A Mass Multivariate Edge-wise Approach for Combining Multiple Connectomes to Improve the Detection of Group Differences

Javid Dadashkarimi[1], Siyuan Gao[2], Erin Yeagle[3], Stephanie Noble[3], and Dustin Scheinost[3(✉)]

[1] Department of Computer Science, Yale University, New Haven, CT, USA
javid.dadashkarimi@yale.edu
[2] Department of Biomedical Engineering, Yale University, New Haven, CT, USA
[3] Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

Abstract. Functional connectivity derived from functional magnetic resonance imaging data has been extensively used to characterize individual and group differences. While these connectomes have traditionally been constructed using resting-state data, recent work has highlighted the importance of combining multiple task connectomes, particularly for identifying individual differences. Yet, these methods have not yet been extended to investigate differences at the group level. Here, we propose a mass multivariate edge-wise approach to improve the detection of group differences by combining connectomes from multiple sources. For each edge, the magnitude of connection strength from each of multiple connectomes are included in statistical hypothesis testing. We evaluate the proposed approach by estimating sex differences in two large, publicly available datasets: the Human Connectome Project and Philadelphia Neurodevelopmental Cohort. Results indicate the proposed mass multivariate edge-wise analysis offers improved detection of group differences compared to univariate analysis, and support the utility of combining multiple connectomes to improve detection of group differences.

1 Introduction

Functional connectomics derived from functional magnetic resonance imaging (fMRI) is a powerful framework to elucidate individual and group differences in brain organization [1]. While connectomes are traditionally generated from resting-state data [2], recent work has shown that connectomes generated from task data offer a significant improvement in detecting individual and group differences [3,4]. Furthermore, combining multiple task connectomes per subject further increases the amount of information useful for detecting these differences [5]. However, these methods have only been used in the context of characterizing individual differences, not group differences. Thus, there remains a need to

M. D. Schirmer et al. (Eds.): CNI 2019, LNCS 11848, pp. 64–73, 2019.
https://doi.org/10.1007/978-3-030-32391-2_7

develop methods that combine connectomes from multiple sources in the context of detecting group differences.

To address this need, we extended a traditional mass univariate edge-wise approach (c.f. [6]) to perform multivariate inferences that include, for each edge, the connectivity strengths of all tasks. We label this approach "mass multivariate edge-wise analysis". We compared our multivariate approach for detecting group differences to the traditional mass univariate approach using task connectomes and the general functional connectivity approach for combining multiple task connectomes into a single connectome using the Human Connectome Project (HCP) and Philadelphia Neurodevelopmental Cohort (PNC) datasets [7,8]. We hypothesize that the proposed mass multivariate approach will detect a greater number of edges exhibiting significant sex effects in comparison to the competing methods. Together, our results support the utility of combining multiple task connectomes to improve the detection of group differences.

2 Related Works

Efforts aimed at combining connectomes from multiple sources is an active field of research. Although an exhaustive review of this field is outside the scope of this paper, we briefly highlight some relevant work. Several studies have combined structural connectomes from diffusion tensor imaging (DTI) with functional connectomes [9,10]. Similarly, work has been done to combine connectomes derived from electroencephalogram (EEG) and fMRI data [11]. Yet, combining multiple connectomes from different tasks has received less attention. To our knowledge only three approaches, based on canonical correlation analysis, ridge regression, and averaging connectomes, respectively, have been proposed [3,5].

3 Methods

In this section, we derive our proposed mass multivariate edge-wise analysis. First, we briefly review multivariate hypothesis testing using the Hotelling's T^2 test in Sect. 3.1; second, we discuss univariate edge-wise analysis in Sect. 3.2; and, finally, we propose our mass multivariate edge-wise analysis in Sect. 3.3. Figure 1 shows an overall schematic of the two analyses.

