Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Junhan Fang Date

Approval Sheet

Between-group comparisons of structural and functional brain connectivity

By

Junhan Fang
Master of Science of Public Health
Biostatistics

_____

_____

Ying Guo, Ph.D., Suprateek Kundu, Ph.D.
Thesis Advisors

_____

Helen Mayberg, MD
Thesis Reader

Abstract Cover Page

Between-group comparisons of structural and functional brain connectivity

By

Junhan Fang

B.A.
Southwestern University of Finance and Economics
2014

Thesis Advisors: Ying Guo, Ph.D.
Suprateek Kundu, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2016

Abstract

Group-level comparisons in structural and functional brain connectivity between two groups of subjects

By Junhan Fang

This thesis mainly focused on detecting the difference of brain connectivity between diseased vs. normal subjects in Philadelphia Neurodevelopmental Cohort (PNC) study. Multimodal neuroimaging including fMRI and diffusion MRI was used to investigate group differences in functional connectivity and structural connectivity with the brain. Probabilistic tractography was performed to estimate structural connectivity and partial correlations of fMRI time series data were used to estimate functional connectivity in subjects' brains. Edge-wise linear regression was then performed to detect the difference of structural and functional connectivity between two groups of subjects. Significant between-group differences were found in 105 region pairs for structural connectivity and in 264 regions pairs for functional connectivity. Two edge-wise linear regression were performed to detect the potential relationship between structural and functional connectivity for each group. More than 150 edges were detected having significant correlation between structural and functional connectivity.

In the second part of the thesis, five graph theoretical metrics in graph theory were evaluated to show the network properties of structural and functional connectivity in each group. The linear regressions, which aim to detect the difference network properties between two groups of subjects, were performed. According to the linear models, characteristic path length had a significantly different between the diseased and normal subjects.

Lastly, a recently developed clustering method, which is one of differentially expressed network methods, was performed to identify clusters of regions that showed significantly different structural and functional connectivity between two groups of subjects. We provided four clustered edges' networks for structural connectivity and functional connectivity to show the significant difference between two groups of subjects.

Abstract Cover Page

Between-group comparisons of structural and functional brain connectivity

By

Junhan Fang

B.A
Southwestern University of Finance and Economics
2014

Thesis Advisors: Ying Guo, Ph.D., Suprateek Kundu, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2016

Contents

List of Tables:

List of Figures:

1. Introduction

Comparing group level difference of the brain network by combining structural connectivity and functional connectivity is a novel topic. Structural connectivity is the structural interrelationship among brain network regions, and functional connectivity is the functional interrelationships among brain network regions (Fingelkurts, Fingelkurts, & Kahkonen, 2005). Different ways to combining structural connectivity and functional connectivity will give diverse analysis angles for brain networks (Rykhlevskaia, 2008).

The existence of fibers which connecting directly between different regions reveal the structural connectivity. True pathways for fibers in people's brain is impossible to observed when he is alive. As a result, diffusion tensor tractography can reconstruct the fiber bundles through diffusion imaging data. Deterministic fiber tracking techniques can be conducted using FACT (fiber assignment by continuous tracking) method (Mori. S, 1999) and the streamline tracking algorithm (Basser PJ, 2000) and other improved methods. However, deterministic fiber tracking methods could ignore that there are multiple orientations of finer bundles in human's brain. With the improvement of diffusion tensor imaging techniques, probabilistic fiber tractography was modelled to represent the complicated multiple orientations of fiber bundles situation (Behrens,2013a and 2013b).

For functional connectivity, functional magnetic resonance imaging (fMRI) data was investigated to reveal human brain's functional connection in many studies (Biswal et al., 1995; Bullmore and Sporns, 2009; Deco et al., 2011; Satterthwaite et al., 2015). Specially, the analysis of resting-state fMRI data was quite important (Dosenbach et al, 2010). There are types of ways to represent the whole brain

networks, pairwise full correlation among regions and partial correlation among regions which excluding other regions' effects (Smith et al, 2011 and Smith et al, 2012). Full correlation, which is also as Pearson's correlation, evaluate the marginal correlation between regions so it cannot reveal the direct functional correlation between the regions. Partial correlation evaluates the direct functional correlation between paired regions after removing third or regions effects. However, it is hard to estimate partial correlations. There are several ways to estimate them, like methods in Merrelec et al. (2006) and Peng et al. (2009).

The relationship between structural connectivity and functional connectivity has not been detected clearly. But we have the hypothesis that functional connectivity can reflect the structural connectivity in some degree (Michael D. et al., 2009). Within the context of clinic research, the relationship between structural and functional connectivity are usually showed as the association between brain structure damages and the fMRI aftereffects. Like the study detected brain connectivity difference between autism spectrum disorder and health children in Rudie et al. (2008). Group-level brain connectivity and comparing group difference in brain connectivity were usually preformed using two sample testing after making some change in groups' brain connectivity. Graph theory analysis is a popular way to use for generating information from original brain connectivity (Rubinov, Sporns, 2010). Global network metric-based method (GMN) and differentially expressed network (DEN) are two main graph analysis methods. GNM use some metrics like, small-worldness to capture the properties of brain network and then produce the statistical testing or regression analysis across the subjects in different groups (Sporns, 2011, 2012; Van Den Heuvel et al., 2010). DEN usually perform edge-wise statistical

analysis at the group level first and then applying optimization algorithm (Zalesky et al., 2010, 2012b) to detect the group difference of brain connectivity.

In this paper, we aim to detect the difference of brain connectivity in disease and no disease group of subjects. Probabilistic tracktography based on the algorithm of Behrens (2013b) was performed for generating structural connectivity and Partial correlation was calculated using Wang et al. (2016) for obtaining the functional connectivity. Then edge-wise linear regression was conducted for structural connectivity and functional connectivity separately across the subjects in two groups independently. General group difference of brain connectivity can be detected using the results of the edge-wise linear regression models. The edge-wise linear regression was conducted again for structural connectivity and functional connectivity to obtain the correlations of SC and FC between each edges. Then for structural connectivity and functional connectivity, we conducted GMN (Rubinov, Sporns, 2010) and DEN methods (Chen et al., 2015) separately to show the difference of brain connectivity between two groups of subjects under a network angle.

2. Method

2.1 Data Acquisition

Our goal is detecting the potential difference of brain connectivity between disease group and no disease group in the Philadelphia Neurodevelopmental Cohort. The Philadelphia Neurodevelopmental Cohort (PNC) is a collaborative project between the Brain Behavior Laboratory at the University of Pennsylvania and the Children's Hospital of Philadelphia (CHOP), funded by NIMH through the American Recovery and Reinvestment Act of 2009, (Satterthwaite et al., 2014; Satterthwaite et al., 2015). The PNC study includes a population-based sample of over 9500 individuals aged 8-21 years selected among those who received medical care at the Children's Hospital of Philadelphia network in the greater Philadelphia area; the sample is stratified by sex, age and ethnicity.

2.1.1 Subjects

All the subjects were recruited through University of Pennsylvania's PNC study. Based on their inclusion criteria, there are total 1,145 subjects completed the MRI scanning progress. After pre-processing steps for Diffusion Tensor Imaging data and fMRI data, there are total 125 subjects which can be used for our analysis.

We define the subjects belong to no disease group if their medical rating is no medical problems or minor but no central nervous system impact. Then we define the subjects belong to disease group if their medical rating is moderate, significant or major problems. In the 125 subjects, 91 of them are belong to no disease group and 34 of them are belong to disease group. Table.1 shows the descriptive of two groups of subjects.

Table 1. Mean, standard deviation and range of sample descriptive

Characteristics	No Disease Group	Disease Group	P-value
Sample Size(N)	91	34	
Number of Females(N)	51	15	0.2426
Age	13.81±3.41	14.76±3.90	0.2228

2.1.2 MRI data

All MRI scans were acquired on a single 3T Siemens TIM Trio whole-body scanner located in the Hospital of the University of Pennsylvania. (Satterthwaite et al, 2014)

The DTI sequence consisted of 64 scans with different diffusion-weighted directions (b=1000 s/mm2), 7 scans with no diffusion sensitization, at b=0. Other parameters were TR=8100 ms, TE=82 ms, GRAPPA on, FOV=240x240 mm, matrix=128x128, with 70 slices, yielding an in-plane voxel dimension of 2×2mm with 2-mmthick axial slices, and total scan time=10 min 56 s.

The Resting State-fMRI (re-fMRI) scans were acquired with 124 volumes, TR=3000ms, TE=32 ms, flip angle=90°, FOV=192x192 mm, matrix 64x64 and effective voxel resolution=3.0x3.0x3.0 mm. The total scan time = 6 min 18 s.

More details of experimental settings and image acquisition can be found in Satterthwaite et al. (2014).

2.1.3 Data Pre-processing

The DTI data are preprocessed using FSL. Eddy current correction are performed using FDT toolbox in FSL for correcting distortions, simple head motions and using affine registration to reference volume. Skull striping is performed for brain extraction. Dtifit toolbox is used to fit diffusion tensor model on corrected data in order to obtaining diffusion tensors and fractional anisotropy (FA) for each subject.

Lastly, bedpostx toolbox is used to obtained the distributions of parameters in each voxel, which would be used in probabilistic tractography.

