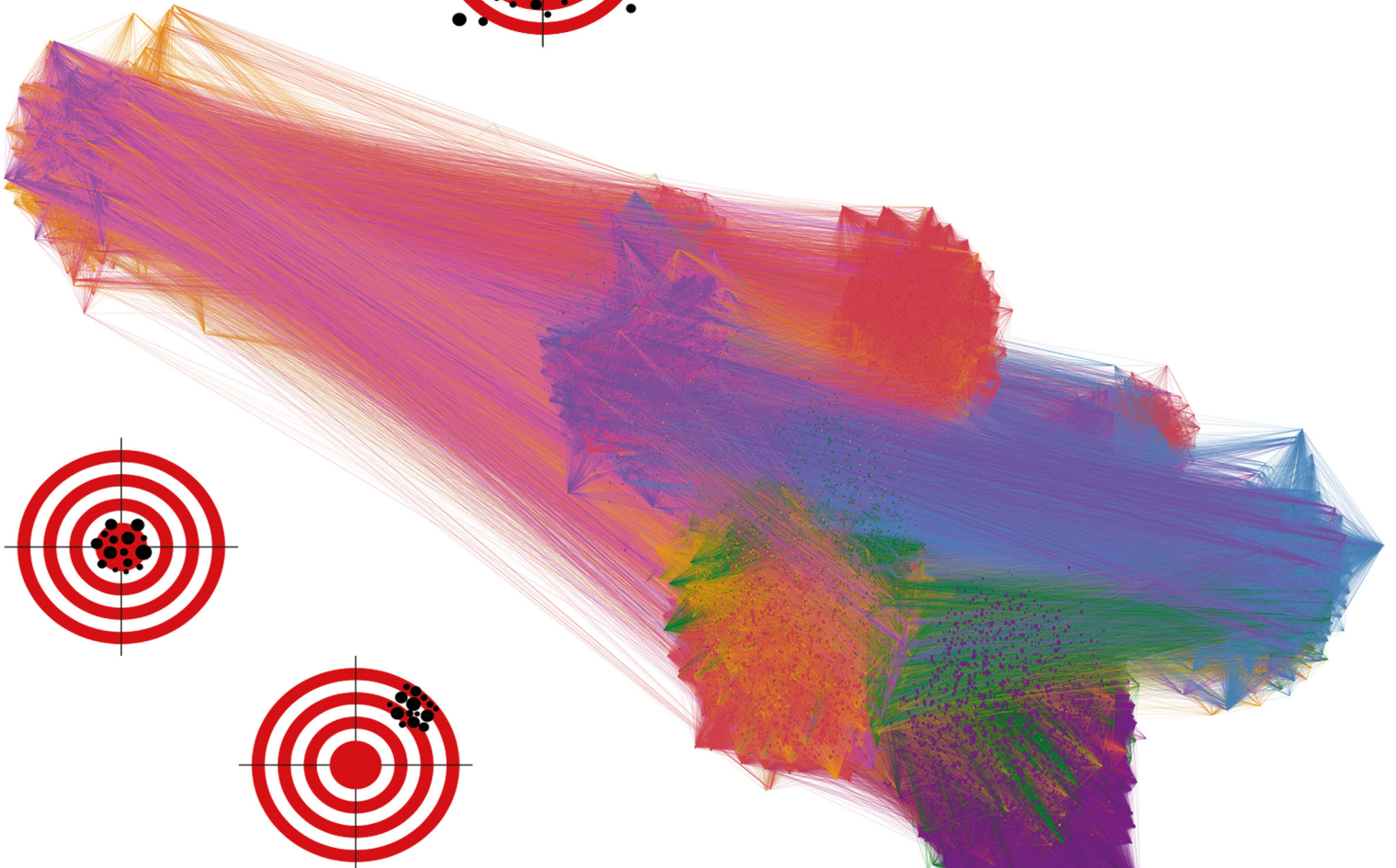
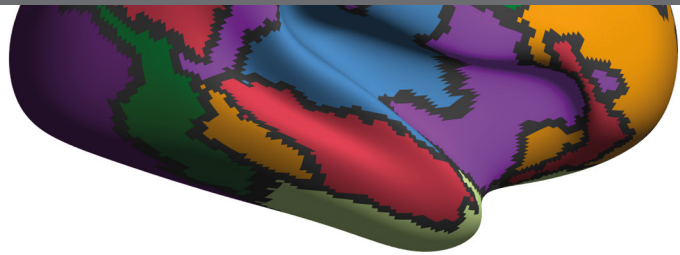




# RELIABILITY AND REPRODUCIBILITY IN FUNCTIONAL CONNECTOMICS

EDITED BY: Xi-Nian Zuo, Bharat B. Biswal and Russell A. Poldrack  
PUBLISHED IN: Frontiers in Neuroscience





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-821-9

DOI 10.3389/978-2-88945-821-9

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



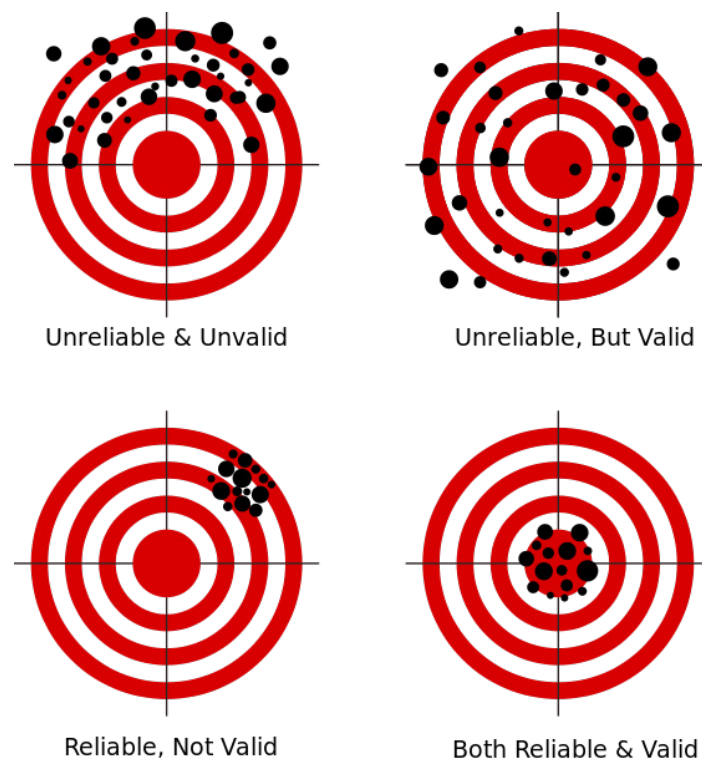
# RELIABILITY AND REPRODUCIBILITY IN FUNCTIONAL CONNECTOMICS

Topic Editors:

**Xi-Nian Zuo**, Chinese Academy of Sciences, China

**Bharat B. Biswal**, New Jersey Institute of Technology, United States

**Russell A. Poldrack**, Stanford University, United States



Shooting Reliable Human Functional Connectomics.

Image and Cover Image: Xi-Nian Zuo.

Functional connectomics enables researchers to monitor interactions among thousands of units within the whole brain simultaneously by using various vivo imaging technologies. For example, resting-state functional magnetic resonance imaging can image low-frequency fluctuations in the spontaneous brain activities, representing a popular tool for macro-scale functional connectomics to characterize individual differences in normal brain function, mind-brain associations, and the various disorders. Reliability and reproducibility represents the most fundamental and critical aspect for the human brain functional connectomics to both research and clinical practice. Unfortunately, lacking a data platform for researchers to rigorously explore the reliability and reproducibility of the functional connectome indices has

been a bottleneck of further development of clinically oriented imaging markers in the field. Recent efforts on open neuroscience, such as Consortium for Reliability and Reproducibility, Human Connectome Project and OpenfMRI, provide the data for the field to refine and evaluate reliability and reproducibility of novel methods as well as those with widespread usage but without sufficient consideration of reliability. This Frontiers Research Topic aims at bringing together contributions from researchers in brain imaging, neuroscience, computer sciences, applied mathematics, psychology and related fields from an interdisciplinary perspective. By focusing on cutting-edge research across these fields, this topic will create new agenda on quantifying the reliability and reproducibility of the myriad connectomics-based measures and informing expectations regarding the potential of biomarker discovery.

**Citation:** Zuo, X-N., Biswal, B. B., Poldrack, R. A., eds. (2019). Reliability and Reproducibility in Functional Connectomics. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88945-821-9



# Table of Contents

- 06** *Editorial: Reliability and Reproducibility in Functional Connectomics*  
Xi-Nian Zuo, Bharat B. Biswal and Russell A. Poldrack
- 10** *Commentary: A Test-Retest Dataset for Assessing Long-Term Reliability of Brain Morphology and Resting-State Brain Activity*  
João R. Sato, Thomas P. White and Claudinei E. Biazoli Jr.

## CHAPTER 1

### PHYSIOLOGICAL FACTORS

- 14** *The Effect of Low-Frequency Physiological Correction on the Reproducibility and Specificity of Resting-State fMRI Metrics: Functional Connectivity, ALFF, and ReHo*  
Ali M. Golestani, Jonathan B. Kwint, Yasha B. Khatamian and J. Jean Chen
- 33** *Improving the Test-Retest Reliability of Resting State fMRI by Removing the Impact of Sleep*  
Jiahui Wang, Junwei Han, Vinh T. Nguyen, Lei Guo and Christine C. Guo
- 46** *Intra- and Inter-scanner Reliability of Scaled Subprofile Model of Principal Component Analysis on ALFF in Resting-State fMRI Under Eyes Open and Closed Conditions*  
Li-Xia Yuan, Jian-Bao Wang, Na Zhao, Yuan-Yuan Li, Yilong Ma, Dong-Qiang Liu, Hong-Jian He, Jian-Hui Zhong and Yu-Feng Zang

## CHAPTER 2

### NOVEL METHODS

- 56** *Improving the Reliability of Network Metrics in Structural Brain Networks by Integrating Different Network Weighting Strategies Into a Single Graph*  
Stavros I. Dimitriadis, Mark Drakesmith, Sonya Bells, Greg D. Parker, David E. Linden and Derek K. Jones
- 73** *Reliability of Static and Dynamic Network Metrics in the Resting-State: A MEG-Beamformed Connectivity Analysis*  
Stavros I. Dimitriadis, Bethany Routley, David E. Linden and Krish D. Singh
- 95** *Test-Retest Reliability of "High-Order" Functional Connectivity in Young Healthy Adults*  
Han Zhang, Xiaobo Chen, Yu Zhang and Dinggang Shen
- 115** *Comparison of IVA and GIG-ICA in Brain Functional Network Estimation Using fMRI Data*  
Yuhui Du, Dongdong Lin, Qingbao Yu, Jing Sui, Jiayu Chen, Srinivas Rachakonda, Tulay Adali and Vince D. Calhoun
- 133** *Psychophysiological Interactions in a Visual Checkerboard Task: Reproducibility, Reliability, and the Effects of Deconvolution*  
Xin Di and Bharat B. Biswal
- 149** *Brain Network Dynamics Adhere to a Power Law*  
Dardo G. Tomasi, Ehsan Shokri-Kojori and Nora D. Volkow

## CHAPTER 3

### CLINICAL CHALLENGES

**158** *Inconsistency in Abnormal Brain Activity Across Cohorts of ADHD-200 in Children With Attention Deficit Hyperactivity Disorder*

Jian-Bao Wang, Li-Jun Zheng, Qing-Jiu Cao, Yu-Feng Wang, Li Sun, Yu-Feng Zang and Hang Zhang

**168** *Investigating Brain Connectomic Alterations in Autism Using the Reproducibility of Independent Components Derived From Resting State Functional MRI Data*

Mohammed A. Syed, Zhi Yang, Xiaoping P. Hu and Gopikrishna Deshpande





# Editorial: Reliability and Reproducibility in Functional Connectomics

Xi-Nian Zuo<sup>1,2,3,4,5,6\*</sup>, Bharat B. Biswal<sup>7,8</sup> and Russell A. Poldrack<sup>9</sup>

<sup>1</sup> Key Laboratory of Brain and Education, Nanning Normal University, Nanning, China, <sup>2</sup> Department of Psychology, University of Chinese Academy of Science, Beijing, China, <sup>3</sup> CAS Key Laboratory of Behavioral Sciences, Institute of Psychology, Beijing, China, <sup>4</sup> Magnetic Resonance Imaging Research Center, CAS Institute of Psychology, Beijing, China, <sup>5</sup> Research Center for Lifespan Development of Mind and Brain, CAS Institute of Psychology, Beijing, China, <sup>6</sup> Institute for Brain Research and Rehabilitation, South China Normal University, Guangzhou, China, <sup>7</sup> The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China, <sup>8</sup> Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, United States, <sup>9</sup> Department of Psychology, Stanford University, Stanford, CA, United States

**Keywords:** test-retest reliability, functional connectomics, open science, dynamic brain theory, big data

## Editorial on the Research Topic

### Reliability and Reproducibility in Functional Connectomics

#### OPEN ACCESS

##### Edited and reviewed by:

Vince D. Calhoun,  
University of New Mexico,  
United States

##### \*Correspondence:

Xi-Nian Zuo  
zuoxn@gxtc.edu.cn;  
zuoxn@psych.ac.cn

##### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 28 October 2018

**Accepted:** 31 January 2019

**Published:** 20 February 2019

##### Citation:

Zuo X-N, Biswal BB and Poldrack RA  
(2019) Editorial: Reliability and  
Reproducibility in Functional  
Connectomics.  
Front. Neurosci. 13:117.  
doi: 10.3389/fnins.2019.00117

Research on functional connectomics of the human brain is exploding (Kelly et al., 2012; Smith et al., 2013), especially for clinical and neurodevelopmental as well as aging studies. However, advances in the reliability and validity of functional connectomics have so far lagged the application of these methods in practice (Zuo and Xing, 2014). In statistical theory, reliability serves as an upper limit of validity and is measurable in practice while validity is more difficult to measure directly (e.g., specific trait and disease) thus often approximated by predictive validity (Kraemer, 2014). Therefore, high reliability is a required standard for both research and clinical use. Of note, excellent reliability ( $>0.8$ ) serves the clinical standard on measurement scales (Streiner et al., 2015). This reflects clinical call of tools with high inter-individual differences (*easily differentiating individuals*) and low intra-individual differences (*high individual stability*) (Fleiss et al., 2003; Zuo and Xing, 2014). This has been recently demonstrated in the anatomy of reliability (Xing and Zuo, 2018). In reliability studies, statistical quantification of reliability is often implemented with intraclass correlation (ICC) regarding its well-developed theory in the field of probability and statistics while the types of ICC are determined by the repeated-measure experimental design (Shrout and Fleiss, 1979; Koo and Li, 2016). Failure of reliability can be an important cause of small statistical power (Button et al., 2013), low reproducibility (Poldrack et al., 2017), puzzlingly high correlations (Vul et al., 2009), and overwhelming need of big data or large sample sizes (Streiner et al., 2015; Hedge et al., 2018). In the field of human brain mapping with magnetic resonance imaging (MRI), structural MRI has clinically-acceptable reliability of mapping brain morphology (Madan and Kensinger, 2017) while most functional MRI measures are challenged by the clinical standard on the reliability (Bennett and Miller, 2010; Zuo and Xing, 2014). This research topic takes action on further steps of improving the reliability of fMRI-based connectomics by publishing 12 papers across experimental design, computational algorithm, and brain dynamics theory.

Given the sensitivity of resting-state fMRI (rfMRI) connectivity measurements to physiological variables, the development of improved strategies for correction of physiological artifacts is imperative. Golestani et al. demonstrated significant improvements of reproducibility of common rfMRI metrics by the low-frequency physiological correction with end-tidal CO<sub>2</sub>. Related to human arousal, as demonstrated in Wang et al., test-retest reliability of human functional connectomics can be significantly improved by removing the impact of sleep using measures of heart rate variability derived from simultaneous electrocardiogram recording. These findings highlight the need of recordings of physiological variables for reproducible functional connectomics. In addition, the use of eyes-open versus eyes-closed resting is an important aspect of rfMRI experimental design and has been of great research interest due to its relationships with visual function (Yang et al., 2007) and arousal (Yan et al., 2009; Tagliazucchi and Laufs, 2014). The study by Yuan et al. provides a novel multivariate method to examine the amplitude differences of brain oscillations between eyes open and eyes close conditions during resting state as well as their scanner-related reliability. Head motion during scanning is another potential source of variability and has been relatively well investigated regarding its impacts on reliability of rfMRI derivatives by using various preprocessing strategies (Yan et al., 2013; Ciric et al., 2017; Parkes et al., 2018). Furthermore, how these variables are modeled and the order in the preprocessing pipelines they are modeled can have significant impacts on results (Chen et al., 2017; Lindquist et al., 2019). These advances have implications on the way of further optimizing the reliability observed (Golestani et al.; Wang et al.).

Many computational algorithms exist for characterizing features of the organization in the functional connectomes across different spatial and temporal scales (Zuo and Xing, 2014). Reliability can guide both methodological choices between these algorithms as well as the validation of new algorithms. Common algorithms have been recently given a state of art review in terms of their test-retest reliability (Zuo and Xing, 2014), indicating that network metrics derived from graph theory applied to rfMRI signal are less reliable (Zuo et al., 2012) than usually required while both local functional homogeneity measure (Zuo et al., 2013) and global network measure with dual regression of independent component analysis (drICA) (Zuo et al., 2010a) almost reach the clinical standard of reliability. This topic offers five studies to illustrate more sophisticated developments of reliability of these algorithms. This topic proposed a novel algorithm for network generation at individual level, using topological filtering based on orthogonal minimal spanning trees to show both functional and structural networks with highly reliable graph theoretical measures using magnetoencephalography (Dimitriadis et al.) and diffusion MRI (Dimitriadis et al.). Reliability evaluations are comprehensively investigated for group information guided ICA, independent vector analysis (IVA) (Du et al.) and other high-order functional connectivity (Zhang et al.). The single-subject spatially-constrained ICA performs favorably compared to IVA (Du et al.) and improves detection of clinical differences compared

to drICA (Salman et al., 2018). Additionally, Di and Biswal warned the field by demonstrating the poor reliability of using psychophysiological interaction analyses in the context of inter-individual correlation or group comparisons.

As commented by Sato et al., open science with sharing of large datasets has paved the way for delineating the fingerprints of human brain function. This is reflected by the fact that most studies in the topic employed the data from Consortium for Reliability and Reproducibility (Zuo et al., 2014), representing a means of accelerating science by facilitating collaboration, transparency, and reproducibility (Milham et al., 2018). To address the reproducibility issue in the field of human brain mapping, the Organization for Human Brain Mapping (OHBM) have created a Committee on Best Practices in Data Analysis and Sharing (COBIDAS) and published its report (Nichols et al., 2017). Beyond the advances, two studies also raised challenges of big-data applications to clinical population, particularly in understanding the high heterogeneity of spontaneous brain activity in ADHD and autism (Wang et al.; Syed et al.). As noted in Button et al. (2013), large samples may produce statistically significant results even for extremely small effects which have little add to diagnostic or clinical utility. These observable but small effects are likely caused by weighing the low measurement reliability with the true effect (Streiner et al., 2015), which could be moderate to large. It is thus very fundamental to estimate effect size in neuroimaging and its relationship with statistical power although most existing studies have not factored the reliability in doing so (Reddan et al., 2017; Geuter et al., 2018). This is particularly valuable for some widely used but less reliable measures (e.g., seed-based functional connectivity) (Shou et al., 2013; Zuo and Xing, 2014; Siegel et al., 2017) to be improved with acceptable reliability ahead of its clinical use (Fox, 2018). Meanwhile, data harmonization techniques such as ComBat (Yu et al., 2018) should be developed to reduce inter-scan or inter-site differences in multi-center big-data studies. One possibility of filling these gaps between empirical computation and clinical application is theoretical development of brain dynamics (Woo et al., 2017). The work by Tomasi et al. demonstrated a power law of the brain network dynamics, which has been framed into a theory of neural oscillations (Buzsáki and Draguhn, 2004). Combination of theory and data via structure-function fusion (Zuo et al., 2010b; Jiang and Zuo, 2016) will remove the reliability barriers of developing clinically useful human brain mapping, which is the final call of the current research topic.

## AUTHOR CONTRIBUTIONS

X-NZ drafted the editorial and worked on the revisions with BB and RP.

## FUNDING

This work was supported in part by the National Basic Research (973) Program (2015CB351702), the Natural Science Foundation of China (81471740, 81220108014), Beijing Municipal Science and Tech Commission (Z161100002616023,



Z171100000117012), the China - Netherlands CAS-NWO Programme (153111KYSB20160020), the Major Project of National Social Science Foundation of China (14ZDB161), the National R&D Infrastructure and Facility Development Program of China, Fundamental Science Data Sharing Platform (DKA2017-12-02-21), and Guangxi BaGui Scholarship (201621 to X-NZ).

## REFERENCES

- Bennett, C. M., and Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Ann. N.Y. Acad. Sci.* 1191, 133–155. doi: 10.1111/j.1749-6632.2010.05446.x
- Butt, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Chen, J. E., Jahani, H., and Glover, G. H. (2017). Nuisance regression of high-frequency functional magnetic resonance imaging data: denoising can be noisy. *Brain Connect.* 7, 13–24. doi: 10.1089/brain.2016.0441
- Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., et al. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* 154, 174–187. doi: 10.1016/j.neuroimage.2017.03.020
- Fleiss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, 3rd edn. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons.
- Fox, M. D. (2018). Mapping symptoms to brain networks with the human connectome. *New Engl. J. Med.* 379, 2237–2245. doi: 10.1056/NEJMra1706158
- Geuter, S., Qi, G., Welsh, R. C., Wager, T. D., and Lindquist, M. A. (2018). Effect size and power in fMRI group analysis. *bioRxiv [Preprint]*. bioRxiv:295048. doi: 10.1101/295048
- Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* 50, 1166–1186. doi: 10.3758/s13428-017-0935-1
- Jiang, L., and Zuo, X. N. (2016). Regional homogeneity: a multimodal, multiscale neuroimaging marker of the human connectome. *Neuroscientist* 22, 486–505. doi: 10.1177/1073858415595004
- Kelly, C., Biswal, B. B., Craddock, R. C., Castellanos, F. X., and Milham, M. P. (2012). Characterizing variation in the functional connectome: promise and pitfalls. *Trends Cogn. Sci.* 16, 181–188. doi: 10.1016/j.tics.2012.02.001
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Kraemer, H. C. (2014). The reliability of clinical diagnoses: state of the art. *Annu. Rev. Clin. Psychol.* 10, 111–130. doi: 10.1146/annurev-clinpsy-032813-153739
- Lindquist, M. A., Geuter, S., Wager, T. D., and Caffo, B. S. (2019). Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human Brain Mapp.* doi: 10.1002/hbm.24528. [Epub ahead of print].
- Madan, C. R., and Kensing, E. A. (2017). Test-retest reliability of brain morphology estimates. *Brain Inform.* 4, 107–121. doi: 10.1007/s40708-016-0060-4
- Milham, M. P., Craddock, R. C., Son, J. J., Fleischmann, M., Clucas, J., Xu, H., et al. (2018). Assessment of the impact of shared brain imaging data on the scientific literature. *Nat. Commun.* 9, 2818. doi: 10.1038/s41467-018-04976-1
- Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., et al. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20, 299–303. doi: 10.1038/nn.4500
- Parke, L., Fulcher, B., Yücel, M., and Fornito, A. (2018). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* 171, 415–436. doi: 10.1016/j.neuroimage.2017.12.073
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., et al. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* 18, 115–126. doi: 10.1038/nrn.2016.167
- Reddan, M. C., Lindquist, M. A., and Wager, T. D. (2017). Effect size estimation in neuroimaging. *JAMA Psychiatry* 74, 207–208. doi: 10.1001/jamapsychiatry.2016.3356
- Salman, M. S., Du, Y., Lin, D., Fu, Z., Damaraju, E., Sui, J., et al. (2018). Group ICA for identifying biomarkers in schizophrenia: ‘adaptive’ networks via spatially constrained ICA show more sensitivity to group differences than spatio-temporal regression. *bioRxiv [Preprint]*. bioRxiv:429837. doi: 10.1101/429837
- Shou, H., Eloyan, A., Lee, S., Zipunnikov, V., Crainiceanu, A. N., Nebel, N. B., et al. (2013). Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (I2C2). *Cogn. Affect. Behav. Neurosci.* 13, 714–724. doi: 10.3758/s13415-013-0196-0
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Siegel, J. S., Mitra, A., Laumann, T. O., Seitzman, B. A., Raichle, M., Corbetta, M., et al. (2017). Data quality influences observed links between functional connectivity and behavior. *Cereb. Cortex* 27, 4492–4502. doi: 10.1093/cercor/bhw253
- Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., et al. (2013). Functional connectomics from resting-state fMRI. *Trends Cogn. Sci.* 17, 666–682. doi: 10.1016/j.tics.2013.09.016
- Streiner, D. L., Norman, G. R., and Cairney, J. (2015). *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5th Edn. New York, NY: Oxford University Press.
- Tagliazucchi, E. and Laufs, H. (2014). Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708. doi: 10.1016/j.neuron.2014.03.020
- Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* 4, 274–290. doi: 10.1111/j.1745-6924.2009.01125.x
- Woo, C. W., Chang, L. J., Lindquist, M. A., and Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* 20, 365–377. doi: 10.1038/nn.4478
- Xing, X. X., and Zuo, X. N. (2018). The anatomy of reliability: a must read for future human brain mapping. *Sci. Bull.* 63, 1606–1607. doi: 10.1016/j.scib.2018.12.010
- Yan, C., Liu, D., He, Y., Zou, Q., Zhu, C., Zuo, X., et al. (2009). Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. *PLoS ONE* 4, e5743. doi: 10.1371/journal.pone.0005743
- Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Martino, A. D., et al. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage* 76, 183–201. doi: 10.1016/j.neuroimage.2013.03.004
- Yang, H., Long, X. Y., Yang, Y., Yan, H., Zhu, C. Z., Zhou, X. P., et al. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *Neuroimage* 36, 144–152. doi: 10.1016/j.neuroimage.2007.01.054
- Yu, M., Linn, K. A., Cook, P. A., Phillips, M. L., McInnis, M., Fava, M., et al. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum. Brain Mapp.* 39, 4213–4227. doi: 10.1002/hbm.24241
- Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* 1:140049. doi: 10.1038/sdata.2014.49

## ACKNOWLEDGMENTS

We would like to thank Dr. Xiu-Xia Xing from School of Applied Sciences, Beijing University of Technology for her work on drafting the first version of this editorial as well as highly valuable comments on the importance of reliability to research and clinical implications.

- Zuo, X. N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., et al. (2012). Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875. doi: 10.1093/cercor/bhr269
- Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010a). Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. *Neuroimage* 49, 2163–2177. doi: 10.1016/j.neuroimage.2009.10.080
- Zuo, X. N., Kelly, C., Martino, A. D., Mennes, M., Margulies, D. S., Bangaru, S., et al. (2010b). Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. *J. Neurosci.* 30, 15034–15043. doi: 10.1523/JNEUROSCI.2612-10.2010
- Zuo, X. N. and Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009
- Zuo, X. N., Xu, T., Jiang, L., Yang, Z., Cao, X. Y., He, Y., et al. (2013). Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage* 65, 374–386. doi: 10.1016/j.neuroimage.2012.10.017

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zuo, Biswal and Poldrack. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Commentary: A test-retest dataset for assessing long-term reliability of brain morphology and resting-state brain activity

João R. Sato<sup>1\*</sup>, Thomas P. White<sup>2</sup> and Claudinei E. Biazoli Jr.<sup>1</sup>

<sup>1</sup> Centre of Mathematics, Computation and Cognition, Universidade Federal do ABC, Santo Andre, Brazil, <sup>2</sup> School of Psychology, University of Birmingham, Birmingham, UK

**Keywords:** replicability, brain networks, brain imaging, reproducibility, fingerprints

## A commentary on

### A test-retest dataset for assessing long-term reliability of brain morphology and resting-state brain activity

by Huang, L., Huang, T., Zhen, Z., and Liu, J. (2016). *Sci. Data* 3:160016. doi: 10.1038/sdata.2016.16

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Institute of Psychology (CAS), China

### Reviewed by:

Lijie Huang,  
Institute of Automation (CAS), China  
Ting Xu,  
Child Mind Institute, USA

### \*Correspondence:

João R. Sato  
joao.sato@ufabc.edu.br

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 13 December 2016

**Accepted:** 07 February 2017

**Published:** 22 February 2017

### Citation:

Sato JR, White TP and Biazoli CE Jr.  
(2017) Commentary: A test-retest  
dataset for assessing long-term  
reliability of brain morphology and  
resting-state brain activity.  
*Front. Neurosci.* 11:85.  
doi: 10.3389/fnins.2017.00085

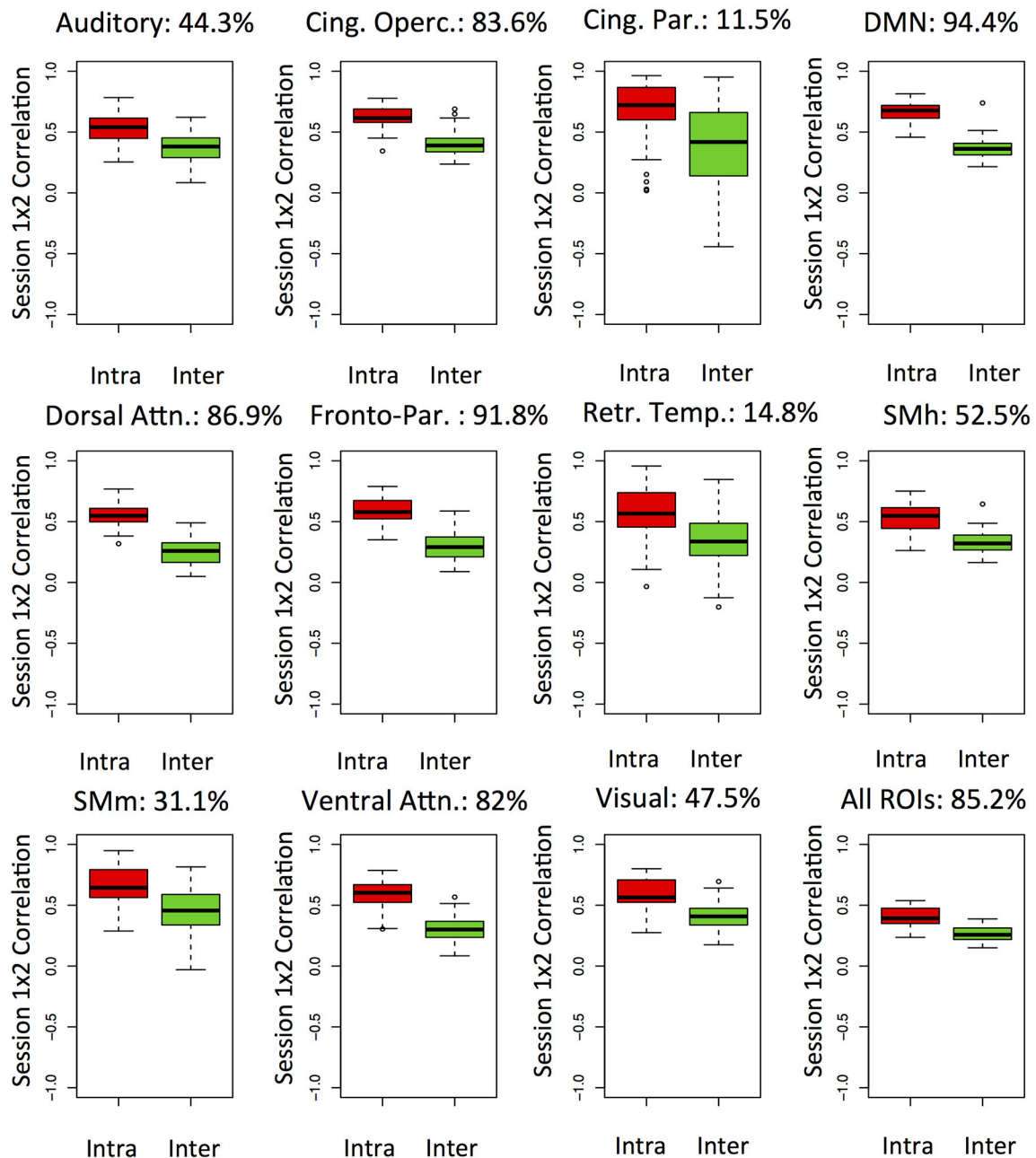
A transformation toward open neuroscience is ongoing (Milham, 2012), and increases the availability of high-quality, open-access neuroimaging datasets (Poline et al., 2012; Mennes et al., 2013). Consequently, a new set of analytical approaches, including discovery science (Biswal et al., 2010) and focus on individual rather than group-level effects (Finn et al., 2015; Miranda-Dominguez et al., 2014), are increasingly accessible. However, moving to single-subject statistics raises specific concerns that must be addressed. Notable amongst these is the test-retest reliability of fMRI-based metrics (Dubois and Adolphs, 2016). Recently, Huang et al. (2016) provided a test-retest neuroimaging dataset (BNU2), with an inter-scan interval in the order of months, allowing investigation of the temporal reliability of features extracted from rs-fMRI. In the spirit of the data-sharing initiatives, the “Consortium for Reliability and Reproducibility in Functional Connectomics” (CoRR) publicly released this data (Zuo et al., 2014).

Recently, Finn et al. (2015) investigated the existence of functional “connectome fingerprints.” The authors hypothesized that, despite overall similarity in connectivity patterns across subjects, portions of brain connectome variability would be fairly singular to each individual (Mueller et al., 2013; Gordon et al., 2015; Laumann et al., 2015; Xu et al., 2016). This notable study demonstrated that, using only the functional connectivity profile extracted from an fMRI scanning session, it was possible to identify the same subjects from their profiles from a second session a few days later. Interestingly, this hypothesis of a functional connectivity fingerprint was also supported by previous work (Miranda-Dominguez et al., 2014), which additionally showed that such individual signatures exist not only in humans but also in non-human primates.

However, the extent to which connectome profile stability can be generalized to more extended timescales remains largely untested. Furthermore, the vast majority of the functional connectome studies to date focus on timescales of seconds to minutes or years to decades (Poldrack et al., 2015; Huang et al., 2016, but see Xu et al., 2016). We thought that the BNU2 dataset is quite suitable to assess the reliability and stability of connectome fingerprints on an intermediate timescale of months. In fact, the released BNU2 dataset consists of anatomical and functional data from 61 healthy adults (19–23 years old) scanned under a resting-state protocol (eyes closed) in two sessions at an interval of 103–189 days. Further information about scanning parameters, demographical and quality metrics data can be found in Huang et al. (2016).

We preprocessed the data and extracted individual functional connectivity estimates using CONN toolbox version 15.g (Whitfield-Gabrieli and Nieto-Castanon, 2012) with standard MNI152 pipeline and parameters. Conservative options (discarding volumes with displacement  $>0.5$  mm and global-signal  $z$ -value  $>3$ ) for scan motion censoring were applied, since motion artifacts are a well-recognized source of error

in functional connectivity studies using fMRI. The pairwise bivariate correlations (functional connectivity) among 333 cortical regions-of-interest (ROIs) were obtained using the Gordon et al. (2016) parcellation. Considering the upper triangular values of the individual correlation matrices as the subject connectivity profile, a functional connectome fingerprinting analyses was then carried out. The similarity



**FIGURE 1 | Functional networks—Box-plots of intra (red) and inter-subject (green) correlations between the first and second resting state fMRI sessions.** The analyses are carried out separated by networks and the connectome fingerprinting accuracy is highlighted as percentage. Cing. Oper., Cingulo-opercular; Cing. Par., Cingulo-parietal; DMN, Default-mode network; Dorsal Attn, Dorsal attention; Fronto-par, Fronto-parietal; Retr. Temp., Retrosplenial-temporal; SMh, somatomotor-hand; SMm, somatomotor-mouth; Ventral Attn, Ventral attention.

between the two profiles was then measured with Spearman's correlation coefficient. The within-subject correlation between the two sessions determines the accuracy as it reflects the proportion of subjects correctly identified. Note that the expected accuracy by chance is  $1/61 = 1.6\%$ .

We expected to reproduce the original results from connectome fingerprint studies (Miranda-Dominguez et al., 2014; Finn et al., 2015) if the individual profiles are stable over months. In order to do so, we attempted to identify the subjects in the second session based on the profiles similarity to the first session. As a second step, we calculated the intra- and inter-subject similarities between the two sessions for each subject. The inter-subject similarity was calculated by random sampling an individual at the second session. We also sought to investigate how large-scale networks connectivity varies within and between subjects in the timescale of months. Each brain parcel was labeled for the conventional resting-state networks, as provided by the Gordon atlas. Thus, we conducted the two previously described analyses' steps considering all ROIs and each network separately. Based on previous findings of within and between subject variability of network connectivity (Mueller et al., 2013; Miranda-Dominguez et al., 2014; Zuo and Xing, 2014; Chen et al., 2015; Finn et al., 2015; Poldrack et al., 2015), we expected increased discriminability of individuals for heteromodal associative networks.

The results are shown in **Figure 1**. A high accuracy of 85% for the whole-brain connectivity profile was found. Moreover, accuracies were above 90% for the default mode and the frontoparietal networks. Interestingly, accuracies for primary sensory and motor networks were lower. It is likely that the ability to uniquely discriminate individuals relies on features with both low within and high between-subject variability over time. For all the networks investigated, we noticed a tendency for higher similarity within subjects than between them. Overall, these results suggest stable connectome fingerprints exist over months and are in agreement with the previously reported inter-individual variability of networks including heteromodal areas. However, caution should be taken when interpreting differences in accuracy between networks as the number and extent of ROIs varies. Since each ROI signal is based on average across

voxels, networks with larger parcels may have superior signal-to-noise ratios. Moreover, the number of ROIs may be related to redundancy of information in the connectivity matrices, which would also affect accuracies. Remarkably, the networks which presented the lowest subjects identification accuracies have  $<8$  ROIs.

Results from individual-based fMRI metrics can be framed by the usual concepts of validity and reliability (Dubois and Adolphs, 2016). However, an inherent issue of the approaches like those proposed here is the extent to which validity and reliability can be disentangled. In other words, it is possible to state that connectome fingerprints are stable over the months, which would constitute a claim for the validity of the underlying neural phenomena. Alternatively, but not mutually exclusively, it is also possible that test-retest reliability varies between subjects and networks. We argue that continuous effort for data-sharing, in the spirit of the CoRR and other initiatives, is of paramount importance as disentangling these factors will ultimately depend on accumulating evidence for the stability of connectome fingerprints across different timescales and with large datasets. Establishing the stability of these measures, in turn, will be essential to investigate true effects of development on the connectomes. Furthermore, adopting comparable acquisition parameters and open and reliable data processing will be necessary to further assure the validity of remarkable findings such as individually unique connectivity profiles.

## AUTHOR CONTRIBUTIONS

JS preprocessed and analyzed the data. All authors wrote the manuscript.

## ACKNOWLEDGMENTS

The authors are grateful to Huang et al. (2016) and the "Consortium for Reliability and Reproducibility in Functional Connectomics" (CoRR) for providing the data. The authors also acknowledge and are thankful to the two reviewers who provided valuable comments to improve this letter.

## REFERENCES

- Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107
- Chen, B., Xu, T., Zhou, C., Wang, L., Yang, N., Wang, Z., et al. (2015). Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month. *PLoS ONE* 10:e0144963. doi: 10.1371/journal.pone.0144963
- Dubois, J., and Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends Cogn. Sci.* 20, 425–443. doi: 10.1016/j.tics.2016.03.014
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671. doi: 10.1038/nn.4135
- Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* 26, 288–303. doi: 10.1093/cercor/bhu239
- Gordon, E. M., Laumann, T. O., Adeyemo, B., and Petersen, S. E. (2015). Individual variability of the system-level organization of the human brain. *Cereb. Cortex* doi: 10.1093/cercor/bhv239. [Epub ahead of print].
- Huang, L., Huang, T., Zhen, Z., and Liu, J. (2016). A test-retest dataset for assessing long-term reliability of brain morphology and resting-state brain activity. *Sci. Data* 3:160016. doi: 10.1038/sdata.2016.16
- Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M. Y., et al. (2015). Functional system and areal organization of a highly sampled individual human brain. *Neuron* 87, 657–670. doi: 10.1016/j.neuron.2015.06.037
- Mennes, M., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2013). Making data sharing work: the FCP/INDI experience. *Neuroimage* 82, 683–691. doi: 10.1016/j.neuroimage.2012.10.064
- Milham, M. P. (2012). Open neuroscience solutions for the connectome-wide association era. *Neuron* 73, 214–218. doi: 10.1016/j.neuron.2011.11.004



- Miranda-Dominguez, O., Mills, B. D., Carpenter, S. D., Grant, K. A., Kroenke, C. D., Nigg, J. T., et al. (2014). Connectotyping: model based fingerprinting of the functional connectome. *PLoS ONE* 9:e111048. doi: 10.1371/journal.pone.0111048
- Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., et al. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron* 77, 586–595. doi: 10.1016/j.neuron.2012.12.028
- Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M. Y., et al. (2015). Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* 6:8885. doi: 10.1038/ncomms9885
- Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Front. Neuroinform.* 6:9. doi: 10.3389/fninf.2012.00009
- Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. doi: 10.1089/brain.2012.0073
- Xu, T., Opitz, A., Craddock, R. C., Wright, M. J., Zuo, X. N., and Milham, M. P. (2016). Assessing variations in areal organization for the intrinsic brain: from fingerprints to reliability. *Cereb. Cortex* 26, 4192–4211. doi: 10.1093/cercor/bhw241
- Zuo, X.-N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* 1:140049. doi: 10.1038/sdata.2014.49
- Zuo, X. N., and Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Sato, White and Biazoli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Effect of Low-Frequency Physiological Correction on the Reproducibility and Specificity of Resting-State fMRI Metrics: Functional Connectivity, ALFF, and ReHo

Ali M. Golestani<sup>1\*</sup>, Jonathan B. Kwinta<sup>1,2</sup>, Yasha B. Khatamian<sup>1</sup> and J. Jean Chen<sup>1,2</sup>

<sup>1</sup> Rotman Research Institute at Baycrest Centre, University of Toronto, Toronto, ON, Canada, <sup>2</sup> Department of Medical Biophysics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

## OPEN ACCESS

### Edited by:

Bharat B. Biswal,  
University of Medicine and Dentistry of  
New Jersey, United States

### Reviewed by:

Sheng Zhang,  
Yale University, United States  
Hidenao Fukuyama,  
Kyoto University, Japan

### \*Correspondence:

Ali M. Golestani  
golestani@psych.utoronto.ca

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 22 January 2017

**Accepted:** 19 September 2017

**Published:** 05 October 2017

### Citation:

Golestani AM, Kwinta JB,  
Khatamian YB and Chen JJ (2017)  
The Effect of Low-Frequency  
Physiological Correction on the  
Reproducibility and Specificity of  
Resting-State fMRI Metrics: Functional  
Connectivity, ALFF, and ReHo.  
Front. Neurosci. 11:546.  
doi: 10.3389/fnins.2017.00546

The resting-state fMRI (rs-fMRI) signal is affected by a variety of low-frequency physiological phenomena, including variations in cardiac-rate (CRV), respiratory-volume (RVT), and end-tidal CO<sub>2</sub> (PETCO<sub>2</sub>). While these effects have become better understood in recent years, the impact that their correction has on the quality of rs-fMRI measurements has yet to be clarified. The objective of this paper is to investigate the effect of correcting for CRV, RVT and PETCO<sub>2</sub> on the rs-fMRI measurements. Nine healthy subjects underwent a test-retest rs-fMRI acquisition using repetition times (TRs) of 2 s (long-TR) and 0.323 s (short-TR), and the data were processed using eight different physiological correction strategies. Subsequently, regional homogeneity (ReHo), amplitude of low-frequency fluctuation (ALFF), and resting-state connectivity of the motor and default-mode networks are calculated for each strategy. Reproducibility is calculated using intra-class correlation and the Dice Coefficient, while the accuracy of functional-connectivity measures is assessed through network separability, sensitivity and specificity. We found that: (1) the reproducibility of the rs-fMRI measures improved significantly after correction for PETCO<sub>2</sub>; (2) separability of functional networks increased after PETCO<sub>2</sub> correction but was not affected by RVT and CRV correction; (3) the effect of physiological correction does not depend on the data sampling-rate; (4) the effect of physiological processes and correction strategies is network-specific. Our findings highlight limitations in our understanding of rs-fMRI quality measures, and underscore the importance of using multiple quality measures to determine the optimal physiological correction strategy.

**Keywords:** resting-state fMRI, physiological noise, test-retest reproducibility, sensitivity, specificity, end-tidal CO<sub>2</sub>, respiratory volume, heart rate variability

## INTRODUCTION

Resting-state fMRI is typically measured through blood oxygenation level dependent (BOLD) contrast, which indirectly measures brain function through blood oxygenation changes following neuronal activity. The typical BOLD-measurement technique is gradient-echo echo-planar-imaging (GE-EPI). However, the BOLD signal contains not only neuronal contributions, but also several physiological contributions, which can either generate BOLD-related hemodynamics or introduce artifacts through interactions with the magnetic field. For instance, respiration and heartbeat generate bulk motion as well as local movement that is most pronounced in the cerebrospinal fluid (CSF), brain stem, and in the vicinity of large blood vessels (Dagli et al., 1999). In addition, respiration causes susceptibility changes in the lungs that interfere with the static magnetic field and induce shifts in the MR image, mainly in the phase-encoding direction (Hu et al., 1995; Raj et al., 2001; Pfeuffer et al., 2002; Murphy et al., 2013)—a major concern for GE-EPI.

Typically, the rs-fMRI sampling rate is  $\sim 0.5$  Hz, which is only appropriate for representing signal changes up to 0.25 Hz. This is much lower than the Nyquist sampling rate required for the fundamental cardiac and respiratory frequencies ( $\sim 1$  and  $\sim 0.3$  Hz, respectively), raising the possibility that physiological contributions to rs-fMRI measures depend on signal sampling rate. Moreover, neuronally-relevant information in rs-fMRI data is commonly identified with the low-frequency range (below 0.1 Hz; Cordes et al., 2001), which is shared with low-frequency physiological fluctuations. The most common examples of these include cardiac rate variation (CRV), respiratory volume per unit time (RVT) and pressure of end-tidal  $\text{CO}_2$  fluctuations ( $\text{PETCO}_2$ ). RVT is mainly localized in the gray matter, specifically in regions with high vascular density, including the occipital region and the default mode network (DMN; Birn et al., 2006). On the other hand, the effect of CRV is strongest in brain regions close to arteries and CSF (Chang et al., 2009). Finally, fluctuations in arterial pressure of  $\text{CO}_2$ , which can be indirectly measured through  $\text{PETCO}_2$ , alter the BOLD signal through vasodilatory and constrictive action. Like the RVT effect, the  $\text{PETCO}_2$  effect is dominant in the gray matter (Wise et al., 2004; Chang and Glover, 2009; Golestani et al., 2015).

In the context of rs-fMRI, these physiological effects have generally been considered as artifacts from non-neuronal sources that can mimic BOLD signal fluctuations and connectivity, potentially reducing the reliability and neuronal-specificity of rs-fMRI measures. Some excellent works in recent years established the theoretical foundation for investigating and removing these physiological effects from the rs-fMRI signal (Birn et al., 2006; Chang et al., 2009). Specifically, the typical procedure is to record the corresponding physiological signals during the rs-fMRI data acquisition, model their effects on the BOLD signal and eliminate them using regression (Birn et al., 2006; Chang et al., 2009; Golestani et al., 2015). However, little is known about the effect of the correction on the quality of rs-fMRI measures, and indeed, the consequences of different physiological corrections.

In the rs-fMRI literature, the accuracy of rs-fMRI measures is typically assessed based on their test-retest reproducibility, commonly quantified through the intra-class correlation coefficient (ICC; Anderson et al., 2011b; Chou et al., 2012; Faria et al., 2012; Zuo and Xing, 2014). ICC is defined as the ratio of inter-subject variance to total variance (inter-subject + inter-session variance). If within-subject inter-session variance were considerably smaller than inter-subject variance, ICC would be close to one, which is inferred as high reproducibility. Previous studies of various rs-fMRI measures have shown moderate to high reproducibility, depending on the measure (Zuo and Xing, 2014). That is, measures such as amplitude of low frequency fluctuations (ALFF; Zuo et al., 2010a) and regional homogeneity (ReHo; Zuo et al., 2013) are highly reproducible across sessions, whereas connectivity metrics derived from graph-theoretical network analysis are considered not very reproducible (Wang et al., 2011). Moreover, the reproducibility of connectivity maps is sensitive to acquisition length, the number of time points included (Birn et al., 2013; Liao et al., 2013), the sampling rate (Liao et al., 2013) and of course the processing steps (Franco et al., 2013; Zuo et al., 2013). The ICC, however, only assesses the reproducibility of the connectivity values but not that of network extent. The latter has previously been assessed using the Dice Similarity Coefficient (Amemiya et al., 2014; Ganger et al., 2015; Jann et al., 2015). The Dice Coefficient compares the spatial extent of different connectivity maps, and a Dice Coefficient close to unity reflects high overlap between two maps, hence high spatial reproducibility.

Notwithstanding the current emphasis on reproducibility as the chief quality measure, high reproducibility does not equal to high measurement quality. For instance, we should also like to be able to distinguish between the areas that are part of a network from those outside of it (i.e., high sensitivity and specificity). Yet, sensitivity and specificity has been largely overlooked in the literature, as they are more difficult to assess. In that respect, while the true individual resting-state connectivity map is unknown, a number of resting-state functional networks have been consistently found in various populations (Damoiseaux et al., 2006; Yeo et al., 2011). The resulting group-based network atlases, which are arguably less affected by physiological artifacts compared to the individual subject-level maps, are presumably more robust and representative of true functional networks. Thus, we may now have a means to quantify the sensitivity and specificity of rs-fMRI connectivity maps.

To the authors' best knowledge, there exists only one prior study addressing the effect of various physiological corrections on rs-fMRI measurement quality, despite the importance of the topic (Birn et al., 2014). Interestingly, the findings suggest that physiological correction may have little or even a negative effect on the reproducibility of the fMRI connectivity patterns. To explain this surprising finding, the authors skilfully demonstrated that physiological correction reduces both within- and between-subject variance, resulting in an overall ICC reduction. Despite this observation, the authors recommend removing the physiological effects from the BOLD signal, as the physiological correction would potentially increase the validity of the rs-fMRI connectivity studies. Moreover, the authors

correctly admitted in the paper that the accuracy of using a global physiological regressor for physiological correction is questionable, given the evident inter-subject and regional variability in the BOLD physiological response (Falahpour et al., 2013; Cordes et al., 2014; Golestani et al., 2015). Nevertheless, this work leaves unanswered a number of important questions. First, it only addressed the effects of CRV and RVT. Given recent evidence of the unique effects of PETCO<sub>2</sub> fluctuations on rs-fMRI (Golestani et al., 2015), the effect of PETCO<sub>2</sub> correction should also be addressed. Second, the study relied solely on reproducibility as a metric of merit, and used ICC as the only measure of reproducibility, neglecting other aspects of rs-fMRI data quality. In addition, the study focused on functional connectivity measurements and did not consider other commonly used rs-fMRI measures such as ReHo and ALFF.

In this paper, we investigate the effect of a number of correction strategies involving three low-frequency physiological signal sources (CRV, RVT, and PETCO<sub>2</sub>) on the rs-fMRI measurements. The novelties of this study are: (1) we study the effect of PETCO<sub>2</sub> correction in addition to CRV and RVT correction; (2) we estimate and eliminate the effect of physiological modulations using a voxel-wise instead of a global approach, accounting for potential inter-subject and inter-regional variability; (3) in assessing reproducibility, we use not only the ICC, but also the Dice Coefficient; (4) in addition to reproducibility, we measure the sensitivity and specificity of the resting-state connectivity maps with the help of a resting-state connectivity template (Yeo et al., 2011); (5) we also assess the separability of the connectivity maps by calculating relative-connectivity of within-network connectivity to between-network connectivity; (6) we include ReHo and ALFF in addition to resting-state connectivity in our assessments; (7) we investigate the effect of fMRI acquisition sampling rate on the efficacy of physiological corrections.

## METHODS

### Participants and Data Acquisition

Nine healthy subjects participated in this study (3 male; mean age =  $26 \pm 5.8$  years). Participants were recruited from Baycrest and local communities through the Baycrest Participants Database. The study was approved by the research ethics board (REB) of Baycrest, and the consent obtained from all participants was both written and informed, in accordance with the Declaration of Helsinki.

All images were acquired using a Siemens TIM Trio 3 Tesla System (Siemens, Erlangen, Germany), with a 32-channel phased-array head coil for reception and body-coil transmission. We acquired rs-fMRI data using multiple repetition times (TR) to investigate the effect of sampling rate. Each TR was used in two sessions to allow assessment of test-retest reproducibility. Specifically, the “long-TR” protocol involved conventional single-shot gradient-echo echo-planar imaging (GRE-EPI; TR = 2,000 ms, TE = 30 ms, flip angle = 90°, 26 slices, 0.6 mm between-slice gap,  $3.44 \times 3.44 \times 4.6$  mm<sup>3</sup> voxels, matrix size:  $64 \times 64 \times 26$ , 240 frames), while the “short-TR” protocol involved slice-accelerated (Feinberg et al., 2010; Setsompop et al., 2012)

single-shot GRE-EPI [TR = 323 ms, TE = 30 ms, flip angle = 40°, 15 slices, 1 mm between-slice gap,  $3.44 \times 3.44 \times 6$  mm<sup>3</sup>, matrix size =  $64 \times 64 \times 15$ , 1,850 frames, acceleration factor = 3, phase encoding shift factor = 2, with “leak block” (Cauley et al., 2014) and a GRAPPA reconstruction kernel of  $3 \times 3$ ]. Participants were instructed to close their eyes but remain awake during the functional scans. Furthermore, T1-weighted anatomical images were collected for cross-subject registration (MPRAGE, TR = 2,400 ms, TE = 2.43 ms, FOV = 256 mm, TI = 1,000 ms, readout bandwidth = 180 Hz/px, voxel size =  $1 \times 1 \times 1$  mm<sup>3</sup>).

### Image Processing

To achieve consistency in data lengths, the initial 2 min of the short-TR data is discarded, yielding 8 min per run for both long- and short-TR datasets. Furthermore, as short-TR and long-TR data acquisitions differed in more than TR, we created a downsampled version of the short-TR data to specifically target the effect of sampling rate. This was done by temporally decimating the original short-TR data to (2,000 ms/323 ms) times the original sample rate, so as to match the sampling interval of the “long-TR” data. Resting-state fMRI processing was carried out using FMRIB software library (FSL, publicly available at [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)). The preprocessing pipeline included motion correction (Jenkinson et al., 2002), brain extraction (Smith, 2002), spatial smoothing (10 mm FWHM), frequency filtering (see section *Resting-State fMRI Measures* for details) and regression of six motion parameters. Time-locked cardiac and respiratory effects were also removed using RETROICOR (Glover et al., 2000) implemented in AFNI (AFNI: <http://afni.nimh.nih.gov/afni>).

### Physiological Monitoring and Correction

The details on measuring, modeling, and correcting for the physiological signals are explained in our previous paper (Golestani et al., 2015). In short, we accounted for the effects of the following three physiological signals:

- *Cardiac-rate variation (CRV)*: The cardiac signal was recorded using the scanner’s built-in pulse oximeter, connected to the subject’s index finger. CRV is defined as the time interval between consecutive R peaks, averaged in a 4-s window (Chang et al., 2009).
- *Respiration volume per unit time (RVT)*: The respiratory-depth correlated signal was recorded using an elastic belt connected to a BioPac system (BioPac, Goleta, USA), placed just below the subject’s ribcage (Birn et al., 2006; Chang and Glover, 2009; Golestani et al., 2015). A piezoelectric sensor within the belt measured the extent of extension and contraction in the belt caused by exhalation and inhalation. It is assumed that the respiration volume is linearly related to the amplitude of the belt signal. RVT was calculated as the ratio of breathing depth (estimated from the local maxima and minima of the respiratory waveform) over a given time period (Birn et al., 2006).
- *End-tidal CO<sub>2</sub> (PETCO<sub>2</sub>)*: CO<sub>2</sub> level in the subject’s breathing was measured using a BioPac system. A mask covering the mouth and nose of the subject was connected via plastic tubing



to the BioPac's CO<sub>2</sub> sensor. PETCO<sub>2</sub> signal was computed as the breath-by-breath maxima of the CO<sub>2</sub> tracing.

At each TR, these three physiological signals were re-sampled to correspond to the sampling rate of the rs-fMRI data. Subsequently, BOLD response functions to the three physiological signals were estimated for each voxel in the brain volume, as explained in our previous work (Golestani et al., 2015). In short, the voxel-wise BOLD responses to the three physiological signals were simultaneously estimated using a Gaussian model. The estimated responses were then used to correct the effect of these physiological signals. The physiological correction involved the convolution of the physiological signals with the corresponding estimated responses, the inclusion of the convolved response into a voxel-wise linear regression and regressing out a given physiological effect of interest from the BOLD signal. In total, eight different physiological correction combinations were applied:

- “Base”: no correction;
- “PETCO<sub>2</sub>,” “CRV,” and “RVT”: each of these three settings involved correcting for only one of the three physiological effects;
- “PETCO<sub>2</sub>+CRV,” “PETCO<sub>2</sub>+RVT,” and “CRV+RVT”: each of these three settings involved correcting for a set of two physiological effects;
- “All”: whereby all of the three physiological signals were corrected for.

We did not orthogonalize the physiological signals with respect to one another as we did in our previous work (Golestani et al., 2015), as the goal is to maximally remove noise instead of estimate their response functions.

## Resting-State fMRI Measures

### Amplitude of Low-frequency Fluctuation (ALFF)

ALFF is defined as the sum of amplitudes of each voxel's signal frequency spectrum within the low-frequency range (Zang et al., 2007) and reflects the amplitude of spontaneous low-frequency fluctuations in the BOLD signal. To eliminate possible effects of low-pass filtering on the rs-fMRI frequency spectrum, datasets with no temporal filtering were used to estimate ALFF. The unfiltered rs-fMRI signal is transformed into the frequency domain using the Fourier transform, and the spectrum in the frequency range of 0.01–0.1 Hz is averaged to calculate ALFF. The Resting-state fMRI Data Analysis Toolkit (REST V1.8, publicly available at <http://restfmri.net>; Song et al., 2011) was used to calculate the ALFF maps. To allow direct comparison of ALFF values generated using long- and short-TR data, each ALFF map was normalized (subtracting the global mean then dividing by the global standard deviation; Xi et al., 2012). This normalization eliminates biases from inter-subject ALFF variability caused by differences in imaging parameters (such as sampling-rate and flip angle) between long- and short-TR acquisitions.

### Regional Homogeneity (ReHo)

ReHo is defined as the Kendall's coefficient of concordance between a given voxel and its 27 neighboring voxels (Zang

et al., 2004) and represents the synchronization between the time series of a given voxel and its neighbors. This measure was also calculated using the REST toolkit. The long-TR data is spatially resampled to the same resolution as the short-TR data prior to ReHo calculations. The rs-fMRI time series was high-pass filtered (to >0.01 Hz) and low-pass filtered (<0.1 Hz) prior to the computations.

### Functional Connectivity: Motor Network

The motor network was the first to be demonstrated using rs-fMRI, as found in the seminal work by Biswal et al. (1995). It can easily be validated based on anatomical landmarks, and the BOLD signal in this region has been shown more affected by respiratory modulations (Birn et al., 2008) than in many other brain regions, including the default-mode network. To simplify the delineation of the motor network, we used seed-based analysis. That is, an ROI with radius of 4 mm was generated over the left motor cortex based on documented coordinates (Van Dijk et al., 2010). The average signal from this motor seed was used to generate correlation-based motor network connectivity maps. These connectivity scores were then corrected using the mixture-model method (Woolrich et al., 2005) as implemented in FSL. The mixture model estimates the distribution of the statistics as a mixture of a null distribution (with zero mean and unity standard deviation) and an alternative distribution. Mixture modeling is typically used when some assumptions in the statistical analysis might not be valid. Specifically, conventional assumptions about the temporal autocorrelation and noise level of the BOLD signal may not be valid in short TR images, leading to inflated z-values. Thus, we used mixture model to overcome this problem and effectively compare long- and short-TR results.

### Functional Connectivity: Default-Mode Network (DMN)

To investigate whether the effect of physiological correction is network-dependent, we also assessed the effect of the physiological correction on the connectivity of the DMN. The DMN is amongst the most widely studied networks in healthy controls (Raichle and Snyder, 2007; Buckner, 2012), and DMN connectivity has been found disrupted in several brain diseases (Buckner et al., 2008; Broyd et al., 2009; Anticevic et al., 2012; Whitfield-Gabrieli and Ford, 2012). Of particular interest to this study is the fact that the spatial pattern of the DMN overlaps with brain regions most affected by low-frequency physiological modulations, particularly RVT and PETCO<sub>2</sub> (Birn et al., 2006; Golestani et al., 2015). Therefore, we used the DMN as a test case to study the effect of the correction for physiological modulations on rs-fMRI functional connectivity (rs-fcMRI). Again, an ROI with a 4 mm radius was generated over the posterior cingulate cortex (PCC) using well-documented coordinates (Van Dijk et al., 2010). The regional average signal from this seed was correlated with all other voxels to generate connectivity maps, as described earlier. As before, each statistical connectivity map was then corrected using FSL's mixture modeling (Woolrich et al., 2005).



## Test-Retest Reproducibility

The maps of all rs-fMRI measures were transformed into the MNI standard space (MNI152, Montreal Neurological Institute). For ALFF and ReHo, ICC was calculated using the maps generated from the two runs of each subject. We assessed the ICC in seven distinct brain networks as defined in the work of Yeo et al. (2011). One realization of the atlas is loosely organized into the visual, somato-motor, dorsal attention, ventral attention, limbic, frontoparietal, and default-mode networks. As this rs-fcMRI atlas was generated from 1,000 subjects based on the most consistent functional connectivity patterns observed across all subjects, it is henceforth referred to as the “1,000-brain atlas.” For rs-fMRI functional connectivity, we chose the motor and default-mode networks only. The following two indices were computed to provide complementary reproducibility quantification.

### Intra-class Correlation Coefficient (ICC)

The ICC is the most common reliability index in fMRI studies (Shehzad et al., 2009; Zuo et al., 2010a,b, 2012, 2013; Anderson et al., 2011b; Wang et al., 2011; Braun et al., 2012; Chou et al., 2012; Faria et al., 2012; Guo et al., 2012; Song et al., 2012; Birn et al., 2013, 2014; Bright and Murphy, 2013; Franco et al., 2013; Liao et al., 2013; Patriat et al., 2013; Wisner et al., 2013; Zhu et al., 2014). It is given by:

$$ICC = \frac{MS_b - MS_w}{MS_b + (k - 1)MS_w} \quad (1)$$

where  $MS_b$  is the inter-subject mean-squared variability,  $MS_w$  is the within-subject inter-session mean-squared variability and  $k$  is the number of runs ( $k = 2$  in our case). As the ICC is sensitive to both inter-subject and within-subject inter-session variability, changes in either of the two would alter the ICC value, which is generally categorized into five reproducibility levels: poor (0–0.2), fair (0.2–0.4), moderate (0.4–0.6), substantial (0.6–0.8), and excellent (0.8–1; Guo et al., 2012; Zuo and Xing, 2014).

### Dice Coefficient

As shown earlier, the ICC reflects the consistency of connectivity values between runs, not necessarily that of the network spatial extent, which is also an important consideration. Thus, we include the Dice Coefficient, which has been commonly used in fMRI studies to evaluate the similarity of two spatial maps (Gorgolewski et al., 2013; Wisner et al., 2013; Zhu et al., 2013; Gross and Binder, 2014). It is defined as:

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (2)$$

where  $A$  and  $B$  are the two spatial maps,  $A \cap B$  is the intersection of the two maps, and  $|A|$  is the size (i.e., the number of voxels) of map  $A$ . We computed the Dice Coefficient between two runs of each subject, with each connectivity map defined as being above a mixture model-corrected  $z$  scores of 0.5.

### Separability Index

Regarding rs-fMRI functional connectivity, maps can be evaluated based on not only reproducibility, but also on

separability. That is, if the rs-fMRI connectivity map were predominantly sensitive to brain function instead of global physiological processes, we would expect it to demonstrate strong distinction between within-network connectivity and global (between-network) connectivity. To embody these two attributes in a single metric, using the “1,000-brain” functional-network atlas framework, the separability index is defined as:

$$SI = \frac{WNC - BNC}{WNC + BNC} \quad (3)$$

where  $WNC$  is the within-network connectivity (average connectivity, e.g.,  $z$ -scores, inside the network of interest) and  $BNC$  is between-network connectivity (average connectivity between the network of interest and the remaining six networks). Separability indices for the motor network and DMN were calculated for each physiological-correction strategy and then averaged across the two runs of each subject.

### Sensitivity and Specificity

Using the 1,000-brain connectivity atlas as the pseudo ground-truth, sensitivity and specificity of the connectivity maps for each physiological correction strategy was calculated. Each connectivity map was defined with a mixture-model corrected threshold of 0.5. Sensitivity was calculated as the ratio of the number of voxels inside the network of interest that is correctly identified (true positives) over the total number of voxels in the network (true positives + false negatives). Specificity was calculated as the ratio of the number of gray-matter voxels outside the network of interest that correctly identified as non-connected (true negatives) over the total number of gray-matter voxels outside of the network (true negatives + false positives).

### Statistical Analysis

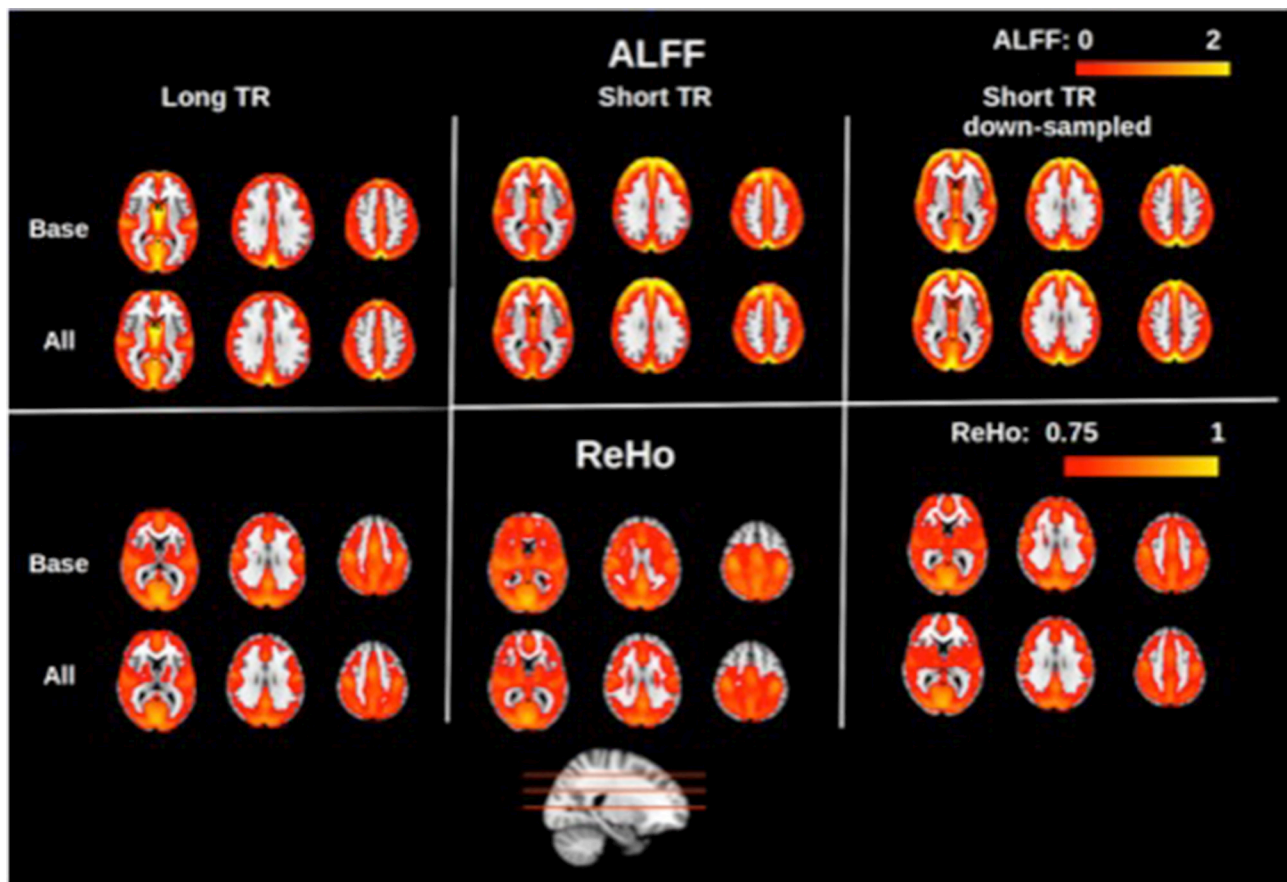
No statistical test was carried out on the ICC values, as the entire subject group would yield a single ICC value. Thus, the ICC values were simply compared among physiological correction methods and sampling rates. For functional connectivity, other measures (Dice Coefficient, separability index, sensitivity, and specificity) were compared within DMN and MN separately using two-factor within-subject ANOVA with physiological correction strategy (“Method”) and sampling-rate (TR) as factors. In case a significant effect was observed, we also assessed the observed effect by performing follow-up  $t$ -tests.

## RESULTS

One of the subjects exhibited considerable head motion during one of the rs-fMRI scans, and was thus excluded from the study, yielding total group size of 8 subjects (3 male, age  $26 \pm 6.2$  years).

### Resting-State fMRI Measures

Group-averaged maps (averaged across all subjects and both sessions) of the rs-fMRI measures are shown in **Figures 1–3**. In **Figure 1**, group-averaged normalized ALFF and ReHo are shown for long-TR, short-TR, and short-TR-down-sampled datasets. Consistent with previous studies (Zang et al., 2007; Zou et al., 2008; Zuo et al., 2010a), ALFF values are higher in the



**FIGURE 1 |** Group-averaged maps of normalized amplitude of low-frequency fluctuations (ALFF) and regional homogeneity (ReHo). Physiological correction does not have noticeable effect on the ALFF and ReHo maps, as shown by the small difference between maps generated with no physiological correction (“Base”) and with all three physiological corrections (“All”). This was the case for long-TR (TR = 2 s), short-TR (TR = 0.323 s), and down-sampled short-TR data. Gray matter is associated with considerably higher ALFF and ReHo values, specifically in the default-mode network as well as frontal and occipital regions. Sampling-rate (TR) does not considerably alters the ALFF and ReHo maps.

gray matter, specifically in the DMN as well as occipital and frontal regions. Likewise, ReHo is considerably stronger in the gray matter than in the white matter and CSF, compatible with the previous study by Long et al. (2008). For the purposes of rs-fMRI, we found ALFF and ReHo maps to be insensitive to the choice of physiological correction method, as ALFF and ReHo maps for different physiological corrections are nearly identical. To demonstrate this point, we contrast the maps derived using no physiological correction (“Base”) with those resulting from correcting for all three physiological signals (“All”). The only noticeable effect of physiological correction is a reduction in white-matter ReHo.

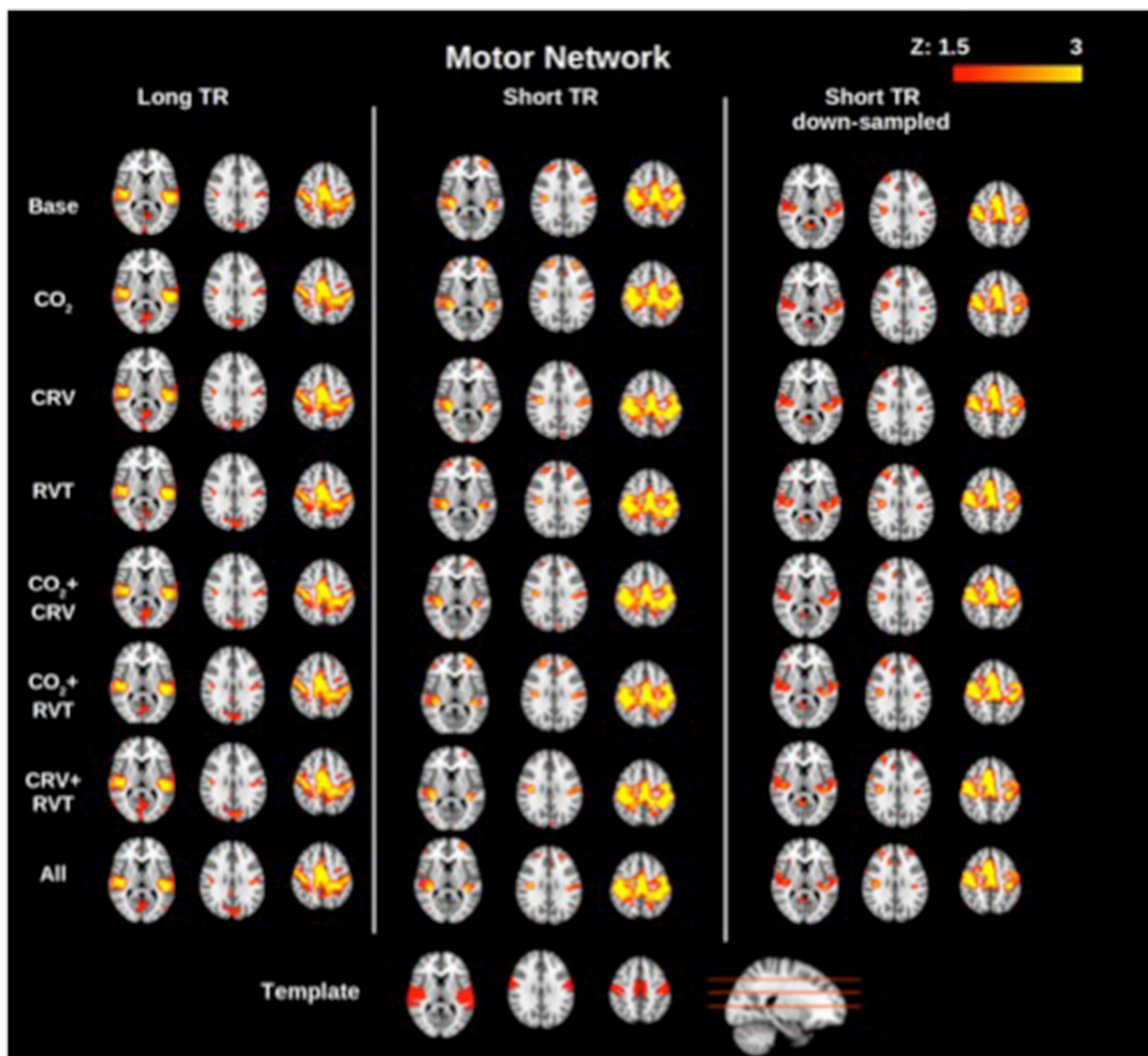
Group-average motor network and DMN connectivity maps corresponding to different physiological correction strategies are shown in **Figures 2, 3**, respectively. Correcting for PETCO<sub>2</sub> and RVT do not appear to have a considerable effect on the connectivity maps. In contrast, for both networks, the involvement of CRV correction was found to have a stronger effect, substantially reducing the size of the connected clusters, specifically outside the network of interest. The effect of

CRV correction on rs-connectivity is more evident in short-TR images. Regardless of the physiological correction method, functional networks are consistently revealed, indicating that our physiological corrections do not significantly compromise the functional information in the BOLD signal. Lastly, sampling-rate does not have considerable effect on the connectivity maps.

## Test-Retest Reproducibility

### ALFF

ICC values associated with the ALFF are shown in **Figure 4A**, generated from long-TR, short-TR, and down-sampled short-TR datasets. Results show ALFF values to be highly reproducible in all cases, with ICC values consistently in the range of 0.65–0.85. For the most part, physiological correction does not considerably alter the reproducibility of ALFF values. Nonetheless, ICC values are higher for the short-TR data. However, the fact that ICC values for down-sampled short-TR data is also higher than the long-TR, shows that the sampling-rate is not the main reason for the higher ALFF reproducibility associated with short-TR data.



**FIGURE 2 |** Group-averaged motor network (MN) connectivity maps generated with different physiological correction strategies, using long-TR, short-TR, and down-sampled short-TR data. A motor network template from the atlas generated by Yeo et al. (2011) is shown at the bottom for reference. CRV correction alters connectivity maps more than correction for PETCO<sub>2</sub> (labeled as CO<sub>2</sub>) and RVT by reducing the extent of connected clusters outside the motor cortex. Sampling-rate does not seem to have a considerable effect on the connectivity maps.

### ReHo

ICC values associated with ReHo are shown in **Figure 4B**. The ICC values (0.5–0.8) demonstrate relatively high reproducibility across different physiological correction methods and sampling-rates. Nonetheless, different trends were observed, specifically in that PETCO<sub>2</sub> and RVT correction increases the ICC whereas CRV correction decreases the ICC. Short-TR down-sampled data are consistently associated with the highest ICC values.

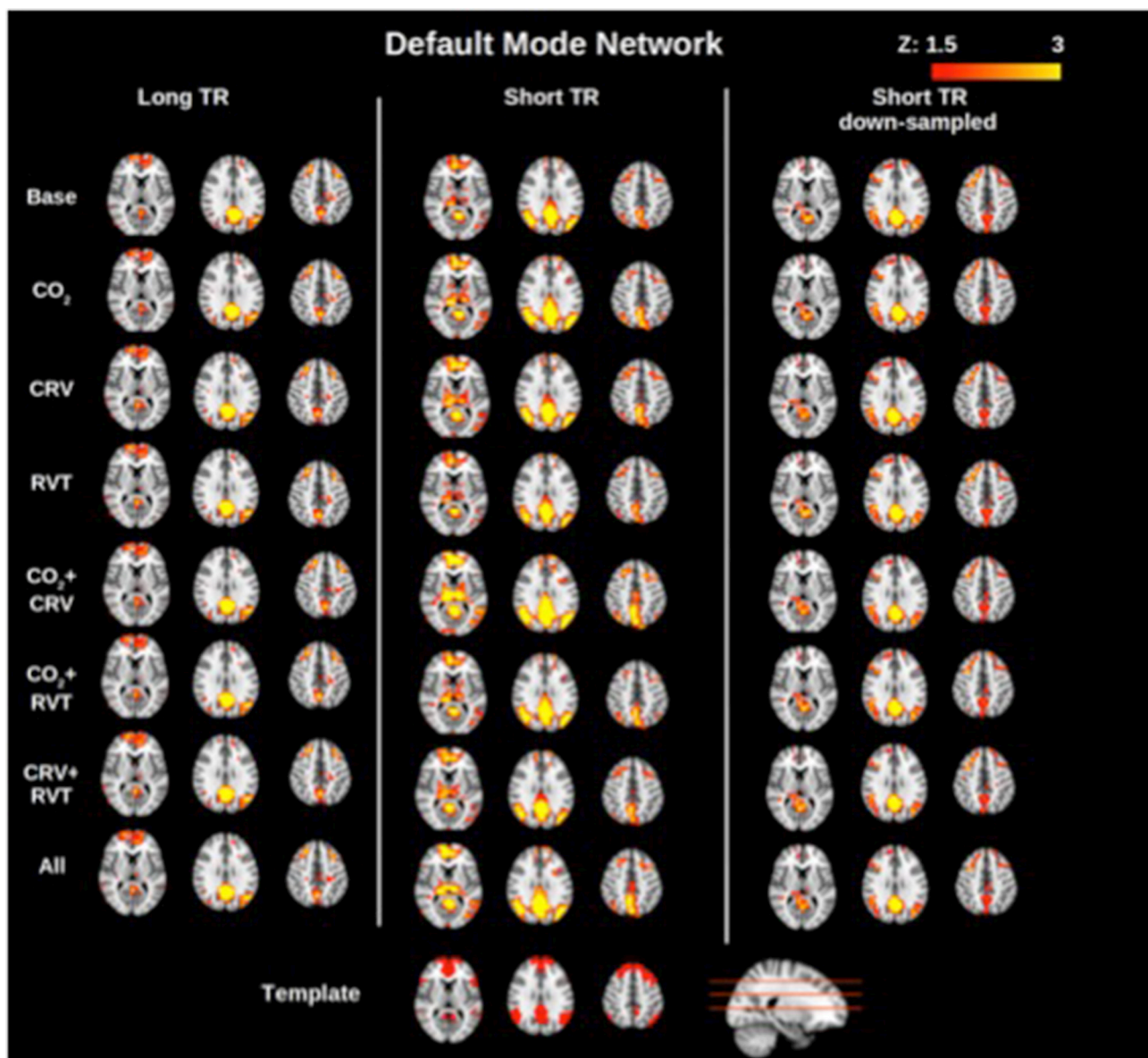
### Functional Connectivity: Reproducibility

Reproducibility measures for motor network (MN) and default-mode network (DMN) connectivity are shown in

**Figure 5**. ICC values are between 0.25 and 0.75 for the DMN, which is substantially higher than the ICC values for the MN (0.10–0.50). Further, short-TR data is associated with lower ICC values, specifically in the DMN, suggesting that data with lower sampling-rates tend to generate connectivity values with higher reproducibility. Broadly speaking, physiological correction does not appear to alter the reproducibility of the connectivity values, except in the DMN, where correction for PETCO<sub>2</sub> and CRV increase the ICC.

Spatial reproducibility of the connectivity maps is shown as Dice Coefficient plots in **Figure 5**. The Dice Coefficient is relatively high (0.7–0.8) showing that the DMN and MN are





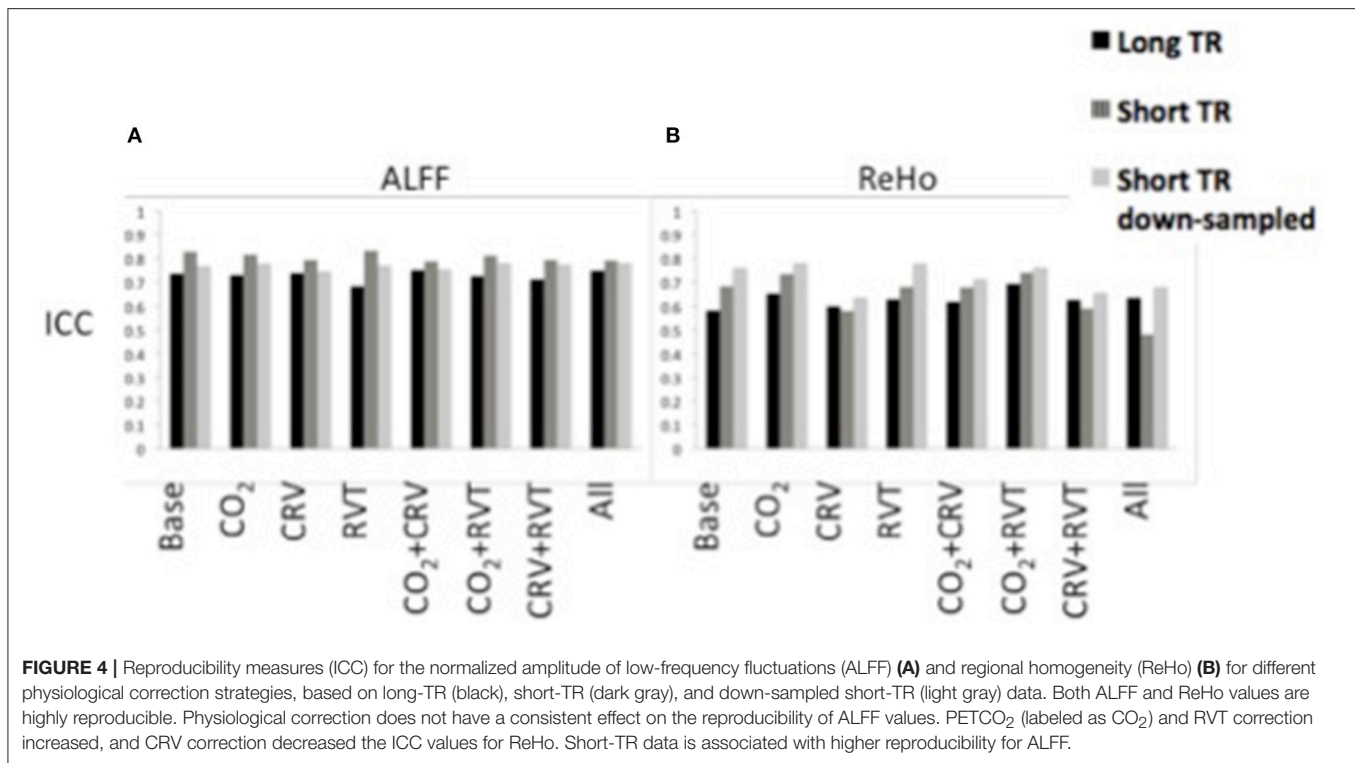
**FIGURE 3 |** Group-averaged default mode network (DMN) connectivity maps generated with different physiological correction strategies, using long-TR, short-TR, and down-sampled short-TR data. A default mode network template from the atlas generated by Yeo et al. (2011) is shown at the bottom for reference. Physiological correction does not noticeably alter connectivity maps. As in the motor network case, connectivity maps generated from data with different sampling rates are comparable.

highly reproducible spatially. The spatial pattern of the MN is slightly more reproducible than that of the DMN. However, statistical analysis (**Table 1**) demonstrated no difference between physiological correction strategies, as well as different sampling-rates.

### Separability Index

Separability indices are summarized in **Figure 6** for the MN and the DMN, generated from long-TR, short-TR, and down-sampled short-TR datasets. The MN exhibited significantly higher separability indices than the DMN, for both long- and short-TR cases ( $0.37 \pm 0.043$  for the MN vs.  $0.31 \pm$

$0.038$  for the DMN,  $p = 0.037$ ). The ANOVA (**Table 2**) showed that the sampling-rate does not significantly influence the separability index in the two networks. With respect to the choice of physiological-correction method, two-way ANOVA within the MN does not reveal any significant effects, but this is not the case for the DMN. Follow-up *t*-tests reveal that correcting for PETCO<sub>2</sub> significantly increases the separability index of the DMN. As indicated by high ICC values ( $ICC > 0.6$ ), the separability indices are highly reproducible. Also, higher sampling rate (short TR) is associated with higher reproducibility of the separability indices, specifically in the MN.



## Sensitivity and Specificity

Sensitivity and specificity of the DMN and MN are shown in **Figure 7**. The MN is associated with significantly higher detection sensitivity than the DMN ( $0.89 \pm 0.054$  for MN vs.  $0.59 \pm 0.082$  for DMN,  $p < 0.001$ ). Specificity of the DMN is higher than for the MN, although the difference is not significant ( $0.66 \pm 0.035$  for DMN and  $0.63 \pm 0.047$  for MN,  $p = 0.23$ ).

In **Table 3** we summarize the statistical test results showing that the MN maps generated from long-TR data are associated with significantly higher sensitivity than those generated from short-TR or down-sampled short-TR data. In contrast, sampling-rate does not affect sensitivity or specificity of mapping the DMN. With respect to the choice of physiological correction method, CRV correction significantly improves the sensitivity of the DMN connectivity maps, but the sensitivity of the MN is not significantly associated with any form of physiological correction.

Statistical results on the specificity of the connectivity maps are shown in **Table 4**. Sampling-rate does not affect on the specificity of the DMN and MN connectivity maps. However, CRV correction tends to reduce the specificity of the connectivity maps in both the DMN and MN, although the effect is not consistently significant.

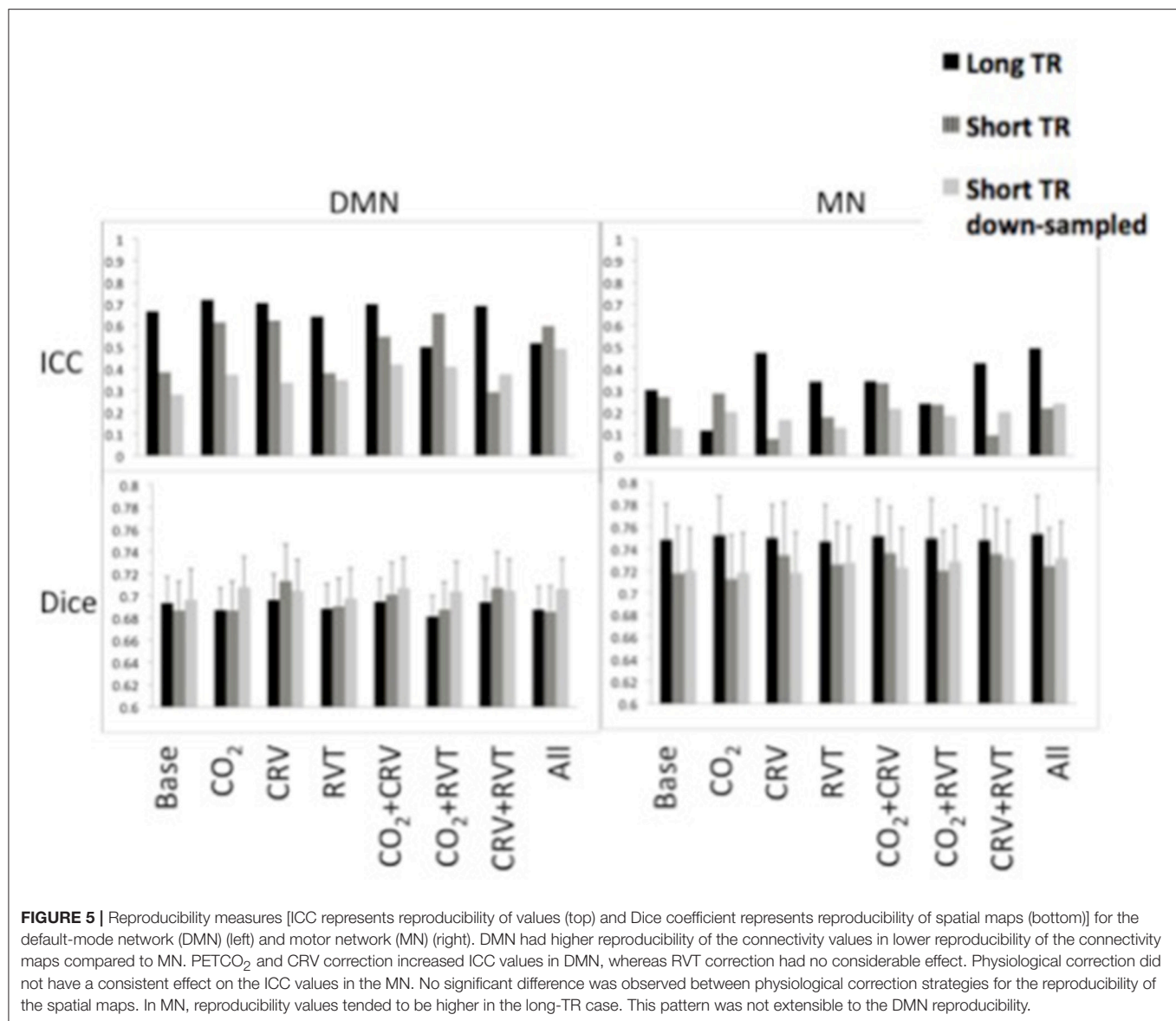
## DISCUSSION

Low-frequency physiological effects can contribute significantly to the BOLD fMRI signal and undermine the accuracy and reliability of fMRI measures, especially in the resting state. However, as mentioned earlier, despite considerable research devoted to studying the reliability of resting-state fMRI measures

(Shehzad et al., 2009; Zuo et al., 2010a,b, 2012, 2013; Anderson et al., 2011b; Wang et al., 2011; Braun et al., 2012; Chou et al., 2012; Faria et al., 2012; Guo et al., 2012; Song et al., 2012; Birn et al., 2013; Bright and Murphy, 2013; Franco et al., 2013; Liao et al., 2013; Patriat et al., 2013; Wisner et al., 2013; Zhu et al., 2014), studies that investigate the effect of physiological correction on the accuracy and reproducibility of rs-fMRI measures are extremely limited.

