·Original Article·

Feature-reduction and semi-simulated data in functional connectivity-based cortical parcellation

Xiaoguang Tian[2,*], Cirong Liu[2,5,*], Tianzi Jiang[4,5], Joshua Rizak[2], Yuanye Ma[1,2,3,6], Xintian Hu[1,2,3,6]

[1]*Yunnan Key Lab of Primate Biomedical Research, China*

[2]*Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China*

[3]*State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China*

[4]*LIAMA Center for Computational Medicine, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China*

[5]*The University of Queensland, Queensland Brain Institute, QLD 4072, Australia*

[6]*Kunming Bio-International*

*These authors contributed equally to this work.

Corresponding authors: Yuanye Ma and Xintian Hu. E-mail: yuanma0716@vip.sina.com; xthu@mail.kiz.ac.cn

ABSTRACT

Recently, resting-state functional magnetic resonance imaging has been used to parcellate the brain into functionally distinct regions based on the information available in functional connectivity maps. However, brain voxels are not independent units and adjacent voxels are always highly correlated, so functional connectivity maps contain redundant information, which not only impairs the computational efficiency during clustering, but also reduces the accuracy of clustering results. The aim of this study was to propose feature-reduction approaches to reduce the redundancy and to develop semi-simulated data with defined ground truth to evaluate these approaches. We proposed a feature-reduction approach based on the Affinity Propagation Algorithm (APA) and compared it with the classic feature-reduction approach based on Principal Component Analysis (PCA). We tested the two approaches to the parcellation of both semi-simulated and real seed regions using the K-means algorithm and designed two experiments to evaluate their noise-resistance. We found that all functional connectivity maps (with/without feature reduction) provided correct information for the parcellation of the semi-simulated seed region and the computational efficiency was greatly improved by both feature-reduction approaches. Meanwhile, the APA-based feature-reduction approach outperformed the PCA-based approach in noise-resistance. The results suggested that functional connectivity maps can provide correct information for cortical parcellation, and feature-reduction does not significantly change the information. Considering the improvement in computational efficiency and the noise-resistance, feature-reduction of functional connectivity maps before cortical parcellation is both feasible and necessary.

Keywords: cortical parcellation; resting-state fMRI; functional connectivity; feature reduction; stimulated data; AP algorithm

INTRODUCTION

There is a consensus that the cerebral cortex can be subdivided into several structurally and functionally distinct regions. The ability to identify these regions is fundamental to evaluation of the normal and/or abnormal brain functions associated with neurological disorders. Although a surge of microarchitectonic and invasive tracer studies has provided substantial microarchitecture and connectivity information regarding cortical parcellation in non-human primates[1,2], similar evidence concerning the parcellation of the human brain is scarce; it is mostly limited to post-mortem observations based on microarchitecture[3-7] or on anatomical landmarks[8]. However, parcellation based on anatomical landmarks and microarchitecture does not always provide accurate functional segregation between distinct areas[9]. In fact, the function of a cortical area is also determined by its extrinsic connections in addition to this intrinsic microarchitecture. In non-human primates, it has been demonstrated that the extrinsic connection of each cortical area represents a unique ‘connectional fingerprint’[10]. Therefore, it has become an important criterion to include the ‘connectional fingerprint’ in the parcellation of functionally distinct brain areas.

Investigation of the ‘connectional fingerprint’ in the human brain is challenging due to the inability to use invasive tracing techniques as well as the limitations of post-mortem anatomical evaluation. The use of resting-state functional magnetic resonance imaging (rs-fMRI) has led to a number of studies in recent years that have investigated this functional ‘connectional fingerprint’ in relation to cortical parcellation[11-22]. By measuring the cross-correlation of low-frequency blood oxygenation level-dependent fluctuations between brain voxels while subjects are at rest[23], one can calculate the functional connectivity between any pair of voxels. This cross-correlation measurement generates functional connectivity maps between voxels, where the functional connectivity of each voxel to all the voxels of the brain is represented. Based on these functional connectivity maps, voxels can then be classified by cluster analysis. This method using rs-fMRI to perform cortical parcellation has provided fine-grained functional subdivisions of the anterior cingulate cortex[11], precuneus[24,25], thalamus[22], basal ganglia[13], lateral parietal cortex[14], medial frontal cortex[12,18], and insular cortex[15,16], as well as the subdivision of the whole brain[17,26]. The fine-grained functional subdivisions of brain regions can be further used to investigate the organization of the human cerebral cortex[26,27].

These parcellation approaches based on functional connectivity maps have proven feasible and promising. However, a whole brain contains a great number of voxels and adjacent voxels are always highly correlated, so functional connectivity maps contain redundant information. For example, a classic preprocessed fMRI dataset with 3-mm isotropic resolution contains >40000 voxels and a region like the cingulate cortex contains up to thousands of voxels. A functional connectivity map for each voxel of the cingulate cortex then involves its functional connectivity with each voxel of the whole brain. The total number of features for each voxel in the cingulate cortex then becomes >40000. Therefore, performing cluster analysis on thousands of voxels in >40000 dimensions becomes computationally inefficient. Furthermore, during clustering, the functional connectivity of each voxel with a target voxel is always regarded as an independent feature, but a voxel is not an independent functional unit of the brain. When we characterize the ‘connectional fingerprint’ of a target region, it is more reasonable to make feature elements at the region level rather than at the voxel level. Under a strict definition, each feature element should be independent and not highly correlated. If we treat each voxel as one feature, we fail to meet the requirement of independence. Considering that each voxel is not independent and is always highly correlated with other voxels, it becomes advantageous to treat voxels in the same functional unit as one feature by averaging these voxels.

In this study, we proposed a feature-reduction approach based on the affinity propagation algorithm (APA)[28] and compared it with the classic approach based on principal component analysis (PCA)[29]. The approach based on APA provides feature-reduction by averaging the time courses of all voxels located within the same functional unit. This allows these voxels to be treated as one feature during the clustering procedure. To evaluate these different approaches, we proposed to build semi-simulated data for connectivity-based parcellation. The semi-simulated data built here were based on real fMRI data, reflecting a complex brain connectivity pattern, and the ground truth was clearly defined. Since we knew the ground truth of which voxel of this ‘semi-simulated’ seed region belonged to which ‘real’ seed region, the conclusions from our analysis were more reliable.