The rs-fMRI data are preprocessed using the preprocessing script released from the 1000 Functional Connectomes Project. Specifically, for remove extra-cranial material, skull stripping was performed on the T1 images. Then we remov the first four volumes of the functional time series to stabilize the signal. The anatomical image is registered to the 8th volume of the functional image and subsequently spatially assigned to the MNI standard brain space. These normalization parameters from MNI space are used for the functional images, which are smoothed using a Gaussian kernel of FWHM 6MM. Motion corrections were applied on the functional images. Nine standard confounding signals (6 motion parameters plus global / White Matter / cerebrospinal fluid) as well as the temporal derivative, quadratic term and temporal derivative of the quadratic of each were regressed out of data. Furthermore, motion-related spike regressors are included to bound the observed displacement. Lastly, the functional time series data are band-pass filtered to retain frequencies between 0.01 and 0.1 Hz which is the relevant frequency range for rs-fMRI. The functional time series were assigned into 90 brain regions using AAL90 Atlas.

Partial correlations were calculated between all the paired of AAL90 regions using the statistical method in Wang et al. (2016). Here, partial correlations are used to instead of full correlation because of its ability to reflect the correlation between two regions after removing other regions effects. It would help us to obtain the true relationship between any paired of regions.

2.2 Probabilistic Tracktography

Structural connectivity matrix can be obtained from the fibers in human's brain. The true fibers cannot be detected directly so a partial volume model is build to calculate the distribution of the fiber orientations of each voxel in human brain in Behrens et al (2007). Based on this model, fibers can be simulated between regions. As a result, simulated fibers numbers between paired of regions can show us the level of strength of two regions connected in order to show structural connectivity indirectly. The more simulated fibers between two regions, the higher probability the two regions have strong connection. In this paper, the structural connectivity represents the strength of two regions connected, which is measured by simulated fiber counts.

Using the partial volume model, the posterior distribution of fiber orientations in each voxel can be used to simulate the fiber tracks in human brain.

In this model, the diffusion-weighted MR signal is split into an infinitely anisotropic component for each fiber orientation, and a single isotropic component.

The predicted signal for each diffusion-weighted measurement at each voxel is:

$$S_i = S_0\left(\left(1 - \sum_{j=1}^{N} f_j\right)\exp(-b_i d) + \sum_{j=1}^{N} f_j \exp(-b_i d r_i^T R_j A R_j^T r_i)\right)$$

where S_0 is the non-diffusion-weighted signal value, d is the diffusivity, b_i is the b-value, and r_i is gradient direction associated with the ith acquisition, and f_j and $R_j A R_j^T$ are the fraction of signal contributed by, and anisotropic diffusion tensor along, the jth fiber orientation (θ_j, φ_j), and N is the maximum number of fibers. A is fixed as:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and R_j rotates A to (θ_j, φ_j). The noise is modelled separately for each voxel as independently identically distributed Gaussian with a mean of zero and standard deviation across acquisitions of σ.

Metropolis Hastings Markov Chain Monte Carlo sampling is applied to estimate the posterior distribution of fiber orientation in each voxel using FDT toolbox in FSL. The prior distributions we assigned to model parameters are:

$$P(S_0) \sim U(0, +\infty)$$

$$P(f_1) \sim U(0,1)$$

$$P(f_2^N) \sim ARD$$

$$P(\theta_1^N) \sim \sin(\theta)$$

$$P(\phi_1^N) \sim U(0,2\pi)$$

$$P(\sigma) = \sigma^{-1}$$

Where, ARD is automatic relevance determination, which is a model selection method used in the field of Neural Networks (MacKay,1995).

After estimated the posterior distribution of the fiber orientation of each voxel, we simulate the fiber direction in each voxel based on this posterior distribution and obtain a simulated fiber tracks in this voxel along this direction to next voxel in order to simulate a fiber streamline in human brain until the simulated fiber was stopped. Then totally N simulated fiber tracks are generated for each voxel. Here, for stopping fiber trajectories, we select a curvature threshold of 80° for the orientation change of fiber at each step (Behrens et al., 2013b).

Since the fiber tracks are simulated, the number of fibers from region A to region B is not same as the number of fibers from region B to region A. As a result, we use the 90th percentile of the voxel-level fiber counts connecting voxels in the seed region to voxels in the target region to reflect the strongest structure connectivity between pairs of regions and using the maximum of the two directional fiber counts for each region pair as the number of fibers counts between two regions (Xue. et al, 2015). The structure connectivity between two regions can be evaluate by the number of fiber tracks between two regions divided by the maximum fibers counts could between the two regions which equals to the total N fiber tracks for each voxel.

Here, the probability of two regions connected is used represent the connectivity between the regions instead of using simulated fiber counts. We measure the probability of two regions connected as following:

$$p_{ij} = \frac{n_{ij}}{N}$$

where, N is the total number of simulated fibers for each voxel, here is 5000, n_{ij} is number of simulated fibers connected region i and j.

All the probability of two regions connected in total 90 AAL regions can construct the structural connectivity matrix for the 90 AAL regions.

2.3 Edge-wise Linear Regression

After pre-processed DTI and resting-state fMRI data, we could generate two matrices to represent the structural connectivity (SC) and functional connectivity (FC) between our regions of interests (ROIs). Our goal is to detect is there any difference between disease group and no disease group on structural connectivity and functional connectivity. Edge-wised linear regression can help us to know the difference between two group of subjects in specific paired of regions. Based on the edge-wised

analysis of the SC and FC, we could learn where are the differences in two group subjects' brain networks.

2.3.1 Edge-wise Linear Regression for SC and FC

In our structural connectivity, we have total 90 regions in the brain network. As a result, there are totally 4005 undirected and unrepeated connections (edges) among there regions. For detecting the potential differences in the 4005 edges between two group subjects, we fit linear models for each edge j:

$$SC_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \beta_{2j}Age_{ij} + \beta_{3j}Sex_{ij} + \varepsilon_{ij}$$

where, SC_{ij} is the structural connectivity of patient i for j_{th} edge and

$$X_{ij} = \begin{cases} 0, & subject\ i\ in\ no\ diease\ group \\ 1, & subject\ i\ in\ diease\ group \end{cases}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Then for each edge j, we can generate a regression p-value for X_{ij}, which is adjusted by covariates: age and sex. Under specific significant level α, we can detect which edges have significant different between the two groups based on the regression p-value of X_{ij}.

Same as structural connectivity, we fit edge-wise linear regression models for functional connectivity to detect the difference between two groups.

$$FC_{ij} = \alpha_{0j} + \alpha_{1j}X_{ij} + \alpha_{2j}Age_{ij} + \alpha_{3j}Sex_{ij} + \varepsilon_{ij}$$

where, FC_{ij} is the functional connectivity of patient i for j_{th} edge and

$$X_{ij} = \begin{cases} 0, & patient\ i\ in\ no\ diease\ group \\ 1, & patient\ i\ in\ diease\ group \end{cases}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Based on this model, we can generate the regression p-value for X_{ij}. Under the specific significant level α, we can know which edges of ROIs are significant different between the two group subjects.

The heat map of the regression p-value would be plot to show the results.

2.3.2 Edge-wise relationship between SC and FC

In previous section, the group difference of SC and FC are detected by edge-wised linear regression. However, these models only detect the SC and FC independently on group level. In this section, we use an edge-wised linear model to detect the relationship between SC and FC on the group level.

In Rudie's (2013) paper, he calculated the correlation between fiber counts and functional connectivity strength across the edges. In my paper, the relationship between fiber counts and functional connectivity strength will be modeled across the subjects. Model would be fit for two group separately and modelling result will be compared between the two groups.

For each edge j, we fitted model for no disease group:

$$FC_{ij1} = \gamma_{0j1} + \gamma_{1j1} SC_{ij1} + \varepsilon_{ij1}$$

for disease group:

$$FC_{ij2} = \gamma_{0j2} + \gamma_{1j2} SC_{ij2} + \varepsilon_{ij2}$$

where,

$$\varepsilon_{ij1} \sim N(0, \sigma^2) \; and \; \varepsilon_{ij2} \sim N(0, \sigma^2)$$

Here, we use structural connectivity as the explanatory variable because SC represent the real world fibers in people's brain. How the fiber shaped in people's brain would decide the potential functional connectivity on the brain.

Using these two model, we could compare the correlation coefficient of between FC and SC between two groups under specific significant level α. The heat map of the correlation coefficient would be plot to show the result for all the 4005 edges.

2.4 Global network metric-based method (GNM)

The probability of brain regions connecting with each other and the functional connectivity among different regions can be explained like a network. As a result, using global network metrics to represent the characteristics of the brain network can help us to know more about the brain. Moreover, this method provides a new angle for us to compare the brain's characteristics between disease group and no disease group. In brain networks, the nodes usually represent the brain regions, the links between nodes represent the connections between brain regions.

2.4.1 Graph theoretical metrics

In network analysis, there are many theoretical metrics can be used to explain the brain networks' properties. These metrics can show us many information about the brain network, such as important nodes, the strength of network connections. Here, five graph theoretical metrics (Rubinov and Sporns, 2009) would be calculated for structural connectivity matrix to measure the network properties. These metrics are: Clustering Coefficient(CC), Characteristic Path Length (CPL), Normalized CC and CPL, and Small-worldness. The network of structural connectivity is weighted undirected networks, which is weighted by the probability of two regions connected with fibers.

