RESEARCH ARTICLE

Classification of schizophrenia by intersubject correlation in functional connectome

Gong-Jun Ji^{1,2,3} | Xingui Chen^{4,2,3} | Tongjian Bai^{4,2,3} | Lu Wang^{4,2,3} | Qiang Wei^{4,2,3} | Yaxiang Gao^{2,3} | Longxiang Tao⁵ | Kongliang He^{6,7} | Dandan Li^{4,2,3} | Yi Dong^{6,8} | Panpan Hu^{4,2,3} | Fengqiong Yu^{1,2,3} | Chunyan Zhu^{1,2,3} | Yanghua Tian^{4,2,3} | Yongqiang Yu⁵ | Kai Wang⁴

¹Department of Medical Psychology, Chaohu Clinical Medical College, Anhui Medical University, Hefei, China

²Laboratory of Cognitive Neuropsychology, Collaborative Innovation Centre of Neuropsychiatric Disorder and Mental Health, Hefei, China

³Anhui Province Key Laboratory of Cognition and Neuropsychiatric Disorders, Hefei, China

⁴Department of Neurology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

⁵Department of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

⁶Anhui Mental Health Center, Hefei, China

⁷The Fourth People's Hospital of Hefei, Hefei, China

⁸The Second Affiliated Hospital of Anhui Medical University, Hefei, China

Correspondence

Kai Wang, Department of Neurology, The First Affiliated Hospital of Anhui Medical University, Hefei 230000, China.
Email: wangkai1964@126.com

Funding information

the National Natural Science Foundation of China, Grant/Award Number: 91432301, 31571149, 2016YFC1300604, 31771222, 91732303, 81673154, 81571308, 81771456, 81803103; Doctoral Foundation of Anhui Medical University, Grant/Award Number: XJ201532; Youth Top-notch Talent Support Program of Anhui Medical University; Collaborative Innovation Center of Neuropsychiatric Disorder and Mental Health of Anhui Province; the Science Fund for Distinguished Young Scholars of Anhui Province, Grant/Award Number: 1808085J23

Abstract

Functional connectomes have been suggested as fingerprinting for individual identification. Accordingly, we hypothesized that subjects in the same phenotypic group have similar functional connectome features, which could help to discriminate schizophrenia (SCH) patients from healthy controls (HCs) and from depression patients. To this end, we included resting-state functional magnetic resonance imaging data of SCH, depression patients, and HCs from three centers. We first investigated the characteristics of connectome similarity between individuals, and found higher similarity between subjects belonging to the same group (i.e., SCH–SCH) than different groups (i.e., HC–SCH). These findings suggest that the average connectome within group (termed as *group-specific functional connectome* [GFC]) may help in individual classification. Consistently, significant accuracy (75–77%) and area under curve (81–86%) were found in discriminating SCH from HC or depression patients by GFC-based leave-one-out cross-validation. Cross-center classification further suggests a good generalizability of the GFC classification. We additionally included normal aging data (255 young and 242 old subjects with different scanning sequences) to show factors could be improved for better classification performance, and the findings emphasized the importance of increasing sample size but not temporal resolution during scanning. In conclusion, our findings suggest that the average functional connectome across subjects contained group-specific biological features and may be helpful in clinical diagnosis for schizophrenia.

KEYWORDS

classification, functional connectome, functional magnetic resonance imaging, multicenter, resting state, schizophrenia

1 | INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a powerful noninvasive technique to investigate how the human brain works. Primarily

based on group-level analysis, we have accumulated abundant knowledge about the universal principles of brain function (Raichle, 2009). However, it is still undetermined how these advances can be

generalized to solve individualized problems, such as clinical diagnosis (Arbabshirani, Plis, Sui, & Calhoun, 2017; Finn et al., 2017).

Recently, it has been demonstrated that the whole-brain pattern of resting-state functional connectivity (rsFC) can be used to identify an individual from a data pool with an accuracy of up to 99% (Anderson, Ferguson, Lopez-Larson, & Yurgelun-Todd, 2011; Finn et al., 2015; Miranda-Dominguez et al., 2014). This finding was further validated by studies with large sample sizes (Waller et al., 2017) and disease conditions (Kaufmann et al., 2017; Rosenberg et al., 2016). Thus, whole-brain rsFC patterns could be viewed as a connectome fingerprint containing an individual's identity features (Miranda-Dominguez et al., 2014). This reliability feature of rsFC patterns within a subject coexists with its variability across subjects. For instance, canonical correlation analysis indicated that the functional connectome pattern across subjects is associated with specific sets of demographic/psychometric measures (Smith et al., 2015). Based on these intrasubject and intersubject features of rsFC, we hypothesized that subjects with similar psychiatric states may have similar connectome patterns; and inversely, their common connectome features (termed as *group-specific functional connectome* [GFC]) (Gratton et al., 2018) may help in discriminating whether an unknown individual has a similar psychiatric state to that of the cohort study. In support of this hypothesis, Parkinson, Kleinbaum, and Wheatley (2018) found that the similarity of neural responses to audiovisual movies decreased with increasing distance between individuals in their shared social network.