In this study, we determine the effect of various low-frequency physiological correction strategies on the reproducibility, sensitivity and specificity of rs-fMRI measures. The main findings are: (1) PETCO<sub>2</sub> correction has the most consistent positive effect on the reproducibility of rs-fMRI metrics; (2) PETCO<sub>2</sub> correction has the most significant positive effect on the separability of functional connectivity maps; (3) the effect of physiological correction is not influenced by fMRI data sampling rate; (4) there is substantial variability between different brain regions and networks in terms of the impact of physiological correction. Specifically: (1) Physiological correction has a stronger effect on the DMN compare to the MN; (2) CRV correction increases the reproducibility but decreases the specificity of the DMN connectivity maps; moreover, it decreases the reproducibility of the ReHo values. These findings are summarized in **Table 5**. Our findings highlight limitations in our understanding of rs-fMRI quality measures, and underscore the importance of using multiple quality measures to determine the optimal physiological correction strategy. In particular, we argue against the simplification of rs-fMRI data quality based on reproducibility alone. We discuss these findings in detail as follows.





**TABLE 1 |** Results of the statistical analysis on between-run Dice coefficient of the connectivity maps for DMN and MN.

Factor	DMN		MN	
	F	p	F	p
TR	$F_{(2, 14)} = 0.23$	0.802	$F_{(2, 14)} = 0.34$	0.718
Method	$F_{(7, 49)} = 1.44$	0.211	$F_{(7, 49)} = 0.78$	0.601

Two-factor, within-subject ANOVA test showed no significant effect of sampling rate (TR) or physiological correction strategy (Method).

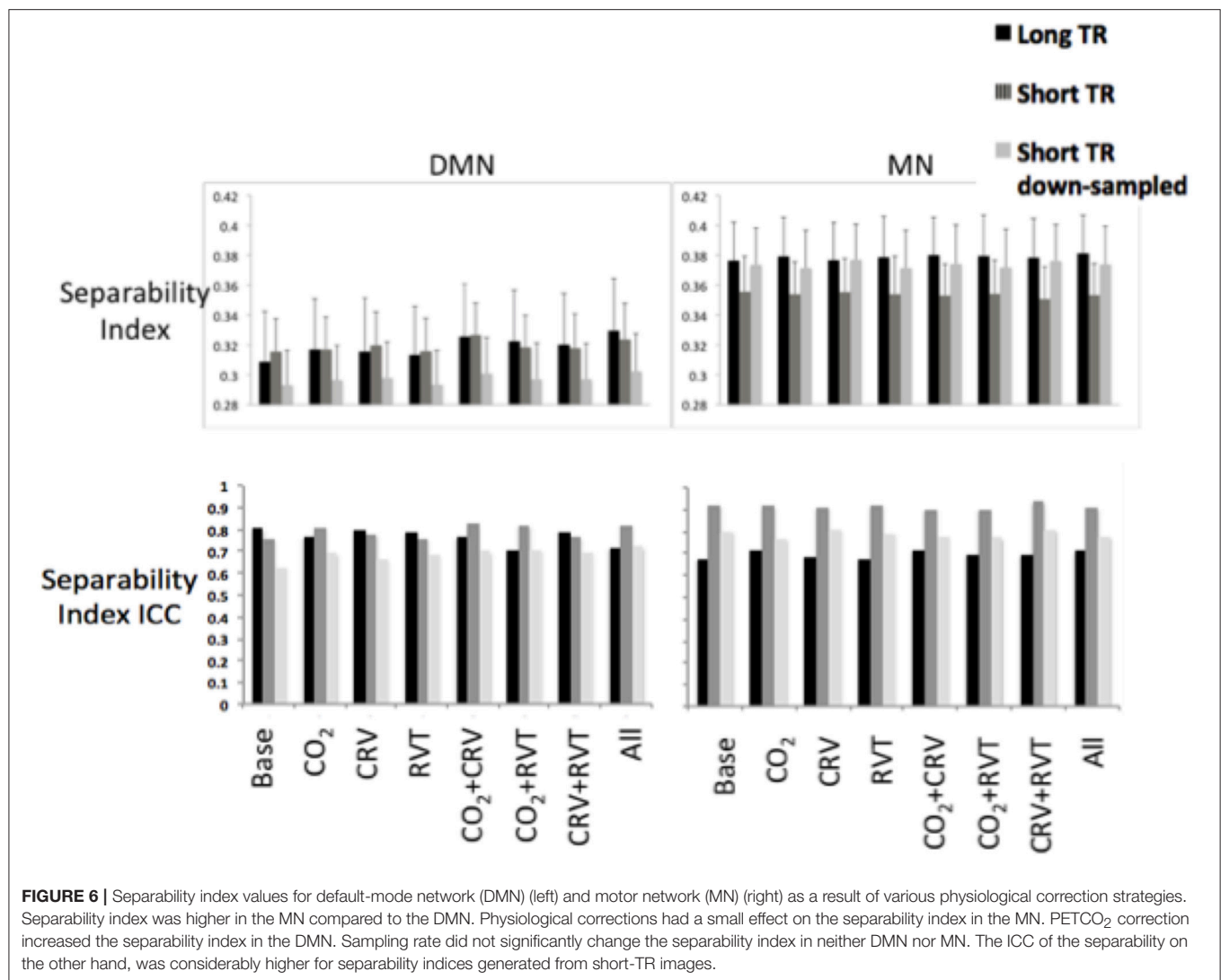
## Data Analysis Quality without Physiological Correction

### Reproducibility

We found the reproducibility of rs-fMRI measures to be highly dependent on the type of rs-fMRI measure in question. The ALFF

shows substantial reproducibility in the gray matter independent of physiological correction method and sampling rate, supported by other studies (Zuo et al., 2010a; Li et al., 2012). As the ALFF measures the power of low-frequency BOLD signal fluctuations, which presumably reflects the magnitude of neural activity (Yang et al., 2007; Zou et al., 2008; Yan and Zang, 2010), we expect to observe higher ALFF values and reproducibility in the gray matter. Likewise, as expected, ReHo values are higher in the gray matter compared to in the white matter and CSF, as ReHo is a measure of local homogeneity in brain activity, which is most meaningfully measured in the gray matter (Li et al., 2012). ReHo is also the most reproducible when based on the down-sampled short-TR data, judging from the ICC values. In fact, this could be related to the fact that the down-sampled data was associated with fewer time points and hence higher ReHo values.

Unlike for ALFF and ReHo, the reproducibility indices of which were measured in all of the seven functional networks,



**FIGURE 6 |** Separability index values for default-mode network (DMN) (left) and motor network (MN) (right) as a result of various physiological correction strategies. Separability index was higher in the MN compared to the DMN. Physiological corrections had a small effect on the separability index in the MN. PETCO<sub>2</sub> correction increased the separability index in the DMN. Sampling rate did not significantly change the separability index in neither DMN nor MN. The ICC of the separability on the other hand, was considerably higher for separability indices generated from short-TR images.

the reproducibility of rs-fMRI functional connectivity was considered within two networks of interest, namely the motor and default-mode networks, which differ vastly in terms of their cytoarchitectonic and functional traits. These networks were chosen to present a snapshot of the network-dependence of functional connectivity measures in our investigation. The moderate ICC values echo findings from previous studies (Braun et al., 2012; Franco et al., 2013; Wisner et al., 2013). Our finding of higher ICC values for the DMN compared to the MN also supports previous findings (Shehzad et al., 2009; Zuo et al., 2010b). Moreover, reproducibility of the connectivity maps measured by the Dice Coefficient is higher for the MN than for the DMN, in agreement with previous studies (Zhu et al., 2013). This suggests that although overall the connectivity values in the DMN are relatively stable, the spatial pattern changes is not as stable. Indeed, it has been reported that DMN connectivity map is more sensitive to the level of vigilance and to uncontrolled brain activations (Kucyi and Davis, 2014; Zalesky et al., 2014). As a case in point, it has been shown that (Demertzi et al.,

2011) hypnosis increases connectivity between middle frontal and angular gyri and decreases connectivity between posterior and parahippocampal structures, which are encompassed in the DMN. Moreover, sleep deprivation may cause disconnection between posterior cingulate and other nodes of the DMN (Wang et al., 2015). MN connectivity maps on the other hand are not known to be affected by factors of this nature.

### Network Separability

The motor network is more spatially separable than the DMN, as represented in **Figures 2, 3**. This point is supported by our quantitative comparison of the separability indices between the motor network and the DMN (**Figure 6**). This may be due to the simpler nature of the motor network, which makes it a good test case for methodological development. Thus, the motor-network results allowed us to establish the effect of physiological correction on the spatial pattern of rs-fcMRI measurements. Nonetheless, with regards to more complex networks such as the DMN, the interpretation of separability

**TABLE 2 |** Results of the statistical analysis on the separability index for DMN and MN.

(A)							
Factor	DMN			MN			
	<i>F</i>	<i>p</i>		<i>F</i>	<i>p</i>		
TR	$F_{(2, 14)} = 0.36$	0.706		$F_{(2, 14)} = 0.99$			0.398
Method	$F_{(7, 49)} = 4.02$	<0.001		$F_{(7, 49)} = 0.07$			0.999
(B)							
Method	CO <sub>2</sub>	CRV	RVT	CO <sub>2</sub> CRV	CO <sub>2</sub> RVT	CRVRVT	All
Base	0.011	0.149	0.245	0.012	0.015	0.113	0.043
CO <sub>2</sub>		0.778	0.275	0.071	0.337	0.661	0.165
CRV			0.324	0.001	0.635	0.753	0.109
RVT				0.025	0.025	0.175	0.052
CO <sub>2</sub> CRV					0.117	0.014	0.824
CO <sub>2</sub> RVT						0.667	0.118
CRVRVT							0.042

(A) Two-factor, within-subject ANOVA test showed the physiological correction method had significant effect on the separability index in the DMN. (B) Follow-up t-test revealed that PETCO<sub>2</sub> correction significantly increased the separability index in the DMN (green shows the cases in which the method in the column gives significantly higher values compared to the method in the row, red shows the cases in which the method in the column gives significantly smaller values compared to the method in the row).

index is less straightforward, and higher separability may not be related to higher accuracy. In fact, there exists episodic connectivity between the DMN and other networks (Smith et al., 2012; Bray et al., 2015), which may increase the correlation-based, overall global connectivity with the DMN. In such cases, the interpretation of functional connectivity measurements themselves becomes less well-defined, prompting us to refer to findings in the motor network for methodological clarifications. The separability index is highly reproducible in both the MN and DMN (Figure 6). This could be due to the normalization factor in the definition of the separability index. That is, relative connectivity is less sensitive to the parameters that might vary between different data acquisition sessions, including signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR; Golestani and Goodyear, 2011).

### Sensitivity and Specificity

The motor network (MN) is associated with high detection sensitivity but only moderate specificity (moderate false positives). In comparison, the detection sensitivity of the DMN is considerably lower (more false negatives). DMN connectivity maps in Figure 3 also confirm presence of false negatives in the DMN connectivity maps. This finding mirrors the DMN's low separability index and is consistent with more variable nature of the DMN (Damoiseaux et al., 2006; Kucyi and Davis, 2014; Zalesky et al., 2014). As mentioned before, the spatial pattern of the DMN maps is dynamic, and some nodes of the DMN might lose their connection to the network sporadically (Demertzi et al., 2011; Kucyi and Davis, 2014; Zalesky et al., 2014; Wang et al.,

2015). In such cases the disconnected nodes would represent as false-negatives, resulting in reduced sensitivity.

### The Effect of PETCO<sub>2</sub> Correction

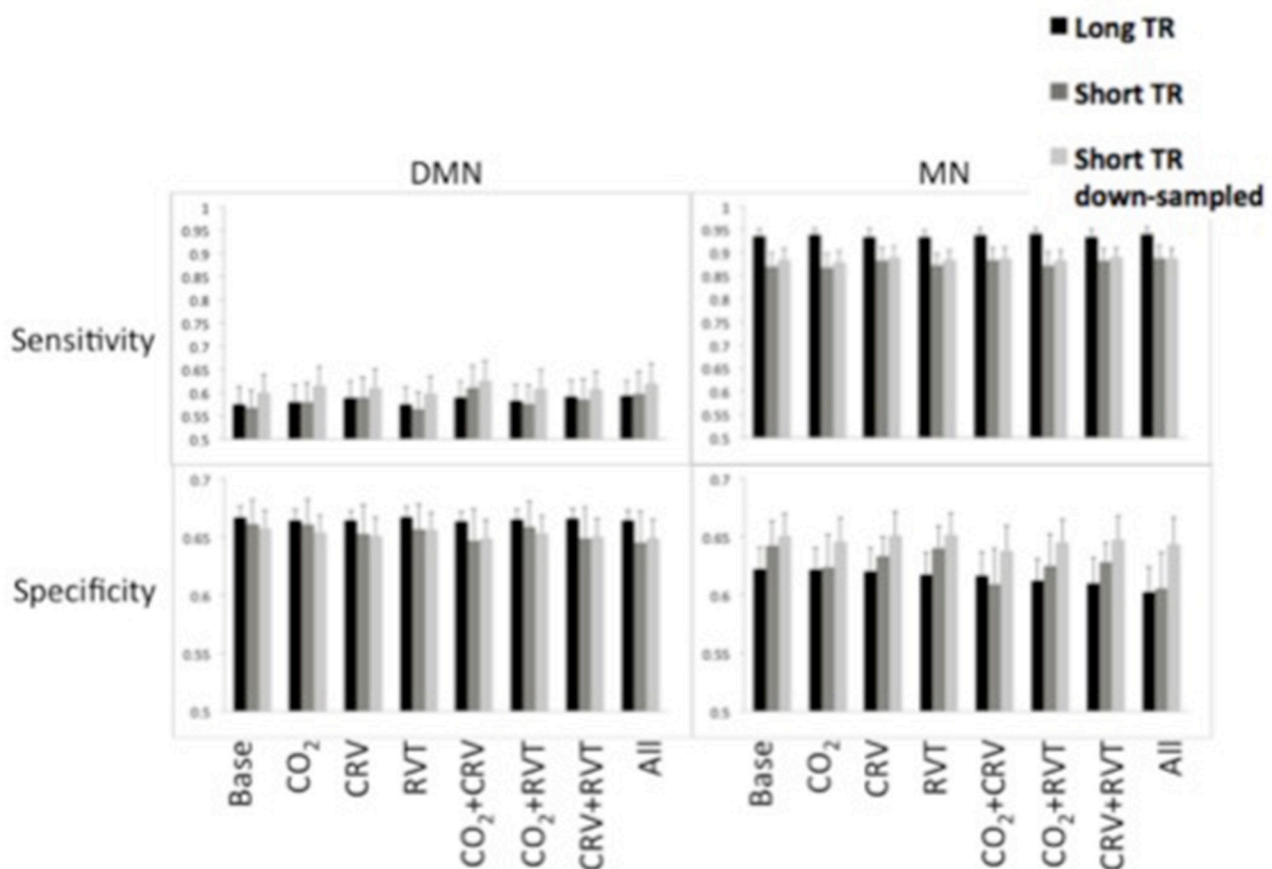
Notwithstanding inter-subject and regional differences, up to 15% of the resting-state BOLD signal is explained by PETCO<sub>2</sub> variations (Golestani et al., 2015). While this is a sizeable contribution, we do not expect that correcting for PETCO<sub>2</sub> fluctuations would dramatically change the fMRI signal. Indeed, ALFF and ReHo (Figure 1) as well as connectivity maps (Figures 2, 3) show that PETCO<sub>2</sub> correction does not qualitatively alter the spatial pattern associated with these metrics. On other hand, the fact that connectivity maps can be consistently generated using data corrected for PETCO<sub>2</sub> demonstrates that correction for PETCO<sub>2</sub> does not jeopardize the bulk of the neuronal information contained in the BOLD signal.

Quantitatively, we found PETCO<sub>2</sub> correction to slightly improve the quality of the rs-fMRI measures, although in a manner that depends on the metric and the network in question. Specifically, PETCO<sub>2</sub> correction distinctly improved the reproducibility of ReHo and DMN functional connectivity values, as well as improving the separability of the DMN. This could be taken as evidence for the successful suppression of reproducible but spurious correlation between non-connected brain regions. On the other hand, the sensitivity and specificity of the resting-state connectivity maps was relatively independent of PETCO<sub>2</sub> correction (Figure 7, Tables 3, 4). We only considered the gray-matter regions in our sensitivity and specificity calculations, as the PETCO<sub>2</sub> effect on the BOLD signal is dominant in the gray matter (Wise et al., 2004; Chang and Glover, 2009; Golestani et al., 2015). Our finding indicates that PETCO<sub>2</sub> correction may have a more global affect that does not distinguish between networks.

We note that in this study, we assume that resting-state PETCO<sub>2</sub> fluctuations are independent of neuronal activity. Indeed, elimination of the PETCO<sub>2</sub> effect did not change resting-state connectivity maps in any major way, but such an assumption may not always hold. In fact, previous studies have shown that the level of arousal is associated with both neural activity level and PETCO<sub>2</sub> level (Dahan and Teppema, 2003; Kotajima et al., 2005). Even a subtle difference in the resting state, such as between eyes-open and eyes-closed states, can alter the vascular reactivity to PETCO<sub>2</sub> (Peng et al., 2013). Nevertheless, the fact that PETCO<sub>2</sub> correction does not considerably alter the rs-fMRI maps shows the possible interaction between PETCO<sub>2</sub> signal and brain activation is not considerable, at least in our experiment. On the other hand, while the improvements in reproducibility brought about by PETCO<sub>2</sub> correction are fairly consistent, one must bear in mind that reproducibility may not always be the best aim, given the natural neural variabilities that were discussed earlier.

### Effects of CRV and RVT Correction

CRV correction appears to improve the quality of resting-state connectivity maps, specifically in the DMN. Many of the brain regions affected by CRV located in the realm of the DMN (Chang et al., 2009; Golestani et al., 2015), e.g., the



**FIGURE 7 |** Sensitivity (top) and specificity (bottom) measures of the default-mode network (DMN) (left) and motor network (MN) (right) for different physiological correction strategies, based on long-TR (black), short-TR (dark gray), and down-sampled short-TR (light gray) data. DMN demonstrated lower sensitivity and higher specificity compared to MN. CRV correction increased sensitivity and decrease specificity of DMN connectivity maps. Physiological correction did not considerably alter the sensitivity of MN connectivity maps and tended to reduce their specificity. For the MN, long-TR data was associated with higher sensitivity and lower specificity.

PCC, medial frontal cortex, and the angular gyrus. The effect of CRV correction on the resting-state connectivity is more pronounced for the short-TR images. Other studies (Faraji-Dana et al., accepted) have also reported stronger CRV contribution to multiband EPI data, as compared to conventional EPI data, the mechanism however is not clear. The reproducibility of the ReHo decreased after CRV correction. On the other hand, the reproducibility of DMN connectivity improved after CRV correction, which is apparently contrary to previous findings by Birn et al. (2014). The most likely explanation for this discrepancy is the fact that we performed voxel-wise estimation of the CRV and RVT response and correction for its effect, whereas Birn et al. (2014) used either no convolution with a response function or in some cases a single global response function averaged across several subjects. As shown in a number of recent works (Falahpour et al., 2013; Cordes et al., 2014; Golestani et al., 2015), inter-subject and inter-regional variance in the CRV and RVT response function is significant, and should be accounted for in physiological corrections. Accurately estimating and eliminating

CRV effect removes a source of signal modulation irrelevant to brain connectivity and generates more reproducible connectivity values.

Despite its positive effect on the reproducibility of DMN connectivity values, CRV correction reduces the specificity of the DMN connectivity maps (Figure 7, Table 4). A potential reason is that the CRV effects may have been over-corrected in regions exhibiting lower CRV dependence, resulting in additional artificial correlations. Therefore, we posit that although correcting for CRV improves measurement reproducibility, it can lead to lower specificity, potentially compromising the accuracy of DMN maps. On the other hand, the 1,000-brain atlas, which served as reference, was generated without CRV correction, and therefore potentially contains CRV-related biases. Nonetheless, CRV correction does not significantly affect motor network connectivity.

In the case of RVT correction, no consistent or significant effect on the rs-fMRI reproducibility is found in our study. This is also in contrast to what has been reported by Birn et al. (2014),

**TABLE 3 |** Results of the statistical analysis on the sensitivity of DMN and MN connectivity maps.

(A)							
Factor	DMN			MN			
	F	p		F	p		
TR	$F_{(2, 14)} = 0.22$	0.802		$F_{(2, 14)} = 5.38$	0.018		
Method	$F_{(7, 49)} = 6.13$	<0.001		$F_{(7, 49)} = 1.77$	0.114		
(B)							
Method	CO <sub>2</sub>	CRV	RVT	CO <sub>2</sub> CRV	CO <sub>2</sub> RVT	CRVRVT	All
Base	0.785	0.010	0.505	0.009	0.212	0.040	0.050
CO <sub>2</sub>		0.434	0.045	0.023	0.181	0.606	0.081
CRV			0.006	0.042	0.272	0.509	0.381
RVT				0.005	0.104	0.013	0.027
CO <sub>2</sub> CRV					0.015	0.010	0.315
CO <sub>2</sub> RVT						0.384	0.037
CRVRVT							0.216
(C)							
TR	Short TR			Resampled			
Long TR			0.038				0.048
Short TR							0.519

(A) Two-factor, within-subject ANOVA test showed the physiological correction method had significant effect on the sensitivity of the DMN. In addition, sampling rate had significant effect on the sensitivity of the MN. (B) Follow-up t-test on the effect of physiological correction method showed that CRV correction significantly increased the sensitivity of DMN connectivity maps, (C) t-tests on the effect of the sampling-rate demonstrated that long-TR data generated MN connectivity maps with higher sensitivity compared to short-TR and down-sampled short-TR data. (Green shows the cases in which the method in the column gives significantly higher values compared to the method in the row, red shows the cases in which the method in the column gives significantly smaller values compared to the method in the row).

whereby RVT correction decreased ICC values. Apart from potential between-subject and between-region differences in the RVT responses that explained before, differences in the resting-state paradigm may also be the cause of this discrepancy. We used eyes-closed resting-state, whereas Birn et al. (2014) used an eyes-open resting-state paradigm. Indeed, recent studies have demonstrated that the RVT signal and hence its effect on the BOLD signal differ between eyes-open and eyes-closed conditions (Yuan et al., 2013). On the other hand, RVT correction does not have a considerable impact on the separability, sensitivity, or specificity of connectivity maps. This is likely due to the rather global nature of RVT effects on the BOLD signal. RVT correction likely eliminates a synchronously oscillating part of the BOLD signal not only from voxels inside DMN but also from elsewhere in the gray matter. Consequently, RVT correction reduces both within- and between-network correlations.

## The Effect of rs-fMRI Sampling Rate

We also targeted the effect of sampling rate using long- and short-TR data. To achieve higher temporal signal to noise ration

**TABLE 4 |** Results of the statistical analysis on the specificity of the DMN and MN connectivity maps.

(A)							
Factor	DMN			MN			
	<i>F</i>	<i>p</i>		<i>F</i>	<i>p</i>		
TR	$F_{(2, 14)} = 0.24$	0.794		$F_{(2, 14)} = 1.13$	0.349		
Method	$F_{(7, 49)} = 2.61$	0.022		$F_{(7, 49)} = 2.99$	0.011		
(B)							
Method	CO <sub>2</sub>	CRV	RVT	CO <sub>2</sub> CRV	CO <sub>2</sub> RVT	CRVRVT	All
Base	0.431	0.088	0.490	0.090	0.359	0.153	0.099
CO <sub>2</sub>		0.248	0.886	0.090	0.666	0.238	0.135
CRV			0.157	0.279	0.346	0.633	0.351
RVT				0.059	0.462	0.096	0.040
CO <sub>2</sub> CRV					0.095	0.349	0.902
CO <sub>2</sub> RVT						0.240	0.078
CRVRVT							0.251
(C)							
Method	CO <sub>2</sub>	CRV	RVT	CO <sub>2</sub> CRV	CO <sub>2</sub> RVT	CRVRVT	All
Base	0.182	0.268	0.447	0.068	0.126	0.037	0.048
CO <sub>2</sub>		0.537	0.351	0.024	0.251	0.741	0.022
CRV			0.727	0.152	0.341	0.244	0.110
RVT				0.106	0.165	0.027	0.057
CO <sub>2</sub> CRV					0.096	0.405	0.274
CO <sub>2</sub> RVT						0.858	0.011
CRVRVT							0.193

(A) Two-factor, within-subject ANOVA test showed the physiological correction method had significant effect on the specificity of DMN and MN. (B) Follow-up t-test on the effect of physiological correction on the DMN specificity. Although not significant, CRV correction tends to deteriorate the DMN specificity. The only observed significant difference was between correction for all three physiological signals and correction for RVT, where correction for all of the physiological signals decreased the specificity of the DMN. (C) Follow-up t-test on the effect of physiological correction on the MN specificity demonstrated that in general, correction for the CRV effect decreased the specificity of the MN. (Red shows the cases in which the method in the column gives significantly smaller values compared to the method in the row).

(tSNR), we used lower flip angle (FA) for short-TR data. A lower FA however not only increases tSNR, but also reduces the effect of physiological noise on the BOLD signal (Gonzalez-Castillo et al., 2011). Moreover, to achieve whole-brain coverage with short-TR acquisitions, we used slightly thicker slice thickness, which might change the through-plane smoothness of images. Furthermore, SMS has also been known to introduce “leakage effects” that may introduce false correlations across slices (Todd et al., 2016), which may bias our findings and are unrelated to sampling rate *per-se*, as we have shown in our previous work (Faraji-Dana et al., accepted). Thus, to determine if any observed difference between the results from short- and long-TR is due to sampling rate and not to other imaging parameters, we created downsampled short-TR data and investigated the effect of sampling rate by comparing the results between the short-TR and downsampled short-TR data.



**TABLE 5 |** Summary of the effects of different physiological corrections on rs-fMRI measures: (A) Reproducibility of ALFF and ReHo, (B) rs-connectivity measures in the default mode network, (C) rs-connectivity measures in the motor network.

(A)					
ICC		ALFF			ReHo
CO <sub>2</sub>		↔			↑
CRV		↔			↓
RVT		↔			↑
(B)					
DMN	ICC	Dice	Separability	Sensitivity	Specificity
CO <sub>2</sub>	↑	↔	↑	↔	↔
CRV	↑	↔	↔	↑	↓
RVT	↔	↔	↔	↔	↔
(C)					
MN	ICC	Dice	Separability	Sensitivity	Specificity
CO <sub>2</sub>	↔	↔	↔	↔	↔
CRV	↔	↔	↔	↔	↓
RVT	↔	↔	↔	↔	↔

(Green arrow shows the cases in which the rs-fMRI measure is increased due to the correction method, red arrow shows cases in which the rs-fMRI measure is decreased due to the correction method).

While RETROICOR correction, which targets the time-locked high-frequency noise components, has demonstrated little effect on the ICC values of functional connectivity measures (Birn et al., 2014), we hypothesized that the effect of correction for time-locked respiration and cardiac effects may be different for long- and for short-TR data. That is, the fundamental frequency peaks of the time-locked effects are captured in the short-TR data and can be directly filtered out, whereas in the long-TR data the effects alias into lower frequencies and become irremovable. Moreover, RETROICOR does not completely remove the time-locked physiological effects even in the absence of aliasing (Golestani et al., 2015). Therefore, it is plausible that long-TR data are more affected by time-locked cardiac and respiratory signals than short-TR data. Alternatively, one may argue that the short-TR data contains more time points, resulting in statistically stronger rs-fMRI maps and potentially higher reproducibility. To our surprise, the effect of sampling rate on rs-fMRI measures and the effectiveness of physiological correction was not as strong as hypothesized. A higher sampling-rate however, appears to improve the reproducibility of some fMRI measures. More specifically, images with higher sampling-rate have more reproducible separability index, specifically in the motor network. Moreover, consistent with the previous study by Zuo et al. (2013), ReHo maps generated from short-TR data is substantially more reproducible than those generated from long-TR data.

Interestingly, the sensitivity of the MN connectivity maps generated from long-TR data is significantly higher than that

associated with short-TR data (Table 3). However, this cannot be directly attributed to sampling-rate, as the sensitivity of the MN connectivity map generated from down-sampled short-TR data is comparable to that of the maps generated from short-TR data (Figure 7). Therefore, other imaging parameter differences might have contributed to the observed phenomenon. For instance, the short-TR data (and by extension in the down-sampled short-TR data) were acquired using a lower flip angle, which is likely to have reduced the fMRI signal to noise ratio (SNR; Gonzalez-Castillo et al., 2011), reducing BOLD signal sensitivity.

We note that existing reproducibility studies assume that the true resting-state connectivity should be stable within subjects and therefore reproducible, whereas noise and artifacts should be more random in nature and hence their elimination would improve reproducibility. However, recent studies have shown that resting-state connectivity is dynamic and variable with time (Chang and Glover, 2010; Schaefer et al., 2014). Moreover, as physiological components in the fMRI data are in fact associated with moderate within-session ICC (Zuo et al., 2010b; Birn et al., 2014), their elimination from the fMRI signal may reduce both intra- and inter-subject variance and hence will affect the ICC in an unpredictable manner. For instance, an ICC reduction could be interpreted as either increased within-subject variance or decreased between-subject variance (Birn et al., 2014). Moreover, the ICC is known to be sensitive to the data range (Muller and Butner, 1994; Lee et al., 2012), and a larger dynamic range is associated with a higher ICC value. This is in fact a limitation of the general practice of using ICC alone to assess reproducibility, and supports our argument that higher reproducibility of rs-fMRI measures does not necessarily translate to higher rs-fMRI measurement accuracy.

In our analyses, we excluded one participant with excessive head motion. The fMRI images from the remaining participants underwent typical motion correction steps (affine motion correction and regression of 6 motion parameters). Recent studies have shown that even a small head motion can create spurious local correlation in resting-state fMRI data (Power et al., 2012). Even-though we did not explicitly correct for such minute motion, we believe our findings are not influenced by head motion or the choice of motion-correction strategy, as the study design uses each data set as its own reference. That is, we assess the impact of physiological corrections only, and do not compare across data sets that may have had different motion contributions or motion correction.

## Limitations

We recognize a number of limitations of this study, many of which are limitations in the field in general.

In the effort to better characterize fMRI data quality, we additionally measured the sensitivity, specificity and separability of the connectivity maps using the 1,000-brain functional-connectivity atlas as pseudo-ground-truth. In doing so, however, we assumed negligible between-subject variability in the spatial pattern of the rs-fMRI connectivity maps. Moreover, we assumed that physiological effects that are more global in nature do not closely reflect neuronal signaling. However, this assumption may only be appropriate in specific networks, such as those related

to lower-level brain function (Anderson et al., 2011a; Manoliu et al., 2013), such as the motor network. Another concern is that when estimating network spatial extent, the necessary z-score thresholding may have affected the outcome (Bennett and Miller, 2010), particularly for the Dice Coefficient, sensitivity, and specificity measures. Here too, we hope that by interpreting our findings based on multiple quality-assessment metrics, we are providing a more complete and less biased picture. To further this line of research, we feel that experimental designs that involve alternative measures of neuronal communication are the most promising avenue.

Under the heading “*The Effect of PETCO<sub>2</sub>*,” we discussed possible interaction between the neuronal activity and PETCO<sub>2</sub> fluctuations. Similar interactions may apply to CRV and RVT, as suggested by a number of previous studies (Shea, 1996; Birn et al., 2009; Macefield, 2009). While these studies and our own have recommended the removal of global physiological signals to improve the reliability of rs-fMRI measures, the relationship between these signals and rs-fMRI signal is still actively investigated.

Moreover, in this study, we used eyes-closed resting-state paradigm. Resting-state connectivity is shown to be more reliable during eyes-open condition (Patriat et al., 2013). Further studies are required to investigate if physiological correction would have different effects on eyes-open vs. eyes-closed fMRI data. As we were not able to gauge the participants’ wakefulness, we are unable to comment on the effect of the vigilance variability in our findings. Notwithstanding, investigating the influence of resting-state condition and arousal level on the physiological artifact correction is part of our future work.

While we used a relatively small sample size ( $N = 8$ ), such sizes are not uncommon amongst fMRI reproducibility studies. For instance, relevant previous studies have used sample sizes of 8 (Chou et al., 2012), 10 (Caceres et al., 2009), 18 (Meindl et al., 2010), 20 (Faria et al., 2012), 22 (Li et al., 2012), and 25 (Birn et al., 2014), respectively, for assessing reproducibility.

Finally, in this work we only investigated functional connectivity within the motor network and the DMN. We chose the DMN because it is strongly affected by physiological signals, specifically by RVT (Birn et al., 2006), and we chose the motor network in part due to its robustness and simplicity (Biswal et al., 1995; Yousry et al., 1995). As stated earlier, these two networks have been better studied and arguably better understood than most of the others in our 7-network template, and our choice is

meant to provide a snapshot of the network-dependence in our measures. However, we recognize that further work is required to thoroughly investigate the effect of physiological correction on resting-state networks in general. This goal would require a better understanding of the neuronal significance of the physiological processes.

## CONCLUSION

In this paper, we investigated the influence of correction for three low-frequency physiological modulations (i.e., PETCO<sub>2</sub>, CRV, RVT) on resting-state fMRI measurements, namely the amplitude of low-frequency fluctuations (ALFF), regional homogeneity (ReHo), and functional connectivity. To that end, we assessed metrics of test-retest reliability, network separability, measurement sensitivity, and specificity. We found that the effect of physiological correction on rs-fMRI measures is network-dependent. First, PETCO<sub>2</sub> correction improved reproducibility and separability of DMN connectivity, with negligible effect on the motor network. Secondly, CRV correction improved the reproducibility but reduced the specificity of DMN connectivity maps. Overall, the motor networks appears to be less sensitive to the choice of physiological correction than the DMN. Based on these general findings, we conclude that the interaction between the rs-fMRI signal and physiological signals is complex and not easily demonstrated. Furthermore, to evaluate the extent of improvement resulting from physiological measures, multiple and complementary metrics should be employed. While further research is necessary to clarify the mechanisms of interactions between BOLD and physiological signals, we suggest correcting for the physiological effects in rs-fMRI studies when possible.

## AUTHOR CONTRIBUTIONS

AG and JC: Designed the study. AG, JK, and YK: Collected the data. AG: Analyzed data. AG and JC: Interpreted the data. AG and JC: Drafted the article.

## ACKNOWLEDGMENTS

This work was supported by grant funding from the National Sciences and Engineering Council of Canada (NSERC, App. 418443, JC) and the Canadian Institutes of Health Research (CIHR). We acknowledge support by the Sandra A. Rotman Program.

## REFERENCES

- Amemiya, S., Kunimatsu, A., Saito, N., and Ohtomo, K. (2014). Cerebral hemodynamic impairment: assessment with resting-state functional MR imaging. *Radiology* 270, 548–555. doi: 10.1148/radiol.13130982
- Anderson, J. S., Ferguson, M. A., Lopez-Larson, M., and Yurgelun-Todd, D. (2011a). Connectivity gradients between the default mode and attention control networks. *Brain Connect.* 1, 147–157. doi: 10.1089/brain.2011.0007
- Anderson, J. S., Ferguson, M. A., Lopez-Larson, M., and Yurgelun-Todd, D. (2011b). Reproducibility of single-subject functional connectivity measurements. *AJNR Am. J. Neuroradiol.* 32, 548–555. doi: 10.3174/ajnr.A2330
- Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X. J., and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* 16, 584–592. doi: 10.1016/j.tics.2012.10.008
- Bennett, C. M., and Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Ann. N. Y. Acad. Sci.* 1191, 133–155. doi: 10.1111/j.1749-6632.2010.05446.x
- Birn, R. M., Cornejo, M. D., Molloy, E. K., Patriat, R., Meier, T. B., Kirk, G. R., et al. (2014). The influence of physiological noise correction on test-retest reliability of resting-state functional connectivity. *Brain Connect.* 4, 511–522. doi: 10.1089/brain.2014.0284
- Birn, R. M., Diamond, J. B., Smith, M. A., and Bandettini, P. A. (2006). Separating respiratory-variation-related fluctuations from

- neuronal-activity-related fluctuations in fMRI. *Neuroimage* 31, 1536–1548. doi: 10.1016/j.neuroimage.2006.02.048
- Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., et al. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83, 550–558. doi: 10.1016/j.neuroimage.2013.05.099
- Birn, R. M., Murphy, K., and Bandettini, P. A. (2008). The effect of respiration variations on independent component analysis results of resting state functional connectivity. *Hum. Brain Mapp.* 29, 740–750. doi: 10.1002/hbm.20577
- Birn, R. M., Murphy, K., Handwerker, D. A., and Bandettini, P. A. (2009). fMRI in the presence of task-correlated breathing variations. *Neuroimage* 47, 1092–1104. doi: 10.1016/j.neuroimage.2009.05.030
- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., et al. (2012). Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 59, 1404–1412. doi: 10.1016/j.neuroimage.2011.08.044
- Bray, S., Arnold, A. E., Levy, R. M., and Iaria, G. (2015). Spatial and temporal functional connectivity changes between resting and attentive states. *Hum. Brain Mapp.* 36, 549–565. doi: 10.1002/hbm.22646
- Bright, M. G., and Murphy, K. (2013). Reliable quantification of BOLD fMRI cerebrovascular reactivity despite poor breath-hold performance. *Neuroimage* 83, 559–568. doi: 10.1016/j.neuroimage.2013.07.007
- Broyd, S. J., Demanuele, C., Debener, S., Helps, S. K., James, C. J., and Sonuga-Barke, E. J. (2009). Default-mode brain dysfunction in mental disorders: a systematic review. *Neurosci. Biobehav. Rev.* 33, 279–296. doi: 10.1016/j.neubiorev.2008.09.002
- Buckner, R. L. (2012). The serendipitous discovery of the brain's default network. *Neuroimage* 62, 1137–1145. doi: 10.1016/j.neuroimage.2011.10.035
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Caceres, A., Hall, D. L., Zelaya, F. O., Williams, S. C., and Mehta, M. A. (2009). Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage* 45, 758–768. doi: 10.1016/j.neuroimage.2008.12.035
- Cauley, S. F., Polimeni, J. R., Bhat, H., Wald, L. L., and Setsompop, K. (2014). Interslice leakage artifact reduction technique for simultaneous multislice acquisitions. *Magn. Reson. Med.* 72, 93–102. doi: 10.1002/mrm.24898
- Chang, C., and Glover, G. H. (2009). Relationship between respiration, end-tidal CO<sub>2</sub>, and BOLD signals in resting-state fMRI. *Neuroimage* 47, 1381–1393. doi: 10.1016/j.neuroimage.2009.04.048
- Chang, C., and Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50, 81–98. doi: 10.1016/j.neuroimage.2009.12.011
- Chang, C., Cunningham, J. P., and Glover, G. H. (2009). Influence of heart rate on the BOLD signal: the cardiac response function. *Neuroimage* 44, 857–869. doi: 10.1016/j.neuroimage.2008.09.029
- Chou, Y. H., Panych, L. P., Dickey, C. C., Petrella, J. R., and Chen, N. K. (2012). Investigation of long-term reproducibility of intrinsic connectivity network mapping: a resting-state fMRI study. *AJNR Am. J. Neuroradiol.* 33, 833–838. doi: 10.3174/ajnr.A2894
- Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., et al. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in “resting-state” data. *Am. J. Neuroradiol.* 22, 1326–1333. doi: 10.1093/cercor/bhr230
- Cordes, D., Nandy, R. R., Schafer, S., and Wager, T. D. (2014). Characterization and reduction of cardiac- and respiratory-induced noise as a function of the sampling rate (TR) in fMRI. *Neuroimage* 89, 314–330. doi: 10.1016/j.neuroimage.2013.12.013
- Dagli, M. S., Ingeholm, J. E., and Haxby, J. V. (1999). Localization of cardiac-induced signal change in fMRI. *Neuroimage* 9, 407–415. doi: 10.1006/nimg.1998.0424
- Dahan, A., and Teppema, L. J. (2003). Influence of anaesthesia and analgesia on the control of breathing. *Br. J. Anaesth.* 91, 40–49. doi: 10.1093/bja/aeg150
- Damoiseaux, J. S., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13848–13853. doi: 10.1073/pnas.0601417103
- Demertzi, A., Soddu, A., Faymonville, M. E., Bahri, M. A., Gosseries, O., Vanhaudenhuyse, A., et al. (2011). Hypnotic modulation of resting state fMRI default mode and extrinsic network connectivity. *Prog. Brain Res.* 193, 309–322. doi: 10.1016/B978-0-444-53839-0.00020-X
- Falahpour, M., Refai, H., and Bodurka, J. (2013). Subject specific BOLD fMRI respiratory and cardiac response functions obtained from global signal. *Neuroimage* 72, 252–264. doi: 10.1016/j.neuroimage.2013.01.050
- Faraji-Dana, A. G., Khatamian, Y., Graham, S., and Chen, J. (accepted). Simultaneous multi-slice resting-state fMRI at 3 tesla: slice-acceleration related biases in physiological noise effects. *Brain Connect.*
- Faria, A. V., Joel, S. E., Zhang, Y., Oishi, K., van Zijl, P. C., Miller, M. I., et al. (2012). Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. *Neuroimage* 61, 613–621. doi: 10.1016/j.neuroimage.2012.03.078
- Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Gunther, M., et al. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS ONE* 5:e15710. doi: 10.1371/journal.pone.0015710
- Franco, A. R., Mannell, M. V., Calhoun, V. D., and Mayer, A. R. (2013). Impact of analysis methods on the reproducibility and reliability of resting-state networks. *Brain Connect.* 3, 363–374. doi: 10.1089/brain.2012.0134
- Ganger, S., Hahn, A., Kublbock, M., Kranz, G. S., Spies, M., Vanicek, T., et al. (2015). Comparison of continuously acquired resting state and extracted analogues from active tasks. *Hum. Brain Mapp.* 36, 4053–4063. doi: 10.1002/hbm.22897
- Glover, G. H., Li, T. Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.* 44, 162–167. doi: 10.1002/1522-2594(200007)44:1<162::AID-MRM23>3.0.CO;2-E
- Golestani, A. M., and Goodyear, B. G. (2011). A resting-state connectivity metric independent of temporal signal-to-noise ratio and signal amplitude. *Brain Connect.* 1, 159–167. doi: 10.1089/brain.2011.0003
- Golestani, A., Chang, C., Kwint, J., Khatamian, Y., and Chen, J. (2015). Mapping the end-tidal CO<sub>2</sub> response function in the resting-state BOLD fMRI signal: spatial specificity, test-retest reliability and effect of fMRI sampling rate. *Neuroimage* 104, 266–277. doi: 10.1016/j.neuroimage.2014.10.031
- Gonzalez-Castillo, J., Roopchansingh, V., Bandettini, P. A., and Bodurka, J. (2011). Physiological noise effects on the flip angle selection in BOLD fMRI. *Neuroimage* 54, 2764–2778. doi: 10.1016/j.neuroimage.2010.11.020
- Gorgolewski, K. J., Storkey, A. J., Bastin, M. E., Whittle, I., and Pernet, C. (2013). Single subject fMRI test-retest reliability metrics and confounding factors. *Neuroimage* 69, 231–243. doi: 10.1016/j.neuroimage.2012.10.085
- Gross, W. L., and Binder, J. R. (2014). Alternative thresholding methods for fMRI data optimized for surgical planning. *Neuroimage* 84, 554–561. doi: 10.1016/j.neuroimage.2013.08.066
- Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., and Seeley, W. W. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. doi: 10.1016/j.neuroimage.2012.03.027
- Hu, X., Le, T. H., Parrish, T., and Erhard, P. (1995). Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magn. Reson. Med.* 34, 201–212. doi: 10.1002/mrm.1910340211
- Jann, K., Gee, D. G., Kilroy, E., Schwab, S., Smith, R. X., Cannon, T. D., et al. (2015). Functional connectivity in BOLD and CBF data: similarity and reliability of resting brain networks. *Neuroimage* 106, 111–122. doi: 10.1016/j.neuroimage.2014.11.028
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132
- Kotajima, F., Meadows, G. E., Morrell, M. J., and Corfield, D. R. (2005). Cerebral blood flow changes associated with fluctuations in alpha and theta rhythm during sleep onset in humans. *J. Physiol.* 568(Pt. 1), 305–313. doi: 10.1113/jphysiol.2005.092577



- Kucyi, A., and Davis, K. D. (2014). Dynamic functional connectivity of the default mode network tracks daydreaming. *Neuroimage* 100, 471–480. doi: 10.1016/j.neuroimage.2014.06.044
- Lee, K. M., Lee, J., Chung, C. Y., Ahn, S., Sung, K. H., Kim, T. W., et al. (2012). Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. *Clin. Orthop. Surg.* 4, 149–155. doi: 10.4055/cios.2012.4.2.149
- Li, Z., Kadivar, A., Pluta, J., Dunlop, J., and Wang, Z. (2012). Test-retest stability analysis of resting brain activity revealed by blood oxygen level-dependent functional MRI. *J. Magn. Reson. Imaging* 36, 344–354. doi: 10.1002/jmri.23670
- Liao, X. H., Xia, M. R., Xu, T., Dai, Z. J., Cao, X. Y., Niu, H. J., et al. (2013). Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. *Neuroimage* 83, 969–982. doi: 10.1016/j.neuroimage.2013.07.058
- Long, X. Y., Zuo, X. N., Kiviniemi, V., Yang, Y., Zou, Q. H., Zhu, C. Z., et al. (2008). Default mode network as revealed with multiple methods for resting-state functional MRI analysis. *J. Neurosci. Methods* 171, 349–355. doi: 10.1016/j.jneumeth.2008.03.021
- Macefield, V. G. (2009). Developments in autonomic research: a review of the latest literature. *Clin. Auton. Res.* 19, 193–196. doi: 10.1007/s10286-009-0024-3
- Manoliu, A., Riedl, V., Doll, A., Bauml, J. G., Muhlau, M., Schwerthoffer, D., et al. (2013). Insular dysfunction reflects altered between-network connectivity and severity of negative symptoms in Schizophrenia during psychotic remission. *Front. Hum. Neurosci.* 7:216. doi: 10.3389/fnhum.2013.00216
- Meindl, T., Teipel, S., Elmouden, R., Mueller, S., Koch, W., Dietrich, O., et al. (2010). Test-retest reproducibility of the default-mode network in healthy individuals. *Hum. Brain Mapp.* 31, 237–246. doi: 10.1002/hbm.20860
- Muller, R., and Buttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Stat. Med.* 13, 2465–2476. doi: 10.1002/sim.4780132310
- Murphy, K., Birn, R. M., and Bandettini, P. A. (2013). Resting-state fMRI confounds and cleanup. *Neuroimage* 80, 349–359. doi: 10.1016/j.neuroimage.2013.04.001
- Patriat, R., Molloy, E. K., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., et al. (2013). The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. *Neuroimage* 78, 463–473. doi: 10.1016/j.neuroimage.2013.04.013
- Peng, T., Niazy, R., Payne, S. J., and Wise, R. G. (2013). The effects of respiratory CO<sub>2</sub> fluctuations in the resting-state BOLD signal differ between eyes open and eyes closed. *Magn. Reson. Imaging* 31, 336–345. doi: 10.1016/j.mri.2012.06.013
- Pfeuffer, J. P. F., Van de Moortele, Ugurbil, K., Hu, X., and Glover, G. H. (2002). Correction of physiologically induced global off-resonance effects in dynamic echo-planar and spiral functional imaging. *Magn. Reson. Med.* 47, 344–353. doi: 10.1002/mrm.10065
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018
- Raichle, M. E., and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage* 37, 1083–1090; discussion: 1097–1089. doi: 10.1016/j.neuroimage.2007.02.041
- Raj, D., Anderson, A. W., and Gore, J. C. (2001). Respiratory effects in human functional magnetic resonance imaging due to bulk susceptibility changes. *Phys. Med. Biol.* 46, 3331–3340. doi: 10.1088/0031-9155/46/12/318
- Schaefer, A., Margulies, D. S., Lohmann, G., Gorgolewski, K. J., Smallwood, J., Kiebel, S. J., et al. (2014). Dynamic network participation of functional connectivity hubs assessed by resting-state fMRI. *Front. Hum. Neurosci.* 8:195. doi: 10.3389/fnhum.2014.00195
- Setsonpop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., and Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.* 67, 1210–1224. doi: 10.1002/mrm.23097
- Shea, S. A. (1996). Behavioural and arousal-related influences on breathing in humans. *Exp. Physiol.* 81, 1–26. doi: 10.1113/expphysiol.1996.sp003911
- Shehzad, Z., Kelly, A. M., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., et al. (2009). The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229. doi: 10.1093/cercor/bhn256
- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062
- Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., et al. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3131–3136. doi: 10.1073/pnas.1121329109
- Song, J., Desphande, A. S., Meier, T. B., Tudorascu, D. L., Vergun, S., Nair, V. A., et al. (2012). Age-related differences in test-retest reliability in resting-state brain functional connectivity. *PLoS ONE* 7:e49847. doi: 10.1371/journal.pone.0049847
- Song, X. W., Dong, Z. Y., Long, X. Y., Li, S. F., Zuo, X. N., Zhu, C. Z., et al. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS ONE* 6:e25031. doi: 10.1371/journal.pone.0025031
- Todd, N., Moeller, S., Auerbach, E. J., Yacoub, E., Flandin, G., and Weiskopf, N. (2016). Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: sensitivity and slice leakage artifacts. *Neuroimage* 124(Pt. A), 32–42. doi: 10.1016/j.neuroimage.2015.08.056
- Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009
- Wang, J., Qiu, S., Xu, Y., Liu, Z., Wen, X., Hu, X., et al. (2011). Graph theoretical analysis reveals disrupted topological properties of whole brain functional networks in temporal lobe epilepsy. *Clin. Neurophysiol.* 125, 1744–1756. doi: 10.1016/j.clinph.2013.12.120
- Wang, Y., Liu, H., Hitchman, G., and Lei, X. (2015). Module number of default mode network: inter-subject variability and effects of sleep deprivation. *Brain Res.* 1596, 69–78. doi: 10.1016/j.brainres.2014.11.007
- Whitfield-Gabrieli, S., and Ford, J. M. (2012). Default mode network activity and connectivity in psychopathology. *Annu. Rev. Clin. Psychol.* 8, 49–76. doi: 10.1146/annurev-clinpsy-032511-143049
- Wise, R. G., Ide, K., Poulin, M. J., and Tracey, I. (2004). Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *Neuroimage* 21, 1652–1664. doi: 10.1016/j.neuroimage.2003.11.025
- Wisner, K. M., Atluri, G., Lim, K. O., and Macdonald, A. W. III. (2013). Neurometrics of intrinsic connectivity networks at rest using fMRI: retest reliability and cross-validation using a meta-level method. *Neuroimage* 76, 236–251. doi: 10.1016/j.neuroimage.2013.02.066
- Woolrich, M. W., Behrens, T. E., Beckmann, C. F., and Smith, S. M. (2005). Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Trans. Med. Imaging* 24, 1–11. doi: 10.1109/TMI.2004.836545
- Xi, Q., Zhao, X. H., Wang, P. J., Guo, Q. H., Yan, C. G., and He, Y. (2012). Functional MRI study of mild Alzheimer's disease using amplitude of low frequency fluctuation analysis. *Chin. Med. J.* 125, 858–862.
- Yan, C., and Zang, Y. (2010). DPARSF: a MATLAB toolbox for “Pipeline” data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Yang, H., Long, X. Y., Yang, Y., Yan, H., Zhu, C. Z., Zhou, X. P., et al. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *Neuroimage* 36, 144–152. doi: 10.1016/j.neuroimage.2007.01.054
- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011
- Yousry, T. A., Schmid, U. D., Jassoy, A. G., Schmidt, D., Eisner, W. E., Reulen, H. J., et al. (1995). Topography of the cortical motor hand area: prospective study with functional MR imaging and direct motor mapping at surgery. *Radiology* 195, 23–29. doi: 10.1148/radiology.195.1.7892475
- Yuan, H., Zotev, V., Phillips, R., and Bodurka, J. (2013). Correlated slow fluctuations in respiration, EEG, and BOLD fMRI. *Neuroimage* 79, 81–93. doi: 10.1016/j.neuroimage.2013.04.068

- Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., and Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10341–10346. doi: 10.1073/pnas.1400181111
- Zang, Y. F., He, Y., Zhu, C. Z., Cao, Q. J., Sui, M. Q., Liang, M., et al. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev.* 29, 83–91. doi: 10.1016/j.braindev.2006.07.002
- Zang, Y., Jiang, T., Lu, Y., He, Y., and Tian, L. (2004). Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22, 394–400. doi: 10.1016/j.neuroimage.2003.12.030
- Zhu, L., Fan, Y., Zou, Q., Wang, J., Gao, J. H., and Niu, Z. (2014). Temporal reliability and lateralization of the resting-state language network. *PLoS ONE* 9:e85880. doi: 10.1371/journal.pone.0085880
- Zhu, S., Fang, Z., Hu, S., Wang, Z., and Rao, H. (2013). Resting state brain function analysis using concurrent BOLD in ASL perfusion fMRI. *PLoS ONE* 8:e65884. doi: 10.1371/journal.pone.0065884
- Zou, Q. H., Zhu, C. Z., Yang, Y., Zuo, X. N., Long, X. Y., Cao, Q. J., et al. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. *J. Neurosci. Methods* 172, 137–141. doi: 10.1016/j.jneumeth.2008.04.012
- Zuo, X. N., and Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009
- Zuo, X. N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., et al. (2010a). The oscillating brain: complex and reliable. *Neuroimage* 49, 1432–1445. doi: 10.1016/j.neuroimage.2009.09.037
- Zuo, X. N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., et al. (2012). Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875. doi: 10.1093/cercor/bhr269
- Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010b). Reliable intrinsic connectivity networks: test-retest approach. *Neuroimage* 49, 2163–2177. doi: 10.1016/j.neuroimage.2009.10.080
- Zuo, X. N., Xu, T., Jiang, L., Yang, Z., Cao, X. Y., He, Y., et al. (2013). Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage* 65, 374–386. doi: 10.1016/j.neuroimage.2012.10.017

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Golestani, Kwinta, Khatamian and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Improving the Test-Retest Reliability of Resting State fMRI by Removing the Impact of Sleep

Jiahui Wang<sup>1</sup>, Junwei Han<sup>1\*</sup>, Vinh T. Nguyen<sup>2</sup>, Lei Guo<sup>1</sup> and Christine C. Guo<sup>2\*</sup>

<sup>1</sup> School of Automation, Northwestern Polytechnical University, Xi'an, China, <sup>2</sup> QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Institute of Psychology (CAS), China

### Reviewed by:

Javier Gonzalez-Castillo,  
National Institute of Mental Health,  
USA  
Veena A. Nair,  
University of Wisconsin-Madison, USA

### \*Correspondence:

Junwei Han  
junwei.han2010@gmail.com  
Christine C. Guo  
christine.cong@gmail.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 27 February 2017

**Accepted:** 18 April 2017

**Published:** 08 May 2017

### Citation:

Wang J, Han J, Nguyen VT, Guo L  
and Guo CC (2017) Improving the  
Test-Retest Reliability of Resting State  
fMRI by Removing the Impact of  
Sleep. *Front. Neurosci.* 11:249.  
doi: 10.3389/fnins.2017.00249

Resting state functional magnetic resonance imaging (rs-fMRI) provides a powerful tool to examine large-scale neural networks in the human brain and their disturbances in neuropsychiatric disorders. Thanks to its low demand and high tolerance, resting state paradigms can be easily acquired from clinical population. However, due to the unconstrained nature, resting state paradigm is associated with excessive head movement and proneness to sleep. Consequently, the test-retest reliability of rs-fMRI measures is moderate at best, falling short of widespread use in the clinic. Here, we characterized the effect of sleep on the test-retest reliability of rs-fMRI. Using measures of heart rate variability (HRV) derived from simultaneous electrocardiogram (ECG) recording, we identified portions of fMRI data when subjects were more alert or sleepy, and examined their effects on the test-retest reliability of functional connectivity measures. When volumes of sleep were excluded, the reliability of rs-fMRI is significantly improved, and the improvement appears to be general across brain networks. The amount of improvement is robust with the removal of as much as 60% volumes of sleepiness. Therefore, test-retest reliability of rs-fMRI is affected by sleep and could be improved by excluding volumes of sleepiness as indexed by HRV. Our results suggest a novel and practical method to improve test-retest reliability of rs-fMRI measures.

**Keywords:** test-retest reliability, resting state, naturalistic paradigm, heart rate variability, sleep

## INTRODUCTION

Resting state functional magnetic resonance imaging (rs-fMRI) paradigm is a widely used tool to explore functional connectivity network in both healthy and clinical population (Biswal et al., 1995; Greicius et al., 2003; Fox et al., 2005; Greicius, 2008; Jafri et al., 2008; Fox and Greicius, 2010; van den Heuvel and Pol, 2010; Friston, 2011; Buckner et al., 2013; Tailby et al., 2015). The task-free nature of rs-fMRI paradigm, with low demand and high tolerance, makes it easy to standardize across study centers and conduct with subjects challenged by task performance (Greicius, 2008). Rs-fMRI has thus become a common tool in clinical studies on brain disorders, and holds great promise as imaging makers for diagnostic and prognostic uses. In addition to connectivity measures between individual brain regions, graph theory has been applied to rs-fMRI connectivity networks to measure higher order characteristics of brain networks, such as degree centrality, clustering coefficient, and modularity (van den Heuvel et al., 2008; Bullmore and Sporns, 2009; Guye et al., 2010; Hayasaka and Laurienti, 2010; He and Evans, 2010; Bullmore and Bassett, 2011; Zuo et al., 2012a).

Rs-fMRI measures, however, have not achieved the level of test-retest reliability as required by clinical imaging. The reliability of functional connectivity and graph measures derived from rs-fMRI ranges from poor to moderate (Telesford et al., 2010; Wang et al., 2011; Braun et al., 2012; Guo et al., 2012; Li et al., 2012; Patriat et al., 2013; Cao et al., 2014), where the unconstrained nature of resting state condition could have a negative impact. Without external stimulation, one problem with resting state paradigm is the excessive head motion and associated scan artifacts (Van Dijk et al., 2012; Yan et al., 2013; Vanderwal et al., 2015). It has been showed that excessive head motion reduces the reliability of fMRI measures and excluding high motion subject or volumes, or regressing out motion related artifacts could improve the reliability of rs-fMRI measures (Schwarz and McGonigle, 2011; Guo et al., 2012; Zuo et al., 2012b; Gorgolewski et al., 2013; Yan et al., 2013; Du et al., 2015).

Sleep was found to affect rs-fMRI measures in previous studies. It was reported that most subjects become drowsy and even fall asleep during resting state paradigms (Tagliazucchi and Laufs, 2014). These sleep episodes during resting state scanning are thought to be mostly non-rapid eye movement (non-REM) sleep, as more than 60 min are required to get into REM sleep (McCarley, 2007). The presence of sleep was found to affect functional connectivity and graph theoretical measures. For example, thalamocortical connectivity was found to reduce at the onset of non-REM sleep, and corticocortical connectivity increase during light sleep before getting disrupted during deep sleep (Massimini et al., 2005; Horowitz et al., 2009; Larson-Prior et al., 2009; Spoormaker et al., 2010; Koike et al., 2011; Tagliazucchi et al., 2012; Picchioni et al., 2014; Tagliazucchi and Laufs, 2014; Hale et al., 2016). Therefore, it seems possible that sleep could also affect the test-retest reliability of rs-fMRI measures, and excluding volumes of high sleepiness might improve the reliability of connectivity measures.

We here investigated this hypothesis using a test-retest fMRI dataset, where 17 participants underwent two identical fMRI sessions 3 months apart. To detect sleep during the scan, we used an established method based on simultaneous ECG recordings during the fMRI acquisition. It is well established that cardiac autonomic regulation alters between wake and different sleep stages (Burgess et al., 1997; Trinder et al., 2012; Tobaldini et al., 2013). Compared with wake condition, non-REM sleep often incurs a marked decrease in heart rate and increase in HRV. The changes start from sleep onset, or when subjects feel drowsy, and continue throughout the non-REM sleep stage. This suggests a general cardiovascular output reduction and a transfer from predominant sympathetic to parasympathetic cardiac modulation during non-REM sleep (Toscani et al., 1996; Elsenbruch et al., 1999; Trinder et al., 2001; Busek et al., 2005; Carrington et al., 2005; de Zambotti et al., 2011, 2014; Cabiddu et al., 2012; Boudreau et al., 2013; Chouchou and Desseilles, 2014; Cellini et al., 2016). HRV could thus be used to detect sleep or drowsiness.

In sleep studies, electroencephalogram (EEG) is recognized as gold standard to identify sleep stages (Rechtschaffen and Kales, 1968; Iber et al., 2007). Nevertheless, it is hard for subjects to fall asleep with EEG scalp on. Lv et al. identified sleep

state using HRV derived from peripheral pulse signals, and observed consistent brain network properties compared to those derived from EEG based studies (Lv et al., 2015). Moreover, HRV measures are widely used, solely or combined with other physiological signal measures, as features in machine learning models to predict and detect the fatigue and sleepiness of drivers. The classification accuracy could reach over 90% (Lal and Craig, 2001; Borghini et al., 2012; Sahayadhas et al., 2012; Abbood et al., 2014). Furthermore, compared to other biosignals used for sleep detection, such as EEG and pupillometry (Abbood et al., 2014), simultaneous recording of cardiac signals, using either ECG or pulse oximetry, is more easily and routinely implemented in fMRI experiments.

Here, we used HRV derived from the ECG to index the level of alertness and sleepiness continuously for each fMRI volume. We then examined the effect on the test-retest reliability of connectivity measures when the volumes of the most extreme HRV values were excluded. To derive a more general conclusion, we used two different HRV measurements—the root mean square of successive difference of normal-to-normal intervals (RMSSD) (Neumann et al., 1941; Malik, 1996) and cardiac vagal index (CVI) (Toichi et al., 1997) to index the level of sleep, independently, and assessed test-retest reliability at both individual unit- and scan-wise levels (Guo et al., 2012).

## MATERIALS AND METHODS

### Participants

Twenty right-handed participants (11 females, 9 males; aged between 21 and 31 years; mean age  $27 \pm 2.7$  years) participated in the study. The participants were recruited from the University of Queensland and provided written informed consent. Participants received a small monetary compensation (\$50) for their participation in the study. The study was approved by the human ethics research committee of the University of Queensland and was conducted according to National Health and Medical Research Council guidelines.

### Experimental Paradigm

The experiment comprised two scanning sessions. For each session, participants underwent an 8-min resting state fMRI exam with eyes closed, and then freely viewed a 20-min short movie “*The Butterfly Circus*.” Resting state condition was always acquired first to avoid potential effect of movie viewing experience on resting state brain activity, and also to reduce the likelihood of fatigue and sleep during resting state. *The Butterfly Circus* is a short film that depicts an intense, emotionally evocative story of a man born without limbs who is encouraged by the showman of a renowned circus to reach his own potential. The movie is live action, color, and shot in high definition. Additional details of the experiment were previously reported (Nguyen et al., 2016b; Wang et al., 2017).

Three months after the first scan session (Session A), participants returned for the second imaging session (Session B) involving an identical protocol of resting state and movie viewing paradigms. Three participants were excluded from the reliability analysis: one was due to technical problems during

data recordings and the other two did not return for the second session. Hence, test-retest reliability analyses were performed on data from the 17 participants who finished both scan sessions.

## Functional Image Acquisition and Preprocessing

Functional and structural images were acquired from a whole-body 3-Tesla Siemens Trio MRI scanner equipped with a 12-channel head coil (Siemens Medical System, Germany). Functional images were acquired using a single-shot gradient-echo Echo Planar-Imaging (EPI) sequence with the following parameters: repetition time (TR) 2,200 ms, echo time (TE) 30 ms, flip angle (FA) 79°, Field of View (FOV) 192 × 192 mm, pixel bandwidth 2,003 Hz, a 64 × 64 acquisition matrix, 44 axial slices, and 3 × 3 × 3 mm<sup>3</sup> voxel resolution. A high-resolution T1-weighted MPRAGE structural image covering the entire brain was also collected for each participant with the following parameters: TE = 2.89 ms, TR = 4,000 ms, FA = 9°, FOV = 240 × 256 mm, and voxel size 1 × 1 × 1 mm<sup>3</sup>.

Functional images were preprocessed using Statistical Parametric Mapping toolbox (SPM12, Wellcome Department of Imaging Neuroscience, Institute of Neurology, London) and a toolbox for Data Processing & Analysis for Brain Imaging (DPABI) (Yan et al., 2016) implemented in Matlab (Mathworks, USA). The first five volumes of each EPI sequence were discarded to allow scanner equilibrium to be achieved. The remaining functional images were slice-time corrected, realigned, co-registered to the T1 structural image of each individual subject, and normalized to the Montreal Neurological Institute (MNI) space without additional smoothing. The images were further regressed out of nuisance signals, bandpass filtered (0.0083–0.15 Hz) and detrended. Nuisance signals include principle components of WM and CSF signals derived using the CompCor method (Behzadi et al., 2007) and Friston-24 motion parameters (Friston et al., 1996; Yan et al., 2013). Additional preprocessing details were previously reported (Wang et al., 2017). After preprocessing, there are total 215 and 530 volumes for resting state and natural viewing conditions, respectively.

## Heart Rate Variability

ECG signals were recorded using Brain Products system (<http://www.brainproducts.com/>). The leads were placed on the back, and the signals were recorded at the sampling rate of 5,000 Hz. Heart beats were first detected automatically using the detection algorithm implemented in QRSTool software (Allen et al., 2007). The detected heart beats were then visually checked and the misidentified ones were manually corrected. Inter-beat intervals (IBI) were then calculated as the time intervals between two successive individual beats. Using HRVAS toolbox (Ramshur, 2010), the resultant IBIs were further cleaned and processed (ectopic values removed, interpolated, and detrended). Finally, the IBIs were used to derive HRV measures: the root mean square of successive difference of IBIs (RMSSD) and Tochi cardiac vagal index (CVI). These two measures are believed to primarily reflect parasympathetic function (Neumann et al., 1941; Malik, 1996; Toichi et al., 1997).

Next we used sliding windows to derive continuous HRV (Guo et al., 2016). Sliding windows were centered in the middle of each TR, moving forward in steps of 1 TR. HRV measures were calculated using the IBIs within each window. We examined a series of window lengths: 4, 8, 12, ..., 50 s, and the proper window length was chosen based on the following criteria: (1) the time-varying HRV is highly consistent with the overall HRV, measured as the ratio of time-varying HRV averaged across all windows and subjects to the overall HRV averaged across all subjects (Thong et al., 2003); (2) the test-retest reliability of the time-varying HRV measures is good. We finally chose RMSSD and CVI with the window length of 16 s for the following analyses, because they are highly consistent with the whole scan HRV (>0.95), relatively reliable (RMSSD: scan-wise ICC: 0.8, unit-wise ICC: 0.672; CVI: scan-wise ICC: 0.693, unit-wise ICC: 0.53. Method of calculating unit- and scan-wise ICC is described below in *Test-retest reliability*), and could still provide satisfactory time resolution.

This continuous HRV was then used as an estimate of the level of sleepiness during each TR. We used a relative threshold of 50% to select the top 50 percentile sleepest (highest HRV values, sleepy-0.5) or most alert (lowest HRV values, alert-0.5) volumes from each session in the reliability analyses. To exclude any non-specific effect due to volume selection, we created a control condition by randomly selecting 50% volumes and taking the average from 5,000 randomizations (random-0.5). To ensure the robustness of our results to the selection threshold of certain state and the window lengths of time varying HRV, we performed additional reliability analyses: (1) using a serial of additional thresholds of data inclusion (0.9, 0.8, 0.7, 0.6, 0.4, 0.3) when HRV was derived using 16 s sliding window; (2) using a serial of window lengths (4, 8, 12, ..., 50 s) to derive the time-varying HRV, then performed reliability analyses for sleepy-0.5 and alert-0.5 conditions. We then used RMSSD derived from 16 s sliding window to derive continuous HRV for movie viewing data, and examined the effects of sleepiness on test-retest reliability in natural viewing conditions. To make the analyses on resting state and natural viewing conditions more comparable, we performed additional analyses on an 8-min segment of the natural viewing data, which matched the duration of the resting state sessions.

## ROI-Based Functional Connectivity Analysis

We first performed functional connectivity analysis using a previous established atlas: the 200 ROI atlas based on Craddock 2012 parcellation (Craddock et al., 2012), as it provides good cortical and subcortical coverage with fine divisions.

ROIs' time series were extracted as the mean signal across all voxels within each ROI from preprocessed fMRI data. Pearson correlation was then computed between each pair of ROIs' time series using the sleepy-0.5, alert-0.5, random-0.5 and whole data separately, resulting in four 200 × 200 connectivity matrices for each subject for each session. For each matrix, the correlation coefficients were transformed to z-scores using Fisher's transformation, averaged across all subjects for each condition, and then reverted to Pearson's r values to derive

group-level connectivity matrices (Zuo et al., 2012a; Vanderwal et al., 2015). To quantitatively evaluate the differences between connectivity matrices at different alertness levels, we performed paired *t*-test across subjects on the connectivity matrices between sleepy-0.5 and alert-0.5 conditions. The results were thresholded using FDR-corrected  $p < 0.05$ .

## Graph Theoretical Analysis

We further derived graph theoretical measures from the ROI connectivity matrices. We produced weighted adjacent matrices by thresholding the fully connected ROI matrices: suprathreshold connections (edge) retained their correlation coefficients denoting edge weights, whereas subthreshold edges were assigned values of 0. To ensure robustness of the threshold chosen, we repeated our analyses using a serial of thresholds ( $T_r = 0.1, 0.3$ , and  $0.5$ ).

We focused on two graph metrics that have been shown to be reliable: degree centrality and clustering coefficient (Braun et al., 2012; Guo et al., 2012; Andellini et al., 2015; Du et al., 2015; Wang et al., 2017). These graph metrics were derived from the weighted adjacency matrices using Brain Connectivity Toolbox (Rubinov et al., 2009). Degree centrality measures the connectedness of each node, computed as the weighted sum of all the edges connected to the node. Clustering coefficient measures the likelihood of the nodes tending to cluster together, calculated as the fraction that the number of edges actually exist to the number of all edges possibly exist. To examine the differences between graph measures with different sleepiness levels, we performed paired *t*-test across subjects on the graph measures between sleepy-0.5 and alert-0.5. The results were thresholded using an FDR-corrected  $p < 0.05$ .

## Test-Retest Reliability

In this paper, we assessed test-retest reliability using intraclass correlation coefficient (ICC) (Shrout and Fleiss, 1979; McGraw and Wong, 1996; Caceres et al., 2009). A one-way ANOVA was applied to the measures of the two scan sessions across subjects, to calculate between-subject mean square ( $MS_b$ ) and within-subject mean square ( $MS_w$ ). ICC values were then calculated as:

$$ICC = \frac{MS_b - MS_w}{MS_b + (d - 1) MS_w}$$

where  $d$  = the number of observations per subject. For every functional connectivity measure, we assessed reliability at both individual unit-wise and scan-wise levels. Unit-wise reliability is commonly reported in the literature (Shehzad et al., 2009; Wang et al., 2011; Braun et al., 2012; Guo et al., 2012; Zuo et al., 2012a; Birn et al., 2013; Liao et al., 2013). Here, one ICC value was calculated for each measurement unit, such as the HRV value of each window, the connectivity score of each ROI pair (edge), or graph metric of each ROI (node). Unit-wise ICC was then produced by averaging the ICC values for all measurement units across the windows or the network to represent reliability at individual level. Additionally, we reported scan-wise reliability, which estimates the reliability of the mean measurement derived from the entire scan session or the whole graph (Guo et al., 2012).

Here, a single ICC value was calculated for the mean HRV values, mean connectivity scores or graph metric averaged across all windows of the whole scan, or edges or nodes of the network.

The reliability results are referred as excellent ( $ICC > 0.8$ ), good ( $0.79 > ICC > 0.6$ ), moderate ( $0.59 > ICC > 0.4$ ), fair ( $0.39 > ICC > 0.2$ ), and poor ( $ICC < 0.2$ ) (Guo et al., 2012).

## Statistical Test

We tested whether ICC values of sleepy and alert conditions were significantly different from corresponding random condition, at both unit- and scan-wise levels. We performed one-tailed permutation test by comparing the true ICC value against the distribution of ICCs from the permuted random conditions (details see *Heart Rate Variability*). A 95% CI for each permutation test was calculated as the highest value (right-tailed test) or lowest (left-tailed test) with an alpha level of 0.05 (Lamotte and Volaufova, 1999; Ernst, 2004).

## Head Motion

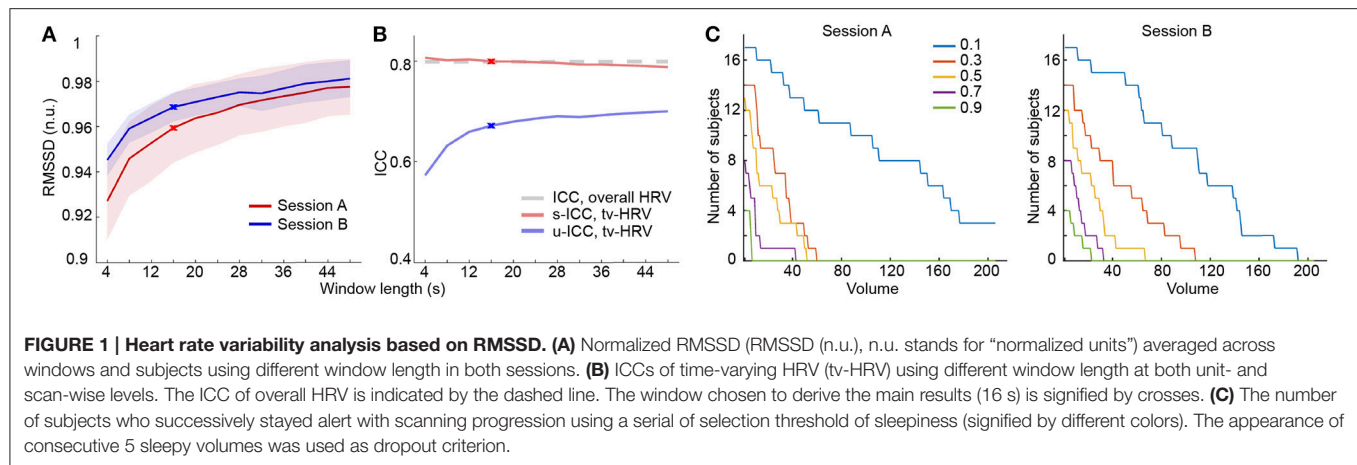
We also examined the amount of head motion during different levels of sleepiness, using framewise displacement proposed by Power et al. (2012). Framewise displacement is a scalar quantity defined as:  $FD_i = |\Delta d_{ix}| + |\Delta d_{iy}| + |\Delta d_{iz}| + |\Delta \alpha_i| + |\Delta \beta_i| + |\Delta \gamma_i|$ , where  $d_{ix}$ ,  $d_{iy}$  and  $d_{iz}$  are translational displacements along X, Y and Z axes, respectively;  $\alpha_i$ ,  $\beta_i$  and  $\gamma_i$  are rotational angles of pitch, yaw and roll, respectively;  $\Delta d_{ix} = d_{(i-1)x} + d_{ix}$ ,  $\Delta d_{iy} = d_{(i-1)y} + d_{iy}$ ,  $\Delta d_{iz} = d_{(i-1)z} + d_{iz}$ ,  $\Delta \alpha_i = \alpha_{(i-1)} + \alpha_i$ ,  $\Delta \beta_i = \beta_{(i-1)} + \beta_i$ ,  $\Delta \gamma_i = \gamma_{(i-1)} + \gamma_i$ . Rotation displacements were converted from degrees to millimeters of distance on a sphere surface (radius: 50 mm, assumed to be the radius of a head). One spike was counted when  $FD_i$  was greater than 0.3 mm (Yan et al., 2013; Vanderwal et al., 2015). We calculated the frequency of spikes as the number of spikes per volume and compared it between the different alert levels using paired *t*-test across subjects. We didn't find any significant influence of sleep on head motion.

## RESULTS

### Heart Rate Variability during Resting State fMRI

HRV is modulated by both sympathetic and parasympathetic nervous systems (Acharya et al., 2006), while the parasympathetic modulation is predominant at rest. We here used two common HRV metrics reflecting mainly parasympathetic modulation—RMSSD (Malik, 1996) and CVI (Toichi et al., 1997) to measure the overall and time-varying HRV during rs-fMRI. The overall HRV measures showed good test-retest reliability (RMSSD: 0.799; CVI: 0.681). Then we used a sliding window method to derive time-varying HRV metrics based on RMSSD and CVI (Guo et al., 2016). With proper window length, time-varying HRV measures were highly consistent ( $>0.95$ ) with overall HRV metrics [Figure 1A; SFigure 1A; results based on RMSSD are presented in main text (Figures), and those based on CVI are in Supplementary Materials (SFigure)], and showed moderate





to good test-retest reliability (RMSSD: scan-wise ICC: 0.8, unit-wise ICC: 0.672; CVI: scan-wise ICC: 0.693, unit-wise ICC: 0.53; **Figure 1B**; SFigure 1B).

It is well established that HRV increases as one gets drowsier, which has been used to detect driver alertness (Lal and Craig, 2001; Borghini et al., 2012; Abbood et al., 2014). Here, we used the time-varying HRV measures as a way to index sleepiness during resting state fMRI scans. Consistent with previous work using EEG for sleep detection (Tagliazucchi and Laufs, 2014), the number of subjects who stayed alert decreased as the scan time increased (**Figure 1C**; SFigure 1C).

## Reliability of Functional Connectivity Measures Affected by Sleep

To examine the effect of sleep on functional connectivity measures and their test-retest reliability, we performed connectivity and reliability analyses using either the 50% of data when subjects were most alert (alert-0.5) or the 50% when subjects were sleepiest (sleepy-0.5). We chose a parcellation scheme of 200 ROIs (Craddock et al., 2012), which covers the entire cortical and subcortical regions, and organized the ROIs into seven networks (Yeo et al., 2011). The seven networks are: visual, somatomotor, dorsal attention, ventral attention, limbic, frontoparietal, default, and other areas (including parts of cerebellums, thalamus, brainstems, and caudate). Overall, group averaged functional connectivity matrices derived from alert-0.5, sleepy-0.5, and whole data conditions showed similar patterns (**Figure 2A**; SFigure 2A). Direct comparison between sleepy-0.5 and alert-0.5 conditions did not detect significant differences (paired *t*-test, FDR-corrected  $p < 0.05$ ).

We then assessed whether the sleepiness affected the reliability of functional connectivity measures. Following previous studies (Guo et al., 2012), unit- and scan-wise ICC measures were used to quantify the test-retest reliability of functional connectivity measures during sleepy-0.5 and alert-0.5 conditions, respectively. Unit-wise ICC refers to that ICC was calculated for each pair of ROI connection, and scan-wise ICC derived from connectivity strengths averaged across the whole connectivity matrix. As reliability decreases with less data volumes (Birn et al., 2013),

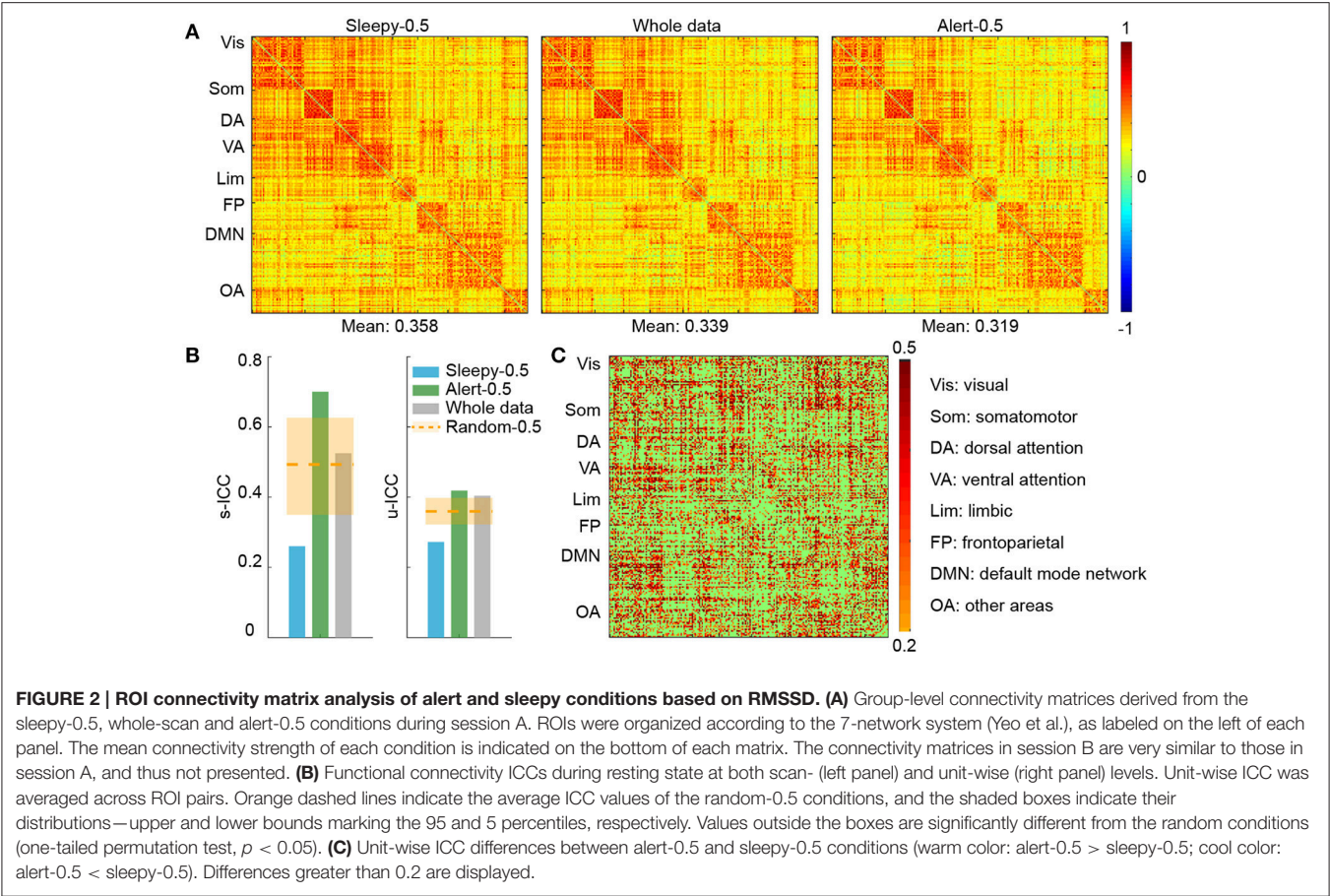
we created a control condition of 50% randomly selected volumes (random-0.5) to compare with the alert-0.5 and sleepy-0.5 conditions. Compared to the random-0.5 condition, the sleepy-0.5 condition resulted in significantly lower ICC and the alert-0.5 condition produced significantly higher ICC for both unit- and scan-wise measures (permutation test,  $p < 0.05$ ; **Figures 2B,C**; SFigure 2B; **Table 1**; **Stable 1**), suggesting that sleepiness during resting state scans reduced the reliability of functional connectivity measures. Even directly compared to the whole data condition, the alert-0.5 condition yielded higher reliability. The ICC values increased by 3.7 and 33.4% at individual unit- and scan-wise levels, respectively (**Figure 2B**; SFigure 2B), further confirming that the volumes with high sleepiness were associated with low reliability.

## Reliability of Graph Theoretical Measures Affected by Sleep

We then assessed the effect of sleep on graph theoretical measures. We focused on the graph metrics known to be reliable: clustering coefficient and degree centrality (Braun et al., 2012; Guo et al., 2012; Wang et al., 2017). To ensure the robustness of our results, graph theoretical measures were derived using a broad range of thresholds:  $T_r = 0.1, 0.3, 0.5$ . Overall, the level of sleepiness did not affect graph theoretical measures (SFigure 3). There was a slight decrease with alert-0.5 condition, but this decrease was not statistically significant (SFigure 3; paired *t*-test, FDR-corrected  $p < 0.05$ ).

We then assessed the test-retest reliability of each graph measure. Similar to functional connectivity, ICCs derived from the sleepy-0.5 condition were significantly lower than those from the random-0.5 condition, while those from the alert-0.5 condition were significantly higher, irrespective of the threshold used (permutation test,  $p < 0.05$ ; **Figure 3A**; SFigure 4A; **Table 1**; **Stable 1**). Furthermore, ICC values derived from the alert-0.5 conditions were also improved when compared to those from the whole data condition, which increased by 29.8% at unit-wise and 37.7% at scan-wise levels averaged across both graph measures and all three thresholds applied (**Figure 3A**; SFigure 4A).





**TABLE 1 | One-tailed permutation tests of the differences in resting state reliability between the sleepy-0.5 or alert-0.5 and the random-0.5 conditions, based on RMSSD.**

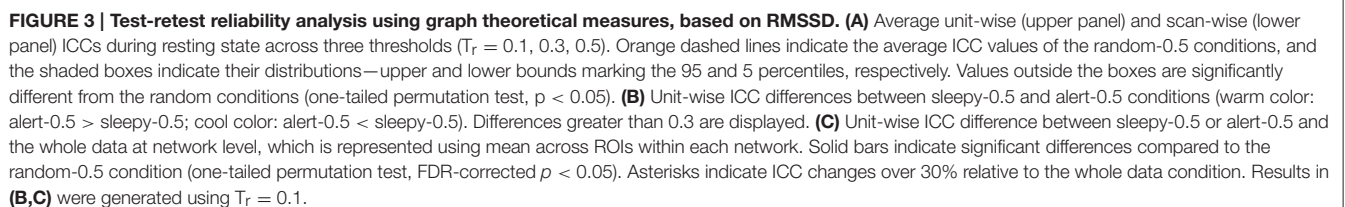
		Unit-wise			Scan-wise		
		Random	Sleepy	Alert	Random	Sleepy	Alert
Functional connectivity	ICC	[0.321, 0.397]	0.273	0.418	[0.349, 0.624]	0.26	0.7
	$p$	—	0.0004	0.0038	—	0.0048	0.003
Clustering coefficient	ICC	[0.314, 0.550]	0.216	0.636	[0.341, 0.628]	0.228	0.707
	$p$	—	0.0024	0.001	—	0.0034	0.0042
Degree centrality	ICC	[0.338, 0.519]	0.257	0.584	[0.351, 0.625]	0.252	0.699
	$p$	—	0.002	0.0014	—	0.0044	0.0032

Graph theoretical metrics were derived with  $T_r = 0.1$ . ICC and  $p$  values are listed for each condition. ICCs of random condition are indicated using upper and lower bounds marking the 95 and 5 percentiles of the random distribution, respectively.

To examine whether these changes in reliability was specific to certain brain networks, we compared unit-wise ICCs across each brain network (Figure 3B; SFigure 4B). The average reliability was calculated as the arithmetic mean across ROIs included in each network. Under alert-0.5 condition, the ICCs increased by more than 30% in most networks for clustering coefficient, and over 25% for degree centrality (Figure 3C; SFigure 4C). To ensure the robustness of the improvement, we also used the median ICCs to represent the average reliability within each network, and observed consistent results (SFigure 7).

**Test-Retest Reliability with Different Data Selection Thresholds**

So far, our results show that the test-retest reliability is improved when excluding the top 50 percentile data of high sleepiness. We then asked what percentage of volumes selection is optimal for improving test-retest reliability. We tested a range of percentiles to select volumes (Figure 4; SFigure 5). When volumes were randomly selected (random conditions), ICC value decreased with less volumes included (Birn et al., 2013). However, when specifically selecting volumes based on HRV,



ICCs increased significantly and continuously with less volumes of high sleepiness included in the calculation, till as much as 60% sleepy volumes were excluded (**Figure 4**; SFigure 5), suggesting that the detrimental effect of sleepiness on reliability outweighed the effect of reduced volumes. In practice, however, it might be desirable to remove the minimal amount of data volume and we found 20% was the least amount of sleepy volumes required to significantly improve test-retest reliability for all three measures (**Figure 4**; SFigure 5).

## Reliability of Natural Viewing Paradigm Not Affected by Sleep

As we showed previously that the reliability of connectivity measures were higher during natural viewing than resting state condition (Wang et al., 2017), we then asked whether it could be further improved by this approach.

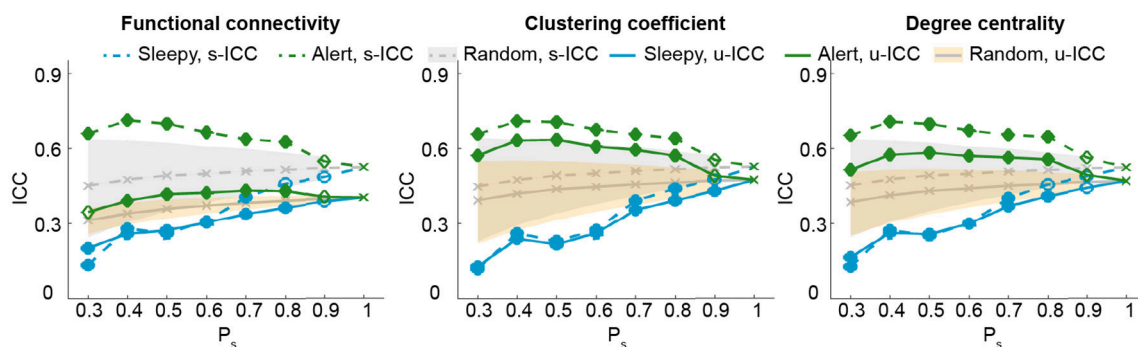
We first examined the measures of HRV during natural viewing. On average, HRV during natural viewing reduced slightly, but this reduction was not significant (paired  $t$ -test,  $p < 0.05$ ; **Figure 5A**). We further derived HRV from the most engaging movie segment based on our previous study (Wang

et al., 2017), and found that HRV during this segment was significantly lower than resting state in session A (paired  $t$ -test,  $p < 0.05$ ; **Figure 5A**). Furthermore, HRV measures were more reliable during natural viewing (0.928) than resting state (0.799), similar to our findings with functional connectivity measures (Wang et al., 2017).

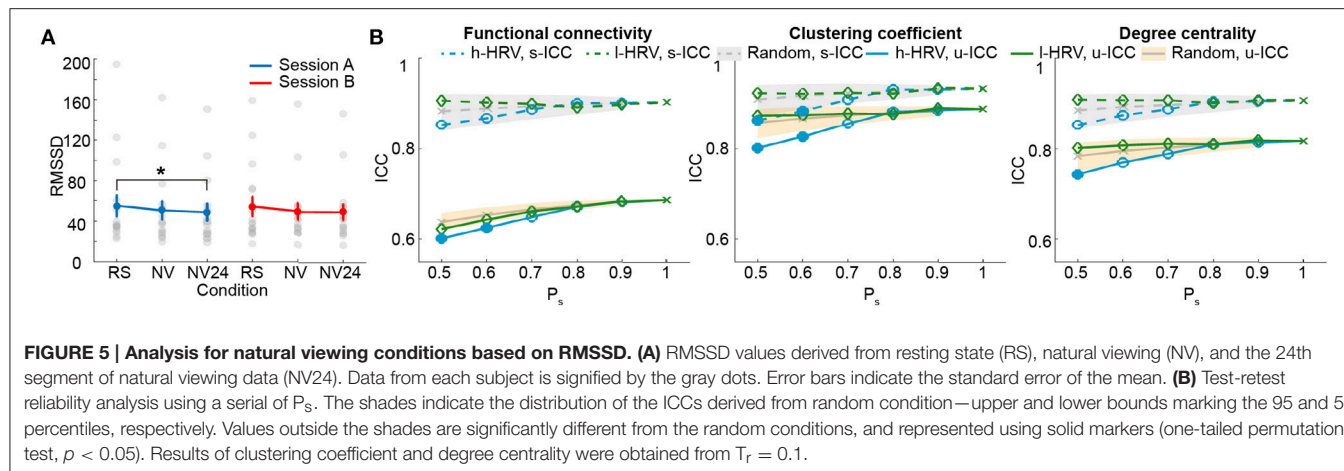
We then compared the unit- and scan-wise ICCs of functional connectivity measures. The results derived from the 8-min segment were similar to the results using the entire natural viewing data (**Figure 5B**; SFigure 8; **Table 2**). While the reliability of conditions with higher HRV level decreased, these changes were much smaller than the ones during resting state conditions. And we did not find consistent and significant increases in reliability with the low HRV conditions (**Figure 5B**; SFigure 8).

## DISCUSSION

In this study, we examined the effect of sleep on test-retest reliability of rs-fMRI connectivity measures. By excluding volumes acquired when participants were sleepy, we could improve the reliability of network connectivity measures during



**FIGURE 4 | Test-retest reliability analysis using a series of volume selection percentiles ( $P_s$ ), based on RMSSD.** The shades indicate the distribution of the ICCs derived from random condition—upper and lower bounds marking the 95 and 5 percentiles, respectively. Values outside the shades are significantly different from the random conditions, and represented using solid markers (one-tailed permutation test,  $p < 0.05$ ). Results of clustering coefficient and degree centrality were obtained from  $T_r = 0.1$ .