Clustering Coefficient measures the clustered connectivity ability around individual nodes in the network. The higher CC of the network has, the higher local efficiency of the network has.

For one subject's brain network,

$$C = \frac{1}{n}\sum_{i \in N} C_i = \frac{1}{n}\sum_{i \in N} \frac{2t_i^w}{k_i(k_i - 1)}$$

where, N is the set of all nodes in the network, w is the weight, n is the number of nodes in the network, k_i is the number of links connected to node i, t_i is the number of triangles around node i.

Characteristic Path Length measures the shortest path length between all pairs of nodes in the network. The smaller the CPL of node i has, the the higher global efficiency of node i has.

For one subject's brain network,

$$L = \frac{1}{n}\sum_{i \in N} L_i = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^w}{n - 1}$$

where, N is the set of all nodes in the network, w is the weight, n is the number of nodes in the network, d_{ij}^w is the weighted shortest path between node $i \; and \; j$.

Normalized CC of subject's network is the ratio of C to the average C from simulated randomized networks. Normalized CPL of subject's network is the ratio of L to the average L from simulated randomized networks. The Brain Connectivity Toolbox in Matlab will be used to simulate one hundred random networks based on the true network for each subjects. The mean C and L of the random networks will be used to as the average C and average L.

Small-worldness of one subject's network is the ratio of normalized CC to normalized CPL. Small-worldness of a network can measure how the network is significantly more clustered than random network.

For subject i's network,

$$S_i = \frac{C_i/C_{rand}}{L_i/L_{rand}}$$

2.4.2 Linear Regression

The goal for calculating these graph theoretical is to detect is there any difference in brain networks properties between disease group and no disease group. Just like the edge-wised linear regression for SC and FC matrix, the graph theoretical metrics can be compared between the two groups by using linear regression.

For two groups subjects, we can fit five linear regression models for CC, CPL, normalized CC, normalized CPL and Small-worldness across the subjects.

$$CC_k = \eta_{0cc} + \eta_{1cc}X_k + \eta_{2cc}Age_k + \eta_{3cc}Sex_k + \varepsilon_{kcc}$$

$$CPL_k = \eta_{0cpl} + \eta_{1cpl}X_k + \eta_{2cpl}Age_k + \eta_{3cpl}Sex_k + \varepsilon_{kcpl}$$

$$Normalized\ CC_k = \eta_{0ncc} + \eta_{1ncc}X_k + \eta_{2ncc}Age_k + \eta_{3ncc}Sex_k + \varepsilon_{kncc}$$

$$Normalized\ CPL_k = \eta_{0ncpl} + \eta_{1ncpl}X_k + \eta_{2ncpl}Age_k + \eta_{3ncpl}Sex_k + \varepsilon_{kncpl}$$

$$Small-worldness_k = \eta_{0sm} + \eta_{1sm}X_k + \eta_{2sm}Age_k + \eta_{3sm}Sex_k + \varepsilon_{ksm}$$

where k is k_{th} subject and

$$X_k = \begin{cases} 0, & subject\ k\ in\ no\ diease\ group \\ 1, & subject\ k\ in\ diease\ group \end{cases}$$

$$\varepsilon_{kcc} \sim N(0,\sigma^2), \varepsilon_{kcpl} \sim N(0,\sigma^2), \varepsilon_{kncc} \sim N(0,\sigma^2), \varepsilon_{kncpl} \sim N(0,\sigma^2), \varepsilon_{knsm} \sim N(0,\sigma^2)$$

According to the regression results of these five models, the potential group difference of network properties can be detected by using the regression p-value of the coefficients of X_k.

2.5 Differentially expressed network(DEN)

In section 2.2, edge-wise linear regression was used to find the differences between two group subjects' SC and FC. For further interests, we use one approach of

Differentially expressed network methods devised by Shuo Chen(2015) to detect which edges are more different in group comparison.

2.5.1 Parsimonious differential brain connectivity network detection Algorithm(Pard)

This paper's goal is to detect is there any difference in brain connectivity between two groups of subjects in PNC dataset. Pard Algorithm can help us to achieve the goal based on the hypothesis:

H_0: Two groups have no difference in connectivity

H_1: There are differentially expressed connectivity networks between the two groups

Pard Algorithm (Shuo Chen et al, 2015) could use two sample T-test p-value for each edges to investigate the significantly differential connectivity expressions (edges) between two groups of subjects with well-controlled false-positive discovery rates. The p-value for each edge will be transformed by using "-log" transformation to show how important the edge is for detecting the difference between two groups of subjects. They investigate the number of disconnected subgraph in the overall graph G with weight matrix W, which comes from the transformed p-value matrix after screening step. For each subgraph, the K-means clustering method would be used to detect the number of clusters in the subgraph.

After using Pard Algorithm, the edges have similar significant effects to detect the difference between two groups would be clustered into same cluster. This method would help us to find the important edges to differentiate two groups of subjects quickly and visually.

In this paper, the regression p-value will instead of T-test p-value as input into the Pard Algorithm. And we will not perform the permutation test and the automatic p_0 selection steps. Following is the details of Pard Algorithm:

1. Calculate the weight matrix W by screening:

$$w_{ij} = \begin{cases} -\log(p_{ij}) & if\ p_{ij} < p_0 \\ 0 & else \end{cases}$$

where p_{ij} is the regression p-value, p_0 is the threshold value we will give

2. Detect disconnected subgraphs in G by eigen decompose the Laplacian matrix. Using the zero eigenvalues and its eigenvectors to decide the number of disconnected subgraphs and their nodes elements in each subgraph.
3. For each subgraph G_q, find the networks which include most significant edges with constrained numbers of nodes for each networks.
4. Try all possible K_q clusters for each subgraph G_q, and then select the optimum number of networks based on Shuo's clustering criteria.

3. Results

3.1 Edge-wise linear regression for SC and FC

After performing probabilistic tracktography, structural connectivity matrix can be constructed by the probabilities two regions connected with each other as the elements in the matrix. Since there are total 90 regions, the structural connectivity matrix is 90 by 90 dimensions.

Based on the mean structural connectivity matrixes (Figure.1), there was tiny different between the groups of subjects under the scale [0,1] when the matrixes were observed by eyes. Moreover, the two groups of subjects have very similar fiber tracks structure.

For detecting the accurate difference of structural connectivity between these two groups of subjects, edge-wise linear regression was performed for these two SC matrixes. There were 105 edges which are significantly different between disease group and no disease group under significant level at 0.05 (Figure.2). This revealed that the two groups of subjects have significantly different fiber structure in this 105 connections between the ROIs after adjusting by age and sex. When under significant level at 0.1, there were 364 edges which are significantly different between these two groups of subjects when adjusting by age and sex.

Partial correlations were calculated using Wang's method (2016) to represent the functional connectivity of two groups of subjects. The mean functional connectivity matrixes were generated from two groups of subjects (Figure.3). Being similar to Structural connectivity matrixes, the two mean functional connectivity matrixes were hard to recognize the difference. Edge-wise linear regression for contracting functional connectivity of two groups of subjects were applied to recognize the

difference. Same as edge-wise linear regression, age and sex were used as adjusting effect to detect the difference between groups.

Under the 5% significant level, 246 edges between paired of regions were significantly different between the two groups of subjects (Figure.4). This revealed that there are 246 different functional connections between paired regions when contracting disease group with no disease group. After increasing the significant level to 10%, 485 edges between paired of regions were significantly different between the disease group and no disease group.

3.2 Relationship between SC and FC

Independent analysis for structural connectivity and functional connectivity revealed the difference magnitude between subjects in disease group and no disease group separately. Edge-wise linear regression between structural connectivity and functional connectivity across the subjects in two groups gave the information from different angle.

According to the regression results (Figure.5), no disease group has almost same positive and negative relationship between SC and FC with disease group. The were 152 edges have significant correlations between SC and FC in no disease group and 167 edges have significant correlations between SC and FC in disease group under 5% significant level. In disease group, 76 of 167 significant correlations were positive and 91 of them were negative. In no disease group, 77 of 152 significant correlations were positive and 77 of them were negative relationships. Comparing all the relationship in details, there were two paired of regions which have negative relationship in disease group but have positive relationship in no disease group. At the same time, there were also two paired of regions which have positive relationship in

disease group but have negative relationship in no disease group. Moreover, there were 181 correlations in disease group are not significant but 77 of them were negative correlation and 74 of them were positive correlation in no disease group. Meanwhile, there were 160 correlations in no disease group are not significant but 73 of them were positive correlation and 87 of them are negative correlation in disease group (Figure.6).

3.3 Graph Theoretical metrics result

After performing linear regression for five graph theoretical metrics across the subjects in two group, the different network properties can be noticed from the results. The results were showed in Table.2. Although there was no significant different in Clustering Coefficient and Small-worldness, Characteristic Path Length was significantly different between two groups of subjects. Subjects in disease group had 0.084 smaller than subjects in no disease group for Characteristic Path Length. This revealed brain region in disease group subjects have higher probability to connect with each other averagely than no disease group.