In this study, we used schizophrenia (SCH) as an example to show the potential of GFC-classification in clinical diagnosis. SCH is a major psychiatry disorder, and many MRI-based algorithms have been developed for its automatic diagnosis. However, a current review (Wolfers, Buitelaar, Beckmann, Franke, & Marquand, 2015) and our summary (Table S1, Supporting Information) indicated that most of these classification studies relied on small sample sizes ($n < 60$) in a single center. A major limitation of the within-center design is the undetermined generalizability of the classification models (Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015). Recently, three multicenter studies with large sample sizes considered this issue in SCH classification. The accuracy of cross-center prediction ranged from 70 to 81% (Nieuwenhuis et al., 2012; Rozycki et al., 2017; Zeng et al., 2018). Although the accuracy was significantly higher than that of a random classification, it was still far from clinical application; at the same time, future directions to improve the effects of these classification models was not thoroughly investigated in these studies.

Recently, performance of diagnosis prediction methods was examined with golden standard predefined by conventional clinical approaches (e.g., The Diagnostic and Statistical Manual of Mental Disorders). However, the psychiatry diagnosis system, largely based on self-report symptoms and subjective assessment (Drysdale et al., 2017; Insel et al., 2010), may lead to a high percent of misdiagnosis (Altamura & Goikolea, 2008; Hirschfeld, Lewis, & Vornik, 2003). On the contrary, there is no doubt that young (<40 years) and old (>50 years) subjects belong to different age groups. Thus, aging may provide a model with golden standard to estimate novel method of classification (Garrett, Kovacevic, McIntosh, & Grady, 2010). Additionally, the abundant public-available datasets for subjects in different age groups (Nooner et al., 2012; Wei et al., 2018) also enable us to

investigate factors (e.g., sample size and scanning sequence) that may affect classification performance, and give implications to further improve classification performance.

Based on resting-state fMRI (rs-fMRI) datasets from multicenters, this study aimed to estimate the performance of GFC-based classification in discriminating SCH patients from HCs and from depression patients. We also investigated the neural mechanism underlying this classification and included aging datasets (young vs. old subjects) to reveal factors important for improving classification performance.

2 | MATERIAL AND METHODS

2.1 | Experiment design and subjects

This study included rs-fMRI datasets from five centers: Anhui Medical University Site 1 (AMU1), AMU Site 2 (AMU2), Southwest University (SWU) (Wei et al., 2018), Nathan Kline Institute (NKI) (Nooner et al., 2012), and the Center of Biomedical Research Excellence (COBRE) (Cetin et al., 2014). This study was approved by the AMU Ethics Committee (see Appendix S1, Supporting Information for data screening). All subjects from AMU were informed of the format of the study and gave written consent. The study design and analyses on each dataset were illustrated in Figure 1.

Patients with mental disorders and healthy controls (HCs) were included from AMU1 (85 HC, 46 major depression patients, and 66 SCH patients), AMU2 (38 HC and 56 SCH patients) and COBRE (62 HC and 62 SCH patients) (Table 1). Depression data were included to test the differential diagnosis ability of GFC-based classification for SCH. Thus, we performed discriminating analysis between SCH and depression patients but not between depression patients and HCs.

FIGURE 1 Study design and analysis. Multicenter SCH and aging data were included to characterize the connectome similarity between subjects and groups, as well as the performance of connectome-based classification (Analysis 1). Particularly, the SWU and NKI data were used to test the influence of sample size and scanning sequence on classification performance (Analysis 2). The findings provide important points for future studies to improve the classification performance. AMU = Anhui Medical University; COBRE = Center of Biomedical Research Excellence; GFC = group-specific functional connectome; HC = healthy control; MD = major depression; NKI = Nathan Kline institute; SCH = schizophrenia; SWU = Southwest University [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Demographic and clinical characteristics of patients and matched controls

Characteristic	AMU1		AMU2	
	SCH vs. HC	SCH vs. MD	SCH vs. HC	COBRE
Sample size	66, 85	46, 46	56, 38	62, 62
Age (yrs) ^a	33.4 (7.2), 33.3 (9.7)	34.4 (7.8), 35.7 (9.2)	24.9 (6.2), 27.3 (7.1)	39.4 (14.0), 38.4 (11.6)
Sex (m/f)	27/39, 44/41	16/30, 16/30	28/28, 14/24	51/11, 51/11
Education (yrs) ^a	10.2 (3.1), 10.8 (3.9)	9.8 (3.0), 9.3(4.2)	11.9 (2.8), 11.6 (2.2)	4.3 (1.3), 4.0 (1.4) ^b
Duration (yrs) ^a	7.2 (6.4), NA	7.3 (6.7), 2.7(3.2)	4.4 (3.6), NA	16.3 (13.2) ^c , NA
BPRS ^a	41.3 (10.6), NA	43.4 (10.8), NA	NA, NA	NA, NA
PANSS total ^a	NA, NA	NA, NA	45.8 (8.8), NA	58.4 (5.0), NA
PANSS positive ^a	NA, NA	NA, NA	11.1 (3.4), NA	14.8 (5.6), NA
PANSS negative ^a	NA, NA	NA, NA	10.1 (3.3), NA	15.2 (15.7), NA
TR (ms)	2,000		2,000	2,000
Total volumes	240		240	150
Voxel size (mm)	3.44 × 3.44 × 4.6		3.75 × 3.75 × 4.0	3.75 × 3.75 × 4.55

BPRS = The Brief Psychiatric Rating Scale; f = female; HC = healthy control; m = male; MD = major depression patient; NA = not available; PANSS = positive and negative syndrome scale; SCH = schizophrenia patient; TR = time of repetition; yrs = years.

^a Data are means, with SE in parentheses.

^b The education levels in the COBRE sample were classified into eight grades. The average scores, 4.3 and 4.0, represent education at the college level.

^c The durations of the COBRE sample were calculated by subtracting the "age at first psychiatric illness" from the "age at the time of experiment."