3.1 Hotelling's T^2 Test

Briefly, the t-test is a statistical method to determine if the means of data from two different groups differ from each other. The Hotelling's T^2 test is a generalization of the t-test that allows for multivariate, rather than univariate, hypothesis testing. For a t-distribution, a confidence interval for a sample of size n, standard deviation of s, and significance level of α is defined as:

$$\bar{x} \pm t_{1-\alpha/2,n-1} \frac{s}{\sqrt{n}} \tag{1}$$

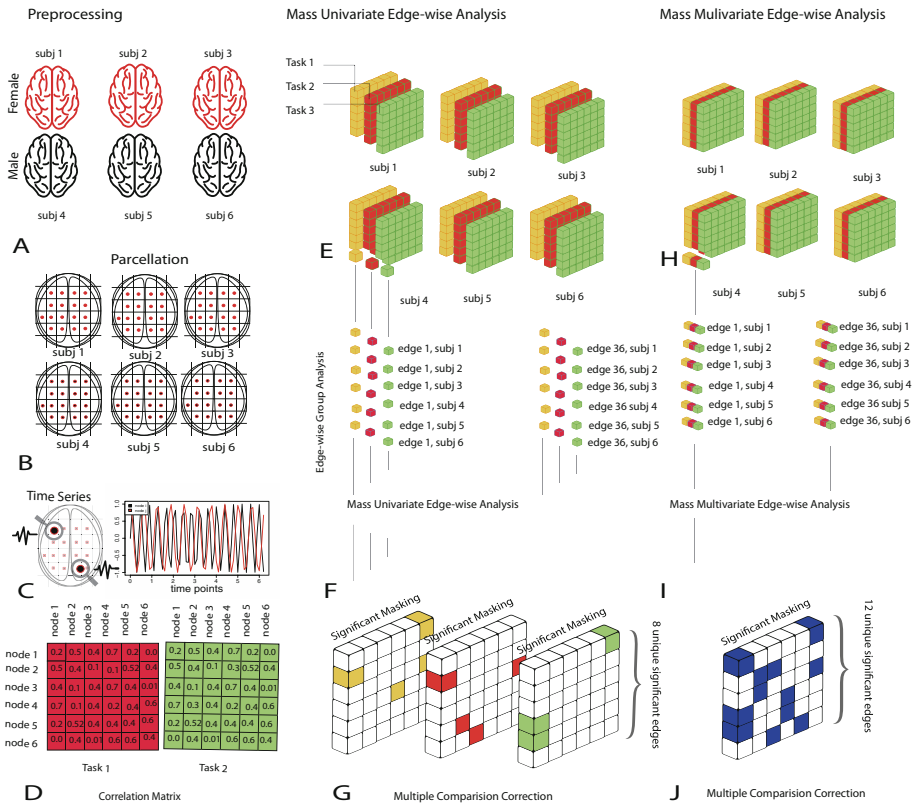


Fig. 1. Mass multivariate edge-wise analysis pipeline. *Common preprocessing steps:* **(A)** fMRI data consisting of k tasks acquired from two groups (*e.g.* males and female) of participants;**(B)** parcellate the brain into N nodes; **(C)** average timeseries for each node; **(D)** generate connectomes $\mathbf{x}_i \in \mathbb{R}^{N\times N}$ for each task and participant using the time series; *Mass Univariate Edge-wise Analysis:* **(E)** independently for each task connectome, create a vector connectivity strength across all participants for each edge; **(F)** apply a t-test on these vectors and corresponding labels $y \in \{0, 1\}$; **(G)** perform hypothesis testing and multiple comparison correction using *a priori* thresholds; *Mass Multivariate Edge-wise Analysis:* **(H)** stack all k connectomes $\mathbf{X} \in \mathbb{R}^{k\times N\times N}$, creating a matrix of connectivity strength across all participants for each; **(I)** apply a Hotelling's T^2 on $\mathbf{X}[:, i, j]$ and $y \in \{0, 1\}$; **(J)** perform hypothesis testing and multiple comparison correction using *a priori* thresholds.

where $t_{1-\alpha/2,n-1}$ is $1-\alpha/2$ fraction of t-distribution with $n-1$ degrees of freedom. In other words, with $1-2\times\frac{\alpha}{2}$ trials, the true mean spans in this interval. To test if the sample mean has a value of μ_0 based on null-hypothesis, we need to verify $T = (\bar{x}-\mu_0)/(s/\sqrt{n}) < t_{\alpha/2,n-1}$ or $T = (\bar{x}-\mu_0)/(s/\sqrt{n}) > t_{1-\alpha/2,n-1}$. We can re-write $T^2 = n\frac{(\bar{x}-\mu_0)^2}{s^2}$ and compare with squared form of t-distribution.