**FIGURE 5 | Analysis for natural viewing conditions based on RMSSD. (A)** RMSSD values derived from resting state (RS), natural viewing (NV), and the 24th segment of natural viewing data (NV24). Data from each subject is signified by the gray dots. Error bars indicate the standard error of the mean. **(B)** Test-retest reliability analysis using a series of  $P_s$ . The shades indicate the distribution of the ICCs derived from random condition—upper and lower bounds marking the 95 and 5 percentiles, respectively. Values outside the shades are significantly different from the random conditions, and represented using solid markers (one-tailed permutation test,  $p < 0.05$ ). Results of clustering coefficient and degree centrality were obtained from  $T_r = 0.1$ .

**TABLE 2 | One-tailed permutation tests of the difference in movie viewing reliability between the sleepy-0.5 or alert-0.5 and the random-0.5 conditions, based on RMSSD.**

		Unit-wise			Scan-wise		
		Random	Sleepy	Alert	Random	Sleepy	Alert
Functional connectivity	ICC	[0.617, 0.658]	0.602	<i>0.622</i>	[0.841, 0.922]	<i>0.854</i>	<i>0.907</i>
	<i>p</i>	–	0.0024	<i>0.1038</i>	–	<i>0.1178</i>	<i>0.1774</i>
Clustering coefficient	ICC	[0.824, 0.889]	0.801	<i>0.874</i>	[0.873, 0.941]	0.863	<i>0.924</i>
	<i>p</i>	–	0.006	<i>0.2194</i>	–	0.022	<i>0.2575</i>
Degree centrality	ICC	[0.750, 0.814]	0.743	<i>0.802</i>	[0.846, 0.922]	<i>0.853</i>	<i>0.909</i>
	<i>p</i>	–	0.0264	<i>0.176</i>	–	<i>0.0894</i>	<i>0.1572</i>

Graph theoretical metrics were derived with  $T_r = 0.1$ . ICC and *p* values are listed for each condition. ICCs of random condition are indicated using upper and lower bounds marking the 95 and 5 percentiles of the random distribution, respectively. Non-significant results are in *italic*.

rs-fMRI paradigm. The improvement of test-retest reliability is robust with removal of as little as 20% of volumes. Noticeably, this improvement on ICC outweighs the opposing effect from reduced volume (Birn et al., 2013). Overall, our results provide a novel and practical way to improve test-retest reliability of rs-fMRI paradigm.

The test-retest reliability of rs-fMRI measures ranges between poor to moderate (Telesford et al., 2010; Wang et al., 2011; Braun et al., 2012; Guo et al., 2012; Li et al., 2012; Patriat et al., 2013; Cao et al., 2014). Many factors contribute to the moderate reliability, including poor signal-to-noise ratio of the blood oxygenation level-dependent (BOLD) signal, excessive head motion, physiological noise, and so on. Previous work has found that test-retest reliability can be improved by removing volumes or subjects with excessive motion, and regressing out motion related artifacts (Schwarz and McGonigle, 2011; Guo et al., 2012; Zuo et al., 2012b; Gorgolewski et al., 2013; Yan et al., 2013; Du et al., 2015). Now we showed that the presence of drowsiness and sleep during scanning is another factor affecting rs-fMRI measures and their reliability. Due to acoustic noise, fatigue, and the lack of stimulation, it is common that subjects fall asleep during rs-fMRI scans (Tagliazucchi and Laufs, 2014). Sleep was found to be associated with changes in brain network, urging caution when interpreting functional connectivity measures during resting state (Massimini et al., 2005; Larson-Prior et al., 2009; Spoormaker et al., 2010; Koike et al., 2011; Picchioni et al., 2014; Tagliazucchi and Laufs, 2014; Hale et al., 2016). Some methods were proved to be effective to prevent subjects from falling asleep, such as requiring subjects to keep eyes open or fixed on a cross (Patriat et al., 2013; Zou et al., 2015), and their test-retest reliability are higher than resting state with eyes closed. However, this impact on connectivity measures and their test-retest reliability appears to differ across brain networks (Patriat et al., 2013; Zou et al., 2015).

Previous studies report decreases in heart rate and increases in HRV at the transition from wake to non-REM sleep. These changes have thus been widely used to detect sleepiness in real life situations (Lal and Craig, 2001; Borghini et al., 2012; Abbood et al., 2014). While previous studies used long ECG data to derive HRV (5 min to 24 h), recent studies have used shorter duration (10–250 s) to improve the temporal resolution (Thong

et al., 2003; Salahuddin et al., 2007; Udi et al., 2011; Chang et al., 2013; Valenza et al., 2014; Guo et al., 2016; Massaro and Pecchia, 2016; Nguyen et al., 2016a). In this study, we examined the robustness and reliability of HRV metrics derived using different window lengths. For both RMSSD and CVI measures, the metrics derived using short data durations are highly consistent ( $>0.95$ ) with the ones derived using the whole 8-min data, and RMSSD achieves good test-retest reliability with the window length of as short as 6 s. These analyses support the use of short-term HRV as time-varying measures. It is increasingly recognized that physiological fluctuations could introduce noise in fMRI signals. It is possible that higher HRV might contribute to greater fMRI noise. Removal of noisy volumes could thus lead to an improvement in reliability. In our current experimental design, it is not possible to discern between the contributions of physiological noise and sleepiness. Irrespective of the source, excluding volumes of high HRV could still provide a valid strategy to improve test-retest reliability of rs-fMRI connectivity.

We excluded volumes of high or low HRV for connectivity and test-retest reliability analyses. This approach is similar to the motion scrubbing method proposed to reduce the impact of motion artifacts (Power et al., 2012). In some study, an average of 58% data were scrubbed for a cohort of children where motion is problematic. In our dataset, after excluding 50% sleepest volumes, we found ICC values increased by 24.9% (0.108) at the unit-wise level and 36.4% (0.187) at the scan-wise level averaged across the three measures we examined (functional connectivity, clustering coefficient and degree centrality), and across all three thresholds for graph measures. The test-retest reliability also improved in higher order brain networks, such as dorsolateral prefrontal cortex, angular gyrus, and cingulate cortex (Figure 3; SFigure 4), reflecting the impact of sleep on these brain regions. In our main analyses, the volume-wise sleepiness level was identified using the time-varying HRV derived from 16 s sliding window, which was chosen based on a tradeoff between time resolution and the robustness of HRV measure itself. We additionally tested the effects of the HRV window length on the reliability of functional connectivity and graph measures (SFigure 6). Our major conclusion, that reliability improved when sleepy volumes were excluded, was consistent across different window lengths. This improvement diminishes,



however, if using a too short or too long window length. ICC of sleepy-0.5 condition decreased in general with longer window length, possibly due to the reduced volume number (Birn et al., 2013). Overall, the method proposed in this work is effective and efficient at improving test-retest reliability of rs-fMRI paradigm.

We additionally examined the effect of sleep on test-retest reliability during natural viewing paradigm. Unlike the effect on resting state measures, excluding volumes with higher HRV had very limited effect on the reliability of natural viewing data with as much as 50% volumes excluded regardless of the data length used (Figure 5B; SFigure 8). During movie viewing, cardiac autonomic activities are likely to be influenced by sustained attention and emotional saliency (Thayer and Lane, 2001) where high HRV does not necessarily reflect sleepiness. The ability of RMSSD to detect sleep is thus diminished. The test-retest reliability of connectivity measures is higher for natural viewing than resting state paradigm (Wang et al., 2017), which might be partially contributed by the high alertness during natural viewing.

There are several limitations to our study. Sleep is a complex physiological condition, and the use of single HRV measure for sleep detection might be oversimplified. In particular, HRV during movie viewing conditions is likely to be influenced by emotional responses rather than sleepiness. Therefore, the method we proposed here is a simple scheme to assess sleep and improve test-retest reliability for rs-fMRI paradigm, and our results on natural viewing should be considered with caution. Various methods have been previously used for sleep detection, such as subjective questionnaires, other physiological signals including EEG (Rechtschaffen and Kales, 1968; Iber et al., 2007), electrooculogram, electromyogram (Abbood et al., 2014), fMRI (Tagliazucchi et al., 2012; Tagliazucchi and Laufs, 2014), and more HRV measures (Sahayadhas et al., 2013). Advanced methods like machine learning have also been applied (Sahayadhas et al., 2012). With more advanced algorithm and/or

additional physiological signals combined, it might be possible to further improve the accuracy of sleep detection, or expand such analysis to more complex conditions. In addition, it would be useful to examine whether HRV derived from pulse oximetry recording could provide similar results as ECG recording. Pulse oximetry is easier to implement and less affected by MR gradient artifact than ECG. While the smooth pulse waveform might offer less precision for peak detection, it has been shown to produce comparable HRV values as ECG (Iyriboz et al., 1991) and used to derive HRV values during rs-fMRI (Lv et al., 2015; Guo et al., 2016). Therefore, pulse oximetry might be used for sleep detection instead of ECG, which could be formally tested in the future studies.

## AUTHOR CONTRIBUTIONS

JW, conducted data analysis, wrote the manuscript; VN, collected the data, provided practical advice for data analysis; CG, initiated the study, designed the experiment, interpreted the results, edited the manuscript; LG and JH, provided administrative and material support.

## FUNDING

This work is supported by QIMR international fellowship, and NHMRC Project (Grant #1098407) and an NHMRC Career Development Fellowship (#1123674) to C.C.G., and the National Science Foundation of China (Grant #61522207) to JH.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2017.00249/full#supplementary-material>

## REFERENCES

- Abbood, H., Al-Nuaimy, W., Al-Ataby, A., Salem, S. A., and Alzubi, H. S. (2014). "Prediction of driver fatigue: approaches and open challenges," in *Computational Intelligence (UKCI) (Bradford: 14th UK Workshop on: IEEE)*, 1–6.
- Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., and Suri, J. S. (2006). Heart rate variability: a review. *Med. Biol. Eng. Comput.* 44, 1031–1051. doi: 10.1007/s11517-006-0119-0
- Allen, J. J., Chambers, A. S., and Towers, D. N. (2007). The many metrics of cardiac chronotropy: a pragmatic primer and a brief comparison of metrics. *Biol. Psychol.* 74, 243–262. doi: 10.1016/j.biopsycho.2006.08.005
- Andellini, M., Cannatà, V., Gazzellini, S., Bernardi, B., and Napolitano, A. (2015). Test-retest reliability of graph metrics of resting state MRI functional brain networks: a review. *J. Neurosci. Methods* 253, 183–192. doi: 10.1016/j.jneumeth.2015.05.020
- Behzadi, Y., Restom, K., Liao, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37, 90–101. doi: 10.1016/j.neuroimage.2007.04.042
- Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., et al. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83, 550–558. doi: 10.1016/j.neuroimage.2013.05.099
- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2012). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Boudreau, P., Yeh, W. H., Dumont, G. A., and Boivin, D. B. (2013). Circadian variation of heart rate variability across sleep stages. *Sleep* 36, 1919–1928. doi: 10.5665/sleep.3230
- Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., et al. (2012). Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 59, 1404–1412. doi: 10.1016/j.neuroimage.2011.08.044
- Buckner, R. L., Krienen, F. M., and Yeo, B. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nat. Neurosci.* 16, 832–837. doi: 10.1038/nn.3423
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575
- Bullmore, E. T., and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annu. Rev. Clin. Psychol.* 7, 113–140. doi: 10.1146/annurev-clinpsy-040510-143934

- Burgess, H. J., Trinder, J., Kim, Y., and Luke, D. (1997). Sleep and circadian influences on cardiac autonomic nervous system activity. *Am. J. Physiol.* 273, 1761–1768.
- Busek, P., Vanková, J., Opavský, J., Salinger, J., and Nevsímalová, S. (2005). Spectral analysis of the heart rate variability in sleep. *Physiol. Res.* 54, 369–376.
- Cabiddu, R., Cerutti, S., Viardot, G., Werner, S., and Bianchi, A. M. (2012). Modulation of the sympatho-vagal balance during sleep: frequency domain study of heart rate variability and respiration. *Front. Physiol.* 3:45. doi: 10.3389/fphys.2012.00045
- Caceres, A., Hall, D. L., Zelaya, F. O., Williams, S. C., and Mehta, M. A. (2009). Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage* 45, 758–768. doi: 10.1016/j.neuroimage.2008.12.035
- Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., et al. (2014). Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage* 84, 888–900. doi: 10.1016/j.neuroimage.2013.09.013
- Carrington, M. J., Barbieri, R., Colrain, I. M., Crowley, K. E., Kim, Y., and Trinder, J. (2005). Changes in cardiovascular function during the sleep onset period in young adults. *J. Appl. Physiol.* 98, 468–476. doi: 10.1152/japplphysiol.00702.2004
- Cellini, N., Whitehurst, L. N., McDevitt, E. A., and Mednick, S. C. (2016). Heart rate variability during daytime naps in healthy adults: autonomic profile and short-term reliability. *Psychophysiology* 53, 473–481. doi: 10.1111/psyp.12595
- Chang, C., Metzger, C. D., Glover, G. H., Duyn, J. H., Heinze, H. J., and Walter, M. (2013). Association between heart rate variability and fluctuations in resting-state functional connectivity. *Neuroimage* 68, 93–104. doi: 10.1016/j.neuroimage.2012.11.038
- Chouchou, F., and Deseilles, M. (2014). Heart rate variability: a tool to explore the sleeping brain? *Front. Neurosci.* 8:402. doi: 10.3389/fnins.2014.00402
- Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 33, 1914–1928. doi: 10.1002/hbm.21333
- de Zambotti, M., Cellini, N., Baker, F. C., Colrain, I. M., Sarlo, M., and Stegagno, L. (2014). Nocturnal cardiac autonomic profile in young primary insomniacs and good sleepers. *Int. J. Psychophysiol.* 93, 332–339. doi: 10.1016/j.ijpsycho.2014.06.014
- de Zambotti, M., Covassin, N., Tona, G. D. M., Sarlo, M., and Stegagno, L. (2011). Sleep onset and cardiovascular activity in primary insomnia. *J. Sleep Res.* 20, 318–325. doi: 10.1111/j.1365-2869.2010.00871.x
- Du, H. X., Liao, X. H., Lin, Q. X., Li, G. S., Chi, Y. Z., Liu, X., et al. (2015). Test–retest reliability of graph metrics in high-resolution functional connectomics: a resting-state functional MRI study. *CNS Neurosci. Ther.* 21, 802–816. doi: 10.1111/cns.12431
- Elsenbruch, S., Harnish, M. J., and Orr, W. C. (1999). Heart rate variability during waking and sleep in healthy males and females. *Sleep* 22, 1067–1071.
- Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Stat. Sci.* 19, 676–685. doi: 10.1214/088342304000000396
- Fox, M. D., and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* 4:19. doi: 10.3389/fnsys.2010.00019
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.* 35, 346–355. doi: 10.1002/mrm.1910350312
- Gorgolewski, K. J., Storkey, A. J., Bastin, M. E., Whittle, I., and Pernet, C. (2013). Single subject fMRI test–retest reliability metrics and confounding factors. *Neuroimage* 69, 231–243. doi: 10.1016/j.neuroimage.2012.10.085
- Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.* 21, 424–430. doi: 10.1097/WCO.0b013e328306f2c5
- Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100
- Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., et al. (2012). One-year test–retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. doi: 10.1016/j.neuroimage.2012.03.027
- Guo, C. C., Sturm, V. E., Zhou, J., Gennatas, E. D., Trujillo, A. J., Hua, A. Y., et al. (2016). Dominant hemisphere lateralization of cortical parasympathetic control as revealed by frontotemporal dementia. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2430–E2439. doi: 10.1073/pnas.1509184113
- Guye, M., Bettus, G., Bartolomei, F., and Cozzone, P. J. (2010). Graph theoretical analysis of structural and functional connectivity MRI in normal and pathological brain networks. *Magn. Reson. Mater. Phys. Biol. Med.* 23, 409–421. doi: 10.1007/s10334-010-0205-z
- Hale, J. R., White, T. P., Mayhew, S. D., Wilson, R. S., Rollings, D. T., Khalsa, S., et al. (2016). Altered thalamocortical and intra-thalamic functional connectivity during light sleep compared with wake. *Neuroimage* 125, 657–667. doi: 10.1016/j.neuroimage.2015.10.041
- Hayasaka, S., and Laurienti, P. J. (2010). Comparison of characteristics between region- and voxel-based network analyses in resting-state fMRI data. *Neuroimage* 50, 499–508. doi: 10.1016/j.neuroimage.2009.12.051
- He, Y., and Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Curr. Opin. Neurol.* 23, 341–350. doi: 10.1097/wco.0b013e32833aa567
- Horovitz, S. G., Braun, A. R., Carr, W. S., Picchioni, D., Balkin, T. J., Fukunaga, M., et al. (2009). Decoupling of the brain's default mode network during deep sleep. *Proc. Natl. Acad. Sci. U.S.A.* 106:11376. doi: 10.1073/pnas.0901435106
- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine Westchester, IL.
- Iyriboz, Y., Powers, S., Morrow, J., Ayers, D., and Landry, G. (1991). Accuracy of pulse oximeters in estimating heart rate at rest and during exercise. *Br. J. Sports Med.* 25, 162–164. doi: 10.1136/bjsm.25.3.162
- Jafri, M. J., Pearlson, G. D., Stevens, M., and Calhoun, V. D. (2008). A method for functional network connectivity among spatially independent resting-state components in schizophrenia. *Neuroimage* 39, 1666–1681. doi: 10.1016/j.neuroimage.2007.11.001
- Koike, T., Kan, S., Misaki, M., and Miyauchi, S. (2011). Connectivity pattern changes in default-mode network with deep non-REM and REM sleep. *Neurosci. Res.* 69, 322–330. doi: 10.1016/j.neures.2010.12.018
- Lal, S. K., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194. doi: 10.1016/S0301-0511(00)00085-5
- Lamotte, L. R., and Volafova, J. (1999). Prediction intervals via consonance intervals. *J. R. Stat. Soc.* 48, 419–424. doi: 10.1111/1467-9884.00200
- Larson-Prior, L. J., Zempel, J. M., Nolan, T. S., Prior, F. W., Snyder, A. Z., and Raichle, M. E. (2009). Cortical network functional connectivity in the descent to sleep. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4489. doi: 10.1073/pnas.0900924106
- Li, Z., Kadivar, A., Pluta, J., Dunlop, J., and Wang, Z. (2012). Test–retest stability analysis of resting brain activity revealed by blood oxygen level-dependent functional MRI. *J. Magn. Reson. Imaging* 36, 344–354. doi: 10.1002/jmri.23670
- Liao, X.-H., Xia, M.-R., Xu, T., Dai, Z.-J., Cao, X.-Y., Niu, H.-J., et al. (2013). Functional brain hubs and their test–retest reliability: a multiband resting-state functional MRI study. *Neuroimage* 83, 969–982. doi: 10.1016/j.neuroimage.2013.07.058
- Lv, J., Liu, D., Ma, J., Wang, X., and Zhang, J. (2015). Graph theoretical analysis of BOLD functional connectivity during human sleep without EEG monitoring. *PLoS ONE* 10:e0137297. doi: 10.1371/journal.pone.0137297
- Malik, M. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use: task force of the European society of cardiology and the North American society for pacing and electrophysiology. *Genes Chromosomes Cancer* 17:354. doi: 10.1111/j.1542-474x.1996.tb00275.x
- Massaro, S., and Pecchia, L. (2016). Heart rate variability (HRV) analysis: a methodology for organizational neuroscience. *Organ. Res. Methods.* doi: 10.1177/1094428116681072
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- McCarley, R. W. (2007). Neurobiology of REM and NREM sleep. *Sleep Med.* 8:302. doi: 10.1016/j.sleep.2007.03.005
- McGraw, K. O., and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1:30. doi: 10.1037/1082-989x.1.1.30

- Neumann, J. V., Kent, R. H., Bellinson, H. R., and Hart, B. I. (1941). The mean square successive difference. *Ann. Math. Stat.* 12, 153–162. doi: 10.1214/aoms/1177731746
- Nguyen, V. T., Breakspear, M., Hu, X., and Guo, C. C. (2016a). The integration of the internal and external milieu in the insula during dynamic emotional experiences. *Neuroimage* 124, 455–463. doi: 10.1016/j.neuroimage.2015.08.078
- Nguyen, V. T., Sonkusare, S., Jane, S., Hu, X., Breakspear, M., and Guo, C. C. (2016b). Distinct cerebellar contributions to cognitive-perceptual dynamics during natural viewing. *Cereb. Cortex* 1–11. doi: 10.1093/cercor/bhw334
- Patriat, R., Molloy, E. K., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., et al. (2013). The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. *Neuroimage* 78, 463–473. doi: 10.1016/j.neuroimage.2013.04.013
- Picchioni, D., Pixa, M. L., Fukunaga, M., Carr, W. S., Horovitz, S. G., Braun, A. R., et al. (2014). Decreased connectivity between the thalamus and the neocortex during human nonrapid eye movement sleep. *Sleep* 37:387. doi: 10.5665/sleep.3422
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018
- Ramshur, J. T. (2010). *HRVAS: Heart Rate Variability Analysis Software*. University of Memphis; Department of Biomedical Engineering.
- Rechtschaffen, A., and Kales, A. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Clin. Neurophysiol.* 26:644. doi: 10.1016/0013-4694(69)90021-2
- Rubinov, M., Kötter, R., Hagmann, P., and Sporns, O. (2009). Brain connectivity toolbox: a collection of complex network measurements and brain connectivity datasets. *Neuroimage* 47:S169. doi: 10.1016/S1053-8119(09)71822-1
- Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors* 12:16937. doi: 10.3390/s121216937
- Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2013). Drowsiness detection during different times of day using multiple features. *Australas. Phys. Eng. Sci. Med.* 36, 243–250. doi: 10.1007/s13246-013-0200-6
- Salahuddin, L., Cho, J., Jeong, M. G., and Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 4656–4659. doi: 10.1109/IEMBS.2007.4353378
- Schwarz, A. J., and McGonigle, J. (2011). Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *Neuroimage* 55, 1132–1146. doi: 10.1016/j.neuroimage.2010.12.047
- Shehzad, Z., Kelly, A. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., et al. (2009). The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229. doi: 10.1093/cercor/bhn256
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86:420. doi: 10.1037/0033-2909.86.2.420
- Spoormaker, V. I., Schröter, M. S., Gleiser, P. M., Andrade, K. C., Dresler, M., Wehrle, R., et al. (2010). Development of a large-scale functional brain network during human non-rapid eye movement sleep. *J. Neurosci.* 30, 11379–11387. doi: 10.1523/JNEUROSCI.2015-10.2010
- Tagliazucchi, E., and Laufs, H. (2014). Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708. doi: 10.1016/j.neuron.2014.03.020
- Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Borisov, S., Jahnke, K., and Laufs, H. (2012). Automatic sleep staging using fMRI functional connectivity data. *Neuroimage* 63, 63–72. doi: 10.1016/j.neuroimage.2012.06.036
- Tailby, C., Masterton, R. A., Huang, J. Y., Jackson, G. D., and Abbott, D. F. (2015). Resting state functional connectivity changes induced by prior brain state are not network specific. *Neuroimage* 106, 428–440. doi: 10.1016/j.neuroimage.2014.11.037
- Telesford, Q. K., Morgan, A. R., Hayasaka, S., Simpson, S. L., Barret, W., Kraft, R. A., et al. (2010). Reproducibility of graph metrics in fMRI networks. *Front. Neuroinform.* 4:117. doi: 10.3389/fninf.2010.00117
- Thayer, J. F., and Lane, R. D. (2001). A model of neurovisceral integration in emotion regulation and dysregulation. *J. Affect. Disord.* 61, 201–216. doi: 10.1016/S0165-0327(00)00338-4
- Thong, T., Li, K., McNames, J., Aboy, M., and Goldstein, B. (2003). “Accuracy of ultra-short heart rate variability measures,” in *Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE: IEEE (Cancun)*, 2424–2427.
- Tobaldini, E., Nobili, L., Strada, S., Casali, K. R., Braghiroli, A., and Montano, N. (2013). Heart rate variability in normal and pathological sleep. *Front. Physiol.* 4:294. doi: 10.3389/fphys.2013.00294
- Toichi, M., Sugiura, T., Murai, T., and Sengoku, A. (1997). A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R-R interval. *J. Auton. Nerv. Syst.* 62, 79–84. doi: 10.1016/S0165-1838(96)00112-9
- Toscani, L., Gangemi, P. F., Parigi, A., Silipo, R., Ragghianti, P., Sirabella, E., et al. (1996). Human heart rate variability and sleep stages. *Neurol. Sci.* 17, 437–439. doi: 10.1007/BF01997720
- Trinder, J., Kleiman, J., Carrington, M., Smith, S. S., Breen, S., Tan, N., et al. (2001). Autonomic activity during human sleep as a function of time and sleep stage. *J. Sleep Res.* 10, 253–264. doi: 10.1046/j.1365-2869.2001.00263.x
- Trinder, J., Waloszek, J., Woods, M. J., and Jordan, A. S. (2012). Sleep and cardiovascular regulation. *Pflügers Arch.* 463, 161–168. doi: 10.1007/s00424-011-1041-3
- Udi, N., Keren, P. E., Keren, K. M. D., Moshe, N. M. D., Shlomo, S. M. D., Volovitz, B., et al. (2011). Reliability of Ultra-Short ECG Indices for heart rate variability. *Ann. Noninvasive Electrocardiol.* 16, 117–122. doi: 10.1111/j.1542-474X.2011.00417.x
- Valenza, G., Citi, L., Lanatà, A., Scilingo, E. P., and Barbieri, R. (2014). Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. *Sci. Rep.* 4:4998. doi: 10.1038/srep04998
- van den Heuvel, M. P., and Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534. doi: 10.1016/j.euroneuro.2010.03.008
- van den Heuvel, M. P., Stam, C. J., Boersma, M., and Pol, H. H. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43, 528–539. doi: 10.1016/j.neuroimage.2008.08.010
- Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., and Castellanos, F. X. (2015). Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage* 122, 222–232. doi: 10.1016/j.neuroimage.2015.07.069
- Van Dijk, K. R., Sabuncu, M. R., and Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59, 431–438. doi: 10.1016/j.neuroimage.2011.07.044
- Wang, J., Ren, Y., Hu, X., Nguyen, V. T., Guo, L., Han, J., et al. (2017). Test-retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Hum. Brain Mapp.* 38, 2226–2241. doi: 10.1002/hbm.23517
- Wang, J., Zuo, X.-N., Gohel, S., Milham, M. P., Biswal, B. B., and He, Y. (2011). Graph theoretical analysis of functional brain networks: test-retest evaluation on short- and long-term resting-state functional MRI data. *PLoS ONE* 6:e21976. doi: 10.1371/journal.pone.0021976
- Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., et al. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage* 76, 183–201. doi: 10.1016/j.neuroimage.2013.03.004
- Yan, C. G., Wang, X. D., Zuo, X. N., and Zang, Y. F. (2016). DPABI: data processing & analysis for (Resting-State) brain imaging. *Neuroinformatics* 14, 339–351. doi: 10.1007/s12021-016-9299-4
- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011
- Zou, Q., Miao, X., Liu, D., Wang, D. J., Zhuo, Y., and Gao, J. H. (2015). Reliability comparison of spontaneous brain activities between BOLD and CBF

- contrasts in eyes-open and eyes-closed resting states. *Neuroimage* 121, 91–105. doi: 10.1016/j.neuroimage.2015.07.044
- Zuo, X. N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., et al. (2012a). Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875. doi: 10.1093/cercor/bhr269
- Zuo, X. N., Xu, T., Jiang, L., Yang, Z., Cao, X. Y., He, Y., et al. (2012b). Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage* 65, 374–386. doi: 10.1016/j.neuroimage.2012.10.017

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Wang, Han, Nguyen, Guo and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Intra- and Inter-scanner Reliability of Scaled Subprofile Model of Principal Component Analysis on ALFF in Resting-State fMRI Under Eyes Open and Closed Conditions

Li-Xia Yuan<sup>1†</sup>, Jian-Bao Wang<sup>2,3,4†</sup>, Na Zhao<sup>2,3,4</sup>, Yuan-Yuan Li<sup>2,3,4</sup>, Yilong Ma<sup>5\*</sup>, Dong-Qiang Liu<sup>6</sup>, Hong-Jian He<sup>1\*</sup>, Jian-Hui Zhong<sup>1</sup> and Yu-Feng Zang<sup>2,3,4\*</sup>

## OPEN ACCESS

### Edited by:

Bharat B. Biswal,  
University of Medicine and Dentistry of  
New Jersey, United States

### Reviewed by:

Pierre Bellec,  
Université de Montréal, Canada  
Tess Li,  
University of Electronic Science and  
Technology of China, China

### \*Correspondence:

Yilong Ma  
yma@northwell.edu  
Hong-Jian He  
hhezju@zju.edu.cn  
Yu-Feng Zang  
zangyf@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 01 September 2017

**Accepted:** 23 April 2018

**Published:** 25 May 2018

### Citation:

Yuan L-X, Wang J-B, Zhao N, Li Y-Y,  
Ma Y, Liu D-Q, He H-J, Zhong J-H  
and Zang Y-F (2018) Intra- and  
Inter-scanner Reliability of Scaled  
Subprofile Model of Principal  
Component Analysis on ALFF in  
Resting-State fMRI Under Eyes Open  
and Closed Conditions.  
Front. Neurosci. 12:311.  
doi: 10.3389/fnins.2018.00311

<sup>1</sup> Key Laboratory for Biomedical Engineering of Ministry of Education, Center for Brain Imaging Science and Technology, College of Biomedical Engineering and Instrumental Science, Zhejiang University, Hangzhou, China, <sup>2</sup> Center for Cognition and Brain Disorders and the Affiliated Hospital, Hangzhou Normal University, Hangzhou, China, <sup>3</sup> Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China, <sup>4</sup> Institutes of Psychological Sciences, College of Education, Hangzhou Normal University, Hangzhou, China, <sup>5</sup> Center for Neurosciences, The Feinstein Institute for Medical Research, Manhasset, NY, United States, <sup>6</sup> Research Center of Brain and Cognitive Neuroscience, Liaoning Normal University, Dalian, China

Scaled Subprofile Model of Principal Component Analysis (SSM-PCA) is a multivariate statistical method and has been widely used in Positron Emission Tomography (PET). Recently, SSM-PCA has been applied to discriminate patients with Parkinson's disease and healthy controls with Amplitude of Low Frequency Fluctuation (ALFF) from Resting-State Functional Magnetic Resonance Imaging (RS-fMRI). As RS-fMRI scans are more readily available than PET scans, it is important to investigate the intra- and inter-scanner reliability of SSM-PCA in RS-fMRI. A RS-fMRI dataset with Eyes Open (EO) and Eyes Closed (EC) conditions was obtained in 21 healthy subjects ( $21.8 \pm 1.8$  years old, 11 females) on 3 visits (V1, V2, and V3), with V1 and V2 (mean interval of 14 days apart) on one scanner and V3 (about 8 months from V2) on a different scanner. To simulate between-group analysis in conventional SSM-PCA studies, 21 subjects were randomly divided into two groups, i.e., EC-EO group (EC ALFF map minus EO ALFF map,  $n = 11$ ) and EO-EC group ( $n = 10$ ). A series of covariance patterns and their expressions were derived for each visit. Only the expression of the first pattern showed significant differences between the two groups for all the visits ( $p = 0.012$ ,  $0.0044$ , and  $0.00062$  for V1, V2, and V3, respectively). This pattern, referred to as EOEC-pattern, mainly involved the sensorimotor cortex, superior temporal gyrus, frontal pole, and visual cortex. EOEC-pattern's expression showed fair intra-scanner reliability ( $ICC = 0.49$ ) and good inter-scanner reliability ( $ICC = 0.65$  for V1 vs. V2 and  $ICC = 0.66$  for V2 vs. V3). While the EOEC-pattern was similar with the pattern of conventional unpaired  $T$ -test map, the two patterns also showed method-specific regions, indicating that SSM-PCA and conventional  $T$ -test are complementary for neuroimaging studies.

**Keywords:** principal component analysis, scaled subprofile model, intra-scanner reliability, inter-scanner reliability, resting-state fMRI

## INTRODUCTION

Identification of reproducible and region-specific effects that characterize normal or diseased brain state is one of the most important goals of brain functional imaging studies. The Scaled Subprofile Model of Principal Component Analysis (SSM-PCA) is one of the earliest multivariate data analytic techniques that are available to recognize significant group-dependent and region-specific effects (Moeller et al., 1987; Moeller and Strother, 1991; Alexander and Moeller, 1994; Eidelberg, 2009). The SSM-PCA is one form of regional covariance analysis, identifying functional interaction patterns among brain regions that are spatially distributed throughout the brain (Moeller et al., 1987). Commonly, the SSM-PCA has been applied to differentiate two groups of subjects (e.g., patients vs. healthy controls) (Alexander and Moeller, 1994; Spetsieris and Eidelberg, 2011; Wu et al., 2013; Tomše et al., 2017). The brain images of the two groups are decomposed to be a linear combination of a series of spatial patterns (i.e., images) by SSM-PCA. Each pattern is expressed in each subject with a Subject Scaling Factor (SSF), which can be prospectively assessed and compared between groups and validated with disease severity and neuropsychological test scores (Alexander and Moeller, 1994; Eidelberg, 2009).

SSM-PCA was first proposed to analyze data from Positron Emission Tomography (PET) (Moeller et al., 1987), and had been widely applied to investigate the effects of neurological and psychiatric illness on brain function, such as Alzheimer's disease (Alexander and Moeller, 1994), Parkinson's disease (Eidelberg et al., 1995), major depressive disorder (Sackeim et al., 1993), acquired immune deficiency syndrome dementia complex (Rottenberg et al., 1987), neoplastic disease (Anderson et al., 1988), and normal aging (Pagani et al., 2016). SSM-PCA was then utilized to deal with structural, perfusion, and diffusion Magnetic Resonance Imaging (MRI) metrics, including white and gray matter density (Brickman et al., 2007, 2008; Bergfield et al., 2010), gray matter volume (Guo et al., 2014; Steffener et al., 2016), cerebral blood flow (CBF) (Asllani et al., 2008; Teune et al., 2014), and fractional anisotropy (Gazes et al., 2016). More recently, SSM-PCA was applied to investigate Parkinson's disease-related covariance brain pattern with a Resting-State Functional MRI (RS-fMRI) metric (Wu et al., 2015), namely Amplitude of Low Frequency Fluctuation (ALFF) (Zang et al., 2007), revealing that the subject's expression of this pattern is capable of discriminating patients from healthy volunteers. RS-fMRI has many metrics and thousands of papers have been published on various brain disorders, however to our best knowledge, only one has utilized SSM-PCA (Wu et al., 2015).

Reliability is the cornerstone of any scientific measurement (Bennett and Miller, 2010). The intra- and inter-scanner reliability is an important metric for quantification of fMRI measurement reliability, given increasing research interest relying on the ability to combine the data from multiple scanners into larger, integrative data sets. For SSM-PCA, we found that only one study measured the test-retest (i.e., intra-scanner) reliability with PET images from two groups of subjects (Ma et al., 2007). That study demonstrated the very high test-retest reliability of the SSF analysis. No study

has investigated inter-scanner reliability. The current studies investigated both intra- and inter-scanner reliability with the following 3 considerations.

First, we simulated a between-group design, i.e., comparison between two groups. SSM-PCA has been widely used to compare two groups of subjects, e.g., patient group vs. healthy group. The reliability of SSM-PCA is different from the reliability of other metrics. For example, the test-retest reliability of ALFF in RS-fMRI is usually tested in a single group of healthy subjects (Zou et al., 2015b), and it is relatively easy to scan a group of healthy subjects twice. But SSM-PCA should be performed on the between-group design to gain the reliability of the pattern as well as its expression for each subject, i.e., SSF. It is rather difficult to scan both the patient group and healthy group twice, especially in two different scanners. Moreover, the brain activity in the patients usually changes more than that in the healthy group over time, which affects the intra- and inter-scanner reliability. Therefore, the current study simulated two groups of subjects with a single group of healthy subjects to investigate both the intra-scanner (i.e., test-retest) reliability and the inter-scanner reliability.

Second, we used a RS-fMRI dataset under Eyes Open (EO) and Eyes Closed (EC) conditions. In RS-fMRI, EO and EC states are two resting physiological states with distinct differences in a few brain regions, and more importantly, these differences are highly reproducible across studies (Yang et al., 2007; Yan et al., 2009; Liu et al., 2013; Yuan et al., 2014; Zou et al., 2015a). EO and EC are usually for within-group designs. By randomly dividing one group into two subgroups and performing subtraction between conditions, e.g., EC-EO group and EO-EC group, between-group designs can be imitated with within-group data.

Third, we compared the spatial patterns generated by the multivariate method of SSM-PCA with that by the univariate statistical method of voxel-wise *T*-test. With a proper threshold, the surviving brain voxels of a *T* map implies the existence of significant difference between two groups (or conditions). By contrast, the surviving brain voxels of a SSM-PCA map mean that these voxels contribute more than other voxels to the difference between groups. We therefore were interested in studying the similarities and differences between the *T* map and SSM-PCA pattern with a certain threshold.

## MATERIALS AND METHODS

### Subjects

The experiment was approved by the ethics committee at the Center for Cognition and Brain Disorders, Hangzhou Normal University (HZNU). Signed informed consent was obtained from all subjects prior to data acquisition. Twenty-one healthy subjects ( $21.8 \pm 1.8$  years old, 11 females) participated in all 3 visits of MRI scans. All subjects were prescreened with a telephone questionnaire to exclude history of neurological illness or psychiatric disorders.

### Data Acquisition

RS-fMRI dataset was obtained on 3 visits (V1, V2, and V3), with V1 and V2 (separated by  $14 \pm 1$  days) on a scanner and V3 ( $230 \pm$

8 days from V2) on a different scanner. For each visit, participants underwent two RS-fMRI scans, during which they were asked to relax with either EC or EO. The order of the two acquisitions was counter-balanced across subjects.

MR images of V1 and V2 were obtained on a GE 3T scanner (MR-750, GE Medical Systems, Milwaukee, WI) with an eight-channel head coil at the Center for Cognition and Brain Disorders of HZNU. To minimize the head movement, subjects laid supine with their heads snugly fixed by straps and foam pads. The Blood-Oxygenation-Level-Dependent (BOLD) images were acquired using a gradient echo Echo-Planar Imaging (EPI) pulse sequence with the following parameters: Repetition Time (TR)/Echo Time (TE) = 2,000/30 ms, Flip Angle (FA) = 60°, 43 slices with interleaved acquisition, thickness/gap = 3.4/0 mm, Field Of View (FOV) = 220 × 220 mm<sup>2</sup> with an in-plane resolution of 3.44 × 3.44 mm<sup>2</sup>. The duration of each resting-state scan was 8 min. A high-resolution 3D volume imaging was performed with a spoiled gradient-recalled pulse sequence (176 sagittal slices, thickness = 1 mm, TR/TE = 8.1/3.1 ms, FA = 9°, FOV = 250 × 250 mm<sup>2</sup>).

Data of V3 were acquired on a Siemens 3T scanner (Prisma, Siemens Healthineers, Erlangen, Germany) at the Center for Brain Imaging Science and Technology of Zhejiang University (ZJU). The BOLD EPI sequence parameters were the same as those on the GE scanner except FA = 90°. The 3D T1-weighted images were acquired with a Magnetization-Prepared Rapid-Acquisition Gradient Echo (MPRAGE) sequence (176 sagittal slices, thickness = 1 mm, TR/TE = 1,800/2.28 ms, inversion time = 755 ms, FA = 8°, echo spacing = 7.1 ms, turbo factor = 208, FOV = 250 × 250 mm<sup>2</sup>).

## Data Preprocessing

Functional MRI data were preprocessed with Resting-State fMRI Data Analysis Toolkit plus V1.2 (RESTplus V1.2, <http://restfmri.net/forum/index.php>). Preprocessing procedures included removal of the first 10 frames, slice-timing correction, realignment to the first image for motion correction, coregistration of individual averaged functional images to T1 images, and spatially normalization into the standard Montreal Neurological Institute (MNI) brain space using the deformation field from segmentation of T1 images. All images were then resampled into 3 × 3 × 3 mm<sup>3</sup> voxels, and smoothed using an isotropic Gaussian filter with a Full Width at Half Maximum (FWHM) of 6 mm. For all the subjects, the maximum translation and rotation were less than 1.5 mm and 1.5°, respectively. After removing the linear drift, ALFF was calculated based on the same procedures reported previously (Zang et al., 2007) with RESTplus. Briefly, the time courses of RS-fMRI signal were first converted to frequency domain with the Fast Fourier Transform (FFT). Then, the averaged amplitude across a frequency band of 0.01–0.08 Hz yielded ALFF. For each subject, the ALFF map was divided by the global mean ALFF value within a whole brain mask in RESTplus.

## SSM-PCA Analysis

In most previous applications of SSM-PCA, image data from both patients and healthy controls were put together for analysis, and

then a disease-related spatial covariance pattern was identified if significant difference in a pattern's expression was found between the patients and healthy controls by two-sample *T*-test (Alexander and Moeller, 1994; Spetsieris and Eidelberg, 2011; Tomšė et al., 2017). The current study was a within-group design, i.e., comparison between two conditions within the same group of subjects, which is also useful for longitudinal follow-up or intervention studies in clinical research. To imitate the analytic procedure in most existing SSM-PCA studies, 21 subjects in this study were randomly divided into two groups with matched age and gender, i.e., EC-EO group (*n* = 11) and EO-EC group (*n* = 10). In detail, for the EO-EC group, we subtracted ALFF map of EC from ALFF map of EO for each subject to generate the difference map, and vice versa for the EC-EO group.

Based on a modified PCA, SSM-PCA decomposes the metric (ALFF in the current study) maps from all the subjects (EC-EO group and EO-EC group here) into a linear combination of orthogonal components. Each component is a whole-brain image, usually named as a “pattern.” Each voxel's value of any component is a weight representing the contribution of that voxel to the corresponding pattern. Voxels with a relatively large weight in the pattern was called “network” in Spetsieris and colleagues' paper on metabolic PET (Spetsieris et al., 2015). The pattern was also termed as Group Invariant Subprofile (GIS) (Moeller and Strother, 1991; Alexander and Moeller, 1994). The projection of each individual's ALFF map onto a pattern is regarded as the pattern's expression in the subject, which is also called Subject Scaling Factor (SSF). The SSFs are then used in further group-level statistical analysis, e.g., *T*-test between two groups or correlation with behavioral variables.

The mathematical basis for SSM-PCA has been previously described in detail (Moeller and Strother, 1991; Alexander and Moeller, 1994; Spetsieris and Eidelberg, 2011). Briefly, the difference ALFF maps were arranged first into an  $M \times N$  dimensional data matrix, where each column represents all the voxels from each subject.  $M$  is the number of voxels and  $N$  is the number of subjects. Secondly, we centered each column to zero, and then acquired the Group Mean Profile (GMP) as the mean value of each row. Thirdly, the data matrix of each row was centered to zero with GMP to obtain the Subject Residual Profile (SRP). As in regular PCA, the reduced Singular Value Decomposition (SVD) was utilized to factorize SRP (Jolliffe, 2002; Spetsieris and Eidelberg, 2011):

$$U \Sigma V^T = \text{SVD}(\text{SRP}), \quad (1)$$

where  $U$  is a  $M \times N$  matrix composed of the left unit-normalized orthogonal singular vectors as columns,  $\Sigma$  is a  $N \times N$  diagonal matrix composed of singular values  $\sigma_k$ , where  $k$  is the component number, and  $V$  is a  $N \times N$  matrix composed of the right unit-normalized orthogonal singular vectors as columns. Then, the GISs (namely patterns) and SSFs (namely patterns' expressions in each subject) can be computed as follows:

$$\text{GIS}_{ik} = U_{ik}, \quad (2)$$

$$\text{SSF}_{jk} = \sum_{i=1}^M (\text{SRP}_{ij} \times \text{GIS}_{ik}) \quad (3)$$

in which  $i$  is the voxel number and  $j$  is the subject number. Variance Accounting For (VAF) represents the ratio of variance corresponding to every GIS to the total variance, calculated by:

$$VAF_k = \sigma_k^2 / \sum_{k=1}^N \sigma_k^2. \quad (4)$$

Two-sample  $T$ -test was then performed on SSFs to assess their difference between the EC-EO group and EO-EC group. GIS, whose SSF with  $p < 0.05$ , was considered to be the EC and EO difference-related spatial covariance pattern (hereafter named as EOEC-pattern). The GISs and SSFs were derived based on the SSMPCA toolbox (<http://www.feinsteinneuroscience.org>).

## Intra-scanner and Inter-scanner Reliability Analysis

The intra-scanner and inter-scanner reliability of SSF from SSM-PCA was measured with Intra-Class Correlation (ICC). ICC for each pair of metrics from two visits was calculated as below (Shrout and Fleiss, 1979):

$$ICC = (BMS - WMS) / (BMS + WMS), \quad (5)$$

where BMS and WMS are the mean squares values of between-target and within-target SSFs. To illustrate the similarity between EOEC-patterns and their SSFs between each pair of MRI scans, Pearson correlation coefficient ( $r$ ) was also calculated (Zang et al., 2017). The effect of group division on the ICC of SSF was investigated by repeating the SSM-PCA with random division of subjects for 1,000 times with bootstrapping.

## EOEC-Pattern Generalization Across Visits

To investigate the generalization of EOEC-pattern across serial MRI datasets, we used Topographic Profile Rating (TPR) algorithm (Eidelberg et al., 1995; Ma et al., 2007). TPR quantifies the expression of a given pattern in an individual subject by the inner product of the pattern and the individual subject's SRP. The individual subject's SRP is acquired by using the GMP image associated with the derivation of the original pattern. For example, an EOEC-pattern was obtained from V1 data and then projected onto V2 and V3 data to compute SSFs. Two-sample  $T$ -test was performed to compare between the EC-EO group and EO-EC group for V2 and V3, respectively. ICC of V2 against V3 was also calculated as an additional way for measuring reliability, as was done by Ma and colleagues (Ma et al., 2007). Similarly, the EOEC-pattern from V2 or V3 was projected onto the data in the other two visits, and intra- or inter-scanner reliability was measured, in addition to the other indicators of reliability described in section Intra-Scanner and Inter-Scanner Reliability Analysis above.

## Comparison Between EOEC-Pattern and Univariate Statistical $T$ Map

In order to compare the EOEC-patterns from SSM-PCA and the  $T$  maps from univariate statistics, Dice Similarity Coefficient (DSC) was utilized. Univariate two-sample  $T$ -test was performed between EC-EO group and EO-EC group. A corrected  $p < 0.05$  was used with AlphaSim in software RESTplus V1.2. This

corrected  $p$  value corresponded to a combined threshold of single voxel  $p < 0.05$  and cluster sizes larger than a certain number of voxels, which was determined on an estimated smoothing kernel size (Full Width at Half Maximum (FWHM) listed in **Table 1**) according to their  $T$  maps. It should be noted that there is currently no widely accepted method to determine the threshold for SSM-PCA pattern maps. To render it more comparable with the  $T$  map, we used the same cluster size threshold (**Table 1**) for the  $z$ -transformed EOEC-pattern map, sorted the absolute  $z$  value, and then determined the  $|z|$  threshold (**Table 1**), by which the total number of voxels of EOEC-patterns were kept almost the same as that of  $T$  maps (**Table 1**). DSC was computed as below (Rombouts et al., 1997):

$$DSC = 2|A \cap B| / (|A| + |B|), \quad (6)$$

where  $|A|$ ,  $|B|$ , and  $|A \cap B|$  are the total voxel numbers of the EOEC-pattern,  $T$  map, and their overlap, respectively.

## RESULTS

### EOEC-Pattern Identification

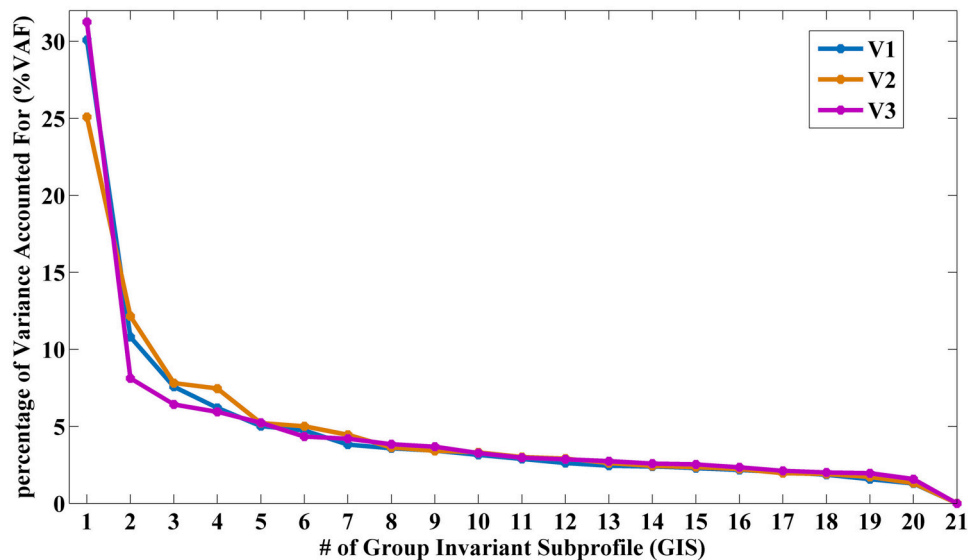
As shown in **Figure 1** and Supplementary Table 1, the VAF of GIS1 for V1, V2, and V3 was remarkably larger than that of GIS2. Only the SSF of GIS1 showed significant difference ( $p < 0.05$ ) between the EC-EO group and EO-EC group for each visit. Thus, GIS1 was named as the EOEC-pattern map for V1–3. As the VAF for GIS21 was zero, GIS21 and its SSF were ignored in  $T$ -test and reliability analysis. **Figures 2A–C** displayed the topography of  $z$ -transformed EOEC-patterns with a threshold of  $|z| > 1$  and their SSFs in V1–3. Combining the SSF distribution and EOEC-pattern, positive  $z$  represented higher ALFF in EC than EO, mainly including the visual cortex, temporal cortex, and sensorimotor cortex. Inversely, negative  $z$  represented lower ALFF in EC than EO, mainly involving frontal pole and posterior parietal cortex. For each of these patterns, SSF values were significantly elevated in the EC-EO group compared to the EO-EC group (**Figures 2D–F**).

**TABLE 1 |** Parameters used in calculating DSC for comparison between EOEC-pattern and univariate statistical  $T$  map.

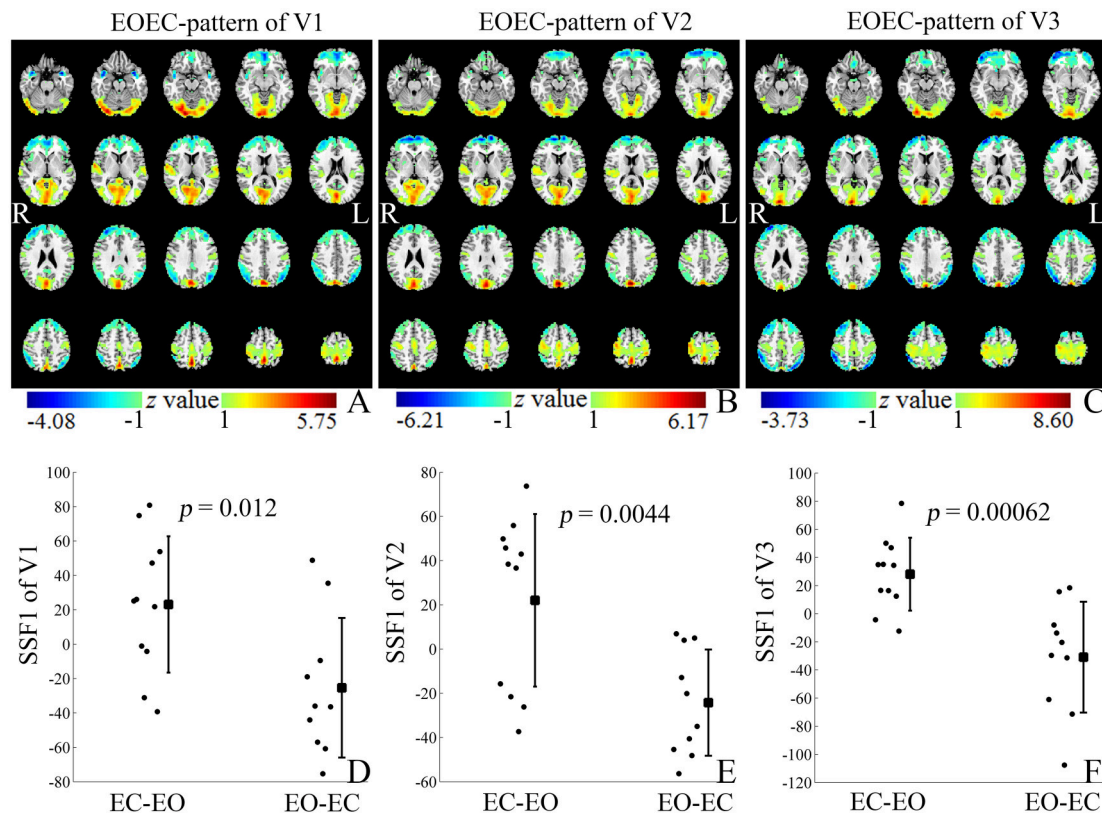
	V1	V2	V3
FWHM (mm) for $T$ map	[12.1 12.9 11.3]	[11.9 12.6 11.2]	[13.0 13.7 13.6]
Cluster size threshold (voxel)	915	952	1137
Voxel number in $T$ map	7669	10246	14913
$ z $ threshold for EOEC-pattern	1.40	1.27	1.01
Voxel number in EOEC-pattern	7668	10246	14911

DSC, Dice Similarity Coefficient; FWHM, Full Width at Half Maximum.





**FIGURE 1** | The percentage of VAF (%VAF) for each GIS in V1, V2, and V3. (VAF, Variance Accounting For; GIS, Group Invariant Subprofile).



**FIGURE 2** | The z-transformed EOE-patterns (with  $|z| > 1$ ) (A–C) and their SSFs (D–F) of datasets V1–3. Positive and negative z values represented higher and lower ALFF in EC than EO, respectively. The z coordinates of each slice were from  $-25$  to  $70$  mm with slice spacing of  $5$  mm (SSF, Subject Scaling Factor).

## Intra- and Inter-scanner Reliability of EOEC-Pattern and its Expression Reliability by ICC

As shown in **Figure 3**, the PCC between both intra- and inter-scanner EOEC-patterns was high (above 0.8,  $p < 0.001$ ) (**Figures 3A–C**). The intra-scanner reliability of EOEC-patterns' expressions, i.e., SSFs, was fair (0.4~0.59) (Cicchetti, 1994). Interestingly, the inter-scanner reliability of SSF was good (0.6~0.74) (**Figures 3D–F**; Cicchetti, 1994). The PCC between each pair of visits was very similar with ICC. Supplementary Table 2 listed the intra- and inter-scanner ICCs of all the SSFs corresponding to GIS1-20. Except for ICC of SSF1, all ICCs of SSF2-20 were smaller than 0.4. The mean value, standard deviation, and 95% confidence interval of ICC of EOEC-patterns' expressions from bootstrapping were listed in **Table 2**. The table demonstrated that the ICC variation is very small compared with the mean value.

### EOEC-Pattern Generalization Across Visits

We calculated the expression (i.e., SSF) of each EOEC-pattern (V1, V2, and V3) on the other two datasets. As shown in **Table 3**, the ICCs between expressions of a given EOEC-pattern in the other two datasets was approximate to those between each pair of SSFs obtained from the EOEC-patterns of their own visits (**Figures 3D–F**). Two-sample *T*-test showed excellent cross-validation results (**Table 3**).

### Comparison Between EOEC-Pattern and Univariate Statistical *T* Map

**Figures 4A–C** showed the univariate statistical *T* maps between the EC-EO group and EO-EC group with a combined  $p$  threshold and cluster size threshold described in **Table 1**. As shown in **Figures 4G–I**, the DSCs for V1, V2, and V3 were 0.27, 0.31, and 0.37, respectively, suggesting that the EOEC-pattern was quite different from the *T* maps. By visual inspection on **Figures 4G–I**, *T*-test detected larger, but not exclusively, areas in the primary sensorimotor area and superior temporal gyrus, whereas EOEC-pattern detected exclusively large area in the occipital lobe.

## DISCUSSION

### EOEC-Pattern Identification

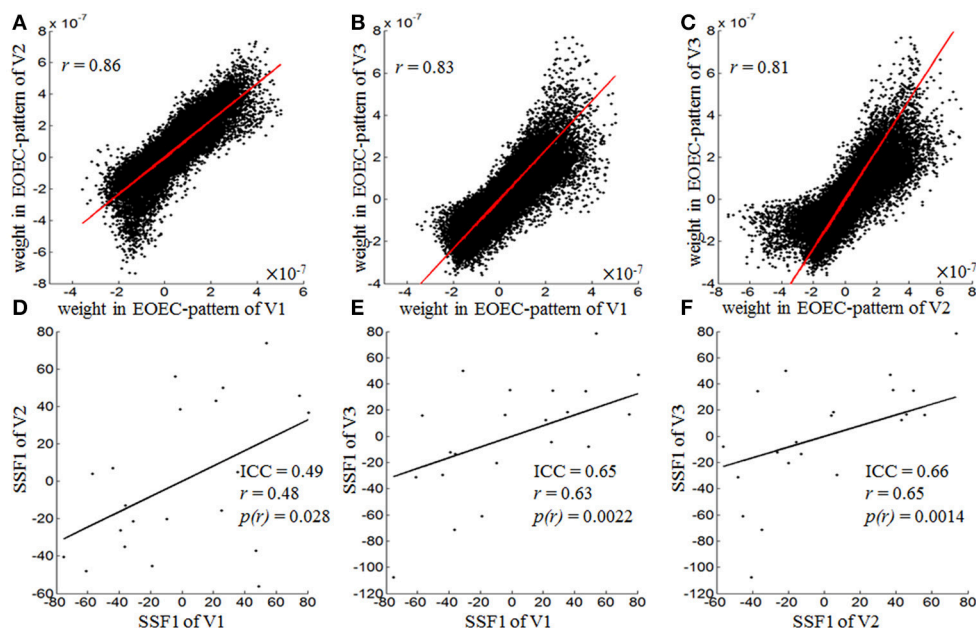
The first pattern, namely EOEC-pattern, accounted substantially more variance than that of the second GIS (GIS2) in each visit (**Figure 1** and Supplementary Table 1). The EOEC-pattern included the primary sensorimotor cortex, visual cortex, and frontal cortex (**Figures 2A–C**). Similar brain areas were also reported in previous EO and EC studies by using paired *T*-tests (Yang et al., 2007; Yan et al., 2009; Liu et al., 2013; Yuan et al., 2014; Zou et al., 2015a). The SSF showed significant differences between the EC-EO group and EO-EC group (**Figures 2D–F**). Many previous PET studies have consistently found that the first pattern was the disease-related pattern (Ma et al., 2007; Pagani et al., 2016; Tomš̃e et al., 2017). After centralization, the voxel-wise similarity between groups was reduced, while the difference between groups was

highlighted (Moeller and Strother, 1991). The current study randomly assigned one group of subjects into EC-EO and EO-EC subgroups, and applied SSM-PCA in the same way as in previous studies. Thus, it is not surprising why the first pattern accounted for the largest portion of total variance, and hence, is the “between-group” difference-related pattern, i.e., EOEC-pattern.

### Intra- and Inter-scanner Reliability of SSM-PCA

We firstly found high similarity among the EOEC-patterns of the 3 visits. We then calculated the reliability of SSF (i.e., the expression) corresponding to EOEC-patterns. Both the intra- and inter-scanner reliability of SSF was fair to good (ICC = 0.49–0.66) (**Figure 3**). Furthermore, we calculated the SSF of an EOEC-pattern of one visit onto the other two visits. For example, the EOEC-pattern of V1 was expressed onto V2 and V3, respectively. We found that each EOEC-pattern of one visit could be successfully applied to other two visits to differentiate the EC-EO group and EO-EC group ( $p < 0.01$ ) (**Table 3**). The ICC value was similar as before (**Table 3** vs. **Figure 3**).

To the best of our knowledge, only one previous study investigated the test-retest reliability of SSM-PCA in two groups of subjects (Ma et al., 2007). Ma and colleagues obtained a Parkinson's disease (PD)-related pattern (PDRP) from a dataset of PD patients and healthy controls. Then this PDRP was expressed onto other datasets to measure the ICC of PDRP' expression (i.e., SSF). They found excellent test-retest reliability over different intervals including 1 h apart (ICC = 0.94 for healthy subjects, and ICC = 0.96 for unmediated PD patients), 1 day apart (ICC = 0.99 for unmedicated early state patients), and 2 months apart (ICC = 0.96 for medicated moderate stage PD patients). These results suggest that the test-retest reliability of FDG PET was higher than that with RS-fMRI ALFF in the present study. This discrepancy might be attributed to differences in the experimental design, imaging modality, as well as computing algorithm of imaging metrics. Simultaneous resting-state PET-fMRI studies have shown that only a small part of brain regions demonstrated significant voxel-level correlation between glucose metabolism and RS-fMRI metrics including ALFF (Aiello et al., 2015; Bernier et al., 2017). Generally speaking, PET and fMRI measure different physiological features. However, the different computation for the two techniques may also account for their discrepancies. The metric for PET glucose metabolism is usually the averaged or integrated value over a period of time, while ALFF of RS-fMRI is the amplitude of fluctuation over time (Zang et al., 2007). A non-invasive perfusion-weighted MRI technique, arterial spin labeling (ASL) is widely used to measure CBF. Some ASL sequences allow calculating both mean CBF over a period of time and CBF-ALFF. A study used ASL and BOLD RS-fMRI to compare between EO and EC states (Zou et al., 2015b). ASL-ALFF and BOLD-ALFF detected similar regions, but CBF-ALFF and CBF-mean detected very different regions. Zou and colleagues also found that CBF-mean showed better test-retest reliability than BOLD-ALFF (Zou et al., 2015a).



**FIGURE 3 |** Pearson correlation of EOEC-patterns (A–C) and reliability of their expressions (D–F), i.e., SSF1, of each pair of visits. The Pearson correlation coefficients for EOEC-patterns of V1 vs. V2 (intra-scanner), V1 vs. V3 (inter-scanner), and V2 vs. V3 (inter-scanner) were 0.86, 0.83, and 0.81, respectively, and the ICCs for EOEC-pattern expressions of V1 vs. V2, V1 vs. V3, and V2 vs. V3 were 0.49, 0.65, and 0.66, respectively (SSF1, Subject Scaling Factor of G1S1, namely EOEC-pattern here; ICC, Intra-Class Correlation;  $r$ , Pearson Correlation Coefficient).

**TABLE 2 |** ICC results of EOEC-pattern's expression from the random group selection for 1,000 times with bootstrapping.

		V1 vs. V2	V1 vs. V3	V2 vs. V3
EOEC-pattern's expression	Mean $\pm$ std	0.48 $\pm$ 0.04	0.61 $\pm$ 0.03	0.66 $\pm$ 0.03
	95% confidence interval	[0.43 0.55]	[0.57 0.66]	[0.62 0.67]

ICC, Intra-Class Correlation.

**TABLE 3 |** EOEC-pattern generalization results across visits.

	Expression of EOEC-pattern of V1		Expression of EOEC-pattern of V2		Expression of EOEC-pattern of V3	
	V2	V3	V1	V3	V1	V2
$p$ -value	0.0082	0.00099	0.0072	0.00038	0.0084	0.0054
$T$ -value	2.95	3.89	3.01	4.31	2.94	3.14
Cohen $d$	1.09	1.30	1.11	1.37	1.09	1.14
ICC	0.64		0.71		0.46	

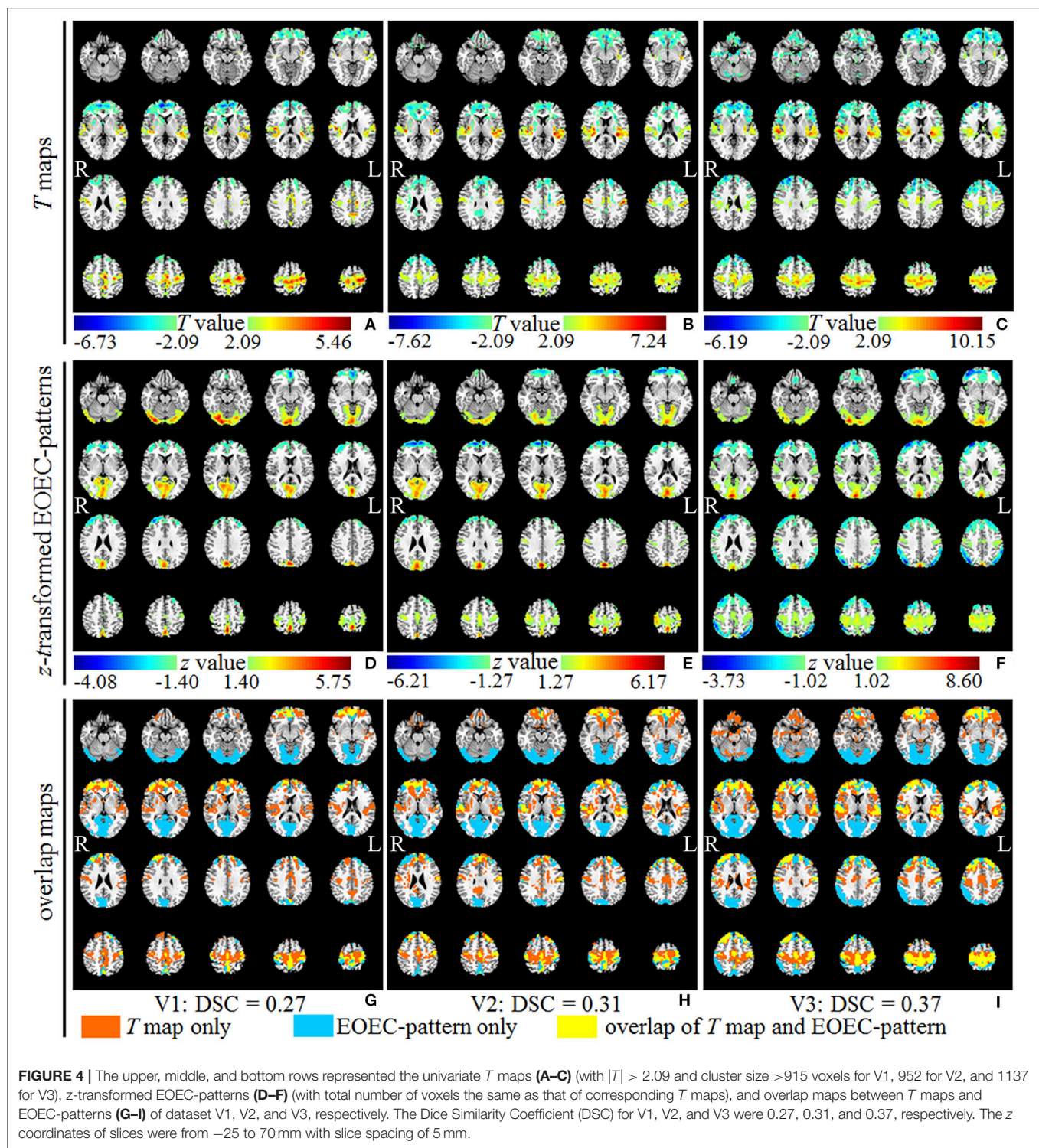
ICC, Intra-Class Correlation. For example, the  $p$  value of 0.0084 meant the  $T$ -test result for the expression of V3's EOEC-pattern in V1 comparing the EC-EO and EO-EC groups. ICC of 0.46 is the ICC between the expressions of V3's EOEC-pattern in V1 and V2 across all subjects.

## Comparison Between EOEC-Pattern and Univariate Statistical $T$ Map

Both SSM-PCA and univariate  $T$ -test are statistical methods, however, they are quite different in their mathematical

foundations. As a multivariate statistical approach, SSM-PCA obtains patterns and the patterns' expression of each subject based on the covariance matrix of all the voxels from all the subjects, which is a kind of pattern analysis (Alexander and Moeller, 1994; Eidelberg, 2009). The patterns are whole-brain images. Then two-sample  $T$ -test is applied to the patterns' expression (i.e., SSF) to assess whether the SSF is different between two groups of subjects. If the difference is significant, the corresponding pattern is then named as difference-related pattern, e.g., EOEC-related pattern in the current study or PD-related pattern (PDRP) in previous studies (Ma et al., 2007; Wu et al., 2013, 2015; Tomše et al., 2017). For a voxel in the difference-related pattern, larger value means more contribution or weight to the difference. On the other hand, for univariate statistical method such as voxel-wise  $T$ -test, comparison is made between the values of each single voxel from two groups of subjects (two-sample  $T$ -test) or two conditions within a group of subjects (one-sample  $T$ -test). The total number of comparisons is very different for the SSM-PCA and  $T$ -test. For SSM-PCA, the total number of comparisons is up to the total number of patterns (up to 21 in the current study), but usually only a few principle patterns are taken into account. And in practice like in the current study and previous studies (Ma et al., 2007; Pagani et al., 2016; Tomše et al., 2017), only the first one component was used because the first component accounts much more variance than the second one. For voxel-wise  $T$ -test, the total number of comparisons is the total number of voxels (70,831 voxels in the current study). Therefore, false discovery problem due to multiple comparisons is much more severe for univariate statistical method in neuroimaging studies (Poldrack et al., 2011).





The aforementioned points reflected merely a theoretical issue from methodological perspective. From neurophysiological perspectives, SSM-PCA is operating on the notion that localized changes engage multiple, interacting brain regions that are widely distributed over the whole brain owe to intrinsic

connectivity in neural substrates. A primary example to support this view is the modulation of SSM-PCA pattern and clinical correlation by neurosurgical interventions delivered locally on any key nodes in the pattern (Peng et al., 2014). On the other hand,  $T$ -test is relying on mean signal in image data to



localize regionally- independent group differences over the whole brain. This method does not explicitly account for important functional interactions between different brain regions except neighborhood autocorrelations on a small scale inherent in image data.

We compared the EOEC-pattern with *T*-test pattern. While the results showed overlapped brain regions, they also showed method-specific brain regions. The *T*-test detected larger brain regions in the primary sensorimotor area and superior temporal gyrus, but SSM-PCA detected exclusively large visual area. In a previous research study with voxel-wise paired *T*-test analysis, it was also found that the mean CBF from ASL technique was significantly lower under EC than EO conditions in the primary visual cortex, which was not detected with ALFF *T*-test (Zou et al., 2015a). Moreover, it is well known that the visual cortex can be activated by visual input, so it is reasonable to detect visual area in the EOEC-pattern. Differences that are not significant enough in *T*-test may show up in pattern from SSM-PCA as reported in PET literatures (Habeck et al., 2008; Ma et al., 2009) and ASL literature (Asllani et al., 2008). Consequently, SSM-PCA and univariate *T*-test are two complementary data analytic approaches for application studies.

## Limitations

One limitation is the experimental design for inter-scanner reliability. The current study aimed to investigate both intra- and inter-scanner reliability. When we kept the interval of the two visits of intra-scanner scanning to be similar across subjects, it is impossible to keep the order of inter-scanner scanning count-balanced. Therefore, the second visit was always before the third visit. It means that the reliability between the second and third scanning is a mixed effect of inter-scanner and test-retest

reliability. Another limitation is using within-group designs to simulate between-group designs for SSM-PCA.

## CONCLUSIONS

Both the intra- and inter-scanner reliability of SSM-PCA of RS-fMRI ALFF was fair to good. The difference-related pattern of SSM-PCA and *T* maps was similar but also showed method-specific brain regions, indicating that the SSM-PCA and *T*-test are two complementary statistical methods.

## AUTHOR CONTRIBUTIONS

Y-FZ, YM, and H-JH conceived and designed the experiment. L-XY, J-BW, and NZ performed the data analysis. L-XY, Y-YL, and D-QL acquired the data. J-BW, YM, and J-HZ provided advices on the analysis and interpretation of the results. L-XY, J-BW, H-JH, YM, and Y-FZ wrote the paper.

## FUNDING

This work was supported by grants from the National Natural Science Foundation of China (81201083, 81661148045, 81520108016, 81271652, 31471084, 81401473, and 91632109), the Fundamental Research Funds for the Central Universities (2017QNA5016) of China. Y-FZ is partly supported by Qian Jiang Distinguished Professor program.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00311/full#supplementary-material>

## REFERENCES

- Aiello, M., Salvatore, E., Cachia, A., Pappatà, S., Cavaliere, C., Prinster, A., et al. (2015). Relationship between simultaneously acquired resting-state regional cerebral glucose metabolism and functional MRI: a PET/MR hybrid scanner study. *Neuroimage* 113, 111–121. doi: 10.1016/j.neuroimage.2015.03.017
- Alexander, G. E., and Moeller, J. R. (1994). Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: a principle component approach to modeling brain function in disease. *Hum. Brain Mapp.* 2, 79–94. doi: 10.1002/hbm.460020108
- Anderson, N. E., Posner, J. B., Sidtis, J. J., Moeller, J. R., Strother, S. C., Dhawan, V., et al. (1988). The metabolic anatomy of paraneoplastic cerebellar degeneration. *Ann. Neurol.* 23, 533–540. doi: 10.1002/ana.410230602
- Asllani, I., Habeck, C., Scarmeas, N., Borogovac, A., Brown, T. R., and Stern, Y. (2008). Multivariate and univariate analysis of continuous arterial spin labeling perfusion MRI in Alzheimer's disease. *J. Cereb. Blood Flow Metab.* 28, 725–736. doi: 10.1038/sj.jcbfm.9600570
- Bennett, C. M., and Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Ann. N.Y. Acad. Sci.* 1191, 133–155. doi: 10.1111/j.1749-6632.2010.05446.x
- Bergfield, K. L., Hanson, K. D., Chen, K., Teipel, S. J., Hampel, H., Rapoport, S. I., et al. (2010). Age-related networks of regional covariance in MRI gray matter: reproducible multivariate patterns in healthy aging. *Neuroimage* 49, 1750–1759. doi: 10.1016/j.neuroimage.2009.09.051
- Bernier, M., Croteau, E., Castellano, C. A., Cunnane, S. C., and Whittingstall, K. (2017). Spatial distribution of resting-state BOLD regional homogeneity as a predictor of brain glucose uptake: a study in healthy aging. *Neuroimage* 150, 14–22. doi: 10.1016/j.neuroimage.2017.01.055
- Brickman, A. M., Habeck, C., Ramos, M. A., Scarmeas, N., and Stern, Y. (2008). A forward application of age associated gray and white matter networks. *Hum. Brain Mapp.* 29, 1139–1146. doi: 10.1002/hbm.20452
- Brickman, A. M., Habeck, C., Zarahn, E., Flynn, J., and Stern, Y. (2007). Structural MRI covariance patterns associated with normal aging and neuropsychological functioning. *Neurobiol. Aging* 28, 284–295. doi: 10.1016/j.neurobiolaging.2005.12.016
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284. doi: 10.1037/1040-3590.6.4.284
- Eidelberg, D. (2009). Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. *Trends Neurosci.* 32, 548–557. doi: 10.1016/j.tins.2009.06.003
- Eidelberg, D., Moeller, J. R., Ishikawa, T., Dhawan, V., Spetsieris, P., Chaly, T., et al. (1995). Assessment of disease severity in parkinsonism with fluorine-18-fluorodeoxyglucose and PET. *J. Nucl. Med.* 36, 378–383.
- Gazes, Y., Bowman, F. D. B., Razlighi, Q. R., O'Shea, D., Stern, Y., and Habeck, C. (2016). White matter tract covariance patterns predict age-declining cognitive abilities. *NeuroImage* 125, 53–60. doi: 10.1016/j.neuroimage.2015.10.016

- Guo, X., Chen, K., Zhang, Y., Wang, Y., and Yao, L. (2014). Regional covariance patterns of gray matter alterations in Alzheimer's disease and its replicability evaluation. *J. Magn. Reson. Imaging* 39, 143–149. doi: 10.1002/jmri.24143
- Habeck, C., Foster, N. L., Perneczky, R., Kurz, A., Alexopoulos, P., Koeppe, R. A., et al. (2008). Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. *NeuroImage* 40, 1503–1515. doi: 10.1016/j.neuroimage.2008.01.056
- Jolliffe, I. (2002). *Principal Component Analysis, Second Edition*. New York, NY: Springer-Verlag.
- Liu, D., Dong, Z., Zuo, X., Wang, J., and Zang, Y. (2013). Eyes-open/eyes-closed dataset sharing for reproducibility evaluation of resting state fMRI data analysis methods. *Neuroinformatics* 11, 469–476. doi: 10.1007/s12021-013-9187-0
- Ma, Y., Tang, C., Moeller, J. R., and Eidelberg, D. (2009). Abnormal regional brain function in Parkinson's disease: truth or fiction? *NeuroImage* 45, 260–266. doi: 10.1016/j.neuroimage.2008.09.052
- Ma, Y., Tang, C., Spetsieris, P. G., Dhawan, V., and Eidelberg, D. (2007). Abnormal metabolic network activity in parkinson's disease: test-retest reproducibility. *J. Cereb. Blood Flow Metab.* 27, 597–605. doi: 10.1038/sj.jcbfm.9600358
- Moeller, J. R., and Strother, S. C. (1991). A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *J. Cereb. Blood Flow Metab.* 11, A121–A135. doi: 10.1038/jcbfm.1991.47
- Moeller, J. R., Strother, S. C., Sidtis, J. J., and Rottenberg, D. A. (1987). Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. *J. Cereb. Blood Flow Metab.* 7, 649–658. doi: 10.1038/jcbfm.1987.118
- Pagani, M., Giuliani, A., Öberg, J., Chincarini, A., Morbelli, S., Brugnolo, A., et al. (2016). Predicting the transition from normal aging to Alzheimer's disease: a statistical mechanistic evaluation of FDG-PET data. *NeuroImage* 141, 282–290. doi: 10.1016/j.neuroimage.2016.07.043
- Peng, S., Eidelberg, D., and Ma, Y. (2014). Brain network markers of abnormal cerebral glucose metabolism and blood flow in Parkinson's disease. *Neurosci. Bull.* 30, 823–837. doi: 10.1007/s12264-014-1472-x
- Poldrack, A., Russell, J., Mumford, A., and Jeanette; Nichols, E., T. (2011). *Handbook of Functional MRI Data Analysis*. New York, NY: Cambridge University Press.
- Rombouts, S. A. R. B., Barkhof, F., Hoogenraad, F. G. C., Sprenger, M., Valk, J., and Scheltens, P. (1997). Test-retest analysis with functional mr of the activated area in the human visual cortex. *Am. J. Neuroradiol.* 18, 1317–1322.
- Rottenberg, D. A., Moeller, J. R., Strother, S. C., Sidtis, J. J., Navia, B. A., Dhawan, V., et al. (1987). The metabolic pathology of the AIDS dementia complex. *Ann. Neurol.* 22, 700–706. doi: 10.1002/ana.410220605
- Sackeim, H. A., Prohovnik, I., Moeller, J. R., Mayeux, R., Stern, Y., and Devanand, D. P. (1993). Regional cerebral blood flow in mood disorders. II. Comparison of major depression and Alzheimer's disease. *J. Nuclear Med.* 34, 1090–1101.
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Spetsieris, P. G., and Eidelberg, D. (2011). Scaled subprofile modeling of resting state imaging data in Parkinson's disease: methodological issues. *NeuroImage* 54, 2899–2914. doi: 10.1016/j.neuroimage.2010.10.025
- Spetsieris, P. G., Ko, J. H., Tang, C. C., Nazem, A., Sako, W., Peng, S., et al. (2015). Metabolic resting-state brain networks in health and disease. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2563–2568. doi: 10.1073/pnas.1411011112
- Steffener, J., Habeck, C., O'Shea, D., Razlighi, Q., Bherer, L., and Stern, Y. (2016). Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiol. Aging* 40, 138–144. doi: 10.1016/j.neurobiolaging.2016.01.014
- Teune, L. K., Renken, R. J., De Jong, B. M., Willemsen, A. T., Van Osch, M. J., Roerdink, J. B. T. M., et al. (2014). Parkinson's disease-related perfusion and glucose metabolic brain patterns identified with PCASL-MRI and FDG-PET imaging. *NeuroImage Clin.* 5, 240–244. doi: 10.1016/j.nicl.2014.06.007
- Tomš, P., Jensterle, L., Grmek, M., Zaletel, K., Pirtošek, Z., Dhawan, V., et al. (2017). Abnormal metabolic brain network associated with Parkinson's disease: replication on a new European sample. *Neuroradiology* 59, 507–515. doi: 10.1007/s00234-017-1821-3
- Wu, P., Wang, J., Peng, S., Ma, Y., Zhang, H., Guan, Y., et al. (2013). Metabolic brain network in the Chinese patients with Parkinson's disease based on 18F-FDG PET imaging. *Parkinsonism Relat. Disord.* 19, 622–627. doi: 10.1016/j.parkreldis.2013.02.013
- Wu, T., Ma, Y., Zheng, Z., Peng, S., Wu, X., Eidelberg, D., et al. (2015). Parkinson's disease-related spatial covariance pattern identified with resting-state functional MRI. *J. Cereb. Blood Flow Metab.* 35, 1764–1770. doi: 10.1038/jcbfm.2015.118
- Yan, C., Liu, D., He, Y., Zou, Q., Zhu, C., Zuo, X., et al. (2009). Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. *PLoS ONE* 4:e5743. doi: 10.1371/journal.pone.0005743
- Yang, H., Long, X. Y., Yang, Y., Yan, H., Zhu, C. Z., Zhou, X. P., et al. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *NeuroImage* 36, 144–152. doi: 10.1016/j.neuroimage.2007.01.054
- Yuan, B. K., Wang, J., Zang, Y., and Liu, D. (2014). Amplitude differences in high-frequency fMRI signals between eyes open and eyes closed resting states. *Front. Hum. Neurosci.* 8:503. doi: 10.3389/fnhum.2014.00503
- Zang, Y., Feng, F., Huo, L., Li, B., Lan, X., Lu, J., et al. (2017). PET-fMRI for precise location of abnormal brain activity: a mini review. *Chin. J. Nucl. Med. Mol. Imaging* 37, 802–808. doi: 10.3760/cma.j.issn.2095-2848.2017.12
- Zang, Y. F., He, Y., Zhu, C., Cao, Q., Sui, M., Liang, M., et al. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev.* 29, 83–91. doi: 10.1016/j.braindev.2006.07.002
- Zou, Q., Miao, X., Liu, D., Wang, D. J. J., Zhuo, Y., and Gao, J. (2015a). Reliability comparison of spontaneous brain activities between BOLD and CBF contrasts in eyes-open and eyes-closed resting states. *NeuroImage* 121, 91–105. doi: 10.1016/j.neuroimage.2015.07.044
- Zou, Q., Yuan, B. K., Gu, H., Liu, D., Wang, D. J. J., Gao, J. H., et al. (2015b). Detecting static and dynamic differences between eyes-closed and eyes-open resting states using ASL and BOLD fMRI. *PLoS ONE* 10:e0121757. doi: 10.1371/journal.pone.0121757

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Yuan, Wang, Zhao, Li, Ma, Liu, He, Zhong and Zang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Improving the Reliability of Network Metrics in Structural Brain Networks by Integrating Different Network Weighting Strategies into a Single Graph

Stavros I. Dimitriadis<sup>1, 2, 3, 4, 5\*</sup>, Mark Drakesmith<sup>2, 5</sup>, Sonya Bells<sup>2, 3</sup>, Greg D. Parker<sup>2, 3</sup>, David E. Linden<sup>2, 5, 6</sup> and Derek K. Jones<sup>2, 5</sup>

<sup>1</sup> Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, United Kingdom, <sup>2</sup> Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom, <sup>3</sup> School of Psychology, Cardiff University, Cardiff, United Kingdom, <sup>4</sup> Neuroinformatics Group, Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom, <sup>5</sup> Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, United Kingdom, <sup>6</sup> MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Institute of Psychology (CAS), China

### Reviewed by:

Veena A. Nair,  
University of Wisconsin-Madison,  
United States  
Xiaoyun Liang,  
Florey Institute of Neuroscience and  
Mental Health, Australia

### \*Correspondence:

Stavros I. Dimitriadis  
stidimitriadis@gmail.com;  
dimitriadisS@cardiff.ac.uk

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 12 June 2017

**Accepted:** 27 November 2017

**Published:** 19 December 2017

### Citation:

Dimitriadis SI, Drakesmith M, Bells S,  
Parker GD, Linden DE and Jones DK  
(2017) Improving the Reliability of  
Network Metrics in Structural Brain  
Networks by Integrating Different  
Network Weighting Strategies into a  
Single Graph.  
Front. Neurosci. 11:694.  
doi: 10.3389/fnins.2017.00694

Structural brain networks estimated from diffusion MRI (dMRI) via tractography have been widely studied in healthy controls and patients with neurological and psychiatric diseases. However, few studies have addressed the reliability of derived network metrics both node-specific and network-wide. Different network weighting strategies (NWS) can be adopted to weight the strength of connection between two nodes yielding structural brain networks that are almost fully-weighted. Here, we scanned five healthy participants five times each, using a diffusion-weighted MRI protocol and computed edges between 90 regions of interest (ROI) from the Automated Anatomical Labeling (AAL) template. The edges were weighted according to nine different methods. We propose a linear combination of these nine NWS into a single graph using an appropriate diffusion distance metric. We refer to the resulting weighted graph as an Integrated Weighted Structural Brain Network (ISWBN). Additionally, we consider a topological filtering scheme that maximizes the information flow in the brain network under the constraint of the overall cost of the surviving connections. We compared each of the nine NWS and the ISWBN based on the improvement of: (a) intra-class correlation coefficient (ICC) of well-known network metrics, both node-wise and per network level; and (b) the recognition accuracy of each subject compared to the remainder of the cohort, as an attempt to access the uniqueness of the structural brain network for each subject, after first applying our proposed topological filtering scheme. Based on a threshold where the network level ICC should be  $>0.90$ , our findings revealed that six out of nine NWS lead to unreliable results at the network level, while all nine NWS were unreliable at the node level. In comparison, our proposed ISWBN performed as well as the best performing individual NWS at the network level, and the ICC was higher compared to all individual NWS at the node level. Importantly, both network and node-wise ICCs of network metrics derived from the topologically filtered ISWBN (ISWBN<sup>TF</sup>),

were further improved compared to the non-filtered ISWBN. Finally, in the recognition accuracy tests, we assigned each single ISWBN<sup>TF</sup> to the correct subject. We also applied our methodology to a second dataset of diffusion-weighted MRI in healthy controls and individuals with psychotic experience. Following a binary classification scheme, the classification performance based on ISWBN<sup>TF</sup> outperformed the nine different weighting strategies and the ISWBN. Overall, these findings suggest that the proposed methodology results in improved characterization of genuine between-subject differences in connectivity leading to the possibility of network-based structural phenotyping.

**Keywords:** connectome, diffusion MRI, structural brain network, tractography, reliability

## INTRODUCTION

Tractography is a popular method for extracting white matter connectivity from diffusion MRI (dMRI) and plays a key role in structural brain connectomics (Fornito et al., 2015). A variety of algorithms have been proposed, with the majority of them using voxel-based assessment of water diffusion to reveal paths/tracts of the white matter bundles. A fundamental problem with tractography is that there is no “ground truth” so it is impossible to separate “true” from spurious false positive and false negative connections (Smith et al., 2012; de Reus and van den Heuvel, 2013; Girard et al., 2014). Any noise in the system can lead to noisy connection matrices, particularly at the single-subject level, leading to numerous false positives (Thomas et al., 2014). It has recently been estimated that false positives are twice as detrimental as false negatives for any network metric derived from binary networks (Zalesky et al., 2016).

Two recent studies have attempted to solve this issue, which is a main obstacle for the application of graph theory to structural brain networks. Drakesmith et al. (2015b), proposed the multi-threshold permutation correction to overcome the effects of false positives and threshold bias. Roberts et al. proposed a consistent thresholding of structural brain networks that attempted to identify highly consistent and highly inconsistent subnetworks across subjects in a targeted cohort (Roberts et al., 2016).

One solution to this bias in structural brain connectivity metrics is to aggregate data over large samples of subjects as a way of increasing the signal to noise ratio, for example, through averaging of brain networks across subjects (Hagmann et al., 2008; Perry et al., 2015). An alternative to this group-averaging approach is to construct a consensus brain network by pooling edges that are derived from a predefined fraction of subjects across the whole cohort (van den Heuvel and Sporns, 2011; de Reus and van den Heuvel, 2013). Consensus brain network is a term derived from consensus clustering where different clusterings that have been obtained from the same dataset, after applying different clustering algorithms, are aggregated to fit a more robust/consistent clustering. Similarly, a consensus brain network maintains the edges that are highly representative across the cohort as a “majority vote” rule.

These aforementioned approaches are problematic because densely seeded tractography leads to dense structural brain

networks and thus, a high level of inherent (but potentially spurious) overlap across subjects. The most common approach to tackling this issue is to adopt a “topological filtering” approach or a “threshold” in order to uncover the backbone of the network topology. Apart from reducing spurious connections, topological filtering of brain connectivity matrices plays a significant role in extracting connection topology (Bullmore and Bassett, 2011). The most common method in this setting is to “threshold” networks to some desired density by keeping only the “strongest” links (Dimitriadis et al., 2010). We recently proposed a data-driven topological filtering scheme based on orthogonal minimal spanning trees (OMST) (Dimitriadis et al., 2017b). It is extremely important that any data-driven filtering approach considers the topology of the brain network and treats both weak and strong connections equally (Gigandet et al., 2008).

Thresholding is widely used in both structural and functional brain network analysis as a step for binarizing the weighted networks (i.e., transforming them into unweighted networks (Dimitriadis et al., 2010, 2015a, 2016a,b,c,d; Rubinov and Sporns, 2010; Antonakakis et al., 2016). While such binarization procedures are recommended for separating strong from weak connections, they are not ideally suited to extracting network metrics. The relative weights on different edges are informative and can give a better characterization of the underlying structural and/or functional topology, potentially leading to better separation of groups or conditions.

Previous studies have attempted to reveal the reliability of network and node-wise network metrics for structural brain networks, using a few edge-weighting strategies. Cheng et al. (2012) assessed test-retest reliability using diffusion tensor MRI (DT-MRI) data from 44 subjects with a focus on the differences between binary and weighted networks. Buchanan et al. (2014), with repeated scans from nine subjects, explored the reliability of network metrics on a network and node-wise level using dMRI and two alternative tractography algorithms, two alternative seeding strategies, a white matter way point constraint and three alternative network weightings (Buchanan et al., 2014). Specifically, Cheng et al. (2012) explored variability of network metrics using DTI and two different weighting network strategies (WS). In the first approach, the weights were computed as the ratio between the sum of the inverse of the fiber length and the mean volume of two Regions of Interests (ROIs) (WS1),



while in the second (WS2), they eliminated the fiber length, counting only the number of fibers normalized by the sum of the voxels in both ROIs. The Intra-Class Correlation Coefficients (ICCs) for six network metrics varies from 0.54 to 0.67 for WS1 but varies from 0.3 to 0.64 for WS2. Buchanan et al. (2014) reported global within-subject differences between 3.2 and 11.9%, with ICCs between 0.62 and 0.76. The mean nodal within-subject differences were between 5.2 and 24.2%, with mean ICCs between 0.46 and 0.62. For 83.3% (70/84) of nodes, the within-subject differences were smaller than between-subject differences.

Both studies demonstrated (ICCs) for network-wise network metrics using a few edge-weighting strategies. However, they did not assess the reliability of network metrics at a node level and more importantly, did not propose a solution for further improving the reliability of the existing methodology in constructing structural brain networks.