For the normal aging dataset, young (20–40 years of age) and old (50–70 years of age) healthy subjects were included from SWU (187 young and 184 old adults) and NKI (68 young and 58 old adults; Table 2).

SCH patients from AMU met the following inclusion criteria: (a) diagnosis of SCH using the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV); (b) on psychotropic medication at steady dosages for at least 8 weeks prior to the study; and (c) a verbal intelligence quotient >85, as measured by a Chinese version of the National Adult Reading Test. Exclusion criteria were as follows: (a) history of significant head trauma or neurological disorders; (b) alcohol or drug abuse; (c) focal brain lesions on T1- or T2-weighted fluid-attenuated inversion-recovery MRIs; (d) recent aggression or other forms of behavioral dysfunction; (e) head motion exceeding 3 mm in translation or 3° in rotation during resting state fMRI scanning; and (7) Hamilton Anxiety Rating Scale or Hamilton Depression Rating Scale Score >7.

Depression patients from AMU were initially recruited for understanding the mechanism of electroconvulsive therapy. Thus, the recruitment criteria were designed for electroconvulsive therapy. All patients met the diagnoses criteria of depression in DSM-IV and showed resistance to drug therapy or a severe suicidal tendency. We

excluded patients with substance dependence, pregnancy, life threatening somatic disease, neurological disorders, or other comorbid mental disorders or MRI-related contraindications (Wei et al., 2014). All the depression patients received electroconvulsive therapy after the MRI data acquisition, and part of the samples have been reported in our previous work (Wei et al., 2014). In this study, this depression data were included for discriminative diagnosis between SCH and depression. The criteria of subject recruitment of COBRE can be found at <http://cobre.mrn.org/phase1/recruitment/index.html>.

2.2 | Image acquisition and preprocessing

All rs-fMRI datasets were collected in 3.0T MRI scanners in which the participants were awake in the scanner, not performing a task, using standard protocols described in Appendix S1, Supporting Information.

fMRI datasets were preprocessed by Data Processing & Analysis for Brain Imaging (DPABI) (Yan, Wang, Zuo, & Zang, 2016), which synthesizes procedures in SPM12 software (www.fil.ion.ucl.ac.uk/spm). The first five images were excluded to ensure steady-state longitudinal magnetization, and the remaining images were corrected for temporal differences and head motion. Images were then normalized to standard Montreal Neurological Institute space (voxel size = $3 \times 3 \times 3$ mm³) and were spatially smoothed (4 mm full-width at half-maximum). To remove physiological noise, we regressed out nuisance variables including 24 parameters obtained by rigid body head motion correction (Friston, Williams, Howard, Frackowiak, & Turner, 1996; Yan et al., 2013) and three average signals from white matter, cerebrospinal fluid, and whole brain area, respectively. To keep consistent with the processing of Finn et al. (2015) and our previous work (Ji et al., 2017; Ji, Liao, Chen, Zhang, & Wang, 2017; Ji, Yu, Liao, & Wang, 2017), we did not remove images according to head motion parameters.

To construct functional connectome, we subdivided the whole brain into 268 regions of interest (ROIs) using the same template as Finn et al. (2015), and performed Pearson's correlation among each

TABLE 2 Demographic characteristics and scanning parameters of healthy participants

Characteristic	SWU sample (young, old)		NKI sample (young, old)	
Age (yrs) ^a	26.3 (5.3), 58.6 (4.7)		25.6 (5.6), 61.5 (6.1)	
Sex (m/f)	76/111, 73/111		34/34, 29/29	
TR (ms)	2,000	645	1,400	2,500
Total volumes	242	900	404	120
Voxel size (mm)	3.4 × 3.4 × 3.0	3.0 × 3.0 × 3.0	2.0 × 2.0 × 2.0	3.0 × 3.0 × 3.0

^a Data are means, with SE in parentheses.

paired ROIs. Finally, unique elements ($268 \times 267/2 = 35,778$) of the correlation matrix were transformed by Fisher's z-transformation for further analysis. Notably, most of the samples from SWU did not cover the whole cerebellum, thus only 216 of the 268 ROIs covered in all subjects during scanning were included to construct correlation matrix.

2.3 | Group-level functional connectome

We defined GFC as the average functional connectome across subjects in the same phenotypic group. This is consistent with current findings that functional brain networks are dominated by stable group and individual factors, not by cognitive or daily variations (Gratton et al., 2018). We then performed the following computations and statistical analyses.

2.3.1 | Intersubject similarity of the connectome

We computed the Pearson's correlation (Fisher's z-transformation) between the functional connectome from each paired subject (Figure 2). Mean correlations of intergroup and intragroup were

compared by two-sample *t* tests. For SCH data, the intergroup condition indicated correlations between SCH patients and HCs, while the intragroup condition indicated correlations between two SCH patients or two HCs. For normal aging data, the intergroup condition indicated correlations between young and old subjects, while the intragroup condition indicated correlations between two young subjects or two old subjects.