PARTICIPANTS AND METHODS

Participants and Data Acquisition

The analysis was performed on the *New York Test-Retest Reliability* dataset of *the 1000 Functional Connectomes Project*[30]. The dataset consists of 6.5-min scans acquired from 25 healthy subjects (10 males and 15 females) at three different time points on a 3T Siemens Allegra scanner using an echo-planar imaging (EPI) sequence (time repetition (TR) = 2 000 ms; time echo (TE) = 25 ms; flip angle = 90°; 39 slices, matrix = 64×64, field of view (FOV) = 192 mm; acquisition voxel size = 3 × 3 × 3 mm^3). The first scanning session of each subject was used in this study. A high-resolution T1-weighted anatomical image using magnetization prepared gradient echo (TR = 2 500 ms; TE = 4.35 ms; flip angle = 8°; 176 slices, FOV = 256 mm) was also obtained for spatial normalization.

fMRI Data Preprocessing

Image preprocessing was performed using AFNI[31] and FSL[32]. In brief, the data were motion-corrected to the mean image volume and were spatially smoothed using a 6-mm FWHM Gaussian kernel, following which the fMRI data were band-pass filtered (0.005 Hz < f < 0.1 Hz), linear and quadratic trends were removed, and the data were de-noised by regressing out the global, white-matter and cerebrospinal fluid signals and 6 motion parameters. The fMRI data were written into MNI152 standard space with concatenated transformations from functional volume to anatomical volume (linear) and spatial normalization of the structural MR images to MNI152 (non-linear), and were restricted using a gray matter mask to reduce the number of non-gray matter voxels and to improve the computational efficiency.

Selection of Seed Regions

Semi-simulated seed region We constructed a semi-simulated seed region by combining six ‘real’ seed regions from different parts of the brain into one ‘semi-simulated’ region. The ground truth was defined since we knew exactly which voxels of this ‘semi-simulated’ seed region belonged to which ‘real’ seed region. Three of the six ‘real’ seed regions were task-positive and were centered on the intraparietal sulcus (-25, -57, -46), the frontal eye field (25, -13, 50), and the middle temporal region (-45, -69, -2)[33]. The remaining three seed regions were task-negative and were centered on the medial prefrontal cortex (-1, 47, -4), posterior cingulate/precuneus (-5, -49, 40), and lateral parietal cortex (-45, -67, 36)[33]. We masked the six “real” seed regions with the same grey-matter mask used in the fMRI data preprocessing and combined the six “real” seed regions into one semi-simulated seed region. The total number of voxels contained in the semi-simulated seed region was 156.

Real seed regions Besides the semi-simulated seed region, we also selected three real seed regions of different sizes – the right supplementary motor area (R-SMA), cingulate cortex, and right prefrontal cortex (R-PFC) to test the different approaches. All seed regions were extracted from the anatomical automatic labeling template[8], the R-SMA being area 20, the cingulate cortex containing areas 31 to 36, and the R-PFC consisting of areas 3-4, 7-8 and 11-14. We masked these seed regions with the same grey-matter mask used in the fMRI data preprocessing so that they would match the preprocessed fMRI data. The final sizes were 516 voxels for the R-SMA, 1975 for the cingulate cortex and 2641 for the R-PFC.

Functional Connectivity Map Analysis

The map for each voxel showed its functional connectivity with all other voxels within the whole brain. Functional connectivity between a pair of voxels was represented by the Pearson correlation coefficient of their preprocessed time series. We averaged the functional connectivity maps of the 25 participants to obtain the group raw functional connectivity map.

Functional Connectivity Maps by Different Feature-reduction Approaches

Based on affinity propagation We first applied a gross parcellation of all voxels within the whole brain using the APA (Fig. 1). Then, an average of the time series of the voxels in the same resulting clusters was taken. Finally, new functional connectivity maps of seed regions were generated by the correlations representing the relationship between raw seed region time series and the new whole-brain time series. Details of the APA are in *Supplemental Data.*

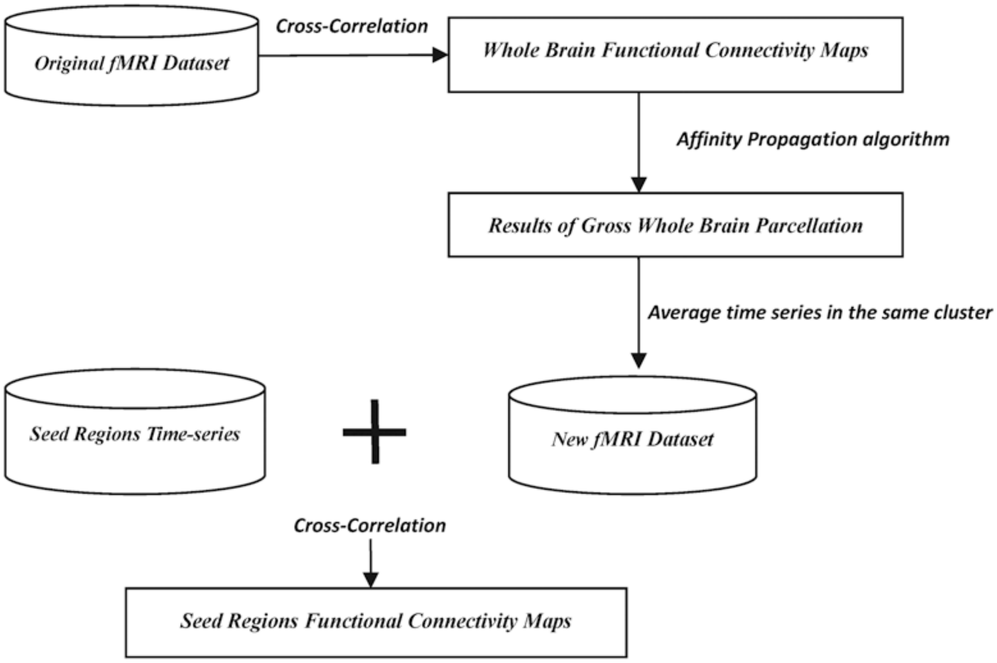
Original fMRI Dataset
Cross-Correlation
Whole Brain Functional Connectivity Maps
Affinity Propagation algorithm
Results of Gross Whole Brain Parcellation
Average time series in the same cluster
Seed Regions Time-series
New fMRI Dataset
Cross-Correlation
Seed Regions Functional Connectivity Maps

Fig. 1. Flow-chart of the APA-based feature-reduction approach.

In this experiment, we first averaged the functional connectivity maps of the whole brain for 25 participants to obtain the group whole-brain functional connectivity map and then performed gross clustering of the whole brain as described above. By averaging the time courses of voxels in the same clusters, we obtained a new whole-brain time course in each participant. The seed region functional connectivity maps were then calculated from the correlations between seed region time courses and the new whole-brain course in each participant. Finally, by averaging the seed region functional connectivity maps of the 25 participants, we obtained the group seed region functional connectivity map with APA-based feature reduction.