Table 2. Regression results for Graph Theoretical Metrics

Properties	Coefficients for group indicator	P-value
Clustering Coefficient (CC)	0.0002	0.55
Characteristic Path Length (CPL)	-0.084	0.04
Normalized CC	0.02	0.51
Normalized CPL	-0.02	0.04
Small-worldness	0.039	0.15

3.4 Differentially expressed network result

As we mentioned in section 2.5, this DEN method could help us to distinguish the most different edges in two groups of subjects' brain connectivity. We used the

regression p-value we obtained from edge-wise regression step as the input for parsimonious age and sex. Based on the threshold p_0 at 0.08, we got 27 clusters for structural connectivity regression p-value matrix after excluding 2 isolated regions. 23 clusters were detected for functional connectivity regression p-value matrix. The results were summarized in Figure.7, and all significant edges tended to be along the diagonal because of shrinkage effect.

Figure 8 and Figure 9 showed the differentially expressed edges for the first four clusters for structural connectivity between two groups of subjects. Figure 10 and Figure 11 showed the differentially expressed edges for the first four clusters for functional connectivity between two groups of subjects. For structural connectivity, the differences were mainly showed on SAM.L, PHG.L, CAL.R, IOG.R, ACR.L, SOG.L, ORBmid.L and ITG.L. For functional connectivity, the differences were mainly showed on MFG.L, PCL.L, MTG.L, IFGtriang.R, IFGoperc.L, IFGoerc.R and SFCmed.R.

4. Discussion

For detecting the difference between disease and no disease group subjects in their brain, we compared the two groups of subjects' brain connectivity. We used the probability of paired of regions connected as the structural connectivity, and partial correlation of fMRI between paired of regions as the functional connectivity. Edge-wise linear regression of structural connectivity reveals that more than hundred edges between paired regions are significantly different between two groups. Edge wise linear regression of functional connectivity also shows that more than two hundred edges between paired regions are significantly different between the two groups. The edge-wise linear regression method can provide a regression p-value which were adjusted by other covariates like age and sex. These p-values can contain more information than the two-simple t-test. For further interest, edge-wise regressions between functional connectivity and structural connectivity were performed. We used structural connectivity to explain functional connectivity since the anatomical structure decide the functional connectivity to some degree. Unlike Rudi et al (2013), we performed linear regression between SC and FC for each edge across the subjects in each group. The results can show us the details about how SC and FC correlated and what edges are significant different between the two groups of subjects.
Structural connectivity and functional connectivity can be treated as networks. Graph theoretical metrics helped us to explore the network properties for two groups of subjects' brain connectivity. Clustering coefficient, characteristic path length and small-worldness were calculated and averaged across all the voxels in each subject's structural connectivity and functional connectivity networks. Linear regressions of graph theoretical metrics were conducted following to detect the network properties

difference between the two groups of subjects. We detected that disease group has a significant difference with no disease group in characteristic path length property. In this paper, we only measured five Graph theoretical metrics. There are some other metrics can be used, such as participation coefficients and modularity Q value, to explain the network properties.

After performing edge-wise linear regressions of structural and functional connectivity, parsimonious differential brain connectivity network detection algorithm helped us to find out which edges in the brain network are most significant different between the two groups of people. We clustered the regions in the structural and functional connectivity networks and reordered them towards to diagonal of the matrix. This method can support us to view the difference between two groups of subjects directly. Obviously, there are some flaws in this paper. The automatically threshold p-value choosing was not applied in this paper. Also, we did not assign a permutation test to control the family-error rate.

The work performed linear regression and graph theory knowledge to detect the difference between disease group and no disease group's brain connectivity. All the methods can effectively identify their difference. However, if we could have a strong explanation in medicine area, it would be better to understand the results. Moreover, we divided all the subjects into two groups based on the simple criteria that subjects have central nervous system problem or not. In fact, PNC study combined different type of subjects which have various mental health problems. Using precise mental health disease group contrasting would have more clearly and significantly different in specific brain regions. And it would be more meaningful to explain the results. Moreover, the edge-wise linear regression may not exactly represent the relationship between structural connectivity and functional connectivity.

Reference

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. Magnetic resonance in medicine,44(4), 625-632.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., ... & Smith, S. M. (2003a). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magnetic resonance in medicine, 50(5), 1077-1088.

Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., ... & Thompson, A. J. (2003b). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature neuroscience, 6(7), 750-757.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?. Neuroimage, 34(1), 144-155.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic resonance in medicine, 34(4), 537-541.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience,10(3), 186-198.

Chen, S., Kang, J., Xing, Y., & Wang, G. (2015). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. Human brain mapping, 36(12), 5196-5206.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. Nature Reviews Neuroscience, 12(1), 43-56.

Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., ... & Barnes, K. A. (2010). Prediction of individual brain maturity using fMRI. Science, 329(5997), 1358-1361.

Fingelkurts, A. A., Fingelkurts, A. A., & Kähkönen, S. (2005). Functional connectivity in the brain—is it an elusive concept?. Neuroscience & Biobehavioral Reviews, 28(8), 827-836.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. Cerebral cortex, 19(1), 72-78.

MacKay, D. J. (1995). Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks.Network: Computation in Neural Systems, 6(3), 469-505.

Marrelec, G., Krainik, A., Duffau, H., Pelegrini-Issac, M., Lehericy, S., Doyon, J., and Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. Neuroimage 32, 228-237.

Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging.Annals of neurology, 45(2), 265-269.

Peng, J., Wang, P., Zhou, N., & Zhu, J. (2012). Partial correlation estimation by joint sparse regression models. Journal of the American Statistical Association.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059-1069.

Rudie, J. D., Brown, J. A., Beck-Pancer, D., Hernandez, L. M., Dennis, E. L., Thompson, P. M., ... & Dapretto, M. (2013). Altered functional and structural brain network organization in autism. NeuroImage: clinical, 2, 79-94.

Rykhlevskaia, E., Gratton, G., & Fabiani, M. (2008). Combining structural and functional neuroimaging data for studying brain connectivity: a review.Psychophysiology, 45(2), 173-187.

Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M.E., Hopson, R., Jackson, C., Keefe, J., and Riley, M. (2014). Neuroimaging of the Philadelphia neurodevelopmental cohort. Neuroimage 86, 544-553.

Satterthwaite, T.D., Wolf, D.H., Roalf, D.R., Ruparel, K., Erus, G., Vandekar, S., Gennatas, E.D., Elliott, M.A., Smith, A., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R.E., and Gur, R.C. (2015). Linked Sex Differences in Cognition and Functional Connectivity in Youth. Cereb Cortex 25, 2383-2394.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., and Woolrich, M.W. (2011). Network modelling methods for FMRI. Neuroimage 54, 875-891.

Smith, S.M. (2012). The future of FMRI connectivity. Neuroimage 62, 1257-1266.

Sporns, O. (2011). The human connectome: a complex network. Annals of the New York Academy of Sciences, 1224(1), 109-125.

Sporns, O. (2012). From simple graphs to the connectome: networks in neuroimaging. Neuroimage, 62(2), 881-886.

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. European Neuropsychopharmacology, 20(8), 519-534.

Wang Y, Kang J, Kemmer PB and Guo Y (2016) An Efficient and Reliable Statistical Method for Estimating Functional Connectivity in Large Scale Brain Networks Using Partial Correlation. *Front. Neurosci*. 10:123. doi: 10.3389/fnins.2016.00123

Xue, W., Bowman, F. D., Pileggi, A. V., & Mayer, A. R. (2015). A multimodal approach for determining brain networks by jointly modeling functional and structural connectivity. Frontiers in computational neuroscience, 9.

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. Neuroimage, 53(4), 1197-1207.

Zalesky, A., Cocchi, L., Fornito, A., Murray, M. M., & Bullmore, E. (2012b). Connectivity differences in brain networks. Neuroimage, 60(2), 1055-1062.

5. Appendix

A. Figure

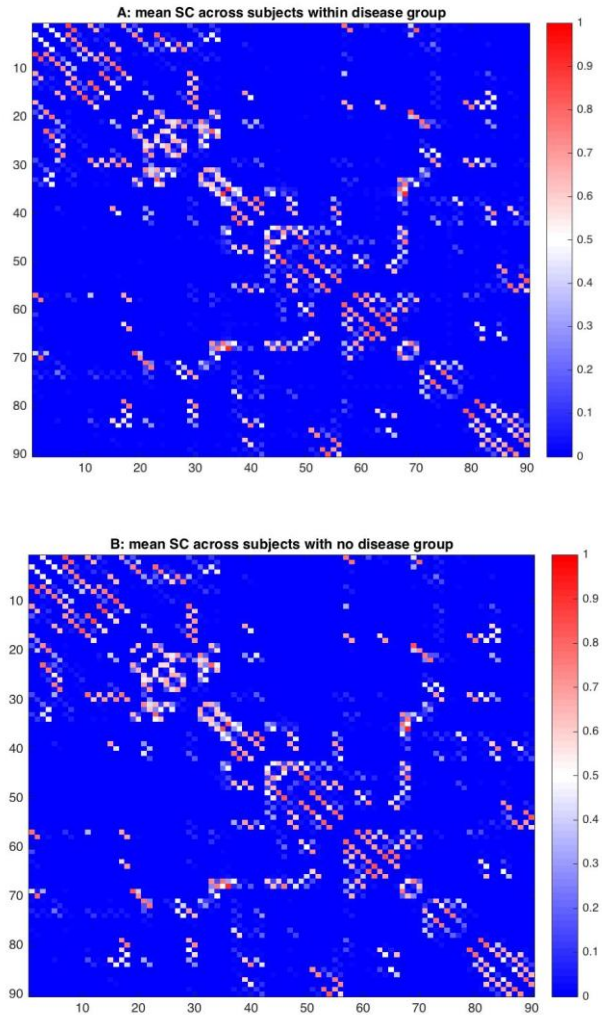


Figure 1. Heat Map of Mean Structural Connectivity

A: Heat map of mean structural connectivity matrix across the subjects in disease group; B: Heat map of mean structural connectivity matrix across the subjects in no disease group.