2.3.2 | Stability of GFC

To test whether the GFC would be more reliable and robust as the sample size increased, a "group identification" analysis for each center was used, as described by Finn et al. (2015). For convenience, we took the SWU sample as an example to describe the analysis. Specifically: (a) *N* subjects were randomly drawn from the young group 499 times and the GFC was calculated each time, which constituted part of a database pool; (b) two independent subgroups, each with *N* subjects, were randomly drawn from the old group. The GFC of one group was added in the database pool (500 GFCs in total) as "target data" and the other group was used

FIGURE 2 Intersubject/group similarity of the functional connectome. (a) Functional connectome similarities were assessed between each of paired subjects. The similarities were then categorized into intergroup (e.g., young–old) and intragroup (e.g., old–old and young–young) conditions. The latter condition showed higher similarity compared with the former across all five centers ($***p < 0.001$, $**p < 0.01$). (b) Taking the normal aging as an example, 499 subgroups with the same sample size (*N*) were randomly selected from the young group and two independent subgroups (target and test group) were separated from the old group. The same procedure was performed for schizophrenia and healthy control groups. The line graph indicated that group identification accuracy increased with the sample size in all seven datasets [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Classification performance within and between centers

Training data	Test data (specificity/sensitivity/accuracy/AUC)			
Young vs. old subjects				
	SWU	NKI ₆₄₅	NKI ₁₄₀₀	NKI ₂₅₀₀
SWU	87/83/85/92	99/24/64/80	91/33/55/72	91/71/82/86
NKI ₆₄₅	88/36/63/74	84/76/80/89	–	–
NKI ₁₄₀₀	90/27/59/69	–	81/76/79/88	–
NKI ₂₅₀₀	96/33/65/80	–	–	90/74/83/87
SCH patients vs. healthy controls				
	AMU1	AMU2	COBRE	
AMU1	77/78/78/84	80/63/73/81	69/86/77/83	
AMU2	76/71/73/82	68/80/76/86	60/69/65/74	
COBRE	65/64/64/69	70/66/68/72	71/82/77/81	

Bold numbers indicate within center performance. AMU1 = Anhui Medical University Site 1; AUC = area under curve; COBRE = Center of Biomedical Research Excellence; HC = healthy control; SCH = schizophrenia patient.

as “test data”; (c) we correlated the test data with each component in the database pool. The “group identification” was successful if the test data showing higher correlation coefficient with the target group than the others in the database pool. (d) we repeated Step “(b)” 500 times and the identification accuracy was defined as the number of successful identifications divided by 500; and (e) finally, identification accuracy under different sample sizes was obtained

by repeating the above steps. The sample size was initially set at two and increased one subject each time. The upper limit of the sample size was less than half the sample size of the old group because: (a) the number of subjects should be equal between “test data” and “target data”; and (b) each subject could only be assigned to one of the two groups (test or target).

2.3.3 | Individual classification

We determined the classification ability of GFC on SCH and aging data by leave-one-out as well as cross-center validation. In each iteration of the leave-one-out cross-validation, one subject (e.g., an old subject) would be left out as a test subject. Pearson's correlation coefficient between test data and the two GFCs (e.g., belonging to the young and old group, respectively) then determined to which group (the one with larger coefficient) the test subject was assigned. Sensitivity, specificity, accuracy, and area under curve (AUC) of the receiver operating characteristic (ROC) were reported for indicating classification performance. Sensitivity and specificity referred to the percentage of cases that the GFC method correctly identified old (or SCH) and young (or HCs) subjects, respectively. Accuracy referred to the total proportion of samples correctly classified. The ROC curves provided information regarding the

FIGURE 3 Classification performance of discriminating old from young adults. High AUCs were observed both within the SWU (a) and the NKI (b) datasets. Classification accuracy could be improved by increasing sample size (c). Scanning sequence had a significant effect on cross-center (d, e) but not within-center classification (b). Cross-center classification indicated high AUC (up to 86%) when the two datasets were acquired by a similar scanning sequence (NKI₂₅₀₀ and SWU) [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 4 Classification performance within and across centers. Within all centers (AMU1, AMU2, and COBRE), the GFC-based method could discriminate schizophrenia (SCH) patients from healthy controls with an AUC > 0.80 (a). Particularly, a high AUC (0.81) was also found when discriminating SCH from major depression patients within the AMU1 center (a). Cross-center classification shows a high AUC when using the (b) AMU1 and (c) AMU2 as training data but shows a relatively low AUC when using COBRE as training data (d) [Color figure can be viewed at wileyonlinelibrary.com]

balance between sensitivity and the false positive rate (1-specificity) across a range of decision thresholds.

Cross-center validation was performed by predicting old (or SCH) in one center based on the GFC information of another center. Our results are primarily based on accuracy and AUC, both of which were analyzed for significance by permutation tests. Specifically, the group labels of each subject in the training data were randomly permuted 1,000 times and based on the distribution of accuracy and the AUC, the discriminative performance in real conditions were determined.

2.3.4 | Edgewise contribution for classification

When performing subject classification, the Pearson's correlation coefficients were computed between the connectivity pattern of a

single subject (i.e., SCH patient) and the GFC from two groups (i.e., HC and SCH groups). The subject was then sorted to the group that had the largest correlation coefficient. Computationally, the Pearson's correlation of two vectors is the sum of element-wise products, given that the two vectors are z-score normalized (0 mean with unit SD). Thus, the classification score (CS) of each edge could be computed as the $R_i \times R_c - R_i \times R_u$, where R_i , R_c , and R_u were the normalized correlation coefficient of one edge in subject i , the group with (R_c) or without (R_u) similar phenotype as subject i , respectively. The positive CS values were summed across classifications to indicate the total contribution of each edge. Negative values were excluded because of their negative contribution to correct classification. Finally, element-wise CS was illustrated within and between networks, as defined in previous studies (Finn et al., 2015; Shen, Tokoglu,

FIGURE 5 Edgewise contribution for classification. Numbers 1–6 indicate the medial frontal, frontoparietal, default mode, subcortical–cerebellar, motor, and visual network, respectively. The matrix shows edgewise contribution for each classification, while the circular diagram indicates the top contributors within (colored lines) and between (gray lines) networks. The circular diagram was designed using DynamicBC software, <http://restfmri.net/forum/DynamicBC> (Liao et al., 2014) [Color figure can be viewed at wileyonlinelibrary.com]

Papademetris, & Constable, 2013): specifically, the medial frontal, frontoparietal, default mode, subcortical–cerebellar, motor, and visual networks. The visual network included three visual components from the work of Finn et al. (2015). The SC for the psychometric model was first computed for the AMU1, AMU2, and COBRE datasets, respectively, and was then averaged across centers. The same processes were performed in the SWU and NKI₂₅₀₀ datasets.