Based on principal component analysis The feature-reduction approach by PCA was directly done on the raw functional connectivity maps. PCA is a mathematical procedure that uses an orthogonal transformation to covert a set of correlated variables into a set of linearly uncorrelated variables (principal components). Details of PCA are in *Supplemental Data*.

In this experiment, we averaged the raw seed region functional connectivity maps of 25 participants to obtain the group functional connectivity map and then performed the feature-reduction by PCA on this map. We selected all principal components (PCA-all) and principal components covering 95% of the cumulative contribution (PCA-95) to build new functional connectivity maps.

Seed Region Parcellation

In order to investigate whether the feature-reduction procedures significantly changed the information in functional connectivity maps, cluster analysis using the K-means algorithm was applied to each seed region (the K-means algorithm was designed to partition *n* observations into k clusters where each observation belongs to the cluster with the nearest mean). We used the standard K-means algorithm of *Matlab*. The algorithm was repeated 1024 times for all seed regions. The goal number of clusters (K) must be defined for this algorithm. Therefore, it was set as K = 6 for the semi-simulated seed region (ground truth was K = 6), K = 2-10 for the R-SMA, K = 2-15 for the cingulate cortex and K = 2-30 for the R-PFC.

To quantitatively evaluate the improvement of computational efficiency brought by the feature-reduction approaches, we compared their time costs during K-means cluster analysis. We recorded the average computational cost of one iteration and the average number of iterations for the convergence of one repetition. Besides, the total computational cost of K-means was also related to the number of repetitions. Therefore, we had to determine the number of repetitions needed for K-means to reach the global minimum solution. To do this, we estimated the probability of the global minimum solution by recording the number of minimum solutions with the defined number of repetitions (1024).

The experimental platform was rack-mounted servers, Dawning I840 with CPUs 12 × 4 AMD Opteron 6174 (2.2 GHz) and 256 GB of memory; the operating system was Mandriva Linux.

We used the inconsistency rate to compare different parcellation results. This rate was calculated as the percentage discrepancy between different connectivity matrices[20]. A connectivity matrix was defined as follows: for the resulting clusters with N voxels in one seed region, a connectivity matrix $M_{N\times N}$ can be generated, where the element $M_{i,j}$ equals 1 if both voxels v_i and v_j are not found in the same cluster and equals 0 otherwise.

For the semi-simulated seed region, we compared the resulting clusters of all functional connectivity maps with the ground truth. Thus, we were able to evaluate whether functional connectivity maps provide effective information for cortical parcellation and whether feature-reduction significantly changes the information of the functional connectivity maps. For the real seed regions, we did not know the ground truth. The inconsistency rate was only used to outline the difference between the resulting clusters based on different functional connectivity maps.

Noise-resistance Analysis

The inconsistency rate cannot tell which feature-reduction approach is better. In this section, we designed simulated data to evaluate the noise-resistance of the feature-reduction approaches. The simulated data were generated based on the semi-simulated seed regions described above (Fig. 2). In brief, we randomly selected one participant and constructed the simulated data comprising the time courses of the semi-simulated seed region in this subject. Then we carried out two experiments by adding noise to the simulated data in different ways. We repeated each experiment 50 times.

(1) Different levels of time-course signal-to-noise ratio (tSNR): randomly-selected time points (the number was 80 or 190) of all voxels were mixed with Gaussian noise under different levels of tSNR (tSNR = 10, 7, 5, 2, 1.5, 1, 0.75, 0.5, 0.25, and 0.125). We randomly selected 50 voxels from the simulated data and treated them as a seed region. We then carried out K-means cluster analysis based on the seed region's functional connectivity maps (APA, PCA-95, and PCA-all/Raw).

(2) Different numbers of time points: A different number of randomly-selected time points for all voxels were mixed with Gaussian noise under a higher or lower tSNR (tSNR = 2.0 or 0.1). The number of points varied from 5 to 190 (5, 10, 30, 50, 80, 120, 150,180, and 190). We then randomly selected 50 voxels from the simulated data, treated them as a seed region, and carried out K-means cluster analysis based on the functional connectivity maps of the seed region (APA, PCA-95, and PCA- all/Raw).

"Eigenmap" of PCA

Moreover, we used the "Eigenmap" concept of PCA to compare the characteristics of different real seed regions. In mapping the modulus of the first eigenvector of PCA to the cortical surface using Caret5 software[34], the parts of the brain that provide the most important features for the parcellation of a certain brain region can easily be identified. To quantitatively compare the first Eigenmaps of different seed regions, we also performed repeated measures one-way ANOVA on the three Eigenmaps, followed by the multiple comparison test with the Bonferroni correction. A whole flowchart of this study was given in Fig. S1.

RESULTS

Feature Reduction

The PCA-based approach The feature-reduction approach based on PCA produces N-1 (N is the number of voxels in each seed region) components responding to non-zero eigenvalues. Therefore, after using all components (PCA-all), the sizes of functional connectivity maps of the four seed regions were reduced to 516 × 515 (R-SMA), 1975 × 1974 (cingulate cortex), 2641 × 2640 (R-PFC), and 156 × 155 (semi-simulated seed region).

The APA-based approach Based on the group whole-

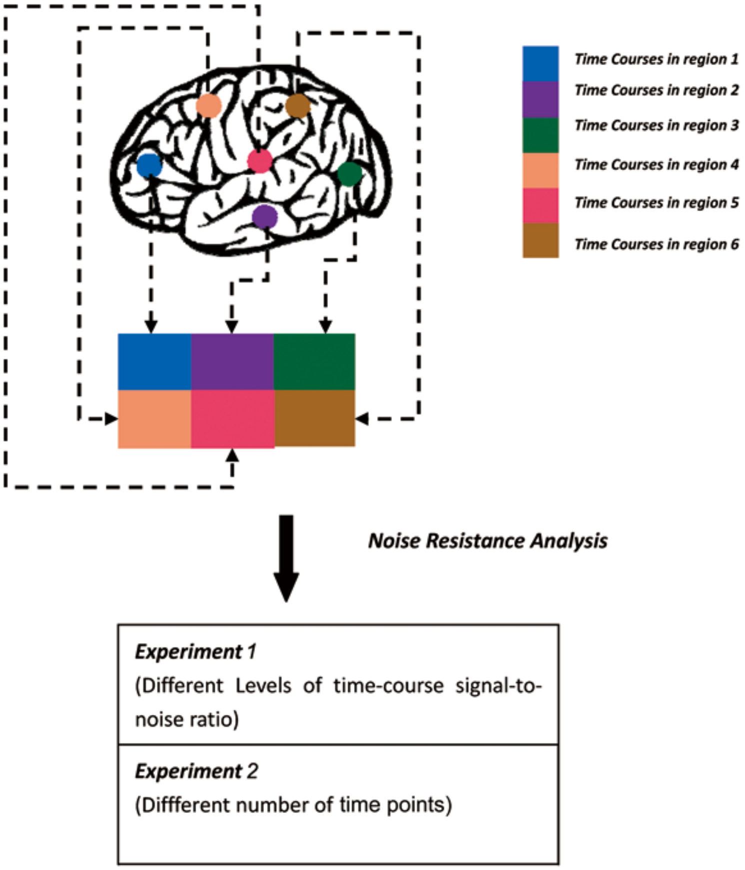

Fig. 2. Flow-chart of noise-resistance analysis. We designed simulated data to evaluate the feature-reduction approaches on their noise-resistance. We randomly selected one participant and constructed the simulated data comprising the time-courses for the semi-simulated seed region. Then we carried out two experiments by adding noise to the simulated data in different ways (A-B). We repeated each experiment 50 times.