Given that $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, .., \bar{x}_k)$ is a vector of k normally distributed variables and $\boldsymbol{\mu}_0 = (\mu_1^0, \mu_2^0, .., \mu_k^0)$ are the corresponding means, we define:

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \tag{2}$$

where $\mathbf{S}$ is the variance-covariance matrix. Equation 2 is exactly comparable to the ratio of between group variance $n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)/(m-1)$ and within group variance $\sum_i \mathbf{S}_i/m$. For hypothesis testing, Hotelling's T^2 is first transformed into an F-statistic using $F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2}T^2 \sim F_{p,n_1+n_2-p-1}$. The null hypothesis at a chosen significance level is rejected if the calculated value is greater than the F-table critical value. Rejecting the null hypothesis means that at least one of the parameters, or a combination of one or more parameters working together, is significantly different between the groups.

While mathematically similar, a Hotelling's T^2 test has a major advantage over a t-test [12]. Since a single comparison is made in the former test, the Type I error rate is well controlled and the relationship between multiple variables is taken into account. In summary, a t-test will denote which variables differ between groups; while a Hotelling's T^2 summarizes the between-group differences.

3.2 Mass Univariate Edge-wise Analysis

Using a single connectome for each participant as input, mass univariate edge-wise analysis involves performing a statistical test (typically, a t-test) to independently compare groups for each edge in a single connectome [6]. This results in a "difference matrix" of test statistics, representing the magnitude of group difference at each edge of that connectome. Multiple comparison correction for the $\frac{N\times(N-1)}{2}$ comparisons needs to be applied, where N is the number of nodes in the parcellation. Finally, for each edge, the null hypothesis can be rejected if the test statistic is greater than the critical value. Figure 1E-G shows an overview of the mass univariate edge-wise analysis.

3.3 Mass Multivariate Edge-wise Analysis

The proposed mass multivariate edge-wise analysis is a multivariate extension of the univariate approach. Using multiple task connectomes for each participant as inputs, this multivariate analysis involves performing a multivariate test (*e.g.* Hotelling's T^2) on connectivity strength for all tasks to create a "difference matrix" of test statistics. In this manner, the connectivity strength of an edge for each task is included in statistical testing. The "difference matrix" can then be corrected for multiple comparisons as above and thresholded for statistical significance. Post-hoc univariate tests (*e.g.* t-test) can be performed on each individual connectome to determine which task or tasks most likely contributed to edges exhibiting significant differences. Figure 1H–J shows an overview of the proposed mass multivariate edge-wise analysis.

4 Experiments

4.1 Datasets

We used two standard datasets in our analysis: the Human Connectome Project (HCP) and Philadelphia Neurodevelopmental Cohort (PNC) [7,8] (see Table 1). We narrowed the participants into the set of participants with mean frame-to-frame displacement less than 0.1 mm and maximum frame-to-frame displacement of less than 0.15 mm. The HCP dataset consists of 9 tasks: gambling (gam), emotion, language, motor, relation, social, working memory (wm), and two resting-state runs (rest1, rest2). The PNC dataset consists of three tasks: emotion, wm, and a resting-state run.

Table 1. Characteristics for the HCP and PNC datasets.

ID	Collection	#male	#female	Size	age	#tasks
HCP	Human Connectome Project	241	274	515	28 ± 3.98	9
PNC	Philadelphia Neurodevelopmental Cohort	251	320	571	15 ± 3.65	3

4.2 Preprocessing

For the HCP dataset, we started with the minimally preprocessed HCP data [13]. For the PNC dataset, functional images were slice-timed and motion-corrected and registered into common space as previously described [4]. Further preprocessing steps were performed using BioImage Suite [14]. These included regressing 24 motion parameters, regressing the mean time courses of the white matter, CSF, and grey matter, removing the linear trend, and low-pass filtering.

Regions were delineated according to the Shen atlas [15]. This atlas, defined in an independent dataset, provides a parcellation of the whole gray matter (including subcortex) into 268 contiguous, functionally coherent regions. These nodes have also been grouped into 10 functionally coherent "networks". For each scan, the average timecourse within each region was obtained, and the Pearson's correlation between the mean timecourses of each pair of regions was calculated. These correlation values provided the edge strengths for a 268×268 symmetric correlation matrix for each combination of subject, session, and run. These correlations were converted to be approximately normally distributed using a Fisher transformation.