2.3.5 | Factors affecting individual classification

Sample size

Using the SWU sample, we tested whether the classification performance could be improved by increasing the sample size. The initial sample size was the least size corresponding to 100% accuracy in “group identification” (i.e., 50), after which the sample size was increased by five subjects in each step until the total number was equal to the total number of old subjects (i.e., 184). In each step, N subjects were randomly drawn from each group 100 times with a guarantee that no significant differences existed in gender between the groups. Finally, the average accuracy and the AUC were reported as measures of classification performance.

MRI scanning sequence

Using the NKI database, we included subjects who took part in all fMRI experiments using three different protocols. According to the “Time of Repetition,” they were termed as protocol NKI₆₄₅, NKI₁₄₀₀, and NKI₂₅₀₀, respectively. The performance of GFC-based classification was estimated by these datasets independently. Additionally, cross-center validation was also performed using the SWU as training data and the NKI as test data. The classification accuracy and the AUC were estimated by permutation tests.

Data processing parameter

Two analyses were performed to test the effect of frame-wise head motion and global brain signal regression on the classification of SCH patients, respectively. In the first analysis, we kept all processing steps except global signal regression. In the second analysis, we performed volume censoring (“scrubbing”) according to frame-wise displacement (FD) (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012). Specifically, image frames with a large FD (>0.5 mm) were identified as bad time points (Yan et al., 2013); along with the bad time points, the one preceding, and the two following points, were deleted to assure exclusion of confounding motion-related bias. Subjects were excluded if the remaining data was less than 100 volumes.

3 | RESULTS

3.1 | Demographics and clinical characteristics

Chi-square tests indicated similar male/female ratios between young and old groups in both SWU ($\chi^2 = 0.04$, $p = 0.85$) and NKI ($\chi^2 = 0$, $p = 1$) samples. No significant age, education, or male/female ratio difference existed between SCH patients and HCs in AMU1 ($t = 0.02$, $p = 0.99$; $t = 1.00$, $p = 0.32$; $\chi^2 = 1.76$, $p = 0.18$), AMU2 ($t = 1.68$, $p = 0.10$; $t = 0.54$, $p = 0.59$; $\chi^2 = 1.59$, $p = 0.21$), or COBRE ($t = 0.41$, $p = 0.68$; $\chi^2 = 0$, $p = 1$). In AMU1, 46 SCH patients and 46 major depression patients were matched for differential diagnosis. No significant age ($t = 0.74$, $p = 0.46$) or male/female ratio ($\chi^2 = 0$, $p = 1$) differences existed between the SCH and depression groups, but disease duration was higher ($t = 4.28$, $p < 0.01$) in the SCH group.

3.2 | Inter-subject similarity of the connectome

The connectome similarity between intergroup subjects (e.g., old-young) and intragroup (e.g., old-old) subjects was compared in the SWU, NKI, AMU1, AMU2, and COBRE samples. All analyses showed significantly higher correlation in intragroup than intergroup conditions (Table S2, Supporting Information and Figure 2a).

3.3 | Stability of the GFC

Overall, we found the accuracy of group identification increased with sample size in an "S" shape (Figure 2b). Within the SWU data, the old group could be identified with 100% accuracy if the sample size was increased to 50 subjects. The highest accuracy of the AMU1, AMU2, and COBRE datasets were 98, 99, and 90%, when the sample size reached 33, 26, and 28 respectively. The curves from the three NKI datasets were quite similar. All accuracies increased to 96% when 29 subjects were in the test/target groups.

3.4 | Diagnosing individuals with SCH

GFC-based classification differentiated SCH patients from HCs in AMU1 datasets with 78% accuracy and with 84% AUC; in AMU2 datasets with 76% accuracy and 86% AUC; and in COBRE datasets with 77% accuracy and 81% AUC (Table 3, all $p < 0.001$; Figure 3a). Additionally, the GFC discriminated SCH from depression patients in the AMU1 dataset with 76% specificity, 74% sensitivity, 75% accuracy, and 81% AUC (all $p < 0.005$, Figure 3a). Cross-center classification showed that the GFC information from AMU1 SCH patients (with a large sample size) could correctly predict for AMU2 SCH patients (accuracy = 73%, AUC = 81%) and for COBRE SCH patients (accuracy = 77%, AUC = 83%), as the within-center performances (all $p < 0.001$). Full cross-center results were presented in Table 3 (Figure 3b–d).

3.5 | Discriminating old subjects from young adults

Based on the GFC information, old individuals could be discriminated from the SWU dataset with significant accuracy (85%, $p < 0.001$) and AUC (92%, $p < 0.001$, Figure 4a). In the NKI datasets, the classification accuracy and the AUC were 80 and 89%, respectively, for NKI₆₄₅; 79 and 88%, respectively, for NKI₁₄₀₀; and 83 and 87%, respectively for NKI₂₅₀₀ (Figure 4b, all $p < 0.001$).