brain functional connectivity maps, the APA divided all the voxels of the brain (45161 voxels in total) into 2881 clusters. The cluster size ranged from 7 to 36 voxels. The time courses of all voxels in each cluster were averaged and all voxels in each cluster were converted into one feature to be used in the following cluster analysis of the seed regions. The sizes of the functional connectivity maps in these regions were reduced from 516 × 45161 to 516 × 2881 (R-SMA), 1975 × 45161 to 1975 × 2881 (cingulate cortex), 264 × 45161 to 2641 × 2881 (R-PFC) and 156 × 45161 to 156 × 2881 (semi-simulated seed region).

Computational Cost of K-means

Table 1 shows a comparison of the computational cost of K-means on functional connectivity maps without feature-reduction (Raw), with APA-based feature-reduction, and with PCA-based feature-reduction selecting all components (PCA-all). For all functional connectivity maps, both the average computational cost of one iteration (t) and the average number of iterations for the convergence of one repetition (Num) increased when the number of clusters (K) and the size of the seed region became larger. Even though there was no evident difference in the average number of iterations for convergence among the three kinds of functional connectivity maps, the average computational cost of one iteration and one repetition for the raw functional connectivity maps was much larger than that of the functional connectivity maps with feature-reduction. Take K = 15 of the cingulate cortex for example. The computational cost of one repetition for the raw functional connectivity maps was 1128.57 s. After using the feature reduction approaches, the computational efficiency

Table 1. The computational cost of K-means based on different functional connectivity maps (Raw, PCA-all, APA)

	Semi-simulated seed region			SMA-R			Cingulate Cortex		
	Num	*t* (s)	*T*(s)= *Num*t*	*Num*	*t* (s)	*T*(s)= *Num*t*	*Num*	*t* (s)	*T*(s)= *Num*t*
Raw									
K=2	-	-	-	8.28	2.12	17.58	7.32	5.03	36.85
K=3	-	-	-	12.03	2.81	33.78	14.11	6.47	91.31
K=4	-	-	-	20.00	3.32	66.44	20.67	7.50	155.11
K=5	-	-	-	21.44	4.18	89.68	27.97	12.25	342.67
K=6	6.42	2.76	17.75	22.50	4.85	109.13	32.47	16.94	549.85
K=7	-	-	-	24.92	5.49	136.80	36.35	12.97	471.37
K=8	-	-	-	24.50	6.13	150.19	31.83	16.00	509.32
K=9	-	-	-	24.23	6.62	160.36	31.75	15.84	502.91
K=10	-	-	-	24.99	7.31	182.70	33.31	18.48	615.67
K=11	-	-	-	-	-	-	36.19	20.61	745.78
K=12	-	-	-	-	-	-	37.90	23.74	899.69
K=13	-	-	-	-	-	-	38.24	24.72	945.51
K=14	-	-	-	-	-	-	40.87	26.01	1063.09
K=15	-	-	-	-	-	-	41.52	27.18	1128.57
PCA-all									
K=2	-	-	-	8.28	0.02	0.18	7.30	0.21	1.51
K=3	-	-	-	12.03	0.03	0.35	14.16	0.28	3.98
K=4	-	-	-	20.01	0.04	0.80	20.52	0.33	6.70
K=5	-	-	-	21.44	0.04	0.92	27.85	0.46	12.86
K=6	6.40	0.01	0.08	22.50	0.06	1.34	33.08	0.54	17.80
K=7	-	-	-	24.92	0.06	1.56	37.01	0.51	18.85
K=8	-	-	-	24.50	0.08	1.98	32.38	0.58	18.68
K=9	-	-	-	24.23	0.10	2.35	32.63	0.63	20.49
K=10	-	-	-	24.99	0.11	2.73	34.14	0.70	23.80
K=11	-	-	-	-	-	-	36.20	0.71	25.66
K=12	-	-	-	-	-	-	37.11	0.77	28.42
K=13	-	-	-	-	-	-	38.85	0.81	31.54
K=14	-	-	-	-	-	-	40.28	0.86	34.70
K=15	-	-	-	-	-	-	40.84	0.86	35.19
APA									
K=2	-	-	-	8.21	0.16	1.30	6.64	0.39	2.59
K=3	-	-	-	9.27	0.23	2.13	16.20	0.48	7.76
K=4	-	-	-	20.80	0.24	5.00	20.63	0.61	12.69
K=5	-	-	-	22.03	0.30	6.57	28.43	0.74	21.01
K=6	6.42	0.20	1.26	24.03	0.37	8.88	30.70	0.90	27.61
K=7	-	-	-	24.42	0.40	9.80	35.50	1.35	47.82
K=8	-	-	-	24.11	0.44	10.49	30.57	1.26	38.67

(To be continued)

(Continued)

K=9	-	-	-	23.97	0.46	11.06	32.03	1.34	42.79
K=10	-	-	-	25.20	0.53	13.37	33.55	1.37	45.87
K=11	-	-	-	-	-	-	36.03	1.41	50.73
K=12	-	-	-	-	-	-	36.20	1.61	58.33
K=13	-	-	-	-	-	-	37.16	1.72	63.89
K=14	-	-	-	-	-	-	39.31	1.82	71.44
K=15	-	-	-	-	-	-	39.86	1.80	71.77

*Num represents the average number of iterations for the convergence of one repetition; *t* represents the average computational cost of one iteration; *T* represents the average computational cost of one repetition; K represents the number of clusters defined in K-means; Raw represents the raw functional connectivity maps; PCA-all represents the functional connectivity maps with PCA-based feature reduction selecting all non-zero components; APA represents the functional connectivity maps with APA-based feature reduction.

improved considerably; the computational cost-saving was 97% with PCA-all and 94% with APA.