In this study, we constructed structural brain networks from five repeat scans of five healthy volunteers by adopting nine different network weighting strategies (NWS) affecting the construction of networks. In each of the nine alternative network weighting scenarios DT-MRI-based weights (Fractional Anisotropy-FA/Radial Diffusivity-RD/Mean Diffusivity-MD), average tract length (ATL), Euclidean distance between the coordinates of the ROIs (ED), the volume of the tract (TV), the number of streamlines (NSTR) and the proportion of streamlines (PSTR) [see section Network Weighting Strategies (NWS) for the definition], we quantified the reliability of six graph-theoretic measures network-wise (characteristic path length, global/local efficiency, radius, diameter, and eccentricity) and two node-wise (global and local efficiency) using (ICC). Since these measures are essential prerequisites for characterizing complex networks, their reliability is crucial to the ultimate interpretation of structural brain networks. Additionally, we propose a methodology for combining the alternative network weighted brain networks into a single integrated weighted structural brain network (IWSBN). We compared the ICCs of the same network metrics both network and node-wise, derived from the IWSBN, with those derived using the nine individual NWS. We also present a data-driven thresholding scheme that can extract the backbone of structural brain networks by optimizing the information flow under the constraint of the overall cost of the selected weighted connections. This topological filtering scheme was applied to the IWSBN, and the ICCs of the network metrics were again estimated. Finally, we tested the NWS-based weighted brain networks, the proposed IWSBN and its topologically filtered version IWSBN<sup>TF</sup> in terms of the ability to match each network to the correct subject out of the whole cohort (i.e., identify which networks are derived from repeat scans of the same subject, which we refer to as “recognition accuracy”). This is important, as it captures the ability to separate intra-individual differences in derived networks (where variance derives from measurement noise), from interindividual differences in networks (reflecting true underlying biological differences). As such, this facilitates the study of individualized structural brain networks without having to resort to group-averaging approaches.

## MATERIALS

### Participants

In total, five healthy subjects participated in this pilot study (mean  $37.1 \pm 4.9$  years std age, five males). The whole procedure involved five repeat scans for each participant 1 week apart from each other. All participants were recruited through the School of Psychology, Cardiff, Wales, UK. All participants were undergoing or had previously completed a university degree course, were right handed as assessed with the Edinburgh Handedness Inventory<sup>3</sup> and of Caucasian origin. Exclusion criteria included a current episode or a history of neurological and psychiatric disorders, drug or alcohol abuse and medication that may have an impact on the structure of the brain. For assessment, the general health questionnaire was used (Goldberg and Huxley, 1980). All subjects provided a written informed consent.

### Structural MRI Scanning

T<sub>1</sub>-weighted structural scans were acquired using an oblique axial, 3D fast-spoiled gradient recalled sequence (FSPGR) with the following parameters: TR = 7.9 ms, TE = 3.0 ms, inversion time = 450 ms, flip angle = 20°, 1 mm isotropic resolution, with a total acquisition time of ~7 min.

### Diffusion MRI Scanning

High angular resolution diffusion-weighted imaging (HARDI) data were acquired in the Cardiff University Brain Research Imaging Centre (CUBRIC) on a 3 T GE Signa HDx system (General Electric, Milwaukee, USA) using a cardiac-gated, peripherally gated twice-refocused spin-echo Echo Planar Imaging (EPI) sequence, with effective TR/TE of 15R-R intervals/87 ms. Sets of 60 contiguous 2.4 mm thick axial slices were obtained, with diffusion-sensitizing gradients applied along 30 isotropically distributed (Jones et al., 1999) gradient directions ( $b = 1,200$  s/mm<sup>2</sup>). For further details of the MRI protocol see (Bracht et al., 2016).

### Diffusion MRI Data Preprocessing

Data were analyzed using Explore DTI 4.8.3 (Leemans et al., 2009). Eddy-current induced distortion and motion correction was performed using an affine registration to the non-diffusion-weighted B<sub>0</sub>-images, with appropriate re-orienting of the encoding vectors (Leemans and Jones, 2009). Field inhomogeneities were corrected for using the approach of Wu et al. (2008). The diffusion-weighted images (DWIs) were non-linearly warped to the T<sub>1</sub>-weighted image using the FA map, calculated from the DWIs, as a reference. Warps were computed using Elastix (Klein et al., 2010) normalized mutual information as the cost function and constraining deformations to the phase-encoding direction. The corrected DWIs were therefore transformed to the same (undistorted) space as the T<sub>1</sub>-weighted structural images. A single diffusion tensor model was fitted to the diffusion data in order to compute quantitative parameters such as FA (Basser et al., 1994). Following the method of Pasternak et al. (2009), a correction for free water contamination of the diffusion tensor based estimates was applied (Pasternak et al., 2009; Metzler-Baddeley et al., 2012). Data quality was

checked by careful visual inspection and by looking at the average residuals per DWI for each participant.

## Tractography

DT-MRI analysis was performed using ExploreDTI (Leemans et al., 2009) following peaks in the fiber orientation density function (fODF) reconstructed from the damped Richardson Lucy algorithm (dRL) (Dell'acqua et al., 2010; Jeurissen et al., 2013). The dRL algorithm estimates multiple fiber orientations in a single voxel and therefore provides a more accurate diffusion profile than DT-MRI-based methods estimating only one fiber orientation per voxel. For each voxel in the dataset, streamlines were initiated along any peak in the (fODF) that exceeded an amplitude of 0.05. A streamline, uniform step-size, algorithm based on that of Basser et al. (2000), but extended to multiple fiber orientations within each voxel (Jeurissen et al., 2011), was used for tractography. Each streamline continued in 0.5 mm steps following the peak in the fODF that subtended the smallest angle to the incoming trajectory. Termination criteria were an angle threshold  $>45^\circ$  and fODF amplitude  $<0.05$ .

## Network Construction

The automated atlas labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) was registered to the HARDI data using a nonlinear transformation (Klein et al., 2010). The streamline termination points were coregistered to each AAL region. The numbers of streamlines connecting each pair of AAL regions were aggregated into a  $90 \times 90$  connectivity matrix.

Connections between regions were computed by identifying the streamlines connecting each pair of gray matter ROIs. The endpoint of a streamline was considered to be the first gray matter ROI encountered when tracking from the seed location.

Streamlines that did not connect to an ROI were discarded. Networks were computed for 13 different thresholds of streamline filtering by minimum contiguous length in white matter, from 0 to 6.0 mm in increments of 0.5 mm (Buchanan et al., 2014). For instance, a threshold of 1 mm discards any streamline that does not pass through at least 1 mm in white matter between gray matter ROIs.

## Network Weighting Strategies (NWS)

In this section, we describe the nine adopted NWS derived from tractography.

Fractional anisotropy (FA) is calculated from the eigenvalues ( $\lambda_1, \lambda_2, \lambda_3$ ) of the diffusion tensor. The eigenvectors ( $\epsilon$ ) give the orientations in which the ellipsoid has major axes and the corresponding eigenvalues give the magnitude of the peak along each axis (Basser and Pierpaoli, 1996). The mean diffusivity (MD) is the average of the three eigenvalues, while the axial and radial diffusivity are given by the largest and average of the two smallest eigenvalues, respectively (Basser et al., 1994).

The fourth NWS was based on average streamline tract length (ATL) leading to ATL-weighted networks. The fifth NWS estimated the Euclidean distance (ED) between the centroids of the two ROIs leading to the ED-weighted network. The Euclidean distance is computed in native space, so will vary across individuals.

The sixth NWS, termed streamline density (SD-weighted), records the interconnecting streamline density corrected for ROI size:

$$w_{ij} = \frac{2}{g_i + g_j} |S_{ij}| \quad (1)$$

where  $S_{ij}$  is the set of all streamlines found between node  $i$  and node  $j$  (and  $S_{ij} = S_{ji}$ ), and  $g_i$  and  $g_j$  are the number of gray matter voxels in nodes  $i$  and  $j$ . This approach leads to the construction of a SD-weighted network.

The seventh NWS is based on the volume of the tract (TV) leading to TV-weighted networks. The tract volume is computed by counting the number of voxels the streamlines of a bundle occupy and multiplying by the voxel size.

Two further NWSs were based on the number and the percentage of streamlines that connected a pair of ROIs. The number of streamlines (NSTR) is the absolute NSTR connecting two regions. The proportion of streamlines (PSTR) is the NSTR between each pair of regions, normalized to the total NSTR across the whole brain.

The adopted NWSs are called the NSTR and PSTR.

Figure 1 illustrates the nine alternative NWS and the corresponding weights from a scan of the first subject.

## Integrating NWS into a Single Graph

We integrated the different NWS via a linear integration based on the best matching of each pair of NWS-based brain networks in terms of their maximum information flow using the graph diffusion distance metric (gDDM) as described in the next section.

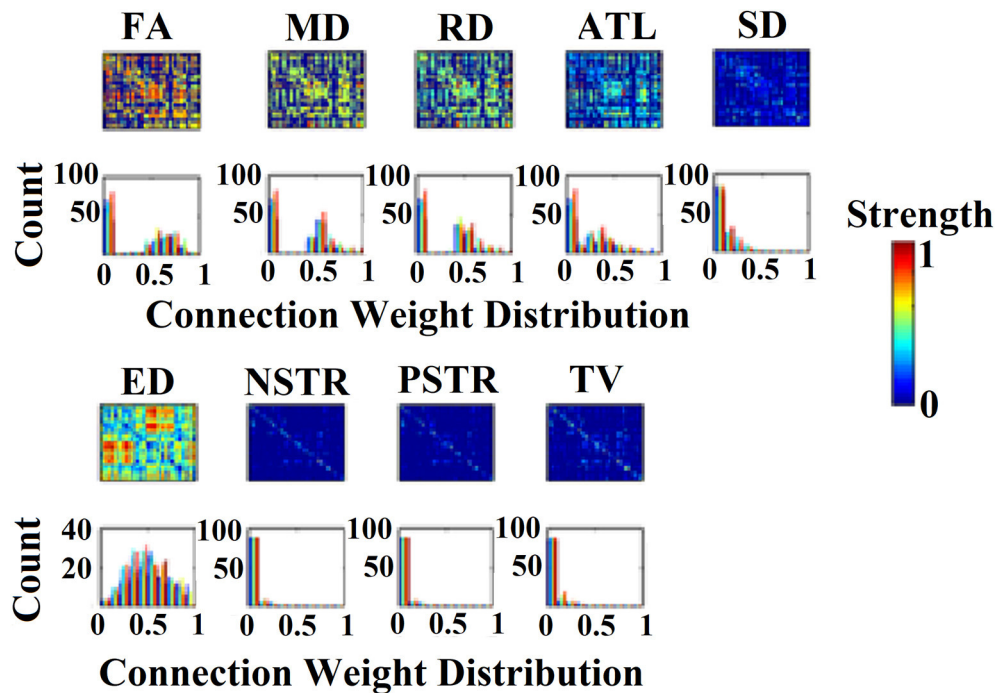
### Graph Diffusion Distance Metric

We computed the dissimilarity distance between every pair of structural brain networks (SBNs) with a novel gDDM Graph Diffusion Distance (GDD), based on a graph Laplacian exponential kernel (Fouss et al., 2012), served as a distance metric.

The graph Laplacian operator of the SBN was defined as  $L = D - SBN$ , where  $D$  is a diagonal degree matrix estimated from the SBN. This method entails modeling hypothetical patterns of information flow among sensors based on each observed (static) SBN. The GDD metric reflects the result of the comparison of such patterns between groups. The diffusion process on the person-specific SBN was allowed for a set time  $t$ ; the quantity that underwent diffusion at each time point is represented by the time-varying vector  $u(t) \in \mathbb{R}^N$ . Thus, for a pair of sensors  $i$  and  $j$ , the quantity  $SBN_{ij}(u_i(t) - u_j(t))$  represents the hypothetical flow of information from  $i$  to  $j$  via the edges that connect them (both directly and indirectly). Summing all these hypothetical interactions for each sensor leads to  $u_j'(t) = \sum_i FCG_{ij}(u_i(t) - u_j(t))$ , which can be written as:

$$u^i(t) = -Lu(t) \quad (2)$$

where  $L$  is the graph Laplacian of SBN. At time  $t = 0$ , Equation (2) has the analytic solution:  $u(t) = \exp(-tL)u^{(0)}$ . Here  $\exp(-tL)$  is a



**FIGURE 1 | (Top)** 90 × 90 connectivity matrices of inter-region connections adopted from a subject for the nine network weighting strategies (NWS). **(Bottom)** The corresponding histograms of the connection weights for each of the nine NWS.

$N \times N$  matrix function of  $t$ , known as a Laplacian exponential diffusion kernel (Fouss et al., 2012), and  $u^{(0)} = e_j$ , where  $e_j \in \mathbb{R}^N$  is the unit vector with all zeros except in the  $j$ th component. Running the diffusion process through time  $t$  produced the diffusion pattern  $\exp(-tL)e_j$  which corresponds to the  $j$ th column of  $\exp(-tL)$ .

Next, a metric of dissimilarity between every possible pair of person-specific diffusion kernelised SBNs ( $SBN_1$ ,  $SBN_2$ ) in the form of the graph diffusion distance  $d_{gdd}(t)$  was computed. The higher the value of  $d_{gdd}(t)$  between the two graphs, the greater the difference in their network topology as well as the corresponding hypothetical information flow. The columns of the Laplacian exponential kernels,  $\exp(-tL_1)$  and  $\exp(-tL_2)$ , describe distinct diffusion patterns, centered at two corresponding sensors within each SBN. The  $d_{gdd}(t)$  function is searching for a diffusion time  $t$  that maximizes the Frobenius norm of the sum of squared differences between these patterns, summed over all sensors, and is computed as:

$$d_{gdd}(t) = \|\exp(-tL_1) - \exp(-tL_2)\|_F^2 \quad (3)$$

where  $\|\cdot\|_F$  is the Frobenius norm.

Given the spectral decomposition  $L = V\Lambda V$ , the Laplacian exponential can be estimated via:

$$\exp(-tL) = V\exp(-t\Lambda)V' \quad (4)$$

where for  $\Lambda$ ,  $\exp(-t\Lambda)$  is diagonal to the  $i$ th entry given by  $e^{-t\Lambda_{i,i}}$ . We computed  $d_{gdd}(SBN_1, SBN_2)$  by first diagonalizing  $L_1$  and

$L_2$  and then applying Equations (3) and (4) to estimate  $d_{gdd}(t)$  for each time point  $t$  of the diffusion process. In this manner, a single dissimilarity value was computed for each pair of SBNs (Hammond et al., 2013).

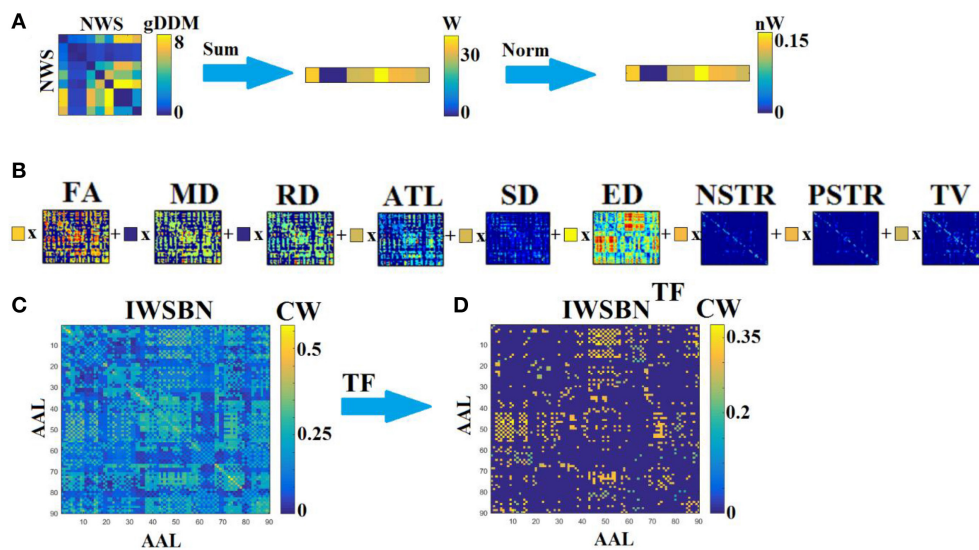
### Linear Integration of the Different NWS-Based SBN into IWSBN

Specifically, adopting a gDDM (see previous section Graph Diffusion Distance Metric), we estimated a dissimilarity matrix  $d_{gDDM}$  between every pair of NWS-based brain networks independently for each scan and subject (Figure 2A). Afterward, we estimated the sum of the rows of  $d_{gDDM}$  and then we normalized this derived vector (such as to have a total sum equal to one), to extract weights,  $l_w$ , for the linear integration of the NWS-based networks into a single graph. Then, we summed across all of these networks weighting each network by  $l_w$  (Figure 2B). The result is an IWSBN that is fully-weighted (Figure 2C). Figure 3 illustrates the topologies of the nine NWS from a single subject from their first scan. We plotted the upper decile 10% of the strongest connections according to the related weight.

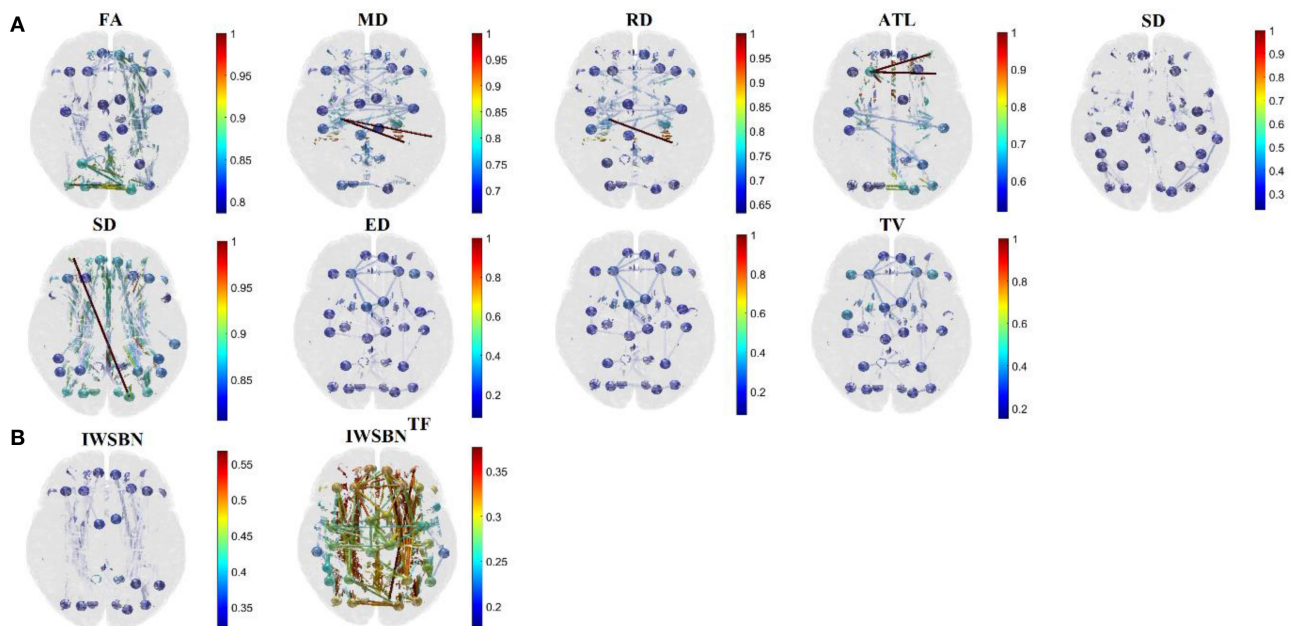
### Topological Filtering of Structural Brain Network

We topologically filtered the IWSBN using a data-driven thresholding scheme that optimizes the information flow over the cost of the surviving/selecting connections. Below, we describe the proposed data-driven topological filtering scheme.





**FIGURE 2 |** Integrated different network weighting strategies into a single weighted structural brain network (IWSBN). **(A)** From the dissimilarity matrix between every pair of NWS-networks to the related linear weights linked to their interrelationship. We first summed the rows from the dissimilarity matrix  $d^{gDDM}$  and then we normalized these weights  $l_W$  such as to have a total sum equal to one. **(B)** Linear integration of the NWS-networks by multiplying (x) each NWS-based SBN with the related weight  $l_W$  derived from (A). **(C)** The suggested IWSBN derived from (C). **(D)** The topological filtered version of IWSBN called IWSBN<sup>TF</sup>.



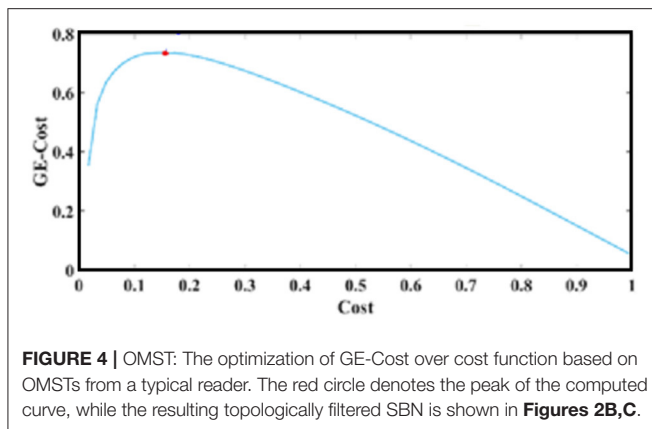
**FIGURE 3 |** Subject 1—First scan: Topographical layouts of **(A)** nine brain networks derived from the related NWS and the **(B)** IWSBN and its topological filtering version IWSBN<sup>TF</sup>. We plotted the 10% of the strongest connections at each structural brain network to enhance the visualization of the network topology.

### Topological Filtering Based on Orthogonal Minimal Spanning Trees (OMST)

In graph theory, a tree is defined as an acyclic connected graph (Estrada, 2011). Acyclic implies that there are no loops (of any length) in the graph. Minimal Spanning Tree (MST) has been shown to be an unbiased, assumption-free method to derive

unique functional brain networks (Meier et al., 2015). However, MST is a tree with only  $V-1$  links, which for large graphs is too sparse to allow reliable discrimination between two (Antonakakis et al., 2016; Dimitriadis et al., 2017a,b) or more groups (Khazaee et al., 2017). Two main algorithms have been described to construct the MST of a weighted graph by Kruskal (1956) and





Prim (1957). In a recent study, we demonstrated a data-driven topological filtering scheme for brain networks using a large number of EEG and fMRI functional connectivity graphs. Our algorithm samples connections from a fully-weighted graph via OMST (Dimitriadis et al., 2017a,b). The objective criterion was the optimization of the Global Cost Efficiency (GCE) = GE-Cost over each round of the OMST. Cost denotes the ratio of the total weight of the selected edges, over multiple iterations of OMST, divided by the total strength of the original fully-weighted graph. The values of GCE range within the limits of an economical small-world network for healthy control participants (Bassett and Bullmore, 2009). The quality formula is described by the following equation:

$$J_{GCE}^{OMSTs} = GE - Cost \quad (5)$$

The curve in **Figure 4** plots Equation (5) over cost after running exhaustive OMSTs until all observed weights were tested, based on data from a typical reader. The maximum of this (always) positive curve reflects the optimization of the proposed OMST algorithm. In the example of **Figure 4**, we applied the algorithm in the IWSBN in **Figure 2C** and the GE-Cost vs. cost function was optimized after four OMSTs leading to a selection of  $4 \times 89 = 356$  connections—a mere 8.9% of the total number of connections that survived the topological filtering approach.

The outcome of this procedure is the IWSBN<sup>TF</sup> presented in **Figures 2D, 3B** which is sparser compared to the IWSBN in **Figures 2C, 3B**. The topological filtering scheme revealed a dense subgraph between frontal areas and calcarine, cuneus, lingual, occipital, and fusiform bilaterally.

## Network Measures

For each of the nine weighted SBNs, we estimated six network metrics at the network level and two at the node level. Specifically, for the network level, we estimated global efficiency, local efficiency, characteristic path length, radius, eccentricity and mean weight. For the node level, we estimated global and local efficiency.

## Test-Retest Statistics

For each metric, an agreement between sessions was computed, via ICC (Shrout and Fleiss, 1979). ICC values were extracted for both network and node level and for every NWS-based brain network, for the IWSBN and also its topological filtering version IWSBN<sup>TF</sup>.

High test-retest reliability is a prerequisite for a connectomic metric to allow for the distinguishing of different individuals (Zuo and Xing, 2014) and also for developing a biomarker of the application of functional connectomics, such as mapping growth charts of human brain function (Dosenbach et al., 2010; Castellanos et al., 2013). Therefore, beyond developing a biomarker, estimations of the test-retest reliability of functional connectomics are valuable for providing a reference regarding how strongly the estimated variables affect the observed results and guiding the significant value of the findings of both normal and abnormal brains (Zuo et al., 2014).

## Classification of Structural Brain Networks

Recognition accuracy was assessed for each individual scan compared to the rest based on a k-nearest neighbor (k-NN) classifier with  $k = 4$  and adopting a leave-one-out cross-validation scheme (LOOCV). Instead of the Euclidean Distance (ED) most commonly used in k-NN classifiers, here we used the proposed gDDM (see section Graph Diffusion Distance Metric). gDDM is a more appropriate metric compared to ED to quantify the distance between two SBNs regarding their distance in terms of information flow based on their topology. The proposed gDDM metric is based on the eigenanalysis of the Laplacian matrices with known attributes in terms of graph theory and diffusion processes (Fouss et al., 2012).

## Discrimination of Healthy Controls from Individuals with Psychotic Experiences via Structural Connectome

To demonstrate the effectiveness of the proposed method in a binary classification problem, we analyzed a large dataset consisting of 123 individuals with psychotic experience and 125 age and gender-matched controls. The details of the cohort and the MRI scanning protocol can be found in the original publication (Drakesmith et al., 2015a).

We followed a binary classification procedure with a 10-fold cross-validation, employing as an input, each weighted SBN separately but also the IWSBN and the topologically filtered version (IWSBN<sup>TF</sup>). As a classifier, we used a tensor subspace analysis to reduce the initial high-dimensionality of the original functional connectivity network to a space of condensed descriptive power (Dimitriadis et al., 2015b,c; Antonakakis et al., 2016). The input on TSA is a 3D tensor-matrix of dimensions (subjects  $\times$  ROIs  $\times$  ROIs). As a classifier, we used a support vector machine with RBF kernel.

## Exploring the Effect of Each Node to the Integrated Graph

The proposed IWSBN<sup>TF</sup> was derived after first linearly combining the nine NWS and after that, topologically filtering

the outcome IWSBN. Our second thought was to weight each node independently within each of the nine NWS first and secondly to weight the whole NWS-network with the proposed methodology. To get the linear weights  $l_w$  for each node within each NWS-network, we followed five strategic network lesion schemes, three based node-wise and two cluster-wise.

The first three node-wise strategies were: (1) zeroing half of the connections of each node; (2) diminishing the weights of the connections of each node by 50%; and (3) combining both where half of the connections of each node were zeroed while the weights of the second half were diminished by 50%. The three node-wise attack strategies were followed for one by one node at every NWS and then we estimated the gDDM distance between the original network and the attacked network. Finally, the derived vector with the 90 gDDM values was normalized such as its sum was equal to one. Then, we multiplied each NWS-network node-wise with this vector and afterward with the network-wise approaches based on the present methodology.

The two cluster-wise lesions were: (1) the distinction of the whole set of hubs into rich-club hubs and non-rich-club hubs (van den Heuvel and Sporns, 2011); and (2) the functional clustering of the NWS-network into distinct clusters using the modularity algorithm (Newman, 2006). The two cluster-wise attack strategies were followed for each cluster at every NWS, based on the three node-wise attack strategies targeting either connections within rc-hub subgraphs and/or non-rc-hub subgraph connections and/or the connections between the non-rc and rc-hubs. Then, we estimated the gDDM distance between the original network and the attacked network. Finally, the derived vector with  $n_{\text{cluster}}$  gDDM values was normalized such as its sum was equal to one.

The whole procedure was added as a first step before the proposed network-wise linear combination of the NWS-network into a single IWSBN. All the NWS-networks were pre-filtered with the proposed data-drive thresholding scheme. The node-wise linear weighting step prior to the proposed network-wise was evaluated based on the ICC values of the adopted network metrics both network and node-wise. Additionally, the recognition accuracy of each subject scan over the rest of cohort was compared to the proposed method.

## RESULTS

### Graph Embedding of the Dissimilarity Matrices Based on $g\text{DDM}$

Dissimilarity matrices (DM) based on each NWS across scans and subjects were estimated based on the gDDM. **Figure 5**

scaling (MDS). The NSTR proved to better discriminate the five subjects compared to the rest of the methods.

### Reliability of Network Level Metrics for NWS and IWSBN

The ICC scores were excellent—ranging from 0.75 to 1—for six out of nine network metrics for the entire set of network metrics. These NWS include the ATL, BIN, SD, ED, NSTR, PSTr, and TV (**Figure 6**). The related group-averaged values of the network metrics for each NWS are shown in **Figure 7**. The proposed IWSBN yielded good ICC values but these were lower than those obtained for each of the six NWS (**Figure 8A**). Significantly, we observed an improvement of the ICC on the IWSBN<sup>TF</sup> which reached the level of the six best NWS in terms of ICC scoring (**Figure 8B**). **Figure 9** demonstrates the group-averaged values of network metrics on the network level.

### Reliability of Network Metrics on a Node Level

The analysis of ICCs on global and local efficiency node-wise on the nine NWS and in both IWSBN and IWSBN<sup>TF</sup> revealed important trends. Firstly, the ICC values for each of the NWS failed to reach a fair value ( $\text{ICC} < 0.1$ ). Secondly, the ICC values derived from the IWSBN showed a large variability but reached on average  $\text{ICC} = 0.68 \pm 0.10$  for global efficiency and  $\text{ICC} = 0.68 \pm 0.17$  for local efficiency (**Figure 10A**). Third, the ICC values for both network metrics were improved in IWSBN<sup>TF</sup>, reaching a mean  $\text{ICC} = 0.75 \pm 0.02$  for global efficiency and a mean  $\text{ICC} = 0.84 \pm 0.02$  for local efficiency (**Figure 10B**). Applying a Wilcoxon Rank Sum Test between the two distributions for each network metrics, we observed significant improvement of ICC values for IWSBN<sup>TF</sup> (global efficiency:  $p = 0.0035 \times 10^{-7}$ , local efficiency:  $p = 0.0067 \times 10^{-10}$ ). **Figure 11** demonstrates the group-averaged values of network metrics on the node level.

### Recognition Accuracy of Structural Brain Networks

Dissimilarity matrices (DM) based on both IWSBN and IWSBN<sup>TF</sup> across scans and subjects were estimated based on the gDDM. **Figure 12** demonstrates the DM for both IWSBN and IWSBN<sup>TF</sup> across repeat scans and the related graph embedding based on multidimensional scaling (MDS). The proposed topological filtering scheme improved the discrimination of the five subjects compared to the original IWSBN.

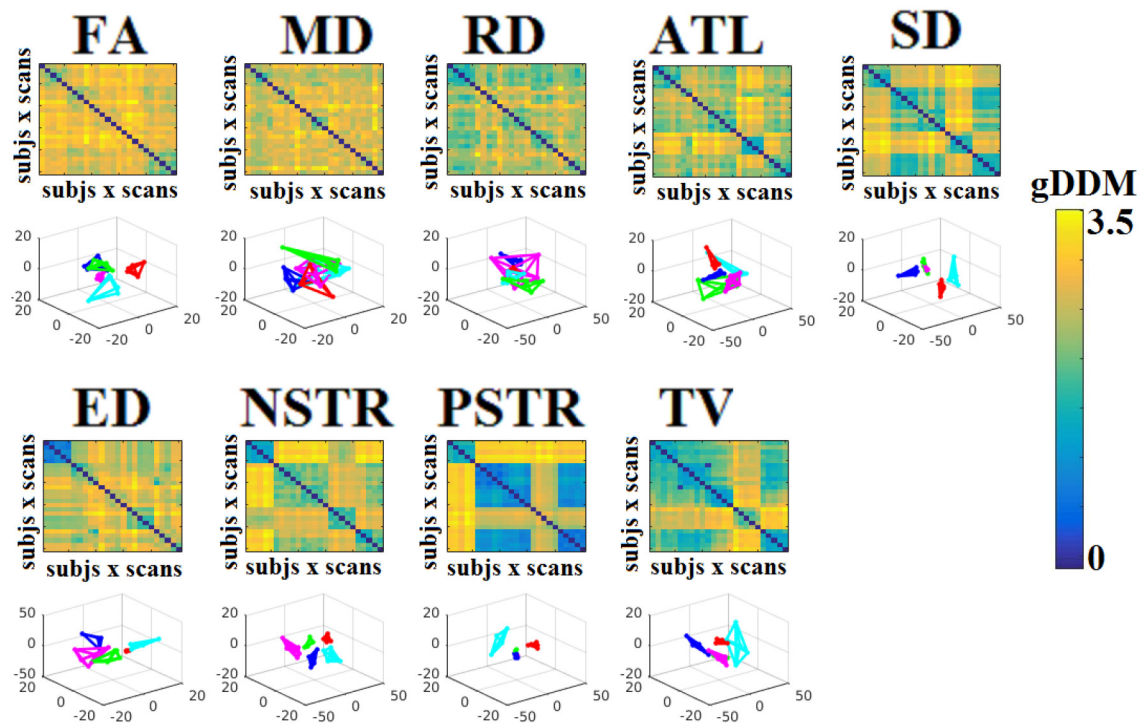
Applying a k-NN classifier with  $k = 4$  and gDDM as the appropriate distance metric under a LOOCV scheme, we succeeded to accurately classify each scan to the right person

$$QCI = \frac{\text{scans} \times (\text{scans} - 1) / 2}{\text{scans} \times \text{scans}} \times \frac{\sum_{su_1=1}^{\text{subjects}} \sum_{su_2=su_1+1}^{\text{subjects}} \sum_{l=1}^{\text{scans}} \sum_{m=1}^{\text{scans}} gDDM(D_{IWSBN^{TF}}(su_1,l), D_{IWSBN^{TF}}(su_2,m))}{\text{subjects} * (\text{subjects} - 1) / 2} \quad (6)$$

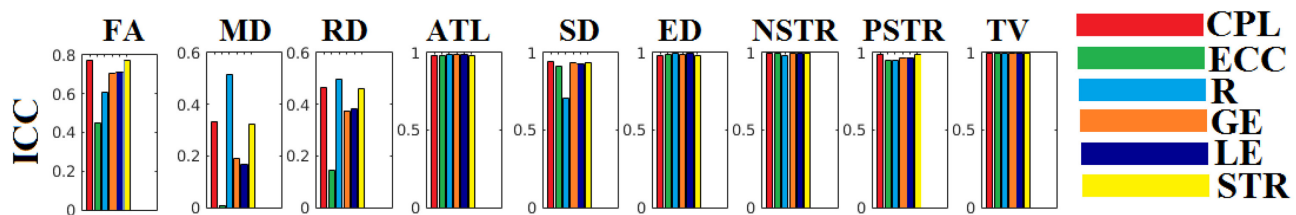
$$\frac{\sum_{su=1}^{\text{subjects}} \sum_{l=1}^{\text{scans}} \sum_{m=l+1}^{\text{scans}} gDDM(D_{IWSBN^{TF}}(su,l), D_{IWSBN^{TF}}(su,m))}{\text{subjects}}$$

demonstrates the DM for each of the NWS across repeat scans and the related graph embedding based on multidimensional

based on IWSBN<sup>TF</sup>. Similar results were also obtained with NSTR (**Figure 5**).



**FIGURE 5 |** Dissimilarity matrices (DM) based on each NWS across scans and subjects based on the gDDM and graph embedding of DM. First and third rows illustrate the DM between every pair of scans across the cohort based on the gDDM metric for each of the nine NWS, while the second and fourth rows demonstrate the embedded DM via the MDS process in a common 3D space. Each color corresponds to a single subject, while lines with the same color interconnect the NWS derived from repeat scans from the same subject (MDS, multidimensional scaling).



**FIGURE 6 |** ICC values for basic network metrics on the network level for each of the nine NWS. CPL, Characteristic Path Length; ECC, Eccentricity; R, Radius; GE, Global Efficiency; LE, Local Efficiency; STR, mean strength.

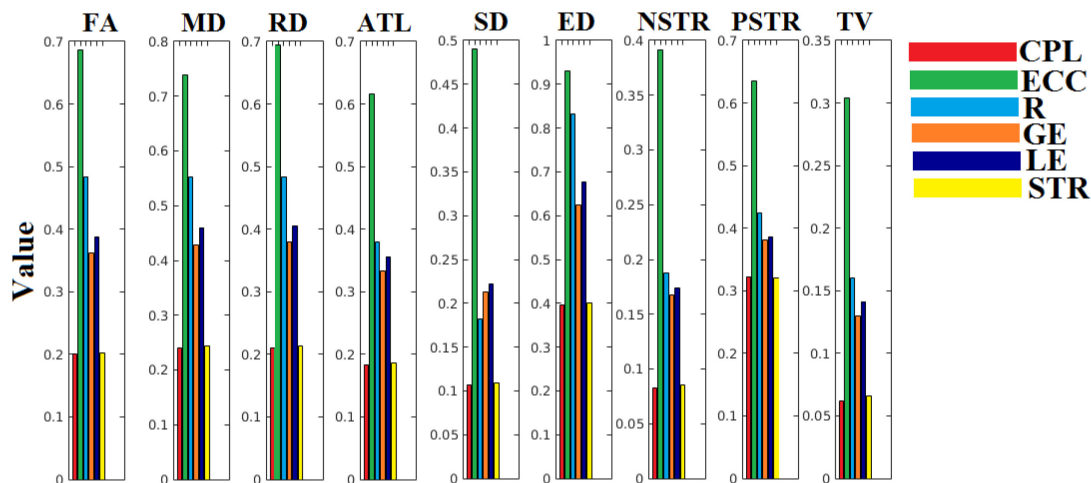
For a better estimation of the discrimination of proposed IWSBN<sup>TF</sup> with NSTR, we defined the following quality of clustering index (QCI) (Dimitriadis et al., 2012):

The QCI quantifies the average similarity of the IWSBN<sup>TF</sup> across scans within each subject (cluster) expressed in the denominator of Equation (6) and the average dissimilarity between every pair of subjects (clusters) across their scans expressed in the numerator. Both similarity and dissimilarity were estimated based on the gDDM. The higher the dissimilarity between the clusters (numerator) and/or the lower the dissimilarity within the clusters (denominator), the higher the QCI. The numerator is averaged across all possible combinations of subjects (clusters) while the denominator across subjects (clusters). The first term was used to equalize the effect

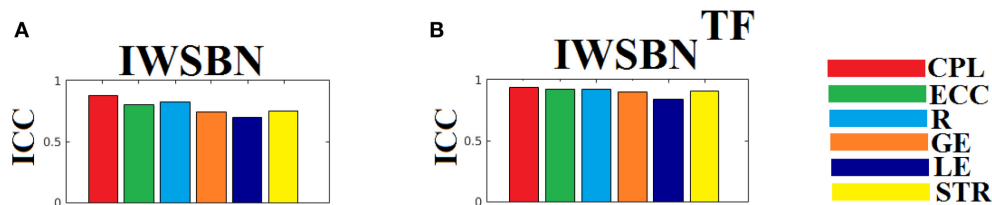
of between-subject (clusters) comparisons vs. within-subject comparisons (clusters). This inversed coefficient guarantees that in the case of all the weights in the DM being equal then it will take a value of one. Therefore, the higher the value of the QCI above one, the higher the separability between the network topologies of the subjects.

We first tabulated all the IWSBN<sup>TF</sup> across scans and subjects into a 4D graph with dimensions equal to [subjects  $\times$  scans  $\times$  Rois  $\times$  Rois] called  $D\_IWSBN^{TF}$ . Afterward, we estimated the QCI for each NWS and for both IWSBN and IWSBN<sup>TF</sup>.

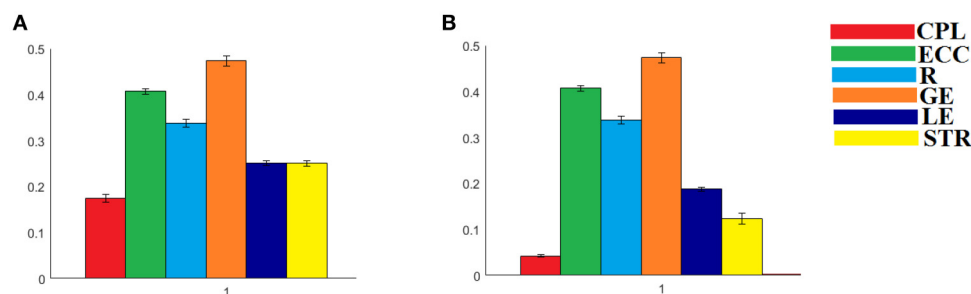
The QCI was 1.45 for IWSBN and 6.45 for IWSBN<sup>TF</sup> while for NSTR the QCI was 4.57. This result can be interpreted as a higher separation of network topologies with our approach compared to the best of NWS.



**FIGURE 7** | Group-averaged values of the adopted network metrics at each NWS. CPL, Characteristic Path Length; ECC, Eccentricity; R, Radius; GE, Global Efficiency; LE, Local Efficiency; STR, mean strength.



**FIGURE 8** | ICC values for basic network metrics on the network level for both (A) IWSBN and (B) IWSBN<sup>TF</sup>. CPL, Characteristic Path Length; ECC, Eccentricity; R, Radius; GE, Global Efficiency; LE, Local Efficiency; STR, mean strength.



**FIGURE 9** | Group-averaged values of the adopted network metrics for both (A) IWSBN and (B) IWSBN<sup>TF</sup>. CPL, Characteristic Path Length; ECC, Eccentricity; R, Radius; GE, Global Efficiency; LE, Local Efficiency; STR, mean strength.

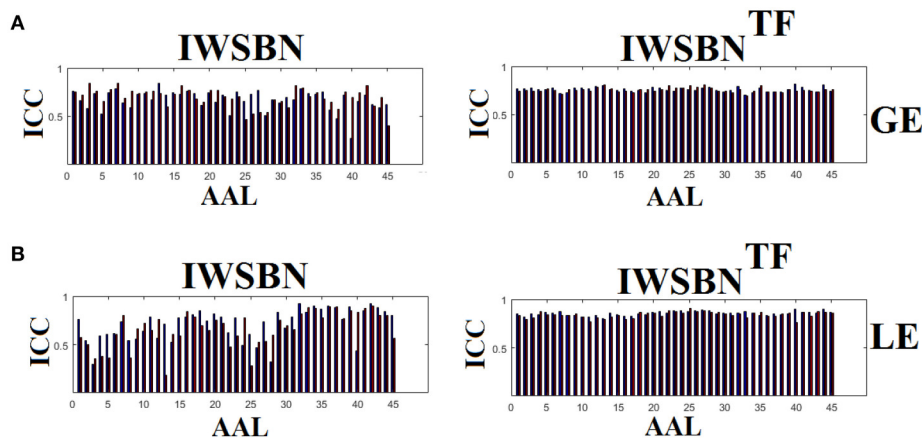
To further enhance the integration of the nine alternative WNS for the construction of an integrated SBN, we repeated the whole procedure splitting the nine weighted versions of SBN into three triads (the first three, the second three and the last three). **Figures 13–15** illustrate the DM and their embedding into a 3D-space. The highest separability between the network topologies of the subjects have been demonstrated for ATL, SD, and ED, while the worst for NSTR, PSTR, and TV, where three subjects overlapped on the same embedding space (**Figure 15B**). The QCI score was lower compared to the

original approach where we combined the nine weighted SBN (**Figure 12**).

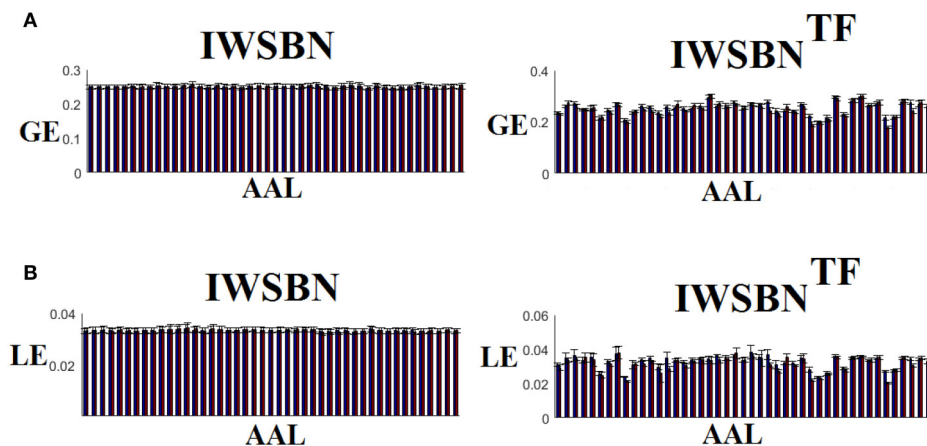
### Structural Connectomic Classification of Healthy Controls (HC) from Individuals with Psychotic Experiences (PE)

Our classification results demonstrated a higher classification accuracy (65.3%) between the two groups for the proposed IWSBN<sup>TF</sup>. The classification performance of the nine





**FIGURE 10 |** ICC values for global and local efficiency on the node level for both IWSBN and IWSBN<sup>TF</sup>. **(A)** ICC node-wise score for global efficiency on **IWSBN and IWSBN<sup>TF</sup>**. **(B)** ICC node-wise score for local efficiency on **IWSBN and IWSBN<sup>TF</sup>**. Blue/red bars refer to Automated Anatomical Labeling (AAL) ROIs for left/right hemisphere correspondingly.



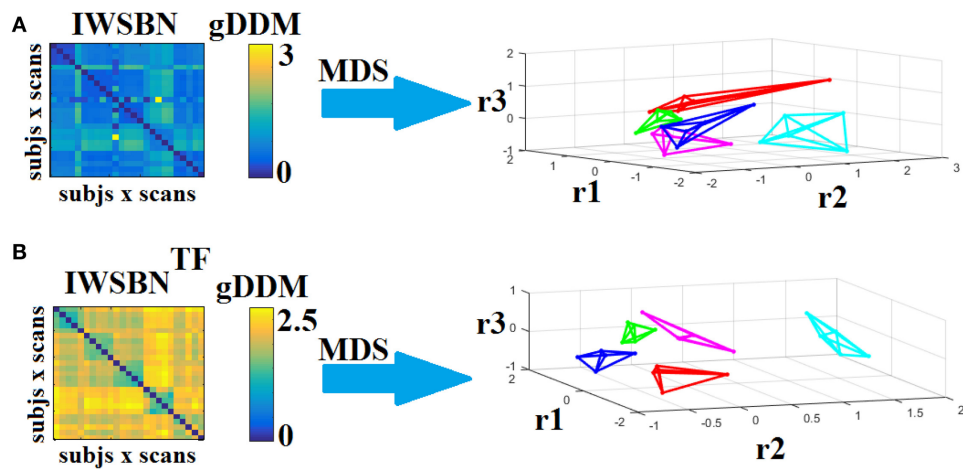
**FIGURE 11 |** Mean/std values for global and local efficiency on the node level for both IWSBN and IWSBN<sup>TF</sup>. **(A)** Mean/Std node-wise GE for global efficiency on **IWSBN and IWSBN<sup>TF</sup>**. **(B)** Mean/Std node-wise LE for local efficiency on **IWSBN and IWSBN<sup>TF</sup>**. Blue/red bars refer to AAL ROIs for left/right hemisphere correspondingly.

weighted strategies was lower than by chance ( $<50\%$ ; see **Table 1**). Additionally, the data-driven topological filtering via the OMST algorithm (Dimitriadis et al., 2017a,b) further improved the classification accuracy (from 57.23 to 65.34%; see **Table 1**).

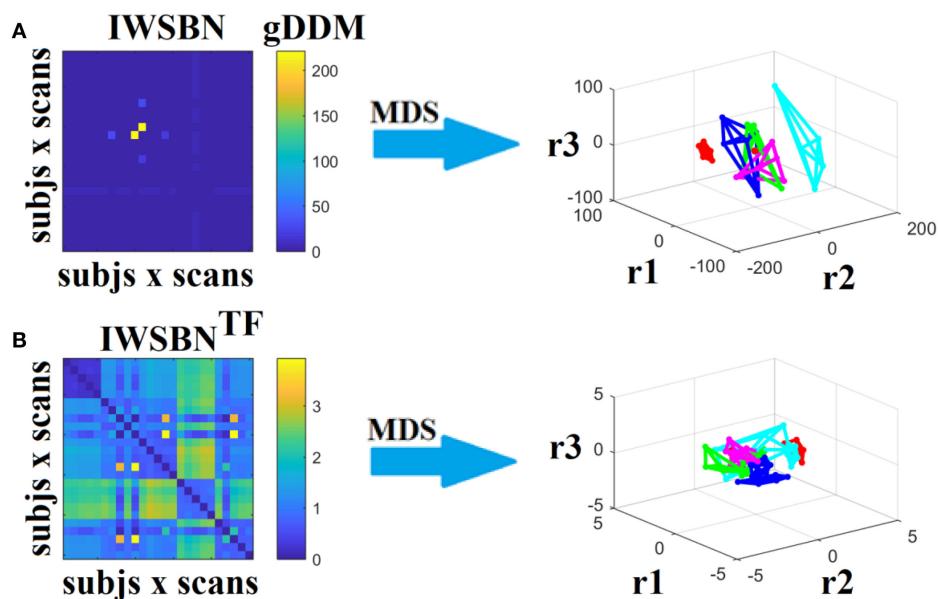
## DISCUSSION

We present, for the first time, the reliability of basic network metrics at both whole-network and node level for nine different NWS. We recruited five subjects who were scanned five times at weekly intervals. The range of age was (mean  $37.1 \pm 4.9$  years of age, five males) to minimize the effect of the age on inter-subject variability. Additionally, for the first time, we propose a completely data-driven algorithm for the linear interpolation of the different NWS-based SBNs

into a single IWSBN. The whole approach is based on a diffusion distance metric that quantifies the maximum distance between two network topologies in terms of their information flow. Complementarily, we propose a completely data-driven topological filtering scheme for extracting the backbone of a SBN on an individual level (scan-based) without attempting to find any consistency among control subjects of a specific age (Roberts et al., 2016). To reveal any gender, age or even individualized differences in terms of dMRI-based SBNs, we should adopt data-driven techniques applied to individual SBNs without any a priori knowledge of the label of a subject's scan (age, gender, HC). Any adopted group or scan consistency as a constraint to the main methodology will diminish individual differences and across scan variability, respectively. Our results can be summarized into the following key points:



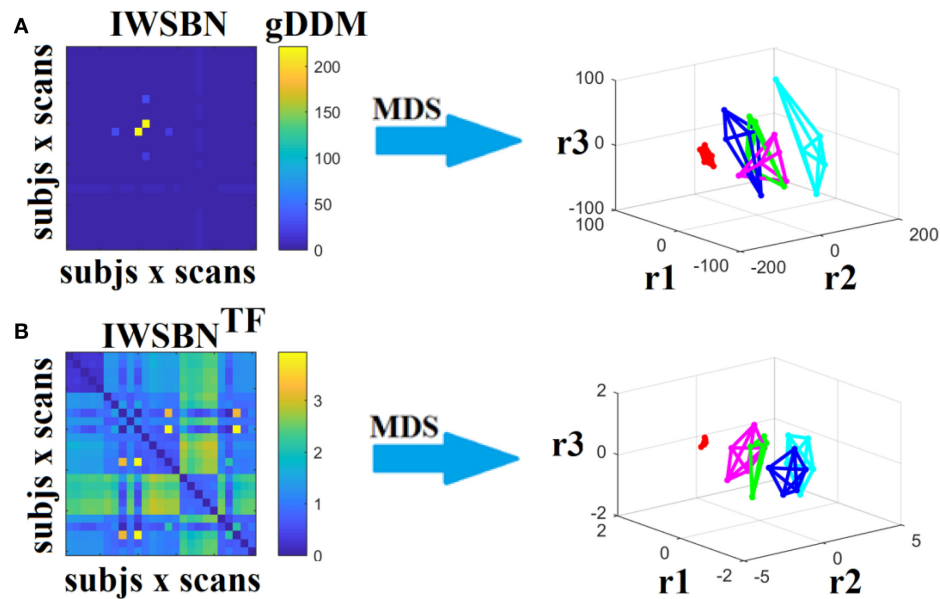
**FIGURE 12 |** Dissimilarity matrices (DM) based on both IWSBN and IWSBN<sup>TF</sup> across scans and subjects using gDDM and graph embedding of DM. **(A)** The first row illustrates the DM based on the IWSBN and its embedding in a 3D space, while the **(B)** second row demonstrates the DM based on the IWSBN<sup>TF</sup> and its embedding in a 3D space. Each color corresponds to a single subject, while lines with the same color interconnect the NWS derived from repeat scans (MDS, multidimensional scaling).



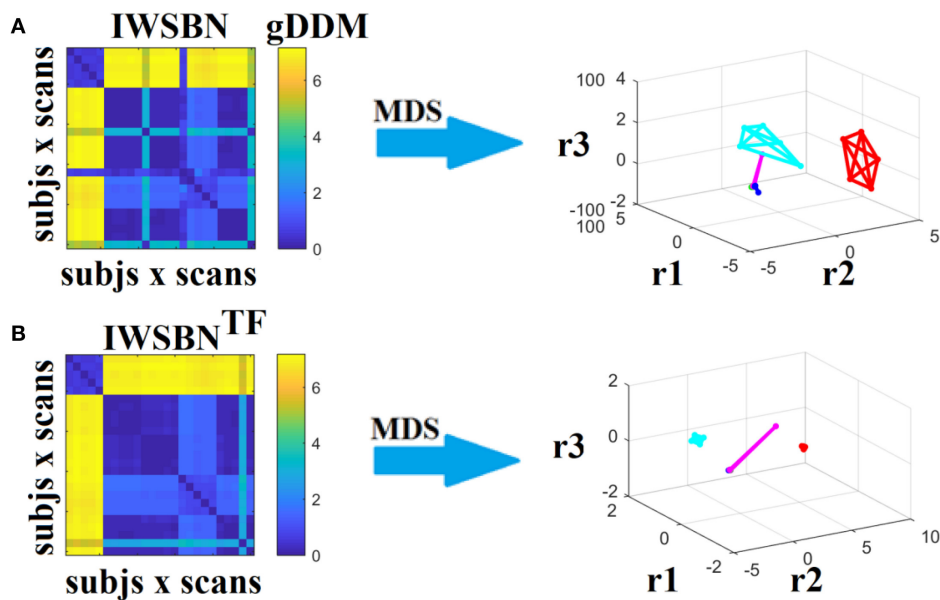
**FIGURE 13 |** Dissimilarity matrices (DM) based on both IWSBN and IWSBN<sup>TF</sup> across scans and subjects using gDDM and graph embedding of DM (as in Figure 12). Both **(A)** IWSBN and **(B)** IWSBN<sup>TF</sup> were constructed based on FA, MD, and RD weighted structural brain networks.

- ICC values of network metrics derived from network levels were high for six out of nine NWS
- ICC values of network metrics derived node-wise were unreliable for all nine NWS
- ICC values for all the network metrics on the network level for IWSBN<sup>TF</sup> were excellent and on the same level as the best NWS
- We observed high ICCs of network metrics node-wise for both IWSBN and IWSBN<sup>TF</sup> compared to each NWS with higher values succeeding based on IWSBN<sup>TF</sup>
- We succeeded in achieving a higher discrimination of each subject compared to the rest of the cohort based on the IWSBN<sup>TF</sup> derived from each scan compared to IWSBN and the best NWS which was the NSTR
- The construction of subject-specific IWSBN<sup>TF</sup> for two large populations (HC and individuals with PE) further improved the classification performance compared to each of the nine weighted versions of their structural connectome.

Previous studies explored different aspects of network reliability using repeat dMRI scans of healthy human volunteers. Hagmann



**FIGURE 14 |** Dissimilarity matrices (DM) based on both IWSBN and IWSBN<sup>TF</sup> across scans and subjects using gDDM and graph embedding of DM (as in **Figure 12**). Both **(A)** IWSBN and **(B)** IWSBN<sup>TF</sup> were constructed based on ATL, SD, and ED weighted structural brain networks.



**FIGURE 15 |** Dissimilarity matrices (DM) based on both IWSBN and IWSBN<sup>TF</sup> across scans and subjects using gDDM and graph embedding of DM (as in **Figure 12**). Both **(A)** IWSBN and **(B)** IWSBN<sup>TF</sup> were constructed based on NSTR, PSTR, and TV-weighted structural brain networks.

et al. (2008) assessed structural networks obtained from diffusion spectrum imaging (DSI), while Vaessen et al. (2010) assessed reproducibility over different sets of diffusion gradient directions using DT-MRI. Bassett et al. (2011) compared reliability in both DT-MRI and DSI, and Cammoun et al. (2012) investigated the effect of network resolution using DSI. Finally, Cheng et al. (2012) assessed test-retest reliability using DT-MRI, with a focus on the differences between binary and

weighted networks. A recent study explored the reliability of network metrics on a network and node-wise level using dMRI and two alternative tractography algorithms, two alternative seeding strategies, a white matter way point constraint and three alternative network weightings (Buchanan et al., 2014). Their best performing configuration, the global within-subject differences, showed ICCs between 0.62 and 0.76, while the mean nodal within-subject differences demonstrated ICCs between

**TABLE 1** | Accuracy, sensitivity, and specificity of the nine weighted strategies, the IWSBN and the IWSBN<sup>TF</sup> following a binary classification of HC vs. individuals with PE via a 10-fold cross-validation strategy.

LOOCV	Accuracy	Sensitivity	Specificity
FA	43.44 ± 5.43	42.12 ± 4.99	42.19 ± 5.08
MD	42.81 ± 4.82	42.72 ± 5.12	43.06 ± 4.91
RD	44.17 ± 5.67	43.65 ± 4.92	43.57 ± 5.24
ATL	45.78 ± 5.61	44.73 ± 5.32	44.82 ± 5.39
SD	45.81 ± 6.71	44.89 ± 5.21	44.71 ± 5.96
ED	46.07 ± 5.92	45.39 ± 5.66	45.87 ± 5.23
NSTR	45.91 ± 5.11	45.44 ± 5.43	45.21 ± 4.89
PSTR	46.17 ± 5.42	46.79 ± 6.13	45.94 ± 5.69
TV	46.38 ± 5.76	46.45 ± 6.32	45.82 ± 5.73
IWSBN	57.23 ± 6.89	56.79 ± 6.20	55.89 ± 5.81
IWSBN <sup>TF</sup>	<b>65.34 ± 6.71</b>	<b>64.36 ± 5.87</b>	<b>65.04 ± 6.05</b>

We underlined with bold font the best classification accuracy succeeded with the proposed IWSBN<sup>TF</sup>.

0.46 and 0.62. In the present study, we revealed higher ICC values, both node and network-wise, based on the IWSBN<sup>TF</sup>. Furthermore, applying our network analysis on IWSBN<sup>TF</sup>, we observed higher between-subject differences compared to within-subject variation (see **Figure 9B**).

Buchanan et al. (2014), concluded that regional reliability of dMRI networks is low suggesting that connections between specific pairs of nodes are unreliable across sessions. Here, we showed the same issue for each of the nine NWS leading to very small ICCs (<0.1). Applying the proposed topological filtering scheme to each of the NWS, we failed to further improve the nodal ICC, which can be interpreted as technical issues derived from tractography (data not shown). Errors in tractography in estimating axonal tracts may reflect both the segmentation of each ROI affecting the streamline construction. Tractography is strongly affected by measurement noise resulting in both false negative and positive connections (Zalesky and Fornito, 2009). Yo et al. (2009) compared different tractography algorithms focusing on the uncertainty of fiber directions in a noisy environment which could be a factor for poor ICC values for node-wise estimated network metrics.

Common poor ICC values for node-wise network metrics for each of the nine NWS and simultaneously excellent ICC values for network-wise network metrics for six out of nine NWS could be interpreted as a common error of the tractography for the former and as a denoising procedure from the latter after integrating across all nodes. It seems that the proposed dual-step scheme for combining NWS into a single IWSBN<sup>TF</sup> diminished any bias of probabilistic tractography and led to a reliable nodal ICC which was higher than that demonstrated in previous work (Buchanan et al., 2014). Both steps proved crucial to simultaneously elevating the ICC values of network metrics network-wise to excellent levels—comparable to the best NWS—and the ICC node-wise values linked to network metrics to fair to good levels (see **Figure 8**).

It is important to mention here that Buchanan et al. (2014) preferred not to threshold the derived weighted SBNs in

order to avoid biasing their results. We completely agree with this approach since, until now, none of the non-data-driven thresholding schemes can work without any bias selection of any criterion. With the present study, we proposed a solution for uncovering the backbone of a SBN by increasing the information flow within the network constrained by the overall cost of the selected connections.

The majority of studies focusing on SBNs have worked on region-to-region connectivity using an anatomical map while they ignored the rich information in the local white matter architecture (Yeh et al., 2016). In this study, the authors analyzed local connectomes, termed connectometry, which tracks the local connectivity patterns along the fiber pathways to further extract the subcomponents of the pathways that are associated with the parameters of study. They demonstrated that connectometry complements global brain networks while they are more sensitive and less affected by fiber tracking issues.

The proposed IWSBN<sup>TF</sup> was derived after first linearly combining the nine NWS and then topologically filtering the derived IWSBN. Our second thought was to independently weight each node within each of the nine NWS first and secondly to weight the whole NWS-network with the proposed methodology. The whole procedure was added as a first preliminary step before the proposed network-wise linear combination of the NWS-network into a single IWSBN. The topological filtering of each NWS-network prior to the node and cluster-wise attacking strategies did not improve the ICC values of network metrics, neither network or node-wise. Additionally, the recognition accuracy was worse compared to the proposed method. One possible interpretation of these results could be that specific connections cause a major effect on the reliability of the whole-network topology. Future strategic artificial lesion approaches on a connection level could reveal where (anatomically) and when (protocols, scanners, other factors) a tractography algorithm produces errors. This methodology could be useful to improve the algorithm between specific tracts.

Future study will shed light on how the proposed dual-step methodology can affect the reliability of connectomic biomarkers in conditions such as Alzheimer's Disease, schizophrenia with a genetic background and dyslexia. Here, we demonstrated the effectiveness of the proposed methodology in discriminating HC from individuals with PE. Additionally, we will compare IWSBN<sup>TF</sup> between different dMRI protocols and also between different scanners with the same or different field strengths (3T and 7T). Finally, large publicly-available dMRI cohorts can be analyzed with this method in order to reveal developmental trends. For all these research questions—looking at individual differences, longitudinal trajectories and case-control difference—a high degree of reliability of the underlying metrics is crucial, and thus our approach could be widely adopted. This data-driven topological filtering algorithm can be a baseline across different studies and big datasets e.g., the Human Connectome Project, UK BIOBANK in order to share metadata in a common feature space across institutes, research centers, universities and research groups.



## Motivations Derived from the Current Study

Analysis of reliability is challenging for neuroimaging as the main results presented in a study depend on both the adopted network metrics and the metrics used to characterize the weight of a connection. Additionally, many neuroimaging studies based on SBNs attempt to shed light on developmental differences, differences between clinical populations and also between a control group and a disease group. A basic reason why all these proposed connectomic biomarkers are not used in daily practice in hospitals is their reliability (Dimitriadis et al., 2015a). A second reason is that there are many studies on the same topic (e.g., brain disease) based on small datasets that adopted different NWS and arbitrary topological filtering schemes. Meaningful aggregation of these, even in the case that their metadata are free available, is impossible. A third reason is that until now, it is not standard practice to assess the reliability of network metrics derived from SBNs across repeat scans on the same population but with different dMRI protocols and scanners (3T vs. 3T or 3T vs. 7T). This is an issue that we would like to investigate in future studies with the proposed scheme.

A basic issue with the ICC is that it needs large samples in order to estimate scores to acceptable precision. A study estimated that for two repeated measures, in order to get an acceptable ICC score of 0.8 with a 95% confidence interval of 0.2 width, then at least 52 subjects are needed (see Table 3 in Shoukri et al., 2004). In a similar vein, for an ICC score of 0.6 with 95% confidence intervals of 0.2 width, repeated measures from 158 subjects would be required. Clearly for most MRI-based studies, this scenario of repeated scans for hundreds of scans is unrealistic. However, our analysis provided a methodology of how to combine different NWSs into a single integrated graph based on a gDDM that counts the network topology as a whole and quantifies the distance between two SBNs in terms of their information flow. The results of the proposed methodology presented here, even in a small dataset, are of paramount importance since they are completely data-driven in any preprocessing step. Simultaneously, they provided novel directions of how to untangle hidden information within dMRI-based brain networks by working on an individualized manner and without any averaging approach (group or scan-wise). Additionally, we proposed a data-driven thresholding scheme applied to the IWSBN that improved the ICCs of basic network metrics at both node and network levels compared to the original IWSBN and each of the adopted NWS. Our data-driven method can be seen as a methodology for improving network reliability on SBNs (Zalesky et al., 2010; Drakesmith et al., 2015b). Complementarily, our approach provided excellent discrimination of the network patterns of the five subjects based on the recognition of each scan to the targeted subject. Finally, our method better separated the five groups of scans based on the topological filtering version of IWSBN compared to the best NWS.

## Limitations of the Study

It is important to mention here the limitations of the current study due to the small dataset for exploring the test-retest reliability statistics. In the era of open science resource where

multisite worldwide neuroimaging labs share neuroimaging datasets, it is important to demonstrate novel techniques that improve the reliability of connectomics in common neuroimaging data (Zuo et al., 2014). A recent Consortium for Reliability and Reproducibility (CoRR) is working to address this gap and establish test-retest reliability as a minimum standard for methods development in functional connectomics (Zuo and Xing, 2014; Zuo et al., 2014) and morphological measurements (MacLaren et al., 2014). Reliability is important to build reliable connectomic biomarkers across multi-site (Nielsen et al., 2013; Abraham et al., 2017) and also longitudinal trajectories of structural and functional brain networks across the life-span (Zuo et al., 2017).

Our future goal is to test the proposed methodology in a larger sample for validating the test-retest reliability of our scheme and also on multi-site diffusion-based structural brain networks for building reliable connectomic biomarkers.

## CONCLUSION

Reliability analysis of both node and network-wise network metrics in IWSBN and its topological filtering version revealed: (1) similar ICC values for all the network metrics on the network level for IWSBN<sup>TF</sup> compared to the best NWS; (2) higher ICC of network metrics node-wise for both IWSBN and IWSBN<sup>TF</sup> compared to each NWS with higher values succeeding based on IWSBN<sup>TF</sup>; and (3) higher discrimination of each subject compared to the rest of the cohort based on the IWSBN<sup>TF</sup> derived from each scan compared to IWSBN and the best NWS which was the NSTR. We thus provided a new approach to identifying highly reliable and discriminative network metrics that can be the basis for studies of interindividual differences, longitudinal trajectories, and pathological changes in structural brain connectivity.

## ETHICS STATEMENT

This study was approved by the ethical committee in the Cardiff University.

## AUTHOR CONTRIBUTIONS

SD: conception of the research, methods and design, and drafting the manuscript; SD, MD, GP, SB: data analysis; MD, GP, SB, DL, DJ: critical revision of the manuscript; Every author read and approved the final version of the manuscript

## ACKNOWLEDGMENTS

SD, DL, and DJ were supported by a MRC grant MR/K004360/1 (Behavioural and Neurophysiological Effects of Schizophrenia Risk Genes: A Multi-locus, Pathway Based Approach).

SD is also supported by a MARIE-CURIE COFUND EU-UK Research Fellowship.

DJ is supported by a Wellcome Trust New Investigator Award and Wellcome Trust Strategic Award.

We would like to acknowledge Cardiff RCUK funding scheme for covering the publication fee.

## REFERENCES

- Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., et al. (2017). Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. *Neuroimage* 147, 736–745. doi: 10.1016/j.neuroimage.2016.10.045
- Antonakakis, M., Dimitriadis, S. I., Zervakis, M., Micheloyannis, S., Rezaie, R., Babajani-Feremi, A., et al. (2016). Altered cross-frequency coupling in resting-state MEG after mild traumatic brain injury. *Int. J. Psychophysiol.* 102, 1–11. doi: 10.1016/j.ijpsycho.2016.02.002
- Basser, P. J., and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* 111, 209–219. doi: 10.1006/jmrb.1996.0086
- Basser, P. J., Mattiello, J., and LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B* 103, 247–254. doi: 10.1006/jmrb.1994.1037
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). *In vivo* fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44, 625–632. doi: 10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O
- Bassett, D. S., and Bullmore, E. T. (2009). Human brain networks in health and disease. *Curr. Opin. Neurol.* 22, 340–347. doi: 10.1097/WCO.0b013e32832d93dd
- Bassett, D. S., Brown, J. A., Deshpande, V., Carlson, J. M., and Grafton, S. T. (2011). Conserved and variable architecture of human white matter connectivity. *Neuroimage* 54, 1262–1279. doi: 10.1016/j.neuroimage.2010.09.006
- Bracht, T., Jones, D., Bells, S., Walther, S., Drakesmith, M., and Linden, D. (2016). Myelination of the right parahippocampal cingulum is associated with physical activity in young healthy adults. *Brain Struct. Funct.* 221, 4537–4548. doi: 10.1007/s00429-016-1183-3
- Buchanan, C. R., Pernet, C. R., Gorgolewski, K. J., Storkey, A. J., Bastin, M. E. (2014). Test-retest reliability of structural brain networks from diffusion MRI. *Neuroimage* 86, 231–243. doi: 10.1016/j.neuroimage.2013.09.054
- Bullmore, E. T., and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annu. Rev. Clin. Psychol.* 7, 113–140. doi: 10.1146/annurev-clinpsy-040510-143934
- Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J. P., Sporns, O., Do, K. Q., et al. (2012). Mapping the human connectome at multiple scales with diffusion spectrum MRI. *J. Neurosci. Methods* 203, 386–397. doi: 10.1016/j.jneumeth.2011.09.031
- Castellanos, F. X., Di Martino, A., Craddock, R. C., Mehta, A. D., and Milham, M. P. (2013). Clinical applications of the functional connectome. *Neuroimage* 80, 527–540. doi: 10.1016/j.neuroimage.2013.04.083
- Cheng, H., Wang, Y., Sheng, J., Kronenberger, W. G., Mathews, V. P., Hummer, T. A., et al. (2012). Characteristics and variability of structural networks derived from diffusion tensor imaging. *Neuroimage* 61, 1153–1164. doi: 10.1016/j.neuroimage.2012.03.036
- de Reus, M. A., and van den Heuvel, M. P. (2013). Estimating false positives and negatives in brain networks. *Neuroimage* 70, 402–409. doi: 10.1016/j.neuroimage.2012.12.066
- Dell'acqua, F., Scifo, P., Rizzo, G., Catani, M., Simmons, A., Scotti, G., et al. (2010). A modified damped Richardson-Lucy algorithm to reduce isotropic background effects in spherical deconvolution. *Neuroimage* 49, 1446–1458. doi: 10.1016/j.neuroimage.2009.09.033
- Dimitriadis, S. I., Antonakakis, M., Simos, P., Fletcher, J. M., Papanicolaou, A. C. (2017a). Data-driven topological filtering based on orthogonal minimal spanning trees: application to multi-group MEG resting-state connectivity. *Brain Connect.* doi: 10.1089/brain.2017.0512. [Epub ahead of print].
- Dimitriadis, S. I., Kanatsoulis, K., Laskaris, N. A., Tsirka, V., Vourkas, M., and Micheloyannis, S. (2012). Surface EEG shows that functional segregation via phase coupling contributes to the neural substrate of mental calculations. *Brain Cogn.* 80, 45–52. doi: 10.1016/j.bandc.2012.04.001
- Dimitriadis, S. I., Laskaris, N. A., Bitzidou, M. P., Tarnanas, I., and Tsolaki, M. N. (2015a). A novel biomarker of amnesic MCI based on dynamic cross-frequency coupling patterns during cognitive brain responses. *Front. Neurosci.* 9:350. doi: 10.3389/fnins.2015.00350
- Dimitriadis, S. I., Laskaris, N. A., Simos, P. G., Fletcher, J. M., and Papanicolaou, A. C. (2016a). Greater repertoire and temporal variability of cross-frequency coupling (CFC) modes in resting-state neuromagnetic recordings among children with reading difficulties. *Front. Hum. Neurosci.* 10:163. doi: 10.3389/fnhum.2016.00163
- Dimitriadis, S. I., Laskaris, N. A., Tsirka, V., Vourkas, M., Micheloyannis, S., and Fotopoulos, S. (2010). Tracking brain dynamics via time-dependent network analysis. *J. Neurosci. Methods* 193, 145–155. doi: 10.1016/j.jneumeth.2010.08.027
- Dimitriadis, S. I., Sun, Y., Kwok, K., Laskaris, N. A., Thakor, N., and Bezerianos, A. (2015b). Cognitive workload assessment based on the tensorial treatment of EEG estimates of cross-frequency phase interactions. *Ann. Biomed. Eng.* 43, 977–989. doi: 10.1007/s10439-014-1143-0
- Dimitriadis, S. I., Sun, Y., Thakor, N., and Bezerianos, A. (2016b). Causal interactions between Frontal $\theta$  - Parieto-Occipital $\alpha 2$  predict behavioral performance in a mental multilevel EEG task. *Front. Hum. Neurosci.* 10:454. doi: 10.3389/fnhum.2016.00454
- Dimitriadis, S. I., Tarnanas, I., Wiederhold, M., Wiederhold, B., and Tsolaki, M. (2016c). Mnemonic-Strategy Training in Mild Cognitive Impairment enhances integration of information processing via cross-frequency-coupling. *Alzheimers Dement.* 2, 241–249. doi: 10.1016/j.trci.2016.08.004
- Dimitriadis, S. I., Sun, Y., Laskaris, N. A., Thakor, N., and Bezerianos, A. (2016d). Revealing cross-frequency causal interactions during a mental arithmetic task through symbolic transfer entropy: a novel vector-quantization approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 1017–1028. doi: 10.1109/TNSRE.2016.2516107
- Dimitriadis, S. I., Zouridakis, G., Rezaie, R., Babajani-Feremi, A., and Papanicolaou, A. C. (2015c). Functional connectivity changes detected with magnetoencephalography after mild traumatic brain injury. *Neuroimage Clin.* 9, 519–531. doi: 10.1016/j.nicl.2015.09.011
- Dimitriadis, S. I., Salis, C., Tarnanas, I., and Linden, D. (2017b). Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel data-driven thresholding scheme based on orthogonal minimal spanning trees (OMSTs). *Front. Neuroinform.* 11:28. doi: 10.3389/fninf.2017.00028
- Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361. doi: 10.1126/science.1194144
- Drakesmith, M., Caeyenberghs, K., Dutt, A., Zammit, S., Evans, C. J., Reichenberg, A., et al. (2015a). Schizophrenia-like topological changes in the structural connectome of individuals with subclinical psychotic experiences. *Hum. Brain Mapp.* 36, 2629–2643. doi: 10.1002/hbm.22796
- Drakesmith, M., Caeyenberghs, K., Dutt, A., Lewis, G., David, A. S., and Jones, D. K. (2015b). Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *Neuroimage* 118, 313–333. doi: 10.1016/j.neuroimage.2015.05.011
- Estrada, E. (2011). *The Structure of Complex Networks: Theory and Applications*. Oxford, UK: Oxford University Press.
- Fornito, A., Zalesky, A., and Breakspear, M. (2015). The connectomics of brain disorders. *Nat. Rev. Neurosci.* 16, 159–172. doi: 10.1038/nrn3901
- Fouss, F., Francoise, K., Yen, L., Pirotte, A., and Saerens, M. (2012). An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification. *Neural Netw.* 31, 53–72. doi: 10.1016/j.neunet.2012.03.001
- Gigandet, X., Hagmann, P., Kuran, M., Cammoun, L., Meuli, R., and Thiran, J.-P. (2008). Estimating the confidence level of white matter connections obtained with MRI tractography. *PLoS ONE* 3:e4006. doi: 10.1371/journal.pone.0004006
- Girard, G., Whittingstall, K., Deriche, R., and Descoteaux, M. (2014). Towards quantitative connectivity analysis: reducing tractography biases. *Neuroimage* 98, 266–278. doi: 10.1016/j.neuroimage.2014.04.074
- Goldberg, D., and Huxley, P. (1980). *Mental Illness in the Community: the Pathway to Psychiatric Care*. London: Tavistock Publications.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6:e159. doi: 10.1371/journal.pbio.0060159
- Hammond, D. K., Gur, Y., and Johnson, C. R. (2013). “Graph diffusion distance: a difference measure for weighted graphs based on the graph laplacian exponential kernel,” in *Global Conference on Signal and Information Processing (GlobalSIP)* (Austin, TX: IEEE), 419–422.
- Jeurissen, B., Leemans, A., Jones, D. K., Tournier, J. D., and Sijbers, J. (2011). Probabilistic fiber tracking using the residual bootstrap with

- constrained spherical deconvolution. *Hum. Brain Mapp.* 32, 461–479. doi: 10.1002/hbm.21032
- Jeurissen, B., Leemans, A., Tournier, J. D., Jones, D. K., and Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum. Brain Mapp.* 34, 2747–2766. doi: 10.1002/hbm.22099
- Jones, D. K., Horsfield, M. A., and Simmons, A. (1999). Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.* 42, 515–525. doi: 10.1002/(SICI)1522-2594(199909)42:3<515::AID-MRM14>3.0.CO;2-Q
- Khazaei A., Ebrahimzadeh A., Babajani-Feremi A., and Alzheimer's Disease Neuroimaging Initiative. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behav. Brain Res.* 322(Pt B), 339–350. doi: 10.1016/j.bbr.2016.06.043
- Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. (2010). Elastix: a toolbox for inhomogeneity-based medical image registration. *IEEE Trans. Med. Imaging* 29, 196–205. doi: 10.1109/TMI.2009.2035616
- Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* 7, 48–50.
- Leemans, A., and Jones, D. K. (2009). The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn. Reson. Med.* 61, 1336–1349. doi: 10.1002/mrm.21890
- Leemans, A., Jeurissen, B., Sijbers, J., and Jones, D. K. (2009). “ExploreDTI: a graphical toolbox for processing, analyzing and visualizing diffusion MR data,” in *Proceedings of the International Society for Magnetic Resonance in Medicine 17th Annual Meeting* (Honolulu, HI), 3536.
- MacLaren, J. Z., Han, S. B., Vos, N., Fischbein, R., and Bammer, R. (2014). Reliability of brain volume measurements: a test-retest dataset. *Sci. Data* 1:140037. doi: 10.1038/sdata.2014.37
- Meier, J., Tewarie, P., and Van Mieghem, P. F. (2015). The union of shortest path trees of functional brain networks. *Brain Connect.* 5, 575–581. doi: 10.1089/brain.2014.0330
- Metzler-Baddeley, C., O'Sullivan, M. J., Bells, S., Pasternak, O., and Jones, D. K. (2012). How and how not to correct for CSF-contamination in diffusion MRI. *Neuroimage* 59, 1394–1403. doi: 10.1016/j.neuroimage.2011.08.043
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8577–8582. doi: 10.1073/pnas.0601602103
- Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., et al. (2013). Multisite functional connectivity MRI classification of autism ABIDE results. *Front. Hum. Neurosci.* 7:599. doi: 10.3389/fnhum.2013.00599
- Pasternak, O., Sochen, N., Gur, Y., Intrator, N., and Assaf, Y. (2009). Free water elimination and mapping from diffusion MRI. *Magn. Reson. Med.* 62, 717–730. doi: 10.1002/mrm.22055
- Perry, A., Wen, W., Lord, A., Thalamuthu, A., Roberts, G., Mitchell, P., et al. (2015). The organisation of the elderly connectome. *Neuroimage* 114, 414–426. doi: 10.1016/j.neuroimage.2015.04.009
- Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell Syst. Tech. J.* 36, 1389–1401. doi: 10.1002/j.1538-7305.1957.tb01515.x
- Roberts, J. A., Perry, A., Roberts, G., Mitchell, P. B., and Breakspear, M. (2016). Consistency-based thresholding of the human connectome. *Neuroimage* 145(Pt A), 118–129. doi: 10.1016/j.neuroimage.2016.09.053
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Shoukri, M. M., Asyali, M. H., and Donner, A. (2004). Sample size requirements for the design of reliability study: review and new results. *Stat. Methods Med. Res.* 13, 251–271. doi: 10.1191/0962280204sm365ra
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Smith, R. E., Tournier, J. D., Calamante, F., and Connelly, A. (2012). Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* 62, 1924–1938. doi: 10.1016/j.neuroimage.2012.06.005
- Thomas, C., Frank, Q. Y., Irfanoglu, M. O., Modi, P., Saleem, K. S., Leopold, D. A., et al. (2014). Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16574–16579. doi: 10.1073/pnas.1405672111
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Vaessen, M. J., Hofman, P. A. M., Tijssen, H. N., Aldenkamp, A. P., Jansen, J. F. A., and Backes, W. H. (2010). The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures. *Neuroimage* 51, 1106–1116. doi: 10.1016/j.neuroimage.2010.03.011
- van den Heuvel, M. P., and Sporns, O. (2011). Rich-club organization of the human connectome. *J. Neurosci.* 31, 15775–15786. doi: 10.1523/JNEUROSCI.3539-11.2011
- Wu, M., Chang, L. C., Walker, L., Lemaitre, H., Barnett, A. S., Marengo, S., et al. (2008). Comparison of EPI distortion correction methods in diffusion tensor MRI using a novel framework. *Med. Image Comput. Comput. Assist. Interv.* 11(Pt 2), 321–329. doi: 10.1007/978-3-540-85990-1\_39
- Yeh, F. C., Badre, D., Verstynen, T. (2016). Connectometry: a statistical approach harnessing the analytical potential of the local connectome. *Neuroimage* 125, 162–171. doi: 10.1016/j.neuroimage.2015.10.053
- Yo, T. S., Anwender, A., Descoteaux, M., Fillard, P., Poupon, C., and Knösche, T. R. (2009). Quantifying brain connectivity: a comparative tractography study. *Med. Image Comput. Comput. Assist. Interv.* 12, 886–893. doi: 10.1007/978-3-642-04268-3\_109
- Zalesky, A., and Fornito, A. (2009). A DTI-derived measure of cortico-cortical connectivity. *IEEE Trans. Med. Imaging* 28, 1023–1036. doi: 10.1109/TMI.2008.2012113
- Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. *Neuroimage* 53, 1197–1207. doi: 10.1016/j.neuroimage.2010.06.041
- Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., van den Heuvel, M. P., and Breakspear, M. (2016). Connectome sensitivity or specificity: which is more important? *Neuroimage* 142, 407–420. doi: 10.1016/j.neuroimage.2016.06.035
- Zuo, X. N., and Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009
- Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* 1:140049. doi: 10.1038/sdata.2014.49. eCollection 2014
- Zuo, X. N., He, Y., Betzel, R. F., Colcombe, S., Sporns, O., and Milham, M. P. (2017). Human connectomics across the life span. *Trends Cogn. Sci.* 21, 32–45. doi: 10.1016/j.tics.2016.10.005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Dimitriadis, Drakesmith, Bells, Parker, Linden and Jones. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Reliability of Static and Dynamic Network Metrics in the Resting-State: A MEG-Beamformed Connectivity Analysis

Stavros I. Dimitriadis<sup>1,2,3,4,5,6\*</sup>, Bethany Routley<sup>1,4</sup>, David E. Linden<sup>1,3,4,5,6</sup> and Krish D. Singh<sup>1,4</sup>

<sup>1</sup> Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom,

<sup>2</sup> Neuroinformatics Group, Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom, <sup>3</sup> Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, United Kingdom, <sup>4</sup> School of Psychology, Cardiff University, Cardiff, United Kingdom, <sup>5</sup> Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, United Kingdom, <sup>6</sup> MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Institute of Psychology (CAS), China

### Reviewed by:

Julia Stephen,  
Mind Research Network (MRN),  
United States  
Chun Kee Chung,  
Seoul National University, South Korea

### \*Correspondence:

Stavros I. Dimitriadis  
stidimitriadis@gmail.com;  
DimitriadisS@cardiff.ac.uk

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 23 January 2018

**Accepted:** 04 July 2018

**Published:** 03 August 2018

### Citation:

Dimitriadis SI, Routley B, Linden DE  
and Singh KD (2018) Reliability of  
Static and Dynamic Network Metrics  
in the Resting-State: A  
MEG-Beamformed Connectivity  
Analysis. *Front. Neurosci.* 12:506.  
doi: 10.3389/fnins.2018.00506

The resting activity of the brain can be described by so-called intrinsic connectivity networks (ICNs), which consist of spatially and temporally distributed, but functionally connected, nodes. The coordinated activity of the resting state can be explored via magnetoencephalography (MEG) by studying frequency-dependent functional brain networks at the source level. Although many algorithms for the analysis of brain connectivity have been proposed, the reliability of network metrics derived from both static and dynamic functional connectivity is still unknown. This is a particular problem for studies of associations between ICN metrics and personality variables or other traits, and for studies of differences between patient and control groups, which both depend critically on the reliability of the metrics used. A detailed investigation of the reliability of metrics derived from resting-state MEG repeat scans is therefore a prerequisite for the development of connectomic biomarkers. Here, we first estimated both static (SFC) and dynamic functional connectivity (DFC) after beamforming source reconstruction using the imaginary part of the phase locking index (iPLV) and the correlation of the amplitude envelope (CorEnv). Using our approach, functional network microstates (FC $\mu$ states) were derived from the DFC and chronnectomics were computed from the evolution of FC $\mu$ states across experimental time. In both temporal scales, the reliability of network metrics (SFC), the FC $\mu$ states and the related chronnectomics were evaluated for every frequency band. Chronnectomic statistics and FC $\mu$ states were generally more reliable than node-wise static network metrics. CorEnv-based network metrics were more reproducible at the static approach. The reliability of chronnectomics have been evaluated also in a second dataset. This study encourages the analysis of MEG resting-state via DFC.

**Keywords:** MEG, resting-state, time-varying network analysis, chronnectomics, functional connectivity microstates, symbolic analysis, reproducibility



## INTRODUCTION

The coordination of spontaneous activity can be characterized with functional connectivity (FC), which refers to statistical dependencies between the activity of distinct brain areas (Pereda et al., 2005) and has been linked to the efficiency of an individual's brain functioning (Baldassarre et al., 2012; Yamashita et al., 2015).

A functional connectivity graph (FCG) can be constructed by estimating the statistical dependencies between the brain activity of all the areas in a pair-wise fashion. An FCG represents statistical or causal relationships measured as cross-correlations, coherence, or information flow (Dimitriadis et al., 2009, 2015d).

Neuroscientists first examined resting-state FC with functional magnetic resonance imaging (fMRI) by correlating blood oxygenation level-dependent (BOLD) signals (Biswal et al., 1995; van den Heuvel et al., 2009; Biswal, 2011, 2012). After 20 years of using fMRI as a dominant neuroimaging tool, the community has succeeded in mapping brain areas to specific brain functions, creating an anatomical-functional atlas (Bandettini, 2012). Although fMRI is of high interest and a key modality to explore human brain function, ultra-slow activity described via BOLD signals is only an indirect measure of brain activity (Logothetis, 2008).

In the last few years, greater attention has been given to explore FC via electro-magneto-encephalography. Even though the spatial resolution of magnetoencephalography (MEG) is lower when compared to fMRI, MEG can capture the multiplexity of human brain activity by providing insight into the spectro-temporo-spatial dynamics of human brain activity. MEG-based FC provides us with a direct measure of neuromagnetic activity with a high temporal resolution (Deco et al., 2011).

Resting-state networks (RSNs) have been successfully extracted with MEG over the past few years using source-space FC (de Pasquale et al., 2010; Brookes et al., 2011a,b; Hipp et al., 2012; Luckhoo et al., 2012; Hall et al., 2013; Wens et al., 2015). Moreover, resting-state MEG FC has been proven to detect abnormal brain functioning in a variety of diseases, including Alzheimer's disease (López et al., 2014, 2017; Engels et al., 2015), multiple sclerosis (Tewarie et al., 2015), in schizophrenia (Bowyer et al., 2015), in dyslexia (Dimitriadis et al., 2013b, 2015c), in mild cognitive impairment (Dimitriadis et al., 2015b) and in mild traumatic brain injury (Dimitriadis et al., 2015c; Dunkley et al., 2015; Antonakakis et al., 2016, 2017).

Several studies have thus captured alterations of MEG parameters in the resting state in order to estimate FC in disease groups compared to controls. However, FC estimates at resting-state could be affected by subject's cognitive, emotional state and other scanning-related systematic differences. For that reason, it is unclear up to which level FC estimates are repeatable for an individual. Moreover, in large studies of hundreds of participants, there is a significant cost, both in financial resources and time, to scan all the subjects two or more times. To establish MEG as a clinically reliable neuroimaging tool that can distinguish disease from healthy populations, the reliability of FC patterns should be explored from repeat scans. Up to date, only a few studies accessed the test-retest reliability

of MEG/electroencephalography (EEG) FC (Jin et al., 2011; Hardmeier et al., 2014; Garcés et al., 2016) while only one study has quantified the test-retest reliability of FC estimates in the source-space MEG (Garcés et al., 2016). Colclough et al. (2016) attempted to report the reliability of every edge-weighted connections with a high number of connectivity estimators but using a split-half strategy from a large pool of subjects. Practically, the results cannot be adopted as reliability of static network metrics since the analysis involved single MEG scan recordings. However, no study has ever explored the reliability of both static and dynamic networks at the source space in MEG.

In the present study, we investigated the test-retest reliability of both static and dynamic FC measures derived from MEG resting-state data. For that purpose, we computed whole-brain FC for 40 subjects who were scanned twice with a 1-week test-retest interval. For each subject and session, MEG-beamformed source activity was estimated and FC was computed between 90 brain areas. FC was estimated with the imaginary part of the phase locking index (iPLV) and the correlation of the amplitude envelope (CorEnv) in both static (SFC) and dynamic models (DFC) by adopting a sliding window approach (de Pasquale et al., 2010; Dimitriadis et al., 2010a, 2012a, 2013a, 2015a, 2016a, 2017a; Dimitriadis and Salis, 2017c). Afterwards, statistical and topological filtering schemes were applied to both SFC and DFC to reveal the true topology (Dimitriadis et al., 2017a,b). For the SFC approach, we estimated well-known network metrics in a node-wise fashion and the reliability was accessed via correlation values between the two measurements and across the cohort. Graph-based reliability was assessed with a novel graph diffusion distance metric.

For the DFC approach, node-wise network metrics were estimated across experimental time. To explore spatio-temporally the derived network activity, we first designed a codebook of prototypical network microstates and then assigned each of the instantaneous connectivity patterns to the most similar code symbol (e.g., functional connectivity graph—FCG) (Dimitriadis et al., 2010a, 2012a, 2013a,b, 2015a, 2016a,b; Dimitriadis and Salis, 2017c). A codebook is a set of prototypical functional connectivity graphs (FCGs). In this way, we derive a unique symbolic time series from each individual where each symbol corresponds to one of the predefined prototypical functional connectivity microstates (FC $\mu$ states). The evolution of these symbol-patterns encapsulates significant state transitions. Furthermore, the evolution of these FC $\mu$ states can be seen as a first order Markovian Chain (MC) that can be modeled representing an individualized state transition model of resting-state FC $\mu$ states. Fractional occupancy of each FC $\mu$ state, transition rates of FC $\mu$ states and MC models are the key features to explore the reliability of chronnectome in MEG source space. The group-consistency of subject-specific FC $\mu$ states was further explored. The whole analysis of dynamic functional connectivity graphs and the definition of FC $\mu$ states have been described in previous paper (Dimitriadis et al., 2013a).

Many techniques have already been proposed to summarize brain activity into short-lived transient brain states using the spectrum of neuromagnetic recordings (Vidaurre et al.,

2016) and also the band-limited amplitude envelopes of source reconstructed MEG data (Baker et al., 2014; O'Neill et al., 2015). In detail, Vidaurre et al. (2016) proposed a combination of multivariate autoregressive model with hidden markovian modeling (MAR-HMM) in order to model the temporal, spectral and spatial properties of MEG reconstructed activity into very short-lived brain states. Similarly, (Baker et al., 2014) modeled resting-state source-reconstructed MEG activity with HMM into distinct spatio-temporal activation profiles called brain states. These brain states were linked to well-known anatomical brain areas. O'Neill et al. (2015) mined MEG source activity from two tasks, a self-paced motor and a Sternberg working memory task. He used a sliding window canonical correlation analysis (CCA) to estimate the functional connectivity at each time-window and a k-means clustering to detect repeatable spatial patterns of connectivity that form transiently synchronizing sub-networks (TSNs) or functional connectivity microstates. Here, we must underline the distinction of summarizing brain activity using the raw time series (ROIs  $\times$  sliding windows; Baker et al., 2014; Vidaurre et al., 2016) which is a 2D matrix and the dynamic functional brain networks (ROIs  $\times$  ROIs  $\times$  sliding windows) which is a 3D matrix (O'Neill et al., 2015). Currently the mapping and the relationship between raw activity and brain connectivity and also the relationship of microstates (raw activity) with functional connectivity microstates (dynamic graphs; Allen et al., 2012; Dimitriadis et al., 2013a,b, 2015a, 2016a,b; Dimitriadis and Salis, 2017c) is still unknown. Further research is needed to explore their mapping at resting-state and during tasks.