We recomputed the functional connectome for the NKI dataset using 216 nodes from the SWU dataset, then performed cross-center prediction for old subjects. Based on the GFC information of the SWU data, the classification accuracy and the AUC were 64 and 80%, respectively, for NKI₆₄₅ (both $p < 0.01$); 55 and 72%, respectively, for NKI₁₄₀₀ (both $p < 0.01$); and 82 and 86%, respectively, for NKI₂₅₀₀ (both $p < 0.001$). Inversely, NKI datasets also showed significant predictions for the SWU dataset in both accuracy and the AUC (Table 3).

3.6 | Edgewise contribution for classification

In the demographic model (young vs. old), edgewise contribution was the average SC from NKI and SWU datasets (Figure 5a). In the psychometric model (SCH vs. HC), edgewise contribution was the

average SC from the AMU1, AMU2, and COBRE datasets (Figure 5b). To show the edges with high contribution in classification, the highest 99.9 percentile of SC are illustrated in Figure 5. The highest contributions in the aging data mainly belonged to the subcortical–cerebellar, motor, and visual networks, while those in the SCH data were distributed across all six networks.

3.7 | Factors affecting individual classification

Sample size

Using the SWU dataset, we found that the average accuracy and AUC increased from 80 to 85% and from 88 to 92%, respectively, as the sample size increased from 50 (corresponding to 100% accuracy of group identification) to 180 (Figure 4c).

MRI scanning sequence

DeLong's test (DeLong, DeLong, & Clarke-Pearson, 1988) did not show a significant AUC difference between each of the paired NKI datasets ($p = 0.36$ for NKI₁₄₀₀ vs. NKI₂₅₀₀; $p = 0.78$ for NKI₆₄₅ vs. NKI₂₅₀₀; $p = 0.51$ for NKI₁₄₀₀ vs. NKI₆₄₅), but cross-center validation (Figure 4d) indicated that the SWU-based prediction showed a larger AUC for NKI₂₅₀₀ than NKI₁₄₀₀ data ($p < 0.001$). The NKI₂₅₀₀-based prediction for the SWU data showed a larger AUC than NKI₁₄₀₀-based prediction ($p < 0.001$). Among the three NKI datasets, only the NKI₂₅₀₀ dataset showed balanced specificity and sensitivity (Table 3).

Data processing parameter

After data scrubbing, one healthy subject in the AMU1 dataset and 73 subjects in the COBRE dataset were excluded because of short time series (<100 time points). No subject was excluded in the AMU2 dataset. Classification of these datasets indicated similar accuracy and AUC with the no-scrubbing findings (Table S3, Supporting Information). In the second analysis, we did not perform global signal regression during preprocessing. The findings were consistent with our main report with regards to accuracy and AUC (Table S3, Supporting Information).

4 | DISCUSSION

In this study, we characterized the functional connectome across subjects by analyzing rs-fMRI data from five centers. We found higher similarity between subjects belonging to the same group (e.g., two SCH patients) than different groups (e.g., patient and HC). These findings suggest that the GFC may help in individual classification. Consistently, significant accuracy (75–85%) and AUC (81–92%) were found in discriminating SCH patients and old subjects from their respective controls by GFC classification. Additional analysis on aging data indicated this performance may be further improved by increasing sample size. Consistently, good performance (accuracy = 73–78%, AUC = 81–84%) was found in the classification of SCH patients when datasets with a large sample size (i.e., AMU1) were used to predict datasets with a small sample size (i.e., AMU2 and COBRE). These findings suggested that GFC information on large sample size may be used to cross-center classification.

In the work of Finn et al. (2015), functional connectomes were demonstrated to be a reliable fingerprint of a given subject. Here, we generalized this concept to intersubject conditions and group level. Intersubject correlation analysis indicated that the connectome pattern was more similar between "patient-patient" (or "old-old") subjects than "patient-control" (or "old-young") subjects. This suggested that subjects with similar behavior features also had similar connectivity signatures in the brain. This was consistent to previous intersubject connectivity analysis using task fMRI (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Simony et al., 2016). For example, a simple movie stimulus significantly improved the functional correlations between subjects (Hasson et al., 2004), suggesting brain function could be temporally synchronized across subjects by the same stimulus. In an rs-fMRI paradigm, no particular task was required for subjects. These findings may reflect a long-term effect of life experience on brain function. Subjects in a similar environment may develop similar behavior as well as brain functional features. The "group identification" analysis further indicated that the GFC was more reliable as the sample size increased. For example, the normal aging effects on brain functions could be reliably captured by at least 50 subjects in both the young and old groups. All the above findings indicated that the average functional connectome across subjects could be a neural signature of a common behavior within the group.

SCH is a disorder to which pattern recognition methods have been commonly applied (Wolfers et al., 2015). However, most of these studies had small sample sizes or relied on dataset from a single center. The potential bias of overfitting in the classification algorithm therefore cannot be excluded (Gabrieli et al., 2015). Using a relatively large sample size from the AMU1, the GFC-based method showed 78% diagnostic accuracy and 84% AUC, which was validated in the AMU2 and COBRE datasets independently. These percentages of performance were comparable to studies using independent datasets for cross-validation (Nieuwenhuis et al., 2012; Ota et al., 2012). More importantly, we found that the GFC information from the AMU1 datasets could accurately classify SCH patients in both the AMU2 and COBRE datasets. The classification performance decreased if COBRE was used to predict the other datasets. This is consistent with our findings of sample-size effect analysis and also to a previous report suggesting "larger sample size, better classification accuracy" (Nieuwenhuis et al., 2012). Taken together, these data indicated that the average functional connectome was a reliable group-specific feature, and implicated high generalizability of the GFC method in cross-center prediction.