The frequency of the minimum solution was estimated based on the results shown in Table 2. As the number of clusters (K) and the size of seed regions increased, the frequency of minimum solution sharply decreased, no

Table 2. Minimum solution of K-means within 1024 repetitions for semi-simulated seed region

	R-SMA			Cingulate Cortex			R-PFC	
	Raw	*PCA-all*	*APA*	*Raw*	*PCA-all*	*APA*	*PCA-all*	*APA*
K=2	1023	1023	1024	1024	1024	1024	13	9
K=3	1024	1024	409	334	327	223	5	24
K=4	18	14	5	17	30	1	2	98
K=5	19	49	19	139	161	60	1	27
K=6	2	6	70	280	299	392	1	11
K=7	3	8	18	88	77	4	3	22
K=8	3	16	9	126	136	3	2	1
K=9	1	2	18	33	28	7	1	19
K=10	1	4	79	2	6	11	1	5
K=11	-	-	-	1	1	19	1	1
K=12	-	-	-	1	1	2	2	1
K=13	-	-	-	3	2	1	1	9
K=14	-	-	-	4	3	2	1	1
K=15	-	-	-	1	2	14	1	1
K=16-30	-	-	-	-	-	-	1	1
	Semi-simulated seed region							
K=6	10	10	8					

*Raw represents the raw functional connectivity maps; PCA-all represents the functional connectivity maps with PCA-based feature-reduction selecting all non-zero components; APA represents the functional connectivity maps with APA-based feature reduction; K represents the number of clusters defined in K-means.

Table 3. Inconsistency rate between resulting clusters of functional connectivity maps without feature reduction (Raw) and those with APA-based feature reduction (APA)

	Raw *vs* PCA-all		Raw *vs* PCA-95		Raw *vs* APA	
	SMA-R	Cingulate Cortex	SMA-R	Cingulate Cortex	SMA-R	Cingulate Cortex
K=2	0	0	0	0	0	0
K=3	0	0	0	0	0.82%	0.31%
K=4	0	0	0	0	0.99%	0.83%
K=5	0	0	0	0.04%	0.97%	2.72%
K=6	0	0	0	0.15%	0.30%	0.81%
K=7	0	0	0.88%	0	2.01%	0.45%
K=8	0	0	0.21%	0.05%	2.19%	0.09%
K=9	0	0	0.27%	0.03%	0.11%	0.30%
K=10	0	0	0.11%	0.15%	0.23%	0.30%
K=11	-	0	-	0.72%	-	0.35%
K=12	-	0.13%	-	0.34%	-	0.34%
K=13	-	0	-	0.80%	-	0.96%
K=14	-	0	-	0.16%	-	0.44%
K=15	-	0.03%	-	0.09%	-	0.20%

*K represents the number of clusters defined in K-means. Examples of resulting clusters were given in Figs. S2 and S3.

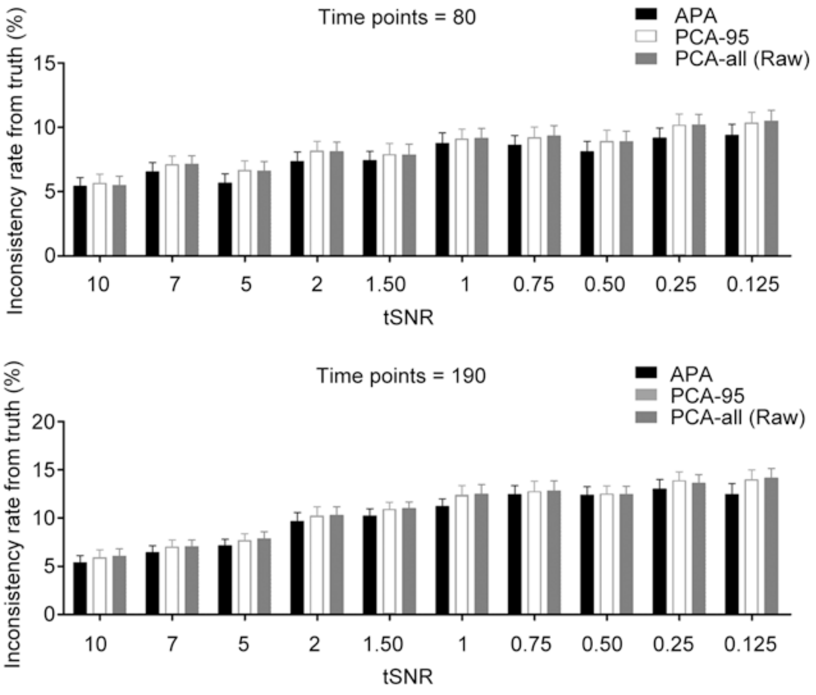

Fig. 3. Results of noise-resistance analysis in experiment 1. Randomly-selected time points of all voxels were mixed with Gaussian noise under different levels of tSNR from 10 to 0.125. We randomly selected 50 voxels from the simulated data and treated them as a seed region. We then carried out K-means cluster analysis based on the seed region's functional connectivity maps (APA, PCA-95, and PCA-all). The inconsistency rate with respect to the ground truth was used to evaluate the resulting clusters. This experiment was repeated 50 times. Data are plotted as the mean with the 95% confidence interval.

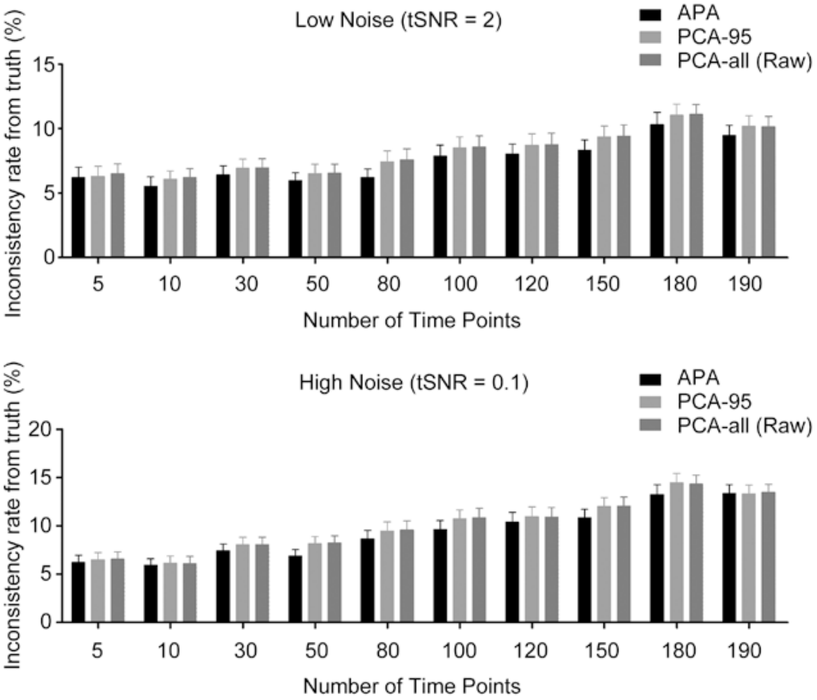

Fig. 4. Results of noise-resistance analysis in experiment 2. Different numbers of randomly selected time points for all voxels were mixed with Gaussian noise under a higher or lower tSNR. The number of time points was varied from 5 to 190. We then randomly selected 50 voxels from the simulated data, treated them as a seed region, and carried out K-means cluster analysis based on the functional connectivity map of the seed region (APA, PCA-95, and PCA-all). The inconsistency rate with respect to the ground truth was used to evaluate the resulting clusters. This experiment was repeated 50 times. Data are plotted as the mean with the 95% confidence interval.