The proposed methodological scheme entails two distinct ways of analyzing dynamic functional connectivity patterns. These patterns are representative brain network topologies across subjects and brain rhythms and are directly linked to a brain state (Buzsáki and Draguhn, 2004). The very first approaches in fMRI constitutes novel contributions to an emerging neuroimaging field called chronnectomics (Allen et al., 2012; Calhoun et al., 2014). Previously, we reported the notion of FC $\mu$ states (Dimitriadis et al., 2013a) and the developmental trends in cognition (Dimitriadis et al., 2015a) using electroencephalographic recordings. The concept of chronnectome is the incorporation of a dynamic view of functional brain connectivity networks and the evolution of revisiting distinct spatio-temporal brain states (functional connectivity microstates—FC $\mu$ states). To the best of our knowledge, this study constitutes the first attempt to assess the test-retest reliability of Dynamic Functional Connectivity at the MEG source level.

Despite growing enthusiasm in the neuroscience community about the potential contribution of neuroimaging and especially brain networks in the designing of connectomic biomarkers for various brain diseases/disorders, many challenges remain open (Stam, 2014). At first level, it is more than significant to explore how reliable are network metrics at both temporal scales (static and dynamic) by analyzing a group of control subjects with repeat scans (e.g., diffusion MRI: Dimitriadis et al., 2017d). Here, we assess evidence of the reliability of neuromagnetic (MEG) based functional connectomics to lead to potential clinically meaningful biomarker identification in target

populations through the lens of the criteria used to evaluate clinical tests.

## MATERIALS AND METHODS

### Subjects

40 healthy subjects (age  $22.85 \pm 3.74$  years, 15 women and 25 men) underwent two resting-state MEG sessions (eyes open) with a 1-week test-retest interval. For each participant, scans were scheduled at the same day of the week and same time of the day. The duration of MEG resting-state was 5 min for every participant. The study was approved by the Ethics Committee of the School of Psychology at Cardiff University, and participants provided informed and written consent.

### MEG-MRI Recordings

Whole-head MEG recordings were made using a 275-channel CTF radial gradiometer system. An additional 29 reference channels were recorded for noise cancellation purposes and the primary sensors were analyzed as synthetic third-order gradiometers (Vrba and Robinson, 2001). Two or three of the 275 channels were turned off due to excessive sensor noise (depending on time of acquisition). Subjects were seated upright in the magnetically shielded room. To achieve MRI/MEG co-registration, fiducial markers were placed at fixed distances from three anatomical landmarks identifiable in the subject's anatomical MRI, and their locations were verified afterwards using high-resolution digital photographs. Head localization was performed before and after each recording, and a trigger was sent to the acquisition computer at relevant stimulus events.

All datasets were either acquired at or down-sampled to 600 Hz, and filtered with a 1-Hz high-pass and a 200-Hz lowpass filter. The data were first whitened and reduced in dimensionality using principal component analysis with a threshold set to 95% of the total variance (Delorme and Makeig, 2004). The statistical values of kurtosis, Rényi entropy and skewness of each independent component were used to eliminate ocular and cardiac artifacts. Specifically, a component was deemed artifactual if more than 20% of its values after normalization to zero-mean and unit-variance were outside the range of  $[-2, +2]$  (Delorme and Makeig, 2004; Escudero et al., 2011; Antonakakis et al., 2016). The artifact-free multichannel MEG resting-state recordings were then entered in the beamforming analysis (see next section).

Subjects further underwent an MRI session in which a 3T GE scanner with an eight-channel receive-only head RF coil T1-weighted 1-mm anatomical scan was acquired, using an inversion recovery spoiled gradient echo acquisition.

### Beamforming

An atlas-based beamformer approach was adopted to project MEG data from the sensor level to source space independently for each brain rhythm. The frequency bands studied were:  $\delta$  (0.5–4 Hz),  $\theta$  (4–8 Hz),  $\alpha_1$  (8–10 Hz),  $\alpha_2$  (10–13 Hz),  $\beta_1$  (13–20 Hz),  $\beta_2$  (20–30 Hz),  $\gamma_1$  (30–45 Hz),  $\gamma_2$  (55–90 Hz). First, the coregistered MRI was spatially normalized to a template MRI using SPM8 (Weiskopf et al., 2011). The AAL atlas was used to anatomically label the voxels, for each participant and session,

**TABLE 1** | Optimization of the width of temporal window and the stepping criterion per frequency band and for both connectivity estimators.

		$\delta$	$\theta$	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$
iPLV	{time window, step}	{2, 0.5}	{3, 0.4}	{4, 0.4}	{4, 0.4}	{6, 0.3}	{6.5, 0.3}	{8, 0.2}	{8.5, 0.2}
CorEnv	{time window, step}	{2, 0.5}	{3, 0.4}	{5, 0.4}	{5, 0.4}	{7, 0.3}	{7, 0.3}	{8, 0.3}	{9, 0.2}

in this template space (Tzourio-Mazoyer et al., 2002). The 90 cortical regions of interest (ROIs) were used for further analysis, as is common in recent studies (Hillebrand and Barnes, 2002; Hillebrand et al., 2016; Hunt et al., 2016). Next, neuronal activity in the atlas-labeled voxels was reconstructed using the LCMV source localization algorithm as implemented in Fieldtrip (Oostenveld et al., 2011).

The beamformer sequentially reconstructs the activity for each voxel in a predefined grid covering the entire brain (spacing 6 mm) by weighting the contribution of each MEG sensor to a voxel's time series—a procedure that creates the spatial filters that can then project sensor activity to the cortical activity. Each ROI in the atlas contains many voxels, and the number of voxels per ROI differs. To obtain a single representative time series for every ROI, we defined a functional-centroid ROI representative by functionally interpolating activity from the voxel time series, within each ROI, in a weighted fashion. Specifically, we estimated a functional connectivity map between every pair of source time series within each of the AALs ROIs (Equation 1) using correlation (Equation 2). We then estimated the connectivity strength of each voxel within the ROI by summing its connectivity values to other voxels within the same ROI (Equation 3) and finally we normalized each strength by the sum of strengths (Equation 4) to estimate a set of weights within the ROI that sum to a value of 1. Finally, we multiplied each voxel time series with their respective weights and we summed across them in order to get a representative time series for each ROI (Equation 5). The whole procedure was applied independently to every quasi-stable temporal segment derived by the settings of temporal window and stepping criterion.

The following Equations 1–5 demonstrated the steps for this functional interpolation.

$$\begin{aligned}
 ROI^{map} &\in \mathbb{R}^{voxels \times voxels}, Voxels \in \text{no of voxel timeseries within} \\
 &\text{each ROI} \quad (1) \\
 S^{Voxels} &= \sum_{k=1}^{Voxels} \sum_{l=k+1}^{Voxels} corr(ROI_k^{map}(t), ROI_l^{map}(t)), \\
 S^{Voxels} &\in ROI \times ROI \quad (2) \\
 SS &= \sum_{k=1}^{Voxels} corr(k, :), S^{voxels} \in 1 \times ROI \quad (3) \\
 W_k &= \frac{SS}{\sum_{k=1}^{Voxels} SS} \quad (4) \\
 ROI^{activity} &= \sum_{k=1}^{Voxels} ROI^{time\ series} * W_k \quad (5)
 \end{aligned}$$

The outline of the methodology is described in **Figure 1**. An exemplar of the representative bandpass filtered ROI based time series is given in **Figure 1**. **Figure 2** illustrates the preprocessing steps described in Equations (1–5).

## Functional Connectivity

Here, functional connectivity was examined among the following 8 brain rhythms of the typical sub-bands of electrophysiological neural signals  $\{\delta, \theta, \alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2\}$ , defined respectively within the ranges  $\{0.5\text{--}4\text{ Hz}; 4\text{--}8\text{ Hz}; 8\text{--}10\text{ Hz}; 10\text{--}13\text{ Hz}; 13\text{--}20\text{ Hz}; 20\text{--}30\text{ Hz}; 30\text{--}45\text{ Hz}; 55\text{--}90\text{ Hz}\}$ . For both static and dynamic approach, we used two estimators: the correlation of the amplitude envelope (CorEnv) and the imaginary part of the phase locking value (iPLV).

## Intra-Frequency Connectivity Estimators

Among the available connectivity estimators, we adopted one based on the imaginary part of phase-locking value (iPLV) (Lachaux et al., 1999) and adjusted properly so as to extract time-resolved profiles of intra-frequency coupling from MEG multichannel recordings at resting state. The original PLV is defined as follows:

$$PLV = \frac{1}{T} \sum_{t=1}^T e^{i(\varphi_k^{(t)}, \varphi_l^{(t)})} \quad (6)$$

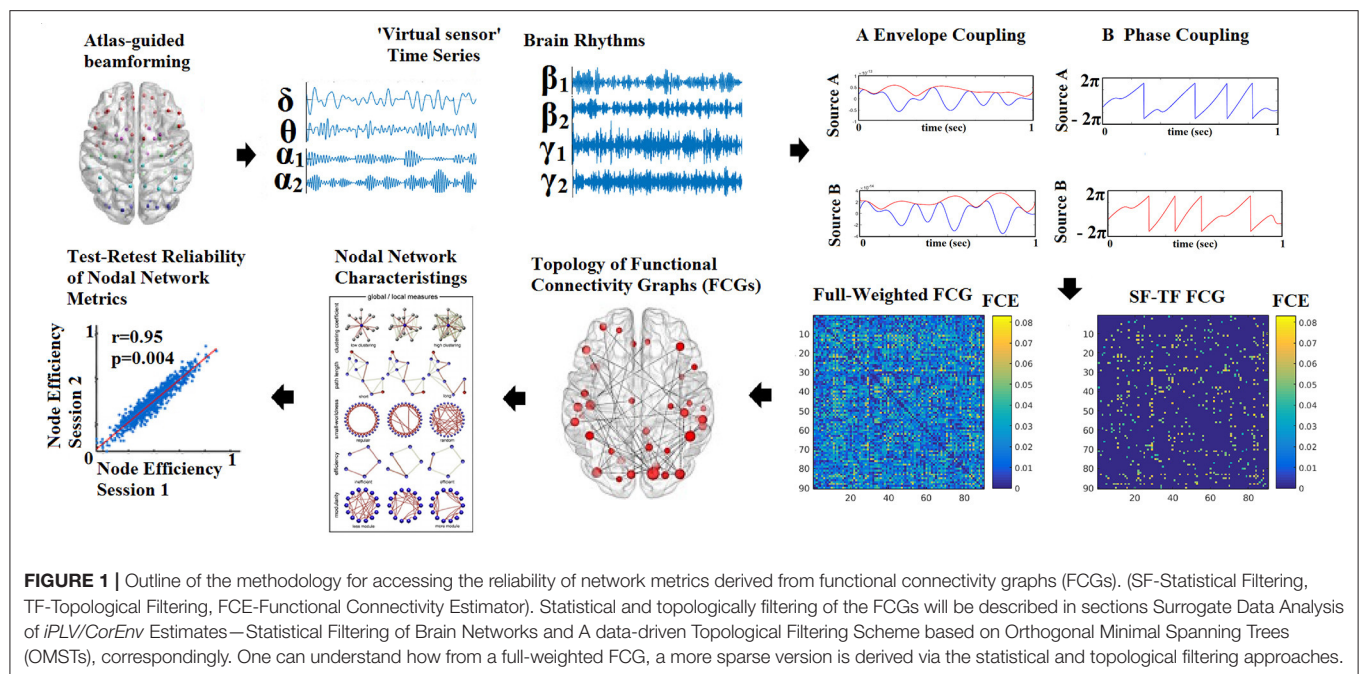
where  $k, l$  denote a pair of MEG sources and the imaginary part of PLV is equal to:

$$Im\{PLV\} = \frac{1}{T} \left| Im \left\{ \sum_{t=1}^T e^{i(\varphi_k^{(t)}, \varphi_l^{(t)})} \right\} \right| \quad (7)$$

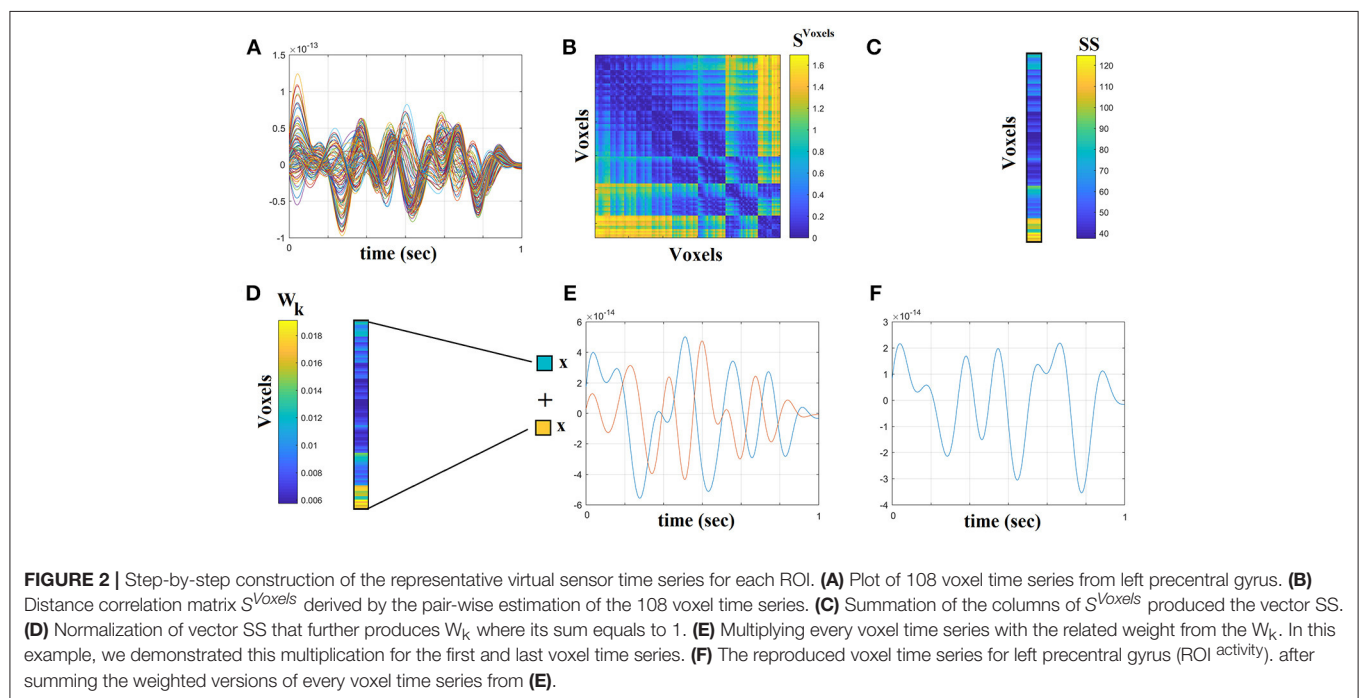
The imaginary part of PLV (iPLV) investigates intra-frequency interactions without putative contributions from volume conductance. In general, the iPLV is mainly sensitive to non-zero-phase lags and for that reason is resistant to instantaneous self-interactions from volume conductance (Nolte et al., 2004). In contrast, it could be sensitive to phase changes that not necessarily imply a PLV oriented coupling.

Correlation of the Envelope coupling (CorEnv) is based upon correlation between the oscillatory envelopes of two frequency band limited sources (Brookes et al., 2012). See **Figure 1** for a schematic diagram of phase and envelope based connectivity analyses based upon neural oscillations. Correlation of the Envelope coupling (CorEnv) is based upon correlation between the oscillatory envelopes of two band limited sources (A) while phase coupling searches for a constant phase lag between signals,





**FIGURE 1** | Outline of the methodology for accessing the reliability of network metrics derived from functional connectivity graphs (FCGs). (SF-Statistical Filtering, TF-Topological Filtering, FCE-Functional Connectivity Estimator). Statistical and topologically filtering of the FCGs will be described in sections Surrogate Data Analysis of iPLV/CorEnv Estimates—Statistical Filtering of Brain Networks and A data-driven Topological Filtering Scheme based on Orthogonal Minimal Spanning Trees (OMSTs), correspondingly. One can understand how from a full-weighted FCG, a more sparse version is derived via the statistical and topological filtering approaches.



in the example a difference of  $\pi$  (B). The time series for the estimation of CorEnv were orthogonalized between each other using the bivariate version of this correction for signal leakage effects (Colclough et al., 2015).

## Static Functional Connectivity Analysis

Using both connectivity estimators, we estimated the fully-weighted ( $90 \times 90$ ) anatomical oriented FCG, one for each subject, recording session and frequency band. To construct the

static FCG (SFCG), we incorporated in the analysis the whole 5 min of the recording session.

## Dynamic iPLV Estimates: The Time-Varying Integrated iPLV Graph ( $^{TV}iPLV$ Graph)

The goal of the analytic procedures described in this section was to understand the repertoire of phase-to-phase interactions and their temporal evolution, while taking into account the quasi-instantaneous spatiotemporal distribution of iPLV estimates.



This was achieved by computing one set of *iPLV* estimates within each of a series of sliding overlapping windows spanning the entire 5-min continuous MEG recording for eyes-open condition. The width of the temporal window and the stepping criterion were optimized for each frequency band separately using as objective criterion the reliability of transition dynamics between scan session 1 and 2 for each brain rhythm (Dimitriadis et al., 2013a; see sections State Transition Rate and Optimizing the Width of the Time-Window and the Stepping Criterion). The center of the stepping window moved forwards every frequency-dependent time-window (see sections Optimizing the Width of the Time-Window and the Stepping Criterion and Tuning Parameters for Dynamic Functional Connectivity Analysis) for the optimization of the parameters) for every intra-frequency interactions and a new functional brain network is re-estimated between every pair of “swifting” temporal segments of MEG activity, from two sources, leading to a “quasi-stable in time” static *iPLV* graph. In this manner, a series of 598 (for  $\delta$ ) to 2,140 (for  $\gamma_2$ ) *iPLV* graph estimates were computed for each frequency (8 within frequency), for each participant and for both repeat scans.

For each subject, a 4D tensor (frequencies bands (8)  $\times$  slides (598–2,140)  $\times$  sources (90)  $\times$  sources (90); see sections Optimizing the Width of the Time-Window and the Stepping Criterion and Tuning Parameters for Dynamic Functional Connectivity Analysis) was created for each condition integrating subject-specific spatio-temporal phase interactions (Figure 3A).

## Surrogate Data Analysis of *iPLV/CorEnv* Estimates—Statistical Filtering of Brain Networks

To identify significant *iPLV/CorEnv*-interactions which were estimated for every pair of frequencies within and between all 90 sources, and at each successive sliding window (i.e., temporal segment), we employed a surrogate data analysis. Accordingly, we could determine (a) if a given *iPLV/CorEnv* value differed from what would be expected by chance alone, and (b) if a non-zero *iPLV/CorEnv* corresponded to non-spurious coupling.

For every temporal segment, sensor-pair and frequency, we tested the null hypothesis  $H_0$ : “the observed *iPLV/CorEnv* value comes from the same distribution as the distribution of surrogate *iPLV/CorEnv*-values”. One thousand surrogate time-series were generated by cutting at a single point at a random location the original time series and exchanging the two resulting time courses (Aru et al., 2015). We restricted the range of the selected cutting point in a temporal window of width up to 10 s in apart from the middle of the recording session (between 140 and 160 s). This surrogate scheme was applied to the original whole time series and not to the signal-segment at every slide. Repeating this procedure leads to a set of surrogates with a minimal distortion of the original phase dynamics, while the non-stationarity of the brain activity is less destroyed compared to shuffling the time series or cutting and rebuilding it in more than one time points.

This procedure ensures that the real and surrogate indices both have the same statistical properties. For each data set, the surrogate *iPLV/CorEnv* ( *$s$ iPLV/ $s$ CorEnv*) was then computed.

We then determined a one-sided *p*-value for each *iPLV/CorEnv* value that corresponded to the likelihood that the observed value could belong to the surrogate distribution. This was done by directly estimating the proportion of “surrogate”  *$s$ iPLV/ $s$ CorEnv* that was higher than the observed *iPLV/CorEnv*. The *p*-value reflected the statistical significance of the observed *iPLV/CorEnv*-level (a very low value revealed that it could not have appeared from processes with no *iPLV* coupling).

At a second level, we applied the FDR method (Benjamini and Hochberg, 1995) to control for multiple comparisons within each snapshot of the dynamic graph (FCG—a  $90 \times 90$  matrix with tabulated *p*-values) with the expected fraction of false positives set to  $q \leq 0.01$ . Finally, for each subject the resulting  *$TV$ iPLV/ $TV$ CorEnv* profiles constituted of two 3D arrays of size [598 to 2,140 for  $\delta$  to  $\gamma_2$  (time windows)  $\times$  90 (sources)  $\times$  90 (sources)] with a value of 0 indicated a non-significant *iPLV/CorEnv* value.

The aforementioned statistical filtering approach was applied independently for each frequency band, session, subject, and connectivity estimator for both static and dynamic functional connectivity graphs.

## A Data-Driven Topological Filtering Scheme Based on Orthogonal Minimal Spanning Trees (OMSTs)

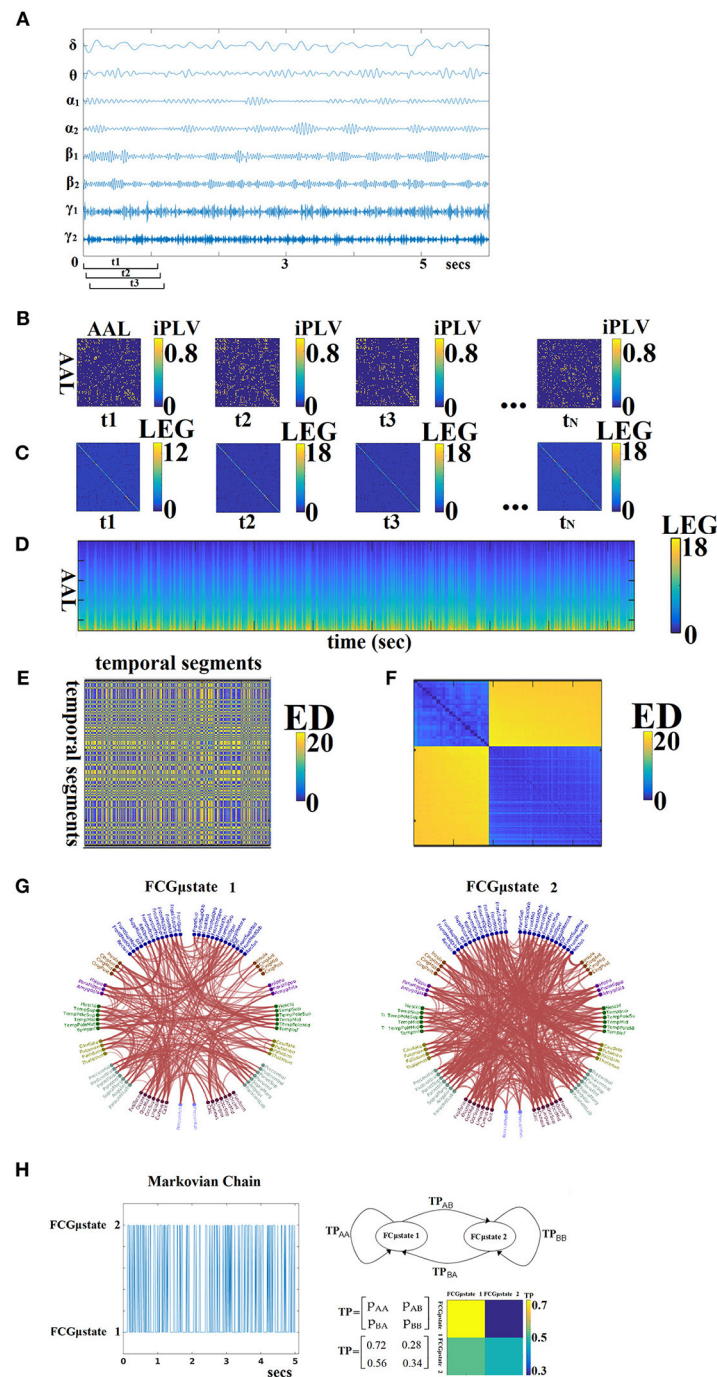
As well as the statistical filtering approach, it is important to adopt a data-driven topological filtering approach in order to reveal the backbone of the network topology over the increment of information flow.

Recently, it was proved that MST is an unbiased method that yields reliable network metrics (Tewarie et al., 2015). In this study, we adopt a variant of this topological filtering scheme called orthogonal minimal spanning trees (OMST), which leads to a better sampling of brain networks, preserving the advantage of MST, that connects the whole network with minimum cost without introducing cycles and without differentiated strong from weak connections compared to the absolute threshold or the density threshold (Dimitriadis et al., 2017a,b). MST is too sparse to capture the “true” network and for that reason leading to the selection of  $N-1$  connections where  $N$  denotes the number of nodes. We introduced OMST which samples the weights of a brain network via the notion of MST and under the optimization of the global information flow under the constraint of the total Cost of preserving the functional connections (Dimitriadis et al., 2017b,d).

Our criterion for topologically filtering a given brain network is based on the maximum value of the following quality formula:

$$J_{GCE}^{OMSTs} = GE - Cost \quad (8)$$

We applied the data-driven topological filtering scheme based on OMST at every static and quasi-instantaneous FCG from the dynamic DFCG. After statistical and topological filtering approaches applied to both SFCG and the DFCG, we estimated network metrics at the node/source level.



**FIGURE 3 |** From dynamic functional connectivity graphs (DFCG) to FC $\mu$ states. **(A)** A characteristic bandpass filtered time series for each of the studying frequency band is given from a ROI. **(B)** Topologies of snapshots of DFCG from the three first temporal segments from  $\delta$  band of subject 1 in order to make clear the estimation of FCG in a dynamic fashion. The first three brain networks refer to the first three temporal segments demonstrated in **(A)**. These functional brain networks were statistically and topologically filtered as described in sections Surrogate Data Analysis of *iPLV/CorEnv* Estimates—Statistical Filtering of Brain Networks and A data-driven Topological Filtering Scheme based on Orthogonal Minimal Spanning Trees (OMSTs). The  $t_N$  refers to the last temporal segment of the DFCG. **(C)** Laplacian matrices for a few snapshots of DFCG. **(D)** The dynamic evolution of the eigenvalues of the laplacian matrices for each frequency band. An example for  $\delta$  frequency band. **(E)** Euclidean Distance matrix of the laplacian eigenvalues between every pair of temporal segments. **(F)** Reordering the correlation matrix in **(E)** to enhance the visualization of the two clusters—FC $\mu$ states illustrated in **(G)**. **(G)** The prototypical FC $\mu$ states in a circular visualization. **(H)** The outcome of this procedure is a symbolic time series that can be seen as a first order Markovian Chain that expresses the evolution of FC $\mu$ states across experimental time. The transition probabilities (TP) of this Markovian Chain based is illustrated in the  $2 \times 2$  colored figure. One can understand that human brain demonstrates a preferred transition from FC $\mu$ states<sup>2</sup> to FC $\mu$ states<sup>1</sup> compared to the opposite direction (see 2D colormap). The chronnectomics were derived from this symbolic time series. ED, Euclidean distance; LEG, Laplacian EiGenvalues; ED, Euclidean distance; AAL, automated anatomical labeling; LEG, Laplacian EiGenvalues.

**Figure 1** demonstrates an example of a full-weighted FCG after applying both statistical and topological filtering approach. Our algorithm was validated over all the existing thresholding schemes with a large EEG dataset over brain fingerprinting and with a multi-scan fMRI dataset over reliability of nodal network metrics (Dimitriadis et al., 2017a). Additionally, we demonstrated the importance of a data-driven topological filtering technique in functional neuroimaging by using OMST in a multi-group study with MEG resting-state recordings (Dimitriadis et al., 2017b). The MATLAB code of the OMST method and also of the majority of existing filtering methods can be downloaded from the: [https://github.com/stdimitr/topological\\_filtering\\_networks](https://github.com/stdimitr/topological_filtering_networks) & researchgate [https://www.researchgate.net/profile/Stavros\\_Dimitriadis](https://www.researchgate.net/profile/Stavros_Dimitriadis).

## Graph Diffusion Distance Metric for Brain Networks

In order to assess group and scan sessions differences in the topologically filtered FCG at the single-case level, we computed the Graph Diffusion Distance as a distance metric (Fouss et al., 2012; Hammond et al., 2013) from the OMST-derived final Functional Connectivity Graphs (FCG). The graph laplacian operator of each subject-specific FCG was defined as  $L = D - FCG$ , where  $D$  is a diagonal degree matrix related to FCG. This method entails modeling hypothetical patterns of information flow among sources based on each observed (static) SFCG. The diffusion process on the person-specific FCG was allowed for a set time  $t$ ; the quantity that underwent diffusion at each time point is represented by the time-varying vector  $u(t) \in \mathbb{R}^N$ . Thus, for a pair of sources  $i$  and  $j$ , the quantity  $FCG_{ij}(u_i(t) - u_j(t))$  represents the hypothetical flow of information from  $i$  to  $j$  via the edges that connect them (both directly and indirectly). Summing all these hypothetical interactions for each sensor leads to  $u'_j(t) = \sum_i FCG_{ij}(u_i(t) - u_j(t))$ , which can be written as:

$$u'(t) = -Lu(t) \quad (9)$$

where  $L$  is the graph laplacian of FCG. At time  $t = 0$  Equation 9 has the analytic solution:  $u(t) = \exp(-tL)u^{(0)}$ . Here  $\exp(-tL)$  is a  $N \times N$  matrix function of  $t$ , known as Laplacian exponential diffusion kernel (Fouss et al., 2012), and  $u^{(0)} = e_j$ , where  $e_j \in \mathbb{R}^N$  is the unit vector with all zeros except in the  $j$ th component. Running the diffusion process through time  $t$  produced the diffusion pattern  $\exp(-tL)e_j$  which corresponds to the  $j$ th column of  $\exp(-tL)$ .

Next, a metric of dissimilarity between every possible pair of person-specific diffusion-kernelized FCGs ( $FCG_1, FCG_2$ ) was computed in the form of the graph diffusion distance  $d_{gdd}(t)$ . The higher the value of  $d_{gdd}(t)$  between two graphs, the more distinct is their network topology as well as the corresponding, hypothetical information flow. The columns of the Laplacian exponential kernels,  $\exp(-tL_1)$  and  $\exp(-tL_2)$ , describe distinct diffusion patterns, centered at two corresponding sources within each FCG. The  $d_{gdd}(t)$  function is searching for a diffusion time  $t$  that maximizes the Frobenius norm of the sum of squared

differences between these patterns, summed over all sources, and is computed as:

$$d_{gdd}(t) = \|\exp(-tL_1) - \exp(-tL_2)\|_F^2 \quad (10)$$

where  $\|\cdot\|_F$  is the Frobenius norm.

Given the spectral decomposition  $L = V\Lambda V$ , the laplacian exponential can be estimated via

$$\exp(-tL) = V\exp(-t\Lambda)V^T \quad (11)$$

where for  $\Lambda$ ,  $\exp(-t\Lambda)$  is diagonal to the  $i$ th entry given by  $e^{-t\Lambda_{i,i}}$ . We computed  $d_{gdd}(FCG_1, FCG_2)$  by first diagonalizing  $L_1$  and  $L_2$  and then applying Equations (9, 10) to estimate  $d_{gdd}(t)$  for each time point  $t$  of the diffusion process. In this manner, a single dissimilarity value was computed for each pair of participants based on their individual characteristic FCGs. For further details see (Hammond et al., 2013). The GDD metric can be downloaded from:

<https://github.com/stdimitr/multi-group-analysis-OMST-GDD>.

## Static Network Metrics

After applying the statistical and topological filtering approach, we estimated the global efficiency for each node in static approach. The static approach leads to 90 (sources) values for each network, metric, frequency band and session per subject. We adopted complementary features that measure the importance of each node in segregation, integration and the information flow within a weighted functional brain network (Dimitriadis et al., 2010a,b, 2013a,b, 2015a). In this study, we estimated four basic network metrics, the global and local efficiency, the strength of each node and the mean first passage time based on random walks (Goñi et al., 2013).

**Network global efficiency (GE)** reflects the overall efficiency of parallel information transfer within the entire set of 90 sources and was estimated as the average source specific GE value over all sources using the following formula (Latora and Marchiori, 2001):

$$GE = \frac{1}{N} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} (d_{ij})^{-1}}{N-1} \quad (12)$$

where  $d$  denotes the shortest path length from  $i$  to  $j$ .

**Local efficiency (LE)** indicates how well the subgraphs exchange information when a particular node is eliminated (Achard and Bullmore, 2007). Specifically, each node is assigned the shortest path length within its subgraph  $G_i G_i$

$$LE = \frac{1}{N} \sum_{i \in N} nodal_{LE_i} = \frac{1}{N} \sum_{i \in N} \frac{\sum_{j, h \in G_i, j, h \neq i} (d_{jh})^{-1}}{k_i(k_i - 1)} \quad (13)$$

where  $k_i$  corresponds to the total number of spatial (first level neighbors) neighbors of the  $i$ -th node, while  $d$  denotes shortest path length.



The **strength** is equal to the total sum of the weights of the connections of each node.

As a fourth candidate network metric, we adopted the mean first passage time (MFPT). Starting a random walk process on a brain network, an analytic expression can give the probability that a single particle departing from a node  $i$  arrives at node  $j$  for the first time within exactly  $L$  steps (Wang and Pei, 2008). This criterion can be applied for each MEG source pair by setting  $L$  to their shortest-path-length. We denote with  $\Pi_G = [\pi_{ij}]$  the  $n \times n$  symmetric matrix containing, for each pair of nodes, the probability of a single particle going from node  $i$  to node  $j$  via the shortest path. Each entry  $\pi_{ij}$  can be computed as

$$\pi_{ij} = 1 - \sum_{v=1}^n [B_j^{\varphi_{ij}}]_{iv}, i \neq j \quad (14)$$

where matrix  $B_j$  is the transition matrix  $P$  introduced above, but with all zeros in the  $j$ -th column, i.e., with  $j$  acting as an absorbing state (Wang and Pei, 2008). Evaluating shortest-path-lengths ensures that  $\forall i, j \quad \pi_{ij} > 0$ . By considering one particle

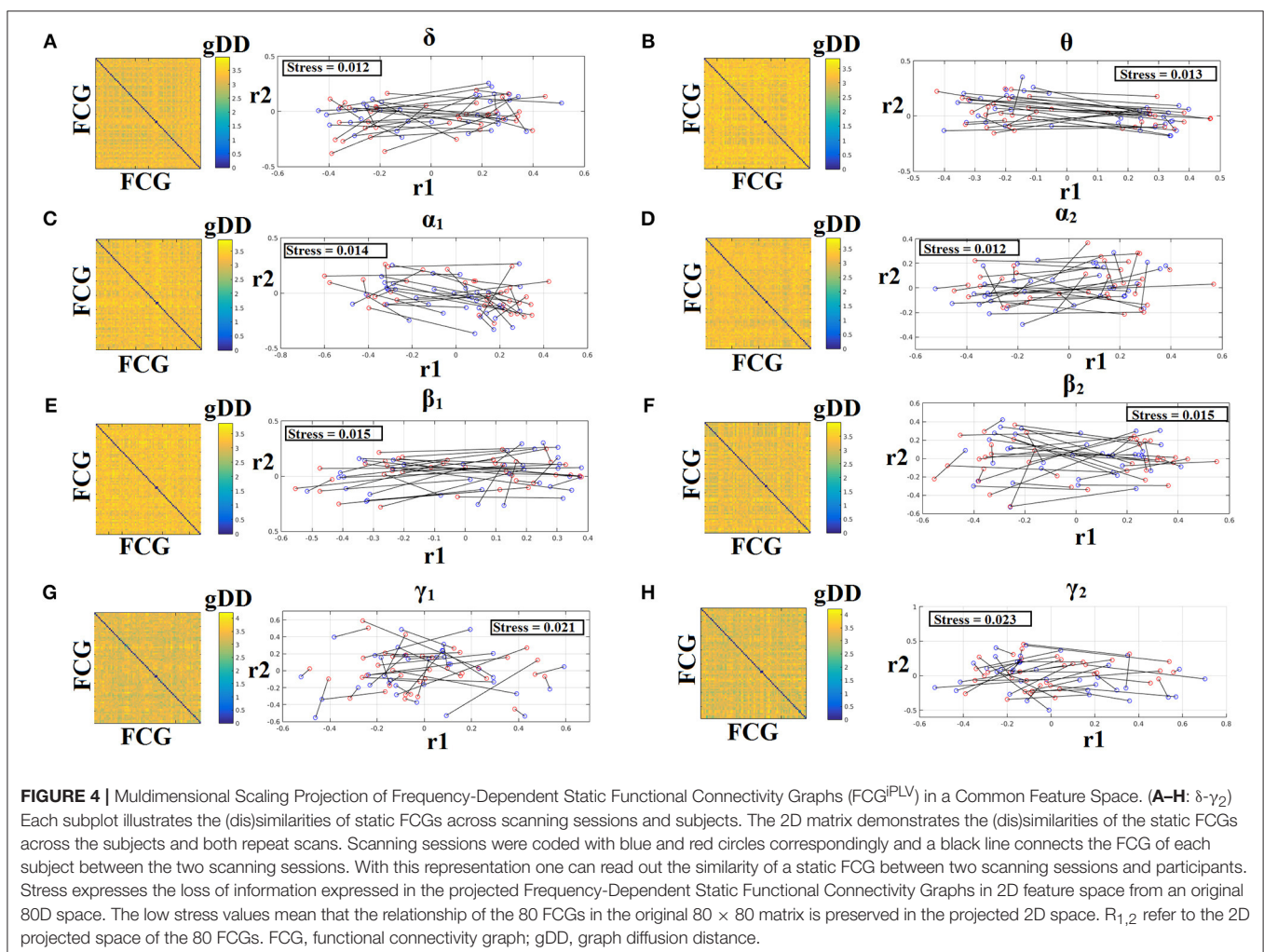
here, the average shortest-path probability of a graph is defined as

$$\Pi_{spl} = \frac{\sum_i \sum_j \pi_{ij}}{n(n-1)}, i \neq j \quad (15)$$

The derived 2D matrix based on nodal NMTs of GE, LE, MFPT and strength will be modeled with the proposed method that is described in the following section.

## Modeling of Dynamic Functional Connectivity Graphs (DFCG) as a 3D Tensor

This subsection serves as a brief introduction to our symbolization scheme, presented in greater details elsewhere (Dimitriadis et al., 2012a,b, 2013a,b). The dynamic functional connectivity patterns can be modeled as prototypical functional connectivity microstates (FCμstates). In a recent study, we demonstrated a better modeling of dynamic functional connectivity graphs (DFCG) based on a vector quantization approach (Dimitriadis et al., 2013a). In our previous work (Dimitriadis et al., 2013a,b, 2015a), we used the neural-gas





algorithm (Martinetz et al., 1993) to learn the 2D matrix (vectorized version of 2D matrix  $\times$  time) leading to a codebook of  $k$  prototypical functional connectivity states (i.e., FC $\mu$ states). This algorithm is an artificial neural network model, which converges efficiently to a small number  $k$  of codebook vectors, using a stochastic gradient descent procedure with a softmax adaptation rule that minimizes the average distortion error (Martinetz et al., 1993). In a recent study, we adopted non-negative matrix factorization (NNMF) as an appropriate learning algorithm of the 2D vectorized version of a dynamic functional brain network (Marimpis et al., 2016).

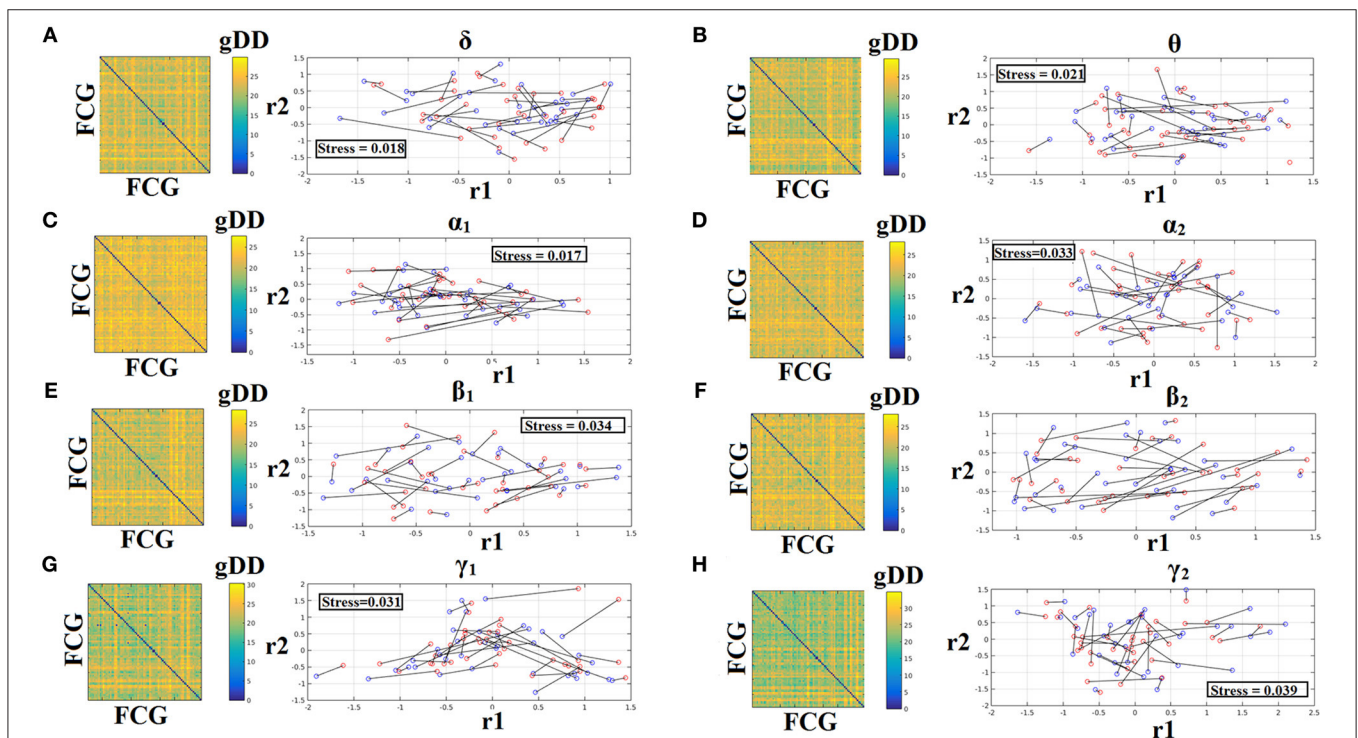
In our previous study where we first demonstrated how to model dynamic functional connectivity graph (dFCG) (Dimitriadis et al., 2013a), we vectorised the upper triangular of each of the quasi-static FCGs building a 2D matrix where the 1st dimension is the number of temporal segments and the 2nd the vectorised version of a static FCG. The final outcome of this approach is to define the so-called functional connectivity microstates (FC $\mu$ states). In a next study, we moved one step further by estimating node-wise global efficiency as the best descriptor to characterize the brain activity. The final outcome of the modeling using the same methodology of neural-gas algorithm was task-based network microstates (Dimitriadis et al., 2015a). Here, the vectorized version of a  $90 \times 90$  FCG produces

a long vector of 4,005 values while the number of temporal segments ranged from 598 to 2,140 which caused the so-called curse of dimensionality where the number of number of the temporal segment over which the modeling will learn the brain states is much smaller compared to the vectorized snapshot of FCG. Simultaneously, the vectorized notion of a brain network didn't maintain the inherent format of a functional brain network which is a 2D matrix, a tensor.

The outline of this procedure is illustrated in **Figure 3**. In **Figure 3A**, a characteristic bandpass filtered time series for each of the studying frequency band is estimated from each ROI. Here, instead of vectorising the upper triangular of an undirected FCG, we used the statistical and topological filtering FCG on its inherent format which is a 2D tensor. In the case of dynamic networks, the dimension is a 3D tensor where the 3rd dimension is the time. **Figure 3B** illustrates a few snapshots of the dFCG for  $\delta$  frequency of the first subject. At the next level, we estimated the laplacian matrix of each quasi-static FCG. Given a FCG, the laplacian matrix is given by:

$$L = D - A, \quad (16)$$

where  $D$  is the degree matrix and  $A$  is the FCG.



**FIGURE 5 |** Multidimensional Scaling Projection of Frequency-Dependent Static Functional Connectivity Graphs (FCG<sup>CorEnv</sup>) in a Common Feature Space. **(A–H:  $\delta - \gamma_2$ )** Each subplot illustrates the (dis)similarities of static FCGs across scanning sessions and subjects. The 2D matrix demonstrates the (dis)similarities of the static FCGs across the subjects and both repeat scans. Scanning sessions were coded with blue and red circles correspondingly and a black line connects the FCG of each subject between the two scanning sessions. With this representation one can read out the similarity of a static FCG between two scanning sessions and participants. Stress expresses the loss of information expressed in the projected Frequency-Dependent Static Functional Connectivity Graphs in 2D feature space from an original 80D space. The low stress values mean that the relationship of the 80 FCGs in the original  $80 \times 80$  matrix is preserved in the projected 2D space.  $R_{1,2}$  refer to the 2D projected space of the 80 FCGs. FCG, functional connectivity graph; gDD, graph diffusion distance.

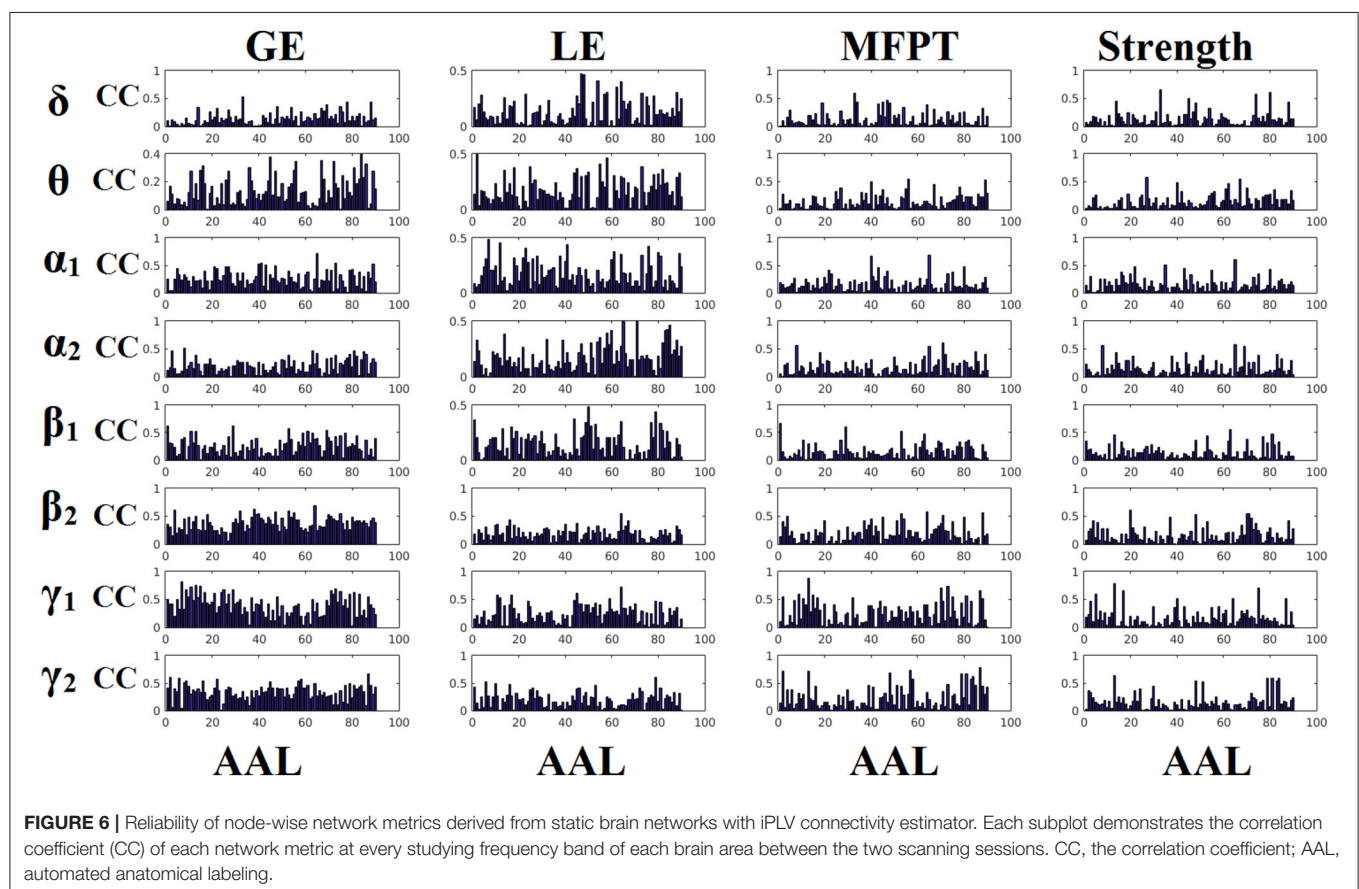
**Figure 3C** demonstrates the laplacian matrix of the FCG in **Figure 3B**. In the main diagonal, the degree of each node is tabulated. Afterward, we applied an eigenanalysis for each of these laplacian matrixes and the eigenvalues of this procedure describes the synchronizability of the original FCG. **Figure 3D** illustrates for the 1st min the eigenvalues for each quasi-static FCG. One can easily detect the abrupt transition between the brain states. Here, neural-gas algorithm was applied on the 2D matrix presented in **Figure 3D** after first concatenated across subjects independently for each frequency band and scan session. The main scope of this codebook learning algorithm is to define FC $\mu$ states.

By estimating the reconstruction error  $E$  between the original 2D matrices ( $90 \times [\text{slides} \times \text{subjects}]$ ) and the one reconstructed via the  $k$  FC $\mu$ states assigned to each snapshot of the DFCG for each predefined threshold, we can detect the optimal threshold  $T$  for each case. In this work, the criterion of the reconstruction error  $E$  was set less than 4%. Practically for all the frequency bands and in both connectivity estimators, the reconstruction error  $E$  was less than 2%. The selected threshold was detected based on the plateau by plotting of reconstruction error  $E$  vs. the threshold  $T$ .

In this way, the richness of information contained in the dynamic connectivity patterns is represented, by a partition matrix  $U$ , with elements  $u_{ij}$  indicating the assignment of input connectivity patterns to code vectors. Following the

inverse procedure, we can rebuild a given time series from the  $k$  FC $\mu$ states, with a small reconstruction error  $E$ . The selection of parameter  $k$  reflects the trade-off between fidelity and compression level. As a consequence, the symbolic time series closely follows the underlying functional connectivity dynamics. The derived symbolic times series that keep the information of network FC $\mu$ states (nFC $\mu$ states) are called hereafter as STS<sup>L-EIGEN</sup> (L:Laplacian – Eigen:Eigenanalysis). **Figure 3E** tabulates the correlation of the eigenvalues between every pair of temporal segments while in **Figure 3F**, the matrix in **Figure 3E** was reordered such as the FC $\mu$ states to be revealed via the neural-gas algorithm. The network topology of the extracted FC $\mu$ states is illustrated in **Figure 3G**. From, **Figure 3F**, one can understand that the two FC $\mu$ states describe the DFCG of this subject.

An exemplar of prototypical FC $\mu$ states is illustrated in **Figure 3G**. The outcome of this clustering procedure is also to extract a symbolic time series per subject, repeat scan and frequency that describes the transition of brain activity between the extracted brain states (FC $\mu$ states; **Figure 3D**). The transition probability  $P$  for this example and for the two FC $\mu$ states is illustrated with a classical figure for first order Markovian Chain. The self-arrows refer to the percentage of sliding windows where the brain stays stable in a FC $\mu$ state without any transition while the directed arrow gives the percentage of transition from one FC $\mu$ state to the other. This symbolic time series can be seen as



a first order Markovian chain where these switching between “quasi-static” FC $\mu$ states can be modeled as a finite Markov chain (Dimitriadis et al., 2013a,b, 2015a; O’Neill et al., 2015; Vidaurre et al., 2016). One can clearly understand that human brain demonstrates a preferred transition from FC $\mu$ states<sup>2</sup> to FC $\mu$ states<sup>1</sup> (off-diagonal lines of the TP) compared to the opposite direction (**Figure 3H**). The sketch of the markovian chain and the colored TP matrix can reveal the aforementioned trend of preferred direction FC $\mu$ states<sup>2</sup> to FC $\mu$ states<sup>1</sup>.

From the symbolic timeseries, specific metrics tailored to the dynamic evolution of FC $\mu$ states were estimated (see next section) and their reliability was assessed via the correlation coefficient between scan session 1 and scan session 2. The whole approach was repeated independently for each frequency band and connectivity estimator by integrating subject and scan-based DFCG.

## Characterization of Time-Varying Connectivity

Once the integrated DFCG is formed and it is modeled via the combination of neural-gas and laplacian eigenanalysis scheme [N-GAS<sup>L-EIGEN</sup> (L:Laplacian - Eigen:Eigenanalysis)], relevant features can be extracted from the data based on the state-transition states. There features are called chronnectomics (chronos—Greek word for time and connectomics for network metrics) which are described in the following section.

## Chronnectomics

The following chronnectomics (dynamic network metrics) will be estimated on the STS<sup>L-EIGEN</sup> which expresses the fluctuation of the FC $\mu$ states.

### State Transition Rate

Based on the state transition vectors STS<sup>L-EIGEN</sup> as demonstrated in **Figure 3A**, we estimated the transition rate (TR) for every pair of states as followed:

$$TR = \frac{\text{no of transitions}}{\text{slides} - 1} \quad (17)$$

where slides denote the number of temporal segments using the sliding window approach.

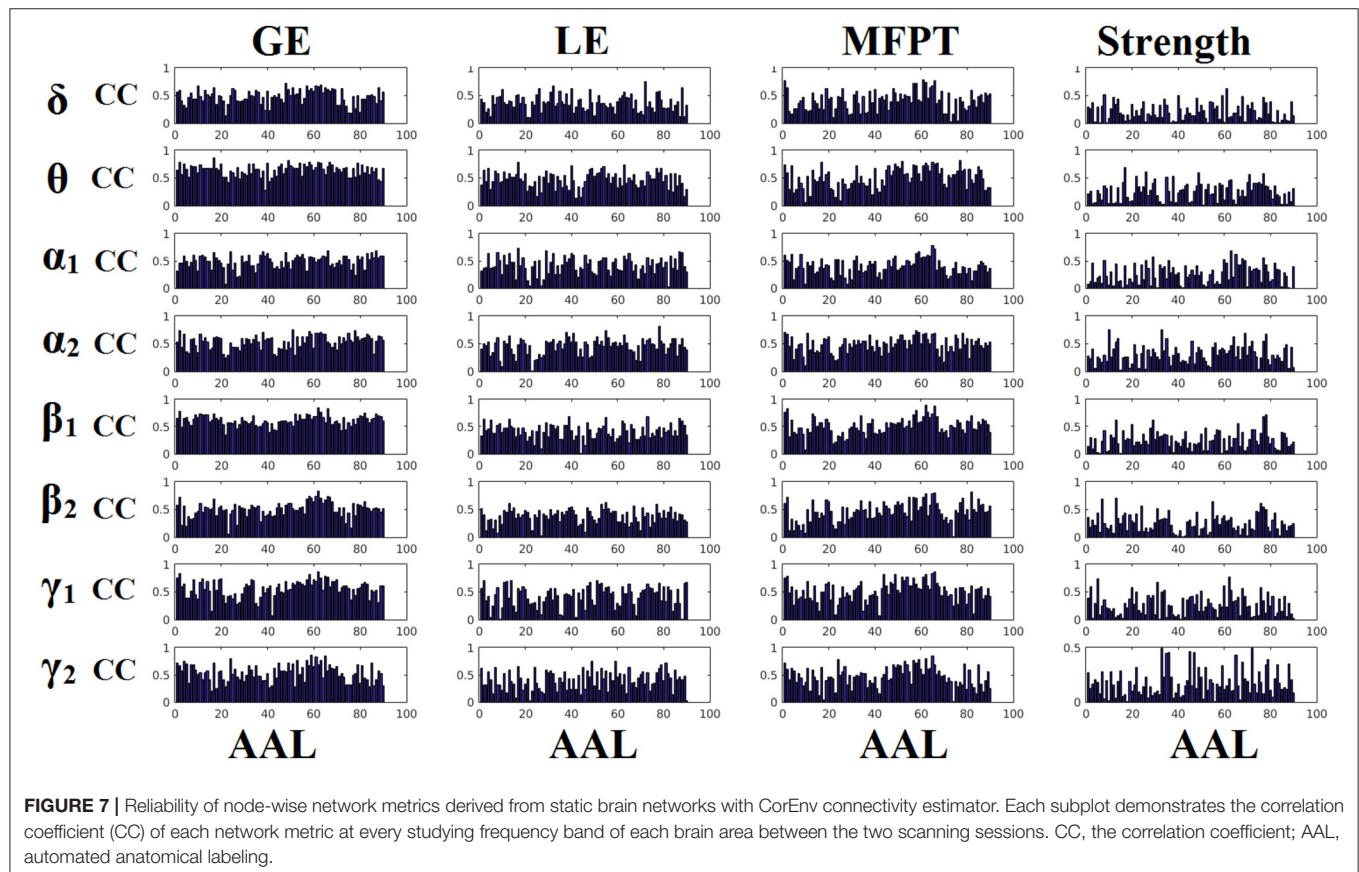
TR yields higher values for increased numbers of “jumps” of the brain between the derived brain states over consecutive time windows. This approach leads to one feature per participant.

### Occupancy Times of the nFC $\mu$ states

Complementary to the aforementioned chronnectomics, we estimated also the occupancy time (OC) of each FC $\mu$ states as the percentage of its occurrence across the experimental time. OC was estimated from STS<sup>L-EIGEN</sup> as follows:

$$OC(k) = \frac{\text{Frequency of Occurance}}{\text{slides}} \quad (18)$$

where  $k$  denotes the FC $\mu$ states.





## Reliability of Static Network Metrics and Chronnectomics

The reliability of static node-wise network metrics and the chronnectomics was assessed with the correlation coefficient between forty values derived from scan session 1 and forty values from scan session 2 for each frequency band, condition and connectivity estimator (see **Figures 1–3**).

## Optimizing the Width of the Time-Window and the Stepping Criterion

We optimized both the width of the time-window and the step criteria for the sliding-window approach based on the maximization of the reliability of TR. The reliability was estimated based on the correlation coefficient of the TR across the whole group between scan session 1 and 2. The whole procedure was followed independently for each brain rhythm. The settings for the width of temporal window and the step were defined as a percentage of the cycles of the studying frequencies: {from 1 up to 10 cycles with step equals to 0.5

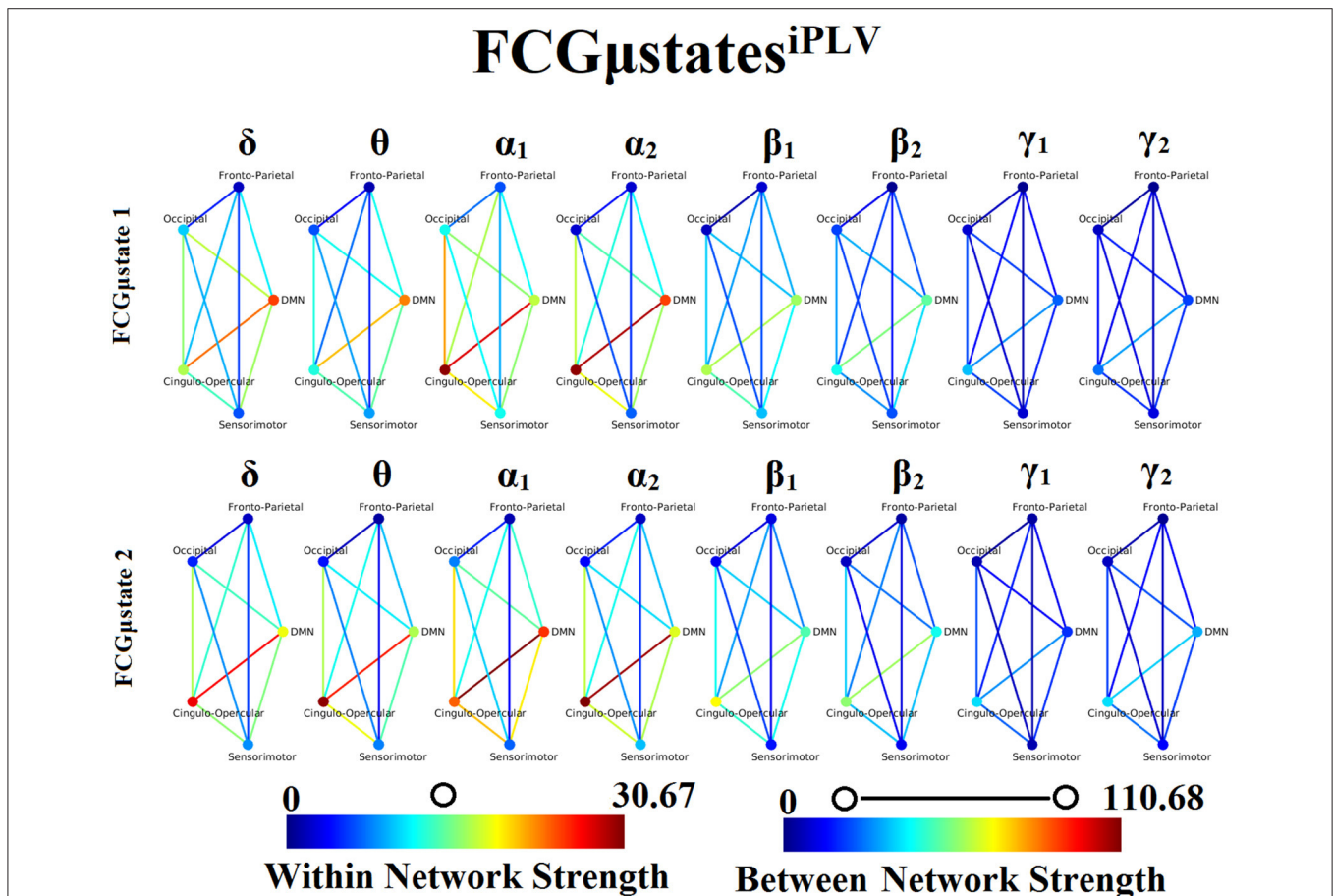
cycle} for the width of the temporal window and {from 0.1 cycles to 2 cycles with step equals to 0.1 cycle} for the step (see **Table 1**).

To avoid overfitting of both TR and OT since, we used TR for both the optimization of the width of the temporal window and the stepping criterion, we used the optimized parameters in an external second repeat scan dataset for further evaluation (see **Supplementary Material**).

## RESULTS

### Tuning Parameters for Dynamic Functional Connectivity Analysis

The optimization of the temporal window and the stepping criterion for each brain rhythm reveals a nice trend for dynamic functional connectivity analysis. The width of the temporal window increased from  $\delta$  to  $\gamma_2$  while the stepping criterion decreased in both connectivity estimators (see **Table 1**).



**FIGURE 8 |** Frequency-dependent  $FCG\mu states^{iPLV}$ . Network topologies of the  $FCG\mu states^{iPLV}$  for each of the studying frequency band. To enhance the visualization and contrast of  $FCG\mu states$  across frequency bands, we adopted a network-wise representation instead of plotting the brain network in a 90 nodes layout. The 90 ROIs of the AAL template were assigned to each of the five networks: default-mode (DMN), fronto-parietal (FPN), occipital (O), sensorimotor (SM), and cingulo-opercular (CO). The color of each node denotes total strength of within network connections while the color of each line the total strength of between network connections. Both strength values were normalized across both frequencies and  $FCG\mu states$ .



## Common Projection Space of Frequency-Dependent Static FCG

To demonstrate the (dis)similarities between sessions and subjects of the frequency-dependent static FCG, we constructed a distance matrix of dimensions  $80 \times 80$  (subjects  $\times$  sessions) using the graph diffusion distance metric. Then, we applied multidimensional scaling (MDS) to project the distance matrix into a common 2D feature space. Using a different colored circle for each scanning session (blue for session 1 and red for session 2) and connecting both of them with a black line for each subject, we further enhanced the (dis)similarities of the static FCGs. **Figures 4, 5** illustrate these FCG-based projections for static FCG<sup>iPLV</sup> and FCG<sup>CorEnv</sup> correspondingly. In **Figure 4G** one can detect a few subjects with high reliable static FCG between the two scan sessions and also subject-specific network topologies that occupied an isolated subarea in the common projection FCG space. The stress index estimated via the MDS approach was low and the relationship of the 80 FCGs in the original  $80 \times 80$  matrix is preserved in the projected 2D space.

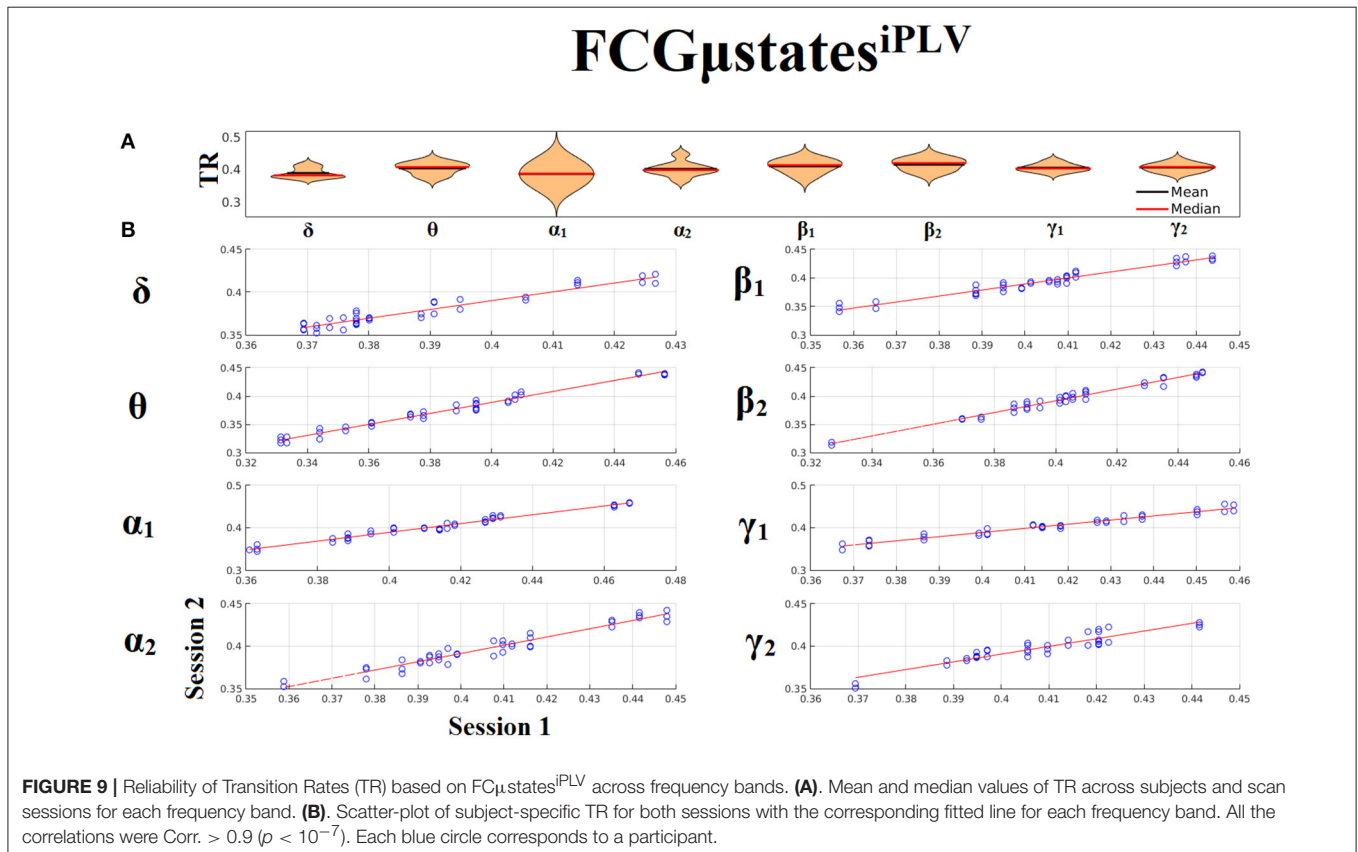
## Reliability of Static Network Metrics

**Figures 6, 7** demonstrate the correlation coefficients for each node-wise network metric between the two scanning sessions for every frequency-dependent static FCG. From the visual comparison of both figures one can clearly reveal that the correlation values are higher for CorEnv compared to iPLV.

Applying Wilcoxon Rank-Sum Test for every frequency and network metric between the 90 correlation values, we detected statistical significant differences in every case ( $p < 0.01$ ,  $p' < p/32$ , Bonferroni Corrected). However, the averaged correlation values did not reach high reliability (e.g.,  $>0.9$ ) even for the CorEnv. It is obvious from the correlation plots that the reliability of node-wise static network metrics has high spatial variability in both connectivity estimators.

## Frequency-Dependent FCG $\mu$ states and Reliability Chronnectomics for iPLV

Our analysis of DFCG based on iPLV revealed two FCG $\mu$ states<sup>iPLV</sup> for each frequency band. The topology of these frequency-dependent FCG $\mu$ states<sup>iPLV</sup> is illustrated in **Figure 8**. We integrated the nodes into five well-known brain networks: default-mode (DMN), fronto-parietal (FPN), occipital (O), sensorimotor (SM), and cingulo-opercular (CO). The mapping between the 90 ROIs of AAL and the five brain networks can be retrieved from section Results in **Supplementary Material**. One can clearly detect that the functional coupling between the default mode network and the cingulo-opercular dominates the coupling strength across the frequency bands and FCG $\mu$ states with less pronounced effect in both  $\gamma$  bands. Complementary, the coupling strength between and within the networks is diminished after  $\alpha_2$  frequency. This behavior can be interpreted as a reduction of the connections



up to the defined threshold following the increment of the frequency. The two  $\text{FCG}\mu\text{states}^{\text{iPLV}}$  showed also a different distribution of strength globally and locally.

Both types of chronnectomics, transition rates (TR) (Figure 9) and occupancy times (OC) (Figure 10) demonstrated high reliability ( $\text{Corr.} > 0.9, p < 10^{-7}$ ) across frequency bands. Similar results, we obtained also for the second external dataset (see section 2 in Supplementary Material).

## Frequency-Dependent $\text{FCG}\mu\text{states}$ and Reliability Chronnectomics for CorEnv

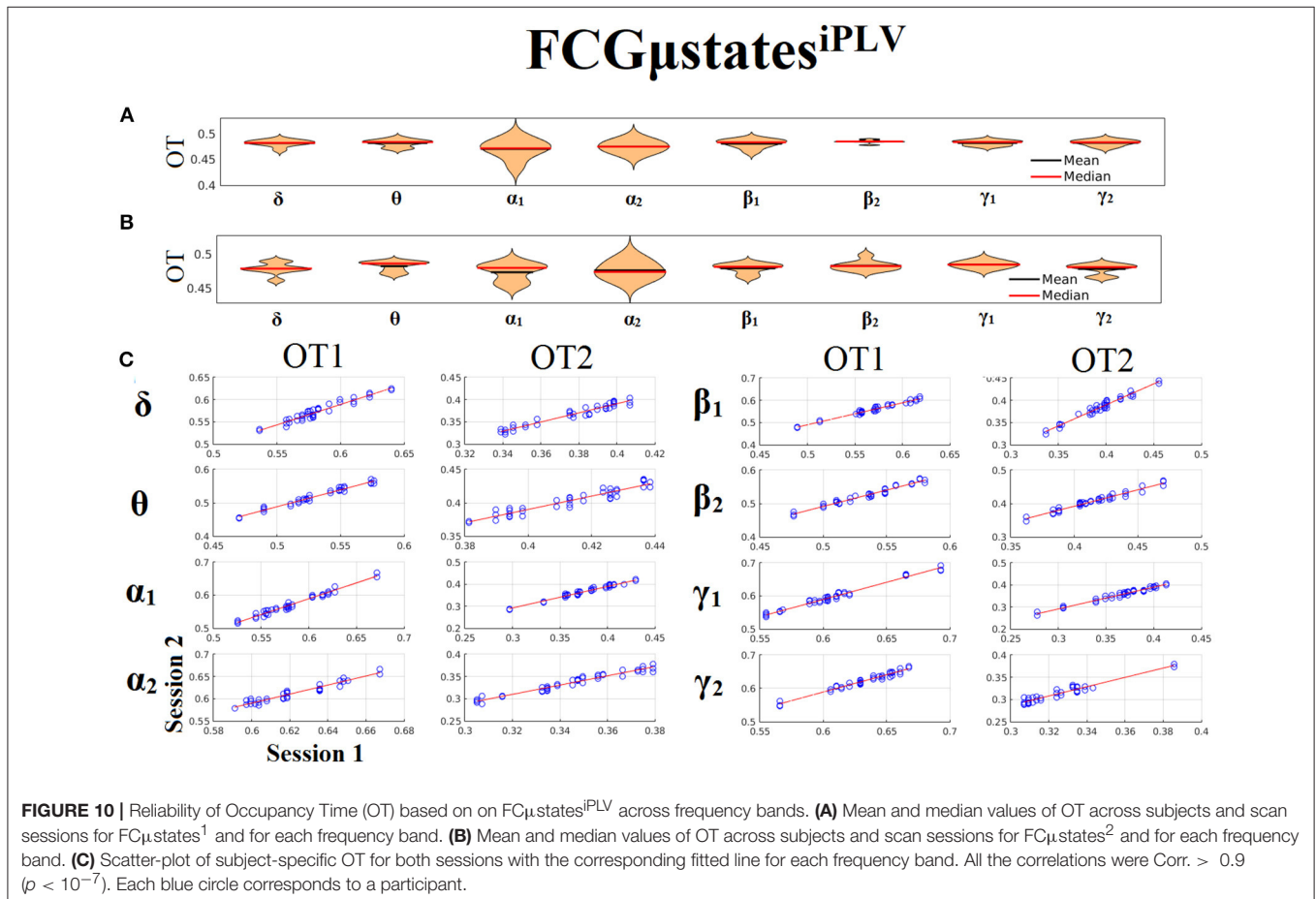
Our analysis of DFCG based on the correlation of the envelope connectivity estimators revealed two  $\text{FCG}\mu\text{states}^{\text{CorEnv}}$  for each frequency band. The topology of these frequency-dependent  $\text{FCG}\mu\text{states}^{\text{CorEnv}}$  is illustrated in Figure 11. The mapping between the 90 ROIs of AAL and the five brain networks can be retrieved from section 2 in Supplementary Material. One can clearly detect that the functional coupling between the default mode network and the cingulo-opercular dominates the coupling strength across the frequency bands and  $\text{FCG}\mu\text{states}$ . Complementary, the coupling strength between and within the networks is diminished after  $\alpha_2$  frequency as it was observed for  $\text{FCG}\mu\text{states}^{\text{iPLV}}$ . Complementary, the network topologies of  $\text{FCG}\mu\text{states}^{\text{CorEnv}}$  between low and high frequencies based on the strength coupling are more common

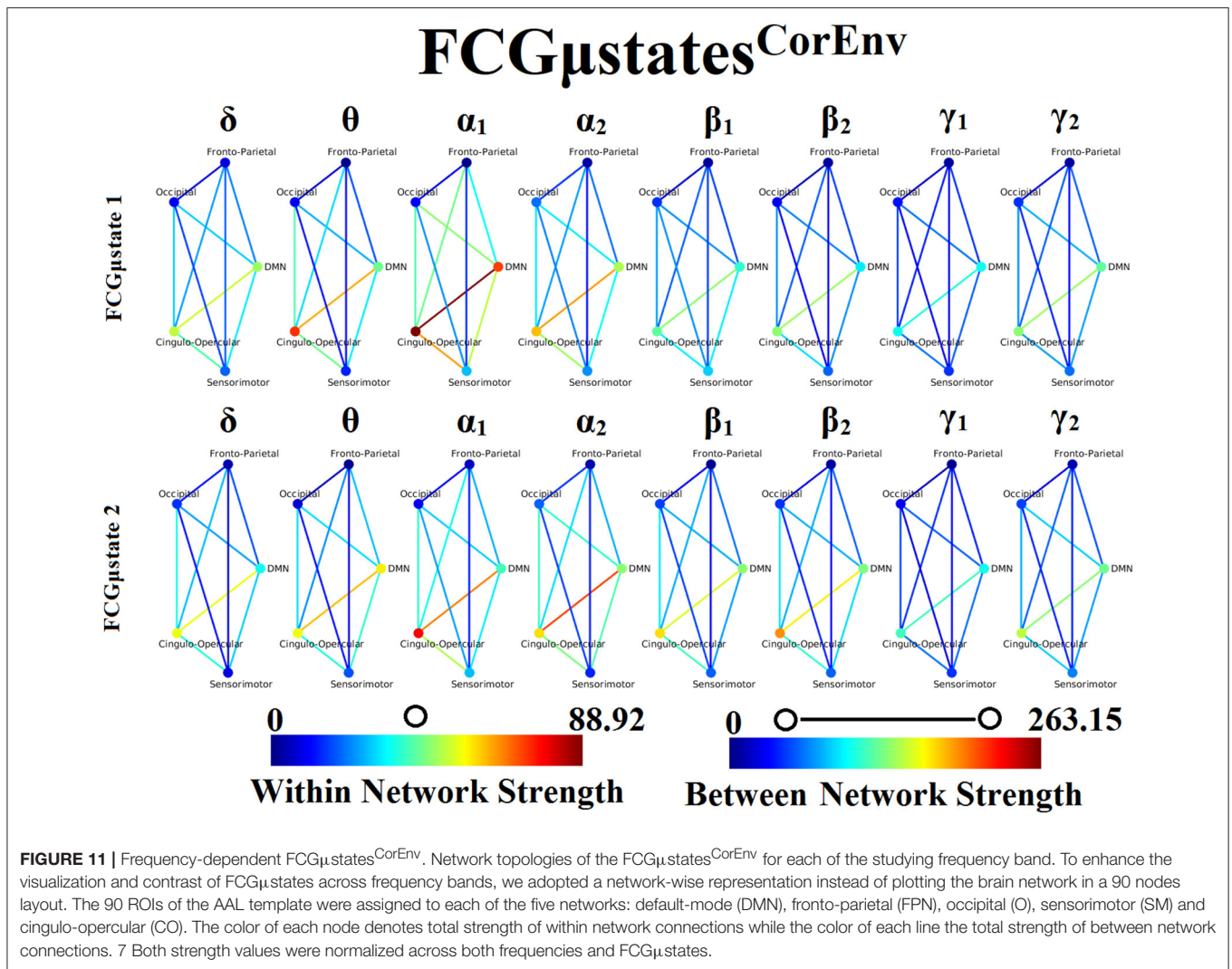
than the  $\text{FCG}\mu\text{states}^{\text{iPLV}}$ . This common substrate across the  $\text{FCG}\mu\text{states}^{\text{CorEnv}}$  is consistent with the general notion that correlation of the envelope is more stacked to the structural connectome compared to the phase-based connectivity patterns which demonstrate higher degrees of freedom (Engel et al., 2013; compare Figure 8 with Figure 11).

Only transition rates (TR) showed high reliability for CorEnv ( $\text{Corr.} > 0.8, p < 10^{-4}$ ) in all the frequency bands with the only exception of  $\beta_1$  (Figure 12). Occupancy times (OT) showed low reliability across the frequency bands ( $p > 0.4$ ) (Figure 13). TR of  $\text{FCG}\mu\text{states}^{\text{CorEnv}}$  increased with the increment of frequency reaching a plateau in  $\gamma_1$ . In contrast, TR of  $\text{FCG}\mu\text{states}^{\text{iPLV}}$  did not show such a frequency-dependent behavior. Similar results, we obtained also for the second external dataset (see section 2 in Supplementary Material).

## DISCUSSION

In the present study, we assessed the reliability of both static and dynamic functional connectivity network descriptors using resting-state MEG data from 40 subjects with repeat scan sessions. This is the first time that the reliability of chronnectomics, at least for the MEG modality, has been taken into account. Source time series were first beamformed



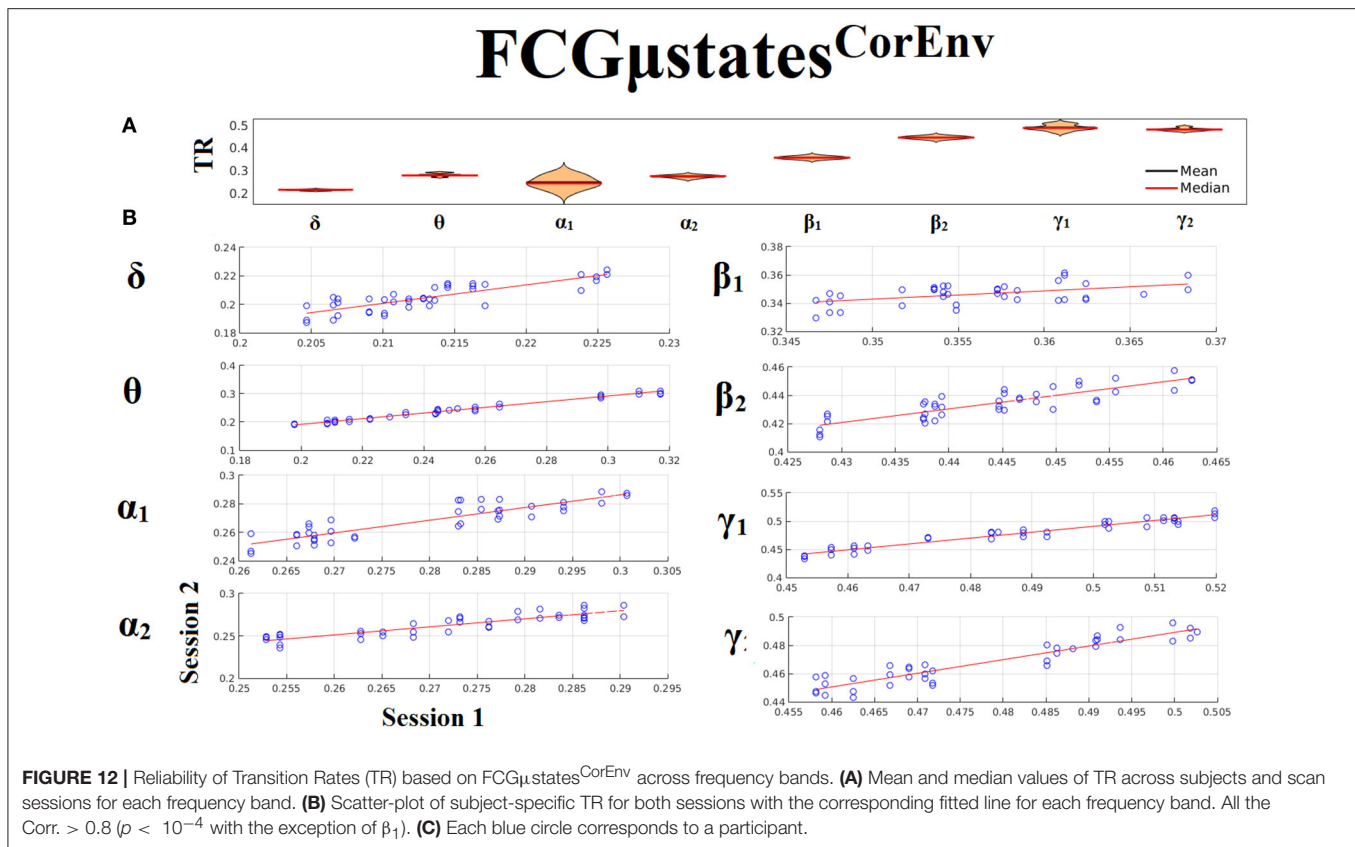


independently for each frequency band (Hillebrand et al., 2005; Schoffelen and Gross, 2009; Brookes et al., 2011b), and then representative voxel time series based on the AAL atlas were extracted using a novel linear interpolation analysis. This procedure produces informative timeseries with a characteristic carrier frequency compared to the noisy time series derived by PCA or by selecting the voxel time series within a ROI that encapsulates the maximum power. Then, both static and dynamic frequency-dependent functional connectivity graphs were computed for each subject and scan session using the imaginary part of phase locking value (iPLV) and the correlation of the amplitude envelope (CorEnv). Both static and dynamic FCG (SFCG-DFCG) were filtered both statistically (surrogates) and topologically (OMST; Dimitriadis et al., 2017a,b).

Here, we adopted a data-driven pipeline of how to estimate both static and dynamic FCG statistically and topologically filtered using an algorithm previously applied to EEG recordings. We explored the reliability of both static network metrics and chronnectomics (dynamic network metrics) by employing two representative connectivity estimators for

the construction of static and dynamic brain networks. Using this pipeline, prototypical FCGμstates were derived which were highly reproducible across subjects and scan sessions in both connectivity estimators and in all frequencies. The reliability of node-wise static network metrics based on four network metrics was low and spatially variable with both connectivity estimators while the CorEnv demonstrates higher ICC values compared to iPLV. The reliability of chronnectomics (TR, OT) for iPLV was high while for CorEnv the reliability of only the TR reaches high acceptable levels. Our results were reproduced also in a second external dataset (see **Supplementary Material**). Our study strongly encouraging the use of DFCG with neuromagnetic recordings that takes the advantage of the nature of MEG modality, its high temporal resolution.

To our knowledge, this is the very first study that explored the reliability of both static and dynamic FCG and the related network metrics and chronnectomics, respectively in neuromagnetic source space at. In static FCG, node-wise network metrics demonstrated poor reliability for iPLV and poor to medium for CorEnv. The node-wise reliability was highly spatial



variable and static FCG have also demonstrated low repeatability in both connectivity estimators and especially in CorEnv. In contrast, prototypical FCG $\mu$ states were highly reproducible across subjects and scan sessions in both connectivity estimators and in all frequencies supporting by the low reconstruction error (<2%) of our brain network learning algorithm. Complementary, the reliability of chronnectomics (TR,OT) for iPLV was high while for CorEnv the reliability of only the TR reaches high acceptable levels. These results strongly encourage the neuroscientists to adopt the notion of DFCG with neuromagnetic recordings that takes the advantage of its high temporal resolution.

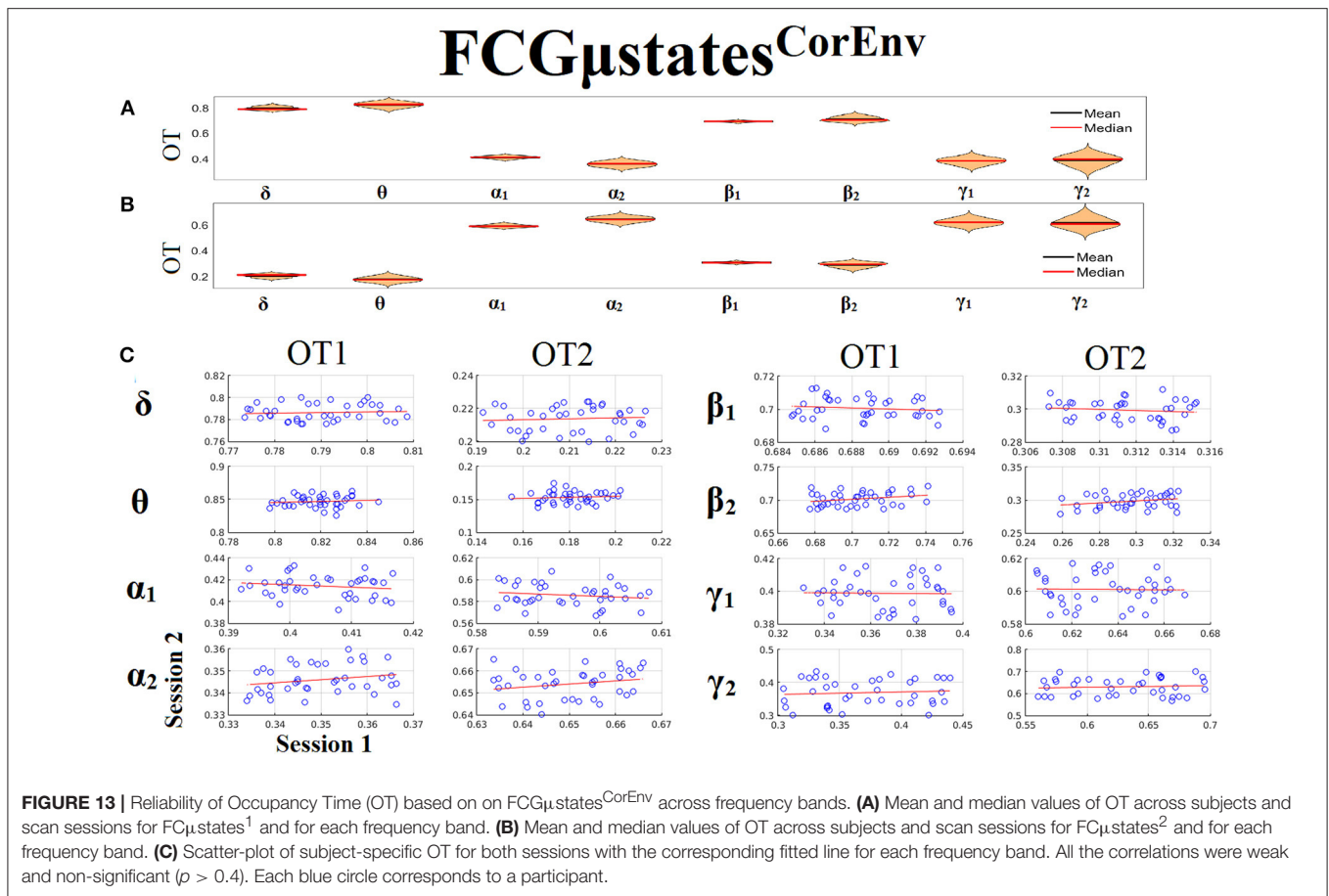
Two main studies explored the reliability of static FCG on the source level using MEG-beamformed resting-state connectivity analysis. Garcés et al. (2016) studied the reliability of resting-state networks using four connectivity estimators: phase-locking value (PLV), phase lag index (PLI), direct envelope correlation (d-ecor), and envelope correlation with leakage correction (lc-ecor). They adopted intra-class correlation coefficient (ICC) and Kendall's W for assessing within and between-subjects agreement respectively. Higher test-retest reliability was found for PLV from  $\theta$  to  $\gamma$ , and for lc-ecor and d-ecor in  $\beta$ . They commented that high ICC in PLV and d-ecor could be artifactual due to volume conduction effects. Colclough et al. (2016) investigated the reliability of static FCG at resting-state using beamformed source static connectivity analysis. They reported high reliability mostly for the partial correlation analysis and the correlation of the envelope among 12 connectivity estimators. Two more

studies, Deuker et al. (2009) estimated the reliability of resting-state network metrics derived from MEG in sensor space using mutual information. They obtained high ICC for clustering, global efficiency and strength at a network level. Jin et al. (2011) found medium ICC for nodal global efficiency, nodal degree and betweenness centrality in  $\alpha$  and  $\beta$  bands.

Our results revealed that nodal network metrics derived from static FCG are less reproducible than their dynamic counterparts. In contrast, chronnectomics are highly reproducible with both adopted connectivity estimators. These results complemented with the results presented in (Colclough et al., 2016) where they adopted multiple connectivity estimators for the construction of static brain networks on the source level using MEG-beamformed resting-state activity. Colclough et al. (2016) showed that the static full-weighted FCG are highly repeatable within the group-level mostly for the correlation of the envelope adopting a split-half strategy on a dataset with single scans. Here, we assessed the reliability of any network metric using a two scan session design per subject. We should state here that edge-weights are significant for the construction of network topology and the reliability of connectomic biomarkers (Dimitriadis et al., 2018).