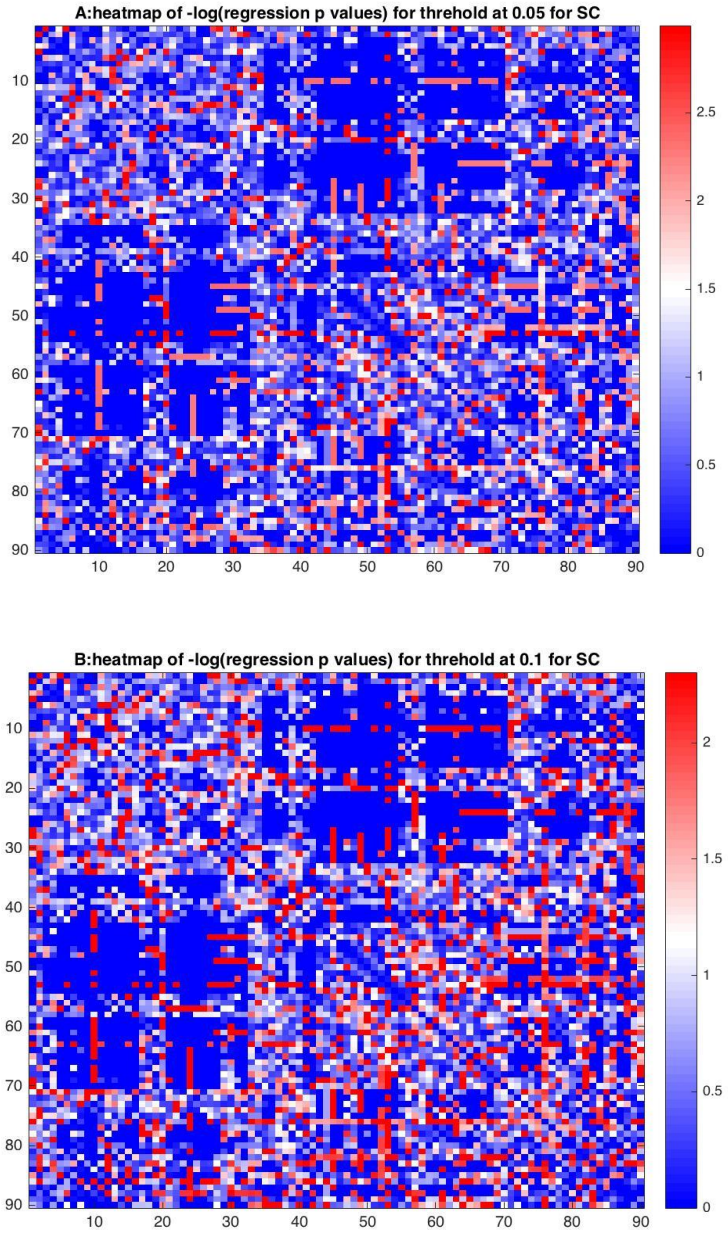


Figure 2. Heat map of –log (edge-wise regression p values) for structural connectivity A: Heat map of –log (edge-wise regression p values) for structural connectivity at 0.05 significant level; B: Heat map of –log (edge-wise regression p values) for structural connectivity at 0.1 significant level;

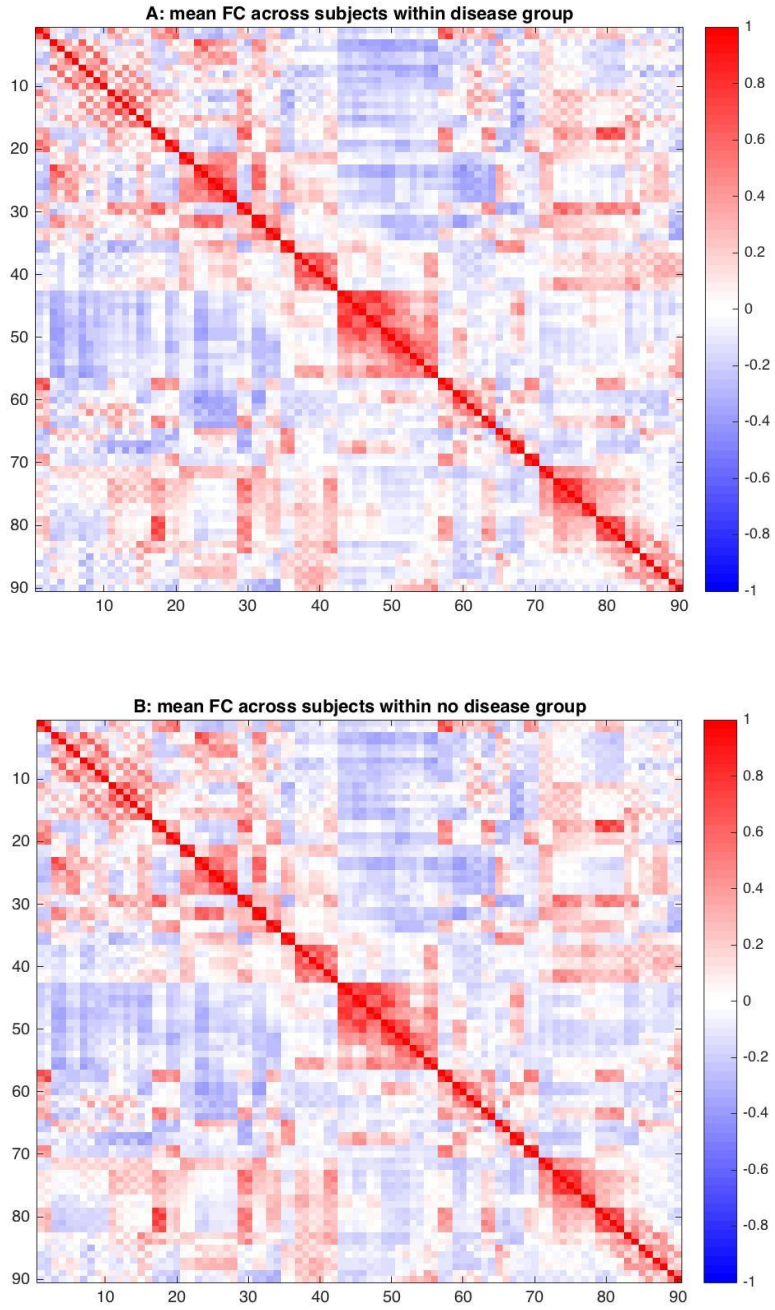


Figure 3. Heat Map of Mean Functional Connectivity

A: Heat map of mean functional connectivity matrix across the subjects in disease group; B: Heat map of mean functional connectivity matrix across the subjects in no disease group.