4.3 Evaluation and Competing Methods

Using the HCP and PNC datasets, we evaluated our mass multivariate edge-wise analysis by quantifying the number of edges exhibiting significant differences between males and females. Significance was defined as $q < 0.005$ using the Storey procedure for positive False Discovery Rate (pFDR) correction [16].

We compared the number of significant edges between male and female participants detected by our mass multivariate edge-wise analysis to the number of significant edges detected by three competing approaches. First, we performed a standard mass univariate edge-wise analysis (as described in Sect. 3.2) for each task connectome, independently. For each task, we quantified the number of edges that exhibited significant differences after correcting for multiple comparisons using pFDR. These results provide a baseline for the amount of significant differences detectable in each connectome. Second, we quantified the union of the significant edges from each task from the first comparison approach. This result produces a comparison for the number of significant edges when naively combining information across all tasks. Third, we combined all task connectomes using the general functional connectivity method from [3] (*i.e.* averaging all connectomes), performed a standard mass univariate edge-wise analysis on this averaged connectome, and quantified the number of edges that exhibited significant differences after correcting for multiple comparisons using pFDR. This result produced a comparison for the number of significant edges when combining information across all tasks with a previously published method [3].

We would like to note that results based on independently performed mass univariate edge-wise analyses (*i.e.* the first and second competing approaches from above) do properly control for type I error as multiple comparison correction is only performed on independent analyses, not accounting the multiple tasks. This will inflate the number of edges detected with these approaches. However, type I error is well controlled for our mass univariate edge-wise analysis as connectomes are combined for a single comparison.

4.4 Visualization of Anatomical Locations of Significant Edges

To visualize anatomical locations of significant edges, we used stacked area plots to explain the probability of finding a significant edge in the network of interest. The 268 nodes were grouped into 10 networks for visualization: limbic system (Limb), default mode network (DMN), cerebellum (CBL), basal ganglia (BG), mediofrontal cortex (MF), motor areas (Mot), subcortical areas (Sc), visual association (VAs), visual-I (VI), visual-II (VII). Within a network of M nodes, the hypergeometric distribution gives the probability of finding a sex difference among all $\binom{M}{2}$ possible edges [17,18]). This is exactly analogous to the traditional example of drawing, without replacement, a white ball from a bag of $\binom{M}{2}$ white balls and $\binom{N}{2} - \binom{M}{2}$ black balls, where M is the number of nodes in a given network and N is the total number of nodes in the connectome.

5 Results

First, we compared the number of edges that were found to significantly differ between male and female participants as a result of the multivariate analysis with the number of edges that differed for each individual task in the univariate analysis (Table 2). A greater number of edges were found to differ between

groups when combining all connectomes compared to any single connectome. In the HCP dataset, the proposed multivariate approach detected significant sex effects in approximately 28% of the total number of edges; whereas, in the PNC dataset, the proposed multivariate approach detected significant sex effects in approximately 4.5% of the total number of edges.

