the probability of AD, this is likely the result of a well known property of LASSO.
- Namely, the property of the LASSO to choose only one of a group of highly correlated predictors to include in the model and drop the others.

Importantly, SNP kgp239829 was also one of the top three SNPs found related to AD.

Table: Study III: Effective Connectivity $\sim$ Genetics

SNP id	Bayes Factor
rs11036476	69.18
kgp239829	**54.95**
rs4910743	14.13
rs7482144	11.75
rs2855039	10.96
kgp10652659	10.23
kgp3458705	9.55
kgp5413383	6.61
kgp4994042	6.61
kgp161879	6.46

Chromosome 11 Signal on DMN

Table: Study II: Disease $\sim$ Genetics

CN - Top SNPs		MCI - Top SNPs		AD - Top SNPs	
ID	$\hat{\beta}$	ID	$\hat{\beta}$	ID	$\hat{\beta}$
kgp324005_G	1.26	rs341080_T	-2.39	kgp5568290_G	1.88
rs175174_G	-0.93	rs442277_A	-1.62	rs7617199_A	1.36
rs9346859_C	0.90	kgp10950119_T	1.07	**kgp239829_C**	**1.33**
rs2759295_G	-0.86	rs4875544_T	0.96	rs10900494_G	-1.21
rs2929981_G	0.83	kgp8323161_C	0.92	rs7781597_C	1.00

Single Representative SNP Picked by Lasso

Top 3 SNPs related to DMN brain connectivity all from chromosome 11 and all in very high linkage. This genetic signal is also found by LASSO to be related to the probability of AD.

Table: Study III: Effective Connectivity $\sim$ Genetics

SNP id	Bayes Factor
rs11036476	**69.18**
kgp239829	**54.95**
rs4910743	**14.13**
rs7482144	11.75
rs2855039	10.96
kgp10652659	10.23
kgp3458705	9.55
kgp5413383	6.61
kgp4994042	6.61
kgp161879	6.46

Summary:

- We have developed a connectome genetics pipeline with two key innovations:

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics
  2. It is the first use of the Bayes factor in imaging genetics

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics
  2. It is the first use of the Bayes factor in imaging genetics
- Starting with 1,220,955 SNPs, we have found a genetic signal from chromosome 11 that in separate analyses looking at different traits is found to both modulate the probability of AD and effective connectivity in the DMN.

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics
  2. It is the first use of the Bayes factor in imaging genetics
- Starting with 1,220,955 SNPs, we have found a genetic signal from chromosome 11 that in separate analyses looking at different traits is found to both modulate the probability of AD and effective connectivity in the DMN.
- Some caution is required here since the number of AD subjects in our study is low.

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics
  2. It is the first use of the Bayes factor in imaging genetics
- Starting with 1,220,955 SNPs, we have found a genetic signal from chromosome 11 that in separate analyses looking at different traits is found to both modulate the probability of AD and effective connectivity in the DMN.
- Some caution is required here since the number of AD subjects in our study is low.
  1. The result for AD is based on LASSO with $n_{AD} = 12$ in this category for the multinomial logistic regression and $n = 112$ subjects in the model.

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics
  2. It is the first use of the Bayes factor in imaging genetics
- Starting with 1,220,955 SNPs, we have found a genetic signal from chromosome 11 that in separate analyses looking at different traits is found to both modulate the probability of AD and effective connectivity in the DMN.
- Some caution is required here since the number of AD subjects in our study is low.
  1. The result for AD is based on LASSO with $n_{AD} = 12$ in this category for the multinomial logistic regression and $n = 112$ subjects in the model.
  2. The results for effective connectivity are based on the Bayes factor ($n = 112$).

Summary:

- We have developed a connectome genetics pipeline with two key innovations:
  1. It is the first example of the use of DCM in imaging genetics
  2. It is the first use of the Bayes factor in imaging genetics
- Starting with 1,220,955 SNPs, we have found a genetic signal from chromosome 11 that in separate analyses looking at different traits is found to both modulate the probability of AD and effective connectivity in the DMN.
- Some caution is required here since the number of AD subjects in our study is low.
  1. The result for AD is based on LASSO with $n_{AD} = 12$ in this category for the multinomial logistic regression and $n = 112$ subjects in the model.
  2. The results for effective connectivity are based on the Bayes factor ($n = 112$).
  3. In our current work we are examining replication and stability.

- We have only considered baseline rs-fMRI data ($n = 112$ scans) and the resulting brain networks in this study.

- We have only considered baseline rs-fMRI data ($n = 112$ scans) and the resulting brain networks in this study.
- Longitudinal rs-fMRI data are available ($n \approx 500$ scans) from ADNI2 and we are working on a longitudinal analysis linking each subject's time sequence of effective connectivity networks to genetics.

- We have only considered baseline rs-fMRI data ($n = 112$ scans) and the resulting brain networks in this study.
- Longitudinal rs-fMRI data are available ($n \approx 500$ scans) from ADNI2 and we are working on a longitudinal analysis linking each subject's time sequence of effective connectivity networks to genetics.
- Estimate white matter connectivity from diffusion tensor imaging data and link to genetics using the pipeline developed here.

Statistical Issues

- Incorporate shrinkage estimation for effective connectivity networks

Statistical Issues

- Incorporate shrinkage estimation for effective connectivity networks
- Incorporate detection of weak genetic signals and investigate interactions

Statistical Issues

- Incorporate shrinkage estimation for effective connectivity networks
- Incorporate detection of weak genetic signals and investigate interactions
- Interesting but perhaps not extremely useful: use the DTI data for subjects that have it to parameterize a computer model that will simulate rs-fMRI data. Then combine the real rs-fMRI data with the computer model generate data.

Statistical Issues

- Incorporate shrinkage estimation for effective connectivity networks
- Incorporate detection of weak genetic signals and investigate interactions
- Interesting but perhaps not extremely useful: use the DTI data for subjects that have it to parameterize a computer model that will simulate rs-fMRI data. Then combine the real rs-fMRI data with the computer model generate data.
  1. Combine the real and simulated fMRI data and run the pipeline as is.

Statistical Issues

- Incorporate shrinkage estimation for effective connectivity networks
- Incorporate detection of weak genetic signals and investigate interactions
- Interesting but perhaps not extremely useful: use the DTI data for subjects that have it to parameterize a computer model that will simulate rs-fMRI data. Then combine the real rs-fMRI data with the computer model generate data.
  1. Combine the real and simulated fMRI data and run the pipeline as is.
  2. Develop an approach for optimal bias-variance trade-off in the analysis of the combined data. This is often done in climate modelling but has not for neuroimaging.

- Look at other resting-state networks - effective connectivity $\sim$ genetics:
  1. **Extended Default Mode Network:** Posterior cingulate/Precuneus, Medial Prefrontal, Left lateral parietal, Right lateral parietal, Left inferior temporal, Right inferior temporal, left hippocampus, right hippocampus
  2. **Executive Control Network:** right dorsolateral prefrontal cortex, Left dorsolateral prefrontal cortex, right ventrolateral prefrontal cortex, left ventrolateral prefrontal cortex, right medial prefrontal cortex (mPFC), right caudate nucleus (CAU), right lateral parietal (LP), left lateral parietal (LP)

Acknowledgement - Funding

- CANSSI Collaborative Research Team on Neuroimaging Statistics
- NSERC Discovery Grant
- NSERC Tier II Canada Research Chair in Biostatistics
- South Africa - Canada Research Chairs Mobility Initiative Grant
- UVic Internal Research Grant

Canada Research Chairs
Chaires de recherche du Canada

Canada