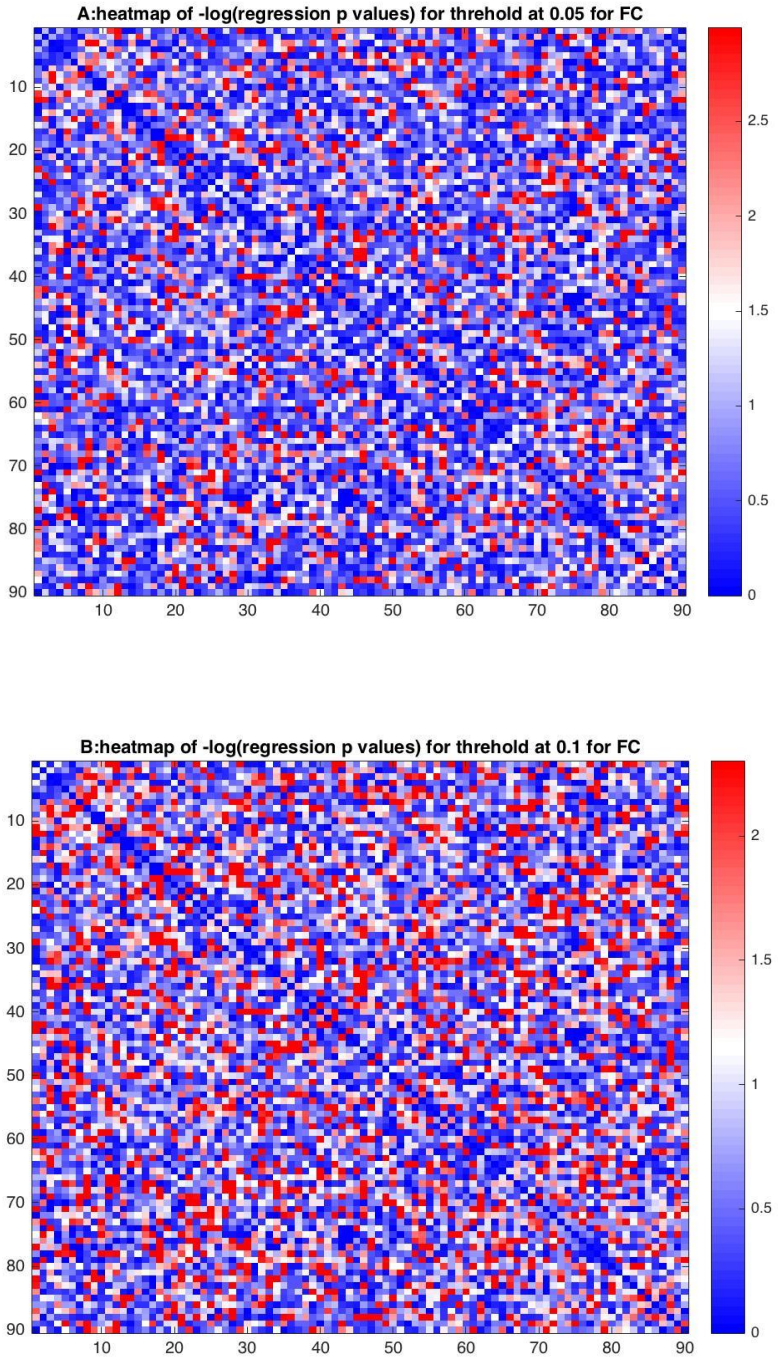


Figure 4. Heat map of –log (edge-wise regression p values) for functional connectivity

A: Heat map of –log (edge-wise regression p values) for functional connectivity at 0.05 significant level; B: Heat map of –log (edge-wise regression p values) for functional connectivity at 0.1 significant level.

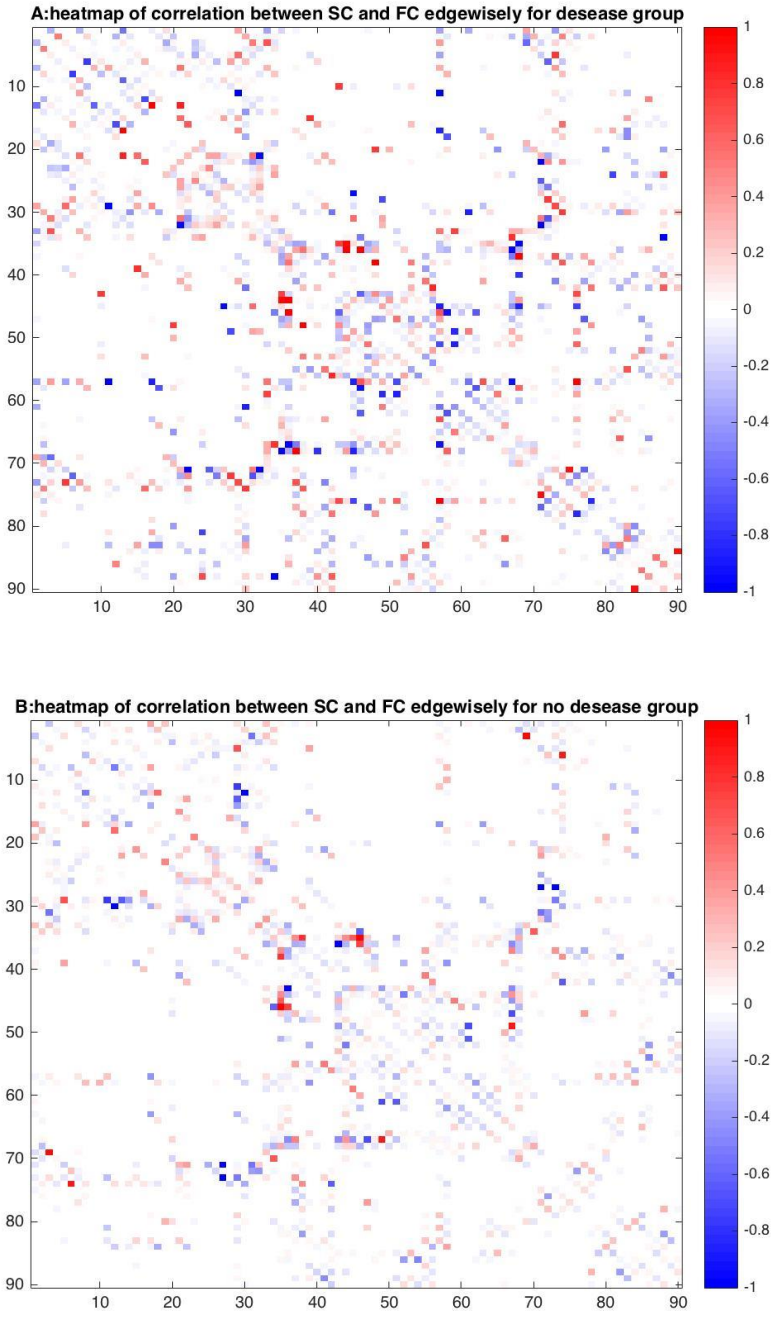


Figure 5. Heat map of edge-wise regression coefficients between SC and FC
A: Heat map of edge-wise regression coefficients between SC and FC for disease group; B: Heat map of edge-wise regression coefficients between SC and FC for no disease group.

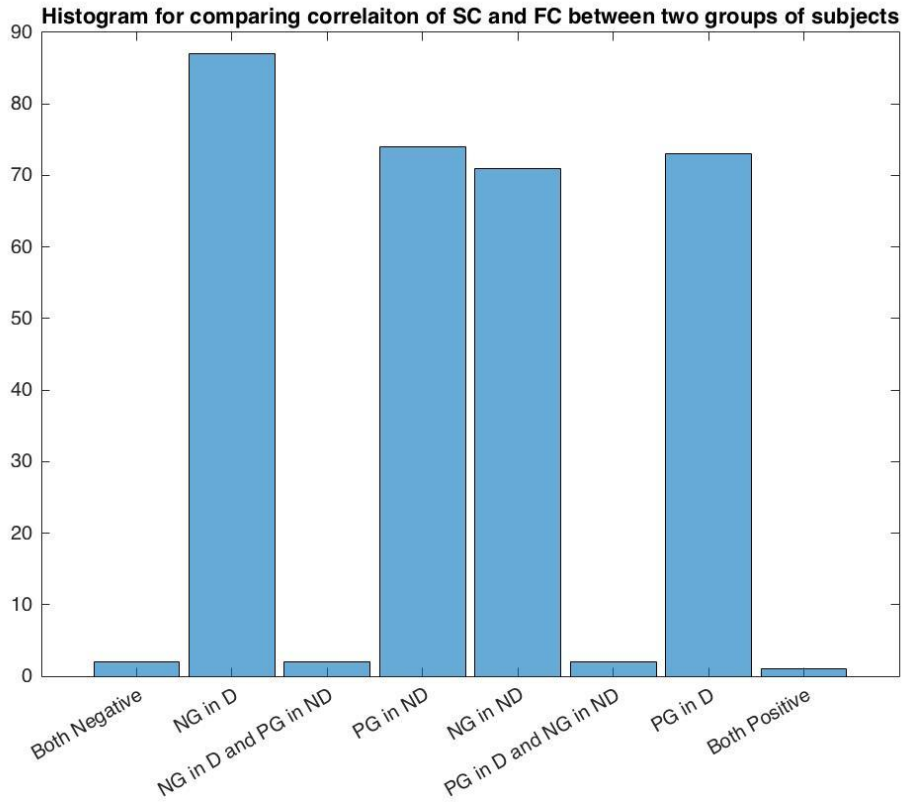


Figure 6. Histogram for comparing significant correlations of SC and FC between two groups of subjects.
NG means negative correlations, PG means positive correlations, D means disease group, ND means no disease group.