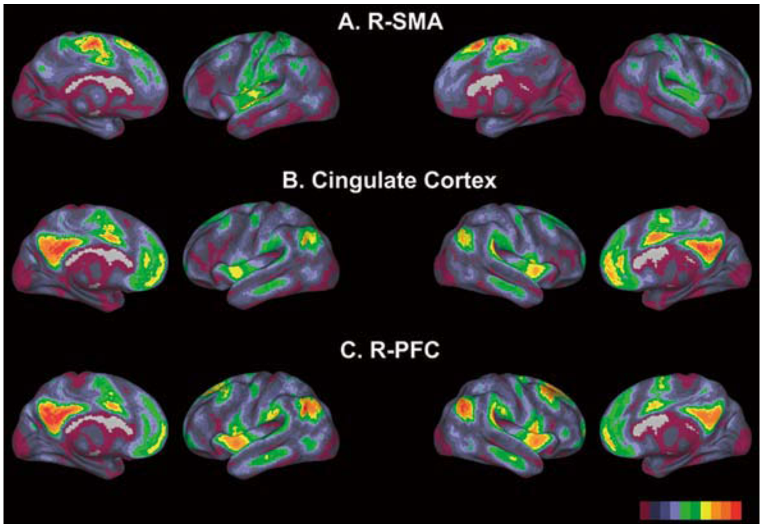

Fig. 5. Surface mapping of the first Eigenmaps of the three real seed regions. The first Eigenmap of each seed region, R-SMA (A), cingulate cortex (B) and R-PFC (C) is displayed. The color bar is defined with the warmest color representing the highest value and the coldest color representing the lowest value.

matter which feature-reduction approach we chose. 1024 repetitions were definitely enough for the global minimum solution when K or the size of seed regions was small, but were not sufficient when K or the size of seed regions was large.

Specifically, the 1024 repetitions were not enough to establish a global minimum solution for the parcellation of the R-PFC. In most circumstances, the number of times the minimum solution was found was only once, and it was not known whether this was a local minimum or a global minimum (probability ≤1/1024). Based on the resulting R-PFC clusters, further evaluation of the inconsistency rate would be inaccurate. Thus, we did not present the results of inconsistency rate for R-PFC.

Comparison of Parcellation Results (Inconsistency Rate)

Semi-simulated seed region All functional connectivity maps provided correct information for the parcellation of the semi-simulated seed region, as the inconsistency rates between the ground truth (K = 6) and the resulting clusters based on different kinds of functional connectivity maps (Raw, PCA-all, PCA-95 and APA) were all zero.

Real seed regions Since the ground truth for the real seed regions was unavailable, we used the clustering results based on Raw as the baseline to calculate the inconsistency rate. Functional connectivity maps with feature-reduction gave results similar or identical to Raw. The resulting clusters based on Raw and those based on PCA-all were identical, except for the low inconsistency rate when K = 12 and K = 15 in the cingulate cortex (Table 3). In most circumstances, the inconsistency between resulting clusters based on Raw and those based on PCA-95 or APA was low.

Noise-resistance Analysis

Experiment 1 was designed to illustrate the noise-resistance of the APA-based approach and the PCA-all approach under different levels of time-course signal-to-noise ratio (tSNR). Under all tSNRs, the performances using the APA-based approach were better than those using PCA-all (Raw) and PCA-95 (Fig. 3).

Experiment 2 was designed to illustrate the noise-resistance of the two approaches under different numbers of noise-added time points. Similar to the above experiments, the APA-based approach outperformed the PCA-all/Raw and PCA-95 approaches (Fig. 4).

First Eigenmaps Comparison

The first Eigenmap of R-SMA differed from those of the cingulate cortex and the R-PFC, whereas the first Eigenmaps of the cingulate cortex and the R-PFC were similar. Areas with high values in the first Eigenmap of R-SMA were limited to the bilateral SMA and posterior insular cortex. Areas with high values in the first Eigenmaps of the cingulate cortex and the R-PFC were similar and more distributed, including the anterior insular cortex, posterior cingulate cortex/precuneus/retrosplenial region, anterior cingulate cortex, medial prefrontal cortex, dorsolateral prefrontal cortex, lateral parietal lobule and interior temporal lobule (Fig. 5).

One-way repeated measures ANOVA revealed significant differences in the three Eigenmaps [F(2,135480) = 5.559, P <0.0001] (Table S1). The *post hoc* test showed that both the first Eigenmap of the R-PFC and that of cingulate cortex differed from the first Eigenmap of the R-SMA (P < 0.01, corrected), but no significant difference was found between the first Eigenmap of the R-PFC and that of the cingulate cortex (Fig. 6).

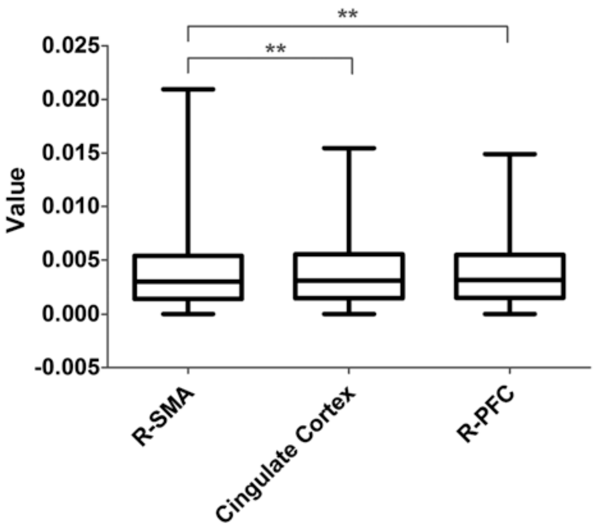

Fig. 6. Box-and-whisker plots of the first Eigenmaps of the three real seed regions. The bottom and top of the box are the lower and upper quartiles, the band near the middle is the median, and the ends of the whiskers represent the minimum and maximum. **significantly different Eigenmaps.