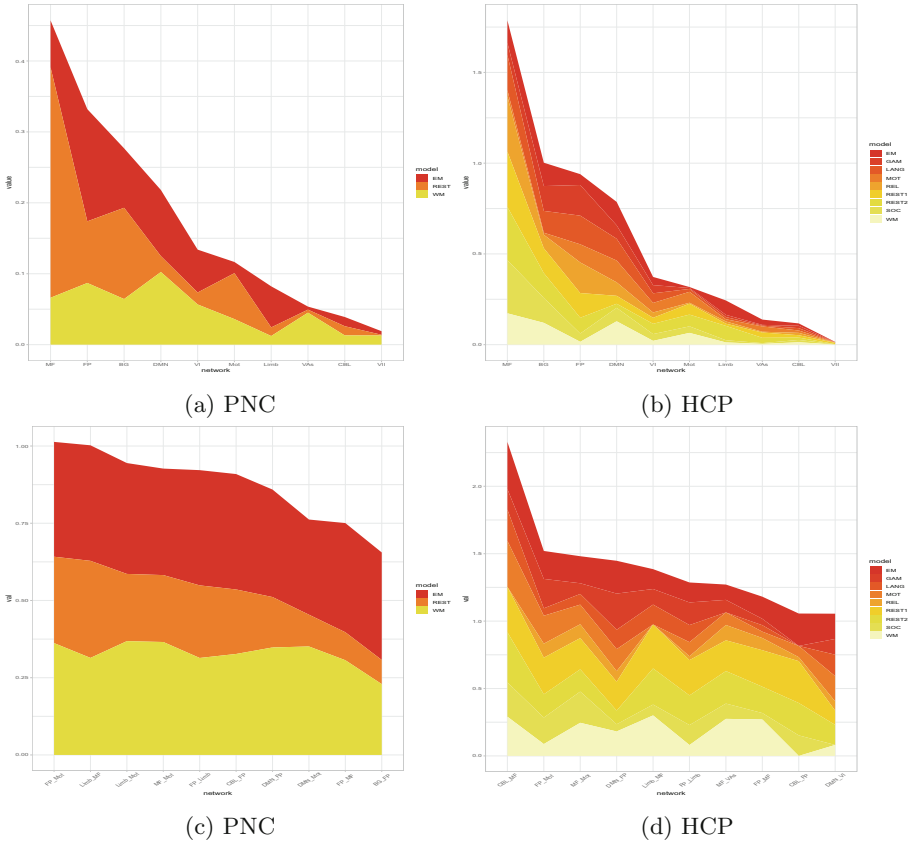
Next, we compared the number of significant edges for the proposed multivariate approach with two competing approaches that also combine all connectomes (Table 3). Our mass multivariate approach resulted in the greatest number of significant edges in HCP dataset. Yet a mass univariate edge-wise analysis on the mean connectome across tasks resulted in the greatest number of significant edges in the PNC dataset.

Table 2. The number of edges exhibiting significant differences between male and female participants for both univariate and multivariate approaches. For each task in the univariate approach, differences between groups were calculated for each connectome individually. For the multivariate approach, differences between groups were calculated using all task connectomes together. Significance was defined as $q < 0.005$ corrected for multiple comparisons using pFDR. Bold numbers show the rows with greatest number of significant edges.

	Task	#sig	%conn		Task	#sig	%conn		Task	#sig	%conn
HCP	gam	1355	3.79%	HCP	motor	1095	3.06%	PNC	emotion	1131	3.16%
	rest1	3331	9.31%		relation	716	2.00%		wm	883	2.47%
	rest2	2207	6.17%		social	1069	2.99%		rest	225	0.63%
	language	1315	3.67%		wm	1873	5.23%		multivariate	**1592**	**4.45%**
	emotion	1695	4.74%		multivariate	**9940**	**27.78%**				

Table 3. The number of significant edges for the 'union' (univariate) column are based on the union of all significant edges each individual univariate edge-wise analysis. The number of significant edges for the 'mean' (also univariate) column are based on the univariate edge-wise analysis using the average connectome of all tasks. The number of significant edges for the multivariate column are based on the proposed mass multivariate edge-wise analysis. Significance was defined as $q < 0.005$ corrected for multiple comparisons using pFDR. Bold numbers show the approaches with greatest number of significant edges.

		Union	Mean	Multivariate
HCP	#sig	8091	5701	**9940**
	%conn	22.61%	15.93%	**27.78%**
PNC	#sig	1740	**1920**	1590
	%conn	4.86%	**5.36%**	4.44%



(a) PNC (b) HCP

(c) PNC (d) HCP

Fig. 2. Area plots for within networks Fig. 2a and b, and between networks Fig. 2c and d. The horizontal axis shows network names and vertical axis shows the value of the hypergeometric function [17,18]. Networks are sorted according to the cumulative values across tasks.

Last, we investigated the anatomical locations of the significant edges detected using the proposed approach. The MF, FP, and BG networks had the greatest total number of significant edges, while the VAs and VII networks had the fewest (Fig. 2a–b). The DMN and FP showed the most consistent between-network association in the HCP dataset, while the Limb and MF networks show the most consistent between network associations in PNC dataset (Fig. 2c–d).