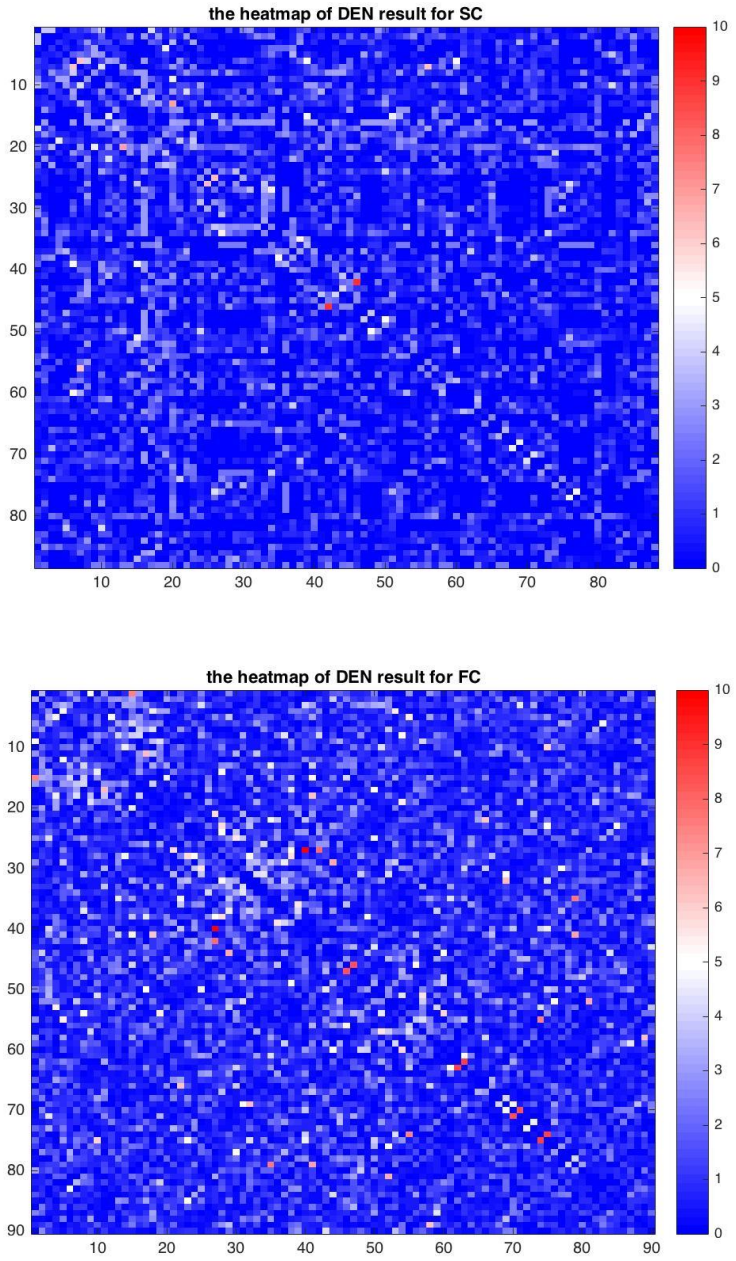
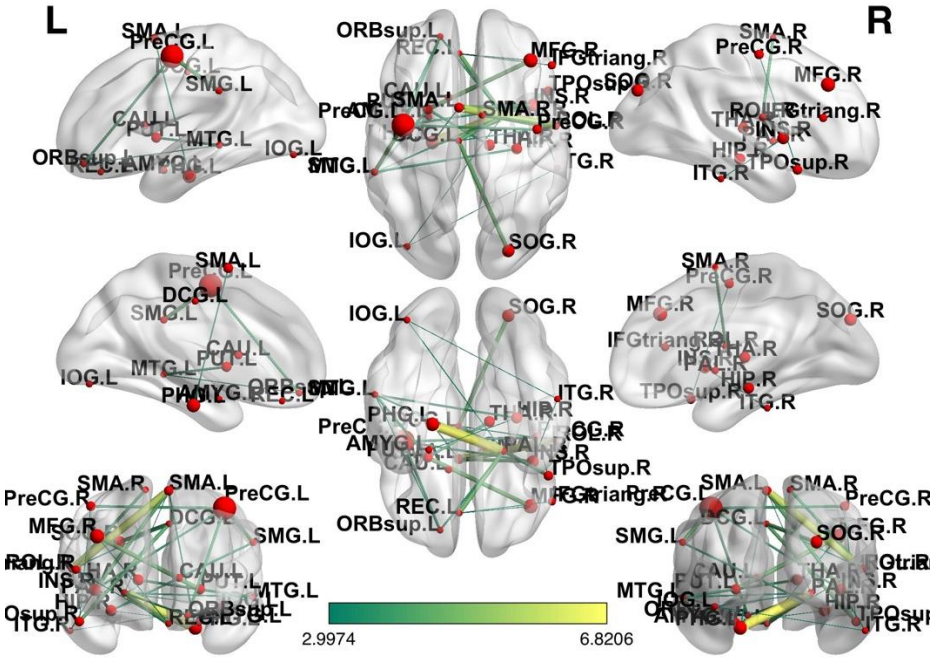


Figure 7. Heat map of differentially expressed network result
Upper Figure: Heat map of differentially expressed network result for structural connectivity matrix based on '-log' transformation; Lower Figure: Heat map of differentially expressed network result for functional connectivity matrix based on '-log' transformation.

Cluster 1



Cluster 2

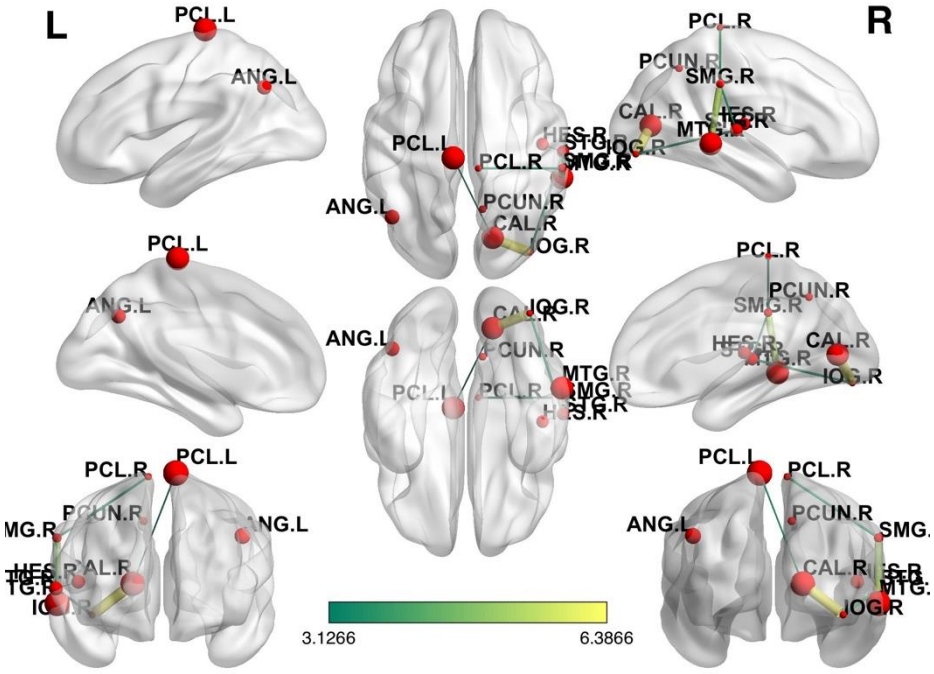
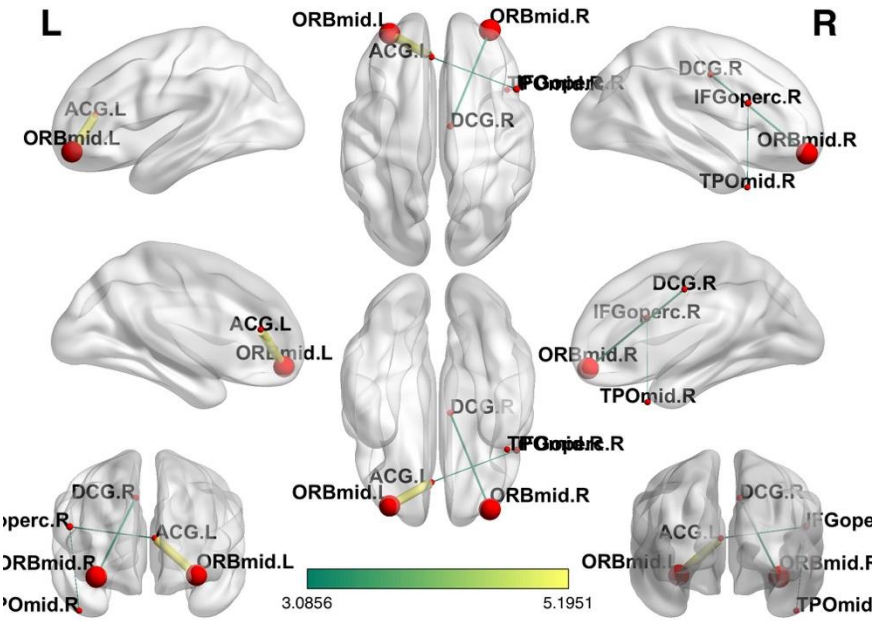


Figure 8. DEN result 1 for Structural Connectivity
The cluster.1 and cluster.2 of the regions which have significant difference expression in structural connectivity matrix based on '-log' transformation.