One of the key findings of our analysis are the frequency-dependent FCG $\mu$ states for each connectivity estimator. Figures 8, 11 illustrate the strength of the coupling within and between brain networks for the prototypical FCG $\mu$ states at every frequency band. It is obvious that the highest strength within a network is observed within the DMN in





both connectivity estimators. The strength between the brain networks is mainly distributed between DMN and the rest of the networks demonstrating the highest value till  $\alpha_2$  and dropped from  $\beta_1$  to  $\gamma_2$  (Figures 8, 11). DMN reignited a high interest the last years for the description of intrinsic ongoing activity in studies of the human brain in health and disease (Raichle, 2015). Disruptions of functional connections within the DMN and between DMN and the rest of brain networks have been linked to various neurological and neuropsychiatric disorders (Mohan et al., 2016). Studies in healthy aging and Alzheimer's disease have revealed the significant role of DMN (Mevel et al., 2011).

Flexible hub theory based on clustering analysis of functional networks gave an explanation of how temporal functional modes exist where one neural region may be switched from a certain network at one time to a different network at another time (Smith et al., 2012). It remains still unclear how the different brain networks are connected together during spontaneous and task-related activity. Dosenbach et al. (2008) proposed that the FPN may serve to initiate and adjust cognitive control, whereas another control-type network, the CO network (CON), provides stable set-maintenance. Cole and colleagues (Cole et al., 2013) helped to untangle the flexible role of the FPN, many questions remain regarding the interaction between the FPN and the CON and also with

other networks such as the DMN, SM and O. In the present study, we characterized the dynamic relationships of the brain networks across time at resting-state in various frequency bands and using representative connectivity estimators. We found that these functional patterns are high reproducible which will help multi groups worldwide to explore these interactions and build reproducible connectomic biomarkers in various diseases and disorders. Understanding the neural basis of intrinsic activity, cognition and structure–function relationships, will further enhance the prognostic/diagnostic abilities in clinical populations.

The interactions of large-scale brain networks at resting-state and during tasks is characterized by the studying frequency. Frequency-specific functional interactions between large-scale brain networks may be individual fingerprints of the idle activity and cognition (Siegel et al., 2012). It will be interesting in the future to explore how the  $FCG\mu states$  from a dynamic integrated functional connectivity graph (Dimitriadis and Salis, 2017c) that incorporates different intrinsic coupling modes both intra and cross-frequency coupling can be used for brain fingerprinting (Engel et al., 2013).

It is critically important to take advantage of new imaging modalities to untangle the mechanisms that produce circuit dysfunctions in many brain diseases and disorders. The development of biomarkers is very important and for that reason

the proposed experimental paradigm and analytics of the meta-data derived from the analysis of human brain activity must be highly reliable and reproducible. Magnetoencephalography (MEG) allows us to measure neuronal events noninvasively with millisecond resolution and recent advanced methods opens new avenues of exploring and answering fundamental key research questions tailored to each brain disease/disorder. MEG can become a pioneering clinical research tool for mental disorders (Bowyer et al., 2015; Grent-T-Jong et al., 2016; Uhlhaas et al., 2017), Alzheimer's disease (López et al., 2014, 2017; Koelewijn et al., 2017), dyslexia (Dimitriadis et al., 2010b, 2013b), traumatic brain injury (Dimitriadis et al., 2015c; Antonakakis et al., 2016, 2017), multiple sclerosis (Tewarie et al., 2015), and other brain diseases. To establish MEG-based biomarkers that can be used for daily clinical practice and clinical evaluation, their reproducibility should be further explored. Complementary, the transition rate and also the occupancy times could be personalized biomarkers of a subject's resting-state condition where more task-related FC $\mu$ states and the related markers derived from them could build a subject specific database for longitudinal studies. Transition rates could be also correlated with IQ scores and also with behavioral performance during execution of cognitive tasks.

In the present study, we proposed a data-driven analytic pathway to assess the reliability of connectomics using MEG-beamformed connectivity analysis. Our results clearly support the notion of dynamic functional connectivity on the source level, the derived prototypical FC $\mu$ states and the related chronnectomics. Last years, many studies explored the dynamic functional connectivity graphs in many modalities (EEG/MEG/fMRI) and in both resting-state and during tasks (Dimitriadis et al., 2010a, 2012a,c, 2013a,b, 2015a,b,c,d, 2016a,b; Bassett et al., 2011; Rosazza and Minati, 2011; Allen et al., 2012; Handwerker et al., 2012; Ioannides et al., 2012; Hutchison et al., 2013; Liu and Duyn, 2013; Braun et al., 2015b; Mylonas et al., 2015; Toppi et al., 2015; Yang and Lin, 2015; Calhoun and Adali, 2016, for reviews see Calhoun et al., 2014). This is the very first study according to authors' knowledge that the reliability of chronnectomics was explored. The outcome of this study opens new avenues in the exploration of human brain dynamics with MEG-beamformed source activity and under the notion of dynamic functional connectivity.

We addressed the key question of the readiness of neuromagnetic-based based functional connectomics to lead to clinically meaningful biomarker identification through the reliability approach that offers a repeat scan study in healthy controls. It is more than significant to customize stable approaches for analyzing neuromagnetic recordings and present reproducible brain connectomics across scans in healthy control populations without sacrificing the individual characteristics that can be used for personalized intervention neuroscience (Gratton et al., 2018). It is highly recommend to access the reliability of any metric derived from any neuroimaging

modality in a repeat scan protocol in healthy control population before applying it to a larger disease group where the cost of scanning is too high (diffusion MRI: Dimitriadis et al., 2017d). Additionally, we will expand this analysis in future efforts to identify disease status alone including clinical variables related to genetic risk (Lancaster et al., 2018), expected treatment response and prognosis.

## CONCLUSIONS

In conclusion, we provided the first source-space test-retest reliability of dynamic functional connectivity of neuromagnetic recordings at resting-state. We computed both static and dynamic functional connectivity based on 90 ROIs according to AAL templated and using two connectivity estimators, the iPLV and the CorEnv. Nodal network metrics were unreliable in both connectivity estimators but with higher reliability demonstrated for CorEnv. Moreover, their reliability demonstrates highly spatial variability. Static FCG were also unreliable and especially for CorEnv. In contrast, prototypical FC $\mu$ states were reliable in both connectivity estimators and across frequency bands. The derived chronnectomics (TR, OT) were highly reproducible for iPLV while only TR was reliable for CorEnv within acceptable levels. Our results strongly encourages future studies with main scope to explore resting-state networks in both healthy control and disease populations to apply a data-driven dynamic functional connectivity analysis using MEG-beamformed source reconstructed brain activity.

## AUTHOR CONTRIBUTIONS

SD: conception of the research analysis, methods and design, data analysis, and drafting the manuscript; BR: data acquisition; BR, DL, and KS: critical revision of the manuscript. Every author read and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

SD and DL were supported by MRC grant MR/K004360/1 (Behavioral and Neurophysiological Effects of Schizophrenia Risk Genes: A Multi-locus, Pathway Based Approach). SD is also supported by a MARIE-CURIE COFUND EU-UK Research Fellowship. BR and the CUBRIC MEG lab are supported by an MRC UK MEG Partnership Grant, MR/K005464/1 and an MRC Doctoral Training Grant, MR/K501086/1. We would like to acknowledge RCUK of Cardiff University and Wellcome Trust for covering the publication fee.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00506/full#supplementary-material>

## REFERENCES

- Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3:e17. doi: 10.1371/journal.pcbi.0030017
- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2012). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Antonakakis, M., Dimitriadis, S. I., Zervakis, M., Micheloyannis, S., Rezaie, R., Babajani-Feremi, A., et al. (2016). Altered cross-frequency coupling in resting-state MEG after mild traumatic brain injury. *Int. J. Psychophysiol.* 102, 1–11. doi: 10.1016/j.ijpsycho.2016.02.002
- Antonakakis, M., Dimitriadis, S. I., Zervakis, M., Papanicolaou, A. C., and Zouridakis, G. (2017). Reconfiguration of dominant coupling modes in mild traumatic brain injury mediated by  $\delta$ -band activity: a resting state MEG study section. *Neuroscience* 356, 275–286. doi: 10.1016/j.neuroscience.2017.05.032
- Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., Singer, W., et al. (2015). Untangling cross-frequency coupling in neuroscience. *Curr. Opin. Neurobiol.* 31, 51–61. doi: 10.1016/j.conb.2014.08.002
- Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J., et al. (2014). Fast transient networks in spontaneous human brain activity. *Elife (Camb.)* 3:e01867. doi: 10.7554/eLife.01867
- Baldassarre, A., Lewis, C. M., Committeri, G., Snyder, A. Z., Romani, G. L., and Corbetta, M. (2012). Individual variability in functional connectivity predicts performance of a perceptual task. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3516–3521. doi: 10.1073/pnas.1113148109
- Bandettini, P. A. (2012). Twenty years of functional MRI: the science and the stories. *Neuroimage* 62, 575–588. doi: 10.1016/j.neuroimage.2012.04.026
- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., and Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7641–7646. doi: 10.1073/pnas.1018985108
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.* 57, 289–300.
- Biswal, B. B. (2011). Resting state functional connectivity. *Biol. Psychiatry* 69, 200S. doi: 10.1016/j.biopsych.2011.03.032
- Biswal, B. B. (2012). Resting state fMRI: a personal history. *Neuroimage* 62, 938–944. doi: 10.1016/j.neuroimage.2012.01.090
- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541.
- Bowyer, S., Gjini, K., Zhu, X., Kim, L., Moran, J., Rizvi, S., et al. (2015). Potential biomarkers of schizophrenia from MEG resting-state functional connectivity networks: preliminary data. *J. Behav. Brain Sci.* 5, 1–11. doi: 10.4236/jbbs.2015.51001
- Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., et al. (2015b). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11678–11683. doi: 10.1073/pnas.1422487112
- Brookes, M. J., Hale, J. R., Zumer, J. M., Stevenson, C. M., Francis, S. T., Barnes, G. R., et al. (2011a). Measuring functional connectivity using MEG: methodology and comparison with fMRI. *Neuroimage* 56, 1082–1104. doi: 10.1016/j.neuroimage.2011.02.054
- Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., et al. (2011b). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16783–16788. doi: 10.1073/pnas.1112685108
- Brookes, M. J., Woolrich, M. W., Barnes, G. R. (2012). Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage. *Neuroimage* 63, 910–920. doi: 10.1016/j.neuroimage.2012.03.048
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Calhoun, V. D., and Adali, T. (2016). Time-varying brain connectivity in fMRI data: whole-brain data-driven approaches for capturing and characterizing dynamic states. *IEEE Signal Process. Mag.* 33, 52–66. doi: 10.1109/MSP.2015.2478915
- Calhoun, V. D., Miller, R., Pearlson, G., and Adali, T. (2014). The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015
- Colclough, G. L., Brookes, M. J., Smith, S. M., and Woolrich, M. W. (2015). A symmetric multivariate leakage correction for MEG connectomes. *Neuroimage* 117, 439–448. doi: 10.1016/j.neuroimage.2015.03.071
- Colclough, G. L., Woolrich, M. W., Tewarie, P. K., Brookes, M. J., Quinn, A. J., and Smith, S. M. (2016). How reliable are MEG resting-state connectivity metrics? *Neuroimage* 138, 284–293. doi: 10.1016/j.neuroimage.2016.05.070
- Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., and Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355. doi: 10.1038/nn.3470
- Deco, G., Jirsa, V. K., and McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat. Rev. Neurosci.* 12, 43–56. doi: 10.1038/nrn2961
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*, 134, 9–21.
- de Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., et al. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6040–6045. doi: 10.1073/pnas.0913863107
- Deuker, L., Bullmore, E. T., Smith, M., Christensen, S., Nathan, P. J., Rockstroh, B., et al. (2009). Reproducibility of graph metrics of human brain functional networks. *Neuroimage* 47, 1460–1468. doi: 10.1016/j.neuroimage.2009.05.035
- Dimitriadis, S. I., Antonakakis, M., Simos, P., Fletcher, J. M., and Papanicolaou, A. (2017b). Data-driven topological filtering based on orthogonal minimal spanning trees: application to multi-group MEG resting-state connectivity. *Brain Connect.* 7, 661–670. doi: 10.1089/brain.2017.0512
- Dimitriadis, S. I., Drakesmith, M., Bells, S., Parker, G. D., Linden, D. E., and Jones, D. K. (2017d). Improving the reliability of network metrics in structural brain networks by integrating different network weighting strategies into a single graph. *Front. Neurosci.* 11:694. doi: 10.3389/fnins.2017.00694
- Dimitriadis, S. I., Kanatsoulis, K., Laskaris, N. A., Tsirka, V., Vourkas, M., and Micheloyannis, S. (2012b). Surface EEG shows that functional segregation via phase coupling contributes to the neural substrate of mental calculations. *Brain Cogn.* 80, 45–52. doi: 10.1016/j.bandc.2012.04.001
- Dimitriadis, S. I., Laskaris, N. A., Bitzidou, M. P., Tarnanas, I., and Tsolaki, M. N. (2015b). A novel biomarker of amnesic MCI based on dynamic cross-frequency coupling patterns during cognitive brain responses. *Front. Neurosci.* 9:350. doi: 10.3389/fnins.2015.00350
- Dimitriadis, S. I., Laskaris, N. A., Del Rio-Portilla, Y., and Koudounis, G. C. (2009). Characterizing dynamic functional connectivity across sleep stages from EEG. *Brain Topogr.* 22, 119–133. doi: 10.1007/s10548-008-0071-4
- Dimitriadis, S. I., Laskaris, N. A., and Micheloyannis, S. (2015a). Transition dynamics of EEG-based Network Microstates unmask developmental and task differences during mental arithmetic and resting wakefulness. *Cogn. Neurodyn.* 9, 371–387. doi: 10.1007/s11571-015-9330-8
- Dimitriadis, S. I., Laskaris, N. A., Simos, P. G., Fletcher, J. M., and Papanicolaou, A. C. (2016b). Greater repertoire and temporal variability of cross-frequency coupling (CFC) modes in resting-state neuromagnetic recordings among children with reading difficulties. *Front. Hum. Neurosci.* 10:163. doi: 10.3389/fnhum.2016.00163
- Dimitriadis, S. I., Laskaris, N. A., Simos, P. G., Micheloyannis, S., Fletcher, J. M., Rezaie, R., et al. (2013b). Altered temporal correlations in resting-state connectivity fluctuations in children with reading difficulties detected via MEG. *Neuroimage* 83, 307–317. doi: 10.1016/j.neuroimage.2013.06.036
- Dimitriadis, S. I., Laskaris, N. A., Tsirka, V., Vourkas, M., and Micheloyannis, S. (2010b). What does delta band tell us about cognitive processes: a mental calculation study. *Neurosci. Lett.* 483, 11–15. doi: 10.1016/j.neulet.2010.07.034
- Dimitriadis, S. I., Laskaris, N. A., Tsirka, V., Vourkas, M., and Micheloyannis, S. (2012a). An EEG study of brain connectivity dynamics at the resting state. *Nonlinear Dyn. Psychol. Life Sci.* 16, 5–22.
- Dimitriadis, S. I., Laskaris, N. A., Tsirka, V., Vourkas, M., Micheloyannis, S., and Fotopoulos, S. (2010a). Tracking brain dynamics via time-dependent network analysis. *J. Neurosci. Methods* 193, 145–155. doi: 10.1016/j.jneumeth.2010.08.027



- Dimitriadis, S. I., Laskaris, N. A., and Tzelepi, A. (2013a). On the quantization of time-varying phase synchrony patterns into distinct functional connectivity microstates (FCμstates) in a multi-trial visual ERP paradigm. *Brain Topogr.* 3, 397–409. doi: 10.1007/s10548-013-0276-z
- Dimitriadis, S. I., Laskaris, N. A., Tzelepi, A., and Economou, G. (2012c). Analyzing functional brain connectivity by means of commute times: a new approach and its application to track event-related dynamics. *IEEE (TBE). Trans. Biomed. Eng.* 59, 1302–1309. doi: 10.1109/TBME.2012.2186568
- Dimitriadis, S. I., López, M. E., Bruña, R., Cuesta, P., Marcos, A., Maestú, F., et al. (2018). How to build a functional connectomic biomarker for mild cognitive impairment from source reconstructed MEG resting-state activity: the combination of ROI representation and connectivity estimator matters. *Front. Neurosci.* 12:306. doi: 10.3389/fnins.2018.00306
- Dimitriadis, S. I., and Salis, C. I. (2017c). Mining time-resolved functional brain graphs to an EEG-based chronnectomic brain aged index (CBAI). *Front. Hum. Neurosci.* 11:423. doi: 10.3389/fnhum.2017.00423
- Dimitriadis, S. I., Salis, C., Tarnanas, I., and Linden, D. (2017a). Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel data-driven thresholding scheme based on orthogonal minimal spanning trees (OMSTs). *Front. Neuroinform.* 11:28. doi: 10.3389/fninf.2017.00028
- Dimitriadis, S. I., Yu, S., Kwok, K., Laskaris, N. A., Thakor, N., and Bezerianos, A. (2015d). Cognitive workload assessment based on the tensorial treatment of EEG estimates of cross-frequency phase interactions. *Ann. Biomed. Eng.* 43, 977–989. doi: 10.1007/s10439-014-1143-0
- Dimitriadis, S. I., Zouridakis, G., Rezaie, R., Babajani-Feremi, A., and Papanicolaou, A. C. (2015c). Functional connectivity changes detected with magnetoencephalography after mild traumatic brain injury. *Neuroimage* 9, 519–531. doi: 10.1016/j.nicl.2015.09.011
- Dimitriadis, S., Sun, Y., Laskaris, N., Thakor, N., and Bezerianos, A. (2016a). Revealing cross-frequency causal interactions during a mental arithmetic task through symbolic transfer entropy: a novel vector-quantization approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 10, 1017–1028. doi: 10.1109/TNSRE.2016.2516107
- Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., and Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn. Sci.* 12, 99–105. doi: 10.1016/j.tics.2008.01.001
- Dunkley, B. T., Da Costa, L., Bethune, A., Jetly, R., Pang, E. W., Taylor, M. J., et al. (2015). Low-frequency connectivity is associated with mild traumatic brain injury. *Neuroimage Clin.* 7, 611–621. doi: 10.1016/j.nicl.2015.02.020
- Engel, A. K., Gerloff, C., Hilgetag, C. C., and Nolte, G. (2013). Intrinsic coupling modes: multiscale interactions in ongoing brain activity. *Neuron* 80, 867–886. doi: 10.1016/j.neuron.2013.09.038
- Engels, M. M., Stam, C. J., van der Flier, W. M., Scheltens, P., de Waal, H., and van Straaten, E. C. (2015). Declining functional connectivity and changing hub locations in Alzheimer's disease: an EEG study. *BMC Neurol* 15:145. doi: 10.1186/s12883-015-0400-7
- Escudero, J., Hornero, R., Abásolo, D., and Fernández, A. (2011). Quantitative evaluation of artifact removal in real magnetoencephalogram signals with blind source separation. *Ann. Biomed. Eng.* 39, 2274–2286. doi: 10.1007/s10439-011-0312-7
- Fouss, F., Francoise, K., Yen, L., Pirotte, A., and Saerens, M. (2012). An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification. *Neural Netw.* 31, 53–72. doi: 10.1016/j.neunet.2012.03.001
- Garcés, P., Martín-Buro, M. C., and Maestú, F. (2016). Quantifying the test-retest reliability of magnetoencephalography resting-state functional connectivity. *Brain Connect.* 6, 448–460. doi: 10.1089/brain.2015.0416
- Gofi, J., Avena-Koenigsberger, A., Velez de Mendizabal, N., van den Heuvel, M. P., Betzel, R. F., and Sporns, O. (2013). Exploring the morphospace of communication efficiency in complex networks. *PLoS ONE* 8:e58070. doi: 10.1371/journal.pone.0058070
- Gratton, C., Laumann, T. O., Nielsen, A. N., Green, D. J., Gordon, E. M., Gillmore, A. W., et al. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* 98, 439–452. doi: 10.1016/j.neuron.2018.03.035
- Grent-T-Jong, T., Rivolta, D., Sauer, A., Grube, M., Singer, W., Wibral, M., et al. (2016). MEG-measured visually induced gamma-band oscillations in chronic schizophrenia: evidence for impaired generation of rhythmic activity in ventral stream regions. *Schizophr. Res.* 176, 177–185. doi: 10.1016/j.schres.2016.06.003
- Hall, E. L., Woolrich, M. W., Thomaz, C. E., Morris, P. G., and Brookes, M. J. (2013). Using variance information in magnetoencephalography measures of functional connectivity. *Neuroimage* 67, 203–212. doi: 10.1016/j.neuroimage.2012.11.011
- Hammond, D., Gur, Y., and Johnson, C. (2013). "Graph diffusion distance: a difference measure for weighted graphs based on the graph Laplacian exponential kernel," in *Proceedings of IEEE GlobalSIP* (Austin, TX), 419.
- Handwerker, D. A., Roopchansingh, V., Gonzalez-Castillo, J., and Bandettini, P. A. (2012). Periodic changes in fMRI connectivity. *Neuroimage* 63, 1712–1719. doi: 10.1016/j.neuroimage.2012.06.078
- Hardmeier, M., Hatz, F., Bousleiman, H., Schindler, C., Stam, C. J., and Fuhr, P. (2014). Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG. *PLoS ONE* 9:e108648. doi: 10.1371/journal.pone.0108648
- Hillebrand, A., Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *Neuroimage* 16, 638–650. doi: 10.1006/nimg.2002.1102
- Hillebrand, A., Prejaas, T., van Dellen, E., Yu, M., Carbo, E. W. S., Douw, L., et al. (2016). Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3867–3872. doi: 10.1073/pnas.1515657113
- Hillebrand, A., Singh, K. D., Holliday, I. E., Furlong, P. L., and Barnes, G. R. (2005). A new approach to neuroimaging with magnetoencephalography. *Hum. Brain Mapp.* 25, 199–211. doi: 10.1002/hbm.20102
- Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., and Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* 15, 884–890. doi: 10.1038/nn.3101
- Hunt, B. A., Tewarie, P. K., Mougins, O. E., Geades, N., Jones, D. K., Singh, K. D., et al. (2016). Relationships between cortical myeloarchitecture and electrophysiological networks. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13510–13515. doi: 10.1073/pnas.1608587113
- Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378. doi: 10.1016/j.neuroimage.2013.05.079
- Ioannides, A. A., Dimitriadis, S. I., Saridis, G., Voultsidou, M., Poghosyan, V., Liu, L., et al. (2012). Source space analysis of event-related dynamic reorganization of brain networks. *Comput. Math. Methods Med.* 15:452503. doi: 10.1155/2012/452503
- Jin, S. H., Seol, J., Kim, J. S., and Chung, C. K. (2011). How reliable are the functional connectivity networks of MEG in resting states? *J. Neurophysiol.* 106, 2888–2895. doi: 10.1152/jn.00335.2011
- Koelewijn, L., Bompas, A., Tales, A., Brookes, M. J., Muthukumaraswamy, S. D., Bayer, A., et al. (2017). Alzheimer's disease disrupts alpha and beta-band resting-state oscillatory network connectivity. *Clin. Neurophysiol.* 128, 2347–2357. doi: 10.1016/j.clinph.2017.04.018
- Lachaux, J. P., Rodriguez, E., Martinerie, J., Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208. doi: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C
- Lancaster, T. M., Dimitriadis, S. I., Tansey, K. E., Perry, G., Ihssen, N., Jones, D. K., et al. (2018). Structural and functional neuroimaging of polygenic risk for schizophrenia: a recall-by-genotype-based approach. *Schizophr. Bull.* doi: 10.1093/schbul/sby037. [Epub ahead of print].
- Latora, V., and Marchiori, M. (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett.* 87, 198701.
- Liu, X., and Duyn, J. H. (2013). Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4392–4397. doi: 10.1073/pnas.1216856110
- Logothetis, L. N. (2008). What we can do and what we cannot do with fMRI. *Nature* 453, 869–878. doi: 10.1038/nature06976



- López, M. E., Bruña, R., Aurretetxe, S., Pineda-Pardo, J. A., Marcos, A., Arrazola, J., et al. (2014). Alpha-band hypersynchronization in progressive mild cognitive impairment: a magnetoencephalography study. *J. Neurosci.* 34, 14551–14559. doi: 10.1523/JNEUROSCI.0964-14.2014
- López, M. E., Engels, M. M. A., van Straaten, E. C. W., Bajo, R., Delgado, M. L., Scheltens, P., et al. (2017). MEG beamformer-based reconstructions of functional networks in mild cognitive impairment. *Front. Aging Neurosci.* 9:107. doi: 10.3389/fnagi.2017.00107
- Luckhoo, H., Hale, J. R., Stokes, M. G., Nobre, C., Morris, P. G., Brookes, M. J., et al. (2012). Inferring task-related networks using independent component analysis in magnetoencephalography. *NeuroImage* 62, 530–541. doi: 10.1016/j.neuroimage.2012.04.046
- Marimpis, A. D., Dimitriadis, S. I., Adamos, D. A., and Laskaris, N. A. (2016). NDMF connectivity microstates: a new approach to represent the dynamic brain coordination. *Front. Neuroinform.* 10. doi: 10.3389/conf.fninf.2016.20.00022
- Martinetz, T. M., Berkovich, S. G., and Schulten, K. J. (1993). “Neural-gas” network for vector quantization and its application to time-series prediction. *IEEE Trans. Neural Netw.* 4, 558–569. doi: 10.1109/72.238311
- Mevel, K., Chételat, G., Eustache, F., and Desgranges, B. (2011). The default mode network in healthy aging and Alzheimer's disease. *Int. J. Alzheimers Dis.* 2011:535816. doi: 10.4061/2011/535816
- Mohan, A., Roberto, A. J., Mohan, A., Lorenzo, A., Jones, K., Carney, M. J. et al. (2016). The significance of the default mode network (DMN) in neurological and neuropsychiatric disorders: a review. *Yale J. Biol. Med.* 89, 49–57.
- Mylonas, D. S., Siettos, C. I., Evdokimidis, I., Papanicolaou, A. C., and Smyrnis, N. (2015). Modular patterns of phase desynchronization networks during a simple visuomotor task. *Brain Topogr.* 29, 118–129. doi: 10.1007/s10548-015-0451-5
- Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307. doi: 10.1016/j.clinph.2004.04.029
- O'Neill, G. C., Bauer, M., Woolrich, M. W., Morris, P. G., Barnes, G. R., and Brookes, M. J. (2015). Dynamic recruitment of resting state sub-networks. *Neuroimage* 115, 85–95. doi: 10.1016/j.neuroimage.2015.04.030
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:9. doi: 10.1155/2011/156869
- Pereda, E., Quiroga, R. Q., and Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* 77, 1–37. doi: 10.1016/j.pneurobio.2005.10.003
- Raichle, M. E. (2015). The brain's default mode network. *Ann. Rev. Neurosci.* 38, 433–447. doi: 10.1146/annurev-neuro-071013-014030
- Rosazza, C., and Minati, L. (2011). Resting-state brain networks: literature review and clinical applications. *Neurol. Sci.* 32, 773–785. doi: 10.1007/s10072-011-0636-y
- Schoffelen, J. M., and Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Hum. Brain Mapp.* 30, 1857–1865. doi: 10.1002/hbm.20745
- Siegel, M., Donner, T. H., and Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* 13, 121–134. doi: 10.1038/nrn3137
- Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., et al. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3131–3136. doi: 10.1073/pnas.1121329109
- Stam, C. J. (2014). Modern network science of neurological disorders. *Nat. Rev. Neurosci.* 15, 683–695. doi: 10.1038/nrn3801
- Tewarie, P., Schoonheim, M. M., Schouten, D. I., Polman, C. H., Balk, L. J., Uitdehaag, B. M. J., et al. (2015). Functional brain networks: linking thalamic atrophy to clinical disability in multiple sclerosis, a multimodal fMRI and MEG Study. *Hum. Brain Mapp.* 36, 603–618. doi: 10.1002/hbm.22650
- Toppi, J., Astolfi, L., Poudel, G. R., Innes, C. R., Babiloni, F., and Jones, R. D. (2015). Time-varying effective connectivity of the cortical neuroelectric activity associated with behavioural microsleeps. *Neuroimage* 124, 421–432. doi: 10.1016/j.neuroimage.2015.08.059
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations inSPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Uhlhaas, P. J., Liddle, P., Linden, D. E. J., Nobre, A. C., Singh, K. D., and Gross, J. (2017). Magnetoencephalography as a tool in psychiatric research: current status and perspective. *Biol. Psychiatry* 2, 235–244. doi: 10.1016/j.bpsc.2017.01.005
- van den Heuvel, M. P., Mandl, R. C., Kahn, R. S., and Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* 30, 3127–3141.
- Vidaurre, D., Quinn, A. J., Baker, A. P., Dupret, D., Tejero-Cantero, A., and Woolrich, M. W. (2016). Spectrally resolved fast transient brain states in electrophysiological data. *Neuroimage* 126, 81–95. doi: 10.1016/j.neuroimage.2015.11.047
- Vrba, J., and Robinson, S. E. (2001). Signal processing in magnetoencephalography. *Methods* 25, 249–271. doi: 10.1006/meth.2001.1238
- Wang, S. P., and Pei, W. J. (2008). First passage time of multiple Brownian particles on networks with applications. *Physica A* 387, 4699–4708. doi: 10.1016/j.physa.2008.03.032
- Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., and Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *Neuroimage* 54, 2116–2124. doi: 10.1016/j.neuroimage.2010.10.023
- Wens, V., Marty, B., Mary, A., Op de Beeck, M., Goldman, S., Van Bogaert, P., et al. (2015). A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. *Hum. Brain Mapp.* 36, 4604–4621.
- Yamashita, M., Kawato, M., and Imamizu, H. (2015). Predicting learning plateau of working memory from whole-brain intrinsic network connectivity patterns. *Sci. Rep.* 5:7622. doi: 10.1038/srep07622
- Yang, C. Y., and Lin, C. P. (2015). Time-varying network measures in resting and task states using graph theoretical analysis. *Brain Topogr.* 28, 529–408. doi: 10.1007/s10548-015-0432-8

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Dimitriadis, Routley, Linden and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Test-Retest Reliability of “High-Order” Functional Connectivity in Young Healthy Adults

Han Zhang<sup>1</sup>, Xiaobo Chen<sup>1</sup>, Yu Zhang<sup>1</sup> and Dinggang Shen<sup>1,2\*</sup>

<sup>1</sup> Department of Radiology and Brain Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, <sup>2</sup> Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Institute of Psychology (CAS), China

### Reviewed by:

Yuhui Du,  
Mind Research Network,  
United States  
Veena A Nair,  
University of Wisconsin-Madison,  
United States

### \*Correspondence:

Dinggang Shen  
dgshen@med.unc.edu

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 10 April 2017

**Accepted:** 18 July 2017

**Published:** 02 August 2017

### Citation:

Zhang H, Chen X, Zhang Y and  
Shen D (2017) Test-Retest Reliability  
of “High-Order” Functional  
Connectivity in Young Healthy Adults.  
Front. Neurosci. 11:439.  
doi: 10.3389/fnins.2017.00439

Functional connectivity (FC) has become a leading method for resting-state functional magnetic resonance imaging (rs-fMRI) analysis. However, the majority of the previous studies utilized pairwise, temporal synchronization-based FC. Recently, high-order FC (HOFC) methods were proposed with the idea of computing “correlation of correlations” to capture high-level, more complex associations among the brain regions. There are two types of HOFC. The first type is *topographical profile similarity-based HOFC* (tHOFC) and its variant, *associated HOFC* (aHOFC), for capturing different levels of HOFC. Instead of measuring the similarity of the original rs-fMRI signals with the traditional FC (low-order FC, or LOFC), tHOFC measures the similarity of LOFC profiles (i.e., a set of LOFC values between a region and all other regions) between each pair of brain regions. The second type is *dynamics-based HOFC* (dHOFC) which defines the quadruple relationship among every four brain regions by first calculating two pairwise dynamic LOFC “time series” and then measuring their temporal synchronization (i.e., temporal correlation of the LOFC fluctuations, not the BOLD fluctuations). Applications have shown the superiority of HOFC in both disease biomarker detection and individualized diagnosis than LOFC. However, no study has been carried out for the assessment of test-retest reliability of different HOFC metrics. In this paper, we systematically evaluate the reliability of the two types of HOFC methods using test-retest rs-fMRI data from 25 (12 females, age  $24.48 \pm 2.55$  years) young healthy adults with seven repeated scans (with interval = 3–8 days). We found that all HOFC metrics have satisfactory reliability, specifically (1) fair-to-good for tHOFC and aHOFC, and (2) fair-to-moderate for dHOFC with relatively strong connectivity strength. We further give an in-depth analysis of the biological meanings of each HOFC metric and highlight their differences compared to the LOFC from the aspects of cross-level information exchanges, within-/between-network connectivity, and modulatory connectivity. In addition, how the dynamic analysis parameter (i.e., sliding window length) affects dHOFC reliability is also investigated. Our study reveals unique functional associations characterized by the HOFC metrics. Guidance and recommendations for future applications and clinical research using HOFC are provided. This study has made a further step toward unveiling more complex human brain connectome.

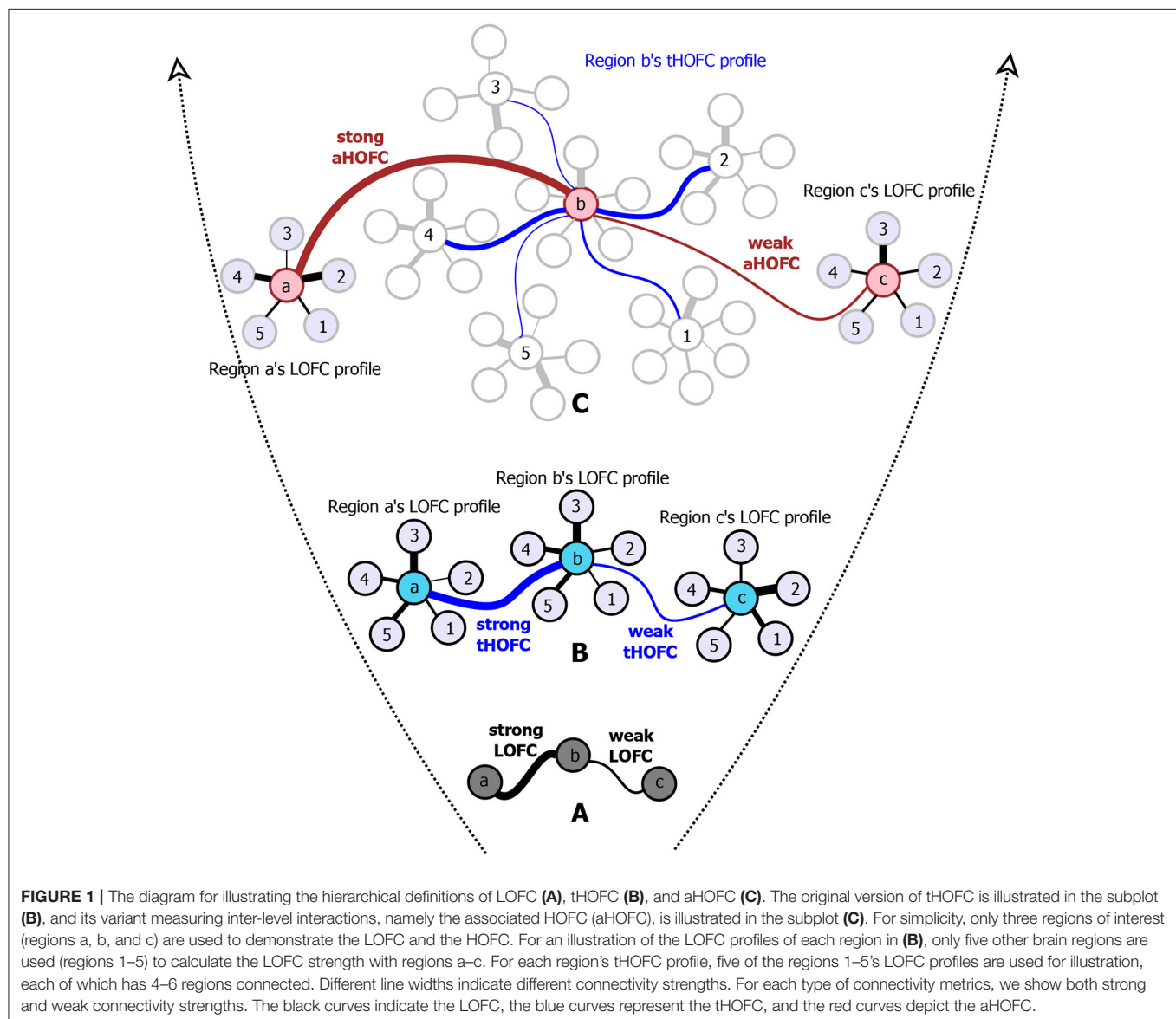
**Keywords:** test-retest, reliability, functional connectivity, high-order connectivity, resting-state fMRI, dynamic connectivity

## INTRODUCTION

Functional connectivity (FC), as originally proposed as the temporal dependence between different spatially-distant brain regions (Friston et al., 1993), has become the major method to analyze resting-state functional magnetic resonance imaging (rs-fMRI) data (Biswal et al., 2010; Fox and Greicius, 2010; Van Dijk et al., 2010; Friston, 2011; Yeo et al., 2011; Fox et al., 2012). Except for the seed-based correlation that mainly focuses on voxel-wise massive one-to-one FC, the mostly adopted FC analysis strategy is pairwise correlation of region-averaged rs-fMRI signals for each pair of  $N$  brain regions, often resulting in an  $N \times N$  FC matrix that represents whole-brain functional connectome. Various post-processing methods can be applied to these matrices to detect the potential connectivity biomarkers for brain diseases, including mass-univariate analyses that reveal group-level FC differences, or pattern cognition

and individualized classification based on the features of all the FCs.

However, such a one-to-one pairwise FC calculation has a well-known limitation since it reveals only simple temporal synchronization between two brain regions (**Figure 1A**). With simple FC, the high-level relationship among the brain regions may not be fully captured. To address this issue, we have proposed several metrics to capture high-level relationship based on “correlation’s correlation,” namely high-order FC (HOFC), among the brain regions. There are two *major types* of HOFC. The first is calculated based on the topological architecture of the complex brain FC networks. As shown in **Figure 1B**, by extracting a regional one-to-all FC profile that constitutes a set of the FC strengths between one region to all other regions, we can characterize the FC topographical similarity for each pair of the brain regions by calculating a second round of correlation on these regional FC profiles (Zhang H. et al., 2016). This

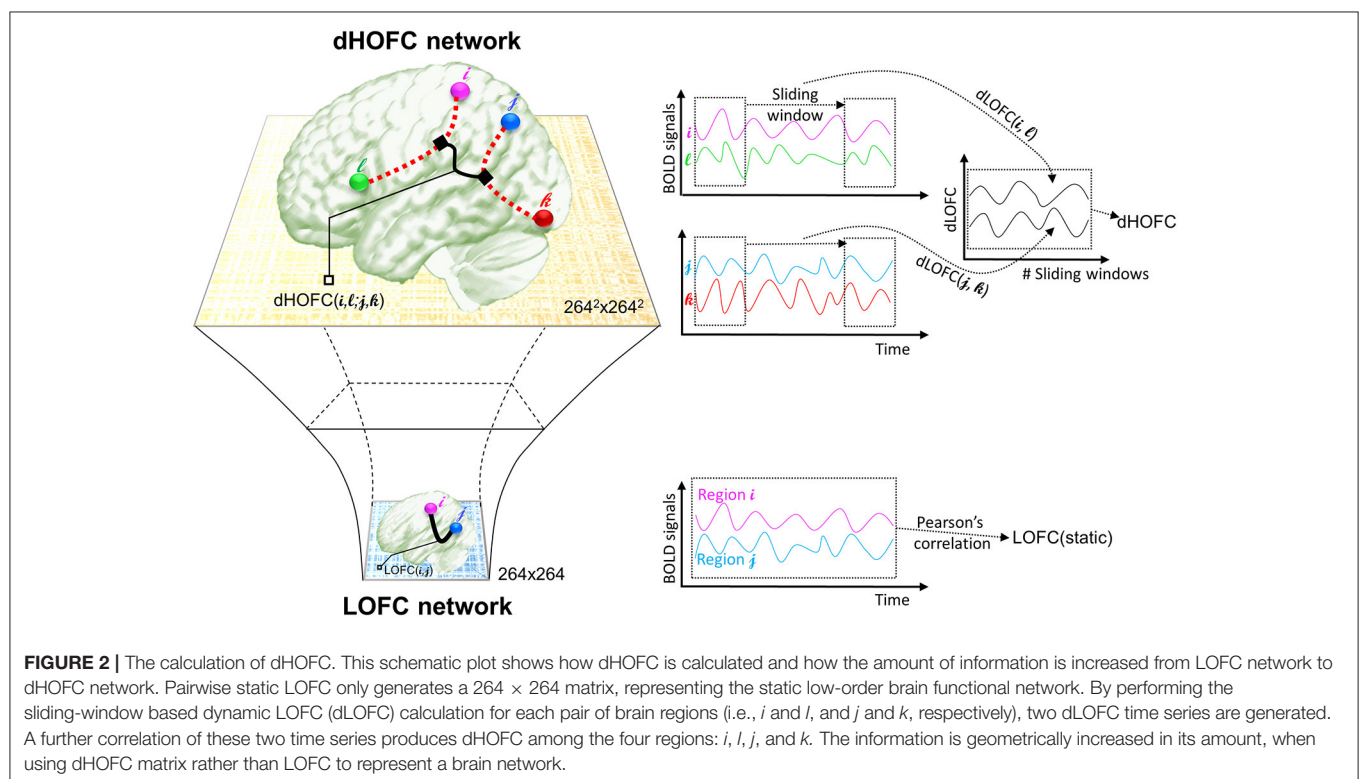


metric captures the high-level functional similarities between two brain regions beyond the traditional temporal synchronization based merely on the raw rs-fMRI signals. We have shown that, with such a “correlation of correlations” strategy, this HOFC metric reveals complementary information to the traditional FC for biomarker detection for brain disease (Zhang H. et al., 2016). From then on, we call the traditional FC as low-order FC (LOFC), and use *topographical profile similarity-based HOFC* (tHOFC) to name this topographical similarity-based HOFC method. If two regions have strong tHOFC, they have quite similar LOFC patterns to all the brain regions but they may have quite distinct rs-fMRI signals. Further comparison of tHOFC between the mild cognitive impairment (MCI) and the healthy elderly groups has unveiled novel potential biomarkers for early Alzheimer’s diseases (AD) detection (Zhang H. et al., 2016). Later on, a variant of tHOFC, named as “associated HOFC (aHOFC),” was also proposed for further assessment of the resemblance between the topographical profile of LOFC and that of the tHOFC (Figure 1C), which indicates a modulation and inter-level functional association between the low- and high-level functional organizations. aHOFC has demonstrated its better performance than LOFC and even the tHOFC in MCI classification (Zhang et al., 2017). Of note, although both tHOFC and aHOFC measure high-level functional association, it is still the pairwise relationship characterized, similar to pairwise LOFC.

The *second* type of HOFC is based on a different interpretation of the “correlations’ correlation” and can measure more complex relationship than a pairwise one. Rather than using the whole length of the rs-fMRI signals to obtain *static* LOFC, we use

a brief segment of the data to conduct LOFC analysis for generating an instantaneous whole-brain LOFC network. By moving the window segment forward, a set of “dynamic” whole-brain LOFC is generated. For each pair of the brain regions, there is a *dynamic* LOFC time series reflecting the time-varying LOFC; it can be further correlated with the *dynamic* LOFC time series from another pair of brain regions, thus measuring high-level, quadruple interactions among four brain regions or two brain region pairs (Chen et al., 2016a). We call this as dynamics-based HOFC (dHOFC), which can be regarded as a “hyperlink” connecting two “hypernodes,” and each of the hypernodes represent a regular link between two brain regions (Figure 2). Since the dynamic LOFC may reflect adaptive and state-related temporary functional architecture, the dHOFC can measure the coherence of such processes, thus can reveal what LOFC cannot find. In addition, as shown in Figure 2, by calculating dHOFC on every quadruplet, we get a larger connectivity matrix of dHOFC compared with the small LOFC matrix. This indicates that we can use dHOFC to further construct more complex brain functional networks with more information introduced. This HOFC method has been successfully applied to early MCI detection (Chen et al., 2016a) and early AD detection (Chen et al., 2016b), as well as prediction of overall survival time of patients with brain gliomas (Liu et al., 2016), all with significantly better accuracy than LOFC.

Despite success in the abovementioned series of studies and the promising future of the HOFC applications, an essential question of how reliable and reproducible of such high-level





FC metrics is still unanswered. Compared to the traditional LOFC with its test-retest reliability systematically assessed, which is fair-to-good when examined in both region- (Wang et al., 2011) and voxel-wise manners (Shehzad et al., 2009; Somandepalli et al., 2015), the state-of-the-art HOFC algorithms still lack comprehensive reliability assessment. Timely evaluation of HOFC reliability is crucial for their broader applications. Only with adequate reliability, we can then expect the detected HOFC biomarkers, or the HOFC-based disease detection, to be reproducible. Notably, the recent revisits of previously famous biomarker detection studies have found that those biomarkers could not be properly reproduced (Horrigan et al., 2017), which has been ringing a warning bell to the field and further increases the urgency of HOFC reliability study. In this paper, we will systematically evaluate the test-retest reliability of both topographical similarity-based HOFC (tHOFC and aHOFC) and dynamics-based HOFC (dHOFC) at the single connection level using repeated rs-fMRI scans. Note that good test-retest reliability of HOFC will indicate that the estimated HOFC from a subject based on one rs-fMRI session can be largely replicated based on the data of the same subject but from another rs-fMRI session. In addition, if a method or metric is proven to be test-retest reliable, its result could be more robust to noise, thus can be more easily to be reproduced by other researchers. Another aim of this HOFC test-retest reliability study is to investigate the underlying neurobiological meaning of the HOFC metrics according to the reliability evaluation. Test-retest reliability and its difference for different connections are informative to let us draw conclusions, especially on differentiating the noise from the signal. For example, previous studies found that the noise-related component derived from independent component analysis (ICA) on the rs-fMRI have lower test-retest reliability than that of the biologically meaningful components representing brain functional networks (Zuo et al., 2010). Finally, as the HOFC is still a new concept to the field, a timely test-retest reliability assessment will provide guidelines to further studies to prevent from unreliable results being misinterpreted.

We hypothesize that all the HOFC metrics (tHOFC, aHOFC, and dHOFC) have at least fair test-retest reliability, which means the major pattern of the HOFC can be largely reproduced based on a repeated rs-fMRI scan, because these metrics were proposed to reflect stable and biologically meaningful brain functional organizations that could thus be consistent. Different from tHOFC and aHOFC, dHOFC is based on dynamic LOFC which captures transient brain states. Although such dynamic LOFC could be different at a different time (such as different rs-fMRI scans), we think that the second round of correlations based on the dynamic LOFC time series could produce stable dHOFC that may reflect higher-level brain functional organization (i.e., synchronized brain state transition). Therefore, we also proposed that the dHOFC is considerably test-retest reliable. As an important influencing factor, whether different parameter settings such as different sliding window lengths could affect the dHOFC reliability will be also investigated. Based on the reliability results, practical guidelines and recommendations are provided for future studies.

## MATERIALS AND METHODS

### Data

We adopted a publicly available test-retest data ([http://dx.doi.org/10.15387/fcp\\_indi.corr.hnu1](http://dx.doi.org/10.15387/fcp_indi.corr.hnu1)) as part of the Consortium for Reliability and Reproducibility (CoRR) (Zuo et al., 2014). This dataset includes 30 healthy adults (aged 20–30 years old, 15 females) with 10 repeated rs-fMRI scans (sessions) within 1 month. The 10-session rs-fMRI scans are essential for more accurate reliability estimation because it constitutes adequate samples at *both* the group (# of repeated scans) *and* individual (# of subjects) levels; however, many previous test-retest reliability studies only used two sessions (Zhang et al., 2011a,b). Based on this dataset, intra-class correlation (ICC) for test-retest reliability evaluation can be more accurately estimated based on multiple repeated scans. Another advantage of this dataset is that the whole period of data collection was completed within 1 month, with each rs-fMRI session separated by 3–4 days. This reduces the potential inference of other longitudinal factors to the reliability estimation, such as development, plasticity, etc. This study was carried out in accordance with the recommendations of the ethics committee of the Center for Cognition and Brain Disorders at Hangzhou Normal University. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of the Center for Cognition and Brain Disorders at Hangzhou Normal University.

The data was acquired by a GE MR750 3.0 Tesla MRI scanner, including both a T1-weighted image (used for rs-fMRI registration) and an rs-fMRI (echo-planar imaging, TR/TE = 2,000/30 ms, voxel size =  $3.4 \times 3.4 \times 3.4 \text{ mm}^3$ , slice number = 43, matrix size =  $64 \times 64 \times 64$ , 10 min, 300 time points). During rs-fMRI, all subjects stared at a fixation point on the screen without falling asleep. For detailed data information, please refer to the data release and CoRR websites.

### Data Preprocessing

The rs-fMRI preprocessing was carried out based on DPARSF v2.3 (Yan and Zang, 2010) with routine procedures following the previous studies (Mao et al., 2015; Yu et al., 2017). It includes: (1) removing the first 5 time points, (2) slice timing correction, (3) head motion correction, (4) unified segmentation of the T1-weighted image after it was aligned to the rs-fMRI data, (5) warping the rs-fMRI data based on the deformation field produced by the previous step to the Montreal Neurological Institute (MNI) standard space, (6) spatially smoothing with a 6-mm Full Width at Half Maximum (FWHM) isotropic Gaussian kernel, (7) band-pass filtering (0.01–0.1 Hz), and 8) regressing out covariate signals including the first- and second-order polynomial functions, averaged signals from the white matter and cerebrospinal fluid (CSF), as well as the *Friston 24*-parameter head motion curves. Similar to our previous works (Chen et al., 2016a), we did not conduct data scrubbing to remove the data with larger frame-wise head motion. Although this step could further reduce head motion effect to LOFC analysis (Power et al., 2014), scrubbing itself will interrupt the temporal structure of the data and probably introduce artifacts into the dynamic LOFC analysis (Hutchison et al., 2013) before the dHOFC calculation. Instead, we used a stringent head motion exclusion criterion.

That is, the subject with head motion larger than 1.5 mm or  $1.5^\circ$  in any rs-fMRI session was discarded. The rs-fMRI sessions with too many ( $>3$ ) subjects discarded were not used for the test-retest reliability estimation. Therefore, sessions #2, #6, and #10 were discarded. Only 7 rs-fMRI sessions and 25 subjects (13 males, 12 females, age  $24.48 \pm 2.55$  years old, ranging from 20 to 30 years old) were finally chosen for the following analysis. We also calculated the percentage of the rs-fMRI frames with excessive ( $>0.5$ ) frame-wise displacement based on Power et al.'s method (Power et al., 2014) for each subject and each session; all the remained subjects have  $< 5\%$  (i.e., 14) frames with excessive micro-head motion. In addition, we further test whether data scrubbing will affect the LOFC, tHOFC, and aHOFC estimation by conducting the similar analysis based on the scrubbed data; as we anticipated, the reliability did not change significantly.

## LOFC: Temporal Synchronization of rs-fMRI Signals

We first calculate the traditional FC (i.e., LOFC) based on the pair-wise temporal correlation of the preprocessed rs-fMRI signals for each of two brain regions using Pearson's correlation. Letting  $x_i(t)$  and  $x_j(t)$  represent the rs-fMRI signals for two brain regions  $i$  and  $j$  at time point  $t$  ( $t = 1, \dots, T$ ), the  $\text{LOFC}_{ij}$  can be defined as

$$\text{LOFC}_{ij} = \frac{\sum_{t=1}^T (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)}{\sqrt{\sum_{t=1}^T (x_i(t) - \bar{x}_i)^2} \sqrt{\sum_{t=1}^T (x_j(t) - \bar{x}_j)^2}}$$

where  $\bar{x}_i$  is the mean of the rs-fMRI signals at region  $i$ . A 264-region brain atlas (Power et al., 2013) was used to parcellate the brain; each region of interest (ROI) was represented by a sphere with 5-mm radius. The mean rs-fMRI signals from each ROI were extracted. LOFC matrix with the size of  $264 \times 264$  was calculated for each subject for each session, which served as a baseline for comparison with the topographical similarity-based HOFC methods (tHOFC/aHOFC).

## tHOFC/aHOFC: Similarity of Topographical Connectivity Profiles

tHOFC and aHOFC calculations are straightforward without free parameters to be estimated, both of which characterize the relationship between two brain regions. However, they characterize different pairwise relationship from that of the LOFC due to the difference in the input "signals" (for LOFC calculation, the input signals are the rs-fMRI time series; but, for tHOFC, they are regional LOFC profile). Different input signals may cause prominent difference between LOFC and tHOFC (or aHOFC) between the same two brain regions. In fact, we have found that two brain regions with little temporal synchronization (indicating weak LOFC) have highly similar topographical LOFC profiles (suggesting strong tHOFC).

Specifically, tHOFC was calculated by column-wise correlation for each pair of the columns (with each column representing the LOFC profile of each brain region) from the  $264 \times 264$  LOFC matrix. Letting  $\text{LOFC}_i$  represents the LOFC profile for region  $i$ , and  $\text{LOFC}_i = \{\text{LOFC}_{ik} | k \in \mathbf{R}, k \neq i\}$  (where  $\mathbf{R}$  is the

set of all brain regions), the  $\text{tHOFC}_{ij}$  can be defined as,

$$\text{tHOFC}_{ij} = \frac{\sum_k (\text{LOFC}_{ik} - \overline{\text{LOFC}_i})(\text{LOFC}_{jk} - \overline{\text{LOFC}_j})}{\sqrt{\sum_k (\text{LOFC}_{ik} - \overline{\text{LOFC}_i})^2} \sqrt{\sum_k (\text{LOFC}_{jk} - \overline{\text{LOFC}_j})^2}}$$

where  $k \in \mathbf{R}, k \neq i, j$ . Before such correlation, all the LOFC values were transformed to  $z$ -scores using Fisher's  $r$ -to- $z$  transformation to satisfy the hypothesis of the second round of Pearson's correlation. Of note, self-connections of the two regions were excluded, i.e., the LOFC profile of each region is a 262-length vector ( $262 = 264 - 2$ ).

The aHOFC is defined further based on the topographical profiles of the tHOFC. It measures the similarity between the LOFC topographical profiles and the tHOFC topographical profiles. Each brain region, when viewed from different levels, could interact with all other regions in both low-level (i.e., LOFC) and high-level (i.e., HOFC) manners. The aHOFC focuses on such a modulatory association between the two levels. During the aHOFC calculation, both LOFC and tHOFC profiles were first transformed into  $z$ -scores; the self-connections were ignored, similar to the calculation of tHOFC. Similarly, we use  $\text{tHOFC}_i$  to represent the tHOFC profile for region  $i$ , where  $\text{HOFC}_i = \{\text{HOFC}_{ik} | k \in \mathbf{R}, k \neq i\}$ . The Pearson's correlation between any  $\text{tHOFC}_i$  and any  $\text{LOFC}_j$  defines aHOFC<sub>ij</sub>:

$$\text{aHOFC}_{ij} = \frac{\sum_k (\text{tHOFC}_{ik} - \overline{\text{tHOFC}_i})(\text{LOFC}_{jk} - \overline{\text{LOFC}_j})}{\sqrt{\sum_k (\text{tHOFC}_{ik} - \overline{\text{tHOFC}_i})^2} \sqrt{\sum_k (\text{LOFC}_{jk} - \overline{\text{LOFC}_j})^2}}$$

where  $k \in \mathbf{R}, k \neq i, j$ . The motivation of aHOFC is that, we think there are not only low-level and high-level FCs in the brain, but also inter-level interactions between LOFC and tHOFC connecting the two levels of FCs, similar to the common observations in many other biological networks, e.g., hierarchical organization and self-resemblance across multiple spatial scales (Guimera et al., 2003). Supposing that, in the human brain, the LOFC may collect and process information and the tHOFC may abstract information via the hierarchy (i.e., correlation's correlation), the possible functions of such inter-level connections could be (1) to facilitate the two levels of information talking to each other, (2) to let the low-level information guide high-level abstraction, and (3) to change the way of low-level information collection for a better high-level information integration. In addition, from robust system point of view, a network or complex biological system could be less fragile and more resilient to the targeted pathological attacks if it has inter-level connections. Taking brain psychophysiological interaction modeling as an example, high-level preset of a psychological status (e.g., attention level) may change sensory information collection, processing, and synthesis. All the evidence together suggests the existence of such an inter-level connection. We have applied the aHOFC to early detection of AD; compared with LOFC, using aHOFC as features not only improved the classification accuracy (Zhang et al., 2017) but also identified different discriminative features as potential AD biomarkers.

Of note, by definition,  $tHOFC_i$  and  $LOFC_i$  could be different, and thus it is also possible to calculate “self-associated HOFC” or  $aHOFC_{ii}$ . Similarly,  $aHOFC_{ij}$  is not necessary to equal to  $aHOFC_{ji}$ . Different from the previous study (Zhang et al., 2017), where the finally obtained aHOFC matrices were converted to be symmetric by adding each subject’s dHOFC matrix with its transpose and dividing the result by two, we did not force the aHOFC matrices derived in this study to be symmetric as our purpose was to assess the aHOFC’s reliability rather than to construct undirected aHOFC networks and make certain neurobiological conclusions. However, to make the mean connectivity matrices comparable among LOFC, tHOFC and aHOFC, we changed the diagonal values in the finally obtained aHOFC matrix to be zeros, i.e., in this study we did not count for the self-associated HOFC. In the future, we should further use the directed aHOFC network with non-zero self-associated HOFC as defined by an asymmetric aHOFC matrix and apply directed network analysis methods on the aHOFC network to reveal more information. **Figure 1** summarize all the three pairwise FC metrics. See the first three columns of the **Table 1** for the summarized differences among these three FC metrics.

$$dHOFC_{il,jk} = \frac{\sum_{\tau=1}^{T-\omega+1} (dLOFC_{il}(\tau) - \overline{dLOFC_{il}}) (dLOFC_{jk}(\tau) - \overline{dLOFC_{jk}})}{\sqrt{\sum_{\tau=1}^{T-\omega+1} (dLOFC_{il}(\tau) - \overline{dLOFC_{il}})^2} \sqrt{\sum_{\tau=1}^{T-\omega+1} (dLOFC_{jk}(\tau) - \overline{dLOFC_{jk}})^2}}$$

## dHOFC: Correlation between Pairwise LOFC Dynamics

The calculation of dHOFC is quite different from that of the topographical similarity-based HOFC metrics (tHOFC/dHOFC), as the tHOFC/aHOFC measures static connectivity but dHOFC is calculated based on dynamic, time-varying LOFC profiles. As for the network topology, dHOFC also differs from tHOFC and aHOFC. As shown in **Figure 2** and summarized in the first three columns of **Table 1**, for dHOFC calculation, dynamic LOFC

for each pair of the brain regions was first calculated using a widely adopted sliding-window strategy (i.e., with window length  $\omega = 30$  time points or 60 s, step size = 1 time point or 2 s); then two dynamic LOFC time series (involving four regions) were correlated using Pearson’s correlation to produce dHOFC between one region pair to another region pair. Letting  $dLOFC(\tau)$  represent the dynamic LOFC strength within a brief time window from  $\tau$  to  $\tau + \omega - 1$ , and the dLOFC time series between region  $i$  and  $l$  can be characterized based on

$$dLOFC_{il}(\tau) = \frac{\sum_{t=\tau}^{\tau+\omega-1} (x_i(t) - \overline{x_i^\tau}) (x_l(t) - \overline{x_l^\tau})}{\sqrt{\sum_{t=\tau}^{\tau+\omega-1} (x_i(t) - \overline{x_i^\tau})^2} \sqrt{\sum_{t=\tau}^{\tau+\omega-1} (x_l(t) - \overline{x_l^\tau})^2}} \\ (\tau = 1, \dots, T - \omega + 1; i, l \in \mathbf{R}, i \neq l)$$

where  $\overline{x_i^\tau}$  represents the mean value of such a brief segment of the rs-fMRI signal starting from  $\tau$ . Similarly, we can define the dLOFC time series between regions  $j$  and  $k$  as  $dLOFC_{jk}(\tau)$  ( $\tau = 1, \dots, T - \omega + 1; j, k \in \mathbf{R}, j \neq k$ ). The further Pearson’s correlation between the two dLOFC time series defines dHOFC between region pairs  $i - l$  and  $j - k$  based on,

where  $\overline{dLOFC_{il}}$  indicates the mean value of the  $dLOFC_{il}$  time series along the whole time. Based on the combination theory, a  $264 \times 264$  LOFC matrix has  $264 \times (264-1)/2 = 34716$  unique region pairs; thus a complete dHOFC network will have  $34716 \times 34716$  in size and over 600 million unique four-region combinations. This will increase the amount of connectomic information and may reveal novel information that cannot be discovered by LOFC/tHOFC/dHOFC. For more details, please see the previous paper (Chen et al., 2016a).

**TABLE 1** | Differences among LOFC and various HOFC metrics.

	Input	Output	Test-retest reliability
LOFC	BOLD signals	Temporal synchronization, functional coherence	Fair-to-good; nearly all connections have fair or better reliability. Within-network connections have better reliability; high-level cognitive function-related connections have better reliability.
tHOFC	Regional LOFC topographical profiles	To what extent two regions share similar LOFC topographical profiles	Fair-to-good; similar to LOFC reliability, but with reduced reliability at within-network connections. Better reliability at inter-network connections (esp. between high-level cognition and primary regions).
aHOFC	Both regional LOFC and regional tHOFC topographical profiles	To what extent topographical LOFC modulates topographical tHOFC	Fair-to-good; similar to LOFC reliability, but with further reduced reliability at within-network connections. Better reliability at inter-network connections.
dHOFC	Dynamic, time varying LOFC time series between two brain regions	Temporal synchronization of two time-varying LOFC time series among four brain regions	Fewer connections have fair or better reliability. Strong (within-network and modulatory) connections have fair-to-moderate reliability. Between-network connections have poor reliability. Shorter window length produces better reliability.

BOLD, Blood-oxygen-level dependent; LOFC, low-order functional connectivity; tHOFC, topographical similarity-based high-order functional connectivity; aHOFC, associated HOFC; dHOFC, dynamics-based HOFC.

In the previous classification-orientated studies, to avoid the curse of dimensionality, dHOFC matrix dimension was further reduced based on hierarchical clustering, which generates relatively fewer clusters by grouping similarly co-varied dynamic LOFC time series together. By doing so, we can detect a few hundreds of the clusters and calculate dHOFC based on the clusters' centroids (Chen et al., 2016a). In the current reliability study, it is not necessary to conduct such a clustering analysis because we are focusing only on the reliability of the dHOFC links, while clustering itself is irrelevant to such a goal and will unnecessarily introduce an additional parameter (i.e., the total number of clusters) which could complicate the current study. Therefore, in this paper, we chose a few ROIs from the total 264 of them to generate a relatively smaller and more interpretable dHOFC network. Specifically, we chose 26 ROIs from the hand-associated sensorimotor areas

for investigation of the dHOFC in the primary functional system; we also chose 17 ROIs from the fronto-parietal task control network (FPN) and 15 from the salience network (SN) to investigate dHOFC in the high-level cognitive function-related brain systems. See **Figure 3** and **Table 2** for the details the ROI definitions. Therefore, we separately generated a dHOFC matrix for the primary areas (with the size of  $325 \times 325$ , where  $325 = 26 \times 25/2$ ) and another dHOFC matrix for two high-level functional areas (with the size of  $496 \times 496$ , where  $496 = 32 \times 31/2$ , since there are totally 32 high-level function-related ROIs,  $32 = 17 + 15$ ). Of note, this is the first paper to systemically investigate the possible neurobiological correlation of the dHOFC in the specific functional systems.

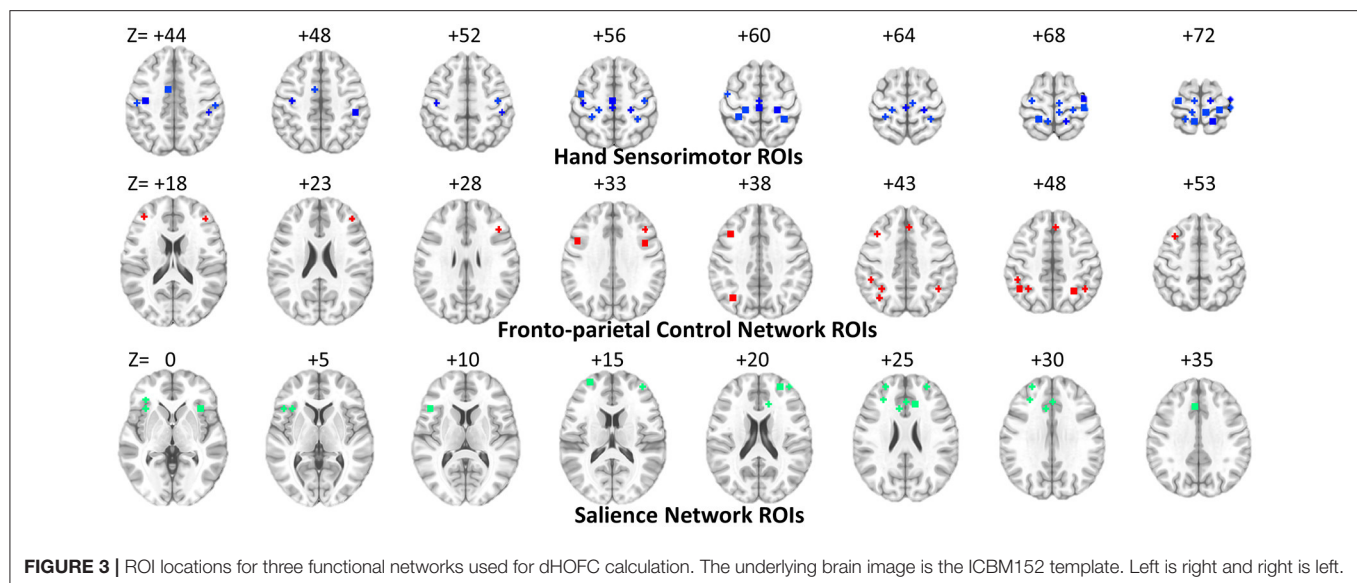
Different window length could affect the accuracy of dynamic LOFC (Hutchison et al., 2013);

**TABLE 2 |** ROI definitions for dHOFC calculation.

#	Orig #	x	y	z	Suggested system	#	Orig #	x	y	z	Suggested system
1	16	10	-2	45	Sensorimotor Hand	1	186	47	10	33	Fronto-parietal Control
2	17	-7	-21	65	Sensorimotor Hand	2	187	-41	6	33	Fronto-parietal Control
3	18	-7	-33	72	Sensorimotor Hand	3	188	-42	38	21	Fronto-parietal Control
4	19	13	-33	75	Sensorimotor Hand	4	189	38	43	15	Fronto-parietal Control
5	20	-54	-23	43	Sensorimotor Hand	5	190	49	-42	45	Fronto-parietal Control
6	21	29	-17	71	Sensorimotor Hand	6	191	-28	-58	48	Fronto-parietal Control
7	22	10	-46	73	Sensorimotor Hand	7	192	44	-53	47	Fronto-parietal Control
8	23	-23	-30	72	Sensorimotor Hand	8	193	32	14	56	Fronto-parietal Control
9	24	-40	-19	54	Sensorimotor Hand	9	194	37	-65	40	Fronto-parietal Control
10	25	29	-39	59	Sensorimotor Hand	10	195	-42	-55	45	Fronto-parietal Control
11	26	50	-20	42	Sensorimotor Hand	11	196	40	18	40	Fronto-parietal Control
12	27	-38	-27	69	Sensorimotor Hand	12	197	-34	55	4	Fronto-parietal Control
13	28	20	-29	60	Sensorimotor Hand	13	198	-42	45	-2	Fronto-parietal Control
14	29	44	-8	57	Sensorimotor Hand	14	199	33	-53	44	Fronto-parietal Control
15	30	-29	-43	61	Sensorimotor Hand	15	200	43	49	-2	Fronto-parietal Control
16	31	10	-17	74	Sensorimotor Hand	16	201	-42	25	30	Fronto-parietal Control
17	32	22	-42	69	Sensorimotor Hand	17	202	-3	26	44	Fronto-parietal Control
18	33	-45	-32	47	Sensorimotor Hand	18	206	31	33	26	Salience Network
19	34	-21	-31	61	Sensorimotor Hand	19	207	48	22	10	Salience Network
20	35	-13	-17	75	Sensorimotor Hand	20	208	-35	20	0	Salience Network
21	36	42	-20	55	Sensorimotor Hand	21	209	36	22	3	Salience Network
22	37	-38	-15	69	Sensorimotor Hand	22	210	37	32	-2	Salience Network
23	38	-16	-46	73	Sensorimotor Hand	23	211	34	16	-8	Salience Network
24	39	2	-28	60	Sensorimotor Hand	24	212	-11	26	25	Salience Network
25	40	3	-17	58	Sensorimotor Hand	25	213	-1	15	44	Salience Network
26	41	38	-17	45	Sensorimotor Hand	26	214	-28	52	21	Salience Network
						27	215	0	30	27	Salience Network
						28	216	5	23	37	Salience Network
						29	217	10	22	27	Salience Network
						30	218	31	56	14	Salience Network
						31	219	26	50	27	Salience Network
						32	220	-39	51	17	Salience Network

Orig #, Original ROI index in the 264 brain region atlas. x, y, z, coordinates of each ROI's center in the MNI space. Suggested system: the functional system suggested by the atlas. We deleted 3 ROIs in the salience network, 8 ROIs in the fronto-parietal control network, and 3 hand sensorimotor ROIs since their belongingness to the suggested functional systems is less replicable across different data sets as suggested by Power et al. (2013).





(Leonardi and Van De Ville, 2015; Zalesky and Breakspear, 2015), thus may further affect dHOFC and its reliability. Therefore, we further examined the relationship between the window length and the test-retest reliability of the dHOFC within the hand sensorimotor areas. The above calculation was repeated with different window length settings (i.e., 20, 40, and 50 time points, corresponding to 40, 80, and 100 s, as  $TR = 2$  s). Of note, the window length of 20 and 30 time points are within the recommended range (30–60 s) from previous studies (Zalesky and Breakspear, 2015), while the larger values (i.e., 80–100 s) are also used in previous dynamic LOFC studies (Leonardi and Van De Ville, 2015).

In addition, according to the previous studies on dynamic LOFC, small correlation values from the dynamic analysis could probably be caused by random noise (Leonardi and Van De Ville, 2015), we think that the reliability of the dHOFC which has weak connectivity strength could also be mainly contributed by noise. To this end, we also used a dHOFC threshold of 0.36 as suggested by Leonardi and Van De Ville (2015) to further filter the dHOFC matrix. If there is a significant modular structure after thresholding, we may be able to draw a conclusion that, although weak dHOFC may be driven by noise, the relatively stronger dHOFC could be biologically meaningful. This is because that, if all dHOFC connections are dominated by noise, the thresholded dHOFC matrix will have a somewhat random spatial pattern rather than a structured one. Similar to the tHOFC and aHOFC, test-retest reliability was also calculated for the relatively strong dHOFC connectivities.

### Intra-Class Correlation for Test-Retest Reliability Evaluation

To investigate test-retest reliability of all types of HOFC connections, we utilized a commonly adopted index called ICC (Shrout and Fleiss, 1979). ICC is a method based on the one-way analysis of variance (ANOVA) which divides the total sum of variance across subjects and repeated rs-fMRI scans into two

parts: between-subject ( $\sigma_b^2$ ) and within-subject (or inter-session variance,  $\sigma_w^2$ ) sum of variance. The theoretical definition of ICC is:

$$ICC = \frac{\sigma_w^2}{\sigma_b^2 + \sigma_w^2};$$

but the estimation of the ICC based on real samples can be written by:

$$ICC = \frac{MS_b - MS_w}{MS_b + (k - 1) \times MS_w},$$

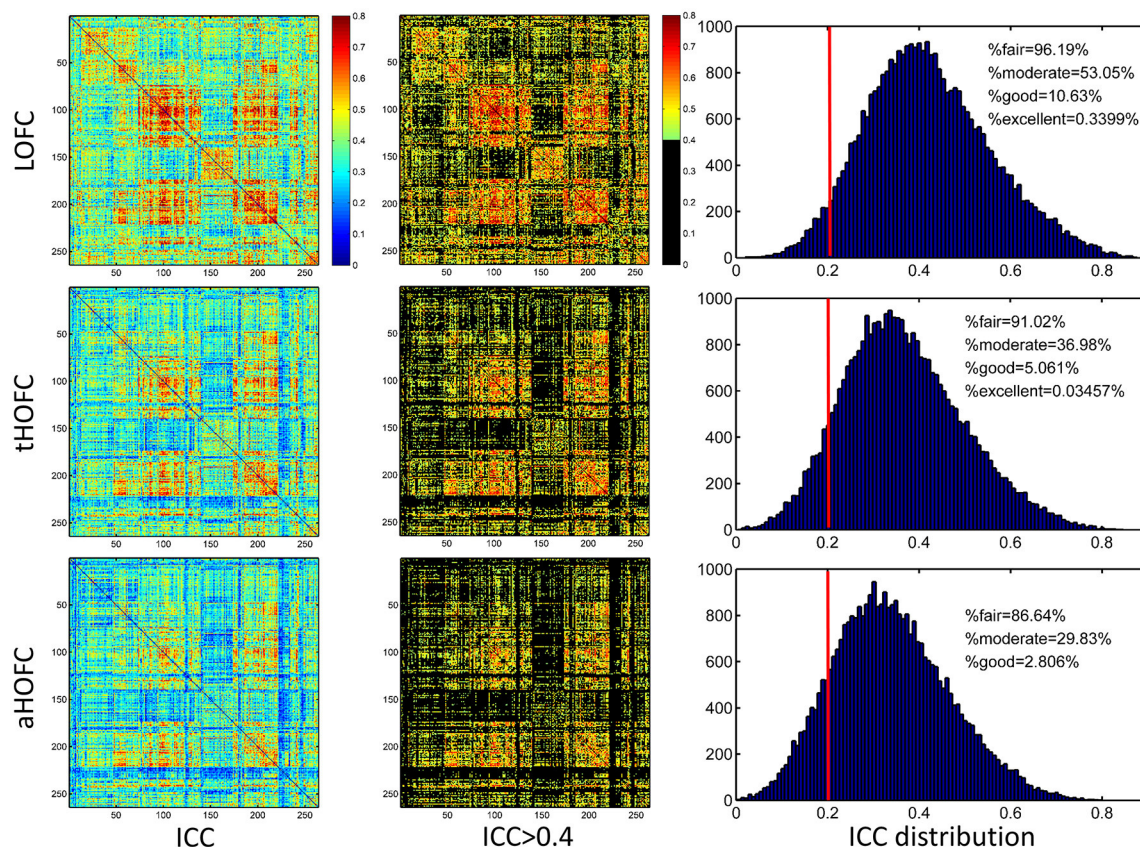
where  $MS_b$  is the mean square of between-subject sum of variance,  $MS_w$  is the mean square of within-subject sum of variance, and  $k$  is the number of repeated rs-fMRI scans (here  $k = 7$ ). ICC is conceptually positive between 0 (not reliable at all) and 1 (perfectly consistent between repeated measurements), but its estimation can be negative in a few cases. We put the negative ICC values to be zeros as always done by previous studies (Zhang et al., 2011a). Based on the value of ICC, reliability is usually categorized as poor ( $ICC = 0-0.2$ ), fair ( $0.2-0.4$ ), moderate ( $0.4-0.6$ ), substantially good ( $0.6-0.8$ ), and excellent ( $>0.8$ ) (Landis and Koch, 1977; Chen et al., 2015).

We first calculated ICC for LOFC, tHOFC and aHOFC, as they are convenient to compare. We then calculate ICC for dHOFC in both primary functional systems (hand sensorimotor areas) and high-level cognition-related functional networks (FPN and SN), to compare the dHOFC in these primary and high-level functional systems.

## RESULTS

### tHOFC and aHOFC Have Moderate-To-Good Test-Retest Reliability

As shown in Figure 4, the test-retest reliability of the tHOFC is generally fair-to-moderate, although slightly lower than that



**FIGURE 4 |** ICC value for each connection of LOFC, tHOFC, and aHOFC. The three rows show test-retest reliability, as assessed by ICC at each connection, for LOFC, tHOFC, and aHOFC, respectively. From left to right are the ICC matrix without thresholding, ICC matrix showing the connectivities with moderate or better reliability (thresholded by  $ICC > 0.4$ ), and the ICC distribution for all connections. In the right of each row, the vertical red line indicates  $ICC = 0.2$ , above which are the connectivities with acceptable (fair or better) reliability; the percentage of the connection with fair, moderate, good and excellent reliability are also shown.

of the LOFC. The test-retest reliability of the aHOFC is similar to that of the tHOFC, with slightly fewer connections having fair-to-moderate ICC. These results indicate that the tHOFC and aHOFC are still reliable metrics. An interesting finding is that the overall pattern of the reliable connections are quite consistent among the LOFC, tHOFC and dHOFC, all of which show prominently better reliability for the connections within default mode network (DMN), as well as those within the FPN and SN, respectively (see those major blocks in the main diagonal and off-diagonal of the **Figure 4**). In addition, we also notice that the off-diagonal connections among the DMN, FPN, and SN have also high reliability. All these high-reliability connections, although a little bit weakened, still exist for tHOFC and dHOFC.

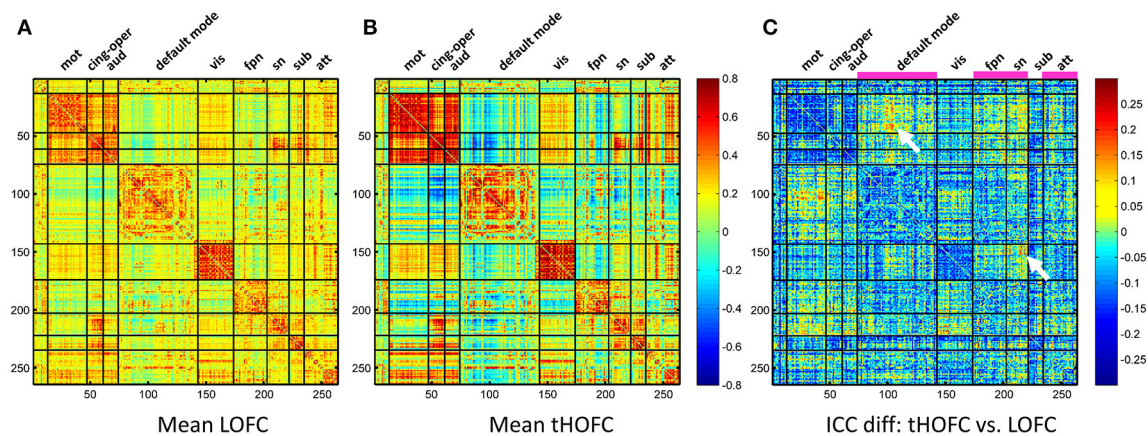
### Links with Increased Reliability for tHOFC and aHOFC, Compared with LOFC

In addition to the overall reduction of reliability for tHOFC/aHOFC compared with LOFC, we further found interesting increased reliability for several tHOFC (**Figure 5**) and aHOFC (**Figure 6**) links. Different from the reduced reliability for mainly intra-network strong connections (see **Figures 5A,B** for the block pattern), the links with increased reliability in

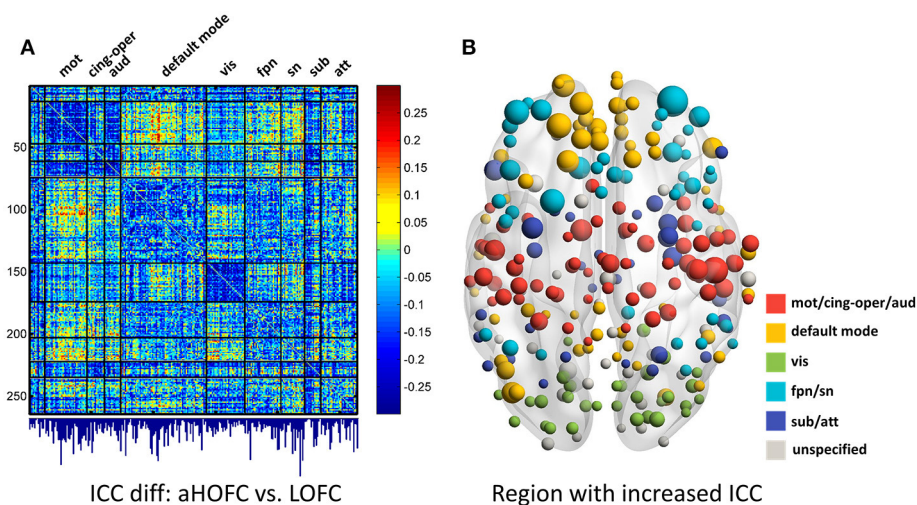
tHOFC compared with LOFC are mainly located at the weak links that connect different systems. Specifically, we found that such links connect *high-level cognition-related* network (DMN, FPN, or SN) and *primary function-related* network (sensorimotor or visual network). For example, as indicated by white arrows in **Figure 5C**, the tHOFC links between the DMN and the hand sensorimotor regions, as well as those between the SN and visual areas, show great (by 0.2) increase in their ICC values. Notably, the group-averaged aHOFC matrix is quite similar to that for LOFC and tHOFC, with the strong aHOFC links mainly located within modules, and the weak aHOFC links between modules (result not shown). Similarly, aHOFC shows the similar result as the tHOFC for the links with increased ICC values, where such increase and reduction in the ICC values are even more prominent (**Figure 6A**).

We further show the specific brain regions with prominent reliability increment by comparing aHOFC with LOFC. To do this, for each brain region, we summarized the extent of ICC increment across all the aHOFC connections to this region with increased ICC. Different regions have various extent of reliability increment (see the bar plot under the matrix of **Figure 6A**). Such differences are further drawn in **Figure 6B** with different sizes





**FIGURE 5 |** The test-retest reliability difference between tHOFC and LOFC. For better understanding which functional system contributes to such ICC increment, we also show the group-averaged LOFC in **(A)** and the group-averaged tHOFC in **(B)** across all subjects and all rs-fMRI sessions. Nine functional systems are shown, with higher intra-system connectivity and sparse inter-system connectivity. The tHOFC with increased reliability **(C)** are located mostly at *inter*-system links (as also highlighted by white arrows). The functional systems with many increased reliability are marked in **(C)** above the matrix and under the names of the functional systems. The abbreviations of the functional systems are: mot (sensorimotor), cing-oper (cingulo-opercular), aud (auditory), vis (visual), fpn (fronto-parietal task control network), sn (salience network), sub (subcortical regions), att (attention-related networks including the dorsal and ventral attentional systems).



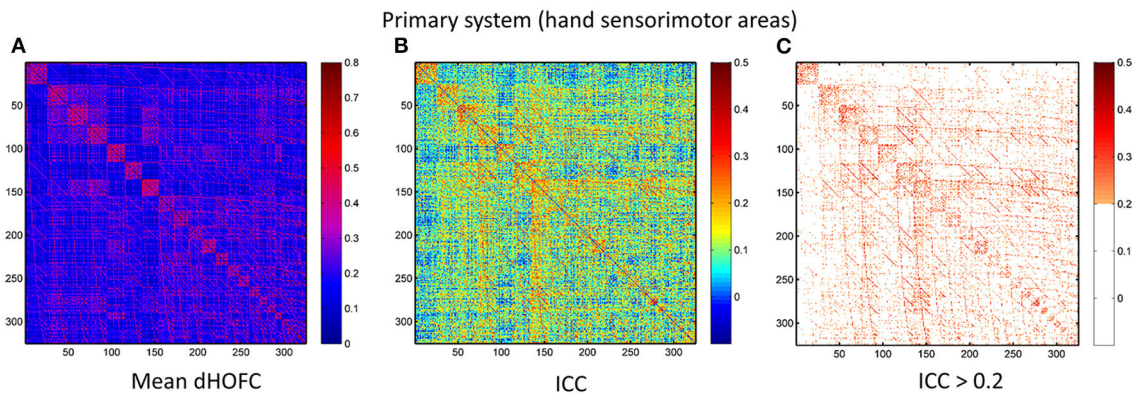
**FIGURE 6 |** The test-retest reliability differences between aHOFC and LOFC. Subplot **(A)** shows the difference in ICC values between aHOFC and LOFC for all connections, with the quantitative measurement of ICC gain for each brain region (i.e., the sum of ICC increment across all the connections to each region) shown as a bar graph under the matrix. Such ICC increment is further visualized as the size of the node for all the brain regions in a brain surface **(B)**. Different colors indicate different functional systems.

of the nodes (with a bigger node indicating greater reliability increment for its aHOFC links). The brain regions with the greatest reliability increment are mainly distributed at the high-level cognitive function-related areas, such as the medial and lateral prefrontal cortices.

Taken together, our results show that both tHOFC and aHOFC have general moderate or better reliability, and that the tHOFC and aHOFC indeed capture novel (mostly high-level cognition-related) information as indirectly reflected by the higher reliability than LOFC.

### Strong dHOFC in the Primary Functional System Has Fair-To-Moderate Reliability

The group-averaged dHOFC within 26 hand sensorimotor ROIs (one of the primary functional system) across all subjects and all sessions is represented by a larger matrix, which shows a significant structure with spatial sparsity (**Figure 7A**). For all the  $325 \times 324/2 = 52650$  dHOFC hyperconnections (by treating the 325 region pairs as hypernodes), their test-retest reliability is shown as an ICC matrix with the same dimension ( $325 \times 325$ ) in **Figure 7B**, with its thresholded (ICC > 0.2) version



**FIGURE 7 |** Test-retest reliability of dynamics-based HOFC (dHOFC) in the primary functional system. **(A)** Averaged dHOFC matrix across all subjects and all imaging sessions, showing the dHOFC strength for every possible high-order links; **(B)** ICC matrix, indicating test-retest reliability for all dHOFC links; **(C)** dHOFC links with fair or better ( $ICC > 0.2$ ) test-retest reliability.

(highlighting only the fairly or better reliable dHOFC) shown in **Figure 7C**. Although many dHOFC links have acceptable reliability as indicated by **Figure 7C**, specific amount of the dHOFC links with  $ICC > 0.2$  is only 11.63% of all possible dHOFC links (**Figure 8A**). We have noted that the dHOFC links with higher reliability tend to be those with greater connectivity strength. If only counting for strong dHOFC (i.e., group mean dHOFC  $> 0.36$ ), a half (49.5%) of such connections will have acceptable reliability (**Figure 8B**). **Figure 9** shows the dHOFC matrix from a randomly selected subject from each of the seven rs-fMRI sessions. We can see that the overall individual dHOFC spatial patterns are consistent across different rs-fMRI sessions. However, there are significant block structures in the group averaged dHOFC matrix (**Figure 7A**) but it is less prominent at the individual-level (**Figure 9**). This difference could be due to the relatively high individual variability in many dHOFC links. While the group average could retain individually consistent dHOFC links, it also suppressed those with relatively high inter-subject variability, thus creating such prominent block structure in the mean dHOFC matrix.

## Strong dHOFC in the High-Level Functional Systems Has Better Reliability

In addition to assessing the reliability for within-primary functional system dHOFC, we also investigated the reliability of high-level cognition-related dHOFC by calculating the dHOFC in the two typical high-order functional systems, i.e., the FPN and SN. **Figure 10A** shows the group-averaged dHOFC in these high-level systems, while **Figure 10B** shows their reliability. Since there are two functional systems involved, the dHOFC can be divided into three main types (see **Figure 10C** and also the summary in **Figure 11**) based on the functional system belongingness of the four brain regions that constitute a dHOFC hyperlink:

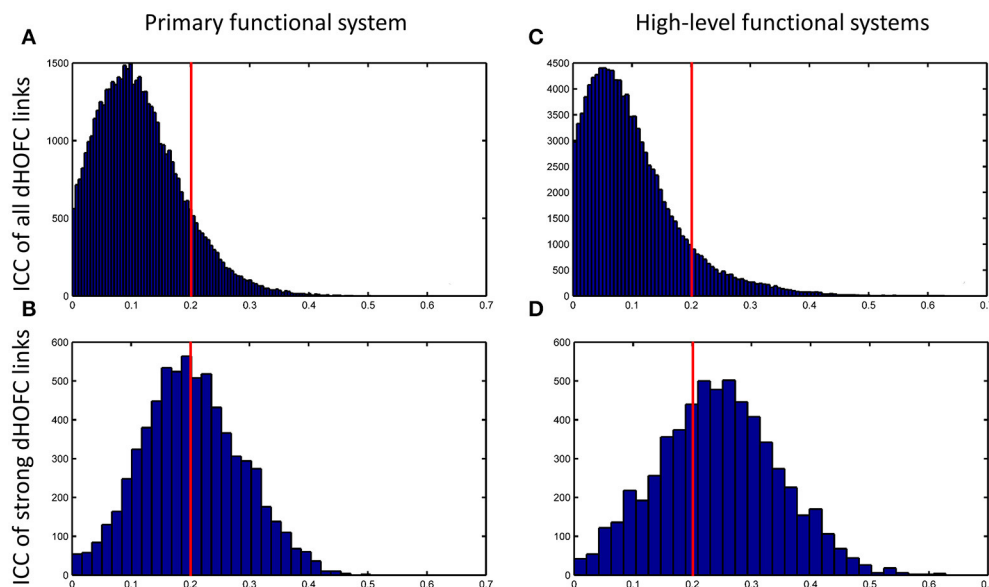
- **Within-network dHOFC.** For each dHOFC consisting of four ROIs, all ROIs belong to the same functional system. For example, a link between two intra-FPN ROIs (regarded as intra-FPN hypernode) has dHOFC with another link between

two intra-FPN ROIs. In this case, both hypernodes are intra-FPN, thus we call this type of dHOFC links *within-FPN* dHOFC. Similarly, we can define *within-SN* dHOFC between two hypernodes that both constitute intra-SN ROIs. This type of the dHOFC characterizes within-network high-order relationship, which has moderate connectivity strength and acceptable reliability (see the first two big blocks in the main diagonal of the matrices in **Figure 10**).

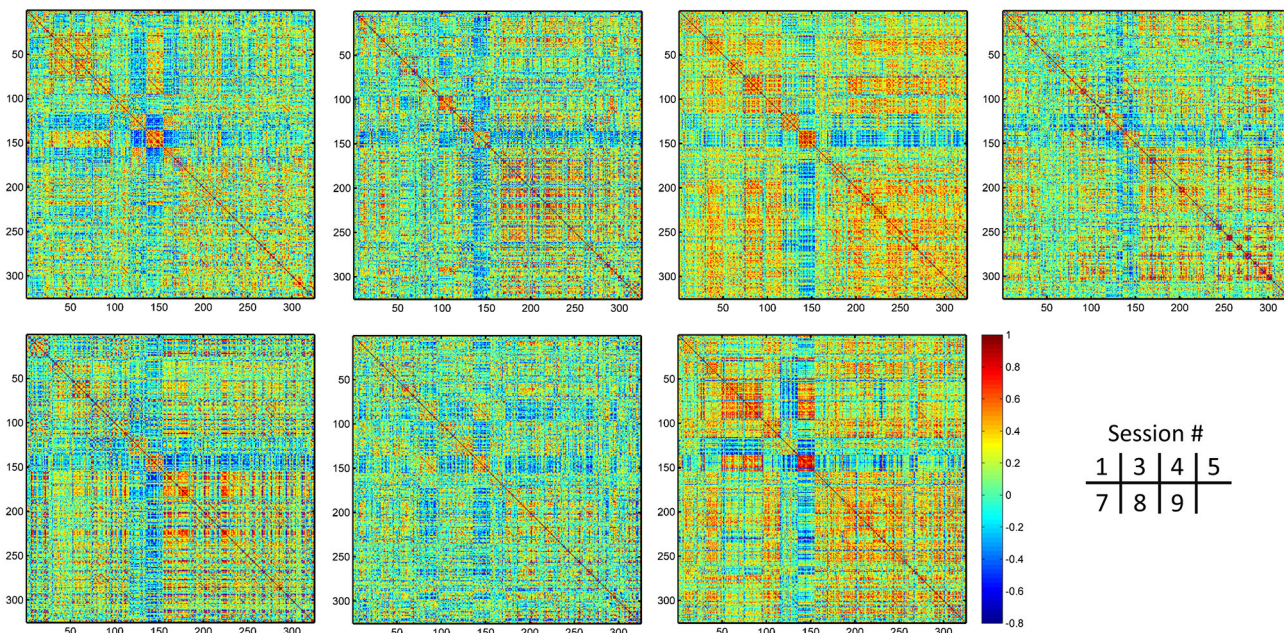
- **Between-network dHOFC.** This type of dHOFC characterizes the high-order relationship *between* two intra-network links (or hypernodes) which belong to different functional networks. For example, a hypernode that connects two FPN ROIs has dHOFC with another hypernode that connects two SN ROIs (i.e., an “intraFPN-to-intraSN” hyperlink). This type of the dHOFC measures high-order functional association between two functional systems. Interestingly, such dHOFC are mostly weak in the connectivity strength and have overall poor reliability (**Figures 10B,C**).
- **Modulatory dHOFC.** This is a new type of connectivity that has not been defined in the previous studies. It contains two hypernodes, at least one of which contains an inter-network link. This type of the dHOFC constitutes the most part of the dHOFC matrix. There are two subtypes for the modulatory dHOFC. The first subtype consists of the dHOFC between one *inter-network* hypernode and one *intra-network* hypernode, e.g., the dHOFC between an intra-FPN hypernode and an FPN–SN link. The second subtype is that both of the hypernodes are the inter-network links. Both of these two cases are able to characterize high-order functional relationships manifesting as “one functional system *modulates* another.” Compared with the first two types of the dHOFC, the modulatory dHOFC show extensive connections (see the third main block in the main diagonal of **Figure 10A**) and acceptable reliability (**Figure 10C**).