In the aging data, high accuracy and AUC were found in both the SWU and NKI datasets when discriminating old from young subjects. The ROC in the three NKI datasets showed similar trajectories suggesting that the multiband sequence did not improve the performance of GFC-based classification significantly as compared to the conventional echo-planar imaging sequence. Analysis on the SWU dataset indicated that classification performance could be further improved if sample size increased. These within-center classifications show similar accuracy as previous investigations (Meier et al., 2012; Vergun et al., 2013). More importantly, the performance of the GFC-classifier remained in the cross-center validation. The GFC information of the SWU dataset accurately classified old subjects in the NKI dataset. Different from our expectation, the best prediction was found in the NKI₂₅₀₀ dataset,

rather than in the NKI₁₄₀₀ or NKI₆₄₅ datasets. This was likely due to our training data (from the SWU dataset), which was acquired by conventional rather than multiband sequence. These results suggest future multicenter studies are needed to acquire data using similar sequences.

There were some limitations to our study. First, although the intersubject similarity was higher in intragroup rather than intergroup conditions, the effect size was relatively small. Future studies may adopt an individualized functional atlas rather than group parcellation for cross-subject alignment (Gordon et al., 2017; Wang et al., 2015). As a result, the connectome matrix built on functionally homogenous ROIs may have low variability of intersubject similarity and thus increase the effect size. Second, although the GFC classification was cross-validated in different centers, the accuracy was far from excellent. According to our findings and to current advances (Horien et al., 2018; Noble et al., 2017), increasing sample size, scanning time/sessions, and controlling head motion may further increase classification performance. Third, besides functional connectome, whole-brain structural connectivity has also been demonstrated as a unique feature to each subject (Yeh et al., 2016). Thus, it would be interesting to combine functional and structural connectomes in individual classification.

5 | CONCLUSIONS

In this study, the intrasubject features of functional connectome (Anderson et al., 2011; Finn et al., 2015; Miranda-Dominguez et al., 2014) were generalized to intersubject/group situations. We found that the GFC-based classification showed good performance in discriminating SCH patients from HCs within three centers. Importantly, this performance was well reproduced in cross-center validations when the GFC was from large sample size. Thus, these findings suggested that the average functional connectome across subjects contains group-specific biological features and which may be helpful in clinical diagnosis for SCH.

ACKNOWLEDGMENTS

This study was funded by the National Natural Science Foundation of China (91432301, 31571149, 2016YFC1300604 and to K.W.; 81771456 to C.Z.; 81673154, 91732303 to Y.T.; 31771222 to F.Y.); Doctoral Foundation of Anhui Medical University (XJ201532 to G.J.J.); Youth Top-notch Talent Support Program of Anhui Medical University (to G.J.J.); the Science Fund for Distinguished Young Scholars of Anhui Province (1808085J23 to Y.T.) and Collaborative Innovation Center of Neuropsychiatric Disorder and Mental Health of Anhui Province. COBRE data were downloaded from the Collaborative Informatics and Neuroimaging Suite Data Exchange tool (COINS; <http://coins.mrn.org/dx>) and data collection was performed at the Mind Research Network, which was funded by a COBRE grant 5P20RR021938/P20GM103472 from the NIH to Dr. Vince Calhoun.