Cluster 3



Cluster 4

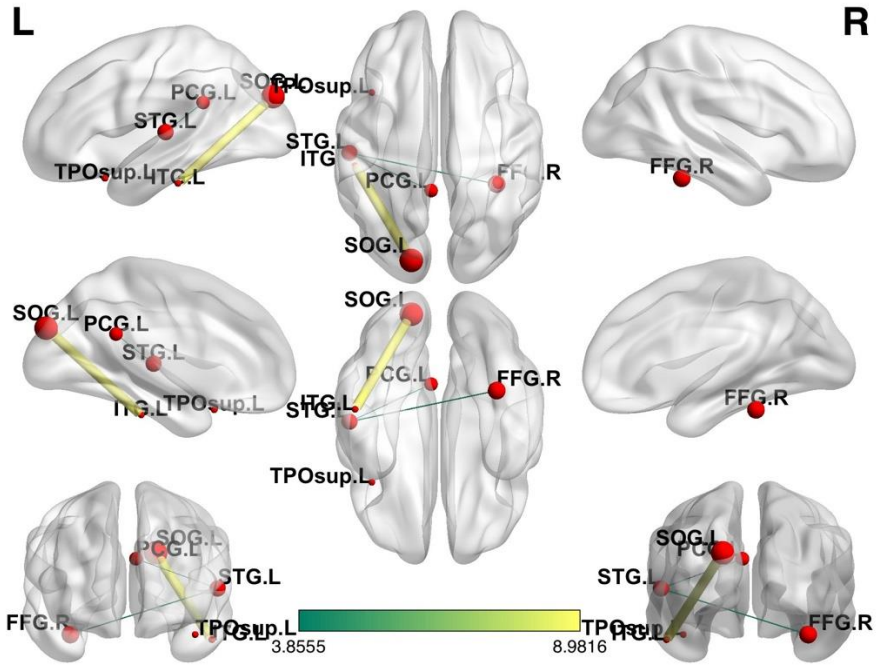
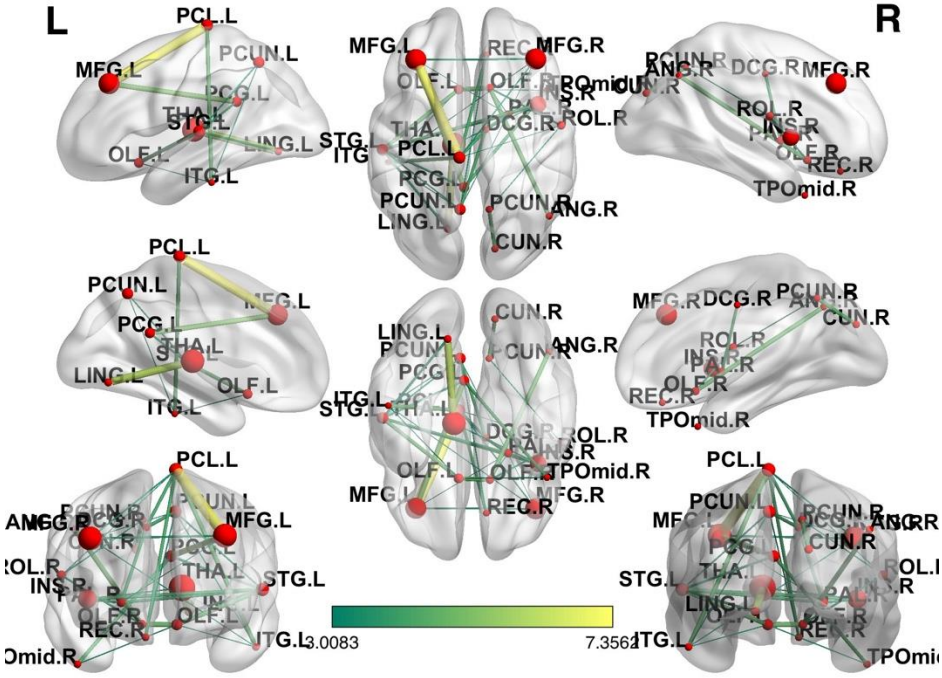


Figure 9. DEN result 2 for Structural Connectivity

The cluster.3 and cluster.4 of the regions which have significant difference expression in structural connectivity matrix based on ‘-log’ transformation.

Cluster 1



Cluster 2

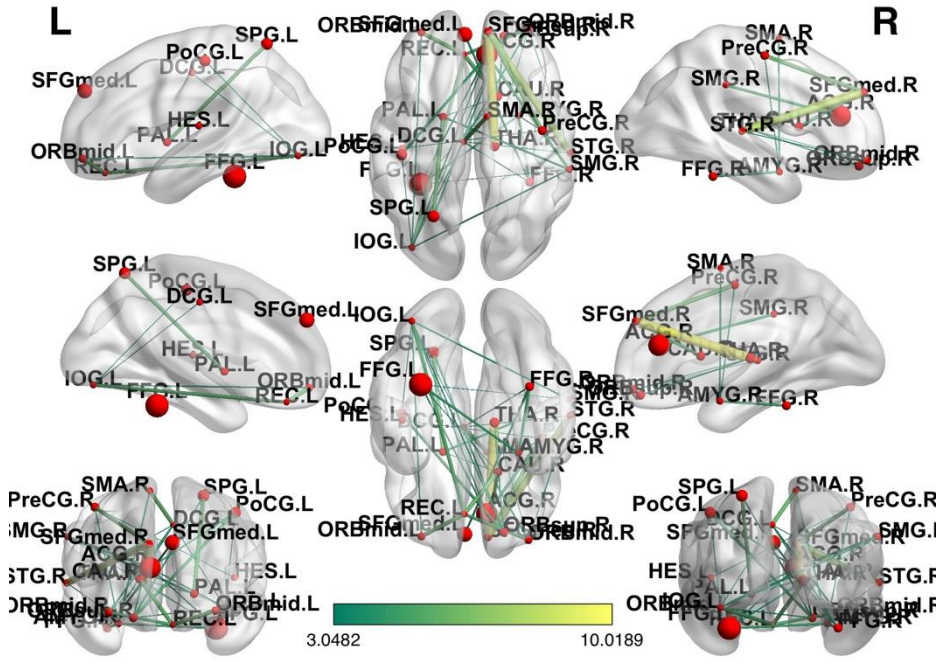
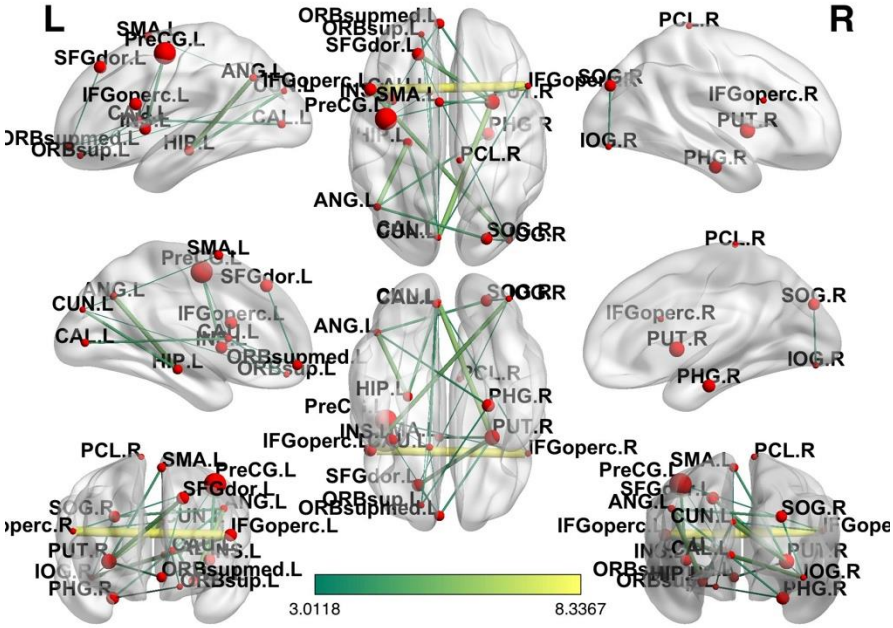


Figure 10. DEN result 1 for Functional Connectivity

The cluster.1 and cluster.2 of the regions which have significant difference expression in functional connectivity matrix based on '-log' transformation.

Cluster 3



Cluster 4

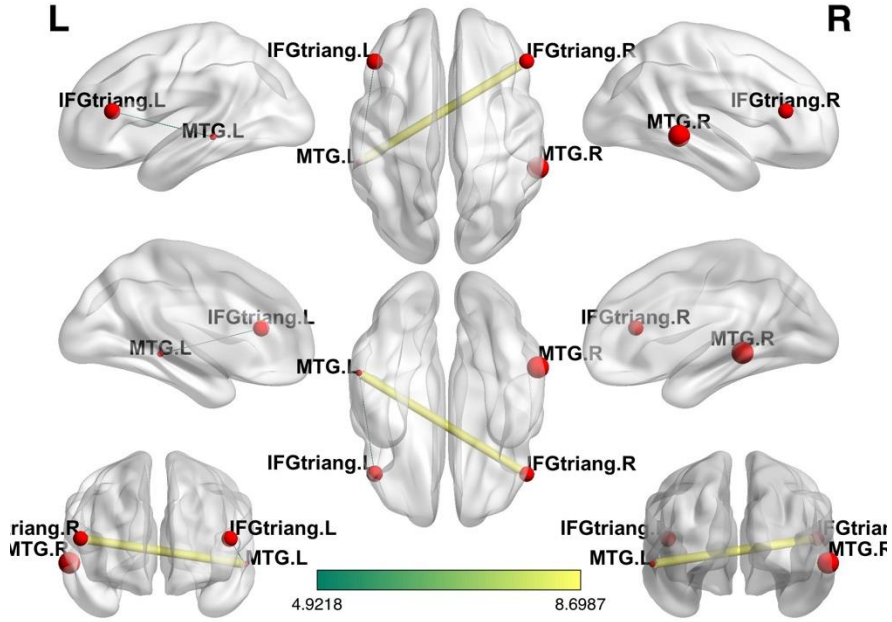


Figure 11. DEN result 2 for Functional Connectivity

The cluster.3 and cluster.4 of the regions which have significant difference expression in functional connectivity matrix based on '-log' transformation.