DISCUSSION

Computational Efficiency

The computational cost of K-means involves three parts: (1) the average computational cost of one iteration; (2) the average number of iterations for the convergence of one repetition; and (3) the number of repetitions needed for a global minimal solution. In this study, we calculated the detailed computational cost for each part and showed how the feature-reduction procedures improved computational efficiency.

Take the cingulate cortex for example. When the number of clusters (K) was set at 15 for the raw functional connectivity maps, the average computational cost of one iteration was 27.18 s and the average number of iterations for one repetition was 41.52. Thus, on average, 1128 s (41.52 × 27.18 s) guaranteed the convergence of one repetition. With the power of modern computing, the average computational cost of one iteration or one repetition for K-means is inexpensive. However, as the number of repetitions increases, the corresponding number of iterations increases many-fold. This leads to a computational cost that modern computing can no long afford, especially when a high data resolution is used. In this study, we set the number of repetitions at 1024, which took >300 h (1024 × 1128 s) to do the parcellation of cingulate cortex with K equal to 15. Moreover, as shown in Table 2, the possibility for a minimum solution for the above example was ≤0.098% (1/1024). In other words, 1024 repetitions could not guarantee the global minimum. However, increasing the number of repetitions may not be a feasible way to obtain a global minimum for the raw functional connectivity maps, especially when seed regions and K are large, because it would require a greatly increased computational cost. Therefore, one feasible solution is to carry out feature-reduction, as its high computational efficiency would then compensate for the computational cost brought by the increased number of repetitions. For example, the PCA-all feature-reduction reduced the average computational cost of one iteration to 0.86 s, 32-fold less than the computational cost of the raw functional connectivity maps. Thus, it only took 35 s to complete one repetition and 10 h to finish 1024 repetitions on average to give identical results.

Besides feature-reduction, efficiency could also be improved by a better clustering algorithm. Nevertheless, reducing the dimensions of the data itself would be a more extensive and basic method and does not conflict with these better algorithms (feature reduction can be applied to these algorithms).

PCA-based Feature Reduction

Due to the specific characteristic of functional connectivity maps that the number of features is greatly larger than the number of voxels, the feature-reduction approach based on PCA produces N-1 (N is the number of voxels in each seed region) components responding to non-zero eigenvalues. Taking all components (PCA-all), the resulting clusters showed little or no difference from the clusters resulting from the raw functional connectivity maps (Table 3). The reason is that the PCA-all was totally based on the raw functional connectivity maps. The variance of PCA-all would be expected to be the same as that of the raw functional connectivity maps, so that the final resulting clusters of both functional connectivity maps would be identical. K = 12 and K = 15 on the cingulate cortex are two cases in which K-means did not reach the global minimum solution so that a low inconsistency rate was found (Table 2). Thus, PCA-all reduced the number of features, improved the computational efficiency, and gave exactly the same results as the raw functional connectivity maps.

It is common to use the first few components for clustering (covering most of the variance) while performing PCA-based feature-reduction because it can further improve computational efficiency and may reduce noise[39]. The noise-resistance of the PCA-based approach depends on the separation of signal and noise in different components and on the selection of the right number of components. However, the performance of PCA-95 in noise-resistance analysis was similar to that of PCA-all (Raw), and did not improve. This indicated that the PCA-based feature-reduction may not separate signal and noise well enough for functional connectivity maps.

Besides, the eigenvectors of PCA can be used to generate “Eigenmaps,” to help visualize important features for the parcellation. The first Eigenmap, which bears most of the variance, plays the most important role in cluster analysis. Eigenmaps also reflected the properties of seed regions. For example, the first Eigenmap of R-SMA showed a high value for limited brain areas including the bilateral SMA and posterior insula (Fig. 5), two areas known to be strongly correlated with the R-SMA[16,18]. In addition, areas with high values in the first Eigenmap of the cingulate cortex and that of the R-PFC were similar and more distributed (Fig. 5). This complicated distribution of these high-value areas reflects important properties of the two regions: both are hetero-modal association regions, taking part in multiple brain networks and functions[35].

APA-based Feature Reduction

The APA-based approach is considerably different from the PCA-based approach because it was not based directly on the raw functional connectivity maps. This approach constructed new whole-brain time-courses by averaging the voxels in the same functional units. Therefore the functional connectivity maps with the APA-based feature-reduction did not contain the same variance as the raw functional connectivity maps. So, the resulting clusters of functional connectivity maps with the APA-based feature-reduction differed from those of Raw (Table 3).

The APA-based approach neutralized noise by averaging voxels, and the noise-resistance of this approach was also better than the PCA-based approach as indicted by the noise-resistance analysis. Meanwhile, the APA-based approach converted voxel-level feature elements to the region level, which improved the independence of features and may contribute to better parcellation results.

However, the APA-based approach required a gross clustering of all voxels within the brain which made it less efficient than the PCA-based approach. From the perspective of computational efficiency, this was not problematic because the APA is a fast algorithm and needs only one repetition. In this experiment, the total computational cost of this gross parcellation was 3.3 h for the brain with a resolution of 3 × 3 × 3 mm^3. This gross parcellation contributed to functional connectivity maps with much less-redundant noise-neutralized seed regions and a much lower computational cost in the seed region parcellation than the raw functional connectivity maps. When both the accuracy and computational efficiency of seed region parcellation were valued, the gross clustering of the whole brain appeared to be a necessary and beneficial step.

The gross parcellation of the APA-based approach reduced ~40000 features into ~2000 features but still gave correct resulting clusters with respect to the ground truth, as shown in the parcellation of the semi-simulated seed region. These results demonstrated that a voxel is not an independent functional unit and functional boundaries exist. Conversely, this characteristic of fMRI data is also the basis for the validity of the APA-based approach.

The functional-connectivity-based parcellation of the brain not only helps to identify distinct functional sub-regions, but may also help to reveal differences in these sub-regions between healthy and psychiatric populations. The abnormal functional connectivity patterns in psychiatric populations may give different parcellations of the brain from the healthy population. Due to its better noise-resistance, the APA-based approach can contribute to more accurate parcellation results compared with the Raw and the PCA-based approach, and thus may help to reveal subtle differences in brain parcellation between different groups of subjects.