6 Discussion and Conclusions

In this work, we proposed a method to combine connectomes from multiple sources in the context of detecting group differences. To accomplish this, we extended a traditional mass univariate edge-wise analysis by incorporating multivariate statistics, where, for each edge, the connectivity strength for each task

is included in hypothesis testing. While connectomes derived from fMRI are typically generated from resting-state data, in this paper, we have shown that combining connectomes generated from multiple sources increases the amount of information useful for characterizing group differences. These results are in agreement with the recent work which has shown that connectomes generated from task data offer a significant improvement in detecting individual and group differences [3,4]. Furthermore, our results suggests that combining multiple connectomes derived from tasks that tap into multiple cognitive dimensions offers greater power than using a connectome from a single source. Although our mass multivariate edge-wise analysis performed better than the competing methods in the HCP dataset, this was not the case for the PNC dataset. One reason for these diverging results is the higher number of tasks in the HCP compared to the PNC dataset. Future work will include investigations into the optimal number of connectomes for mass multivariate edge-wise analysis. Additional future work will involve extending our mass multivariate edge-wise analysis to use multivariate analysis of variance (MANOVA)–the multivariate extension of an analysis of variance (ANOVA)–to compare multiple groups. Finally, our mass multivariate edge-wise analysis generalizes across different sources of connectomes as long as all connectomes have the same size. We will explore incorporating structural connectomes generated from DTI data and functional connectomes generated from EEG data. In conclusion, our results support the utility of combining multiple task connectomes to improve the detection of group differences.

Acknowledgements. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54 MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The remainder of the data used in this study were provided by the Philadelphia Neurodevelopmental Cohort (Principal Investigators: Hakon Hakonarson and Raquel Gur; phs000607.v1.p1). Support for the collection of the data sets was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia.

References

1. Dubois, J., Adolphs, R.: Building a science of individual differences from fMRI. Trends Cogn. Sci. **20**(6), 425–443 (2016)
2. Song, X.-W., et al.: REST: a toolkit for resting-state functional magnetic resonance imaging data processing. PLoS ONE **6**(9), e25031 (2011)
3. Elliott, M.L.: General functional connectivity: shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. NeuroImage **189**, 516–532 (2019)
4. Greene, A.S., Gao, S., Scheinost, D., Constable, R.T.: Task-induced brain state manipulation improves prediction of individual traits. Nat. Commun. **9**(1), 2807 (2018)

5. Gao, S., Greene, A., Constable, T., Scheinost, D.: Combining multiple connectomes improves predictive modeling of phenotypic measures. Neuroimage **201**, 116038 (2019)
6. Zalesky, A., Fornito, A., Bullmore, E.T.: Network-based statistic: identifying differences in brain networks. NeuroImage **53**(4), 1197–1207 (2010)
7. Van Essen, D.C., et al.: The WU-Minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)
8. Satterthwaite, T.D., et al.: The Philadelphia neurodevelopmental cohort: a publicly available resource for the study of normal and abnormal brain development in youth. Neuroimage **124**, 1115–1119 (2016)
9. Ystad, M.: Cortico-striatal connectivity and cognition in normal aging: a combined DTI and resting state fMRI study. Neuroimage **55**(1), 24–31 (2011)
10. Wang, J., et al.: Alternations in brain network topology and structural-functional connectome coupling relate to cognitive impairment. Front. Aging Neurosci. **10**, 404 (2018)
11. Deligianni, F., Centeno, M., Carmichael, D.W., Clayden, J.D.: Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. Front. Neurosci. **8**, 258 (2014)
12. Hotelling, H., et al.: A generalized t test and measure of multivariate dispersion. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. The Regents of the University of California (1951)
13. Glasser, M.F., et al.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)
14. Joshi, A.: Unified framework for development, deployment and robust testing of neuroimaging algorithms. Neuroinformatics **9**(1), 69–84 (2011)
15. Shen, X., Tokoglu, F., Papademetris, X., Constable, R.T.: Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage **82**, 403–415 (2013)
16. Storey, J.D.: A direct approach to false discovery rates. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **64**(3), 479–498 (2002)
17. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables, vol. 55. Courier Corporation, Chelmsford (1965)
18. Luke, Y.L.: Special Functions and their Approximations, vol. 2. Academic press, Cambridge (1969)