As shown in **Figure 10**, the mean dHOFC strength matrix and the dHOFC reliability matrices have highly similar *structured* and *blocked* patterns. Please note that we did not re-arrange





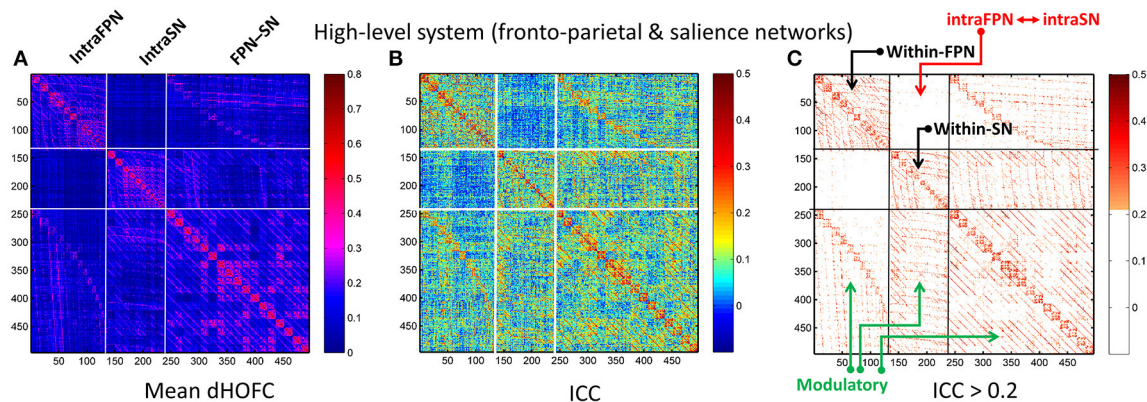
**FIGURE 8 |** Distribution of test-retest reliability (ICC values) for the dHOFc links. **(A,B)** Results for the primary functional system; **(C,D)** results for the high-level functional systems. **(A,C)** show the distribution of the ICC values for all dHOFc links, while **(B,D)** show the distribution of the ICC values for the relatively strong (i.e., mean dHOFc > 0.36) dHOFc links. We selected the hand sensorimotor areas as an example of the primary functional system, and selected both fronto-parietal task control network and salience network as examples of the high-level cognition-related functional systems. Again, the red line indicates the same ICC threshold of 0.2; the bars on the right side of the red line are the numbers of dHOFc links with fair or better reliability.



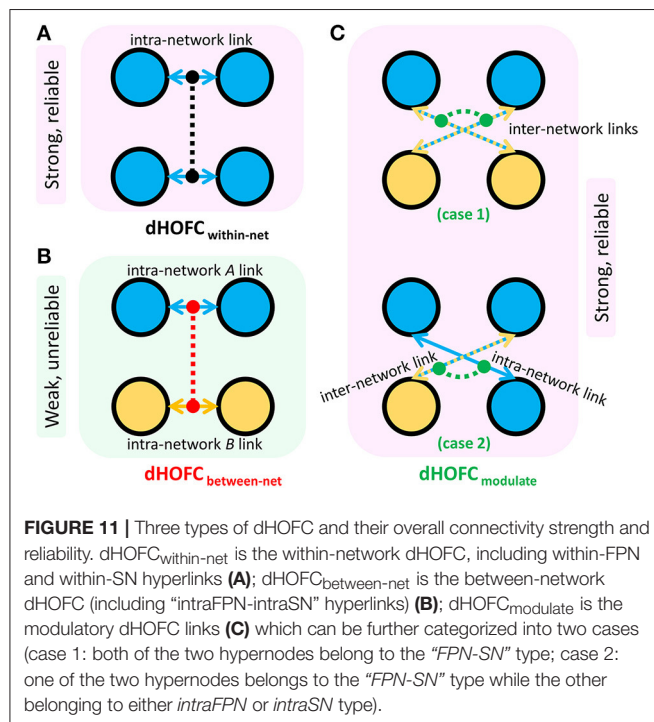
**FIGURE 9 |** Individual dHOFc matrices for all 7 sessions. The dHOFc matrices for the hand sensorimotor network of a randomly selected subject (#9) are plotted.

the columns and the rows of these matrices in a *post hoc* way (e.g., based on module detection using the dHOFc strength); instead, we just grouped the same type of the hypernodes (three types: intra-FPN, intra-SN, and FPN-to-SN) together

before calculating dHOFc and re-arranging the columns and the rows of these matrices in an order of firstly intra-FPN, then intra-SN, and finally FPN-to-SN. Merely through this a priori grouping and rearranging could we reveal such an interesting



**FIGURE 10 |** Test-retest reliability of dHOFC in the two high-level cognition-related functional systems. We selected the fronto-parietal task control network and salience network as examples of the high-level functional system. **(A)** Averaged dHOFC matrix; **(B)** ICC matrix for all dHOFC links; **(C)** ICC matrix for the dHOFC links with fair or better reliability ( $ICC > 0.2$ ). The order of dHOFC links in the matrices is rearranged according to the types of the “hypernodes” (where a hypernode represents a dynamic link between two brain regions). If the hypernode consists of two brain regions that are both from the fronto-parietal task control network, we call it “*intraFPN*” hypernode and re-order them into the first 136 ( $136 = 17 \times 16/2$ ) columns of the dHOFC matrix. We further re-group the 105 ( $105 = 15 \times 14/2$ ) hypernodes which consist of two brain regions both from the salience network (*intraSN*) and put them after the *intraFPN* hypernodes. At last, we put all the remaining 255 ( $255 = 17 \times 15$ ) hypernodes (consisting of one region from FPN and the other from SN, thus called inter-network “*FPN-SN*” hypernodes) after the *intraSN* hypernodes. In this way, the dHOFC matrix is rearranged. According to different types of hypernodes, there are also three different types of (dHOFC) hyperlinks. Among them, the “*within-FPN*” (with both hypernodes being *intraFPN* nodes) and “*within-SN*” (with both hypernodes being *intraSN* nodes) are both indicated by black arrows; the between-network dHOFC hyperlinks (named here as “*intraFPN-intraSN*,” with one hypernode from *intraFPN* and another from *intraSN*) are indicated by the red arrows; and all the remaining dHOFC hyperlinks are named as “*modulatory*” dHOFC (with at least one hypernode belonging to the “*FPN-SN*” type) as indicated by the green arrows.



**FIGURE 11 |** Three types of dHOFC and their overall connectivity strength and reliability. dHOFC<sub>within-net</sub> is the within-network dHOFC, including within-FPN and within-SN hyperlinks **(A)**; dHOFC<sub>between-net</sub> is the between-network dHOFC (including “*intraFPN-intraSN*” hyperlinks) **(B)**; dHOFC<sub>modulate</sub> is the modulatory dHOFC links **(C)** which can be further categorized into two cases (case 1: both of the two hypernodes belong to the “*FPN-SN*” type; case 2: one of the two hypernodes belongs to the “*FPN-SN*” type while the other belonging to either *intraFPN* or *intraSN* type).

structured and block-like pattern for both dHOFC strength and reliability.

Different from the dHOFC in the primary functional system (Figure 7), the dHOFC of the two high-level functional

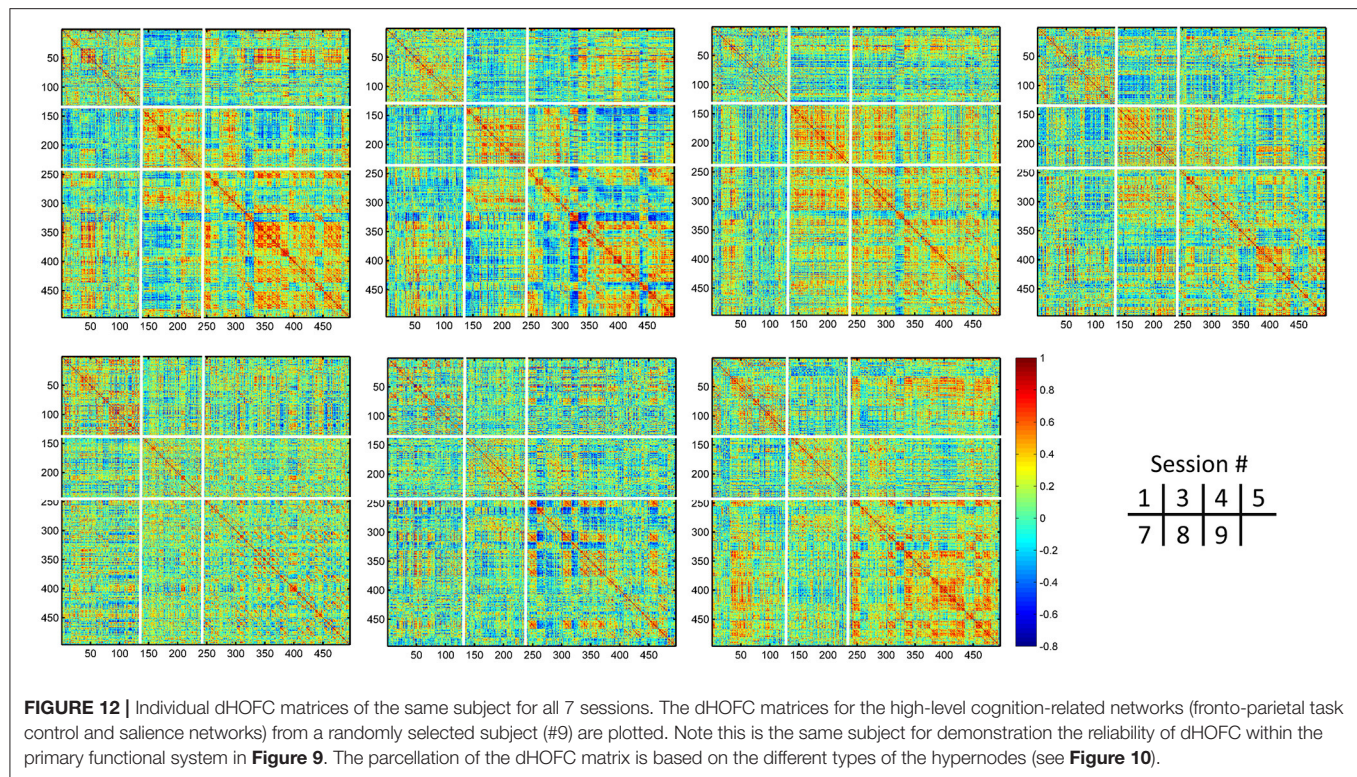
systems show meaningful and visually detectable and systematic differences in the test-retest reliability, which becomes more prominent when only looking at the connections with fair or better reliability (Figure 10C). That is, for between-network dHOFC, their connectivity strength is weak, and their connectivity reliability is also poor, while the other two types of the dHOFC have both greater strength and better reliability. At the subject level, Figure 12 shows the dHOFC matrices derived from all the seven rs-fMRI sessions of the same subject (subject #9), with a roughly stable pattern.

Compare with the dHOFC in the primary functional system, those in the high-level functional systems have *better* reliability for several (mainly the between-network and the modulatory) connections while lower reliability for several other (mainly within-network) connections (Figure 8C). When only looking at the strong and putative connections, dHOFCs in the high-level system are more reliable (Figure 8D), with more (66.4%) connections characterized as fairly reliable or better.

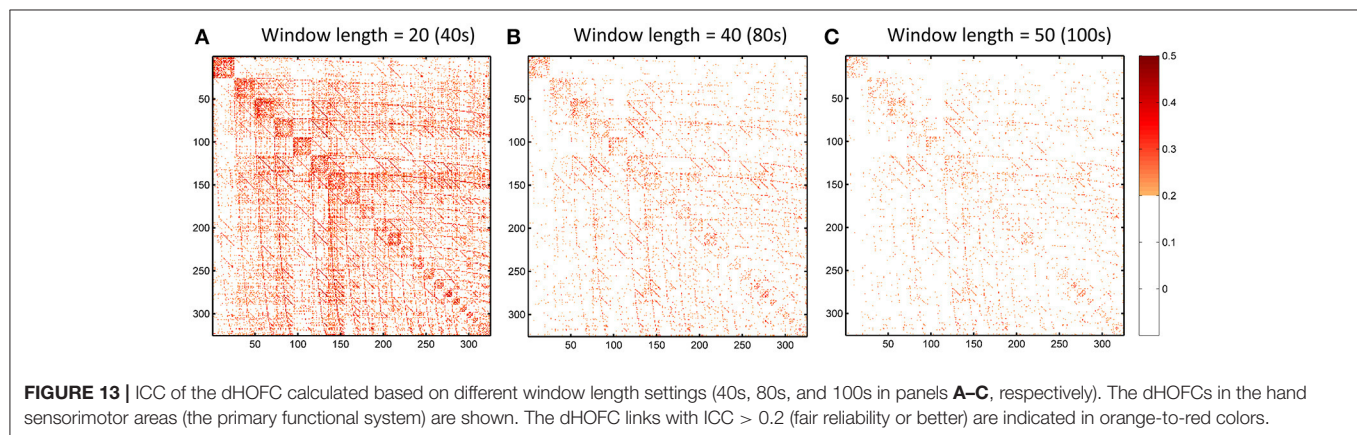
## Sliding Window Length Significantly Affects dHOFC Reliability

We further show how the length of sliding window (or the window width), an important parameter for both dynamic LOFC and dHOFC analyses, will affect dHOFC reliability. The ICC matrices based on different window lengths of 40, 80, and 100 s are shown in Figure 13. Together with the main dHOFC ICC result using a window length of 60 s (Figure 7C), we, for the first time, revealed that the setting of sliding-window length significantly affected the dHOFC test-retest reliability. Shorter





**FIGURE 12 |** Individual dHOFc matrices of the same subject for all 7 sessions. The dHOFc matrices for the high-level cognition-related networks (fronto-parietal task control and salience networks) from a randomly selected subject (#9) are plotted. Note this is the same subject for demonstration the reliability of dHOFc within the primary functional system in **Figure 9**. The parcellation of the dHOFc matrix is based on the different types of the hypernodes (see **Figure 10**).



**FIGURE 13 |** ICC of the dHOFc calculated based on different window length settings (40s, 80s, and 100s in panels **A–C**, respectively). The dHOFcs in the hand sensorimotor areas (the primary functional system) are shown. The dHOFc links with ICC > 0.2 (fair reliability or better) are indicated in orange-to-red colors.

window length generated better reliable dHOFc. Based on the ICC values, the window length of 40, 60, 80, and 100 produces 77.3, 49.5, 30.3, and 18.9% fairly reliable (ICC > 0.2) dHOFc links among all the strong dHOFc links (**Figure 14**).

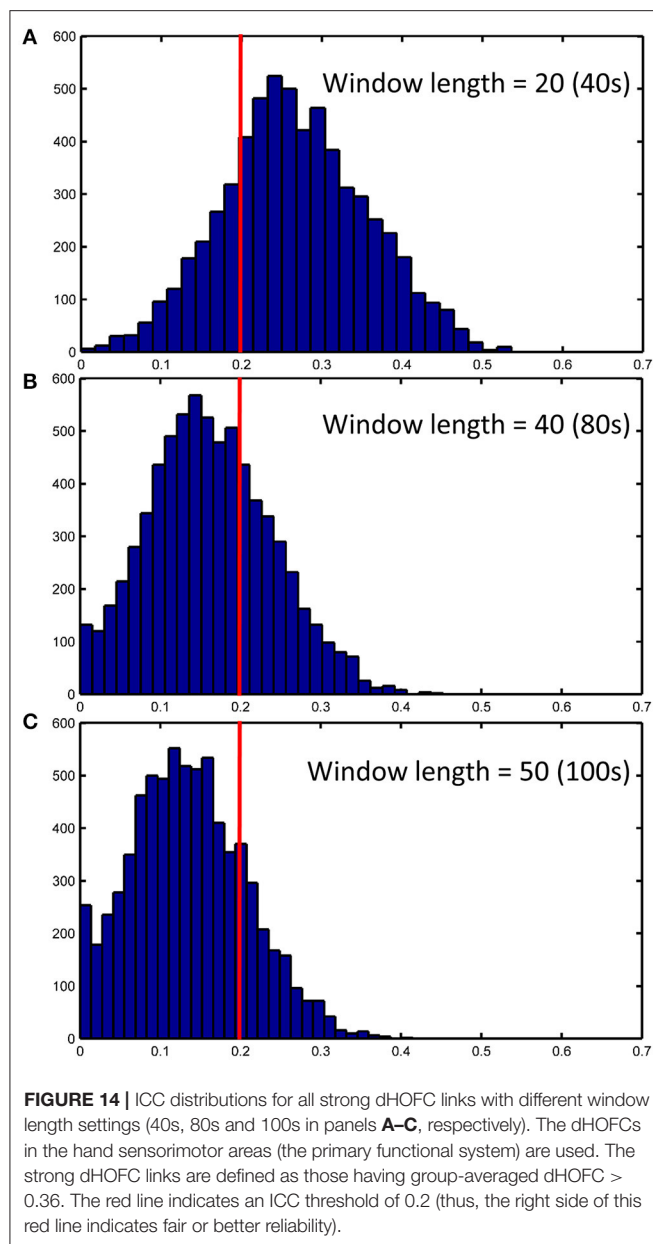
We still chose the window length setting of 60 s as the main dHOFc result, because the previous comprehensive simulated experiments have shown that a too short window length setting may cause limited sample size in the calculation of dynamic LOFC within each window and could *overestimate* the dynamic FC. In other words, with small window length, we may inflate the window-based LOFC estimations and increase the possibility of type-I error in finding the significant dynamic FC. This will, in turn, compromise the dHOFc calculation because dHOFc is based on the second round of the correlation analysis on the

dynamic LOFC time series, and the overestimated LOFC changes may cause bias in the following dHOFc calculation.

## DISCUSSION

### General Discussion

In this paper, we assessed the test-retest reliability of all existing HOFC (high-order FC) metrics extracted from young healthy adults. **Table 1** summarized all definitions and potential biological meanings for all the HOFC metrics involved. We found that, in general, all the methods have acceptable test-retest reliability. Please also see **Table 1** for a summary of all reliability assessment results, and **Figure 11** for specially summarized connectivity strength and reliability characteristics



for different types of the dHOFC links. The goal of presenting such reliability analysis results is to obtain new knowledge based on the reliability analysis for better understanding the biological meaning of different types of HOFC, deriving guidance for future HOFC studies, and accelerating wider clinical applications using HOFC.

To our best knowledge, there is no such reliability study before on the HOFC metrics. We note that there is a recent study investigating the reproducibility of dynamics LOFC-based brain transient status detection across different data sets (Abrol et al., 2016), which suggested that a few transient LOFC patterns are reproducible; but this study didn't go further to analyze the high-order FC and its reproducibility. Here, we use a dedicated dataset with amply repeated scans and sample size to produce an accurate

estimation of the HOFC's test-retest reliability. We believe that this novel test-retest reliability studies on such state-of-the-art connectomic metrics could have instructive meanings toward understanding how the human brain is functionally organized.

## Why Focusing on HOFC's Test-Retest Reliability?

Besides characterizing pair-wise temporal synchronization of rs-fMRI BOLD signals and building such traditional LOFC brain networks, researchers are also eager to look for the methods that can capture more complex functional organization of the human brain, i.e., HOFC. The HOFC may have more generalized definition, as long as it captures more complex functional organization, e.g., hierarchical FC architectures (Cordes et al., 2002), modularity/rich-club from deep analysis to the LOFC networks (van den Heuvel and Sporns, 2011), hypergraph consisting of hypernodes and hyperlinks (Jie et al., 2016), cross-modality association (Honey et al., 2009) and context-sensitive divergence (Hermundstad et al., 2013), but here we only focus on the narrowly defined HOFC metrics, which are the metrics that have been explicitly proposed to be "high order" based on "correlation's correlation." Of note, a previous study first calculated dynamic local LOFC and then calculated regional covariance of the regional dynamic local LOFC time series (Deng et al., 2016), which is somewhat also based on the correlation of correlations. We think that this method is more like the dHOFC, but still characterizing the pairwise relationship since the first round of correlations are collapse into regional time series. Although this paper did not provide reliability or reproducibility results, it did show a highly structured high-level functional organization. Another recent work also calculated topographical LOFC profiles (Zhang J. et al., 2016) and their dynamics, but they further calculated the similarity among each brain region's topographical LOFC profiles across time to define a variation-based metric for each brain region. Therefore, they did not use inter-regional topographical similarity to define HOFC but rather using intra-regional time varying topographical information to capture brain function. All these state-of-the-art studies have indicated that characterizing high-order brain functional organization is the common research interest and also a hot topic. Therefore, test-retest reliability on these HOFC metrics is highly necessary.

Of all the studies which explicitly defined or adopted HOFC, the *tHOFC* characterizes similarity of the topographical LOFC profile between any two brain regions; the *aHOFC* defines a different pair-wise topographic profile similarity which is actually a cross-level (i.e., the modulation between the low-level and the high-level FC organizations among brain regions) HOFC measurement; and the *dHOFC* defines an even more complex, i.e., four region-based functional relationships by adopting dynamic LOFC profiles, where the covariance of two LOFC dynamic time series naturally reflect a modulatory interaction. Based on the network belongingness of every four brain regions, we have the opportunity to *explicitly* define different types of high-level modulation rather than just *inherently* considering such high-level functional coherence like most of the existing



LOFC dynamics studies on brain “status.” In summary, all the HOFC metrics are methodologically innovative and state-of-the-art. Most importantly, these metrics may characterize different aspects of biologically meaningful functional organization architecture, which is systematically different from LOFC. In order to further validate this argument, we need to assess their reliability to add further support to this hypothesis.

## Reasons and Factors That May Cause Variation in HOFC Reliability

There are several factors that could cause the difference in reliability among the HOFC metrics. Next, we will discuss the possible contributing factors that may lead to such differences in HOFC reliability, which include:

- *Various types of noise and artifacts* (e.g., cardiac pulsation and head motion) in the rs-fMRI data may interfere LOFC estimation (Chang and Glover, 2009; Power et al., 2014), which often leads to overestimated LOFC due to the structured and spatially overspread noise. Since HOFC is calculated based on LOFC, the noise and artifacts can interfere HOFC as well, although the effect will be different for HOFC compared to LOFC.
- *The complexity of the algorithm.* First, noise problem can be exaggerated when there are more operations (“correlation’s correlation”) applied on the data. In other words, the noise-induced error may propagate and increase by further steps of correlation analysis. Similar substantial reliability reduction has been witnessed in the previous reliability studies on the graph-theoretic analysis-derived network properties from the LOFC network (Wang et al., 2011). Moreover, dHOFC has several freely estimable parameters, one of which is the sliding window length. From the **Figure 7C** and **Figures 13–14**, we can see that window length indeed affects dHOFC reliability (with the shorter window length leading to more reliable dHOFC). Therefore, if using dHOFC to detect potential disease biomarkers, we may *not only* have a risk in the reliability reduction due to the computational complexity, *but also* have to decide the optimal parameter setting. On the other hand, tHOFC and aHOFC do not have free parameters as long as the region-averaged rs-fMRI signals are obtained.
- *The HOFC strength itself.* An interesting finding for all types of HOFC (and the LOFC previously) is that, generally, connections with greater strength may be more reliable, and vice versa. Such a phenomenon is more prominent for the dHOFC. This may be because weaker connectivities are more likely to be affected by the noise and artifacts. Of note, it is difficult to determine the threshold for weak/strong dHOFC as the parametric testing, such as the *t*-tests tends to overestimate the “significant” dHOFC, i.e., even a small dHOFC could be significantly large due to a large number of sliding windows and the statistical dependence among nearby windows. Based on the suggestion of previous dynamic FC study (Leonardi and Van De Ville, 2015), even a large dynamic FC could be purely induced by noise. Thus, we use a relatively large threshold to determine strong dHOFC ( $>0.36$ ). In future, non-parametric analysis, such as permutation test can also be used to generate the “null model” of dHOFC and determine which is significantly strong. Here, to make fair comparison among different window lengths, we use the predefined threshold of dHOFC  $> 0.36$  to identify strong dHOFC links. However, such a rule does not apply to several LOFC, tHOFC and aHOFC links, such as the connections among the DMN, FPN, and SN; interestingly, their weak connectivities are astonishingly stable across repeated scans (see **Figures 4–6**).
- *The subject’s varying status.* Recently, studies on brain LOFC dynamics have revealed that the brain functional network is not a static but a continuously changing system (Hutchison et al., 2013; Calhoun et al., 2014; Preti et al., in press). Decompositions to the LOFC spatiotemporal dynamics have revealed a few instantaneous LOFC network patterns that occur from time to time and switch to each other with certain transformation probability, which may represent different brain “statuses” (Allen et al., 2014). The occurrence frequency and the dwelling time of the status may be substantially different in different rs-fMRI sessions; moreover, several statuses may not occur at all during a particular scanning session (Abrol et al., 2016). Such a variation could be larger if the interval of the repeated scans is longer. Although our test-retest data were acquired within a month, such a period will still allow unneglectable changes in subject’s physical and mental conditions (e.g., drowsiness) to happen and lead to differences in status switching and their occurrence frequency. Since HOFC is proposed to measure high-level brain functional architecture, a small variation may still affect its reliability. We think that dHOFC could be affected more because this metric *per se* is directly estimated on the basis of dynamic analysis.
- *Head motion.* Although we had stringently controlled head motion effect according to the strict data inclusion criteria, head motion can still be a source of the reduced reliability. We believe that the head motion will have more effect on dHOFC estimation because sliding window-based analysis uses fewer samples to conduct temporal correlation, such that the robustness to the head motion-related artifacts could drop. This argument has been supported by both previous studies (Laumann et al., 2016) and the leftward shift of the ICC histogram from **Figure 4** to **Figures 8A,C**.
- *Other unavoidable factors*, such as the changing condition and status of the MRI scanner, will likely to affect the test-retest reliability.

## Biological Meaning of HOFC Indicated by Reliability Assessment Result

Based on reliability analysis, we may have a chance to revisit the underlying biological meaning of the HOFC. Our result has four major implications. First, we examined which HOFC links have reliability gain when comparing tHOFC (and aHOFC, with the similar result) and LOFC. We found that the links with better reliability than those of the LOFC are highly *structured* with highly *specified* anatomical location. Most of them are the inter-network connectivities between the high-level and the primary functional systems (**Figure 5C**). The primary systems are the

sensorimotor and visual areas, while the high-level functional systems include the DMN, FPN and SN, which have a perfect agreement with so-called “triple networks” (Menon, 2011). The triple networks have been proposed to be responsible for high-order cognitive functions, such as task control, attention, self-awareness, etc. Meanwhile, many neurological and psychiatric diseases (such as AD and schizophrenia) have abnormalities commonly located at such three networks. The increment of test-retest reliability for the tHOFC and aHOFC indicates that the tHOFC can *more reliably* estimate the connections between the high-level and low-level brain networks. These results support the previous finding using the tHOFC, that is, the topographical LOFC profile can suppress noise in several links (Zhang H. et al., 2016). Because these “reliability enhanced” links are mostly the weak connections, if the noise level is not favorable, these connectivities cannot be used for biomarker detection and disease classification due to the noise-induced reliability reduction. Our result suggests that tHOFC and aHOFC could be more suitable for such studies if these particular (although weak) connections are of interest. From another viewpoint, this result indicates that tHOFC and aHOFC are able to model the feedforward and feedback functional relationships, which may reflect information exchange between the high-level and the primary areas.

Second, after visualizing the extent of the reliability gain for each brain region, we found that the mostly benefiting nodes are the medial frontal regions in the DMN and the lateral frontal regions in the FPN and SN (Figure 6B), indicating the importance of these areas in such a cross-level information exchange. Moreover, we, for the first time, show that these medial and lateral frontal regions could be functionally important based on the reliability gain against LOFC. In the future, more efforts should be made on these putative but weak high-order cross-level interactions between the triple networks to the primary functional areas. The importance of such a type of HOFC links could be diminished if only traditional LOFC is used. Based on this finding, we have a further tentative assumption that, for the neurodegenerative diseases, such as AD and the neurodevelopmental disorders, such as autism spectrum disorder, at the very beginning, the pathological attack (such as neurofibrillary tangles and amyloid beta-peptide deposition in AD) could first occur at these frontal areas (Braak and Braak, 1991). At this early stage, there is usually no significant cognitive abnormalities for the patients. We hypothesize that it is such high-order cross-level feedback and feedforward connections that could be affected at this period, and the high-level to primary information exchanges are likely to be already changed. Traditional LOFC is less reliable for such connections, thus early detection is difficult and less sensitive. If tHOFC, especially aHOFC, is used as connectivity-based metrics, we could have much larger chances to detect such early but subtle changes.

Third, as shown in Figure 7A, the group-level dHOFC matrix in the hand sensorimotor system shows the prominent modular structure (i.e., small blocks along the main diagonal of the dHOFC matrix). The dHOFC strength within modules is higher than that between modules. Further investigation revealed that the higher dHOFC in each of the block or module had

a brain region acting as a common driving source, so that any dLOFC links sharing the same driving region had quite similar dynamic patterns along time. For example, the dHOFC among dLOFC<sub>12</sub>, dLOFC<sub>13</sub>, ..., dLOFC<sub>1R</sub> (which all share the region #1) are stronger than the dHOFC between dLOFC<sub>12</sub> and dLOFC<sub>34</sub> (because they share no region). This could indicate the organization architecture of the dHOFC in the sensorimotor system; that is, many strong dHOFC hypernodes (dLOFC links) share a common driving source from a single brain region and this can form a “star-shaped” local topological structure. This star-shaped cluster could be the basic unit for high-level brain functional organization. Traditionally, it is impossible to reveal such a high-level spatiotemporal organization architecture.

Fourth, in this study, we have included two high-level functional systems (FPN and SN) for dHOFC analysis. The reliability matrix has shown a structured and inherently well-organized pattern (Figures 10B,C), consistent with the pattern of the dHOFC strength (Figure 10A). Based on the complexity of the dHOFC's definition (involving four regions for characterizing a hyperlink of dHOFC), we have further separated dHOFC hyperlinks into the *within-network*, *between-network* and, completely new, *modulatory* types (containing hybrid inter-network connection(s) as hypernode(s); see Figure 11C). Note that, previously, there is no study on the third type of the connections. We found that the between-network dHOFC, which consists of two intra-network hypernodes for each of the two networks, respectively, are nearly zero (weak connections). This result indicates that the two high-level functional systems, as shown by their respective nearly *uncorrelated* dynamic connectivity profiles, may work quite *independently*. The reliability of such type of dHOFC is also poor, meaning that such weak high-order connectivities are prone to be affected by noise. However, the within-network dHOFC, similar to previous findings for the within-network LOFC, is relatively strong and much more reliable than the between-network dHOFC or LOFC. The most interesting finding is that the *modulatory* dHOFC, especially when both hypernodes are inter-network connections (with the two ROIs of each hypernode belonging to two different functional systems), are also relatively strong with better reliability. This result indicates that the brain functional organization is not in a one-by-one or pairwise manner. The two high-level functional networks may *not only* interact with each other via pairwise LOFCs, *but also* have extensive and deep modulatory relationship in a high-order way. Such a high-order relationship can be further divided into two subtypes (Figure 11C), reflecting different modulatory interactions. In this sense, the dHOFC may be able to model more complex interactions among the brain networks that cannot be easily modeled using the traditional inter-network LOFC.

Finally, as shown by Figure 10A, there are strong off-diagonal connections for the case 1 of the modulatory dHOFC, indicating that the two high-level cognitive function-related networks indeed communicate with each other more in a more complex manner than any LOFC can capture. However, when compared the connectivity strength of the similar off-diagonal LOFC (i.e., the mean inter-network LOFC between the FPN and SN), for the strongest 50 connections, we found that the dHOFC values are significantly ( $p < 0.0001$ ) larger than

LOFC (Figure 15). Moreover, such a type of the dHOFCs has acceptable reliability. Therefore, we propose that the modulatory dHOFC with each of the two hypernodes connecting with both networks (see Figure 11C, case 1) can better characterize inter-network functional association via complex high-order modulatory interactions. In the future, this type of dHOFC could be specifically selected as features to search for potential biomarkers of brain disease if inter-network connectivity is the main target.

## Suggestions and Guideline to Future HOFC Study

Based on the findings on HOFC reliability, we give several suggestions to future studies that focus on high-order brain functional organization or its modulation by different experimental states or diseases:

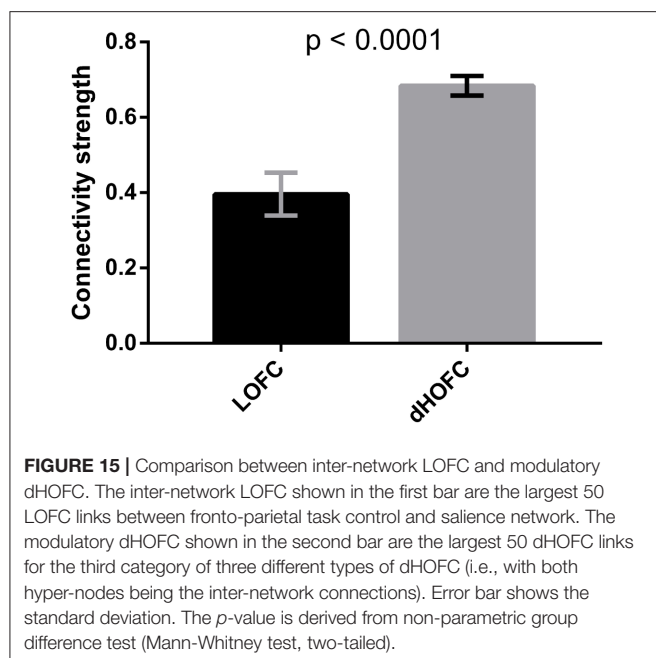
- The tHOFC and aHOFC may have less reliability for within-network connections than LOFC, but are still moderately reliable. If interested in the within-network connectivity, it's better to use LOFC.
- The tHOFC and aHOFC have higher reliability for the weak between-network connections. If interested in such type of connections, it's better to use tHOFC or aHOFC. The tHOFC seems to have *both* acceptable reliability for within-network connectivity *and* more reliable between-network connectivity. Due to such a trade-off, the tHOFC may be more suitable for the exploratory whole-brain network analysis, including both within- and between-network connectivities.
- The aHOFC is especially reliable for modeling high-level feedback and feedforward relationship between the high-level cognition-related and the primary functional systems, which is suitable for studies on top-down or bottom-up connectivities.
- dHOFC implementation should be careful due to its lower reliability compared with that for static LOFC or HOFC. However, within-network and modulatory dHOFC or relatively strong dHOFC are still sufficiently reliable. Future dHOFC studies should focus on these dHOFC links.
- Data processing parameters, such as sliding-window length should be carefully determined for dHOFC calculation. Too small window length may be less robust to noise and may lead to spuriously high “reliability.” A window length of 60 s is a recommended choice for robust dHOFC estimation with adequate reliability.
- For early diagnosis studies, in order to increase detection sensitivity, it's better to choose a certain type of HOFC to characterize the subtle connectivity abnormalities. For example, weak connections might be more likely to be affected by the pathological attacks than strong connections; all the HOFC metrics have satisfactory reliability for the weak connections.
- While static LOFC does not have adequate sensitivity for biomarker detection, modulatory dHOFC, especially the case 1, could be an alternative approach to estimate those deeply inherent inter-network interactions.

## LIMITATIONS AND FUTURE WORKS

First, in this paper, we only focused on the reliability assessment of the connectivity strength without going further to assess the reliability of graph-theoretical analysis-based network properties, which we think deserves a dedicated research after more suitable complex network construction approach for the HOFC is proposed. Second, this paper is dedicated to investigating HOFC reliability, the further study on the biological relationship (validity) between the HOFC strength and neurocognitive measurements or disease states are not our main goal and will be investigated in the future. Third, the test-retest reliability with varied inter-scan interval (especially the intra-session reliability) will better disentangle the mixed effect of influencing factors on the HOFC reliability. This is especially important for the dHOFC, because it is based on the dynamic LOFC which is theoretically expected to be fluctuating. Although the dHOFC calculates the coordination of the dynamic LOFC, making this HOFC metric more like a measurement of “trait” than “state,” a dedicated study on how the changing brain “state” may affect the trait characterization is highly required. Finally, due to the increased dimensionality, we only calculate dHOFC for a few functional systems. In the future, the better algorithm needs to be proposed to overcome such a limitation and extend our understanding of the neurobiological meaning of the dHOFC in the whole-brain level.

## AUTHOR CONTRIBUTIONS

HZ, YZ, and XC analyzed the data. HZ drafted the manuscript. DS conceived this work, led the project, and revised the paper.



## FUNDING

This work was supported in part by NIH grants (EB006733, EB008374, MH100217, MH108914, AG041721, AG049371, AG042599, AG053867, EB022880, MH110274), and National Natural Science Foundation of China (NSFC) Grant (61473190).

## REFERENCES

- Abrol, A., Chaze, C., Damaraju, E., and Calhoun, V. D. (2016). "The chronnectome: Evaluating replicability of dynamic connectivity patterns in 7500 resting fMRI datasets," in *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Orlando, FL), 5571–5574.
- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107
- Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–259. doi: 10.1007/BF00308809
- Calhoun, V. D., Miller, R., Pearlson, G., and Adali, T. (2014). The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015
- Chang, C., and Glover, G. H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *Neuroimage* 47, 1448–1459. doi: 10.1016/j.neuroimage.2009.05.012
- Chen, B., Xu, T., Zhou, C., Wang, L., Yang, N., Wang, Z., et al. (2015). Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month. *PLoS ONE* 10:e0144963. doi: 10.1371/journal.pone.0144963
- Chen, X., Zhang, H., Gao, Y., Wee, C. Y., Li, G., and Shen, D. (2016a). High-order resting-state functional connectivity network for MCI classification. *Hum. Brain Mapp.* 37, 3282–3296. doi: 10.1002/hbm.23240
- Chen, X., Zhang, H., and Shen, D. (2016b). *Ensemble Hierarchical High-Order Functional Connectivity Network for MCI Classification*. Athens: MICCAI.
- Cordes, D., Haughton, V., Carew, J. D., Arfanakis, K., and Maravilla, K. (2002). Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magn. Reson. Imaging* 20, 305–317. doi: 10.1016/S0730-725X(02)00503-9
- Deng, L., Sun, J., Cheng, L., and Tong, S. (2016). Characterizing dynamic local functional connectivity in the human brain. *Sci. Rep.* 6:26976. doi: 10.1038/srep26976
- Fox, M. D., and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* 4:19. doi: 10.3389/fnsys.2010.00019
- Fox, M. D., Halko, M. A., Eldaief, M. C., and Pascual-Leone, A. (2012). Measuring and manipulating brain connectivity with resting state functional connectivity magnetic resonance imaging (fcMRI) and transcranial magnetic stimulation (TMS). *Neuroimage* 62, 2232–2243. doi: 10.1016/j.neuroimage.2012.03.035
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13, 5–14. doi: 10.1038/jcbfm.1993.4
- Guimerà, R., Danon, L., Diaz-Guilera, A., Giral, F., and Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 68:065103. doi: 10.1103/PhysRevE.68.065103
- Hermundstad, A. M., Bassett, D. S., Brown, K. S., Aminoff, E. M., Clewett, D., Freeman, S., et al. (2013). Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6169–6174. doi: 10.1073/pnas.1219562110
- Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040. doi: 10.1073/pnas.0811168106
- Horrigan, S. K., Courville, P., Sampey, D., Zhou, F., and Cai, S. (2017). Replication study: melanoma genome sequencing reveals frequent PREX2 mutations. *eLife* 6:e21634. doi: 10.7554/eLife.21634
- Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378. doi: 10.1016/j.neuroimage.2013.05.079
- Jie, B., Wee, C. Y., Shen, D., and Zhang, D. (2016). Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* 32, 84–100. doi: 10.1016/j.media.2016.03.003
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., et al. (2016). On the stability of BOLD fMRI correlations. *Cereb. Cortex*. doi: 10.1093/cercor/bhw265. [Epub ahead of print].
- Leonardi, N., and Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage* 104, 430–436. doi: 10.1016/j.neuroimage.2014.09.007
- Liu, L., Zhang, H., Rekik, I., Chen, X., Wang, Q., and Shen, D. (2016). *Outcome Prediction for Patients with High-Grade Gliomas from Brain Functional and Structural Networks*. Athens: MICCAI.
- Mao, D., Ding, Z., Jia, W., Liao, W., Li, X., Huang, H., et al. (2015). Low-frequency fluctuations of the resting brain: high magnitude does not equal high reliability. *PLoS ONE* 10:e0128117. doi: 10.1371/journal.pone.0128117
- Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn. Sci.* 15, 483–506. doi: 10.1016/j.tics.2011.08.003
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048
- Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N., and Petersen, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron* 79, 798–813. doi: 10.1016/j.neuron.2013.07.035
- Preti, M. G., Bolton, T. A., and Van De Ville, D. (in press). The dynamic functional connectome: state-of-the-art and perspectives. *Neuroimage*. doi: 10.1016/j.neuroimage.2016.12.061
- Shehzad, Z., Kelly, A. M., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., et al. (2009). The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229. doi: 10.1093/cercor/bhn256
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Somandepalli, K., Kelly, C., Reiss, P. T., Zuo, X. N., Craddock, R. C., Yan, C. G., et al. (2015). Short-term test-retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. *Dev. Cogn. Neurosci.* 15, 83–93. doi: 10.1016/j.dcn.2015.08.003
- van den Heuvel, M. P., and Sporns, O. (2011). Rich-club organization of the human connectome. *J. Neurosci.* 31, 15775–15786. doi: 10.1523/JNEUROSCI.3539-11.2011
- Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009
- Wang, J. H., Zuo, X. N., Gohel, S., Millham, M. P., Biswal, B. B., and He, Y. (2011). Graph theoretical analysis of functional brain networks: test-retest evaluation



- on short- and long-term resting-state functional MRI data. *PLoS ONE* 6:e21976. doi: 10.1371/journal.pone.0021976
- Yan, C. G., and Zang, Y. F. (2010). DPARSF: a MATLAB toolbox for "Pipeline" data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011
- Yu, R., Zhang, H., An, L., Chen, X., Wei, Z., and Shen, D. (2017). Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification. *Hum. Brain Mapp.* 38, 2370–2383. doi: 10.1002/hbm.23524
- Zalesky, A., and Breakspear, M. (2015). Towards a statistical test for functional connectivity dynamics. *Neuroimage* 114, 466–470. doi: 10.1016/j.neuroimage.2015.03.047
- Zhang, H., Chen, X., Shi, F., Li, G., Kim, M., Giannakopoulos, P., et al. (2016). Topographical information-based high-order functional connectivity and its application in abnormality detection for mild cognitive impairment. *J. Alzheimers Disease* 54, 1095–1112. doi: 10.3233/JAD-160092
- Zhang, H., Duan, L., Zhang, Y. J., Lu, C. M., Liu, H., and Zhu, C. Z. (2011a). Test-retest assessment of independent component analysis-derived resting-state functional connectivity based on functional near-infrared spectroscopy. *Neuroimage* 55, 607–615. doi: 10.1016/j.neuroimage.2010.12.007
- Zhang, H., Zhang, Y. J., Duan, L., Ma, S. Y., Lu, C. M., and Zhu, C. Z. (2011b). Is resting-state functional connectivity revealed by functional near-infrared spectroscopy test-retest reliable? *J. Biomed. Opt.* 16:067008. doi: 10.1117/1.3591020
- Zhang, J., Cheng, W., Liu, Z., Zhang, K., Lei, X., Yao, Y., et al. (2016). Neural, electrophysiological and anatomical basis of brain-network variability and its characteristic changes in mental disorders. *Brain: a journal of neurology*, 139, 2307–2321. doi: 10.1093/brain/aww143
- Zhang, Y., Zhang, H., Chen, X., and Shen, D. (2017). *Hybrid High-order Functional Connectivity Networks Using Resting-State Functional MRI for Mild Cognitive Impairment Diagnosis*. Scientific Report.
- Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific Data* 1:140049. doi: 10.1038/sdata.2014.49
- Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010). Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. *Neuroimage* 49, 2163–2177. doi: 10.1016/j.neuroimage.2009.10.080

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Zhang, Chen, Zhang and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparison of IVA and GIG-ICA in Brain Functional Network Estimation Using fMRI Data

Yuhui Du<sup>1,2\*</sup>, Dongdong Lin<sup>1</sup>, Qingbao Yu<sup>1</sup>, Jing Sui<sup>1,3</sup>, Jiayu Chen<sup>1</sup>,  
Srinivas Rachakonda<sup>1</sup>, Tulay Adali<sup>4</sup> and Vince D. Calhoun<sup>1,5</sup>

<sup>1</sup> The Mind Research Network, Albuquerque, NM, USA, <sup>2</sup> School of Computer and Information Technology, Shanxi University, Taiyuan, China, <sup>3</sup> Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, <sup>4</sup> Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD, USA, <sup>5</sup> Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

## OPEN ACCESS

### Edited by:

Bharat B. Biswal,  
Rowan University School of  
Osteopathic Medicine, USA

### Reviewed by:

Xu Lei,  
Southwest University, China  
Veena A. Nair,  
University of Wisconsin-Madison, USA  
Han Zhang,  
University of North Carolina at Chapel  
Hill, USA

### \*Correspondence:

Yuhui Du  
ydu@mrn.org

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 31 December 2016

**Accepted:** 25 April 2017

**Published:** 19 May 2017

### Citation:

Du Y, Lin D, Yu Q, Sui J, Chen J,  
Rachakonda S, Adali T and  
Calhoun VD (2017) Comparison of IVA  
and GIG-ICA in Brain Functional  
Network Estimation Using fMRI Data.  
Front. Neurosci. 11:267.  
doi: 10.3389/fnins.2017.00267

Spatial group independent component analysis (GICA) methods decompose multiple-subject functional magnetic resonance imaging (fMRI) data into a linear mixture of spatially independent components (ICs), some of which are subsequently characterized as brain functional networks. Group information guided independent component analysis (GIG-ICA) as a variant of GICA has been proposed to improve the accuracy of the subject-specific ICs estimation by optimizing their independence. Independent vector analysis (IVA) is another method which optimizes the independence among each subject's components and the dependence among corresponding components of different subjects. Both methods are promising in neuroimaging study and showed a better performance than the traditional GICA. However, the difference between IVA and GIG-ICA has not been well studied. A detailed comparison between them is demanded to provide guidance for functional network analyses. In this work, we employed multiple simulations to evaluate the performances of the two approaches in estimating subject-specific components and time courses under conditions of different data quality and quantity, varied number of sources generated and inaccurate number of components used in computation, as well as the presence of spatially subject-unique sources. We also compared the two methods using healthy subjects' test-retest resting-state fMRI data in terms of spatial functional networks and functional network connectivity (FNC). Results from simulations support that GIG-ICA showed better recovery accuracy of both components and time courses than IVA for those subject-common sources, and IVA outperformed GIG-ICA in component and time course estimation for the subject-unique sources. Results from real fMRI data suggest that GIG-ICA resulted in more reliable spatial functional networks and yielded higher and more robust modularity property of FNC, compared to IVA. Taken together, GIG-ICA is appropriate for estimating networks which are consistent across subjects, while IVA is able to estimate networks with great inter-subject variability or subject-unique property.

**Keywords:** functional magnetic resonance imaging (fMRI), brain functional networks, independent component analysis (ICA), group information guided ICA (GIG-ICA), independent vector analysis (IVA)

## INTRODUCTION

There is a rapidly increasing interest in using functional magnetic resonance imaging (fMRI) data to characterize brain functional networks. Independent component analysis (ICA), a data-driven method, has been widely used to analyze fMRI data without requiring the definition of brain regions (or nodes). Spatial ICA (sICA) (McKeown et al., 1998), a popular method for analyzing fMRI data, decomposes fMRI data into a linear mixture of spatially independent components (ICs) some of which are subsequently identified as brain functional networks. Despite success of ICA in fMRI data analyses, ICA faces some challenges. The order of resulting ICs from individual-subject ICA is arbitrary, increasing the difficult to establish correspondence among ICs estimated from different subjects. Another issue is that the estimated ICs include not only meaningful functional networks, but also various artifact-related ICs resulting from imaging and non-neural physiological activity. In addition, the number of sources is unknown, so the number of ICs always needs to be estimated (Li et al., 2007); however the numbers estimated using different criteria are varied (Zuo et al., 2010). These shortcomings of ICA bring difficulties to multiple-subject fMRI data analyses, especially when shared networks across subjects are expected for subsequent group analyses. To address the problems, group independent component analysis (GICA) and independent vector analysis (IVA) have been proposed.

Several GICA frameworks have been proposed for fMRI data analyses, including using the spatial concatenation (Svensén et al., 2002), temporal concatenation (Calhoun et al., 2001, 2009; Beckmann et al., 2009) and tensor organization (Beckmann and Smith, 2005) strategies. Relative to ICA on each individual-subject's data, one advantage of GICA is that it can build direct correspondence of ICs across subjects. Among the GICA approaches, the temporal concatenation based method is most widely used. This approach first estimates the group-level ICs by performing ICA on the time points-concatenated fMRI data of all subjects, and then back-reconstructs the subject-specific ICs mainly using principal component analysis (PCA) based (Calhoun et al., 2001; Erhardt et al., 2011) or regression based (Beckmann et al., 2009; Erhardt et al., 2011) algorithms. More recently, in order to improve the accuracy of the subject-specific ICs estimation, group information guided independent component analysis (GIG-ICA) (Du and Fan, 2013; Du et al., 2015c, 2016a) as a variant of GICA has been proposed. GIG-ICA estimates the subject-specific ICs under the guidance of the group-level ICs by using a multi-objective function optimization framework, which simultaneously optimizes the independence among multiple ICs of each subject and the correspondence between each group-level IC and the associated subject-specific IC. The optimization of independence of multiple components for each subject's data benefits yielding accurate subject-specific functional networks. The optimization of correspondences between one group-level IC and the associated subject-specific ICs guarantees that the obtained individual networks have the same physiological meanings and then are comparable across subjects. Therefore, compared to the traditional PCA-based and regression-based back-reconstruction techniques that ignore

independence of individual ICs to some extent, GIG-ICA can yield more accurate individual networks and the associated time courses (Du and Fan, 2013; Du et al., 2016a) while still preserving correspondence and comparability of shared networks across different subjects. Notably, GIG-ICA can estimate individual networks for new data by utilizing *a priori* spatial network maps as guidance. Our previous work (Du et al., 2015c) has shown the promise of GIG-ICA to estimate spatial functional networks from fMRI data. By applying GIG-ICA to resting-state fMRI data (Du et al., 2014, 2015c), we found potential biomarkers in several functional networks for distinguishing schizophrenia, bipolar disorder and schizoaffective disorder. In addition, GIG-ICA (Du et al., 2015b, 2017) has the ability to extract functional connectivity states from time-varying functional connectivity (Calhoun et al., 2014; Du et al., 2015a, 2016b). Our work (Du et al., 2015b, 2017) revealed interesting biomarkers of schizophrenia, bipolar disorder and schizoaffective disorder in multiple connectivity states. In this paper, we only focus on the application of GIG-ICA in estimating functional networks from fMRI data.

Independent vector analysis (IVA), an alternative method to achieve an independent decomposition (Adali et al., 2014), has been applied to analyzing fMRI data of schizophrenia patients (Gopal et al., 2016) as well as patients with stroke (Laney et al., 2015a,b). The approach models both the independence of individual components and the dependence of similar components across subjects. Several advancements of IVA have been made for achieving reliable source separation for linearly dependent Gaussian and non-Gaussian sources (Anderson et al., 2010, 2014; Dea et al., 2011; Li et al., 2011; Adali et al., 2014; Boukouvalas et al., 2015). Among those, IVA-GL (IVA with multivariate Gaussian source component vectors plus IVA with Laplace source component vectors), which is a combination of two IVA algorithms, IVA with multivariate Gaussian component vectors (IVA-G) (Anderson et al., 2012) and IVA with multivariate Laplace component vectors (IVA-L) (Lee et al., 2008), provides an attractive tradeoff in terms of complexity and performance and has been the algorithm used in previous applications of IVA to fMRI data (Laney et al., 2015a,b; Gopal et al., 2016). Previous studies (Dea et al., 2011; Ma et al., 2013; Michael et al., 2014; Laney et al., 2015a,b) compared IVA and traditional GICA under different levels of subject variability and parameters, and showed outperformance of IVA over GICA in terms of capturing subject-specific variability.

Both IVA and GIG-ICA are able to optimize the independence among intra-subject components and dependence among inter-subject components, and showed advantages over traditional GICA in several comparison studies (Dea et al., 2011; Du and Fan, 2013; Ma et al., 2013; Michael et al., 2014; Du et al., 2016a). However, a full comparison between IVA and GIG-ICA has not been well studied, especially in neuroimaging application. In this paper, we compare their performance using both simulations and real fMRI data. We evaluate the two methods with respect to the estimation accuracy of components and time courses by using simulated data with different quality and quantity, data with varied number of sources generated, inaccurate number of components for computation, as well as

data with subject-unique sources. In addition, test-retest resting-state fMRI datasets are also utilized to compare the two methods in terms of estimated functional networks and interaction among networks. We assess if IVA and GIG-ICA can yield reliable network maps and functional network connectivity (FNC) using the test-retest data. With these detailed comparisons, we expect to gain more knowledge of both methods in analyzing fMRI data under different scenarios and thus provide guidance for researchers in the field.

## MATERIALS AND METHODS

### IVA and GIG-ICA

As for IVA, IVA-GL algorithm was adopted to estimate components for comparisons in this work. It can be accessed in Group ICA for fMRI toolbox (GIFT) (<http://mialab.mrn.org/software/gift/index.html>). There are mainly two steps: (1) performing subject-level PCA on each subject's data; (2) applying IVA-GL to estimate the subject-specific components and time courses (TCs). The estimated components are then Z-scored. A free parameter is the number of components used for the subject-level PCAs, denoted as  $I1$ .

GIG-ICA (Du and Fan, 2013; Du et al., 2016a), also included in GIFT, involves the following steps: (1) performing subject-level PCA reduction on each subject's data and group-level PCA on the temporal concatenation of subject-level PCAs reduced data; (2) applying group-level ICA to the reduced data using Infomax algorithm (Bell and Sejnowski, 1995); (3) identifying and removing artifact-related group-level ICs; (4) computing each subject-specific IC via a multi-objective function optimization based on the individual-subject data and each remaining non-artifact group-level IC (Du and Fan, 2013) using a deflation manner; and finally (5) estimating the subject-specific TCs. In step (4), GIG-ICA simultaneously optimizes the independence of each subject-specific IC, measured by negentropy, as well as the correspondence between each subject-specific IC and each group-level IC, measured by their correlation (Du and Fan, 2013), automatically resulting in Z-scored subject-specific ICs. Relevant parameters include the number of principal components (PCs) used for the subject-level PCAs, denoted as  $G1$ , and the number of PCs/ICs used for the group-level PCA/ICA, denoted as  $G2$ .

It is worth noting that in order to decrease computation load, GIG-ICA can remove artifact-related group-level ICs before estimating individual components (Du et al., 2016a), only yielding subject-specific meaningful networks. However, IVA has to compute all components and remove artifact-related components in a subsequent postprocessing stage. To facilitate comparison between GIG-ICA and IVA, we computed all components in GIG-ICA without performing artifact removal after the group-level ICA step. In experiments using real fMRI data, we matched components between the two methods after obtaining the individual results and then removed the corresponding artifact-related components for both methods. For comparison, we also set  $G1 = G2 = I1$ , resulting in equivalent numbers of components for the two methods. In this work, Infomax algorithm

employed in the first step (i.e., the group-level ICA) of GIG-ICA and IVA-GL algorithm are comparable, since both methods use fixed nonlinearity matched to super-Gaussian sources.

### Experiments Using Simulations

Due to that there is no ground truth in practical applications, simulation-based tests are necessary for evaluating different methods. In order to comprehensively compare IVA and GIG-ICA, we performed several experiments to assess the accuracy of the estimated individual components and TCs under different conditions, including various data quality and quantity (Experiment 1), varied number of sources and inaccurate number of components for computation (Experiment 2), and spatially subject-unique sources (Experiment 3). In each experiment, we simulated fMRI-like data of multiple subjects using the SimTB toolbox (Allen et al., 2012; Erhardt et al., 2012). The number of subjects  $M$  was simulated to be 10. For each subject,  $C$  source images ( $148 \times 148$  pixels) and their corresponding TCs (150 or less time points in length) were simulated and then used to generate data by a linear mixture model. In our experiments, we set  $C$  to be 7 or 8. Rician noise was then added to data with a specified contrast-to-noise ratio (CNR). Repetition time (TR) was 2 s/sample. Among  $C$  sources, some sources were similar across all subjects with slight variance (i.e., subject-common), while the other sources were unique and only present on specific subject (i.e., subject-unique). These subject-unique sources were generated to simulate significant source variability across subjects. The parameters of our experiments are summarized in Table 1.

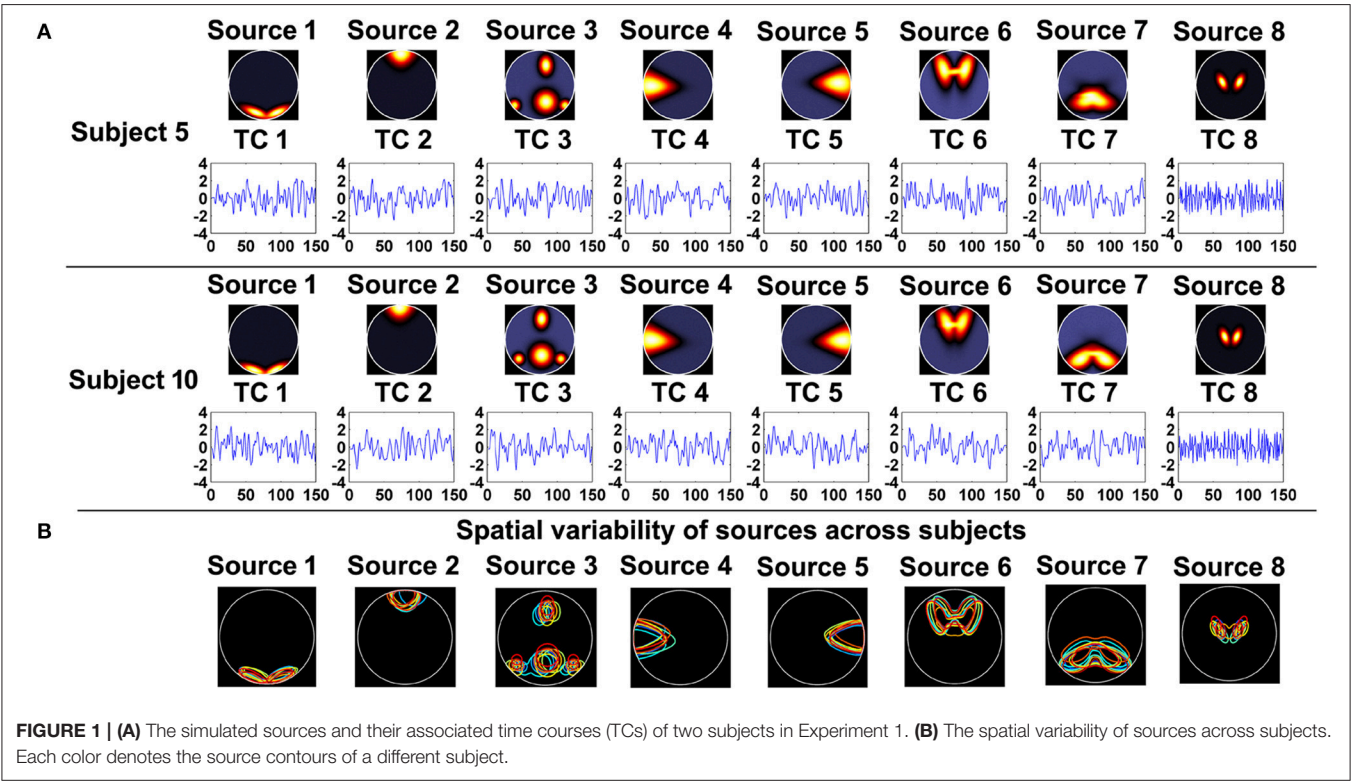
#### Experiment 1: Comparing IVA and GIG-ICA Using Data with Different Quality and Quantity

As shown in Figure 1A, 8 sources and their associated TCs were simulated for each subject. Similar to previous work, each of the 8 sources was generated from a common map with added spatial variability across subjects by random translations [mean = 0 pixel; standard deviation (SD) = 5 pixels], rotations (mean of 0 degree; SD = 3°), and spatial extents (i.e., spreads) of the common spatial map (mean = 3 magnification; SD = 0.03 magnification). Thereby, sources were spatially consistent across subjects but showing moderate subject-specific variability, as shown in Figure 1B. Additionally, temporal variation was applied in simulation of TCs. In order to evaluate the effect of data quality, for each subject, we simulated 16 datasets with different CNRs (ranging from 0.5 to 2 with the step of 0.1) and a fixed number of time points (i.e., 150). Subsequently, regarding each specific CNR (e.g., the CNR = 1), we performed IVA and GIG-ICA on the associated data of all subjects, respectively. To investigate the effect of data quantity, for each subject, we simulated 5 datasets by varying the time points from 40 to 120 in steps of 20 while the CNR was fixed at 2. Afterwards, in terms of each given number (e.g., the number of time points = 100), each method (IVA or GIG-ICA) was applied to the relevant data of all subjects. For these experiments, we set the number of



TABLE 1 | Parameters of simulations and methods in the following simulation-based experiments.

	Experiment 1		Experiment 2		Experiment 3
	Different data quality	Different data quantity	Varied number of sources generated	Inaccurate number of components used in computation	Spatially subject-unique source
Number of subjects	10	10	10	10	10
Number of sources in each subject's data	8	8	8 ( $i = 1, \dots, 5$ ), 7 ( $i = 6, \dots, 10$ ), $i$ denotes the subject index	8	8
CNR in each subject's data	0.5–2 with the step of 0.1	2	2	2	2
Number of time points in each subject's data	150	40–120 with the step of 20	150	150	150
Source type	similar across subjects	similar across subjects	similar across subjects	similar across subjects	one subject-unique source
Number of components used in computation	8	8	7 and 8	6, 8, and 10	8



components (i.e.,  $G1$ ,  $G2$ ,  $I1$ ) used in the analyses to be the same as the number of true sources (i.e., 8).

### Experiment 2: Comparing IVA and GIG-ICA under Conditions of Varied Number of Sources and Inaccurate Number of Components

In this section, we first assessed the performance of the two methods using data with varied number of sources across different subjects (see Table 1 for the detailed parameters). Five subjects' datasets were simulated with 8 sources, while the other five subjects' datasets were simulated with 7 sources. Among

the sources, each of 7 sources had a similar spatial pattern across all subjects with slight inter-subject variability, while the other source was only present in five subjects with small spatial variation. Considering the difference in the simulated number of sources across different subjects, we performed two comparisons by setting the same number of components in IVA and GIG-ICA to 7 and 8 separately.

It is known that prior to ICA, the number of components is a free parameter, typically either selected by the user or estimated by some information-based criteria (Li et al., 2007; Fu et al., 2014). This parameter may influence the

source decomposition since the measure of total component independence may change and thus converge to different solution. In order to evaluate the effect of an inaccurate component number in both methods, based on the data generated with 8 sources (i.e., the data from Experiment 1 with the  $\text{CNR} = 2$ ) we examined each method by setting the number of estimated components to 6, 8, and 10, respectively.

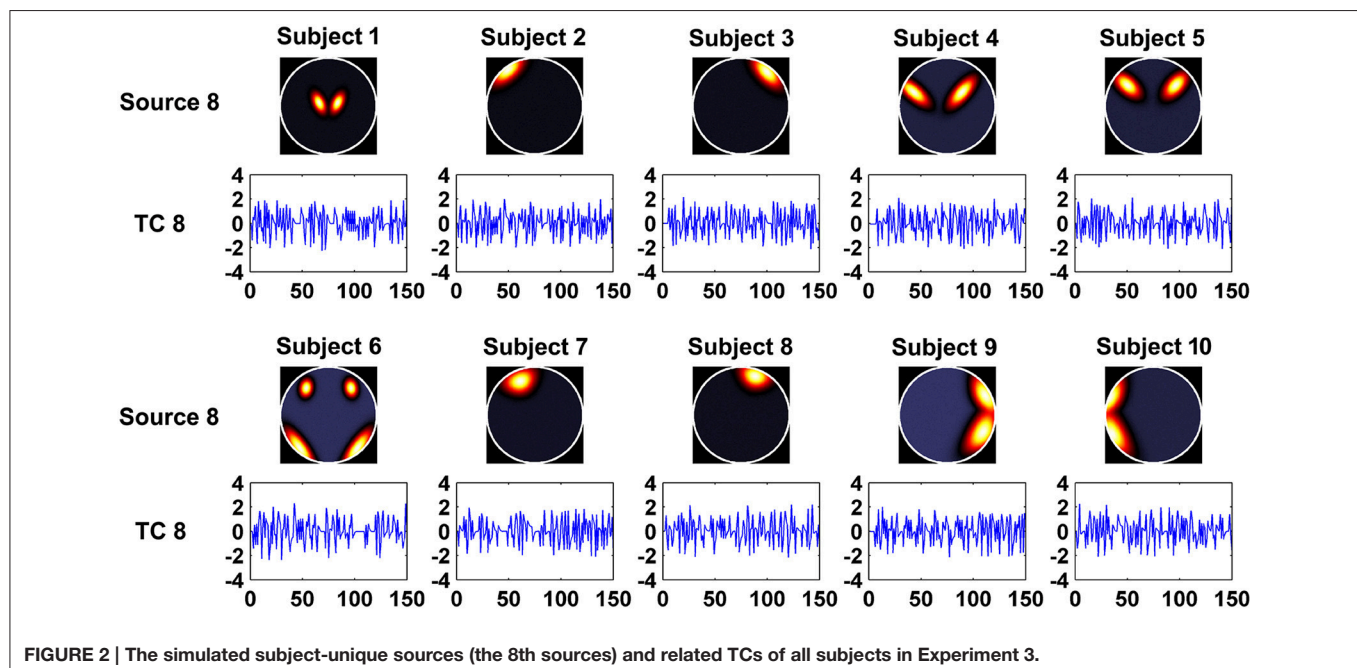
### Experiment 3: Comparing IVA and GIG-ICA Using Data with Spatially Subject-Unique Sources

In this experiment, we aimed to evaluate the ability of the two methods in recovering both subject-common and subject-unique sources. Each subject's data was simulated by using 7 sources each of which was similar across subjects and one additional source unique for each individual subject. The 7 sources had similar patterns with the first 7 sources generated in Experiment 1. **Figure 2** shows the simulated subject-unique sources (i.e., the 8th sources) and related TCs of all subjects. The number of components used in computation was specified as the real number of sources (i.e., 8).

### Evaluation Metrics in Simulation-Based Experiments

To compare the performance of IVA and GIG-ICA on simulations, we evaluated accuracy of the estimated subject-specific components and TCs using correlation between estimation and ground truth, consistent to many prior studies (Schmithorst and Holland, 2004; Allen et al., 2012; Du and Fan, 2013; Michael et al., 2014; Du et al., 2016a). We firstly matched the estimated subject-specific components with the simulated subject-specific ground-truth (GT) sources as follows. Regarding each source in Experiment 1 and 2, the corresponding GT sources of all subjects were averaged, and

then the mean GT sources were used as source templates. Next, for GIG-ICA method, we matched the group-level ICs with the source templates using a greedy rule (see the Supplementary Material for details). Similarly, for IVA method, we averaged the corresponding components from all subjects to represent the group-level components, which were then matched with the sources templates. For each method, based on the matched results between the group-level components and the source templates, the estimated subject-specific components/TCs were then accordingly matched to the subject-specific GT sources/TCs. For Experiment 3 which involved a subject-unique source in data, we first averaged each of 7 subject-common GT sources across subjects to get its source template, and then matched the 8 group-level component maps obtained from each method with the 7 source templates, consequently constructing correspondence between 7 components and 7 GT sources for those subject-common sources of each subject. Thus, one additional subject-unique component can be matched to the subject-unique GT source for each subject. After matching, we computed the absolute value of Pearson correlation coefficient between each estimated component/TC and its matched GT source/TC to measure the component/TC accuracy. In Experiment 1 and 2, we further calculated the mean of all components/TCs accuracy measures of each subject to reflect its overall component/TC accuracy. In Experiments 1 and 2, for each setting, a two-tailed paired *t*-test was performed to compare the overall component (or TC) accuracy metrics of all subjects from IVA with that from GIG-ICA. In Experiment 3, for each component, we compared the spatial (or temporal) accuracy of all subjects between IVA and GIG-ICA using one two-tailed paired *t*-test. The results were corrected using  $p < 0.05$  with Bonferroni correction.



## Experiments Using Test-Retest Resting-State fMRI Data

Seventy five resting-state fMRI datasets (Zuo et al., 2010) comprising 25 healthy participants with three scans were adopted in the experiment. Each dataset consisted of 197 contiguous EPI functional volumes ( $TR = 2,000$  ms;  $TE = 25$  ms; flip angle =  $90^\circ$ , 39 slices, matrix =  $64 \times 64$ ; FOV = 192 mm; acquisition voxel size =  $3 \times 3 \times 3$  mm). The first scan (scan 1) is in a scan session. Five to Sixteen months (mean  $11 \pm 4$ ) after scan 1, scan 2 and scan 3 were conducted with short interval (about 45 min). The fMRI images were preprocessed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>). The first 10 images were discarded, and the remaining 187 images were slice-time corrected and realigned to the first volume for head-motion correction. Subsequently, the images were spatially normalized to the Montreal Neurological Institute (MNI) EPI template and spatially smoothed with a 6 mm FWHM Gaussian kernel.

IVA and GIG-ICA were applied to all 75 preprocessed datasets, respectively, to estimate brain functional networks and their associated TCs of each dataset. For a comprehensive evaluation of these two methods, we used both low and high numbers of components for analyses. When a low number is used, it makes more sense to think of each meaningful component itself as a brain functional network. Many studies (Meda et al., 2014; Du et al., 2015c) have conducted analyses on spatial maps of networks revealed by ICA with low model order, aiming to explore disease biomarkers. In contrast, if a high number is used, the meaningful networks were then usually used as nodes for computing consequent FNC (Allen et al., 2011). Each FNC matrix, which is computed based on the individual-subject's TCs of networks, reflects interaction among different networks. To be consistent with previous studies (Allen et al., 2011; Du and Fan, 2013; Du et al., 2015c, 2016a), we specified 20, 25, and 30 as low model order settings, and 75 and 100 as high model order settings. For simplification, we assessed the results from both low and high model orders using the same manner by considering properties of both networks' spatial maps and interaction among networks (i.e., FNC). Regarding results from each model order setting, we first matched the obtained components from the two methods based on their group-level component maps using a greedy rule (see the Supplementary Materials). Then, based on the matched components with high similarity (correlation  $> 0.5$ ) between the two methods, we only selected the meaningful networks by manually inspecting spatial and temporal information of the matched components (Allen et al., 2014; Du et al., 2016a) for further investigation. Next, the following evaluations in terms of network maps and FNC were performed on the selected networks for each method. Finally, the performances of the two methods under different model orders were compared.

For each selected network, we evaluated its reliability based on the estimated individual networks from 75 datasets as follows, which is consistent to previous studies (Zuo et al., 2010; Du et al., 2016a). First, voxel-wise right-tailed one-sample  $t$ -tests ( $p < 0.01$  with false discovery rate (FDR) correction) were performed on the corresponding networks of all 75 datasets. Next, since the data

from scan 2 and scan 3 were collected with short intervals, voxel-wise intra class coefficients (ICCs) (Zuo et al., 2010) between the corresponding 25 networks from scan 2 and the corresponding 25 networks from scan 3 were calculated to assess the short-term reliability of the network, resulting one 3D ICC map reflecting short-term reliability of the network. In our work, ICC of each voxel was computed using a model (Zuo et al., 2010; Guo et al., 2012) based on one-way analysis of variance (ANOVA), due to that those subjects were scanned using the same scanner.

The used equation was:  $ICC = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}$ , where  $\sigma_p^2$  denotes the variance of inter-subject effect and  $\sigma_e^2$  denotes the variance of measurement error. As mentioned above, the data of scan 2 and scan 3 were collected after several months of scan 1. So, we computed ICCs between the corresponding 25 networks from scan 1 and the averaged 25 networks from scan 2 and scan 3 to assess the long-term reliability of the network, resulting in one 3D ICC map reflecting long-term reliability. Based on each ICC map reflecting short-term or long-term reliability of the network, the ICC values were then averaged across voxels within a specific mask which included statistically significant voxels for both methods based on the one-sample  $t$ -tests results after FDR correction, to summarize the short-term or long-term reliability of the network.

To investigate network interaction, we calculated FNC for each of the 75 datasets, and then evaluated graph-theory based measures using the brain connectivity toolbox (<https://sites.google.com/site/bctnet/>) as well as reliability in both connectivity and modularity. First, for each dataset, we obtained one FNC matrix by computing Pearson correlation coefficients between the associated TCs of any paired networks. Next, we averaged the FNC matrix across all 75 datasets. Based on the mean FNC matrix, we detected its modules (i.e., network communities) using the most applied eigenvector-based method (Newman M. E., 2006; Newman M. E. J., 2006), where the modularity  $Q$ -value reflects the accuracy or quality of a community structure. Greater  $Q$ -value represents stronger modular structure. Subsequently, modularity analysis was also performed on each individual FNC matrix, resulting in a module segmentation and related  $Q$ -value for each dataset. Since different datasets may have greatly varied modular brain networks, we measured modularity similarity between any pair of datasets using the adjusted mutual information (AMI), consistent to a recent study (Liao et al., 2017). The mean of AMI values computed between datasets in scan 2 and datasets in scan 3 was used to measure the short-term modularity reliability. The mean of AMI values obtained between datasets in scan 1 and datasets in scan 2 or 3 was used to reflect the long-term modularity reliability. Additionally, each connectivity's short-term and long-term reliability in FNC was examined using ICC. Specifically, for each connectivity (i.e., one element in FNC matrix), ICC between the corresponding 25 connectivity strengths from scan 2 and the corresponding 25 connectivity strengths from scan 3 was calculated to assess the short-term reliability of the connectivity; ICC between the corresponding 25 connectivity strengths from scan 1 and the corresponding 25 connectivity strengths averaged between scan 2 and 3 was calculated to assess the long-term reliability of the



connectivity. Finally, we calculated the averaged node strength, clustering coefficient, global efficiency, and local efficiency (Rubinov and Sporns, 2010) based on each individual FNC matrix, the elements of which were first changed to their absolute values and thresholded to preserve half elements with higher values (sparsity = 0.5) (Du et al., 2016b).

## RESULTS

### Results from Simulation-Based Experiments

#### Component and Time Course Accuracy Estimated from IVA and GIG-ICA in Experiment 1

**Figure 3** shows the components/TCs of one subject estimated by IVA and GIG-ICA from the simulated data with the CNR = 1. For this case, we can see that both methods can generally recover all of spatial components and TCs. For some components (e.g., component 3, 6, and 7), GIG-ICA had slightly higher component/TC accuracy than IVA. **Figure 4A** summarizes the comparison results across 10 subjects under the condition of different CNRs. It can be observed that the recovery accuracy of components by both methods was improved along with the increasing of CNR while TCs recovery was relatively insensitive to different CNRs. Measured by the mean accuracy, GIG-ICA outperformed IVA across most of CNR settings in terms of component/TC accuracy. **Figure 4B** demonstrates results from evaluating the influence of different numbers of time points. Both methods showed increasing recovery accuracy of components with more time points used. Paired *t*-test results (**Table 2**) show that for most of the CNR and time point settings tested, GIG-ICA showed significantly increased accuracy (especially the spatial accuracy) than IVA. Our results indicate the advantage of GIG-ICA in recovering subject-common sources than IVA, and GIG-ICA can yield components with higher accuracy even under the case of low quality and quantity of data.

#### Component and Time Course Accuracy Estimated from IVA and GIG-ICA in Experiment 2

**Figure 5A** shows the accuracy results obtained from data with varied numbers of sources across different subjects. We can see that under all model orders (i.e., different numbers of components), GIG-ICA showed significantly better performance (see **Table 3**) than IVA, indicating that GIG-ICA is able to tolerate source number variation and is also not very sensitive to the number of components used.

The results demonstrated in **Figure 5B** were obtained using data generated with 8 sources and different numbers of components for computation (i.e., 6, 8, and 10). It can be observed that GIG-ICA significantly outperformed IVA under all model orders in terms of the spatial accuracy (see **Table 3**). Regarding the temporal accuracy, GIG-ICA had significantly greater accuracy using the model order 6, but slightly decreased accuracy using the model order 10, compared to IVA. For the model order 8, the TC results of the two methods are statically close. When the used number of components was the same as the real source number (i.e., 8), both methods achieved the best estimation. When the number of components was 10, there was a

slight decrease in recovering components/TCs for both methods, compared to the results of the model order as 8. However, when the number of sources was underestimated (i.e., 6), there was a significant drop of accuracy for the IVA but a slight decrease for GIG-ICA. Because the accurate number of components is very difficult to estimate correctly in practice, the relative insensitivity of GIG-ICA to the model order may provide an important benefit.

#### Component and Time Course Accuracy Estimated from IVA and GIG-ICA in Experiment 3

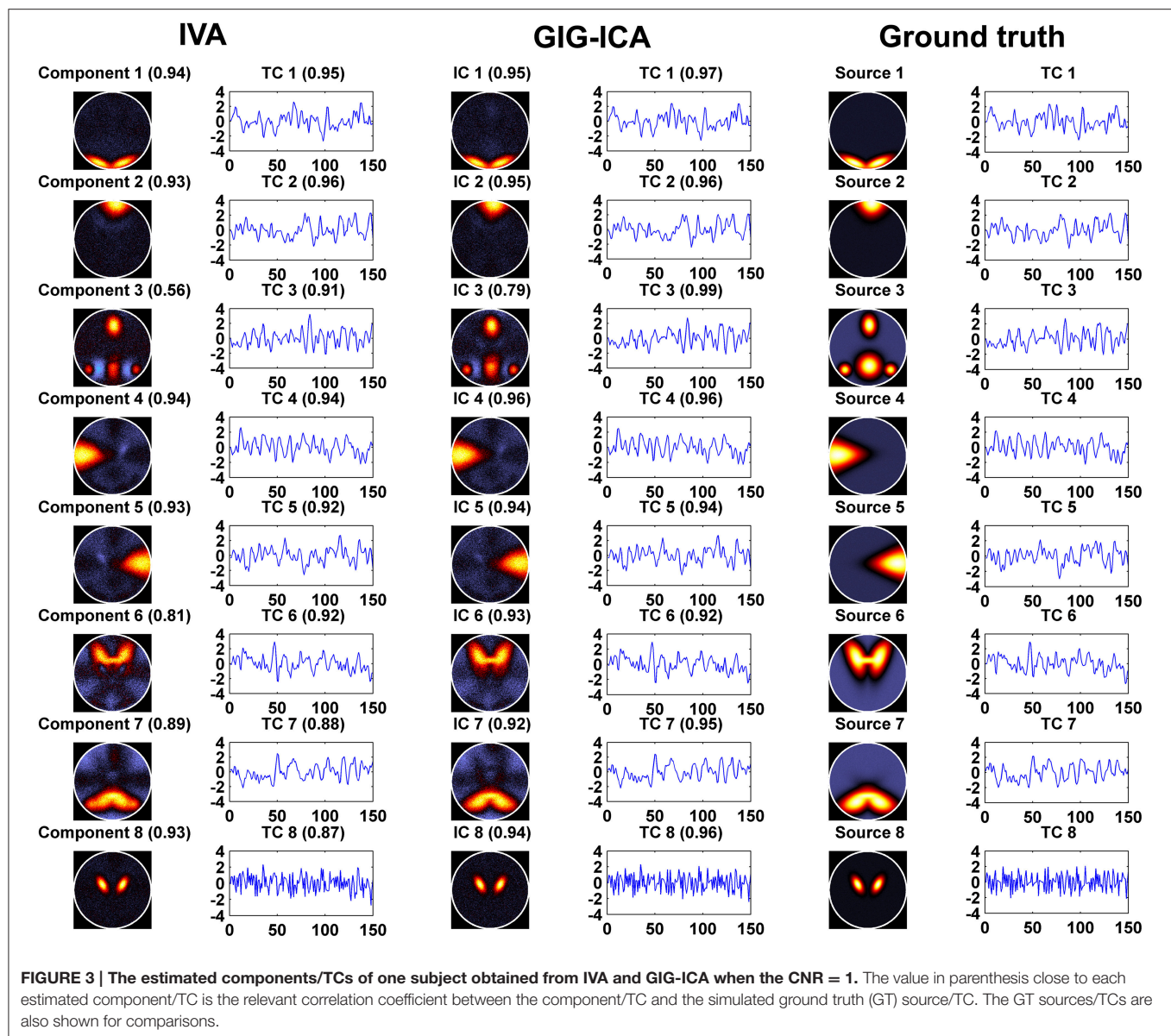
In this experiment, we tested the two methods using datasets where each subject had a spatially unique source. Accuracy of each estimated individual component/TC is shown in **Figure 6**. It is seen that for the estimated spatial components, measured by the mean accuracy across subjects, GIG-ICA had a better performance for the subject-common sources (i.e., the first 7 sources), but showed a worse estimation for the subject-unique source (i.e., the 8th source) than IVA. Regarding the estimated eight TCs, measured by the mean accuracy across subjects, GIG-ICA had higher TC accuracy in four TCs and decreased TC accuracy in terms of the subject-unique source compared to IVA. Using paired *t*-tests (see **Table 4**), among the 7 subject-common sources, four components and three TCs were significantly more accurate using GIG-ICA than using IVA. IVA outperformed GIG-ICA in estimating the subject-unique sources (passing  $p < 0.05$  with correction). Our results suggest that for the data generated with subject-unique sources, in general GIG-ICA still performed well for the similar sources but did not work well for the unique source. In contrast, IVA can estimate the subject-unique source and its associated TC with high accuracy.

### Results from Test-Retest Resting-State fMRI Data

Using the test-retest resting-state fMRI datasets, we assessed the individual-level spatial networks in terms of their short-term and long-term reliability. **Figure 7** shows the one-sample *t*-tests results of the 12 matched networks for the two methods under the condition of the model order as 30. We found that compared to IVA, GIG-ICA in general showed higher *t*-values for all networks. For the case of the model order as 30, the short-term and long-term reliability measures of each network are demonstrated in **Figures 8A,B**, respectively. Results indicate that for most of the networks, greater reliability measures were obtained using GIG-ICA compared to IVA, although there were also four networks (including Network 1, Network 5, Network 7, and Network 11) showing slightly higher short-term or long-term reliability in IVA than GIG-ICA. Furthermore, some networks including the sensorimotor and cerebellum-related networks from IVA had very low reliability. To summarize, we show the short-term and long-term reliability of all networks estimated with different model orders in **Figures 8C,D**. It can be seen that measured by the mean values of reliability measures across all networks, the higher network reliability was achieved by GIG-ICA than IVA for all model order settings.

We also compared the two methods in constructing functional interaction among networks. Under a model order of 100,

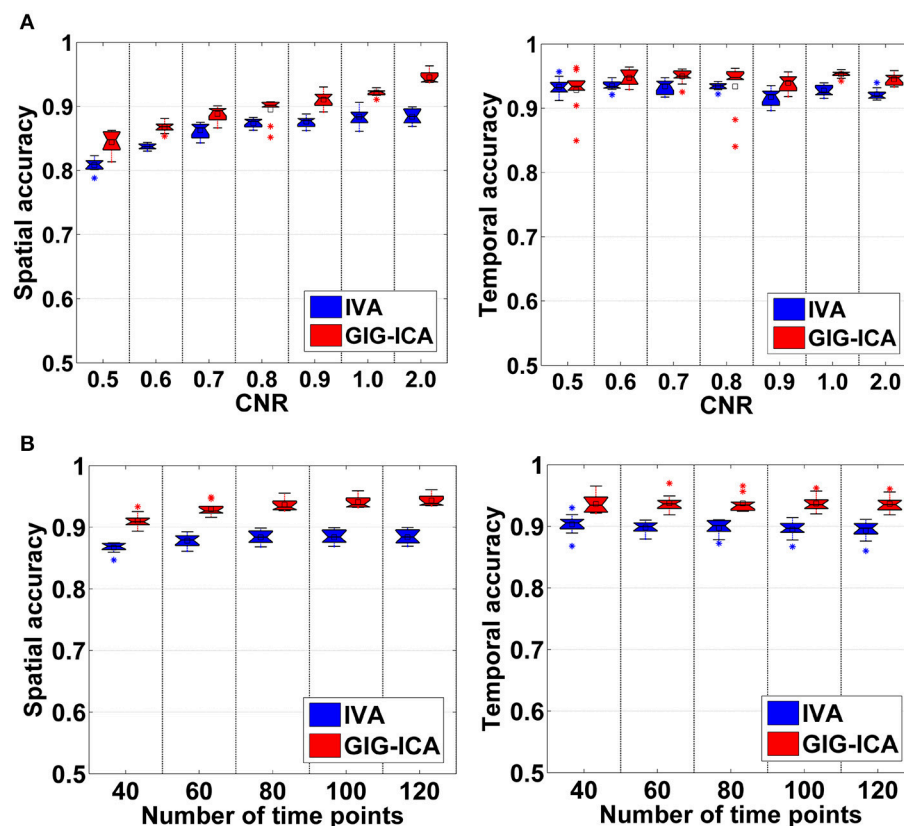




22 networks were highly matched between the two methods. For each dataset, one FNC matrix was generated based on the associated TCs of the 22 networks. **Figures 9A,B** show the mean FNC matrix across all 75 datasets for IVA and GIG-ICA, respectively. It is observed that the two FNC matrices generally showed a similar pattern. However, the contrast in FNC appeared higher in GIG-ICA than IVA. According to the modularity segmentation of networks, we reorganized the mean FNC matrix's structure for IVA (**Figure 9C**) and GIG-ICA (**Figure 9D**). The identified modules for the two methods were demonstrated in **Figures 9E,F**, respectively. Three modules mainly relating to the default mode network (module 1), the cognitive control, sensorimotor and auditory functions (module 2), and the vision function (module 3) were found using GIG-ICA. Module 1 and 2 showed anti-correlations in their connectivities. Regarding

IVA, two modules were detected, while the vision-associated networks were separated into two modules. Furthermore, the modularity quality was greater in GIG-ICA ( $Q = 1.33$ ) compared to IVA ( $Q = 0.51$ ) when the number of components was 100.

Furthermore, GIG-ICA showed an equivalent or higher modularity  $Q$ -value of the mean FNC than IVA for the model order settings tested (see **Figure 10A**). Regarding individual FNC's modularity, **Figure 10B** demonstrates that excepting the low model order 20, GIG-ICA with a greater mean  $Q$ -value outperformed IVA for most of the cases. Moreover, measured by the AMI, both the short-term and the long-term modularity reliability metrics were greater in GIG-ICA than IVA for all tests, as shown in **Figures 10C,D**. Both the short-term and long-term ICC measures (**Figures 10E,F**) support that the connectivity strengths in FNC were



**FIGURE 4 |** (A) Spatial and temporal accuracy measure obtained from IVA and GIG-ICA under different CNRs ranging from 0.5 to 2. (B) Spatial and temporal accuracy measure obtained from IVA and GIG-ICA under different numbers of time points. The x-axis in each boxplot denotes CNR in (A) or number of time points in (B). The y-axis denotes the mean of spatial/temporal correlation coefficients between one subject's estimated components/TCs and the corresponding ground truth sources/TCs, which was used to measure the overall spatial/temporal accuracy of one subject's result. Each point in a given boxplot corresponds to the overall spatial/temporal accuracy of one subject. For each boxplot, the central line is the median, and the edges of the box are the 25 and 75th percentiles. The whiskers extend to 1 inter-quartile range, and each outlier is displayed with a "\*" sign. The mean value is indicated by a square. Subsequent boxplots are formatted similarly.

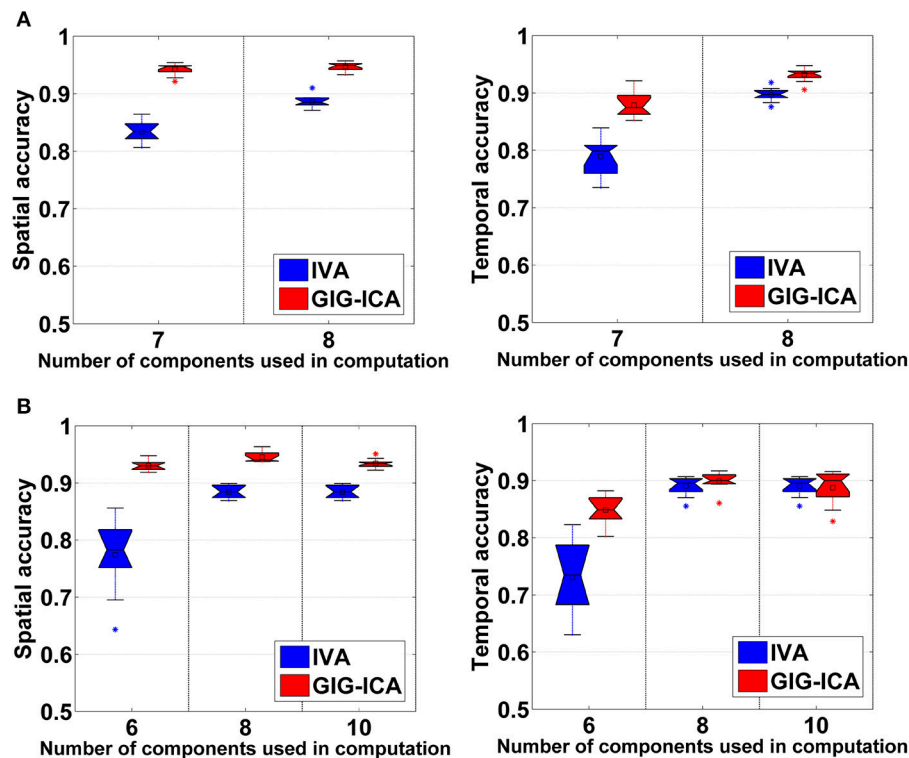
**TABLE 2 | Results of the estimation accuracy using paired *t*-tests for Experiment 1.**

Data with different quality	CNR = 0.5	CNR = 0.6	CNR = 0.7	CNR = 0.8	CNR = 0.9	CNR = 1.0	CNR = 2
<i>p</i> -value in spatial accuracy	5.43e-05	1.16e-06	3.05e-06	7.96e-03	1.42e-07	2.79e-06	1.77e-07
<i>t</i> -value in spatial accuracy	-7.14	-11.44	-10.19	-3.39	-14.61	-10.30	-14.24
<i>p</i> -value in temporal accuracy	0.61	0.002	4.34e-05	0.98	3.12e-05	6.20e-06	7.14e-06
<i>t</i> -value in temporal accuracy	0.52	-4.12	-7.35	0.014	-7.66	-9.36	-9.20
Data with different quantity	Number of time points = 40	Number of time points = 60	Number of time points = 80	Number of time points = 100	Number of time points = 120		
<i>p</i> -value in spatial accuracy	2.21e-05	2.01e-06	9.17e-07	4.71e-07	2.98e-07		
<i>t</i> -value in spatial accuracy	-8.00	-10.71	-11.76	-12.71	-13.41		
<i>p</i> -value in temporal accuracy	3.74e-03	2.0e-3	1.58e-3	8.52e-05	5.64e-05		
<i>t</i> -value in temporal accuracy	-3.88	-6.01	-6.21	-6.73	-7.10		

generally more robust using GIG-ICA method, compared to IVA.