ORCID

Gong-Jun Ji <https://orcid.org/0000-0002-7073-5534>

Tongjian Bai <https://orcid.org/0000-0002-2185-0042>

REFERENCES

- Altamura, A. C., & Goikolea, J. M. (2008). Differential diagnoses and management strategies in patients with schizophrenia and bipolar disorder. *Neuropsychiatric Disease and Treatment*, 4, 311–317.
- Anderson, J. S., Ferguson, M. A., Lopez-Larson, M., & Yurgelun-Todd, D. (2011). Reproducibility of single-subject functional connectivity measurements. *AJNR. American Journal of Neuroradiology*, 32, 548–555.
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, 145, 137–165.
- Cetin, M. S., Christensen, F., Abbott, C. C., Stephen, J. M., Mayer, A. R., Canive, J. M., ... Calhoun, V. D. (2014). Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *NeuroImage*, 97, 117–126.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845.
- Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., ... Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, 23, 28–38.
- Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, 160, 140–151.
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18, 1664–1671.
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35, 346–355.
- Gabrieli, J. D., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, 85, 11–26.
- Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2010). Blood oxygen level-dependent signal variability is more than just noise. *The Journal of Neuroscience*, 30, 4914–4921.
- Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., ... Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, 95, 791–807 e7.
- Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., ... Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, 98, 439–452 e5.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303, 1634–1640.
- Hirschfeld, R. M., Lewis, L., & Vornik, L. A. (2003). Perceptions and impact of bipolar disorder: How far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *The Journal of Clinical Psychiatry*, 64, 161–174.
- Horien, C., Noble, S., Finn, E. S., Shen, X., Scheinost, D., & Constable, R. T. (2018). Considering factors affecting the connectome-based identification process: Comment on Waller et al. *NeuroImage*, 169, 172–175.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, 167, 748–751.
- Ji, G. J., Liao, W., Chen, F. F., Zhang, L., & Wang, K. (2017). Low-frequency blood oxygen level-dependent fluctuations in the brain white matter: More than just noise. *Science Bulletin*, 62, 656–657.
- Ji, G. J., Yu, F., Liao, W., & Wang, K. (2017). Dynamic aftereffects in supplementary motor network following inhibitory transcranial magnetic stimulation protocols. *NeuroImage*, 149, 285–294.
- Ji, G. J., Yu, Y., Miao, H. H., Wang, Z. J., Tang, Y. L., & Liao, W. (2017). Decreased network efficiency in benign epilepsy with centrotemporal spikes. *Radiology*, 1, 186–194.
- Kaufmann, T., Alnaes, D., Doan, N. T., Brandt, C. L., Andreassen, O. A., & Westlye, L. T. (2017). Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nature Neuroscience*, 20, 513–515.
- Liao, W., Wu, G., Xu, Q., Ji, G. J., Zhang, Z., Zang, Y. F., & Lu, G. (2014). DynamicBC: A MATLAB toolbox for dynamic brain connectome analysis. *Brain Connectivity*, 4, 780–790.
- Meier, T. B., Desphande, A. S., Vergun, S., Nair, V. A., Song, J., Biswal, B. B., ... Prabhakaran, V. (2012). Support vector machine classification and characterization of age-related reorganization of functional brain networks. *NeuroImage*, 60, 601–613.
- Miranda-Dominguez, O., Mills, B. D., Carpenter, S. D., Grant, K. A., Kroenke, C. D., Nigg, J. T., & Fair, D. A. (2014). Connectotyping: Model based fingerprinting of the functional connectome. *PLoS One*, 9, e111048.
- Nieuwenhuis, M., van Haren, N. E., Hulshoff Pol, H. E., Cahn, W., Kahn, R. S., & Schnack, H. G. (2012). Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *NeuroImage*, 61, 606–612.
- Noble, S., Spann, M. N., Tokoglu, F., Shen, X., Constable, R. T., & Scheinost, D. (2017). Influences on the test-retest reliability of functional connectivity MRI and its relationship with behavioral utility. *Cerebral Cortex*, 27, 5415–5429.
- Nooner, K. B., Colcombe, S. J., Tobe, R. H., Mennes, M., Benedict, M. M., Moreno, A. L., ... Milham, M. P. (2012). The NKI-Rockland sample: A model for accelerating the pace of discovery science in psychiatry. *Frontiers in Neuroscience*, 6, 152.
- Ota, M., Sato, N., Ishikawa, M., Hori, H., Sasayama, D., Hattori, K., ... Kunugi, H. (2012). Discrimination of female schizophrenia patients from healthy women using multiple structural brain measures obtained with voxel-based morphometry. *Psychiatry and Clinical Neurosciences*, 66, 611–617.
- Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, 9, 332.
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59, 2142–2154.
- Raichle, M. E. (2009). A paradigm shift in functional brain imaging. *The Journal of Neuroscience*, 29, 12729–12734.
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, 19, 165–171.
- Rozycki, M., Satterthwaite, T. D., Koutsouleris, N., Erus, G., Doshi, J., Wolf, D. H., ... Davatzikos, C. (2017). Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. *Schizophrenia Bulletin*, 4, 1035–1044.
- Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Group-wise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82, 403–415.
- Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, 7, 12141.
- Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E., Glasser, M. F., ... Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, 18, 1565–1567.
- Vergun, S., Deshpande, A. S., Meier, T. B., Song, J., Tudorascu, D. L., Nair, V. A., ... Prabhakaran, V. (2013). Characterizing functional connectivity differences in aging adults using machine learning on resting state fMRI data. *Frontiers in Computational Neuroscience*, 7, 38.
- Waller, L., Walter, H., Kruschwitz, J. D., Reuter, L., Muller, S., Erk, S., & Veer, I. M. (2017). Evaluating the replicability, specificity, and generalizability of connectome fingerprints. *NeuroImage*, 158, 371–377.
- Wang, D., Buckner, R. L., Fox, M. D., Holt, D. J., Holmes, A. J., Stoecklein, S., ... Liu, H. (2015). Parcellating cortical functional networks in individuals. *Nature Neuroscience*, 18, 1853–1860.
- Wei, D., Zhuang, K., Ai, L., Chen, Q., Yang, W., Liu, W., ... Qiu, J. (2018). Structural and functional brain scans from the cross-sectional South-west University adult lifespan dataset. *Scientific Data*, 5, 180134.

- Wei, Q., Tian, Y., Yu, Y., Zhang, F., Hu, X., Dong, Y., ... Wang, K. (2014). Modulation of interhemispheric functional coordination in electroconvulsive therapy for depression. *Translational Psychiatry*, 4, e453.
- Wolters, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., & Marquand, A. F. (2015). From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience and Biobehavioral Reviews*, 57, 328–349.
- Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., ... Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76, 183–201.
- Yan, C. G., Wang, X. D., Zuo, X. N., & Zang, Y. F. (2016). DPABI: Data processing & analysis for (resting-state) brain imaging. *Neuroinformatics*, 14, 339–351.
- Yeh, F. C., Vettel, J. M., Singh, A., Poczos, B., Grafton, S. T., Erickson, K. I., ... Verstynen, T. D. (2016). Quantifying differences and similarities in whole-brain white matter architecture using local connectome fingerprints. *PLoS Computational Biology*, 12, e1005203.
- Zeng, L. L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., ... Hu, D. (2018). Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. *EBioMedicine*, 30, 74–85.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ji G-J, Chen X, Bai T, et al. Classification of schizophrenia by intersubject correlation in functional connectome. *Hum Brain Mapp.* 2019;40:2347–2357. <https://doi.org/10.1002/hbm.24527>