Semi-simulated Data

The best way to evaluate a pattern-recognition problem is to compare its results with the ground truth. The ground truth can be defined by constructing simulated data or manually defining it in real data. For example, when we evaluated the different algorithms for the segmentation of brain tissues (grey matter, white matter) on anatomical images, we were able to manually draw the boundaries of the different tissues to build the ground truth. However, it was difficult to define the ground truth in connectivity-based parcellation because it was not known which brain voxels belonged to which clusters. Furthermore, simulating a complex brain connectivity pattern to address this issue was even more difficult. Though recent years have witnessed a growing number of studies related to connectivity-based parcellation (both fMRI and DTI), few studies have attempted to construct simulated data or define the ground truth to evaluate different methods. Here, we attacked this challenge and proposed an approach to building semi-simulated data for connectivity-based parcellation. The semi-simulated data built here reflected the complex brain connectivity pattern, and the ground truth could be clearly defined. We remain optimistic that this approach of building semi-simulated data is also applicable to the DTI connectivity-based parcellation in addition to fMRI data as well as other related problems.

CONCLUSION

The clustering results of the semi-simulated data suggest that functional connectivity maps or the ‘connectional fingerprint’ can provide correct information for cortical parcellation, and feature-reduction does not significantly change the ‘connectional fingerprint’ information. Considering the improvement in computational efficiency (the APA-based approach and the PCA-based approach) and the noise-resistance (the APA-based approach), feature-reduction of functional connectivity maps before cortical parcellation is both feasible and necessary.

SUPPLEMENTAL DATA

Supplemental data include detailed methods for functional connectivity maps by different feature reduction approaches, a whole flowchart of the study, two figures of examples of resulting clusters of the R-SMA and cingulate cortex, and one table of the statistics on the three Eigenmaps. They can be found online at http://www.neurosci.cn/epData.asp?id=90.

ACKNOWLEDGMENTS

We thank The 1000 Functional Connectomes Project for sharing the data. This work was supported by the National Basic Research Development Program (973 Program) of China (2012CBA01304, 2011CB707800), the National High Technology Research and Development Program (863 Program) of China (2012AA020701), the National Natural Science Foundation of China (31271167, 31271168, 81271495, 31070963, 31070965), the Strategic Priority Research Program of the Chinese Academy of Science, China (XDB02020500), and the Development and Reform Project of Yunnan Province, China.

Received date: 2012-06-29; Accepted date: 2012-08-10

REFERENCES

[1] Stephan KE, Kamper L, Bozkurt A, Burns GA, Young MP, Kötter R. Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). Philos Trans R Soc London B Biol Sci 2001, 356: 1159–1186.

[2] Kotter R. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinformatics 2004, 2: 127–144.

[3] Brodmann K. Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenaufbaues. Leipzig: Barth, 1909.

[4] Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain: 3-Dimensional proportional system—An approach to cerebral imaging. New York: Thieme, 1988.

[5] Vogt O. Die myeloarchitektonische Felderung des menschlichen Stirnhirns. J Psychol Neurol 1910, 15: 221–232.

[6] Vogt O. Die Myeloarchitektonik des Isocortex parietalis. J Psychol Neurol 1911, 18: 379–390.

[7] Zilles K, Amunts K. Receptor mapping: architecture of the human cerebral cortex. Curr Opin neurol 2009, 22: 331–339.

[8] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N*, et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002, 15: 273–289.

[9] Roland PE, Zilles K. Structural divisions and functional fields in the human cerebral cortex. Brain Res Brain Res Rev 1998, 26: 87–105.

[10] Passingham RE, Stephan KE, Kotter R. The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 2002, 3: 606–616.

[11] Margulies DS, Kelly AM, Uddin LQ, Biswal BB, Castellanos FX, Milham MP. Mapping the functional connectivity of anterior cingulate cortex. Neuroimage 2007, 37: 579–588.

[12] Kim JH, Lee JM, Jo HJ, Kim SH, Lee JH, Kim ST*, et al.* Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: Functional connectivity-based parcellation method. Neuroimage 2010, 49: 2375–2386.

[13] Barnes KA, Cohen AL, Power JD, Nelson SM, Dosenbach YB, Miezin FM*, et al.* Identifying Basal Ganglia divisions in individuals using resting-state functional connectivity MRI. Front Syst Neurosci 2010, 4: 18.

[14] Nelson SM, Cohen AL, Power JD, Wig GS, Miezin FM, Wheeler ME*, et al.* A parcellation scheme for human left lateral parietal cortex. Neuron 2010, 67: 156–170.

[15] Cauda F, D'Agata F, Sacco K, Duca S, Geminiani G, Vercelli A. Functional connectivity of the insula in the resting brain. Neuroimage 2011, 55: 8–23.

[16] Deen B, Pitskel NB, Pelphrey KA. Three systems of insular functional connectivity identified with cluster analysis. Cereb Cortex 2011, 21: 1498–1506.

[17] Craddock RC, James GA, Holtzheimer PE 3rd, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp 2012, 33(8): 1914–1928.

[18] Zhang S, Ide JS, Li CS. Resting-state functional connectivity of the medial superior frontal cortex. Cereb Cortex 2012,

22(1): 99–111.

[19] Cohen AL, Fair DA, Dosenbach NUF, Miezin FM, Dierker D, Van Essen DC*, et al.* Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage 2008, 41: 45–57.

[20] van den Heuvel M, Mandl R, Pol HH. Normalized cut group clustering of resting-state fMRI data. PLoS One 2008, 3: e2001.

[21] Shen X, Papademetris X, Constable RT. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. Neuroimage 2010, 50: 1027–1035.

[22] Zhang DY, Snyder AZ, Fox MD, Sansbury MW, Shimony JS, Raichle ME. Intrinsic functional relations between human cerebral cortex and thalamus. J Neurophysiol 2008, 100: 1740–1748.

[23] Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using Echo-Planar MRI. Magn Reson Med 1995, 34(4): 537–541.

[24] Margulies DS, Vincent JL, Kelly C, Lohmann G, Uddin LQ, Biswal BB*, et al.* Precuneus shares intrinsic functional architecture in humans and monkeys. Proc Natl Acad Sci U S A 2009, 106: 20069–20074.

[25] Zhang S, Li CS. Functional connectivity mapping of the human precuneus by resting state fMRI. Neuroimage 2012, 59(4): 3548–3562.

[26] Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M*, et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 2011, 106: 1125–1165.

[27] Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA*, et al.* Functional network organization of the human brain. Neuron 2011, 72: 665–678.

[28] Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007, 315: 972–976.

[29] Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933, 24: 417–441.

[30] Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM*, et al.* Toward discovery science of human brain function. Proc Natl Acad Sci U S A 2010, 107: 4734–4739.

[31] Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996, 29: 162–173.

[32] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H*, et al.* Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004, 23: S208–S219.

[33] Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 2005, 102: 9673–9678.

[34] Van Essen DC. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. Neuroimage 2005, 28: 635–662.

[35] Paxinos G, Mai JK. The Human Nervous System (2nd ed.). San Diego: Elsevier Academic Press, 2004.