As mentioned in the method section, we also examined other graph-theory based metrics for individual FNC. The summarized

results for the averaged node strength, clustering coefficient, global efficiency, and local efficiency are shown in (Figures 11), suggesting that GIG-ICA resulted in higher mean values in all these graph metrics than IVA under all model order settings.



**FIGURE 5 | (A)** Spatial and temporal accuracy obtained from IVA and GIG-ICA under different model orders for datasets with varied numbers of sources across subjects. **(B)** Spatial and temporal accuracy obtained from IVA and GIG-ICA under the model order as 6, 8, and 10 for datasets generated with 8 sources. The x-axis in each boxplot denotes the number of components used in computation. The y-axis denotes the mean of spatial/temporal correlation coefficients between one subject's estimated components/TCs and the corresponding ground truth sources/TCs, which was used to measure the overall spatial/temporal accuracy of one subject's components/TCs.

**TABLE 3 | Results of the estimation accuracy using paired *t*-tests for Experiment 2.**

	Data with varied source number		Inaccurate number of component		
	IC number = 7	IC number = 8	IC number = 6	IC number = 8	IC number = 10
<i>p</i> -value in spatial accuracy	1.32e-09	2.47e-09	2.24e-05	1.76e-07	1.33e-07
<i>t</i> -value in spatial accuracy	-24.86	-23.18	-7.99	-14.25	-14.72
<i>p</i> -value in temporal accuracy	8.10e-06	7.25e-05	4.10e-04	0.10	0.77
<i>t</i> -value in temporal accuracy	-9.06	-6.88	-5.44	-1.85	0.30

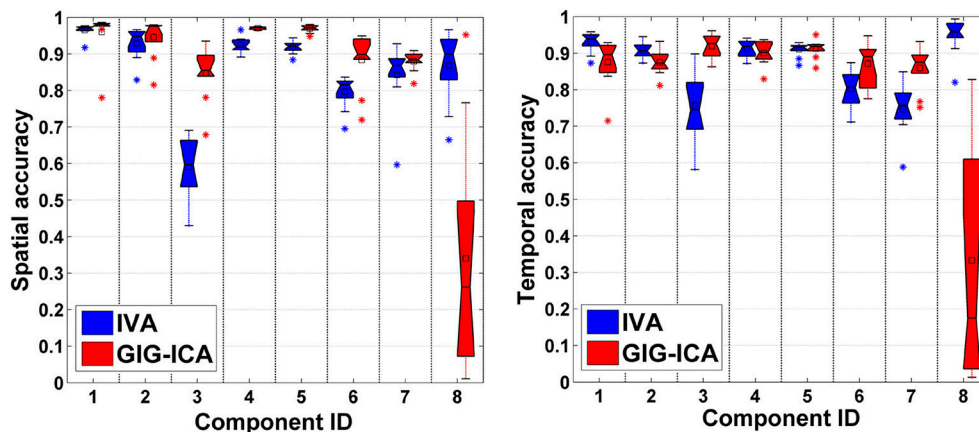
**TABLE 4 | Results of the estimation accuracy using paired *t*-tests for Experiment 3.**

	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6	Component 7	Component 8
<i>p</i> -value in spatial accuracy	0.70	0.02	4.34e-08	1.04e-04	2.01e-07	1.01e-03	0.21	8.95e-04
<i>t</i> -value in spatial accuracy	0.39	-2.66	-16.74	-6.56	-14.04	-4.78	-1.33	4.86
<i>p</i> -value in temporal accuracy	0.03	0.03	3.81e-04	0.20	0.78	5.53e-03	3.29e-05	1.96e-04
<i>t</i> -value in temporal accuracy	2.53	2.43	-5.50	1.38	-0.29	-3.62	-7.61	6.03

## DISCUSSION

In this work, we compared two promising approaches (i.e., IVA and GIG-ICA) for analyzing multi-subject fMRI data.

Both methods can estimate subject-specific brain functional networks with correspondence across different subjects. IVA considers both the independence of individual components and the dependence of similar components across subjects. GIG-ICA



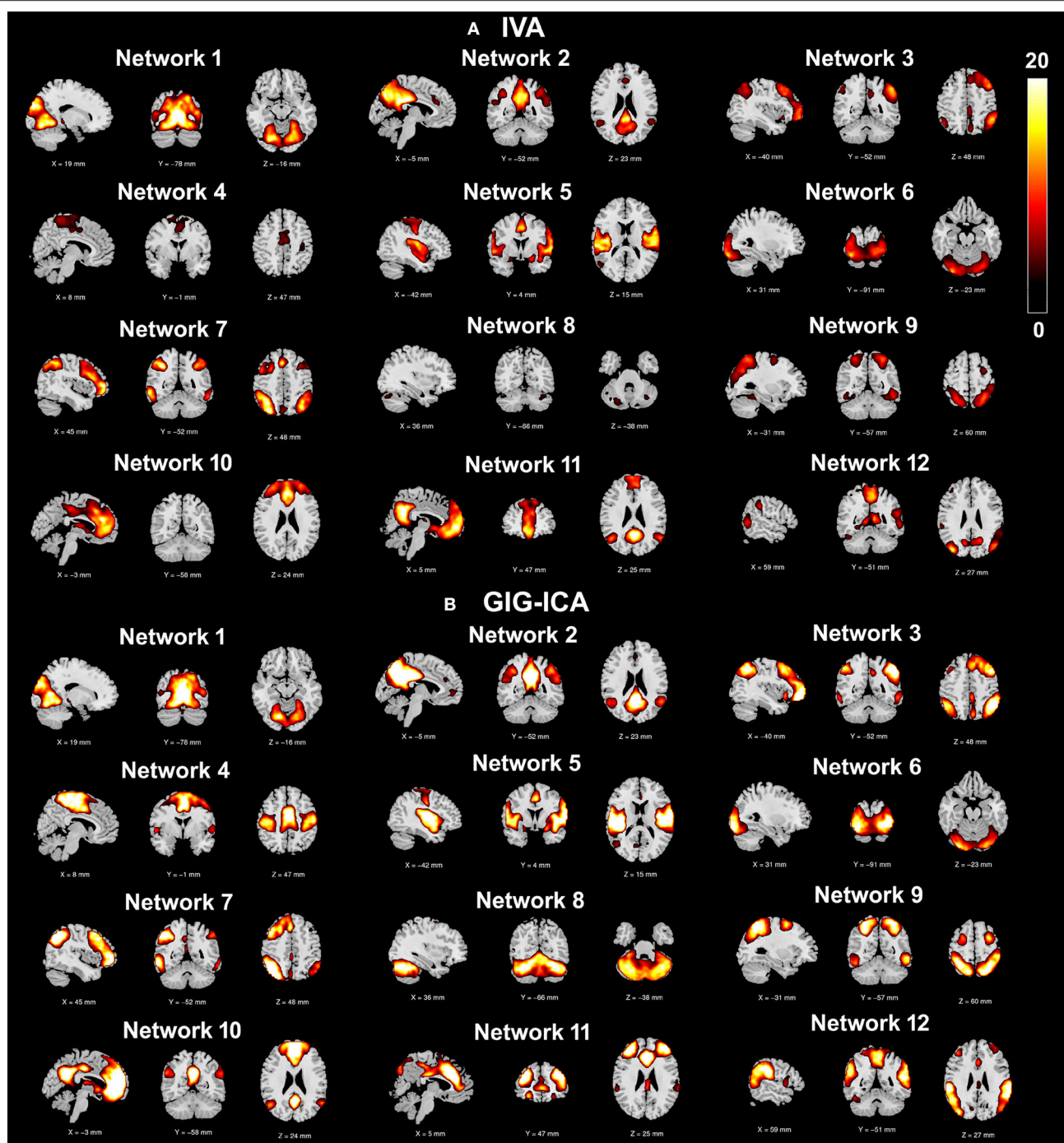
**FIGURE 6 | Spatial and temporal accuracy of each estimated component and TC obtained from IVA and GIG-ICA for datasets with subject-unique sources.** The 8th component is subject-unique. The y-axis denotes the spatial/temporal correlation between one subject-specific component/TC and the corresponding ground truth source/TC. The accuracy metrics of each component/TC from all subjects are shown using one boxplot. Each point in one boxplot corresponds to the spatial/temporal accuracy of one component/TC for one subject.

first estimates the group-level ICs from all data and then computes the subject-specific ICs with the group-level ICs as guidance. Using simulations, we investigated if the two methods can yield accurate individual-level components and time courses under different conditions, including different data quality (i.e., CNR) and data quantity (i.e., number of time points), varied number of sources and inaccurate number of components, as well as presence of spatially subject-unique sources. Furthermore, we assessed their performance using test-retest resting-state fMRI data with respect to spatial networks' reliability and graph-theory based metrics of FNC under different model orders.

In Experiment 1 using simulations, we evaluated the two methods using data with various quality and quantity. Our results suggest that both IVA and GIG-ICA showed improved performance along with the increased CNR and time points of data. For the sources with slight inter-subject spatial variability, GIG-ICA obtained components with higher accuracy than IVA, and performed very well under the case of low CNR and less time points. It is known that both IVA and GIG-ICA require a fixed number of components for computation, generating the same number of components for all subjects. When datasets of different subjects are simulated using different numbers of sources, the resulting components of some subjects have different numbers with the real number of sources. So, in Experiment 2, we simulated varied number of sources between different groups and also investigated the influence of inaccurate number of components. Our results suggest that GIG-ICA showed a relatively better performance and was stable to the various numbers of sources under this case. We also tested the two methods in terms of the effect of the number of components, indicating that IVA gave rise to a significant reduced accuracy when the model order was underestimated while GIG-ICA was not very sensitive to the inaccurate model order.

All the above mentioned experiments were applied to the datasets generated using sources that were similar across subjects. In Experiment 3, using datasets where all subjects had a subject-unique source with large inter-subject spatial variability, we found that IVA significantly showed a better performance in the component/TC accuracy of the unique source than GIG-ICA, although GIG-ICA in general still performed better for other subject-common sources compared to IVA. This is likely due to that the two methods are different in algorithm level. GIG-ICA first extracts the group-level components, and then estimates the corresponding individual-level components for each individual-subject's data. In contrast, IVA simultaneously estimates the individual-subject's components and optimizes the dependence of components across different subjects. Therefore, we suggest using GIG-ICA to estimate networks that are consistent across subjects, while IVA is more appropriate for networks with significant inter-subject variability. IVA's superiority in estimating subject-unique sources possibly enables it to be more suitable to data from patients with particular brain structure damage, such as patients suffering from brain tumor that could result in greatly different functional networks. Our previous work (Du et al., 2014, 2015c, 2017) showed that GIG-ICA performed well for fMRI data from healthy controls and patients with mental disorders, which are supposed to have similar network patterns but subtle differences. In fact, all GICA approaches (Calhoun et al., 2001, 2009; Beckmann et al., 2009; Erhardt et al., 2011) have the same limitation with GIG-ICA, since all of them back-reconstruct individual-subject's ICs based on the group-level ICs. However, as our previous work (Du et al., 2016a) suggested, GIG-ICA is a powerful approach for main fMRI researches, due to the fact that the subject-common networks can be estimated and denoised without having to accurately estimate the artifacts. In future, a general framework that leverages the strengths of IVA and GIG-ICA is expected for achieving high accuracy of both subject-common and subject-unique networks.

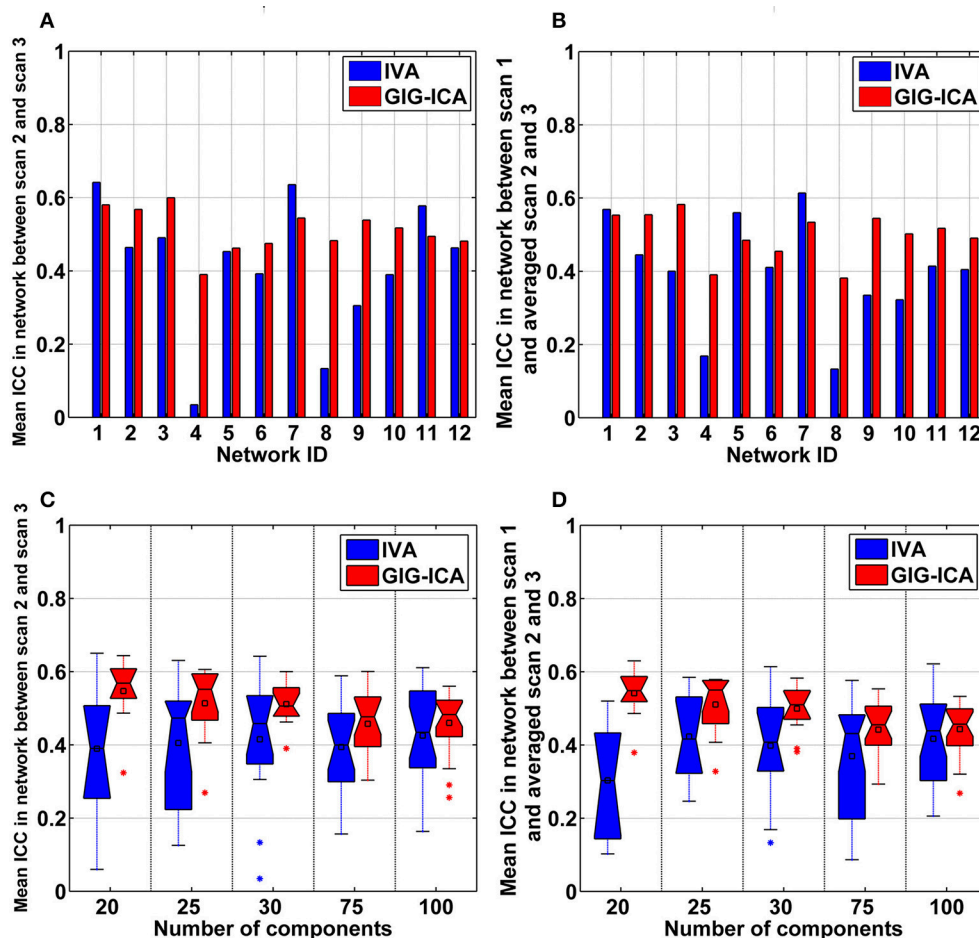




**FIGURE 7 | One-sample  $t$ -test  $t$ -value maps ( $p < 0.01$  with FDR correction) of the 12 matched networks, obtained by (A) IVA and (B) GIG-ICA under the case of the model order as 30. The 12 matched networks shown are sorted according to the similarity (i.e., correlation) between networks from the two methods.**

Our experiments using healthy participants' test-retest resting-state fMRI data revealed that regardless of low model order and high model order, GIG-ICA in general obtained functional networks with relatively greater short-term and long-term reliability compared to IVA, although a few networks showed slightly higher reliability in IVA than GIG-ICA. In terms of the interaction among networks represented by FNC, we found

that the mean FNC matrix from the two methods showed a similar pattern to some extent. However, both the mean FNC and the individual-level FNC showed stronger modularity (i.e.,  $Q$ -value) using GIG-ICA compared to IVA for most of the model order settings examined. Measured by the AMI, the modular structure was more reliable during short-term and long-term rescanning using GIG-ICA for all tests, compared to using IVA.

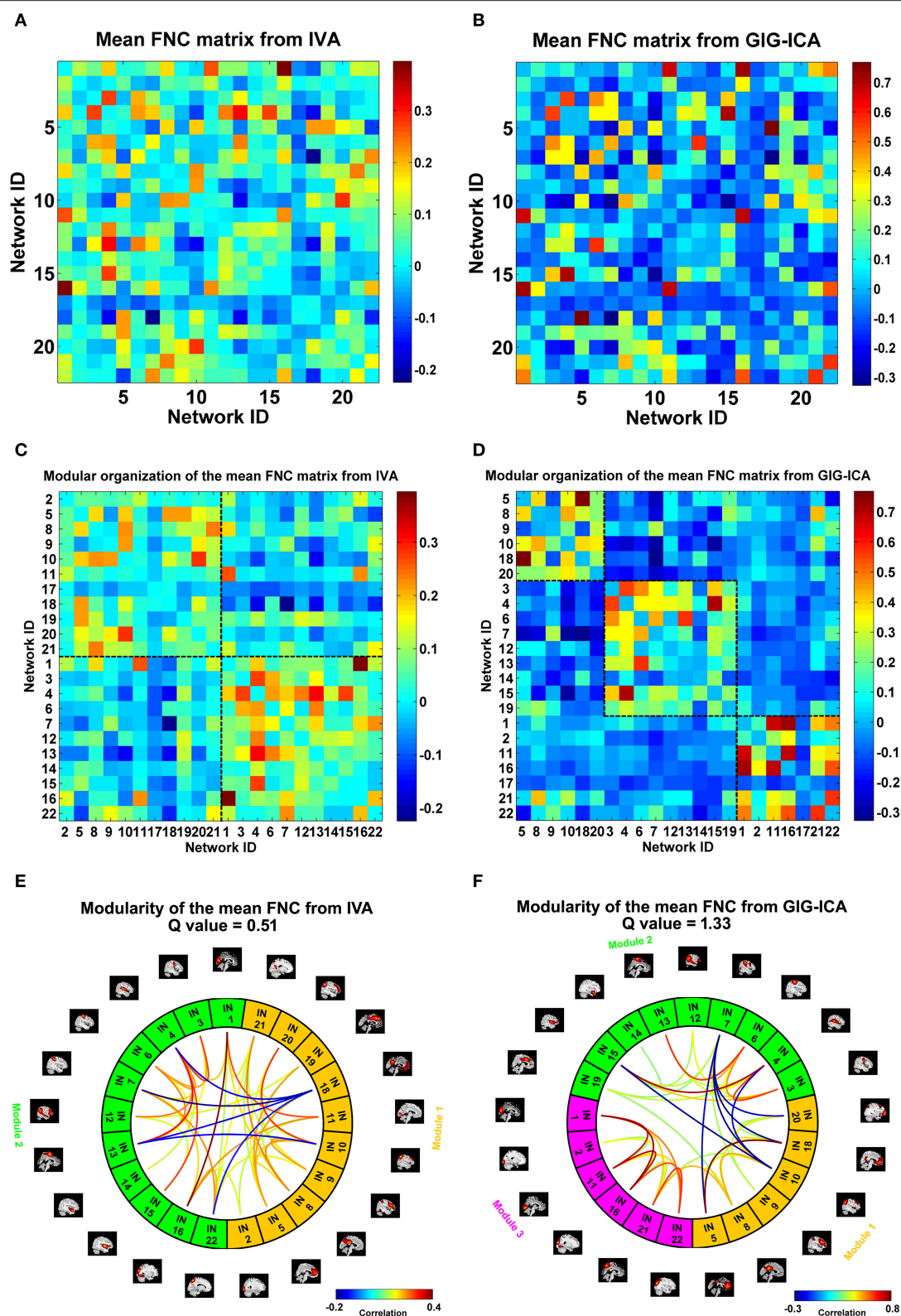


**FIGURE 8 | (A,B)** Reliability measures of the 12 selected networks for IVA and GIG-ICA under the case of the model order as 30. The x-axis denotes the network ID which corresponds to that in **Figure 7**. **(A)** Mean ICC value in each network reflecting the short-term reliability of the network. The value was obtained by first computing ICCs between the corresponding networks of scan 2 and that of scan 3, and then averaging ICCs in the significant voxels. **(B)** Mean ICC value in each network reflecting the long-term reliability of the network. The value was obtained by first computing ICCs between the corresponding networks from scan 1 and the mean networks of scan 2 and 3, and then averaging ICCs in the significant voxels. **(C,D)** The summarized network reliability measures for IVA and GIG-ICA under different model orders (i.e., different numbers of components). **(C)** Short-term reliability of networks. **(D)** Long-term reliability of networks. Each boxplot shows the reliability measures of different networks using IVA or GIG-ICA with one given model order. For the model order 20, 25, 30, 75, and 100, the number of matched networks between the two methods were 9, 10, 12, 19, and 22, respectively.

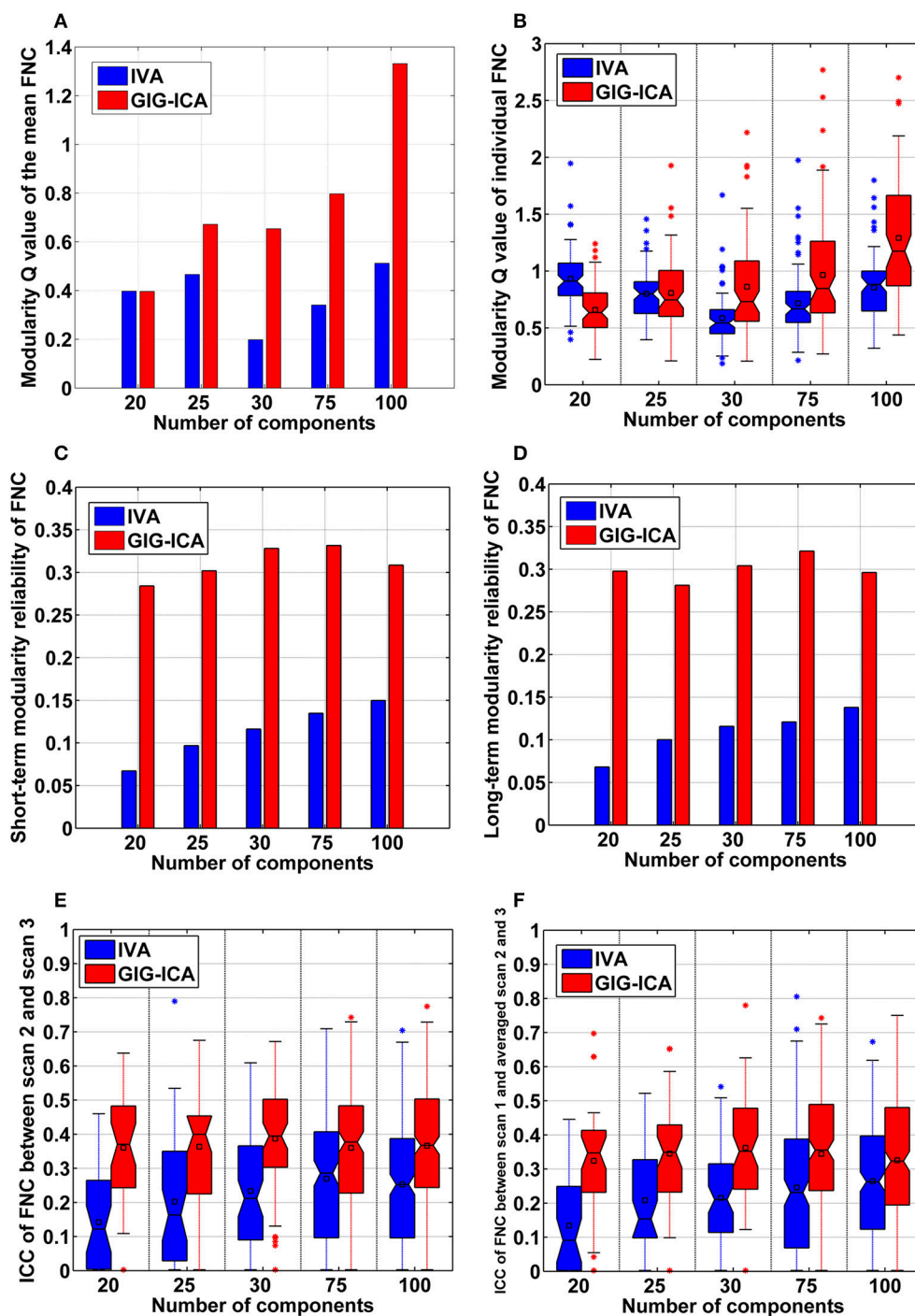
Despite short-term and long-term, ICC measures demonstrate that connectivity strengths were generally more robust using GIG-ICA method, compared to IVA. Moreover, FNC obtained from GIG-ICA showed relatively higher values in the averaged node strength, clustering coefficient, global efficiency, and local efficiency, indicating stronger interaction among brain functional networks.

There are some limitations in our work. (1) The simulations are quite simple. Only eight sources and ten subjects were simulated, while the proportion in fMRI data is certainly greater. In practical applications, there exist more complex situations that could involve many subject-unique sources, high diversity in source number, and great bias in model order estimation. Therefore, it's possible that conclusions we draw from simulations are over-simplified and of limited applicability. However, we also evaluated the two methods using data with

more subject-common and subject-unique sources. The results are included in the Supplementary Materials (Figures S2, S4). Our results suggest that the performances of both methods were affected by greater spatial overlapping among sources, and the presence of more subject-unique sources may slightly influence the estimations of the subject-common sources in GIG-ICA to some extent. (2) The number of sources in real data is difficult to estimate accurately. Therefore, we don't know the appropriate model orders at which to compare these two methods in real data. We compared the two methods using different numbers of components and found similar results, but these methods may yield different performances with other model orders. (3) Since IVA involves a more complicated optimization task, performance might improve if a best run selection mechanism as in previous work (Ma et al., 2011) is used to select the most reliable run across multiple runs. However, we did not perform



**FIGURE 9 | The mean FNC matrix and its modularity result with the model order as 100. (A,B)** The mean FNC matrix across subjects derived from IVA and GIG-ICA. There were 22 matched networks between the two methods. **(C,D)** Modular organization of the mean FNC matrix from IVA and GIG-ICA. **(E,F)** The connectogram representation of the modularity of the mean FNC obtained from IVA and GIG-ICA. In **(E,F)**, the intrinsic networks (INs) belonging to the same modular are shown using the same color, and only top 20% of the connectivities with higher absolute connectivity strengths among networks are shown using lines.

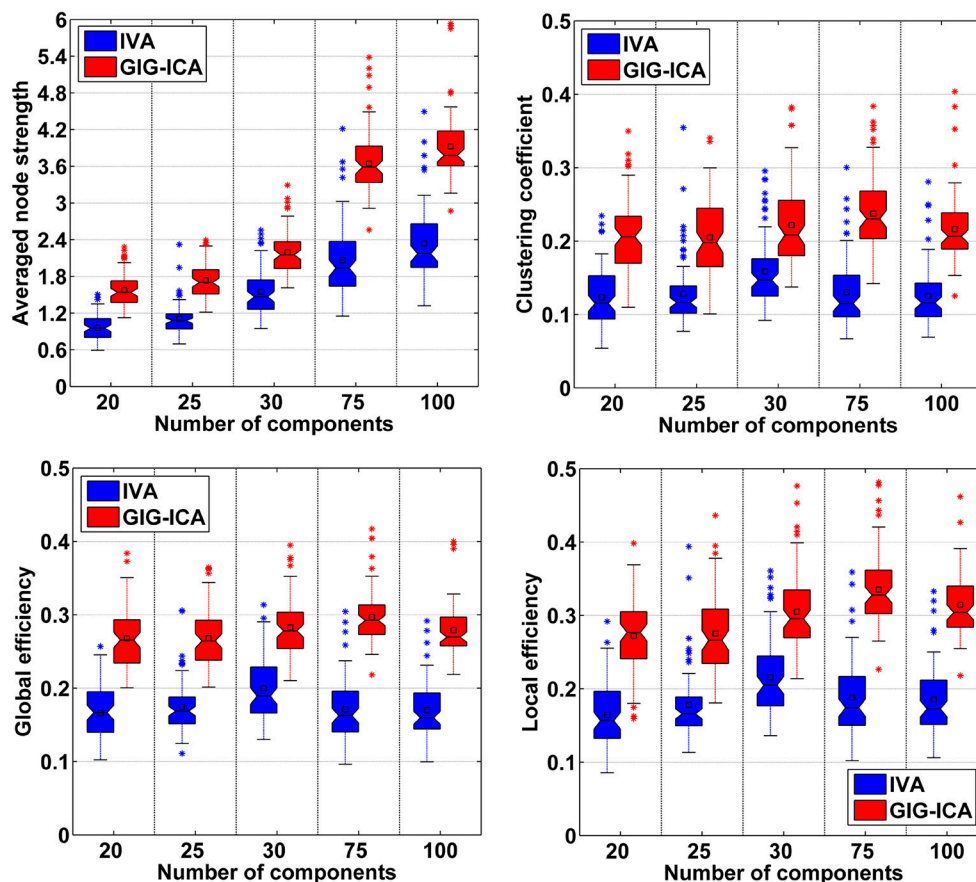


**FIGURE 10 | The summarized modularity and reliability measures of FNC for IVA and GIG-ICA under different model orders (i.e., different numbers of components).** (A) The modularity Q-value of the mean FNC. (B) Individual FNC's modularity Q-value. The Q-values of FNC matrices from all datasets are shown using one boxplot, each point of which corresponds to a Q-value of one individual FNC matrix. (C) The short-term modularity reliability. (D) The long-term modularity reliability. (E) The short-term reliability of connectivity strengths in FNC. (F) The long-term reliability of connectivity strengths in FNC. In (E,F), each boxplot includes ICC values of all connectivities.

estimation of multiple runs due to the computation load that would significantly increase the computation time. Similarly use of a more powerful IVA algorithm such as the one proposed

in Boukouvalas et al. (2015) might improve the estimation performance at the expense of computation cost. (4) Using healthy participants' test-retest resting-state fMRI data, GIG-ICA





**FIGURE 11 | Comparison of graph theory based metrics of FNC under different model orders.** In the left-top subfigure, the averaged node strength values computed based on all 75 datasets' FNC matrices from IVA/GIG-ICA with one specific model order are shown in one boxplot. Other boxplots are formatted similarly.

obtained higher network reliability as well as stronger and more reliable modularity than IVA. Network reliability is regarded as a desirable property since the fMRI data in our experiments were from healthy subjects' test-retest scans (Shehzad et al., 2009; Zuo and Xing, 2014). Previous researches (Wang et al., 2010; Bullmore and Bassett, 2011) have supported that healthy brain's intrinsic activity is organized as a small-world, highly efficient network with highly connected brain regions. Nevertheless, the truths regarding both network reliability and integration are unknown for real data. In the future, we will employ fMRI data from both healthy controls and patients with mental disorders to examine the ability of the two methods in identifying potential biomarkers.

## ETHICS STATEMENT

The fMRI data we used are open. These data are fully available via [http://www.nitrc.org/projects/nyu\\_trt](http://www.nitrc.org/projects/nyu_trt). Data were collected according to protocols approved by the institutional review boards of New York University (NYU) and the NYU School of Medicine. The related information can be found in Zuo

et al. (2010). Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach.

## AUTHOR CONTRIBUTIONS

YD designed the study; analyzed and interpreted the data; revised the manuscript; and gave final approval. DL interpreted the results; drafted the manuscript and gave final approval. QY interpreted the results; revised the manuscript and gave final approval. JS interpreted the data; revised the manuscript and gave final approval. JC interpreted the results; revised the manuscript and gave final approval. SR helped with IVA implementation, interpreted the data; drafted and revised the manuscript; gave final approval. TA interpreted the data; drafted and revised the manuscript; gave final approval. VC designed the study; interpreted the data; drafted and revised the manuscript; gave final approval.

## ACKNOWLEDGMENTS

This work was partially supported by National Institutes of Health grants R01EB006841 (to VC) and R01EB020407 (to VC),

National Science Foundation (NSF) grants 1016619 and 1539067, and a Centers of Biomedical Research Excellence (COBRE) grant P20RR021938/P20GM103472 (to VC), and by the NSF grants 1618551 and 1631838 (to TA), and by natural science foundation of Shanxi (2016021077, to YH).

## REFERENCES

- Adali, T., Anderson, M., and Fu, G. S. (2014). Diversity in independent component and vector analyses: identification, algorithms, and applications in medical imaging. *IEEE Signal Process. Mag.* 31, 18–33. doi: 10.1109/MSP.2014.2300511
- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., and Silva, R. F. (2011). A baseline for the multivariate comparison of resting-state networks. *Front. Syst. Neurosci.* 5:2. doi: 10.3389/fnsys.2011.00002
- Allen, E. A., Erhardt, E. B., Wei, Y., Eichele, T., and Calhoun, V. D. (2012). Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study. *Neuroimage* 59, 4141–4159. doi: 10.1016/j.neuroimage.2011.10.010
- Anderson, M., Adali, T., and Li, X. L. (2012). Joint blind source separation with multivariate gaussian model: algorithms and performance analysis. *IEEE Trans. Signal Proc.* 60, 1672–1683. doi: 10.1109/TSP.2011.2181836
- Anderson, M., Fu, G. S., Phlypo, R., and Adali, T. (2014). Independent vector analysis: identification conditions and performance bounds. *IEEE Trans. Signal Proc.* 62, 4399–4410. doi: 10.1109/TSP.2014.2333554
- Anderson, M., Li, X. L., and Adali, T. (2010). Nonorthogonal independent vector analysis using multivariate gaussian model. *Latent Variable Anal. Signal Sep.* 6365, 354–361. doi: 10.1007/978-3-642-15995-4\_44
- Beckmann, C. F., and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage* 25, 294–311. doi: 10.1016/j.neuroimage.2004.10.043
- Beckmann, C., Mackay, C., Filippini, N., and Smith, S. (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *Neuroimage* 47(Suppl. 1), S148. doi: 10.1016/s1053-8119(09)71511-3
- Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129
- Boukouvlas, Z., Fu, G. S., and Adali, T. (2015). “An Efficient multivariate generalized gaussian distribution estimator: application to IVA,” in *2015 49th Annual Conference on Information Sciences and Systems (CISS)* (Baltimore, MD).
- Bullmore, E. T., and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annu. Rev. Clin. Psychol.* 7, 113–140. doi: 10.1146/annurev-clinpsy-040510-143934
- Calhoun, V. D., Adali, T., Pearson, G. D., and Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151. doi: 10.1002/hbm.1048
- Calhoun, V. D., Liu, J., and Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45, S163–172. doi: 10.1016/j.neuroimage.2008.10.057
- Calhoun, V. D., Miller, R., Pearson, G. D., and Adali, T. (2014). The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuro* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015
- Dea, J. T., Anderson, M., Allen, E., Calhoun, V. D., and Adali, T. (2011). “IVA for Multi-Subject Fmri Analysis: a comparative study using a new simulation toolbox,” in *2011 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (Santander).
- Du, Y. H., Allen, E. A., He, H., Sui, J., Wu, L., and Calhoun, V. D. (2016a). Artifact removal in the context of group ICA: a comparison of single-subject and group approaches. *Hum. Brain Mapp.* 37, 1005–1025. doi: 10.1002/hbm.23086
- Du, Y. H., and Fan, Y. (2013). Group information guided ICA for fMRI data analysis. *Neuroimage* 69, 157–197. doi: 10.1016/j.neuroimage.2012.11.008
- Du, Y. H., He, H., Wu, L., Yu, Q. B., Sui, J., and Calhoun, V. D. (2015a). “Dynamic default mode network connectivity diminished in patients with schizophrenia,” in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)* (New York, NY), 474–477.
- Du, Y. H., Liu, J. Y., Sui, J., He, H., Pearson, G. D., and Calhoun, V. D. (2014). “Exploring difference and overlap between schizophrenia, schizoaffective and bipolar disorders using resting-state brain functional networks,” in *The 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Chicago, IL), 1517–1520.
- Du, Y. H., Pearson, G. D., He, H., Wu, L., Y., C. J., and Calhoun, V. D. (2015b). “Identifying brain dynamic network states via GIG-ICA: Application to schizophrenia, bipolar and schizoaffective disorders,” in *IEEE 12th International Symposium on Biomedical Imaging (ISBI)* (New York, NY), 478–481.
- Du, Y. H., Pearson, G. D., Liu, J. Y., Sui, J., Yu, Q. B., He, H., et al. (2015c). A group ICA based framework for evaluating resting fMRI markers when disease categories are unclear: application to schizophrenia, bipolar, and schizoaffective disorders. *Neuroimage* 122, 272–280. doi: 10.1016/j.neuroimage.2015.07.054
- Du, Y. H., Pearson, G. D., Yu, Q., He, H., Lin, D. D., Sui, J., et al. (2016b). Interaction among subsystems within default mode network diminished in schizophrenia patients: a dynamic connectivity approach. *Schizophr. Res.* 170, 55–65. doi: 10.1016/j.schres.2015.11.021
- Du, Y. H., Pearson, G., Lin, D. D., Sui, J., Chen, J. Y., Salman, M., et al. (2017). Identifying dynamic functional connectivity biomarkers using GIG-ICA: application to schizophrenia, schizoaffective disorder and psychotic bipolar disorder. *Hum. Brain Mapp.* 38, 2683–2708. doi: 10.1002/hbm.23553
- Erhardt, E. B., Allen, E. A., Wei, Y., Eichele, T., and Calhoun, V. D. (2012). SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *Neuroimage* 59, 4160–4167. doi: 10.1016/j.neuroimage.2011.11.088
- Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adali, T., and Calhoun, V. D. (2011). Comparison of multi-subject ICA methods for analysis of fMRI data. *Hum. Brain Mapp.* 32, 2075–2095. doi: 10.1002/hbm.21170
- Fu, G. S., Anderson, M., and Adali, T. (2014). Likelihood estimators for dependent samples and their application to order detection. *IEEE Trans. Signal Proc.* 62, 4237–4244. doi: 10.1109/TSP.2014.2333551
- Gopal, S., Miller, R. L., Michael, A., Adali, T., Cetin, M., Rachakonda, S., et al. (2016). Spatial variance in resting fMRI networks of schizophrenia patients: an independent vector analysis. *Schizophr. Bull.* 42, 152–160. doi: 10.1093/schbul/sbv085
- Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., et al. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. doi: 10.1016/j.neuroimage.2012.03.027
- Laney, J., Adali, T., Waller, S. M., and Westlake, K. P. (2015a). Quantifying motor recovery after stroke using independent vector analysis and graph-theoretical analysis. *Neuroimage Clin.* 8, 298–304. doi: 10.1016/j.nicl.2015.04.014
- Laney, J., Westlake, K. P., Ma, S., Woytowicz, E., Calhoun, V. D., and Adali, T. (2015b). Capturing subject variability in fMRI data: a graph-theoretical analysis of GICA vs. IVA. *J. Neurosci. Methods* 247, 32–40. doi: 10.1016/j.jneumeth.2015.03.019
- Lee, J. H., Lee, T. W., Jolesz, F. A., and Yoo, S. S. (2008). Independent vector analysis (IVA): multivariate approach for fMRI group study. *Neuroimage* 40, 86–109. doi: 10.1016/j.neuroimage.2007.11.019

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2017.00267/full#supplementary-material>

- Li, X. L., Adali, T., and Anderson, M. (2011). Joint blind source separation by generalized joint diagonalization of cumulant matrices. *Signal Proc.* 91, 2314–2322. doi: 10.1016/j.sigpro.2011.04.016
- Li, Y. O., Adali, T., and Calhoun, V. D. (2007). Estimating the number of independent components for functional magnetic resonance imaging data. *Hum. Brain Mapp.* 28, 1251–1266. doi: 10.1002/hbm.20359
- Liao, X., Cao, M., Xia, M., and He, Y. (2017). Individual differences and time-varying features of modular brain architecture. *Neuroimage* 152, 94–107. doi: 10.1016/j.neuroimage.2017.02.066
- Ma, S., Correa, N. M., Li, X. L., Eichele, T., Calhoun, V. D., and Adali, T. (2011). Automatic identification of functional clusters in fMRI data using spatial dependence. *IEEE Trans. Biomed. Eng.* 58, 3406–3417. doi: 10.1109/TBME.2011.2167149
- Ma, S., Phlypo, R., Calhoun, V. D., and Adali, T. (2013). “Capturing group variability using IVA: A Simulation study and graph-theoretical analysis,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Vancouver, BC), 3128–3132.
- McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., et al. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188. doi: 10.1002/(SICI)1097-0193(1998)6:3<160::AID-HBM5>3.0.CO;2-1
- Meda, S. A., Ruano, G., Windemuth, A., O’Neil, K., Berwise, C., Dunn, S. M., et al. (2014). Multivariate analysis reveals genetic associations of the resting default mode network in psychotic bipolar disorder and schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6864–6864. doi: 10.1073/pnas.1313093111
- Michael, A. M., Anderson, M., Miller, R. L., Adali, T., and Calhoun, V. D. (2014). Preserving subject variability in group fMRI analysis: performance evaluation of GICA vs. IVA. *Front. Syst. Neurosci.* 8:106. doi: 10.3389/fnsys.2014.00106
- Newman, M. E. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8577–8582. doi: 10.1073/pnas.0601602103
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 74, (3 Pt 2), 036104. doi: 10.1103/physreve.74.036104
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Schmithorst, V. J., and Holland, S. K. (2004). Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *J. Magn. Reson. Imaging* 19, 365–368. doi: 10.1002/jmri.20009
- Shehzad, Z., Kelly, A. M., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., et al. (2009). The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229. doi: 10.1093/cercor/bhn256
- Svensén, M., Kruggel, F., and Benali, H. (2002). ICA of fMRI group study data. *Neuroimage* 16, 551–563. doi: 10.1006/nimg.2002.1122
- Wang, J., Zuo, X., and He, Y. (2010). Graph-based network analysis of resting-state functional MRI. *Front. Syst. Neurosci.* 4:16. doi: 10.3389/fnsys.2010.00016
- Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010). Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. *Neuroimage* 49, 2163–2177. doi: 10.1016/j.neuroimage.2009.10.080
- Zuo, X. N., and Xing, X. X. (2014). Test-retest reliabilities of resting-state fMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Du, Lin, Yu, Sui, Chen, Rachakonda, Adali and Calhoun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Psychophysiological Interactions in a Visual Checkerboard Task: Reproducibility, Reliability, and the Effects of Deconvolution

Xin Di and Bharat B. Biswal\*

Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, United States

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Institute of Psychology (CAS), China

### Reviewed by:

Donald G. McLaren,  
University of Wisconsin-Madison,  
United States  
Sheng Zhang,  
Yale University, United States

### \*Correspondence:

Bharat B. Biswal  
bbiswal@yahoo.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 02 June 2017

**Accepted:** 02 October 2017

**Published:** 17 October 2017

### Citation:

Di X and Biswal BB (2017)  
Psychophysiological Interactions in a  
Visual Checkerboard Task:  
Reproducibility, Reliability, and the  
Effects of Deconvolution.  
Front. Neurosci. 11:573.  
doi: 10.3389/fnins.2017.00573

Psychophysiological interaction (PPI) is a regression based method to study task modulated brain connectivity. Despite its popularity in functional MRI (fMRI) studies, its reliability and reproducibility have not been evaluated. We investigated reproducibility and reliability of PPI effects during a simple visual task, and examined the effect of deconvolution on the PPI results. A large open-access dataset was analyzed ( $n = 138$ ), where a visual task was scanned twice with repetition times (TRs) of 645 and 1,400 ms, respectively. We first replicated our previous results by using the left and right middle occipital gyrus as seeds. Then regions of interest (ROI)-wise analysis was performed among 20 visual-related thalamic and cortical regions, and negative PPI effects were found between many ROIs with the posterior fusiform gyrus as a hub region. Both the seed-based and ROI-wise results were similar between the two runs and between the two PPI methods with and without deconvolution. The non-deconvolution method and the short TR run in general had larger effect sizes and greater extents. However, the deconvolution method performed worse in the 645 ms TR run than the 1,400 ms TR run in the voxel-wise analysis. Given the general similar results between the two methods and the uncertainty of deconvolution, we suggest that deconvolution may be not necessary for PPI analysis on block-designed data. Lastly, intraclass correlations (ICC) between the two runs were much lower for the PPI effects than the activation main effects, which raise cautions on performing inter-subject correlations and group comparisons on PPI effects.

**Keywords:** reproducibility, reliability, test-retest, psychophysiological interaction, deconvolution

## INTRODUCTION

Psychophysiological interaction (PPI) is a widely used method to study task related brain functional connectivity changes (Friston et al., 1997). It employed simple regression-based method to model task modulated connectivity effects, thus enabling whole brain exploratory analysis. Therefore, even though there are more sophisticated methods available, e.g., dynamic causal modeling (Friston et al., 2003), PPI is still a valuable method for functional MRI (fMRI) data, given that our knowledge on large-scale task related connectivity is still quite limited. Several modifications of the PPI method have been made after it was proposed, including adding a deconvolution step to deal with the asynchrony between task design and fMRI hemodynamic response (Gitelman et al., 2003) and introducing a generalized framework to model more than two experimental conditions (McLaren et al., 2012).



A PPI effect is defined as an interaction between the time series of a brain region (physiological variable) and a (or more) task design variable (psychological variable). Noises of both the physiological and psychological variables go into the interaction term, so that the interaction effect is much noisier than the main effects of task free connectivity (physiological main effect) and task activation (psychological main effect). This makes PPI analysis having lower statistical power than simple connectivity and conventional activation analysis. Since PPI analysis has been increasingly used to study group differences and inter-subjects variability, it is important to evaluate the reproducibility and reliability of the PPI methods (Vul et al., 2009; Dubois and Adolphs, 2016). Voxel-based meta-analysis has been used to examine consistency of PPI results across studies (Di et al., 2017a). However, because the tasks used in different studies varied greatly, the motivation of a meta-analysis on PPI was rather to identify different connectivity that were modulated by different tasks, than to simply identify consistent connectivity cross studies with different tasks (Di et al., 2017a). Nevertheless, the reliability of PPI effect has not been directly examined.

One critical step for the PPI method is to properly deal with the asynchrony between task design and observed blood-oxygen-level dependent (BOLD) signals. An earlier solution is to convolve the psychological variable with hemodynamic response function (HRF). Then the PPI term  $x_{PPI}^1$  could be expressed as:

$$x_{PPI}^1 = x_{physio} \cdot (z_{psych} * hrf) \quad (1)$$

where  $x_{physio}$  represents the physiological variable,  $z_{psych}$  represents the psychological design variable, and  $*$  represents convolution operator. However, this calculation is not appropriate if the interaction happened faster than the slow hemodynamic response. Therefore, a deconvolution procedure is required (Gitelman et al., 2003) to find a variable  $z_{physio}$  that:

$$x_{physio} = z_{physio} * hrf \quad (2)$$

If this could be achieved, then the interaction could be calculated at the neuronal level and then convolve with HRF:

$$x_{PPI}^2 = (z_{psych} \cdot z_{physio}) * hrf \quad (3)$$

We can also put Equation (2) to Equation (1), so that:

$$x_{PPI}^1 = (z_{psych} * hrf) \cdot (z_{physio} * hrf) \quad (4)$$

Mathematically,  $x_{PPI}^1$  and  $x_{PPI}^2$  are not equivalent. Therefore, deconvolution seems necessary. However, effective deconvolution depends on assumptions such as, known HRF and noise characteristics in the BOLD signals (Roebroeck et al., 2011; O'Reilly et al., 2012). Unfortunately, there are substantial amount of variability in HRF both across brain regions and across subjects (Handwerker et al., 2004). On the other hand, if a task design is slower than the hemodynamic response, e.g., a blocked design, the PPI terms calculated from the above mentioned two methods could be very similar. We have demonstrated that the PPI results of a block-designed visual task

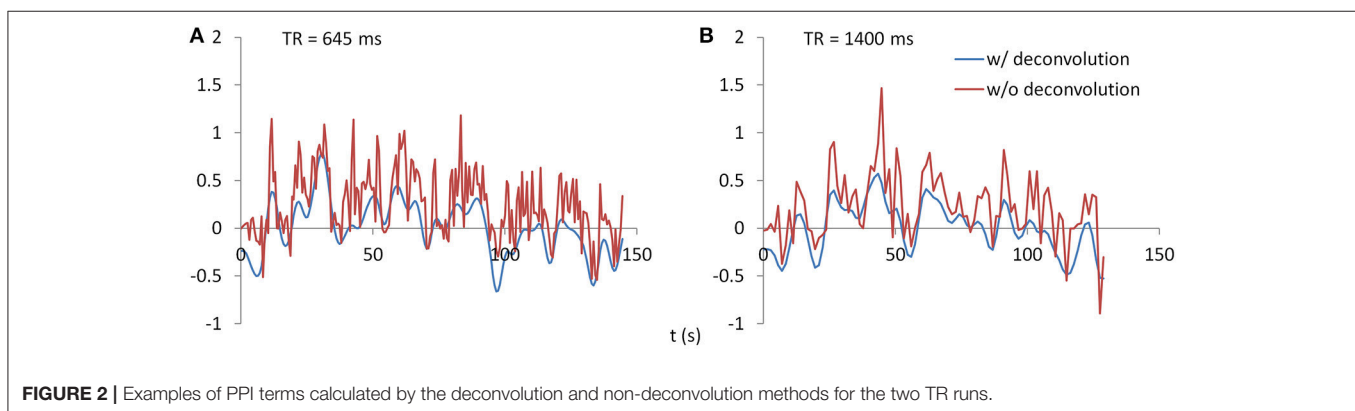
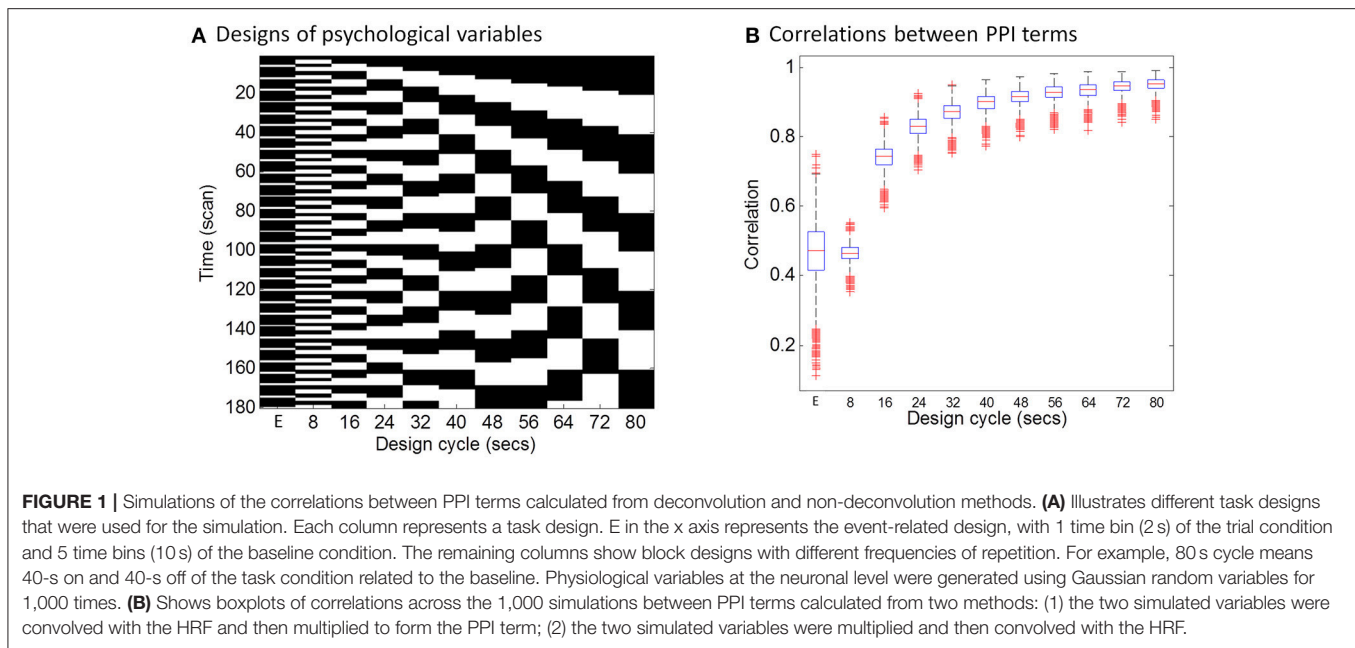
are spatially corresponding very well between the deconvolution and non-deconvolution PPI methods (Di et al., 2017b). Whether to perform deconvolution then needs to compromise between the deviation between the PPI terms calculated in different ways and the uncertainty of deconvolution (Di et al., 2017b). Therefore, it might be better to not perform deconvolution for a block-designed task, which is actually recommended by FSL (FMRIB Software Library; Jenkinson et al., 2012; O'Reilly et al., 2012). For event-related designed task, however, deconvolution may be still necessary, because the PPI terms calculated from the deconvolution and non-deconvolution methods may be dramatically different.

We recently demonstrated negative PPI effects (reduced connectivity) between the middle occipital gyrus to the fusiform gyrus and supplementary motor areas in a simple block-designed checkerboard task compared with a fixation baseline (Di et al., 2017b). Here, we further analyzed a larger sample of checkerboard data ( $n = 138$ ) of two separate runs with two repetition times (TR: 645 and 1,400 ms; Nooner et al., 2012). The aims of the current study are to first evaluate reproducibility and reliability of PPI effects in the checkerboard task. Additionally, we investigated the impact of PPI calculation methods on the PPI results and their reproducibility and reliability. We operationally defined reproducibility as whether previously reported clusters could be observed in the current analysis, and whether the clusters reported in one run could be observed in the other run. Quantitatively, we utilized Dice coefficient to quantify overlaps of voxels on thresholded maps (Rombouts et al., 1998; Taylor et al., 2012). Next, we used intraclass correlation (ICC) to quantify test-retest reliability. Because the short TR run has about twice the number of time points as the long TR one, we predict that statistical results would be better for the short TR run compared with the long TR run. In addition, shorter sampling rate may provide more accurate estimate of hemodynamic response, therefore deconvolution PPI method should work better for the short TR than the long TR runs.

## METHODS

### Simulations on the Correlations between PPI Terms

The hemodynamic response is a slow response compared with neuronal events, which can be understood as a low-pass filter. Intuitively, if a task design is slow enough, e.g., a blocked design, the convolution with the HRF may not affect PPI calculations much. To directly demonstrate this relationship between design alternating length and the effect of convolution on PPI calculation, we firstly performed a simulation. In this simulation, we defined a simple block-designed task with equal on and off periods with different cycle lengths (from 8 to 80 s), and a simple event-related design with fixed inter trial interval of 12 s (Figure 1A). We used a typical sampling rate of 2 s, so that the event-related design could be expressed as alterations of one time bin (2 s) of a trial and five time bins (10 s) of the baseline condition (The first column in Figure 1A). The remaining columns in Figure 1A show block designs with



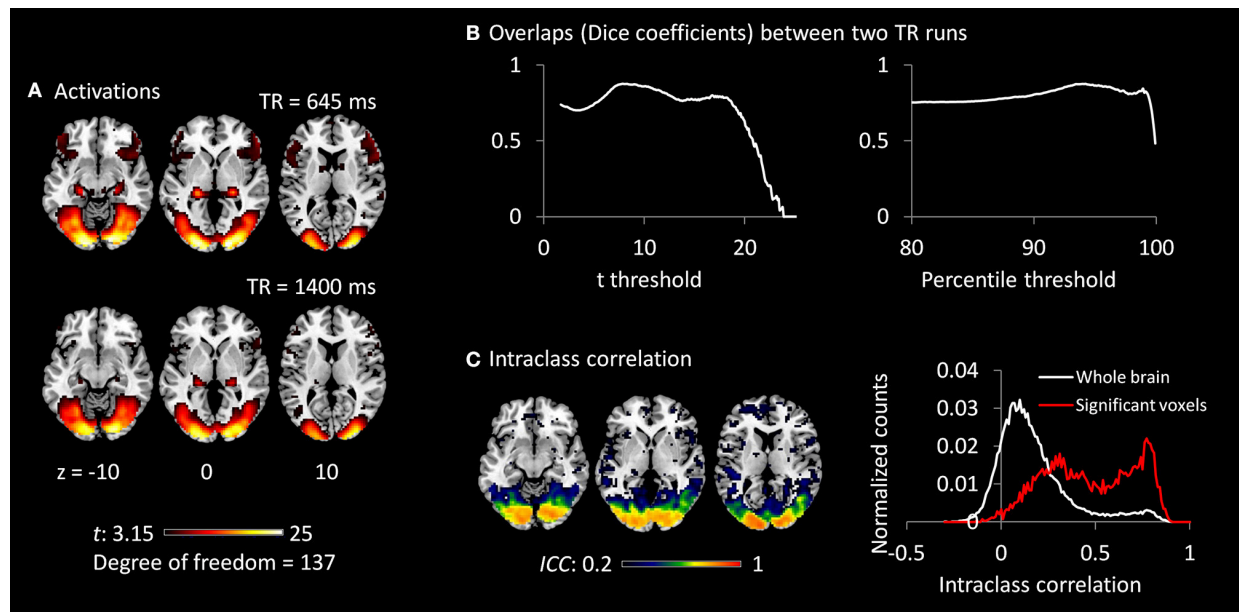
different frequencies of repetition. For example, 80 s cycle means 40-s on and 40-s off of the task condition related to the baseline. We then simulated the physiological variable of neuronal activities as a Gaussian variable for 1,000 times. For each design and simulated “neuronal” physiological variable, we calculated PPI terms using two ways: (1) each variable convolved with the canonical HRF and then the two convolved variables were multiplied to form a PPI term (corresponding to  $x_{PPI}^1$  in Equation 4); (2) the two variables were multiplied and then convolved with the canonical HRF (corresponding to  $x_{PPI}^2$  in Equation 3). We then calculated the correlations of the PPI terms calculated from the two methods. The code for this simulation can be found at: [https://github.com/dixy0/PPI\\_correlation\\_demo](https://github.com/dixy0/PPI_correlation_demo).

## fMRI Data and Task Design

We used the checkerboard fMRI data with TRs of 645 and 1,400 ms from the release 1 of Enhanced Nathan Kline Institute—Rockland Sample ([http://fcon\\_1000.projects.nitrc.org/indi/enhanced/](http://fcon_1000.projects.nitrc.org/indi/enhanced/)). One hundred and forty-six subjects’

data with age equal or larger than 20 years old were included for analysis. Six subjects’ data were discarded due to large head motion during fMRI scanning in any of the two scans (maximum frame-wise displacement (FD) (Di and Biswal, 2015)  $> 1.5$  mm or  $1.5^\circ$ ). One subject’s data were deleted because of poor coverage of the lower occipital lobe, and another subject’s data were deleted because of failure of coregistration and normalization. The effective number of subjects was 138 (89 females, 45 males, 1 unidentified). The mean age of the sample was 47.8 years (20–83 years).

The checkerboard task consisted of 20 s fixation block and 20 s flickering checkerboard block repeated three times. A blank screen was presented after the third checkerboard block until fMRI scan was complete. The task was scanned for two separate runs with two TRs: 645 and 1,400 ms, respectively. For the 645 ms run, 239 or 240 fMRI images were scanned for each subject. The following parameters were used: TR = 645 ms; TE = 30 ms; flip angle =  $60^\circ$ ; voxel size =  $3 \times 3 \times 3$  mm<sup>3</sup> isotropic; number of slices = 40. For the 1,400 ms run, 98 fMRI images were



**FIGURE 3 | (A)** Activations ( $t$  maps) of visual checkerboard presentation for the 645 ms TR run (upper) and 1,400 ms TR run (lower). The threshold  $t$ -value corresponds to one-tailed significance at  $p < 0.001$ . **(B)** Overlaps (Dice coefficients) between the two TR runs using  $t$  threshold (left) and percentile threshold (right). **(C)** Test-retest reliability map (intraclass correlations, ICC) of activations between the two runs is shown on the left, which were thresholded at  $\text{ICC} > 0.2$ . The histograms of ICC of activations between the two TR runs in significant voxels and whole brain are shown on the right. The significant voxels were determined using intersection of the two runs each thresholded at  $p < 0.001$ .

scanned for each subject. The following parameters were used: TR = 1,400 ms; TE = 30 ms; flip angle =  $65^\circ$ ; voxel size =  $2 \times 2 \times 2 \text{ mm}^3$  isotropic; number of slices = 64. Anatomical T1 images were scanned using MPRAGE (magnetization-prepared rapid acquisition with gradient echo) sequence with the following parameters: TR = 1,900 ms; TE = 2.52 ms; flip angle =  $9^\circ$ ; voxel size =  $1 \times 1 \times 1 \text{ mm}^3$  isotropic. More information of the data can be found in Nooner et al. (2012).

## fmRI Data Analysis

### fmRI Data Preprocessing

Functional MRI (fmRI) data preprocessing and analysis were performed using SPM12 software (<http://www.fil.ion.ucl.ac.uk/spm/>) under MATLAB environment (<http://www.mathworks.com/>). For the 645 ms run, the first 14 images (9 s) were discarded from analysis, resulting in 225 images for each subject. For the 1,400 TR run, the first five images (7 s) were discarded from analysis, resulting in 93 images for each subject. The functional images were motion corrected, and coregistered to subject's anatomical images. The anatomical images were segmented, and the deformation field images were used to normalize the functional images into MNI space. The data from the two TR runs were both resliced and resampled at a spatial resolution of  $3 \times 3 \times 3 \text{ mm}^3$ . Lastly, the functional images were smoothed using a 6 mm full width at half maximum (FWHM) Gaussian kernel.

### Activation Analysis

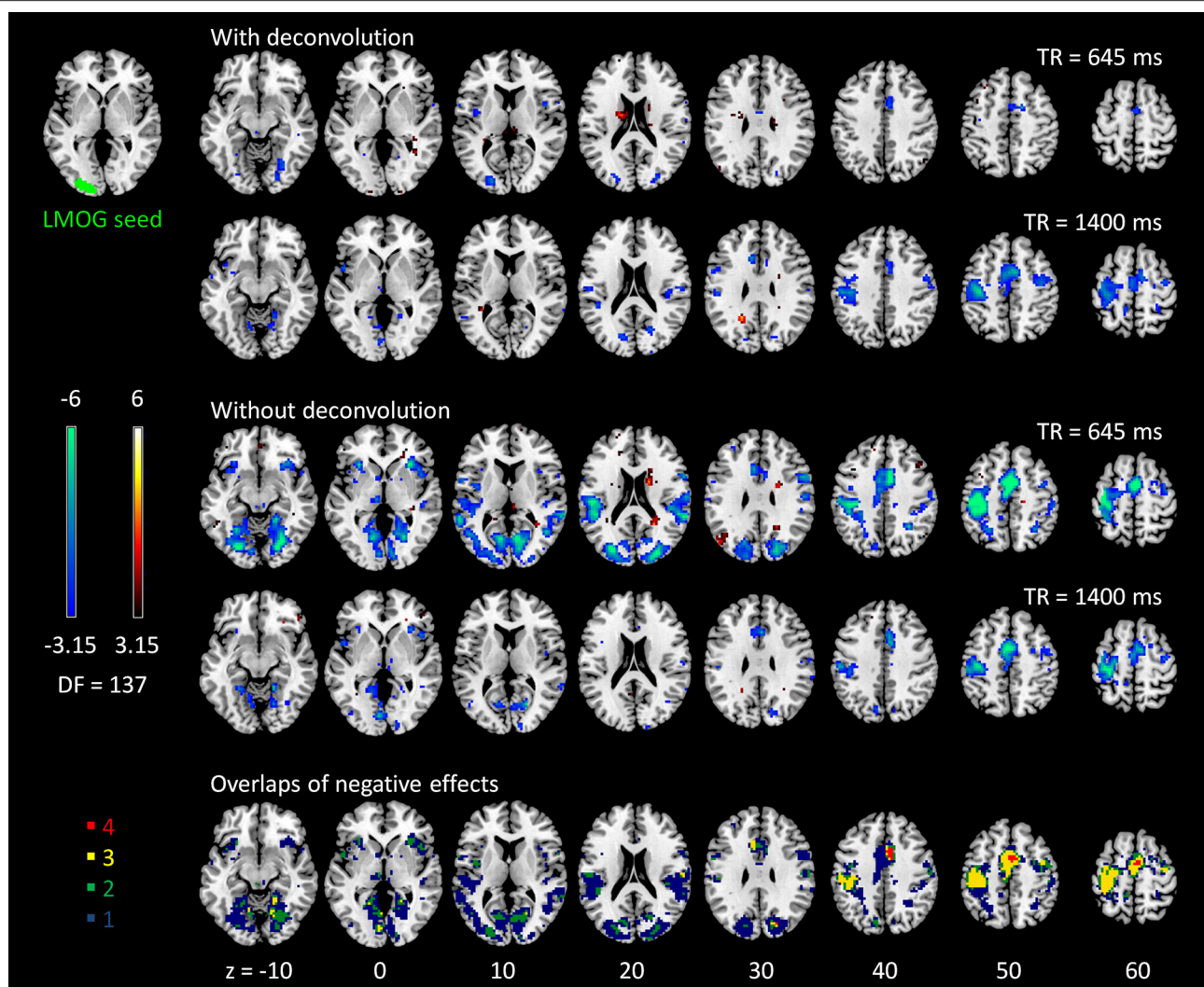
We first defined functional ROIs of the visual thalamus and lower visual area by performing general linear model (GLM) analysis

on the checkerboard task. The checkerboard task was modeled as a box-car function, with 1 representing the checkerboard condition and 0 representing the fixation or blank screen. The box-car function was convolved with the canonical HRF to form a predictor of BOLD responses. Two regressors of the first eigenvariate of BOLD signals in white matter and cerebrospinal fluid (CSF), and 24 regressors of Friston's autoregressive head motion model (Friston et al., 1996) were also added in the model as covariates. An implicit high-pass filter of 1/128 Hz was also implemented in the model. The high-pass filtering is accomplished in SPM by using discrete cosine transform functions. The effective high-pass filtering cutoffs were then 0.0069 Hz (1/145.125 s) for the 645 ms TR run and 0.0077 Hz (1/130.2 s) for the 1,400 ms TR run. The GLM model was estimated for each voxel in the brain to identify regions that showed similar patterns of activations as the task design. The beta maps of task activation were used for group level analysis using a one sample  $t$ -test model. Statistical significant clusters were identified by using cluster level statistics based on random field theory. Clusters were first identified using a one-tailed  $t$ -test at  $p < 0.001$ , and cluster extent was determined using false discovery rate (FDR) at  $p < 0.05$ .

### Definition of Regions of Interest

We performed two types of PPI analyses, voxel-wise analysis using seed regions that were activated by the checkerboard task and ROI-based analysis among visual thalamus and cortical visual areas independently defined from other toolbox. In the activation analysis of the current data, the posterior visual cortex



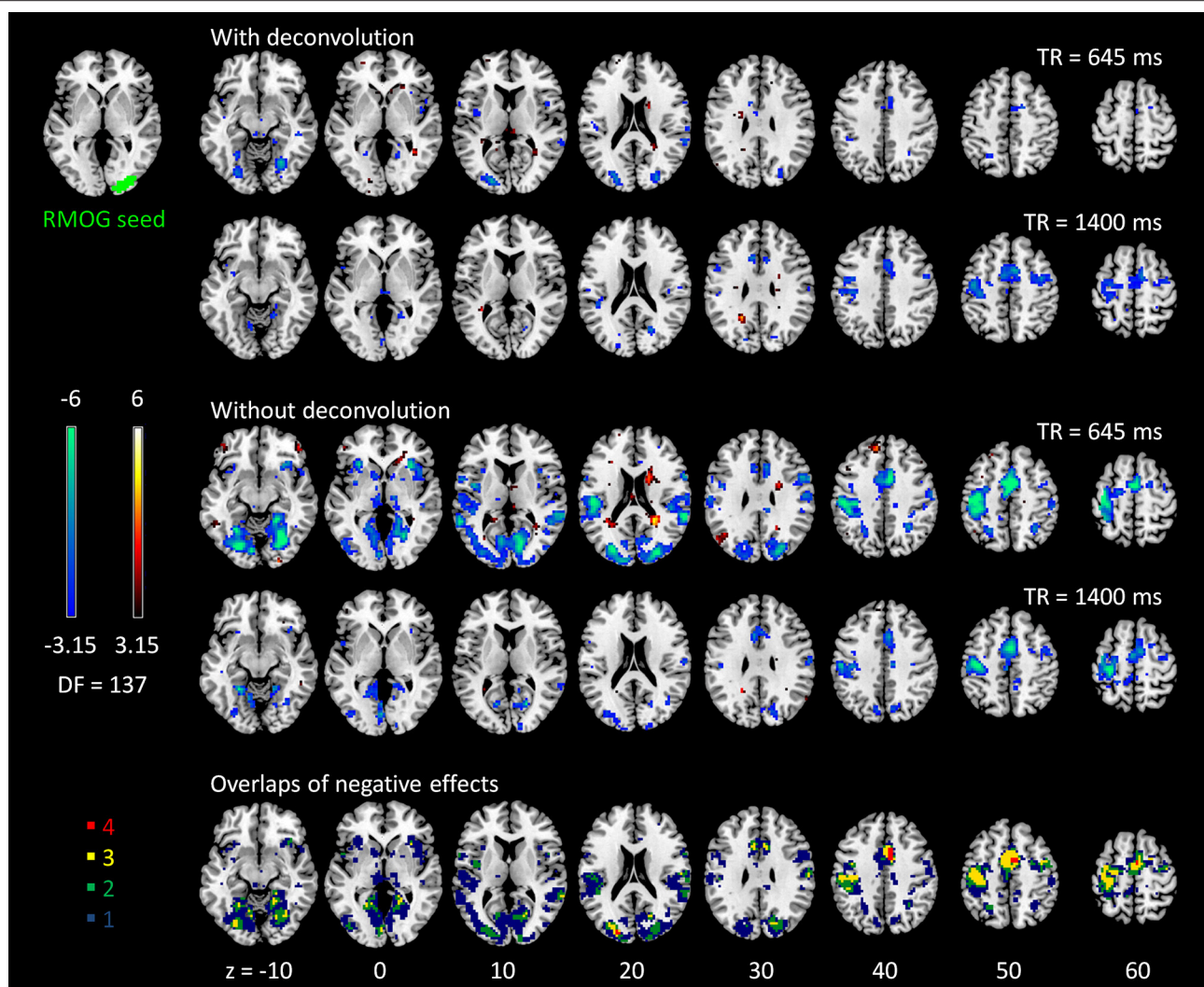


**FIGURE 4 |** Psychophysiological interaction (PPI) results for the left middle occipital gyrus (LMOG) seed during checkerboard presentation in the two runs of TR (repetition time) 645 ms and TR 1,400 ms. The resulting clusters were thresholded at  $p < 0.001$  (approximated  $t = 3.15$ ), with DF (degree of freedom) of 137. The last row illustrates the number of overlapped negative PPI results in the four scenarios. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

and the posterior portion of the thalamus were robustly activated by the visual checkerboard stimulation in both TR runs. We therefore defined the left and right middle occipital gyrus (LMOG and RMOG) and the thalamus as regions of interest (ROIs) based on the activations. To define the ROIs with proper size, we increase the threshold to  $t > 16$  to define the LMOG and RMOG, and made an intersection between the two runs. The size of LMOG was 222 voxels, and the size of RMOG was 259 voxels. Thalamus was defined using a threshold of  $p < 0.001$ , with an intersection between the two runs. Because the visual thalamus is small, left and right ROIs were combined to form a single thalamus ROI (171 voxels). Different thresholds were chosen to ensure that these ROIs are similar in size. The eigenvariate of a ROI was extracted with adjustment of effects of no interests (head motion, WM/CSF variables, and low frequency drifts).

For the ROI-based analysis, we defined the visual thalamus as the regions that show functional associations with the lateral visual network in resting-state (Yuan et al., 2016). Cortical visual areas were defined by using probabilistic cytoarchitectonic maps. These areas include the OC1/OC2 (occipital cortex; Amunts et al., 2000), ventral and dorsal OC3 and OC4 (Rottschy et al., 2007; Kujovic et al., 2013), OC5 (Malikovic et al., 2006), and FG1/FG2 (fusiform gyrus; Caspers et al., 2013). For the probabilistic maps of these regions, we first performed a winner-takes-all algorithm to define unique regions of each area, and then split them into left and right regions. As a result, there are 20 ROIs (left and right thalamus, OC1, OC3, OC3d, OC3v, OC4d, OC4v, OC5, FG1, and FG2). The eigenvariate of a ROI was extracted with adjustment of effects of no interests (head motion, WM/CSF variables, and low frequency drifts).





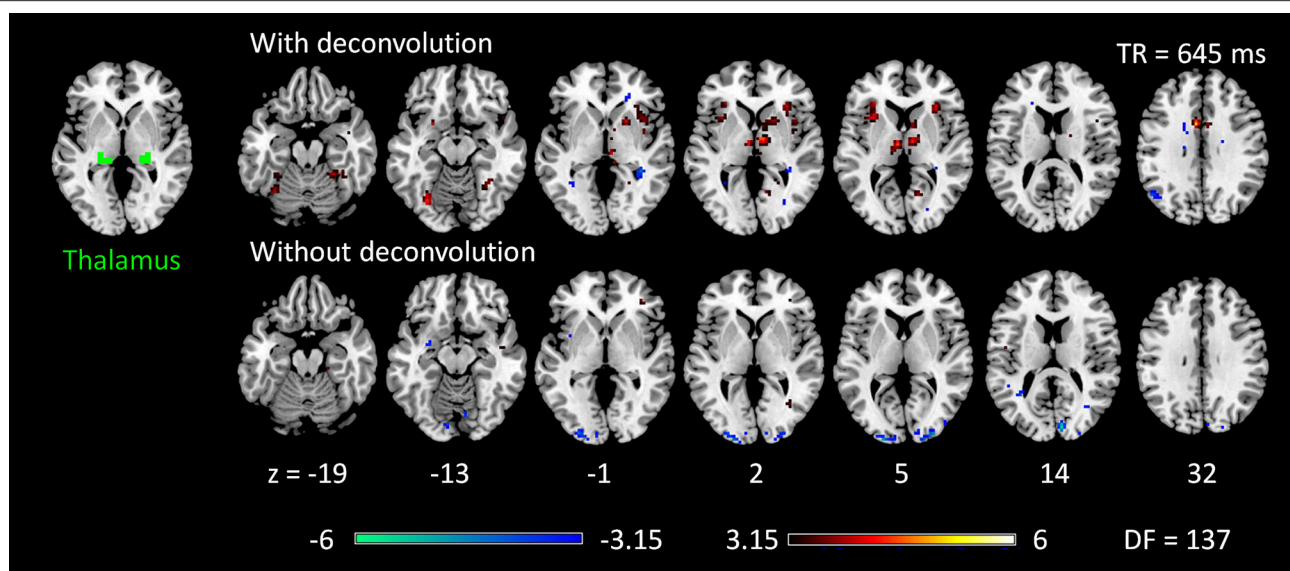
**FIGURE 5 |** Psychophysiological interaction (PPI) results for the right middle occipital gyrus (RMOG) seed during checkerboard presentation in the two runs of TR (repetition time) 645 ms and TR 1,400 ms. The resulting clusters were thresholded at  $p < 0.001$  (approximated  $t = 3.15$ ), with DF (degree of freedom) of 137. The last row illustrates the number of overlapped negative PPI results in the four scenarios. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

### Psychophysiological Interaction Analysis

PPI analysis was performed using SPM12 with updates 6685. PPI terms were calculated by using both deconvolution method and non-deconvolution method. For the deconvolution method, the time series of a seed region was deconvolved with the canonical HRF, multiplied with the centered psychological box-car function, and convolved back with the HRF to form a predicted PPI time series at hemodynamic response level. For the non-deconvolution method, the box-car function of psychological design was convolved with the HRF to form a psychological variable, and it was centered and multiplied with the raw seed time series. **Figure 2** shows examples of PPI terms calculated from the two methods in the two TR runs.

For voxel-wise PPI analysis, separate GLMs were built for the LMOG, RMOG, and thalamus seeds, and for the two TR

runs. The models included one regressor representing the task activation, one regressor representing the seed time series, one regressor representing the PPI term, and the covariates the same as the activation GLMs described above. Group-level one sample  $t$ -test was used on the corresponding PPI effects, to test where in the brain showed consistent PPI effects with a seed region. For both positive and negative contrasts, a one-tailed  $t$ -test of  $p < 0.001$  was first used to define clusters, and then a FDR cluster threshold of  $p < 0.05$  was used to identify statistical significant clusters. For the ROI-wise analysis, PPI GLM models were built for each of the 20 ROIs, and applied to all other ROIs as a dependent variable. The GLM model included one psychological variable, one physiological variable, one PPI variable, and one constant term. The covariates were not included because they have already been regressed out from all ROI time



**FIGURE 6 |** Psychophysiological interaction (PPI) results for the thalamus seed during checkerboard presentation in the TR (repetition time) run of 645 ms. There is no significant PPI effects of the thalamus seed in TR run of 1,400 ms. The resulting clusters were thresholded at  $p < 0.001$  (approximated  $t = 3.15$ ), with DF (degree of freedom) of 137. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

series. PPI effects were calculated between each pair of ROIs, resulting in a  $20 \times 20$  matrix of beta values for each subject. The matrices were symmetrized by averaging corresponding upper and lower diagonal elements (Di et al., 2017b), with a total of 190 ( $20 \times 19/2$ ) unique effects. Group-level one-sample  $t$ -test was performed on each element of the matrix. For both positive and negative contrasts, a one-tailed  $t$ -test of  $p < 0.001$  was used to identify significant PPI effects. This threshold was chosen to match with voxel-wise analysis. We also used FDR correction on the total of 190 effects. And the results are similar to what using a  $p < 0.001$  threshold. However, FDR depends on the distribution of all tested  $p$ -values, making it difficult to compare between two runs. Therefore, we adopted  $p < 0.001$  to report ROI-based PPI results.

### Reproducibility and Reliability

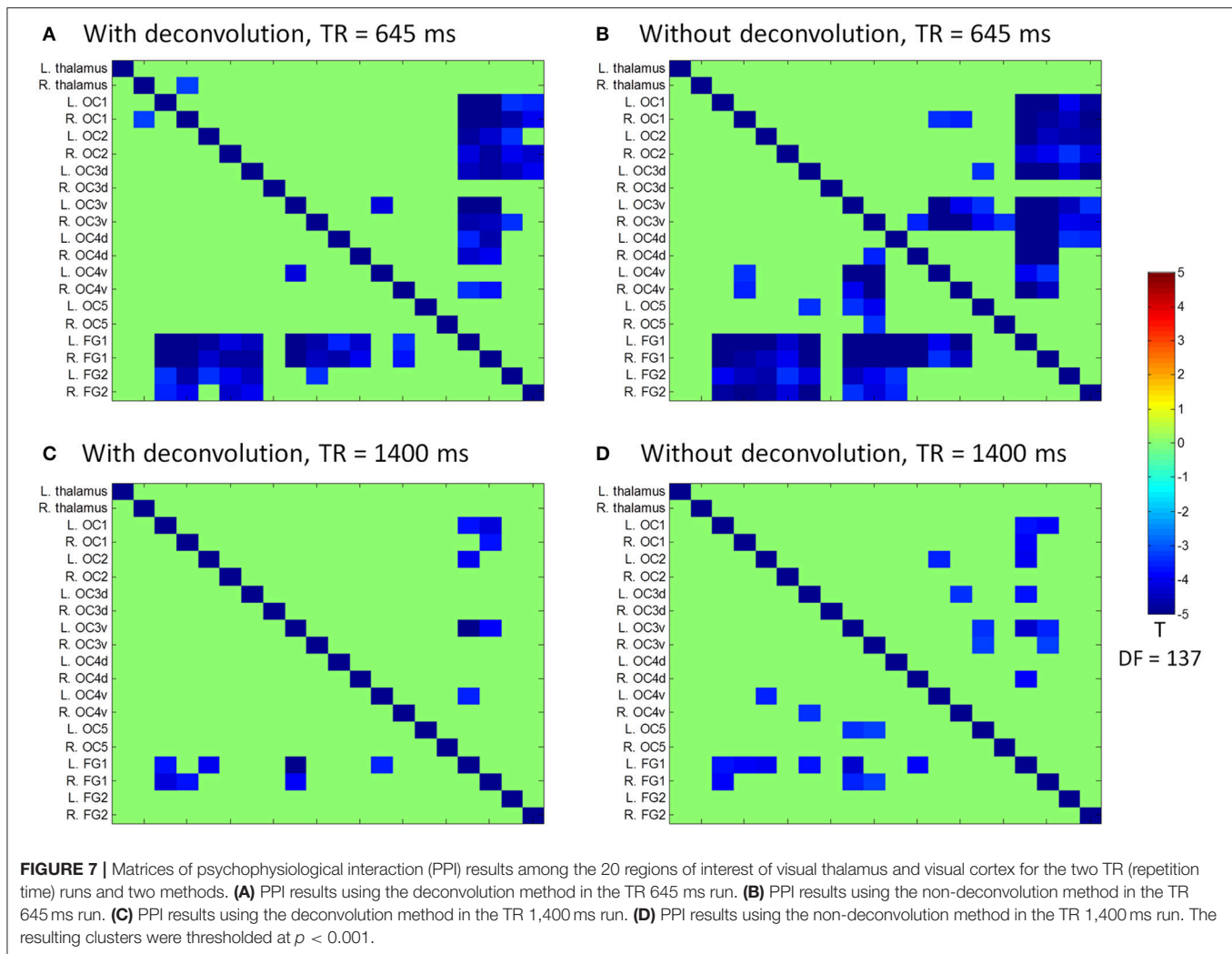
We operationally define reproducibility as overlaps of supra-threshold clusters. Dice coefficient was used to quantify reproducibility (Rombouts et al., 1998). Two strategies were used to threshold the maps or matrix from the two TR runs. First, statistical  $t$  maps or  $t$  matrices from the two TR runs were thresholded using a common  $t$ -value, ranging from 1.7 (approximately corresponds to  $p < 0.05$ ) to 7. However, it is possible that the effect sizes in the two TR runs are systematically different, so that using a same  $t$ -value could generate very different numbers of supra-threshold voxels or elements in the two runs. Therefore, we also thresholded  $t$  maps or  $t$  matrices based on the percentile of  $t$ -values within a map or matrix. This could ensure that the numbers of supra-threshold voxels or elements are the same between the two TR runs.

We operationally define reliability as test-retest reliability between the two TR runs, as quantified as ICC (Zuo et al.,

2010a). Voxel-wise ICC maps or each ROI and ICC matrices across 20 ROIs were calculated between two TR runs for each PPI method. At each voxel or matrix element, ICC was calculated from a 138 (subject) by 2 (run) matrix by using a MATLAB function written by Zuo et al. (2010a). Because only voxels that have significant effects might show meaningful reliability, we displayed histograms of ICCs within significant voxels or elements with reference to those in the whole brain. For task activations, the significant voxels were determined using intersection of the two TR runs each thresholded at  $p < 0.001$ . For PPI effects of each ROI, the significant voxels were determined using intersection of the two TR runs and two methods each thresholded at  $p < 0.01$ . This slightly liberal threshold was chosen to ensure enough number of voxels survived in the conjunction of the four scenarios. The whole brain mask was determined as all voxels in the brain, including WM and CSF.

### Coefficient of Variation

We calculated coefficient of variation to estimate measurement error of task activations and PPI effects. Coefficient of variation was calculated in ROIs that showed significant activation effects. Specifically, the LMOG, RMOG, and thalamus ROIs that were used as seed in the PPI analysis were used to represent activation effects. For the PPI results, we performed a conjunction analysis of the voxel-wise negative PPI effects across all the eight contrasts ( $2$  PPI methods  $\times 2$  TR runs  $\times 2$  seeds) using a threshold of  $p < 0.01$ , and identified 27 ROIs that showed common negative PPI effects. Beta values of activations or PPI effects of these ROIs were extracted. Coefficient of variation was calculated based on the method assuming the variation is proportional to the mean (Bland and Altman, 1996). It measures within subject variations (across the two TR runs in the current case) relative to the



mean effects of the two runs. Specifically, coefficient of variation was calculated based on a 138 (subject)  $\times$  2 (run) matrix. The beta values were first logarithmic transformed. Variation was then calculated for each subject, and a square root of mean variations across subjects was calculated. The resulting value was then transformed back using an exponential function, and subtracted by 1. The script for calculating coefficient of variation is available at: [https://github.com/dixy0/PPI\\_correlation\\_demo](https://github.com/dixy0/PPI_correlation_demo). The resulting value represents the percentage of variation of a measure relative to the mean. Coefficients of variation were calculated on the LMOG, RMOG, and thalamus ROIs to reflect measurement errors of the task activations, and were calculated on the 27 ROIs from the analyses of the LMOG and RMOG seeds to reflect measurement errors of the PPI effects.

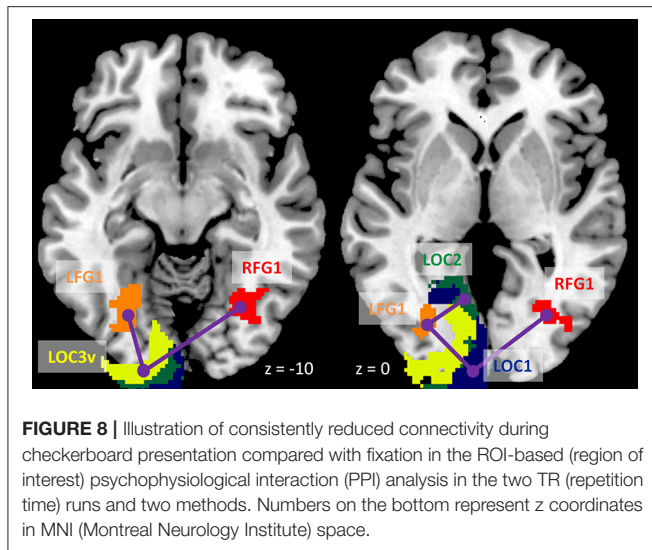
## RESULTS

### Simulations on the Correlations between PPI Terms

The distributions of PPI correlations for each task design are shown in **Figure 1B**. For the block designs, the PPI correlations

are a function of block cycle length. With longer design cycle, e.g.,  $>40$  s (20-s on and 20-s off), the correlations of PPI terms could be higher than 0.9. Practically, most of the block-designed fMRI experiments have longer block cycles than 20-s on and 20-s off. If the block alterations become faster, the correlation between PPI terms decreased. And for the event-related design, the mean PPI correlations were below 0.5 and with large variations. This simulation demonstrates that if a neuronal activity time series is known, using convolved time series to calculate PPI term (i.e.,  $x_{PPI}^1$ ) could be very similar to what calculated by first multiplying the two variables and then convolving (i.e.,  $x_{PPI}^2$ ) for typical block designed experiments. In real fMRI data, the “neuronal” physiological variable is not known, and has to be estimated by using deconvolution. Considering the similarities of the PPI terms and the caveats of deconvolution, PPI calculations without deconvolution may be a better choice for block designed experiments. On the other hand, the PPI correlations in the event-related design are much smaller ( $r < 0.5$ , meaning  $<25\%$  of shared variance). So that deconvolution is still a necessary step for PPI analysis in event-related designed experiments.





## Activations of the Checkerboard Task

Both TR runs showed highly significant activations in the visual cortex, as well as in the posterior portion of the thalamus (**Figure 3A**). The overlaps (Dice coefficients) of thresholded  $t$  maps between the two TR runs were as high as 0.7 (**Figure 3B**) at most of the shown  $t$  range or percentile range. And Dice coefficients went down when only extremely activated voxels were thresholded. The visual cortex regions also showed high test–retest reliability ( $ICC > 0.7$ ; **Figure 3C**). However, the activations of the thalamus only showed small test–retest reliability around 0.2. The histograms of ICCs in the significant voxels and in the whole brain are shown on the right of **Figure 3C**.

## Psychophysiological Interactions

The voxel-wise PPI analysis of the LMOG and RMOG seeds conveyed very similar patterns. The PPI effects of the LMOG seed for the two TRs and two methods are shown in **Figure 4**. We first observed that even though spatial extents of PPI effects varied across the two TR runs and two PPI methods, the negative PPI effects in previously reported regions, i.e., supplementary motor area and higher visual cortex, could be observed from all four scenarios. The deconvolution method in 645 ms TR run had the smallest spatial extent and statistical significance, while the non-deconvolution method in 645 ms TR run had the largest spatial extent and strongest statistical significance. Both methods in TR of 1,400 ms showed similar spatial extent and significance levels. The last row in **Figure 4** demonstrates the overlaps of negative effects in the four scenarios. Similar results were found in the analysis of the RMOG seed (**Figure 5**).

The voxel-wise PPI analysis of the thalamus seed only showed significant effects in the 645 TR run, but with different brain regions with opposite effects in the two PPI methods (**Figure 6** and Table S1). With deconvolution method, the thalamus seed showed significant positive PPI effects with the middle cingulate gyrus, anterior portion of the thalamus, bilateral anterior insula,

basal ganglia, and right fusiform gyrus. Whereas, with non-deconvolution method, the thalamus seed showed significant negative PPI effects with the bilateral occipital pole regions. There were no consistent results between two TR runs and two methods. Therefore, subsequent analysis was only performed on the LMOG and RMOG seeds.

We next performed ROI-based PPI analysis among the 20 regions of visual thalamus and cortical visual areas (**Figure 7**). The 645 ms TR run showed more significant PPI effects than the 1,400 ms TR run. And non-deconvolution method showed more significant PPI effects than the deconvolution method. A prominent number of connectivity changes are between the bilateral FG1 regions and other lower level visual areas ranging from OC1, OC2, to OC4. We performed a conjunction analysis of PPI results across the four scenarios, and identified five connections with reduced connectivity in checkerboard than in fixation. The regions and connections are highlighted in **Figure 8**.

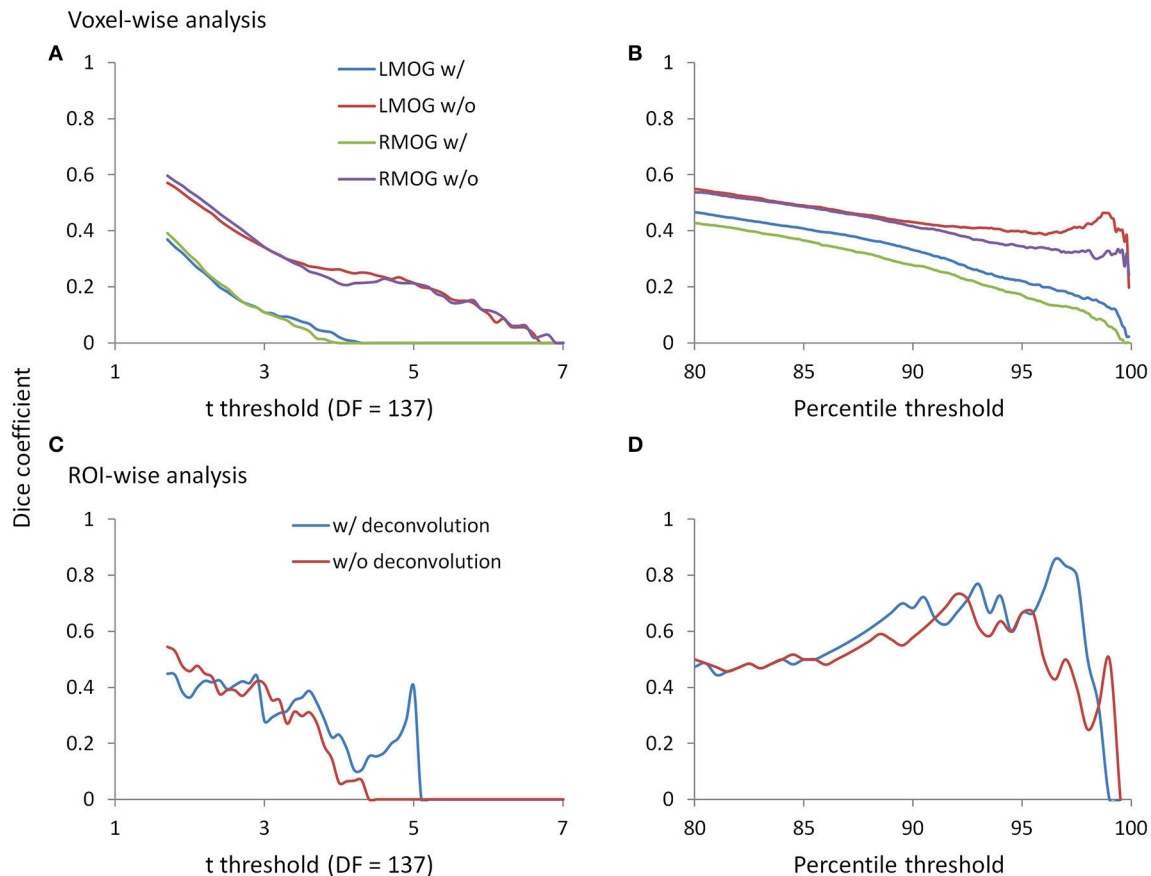
## Reproducibility of PPI Effects

Since we observed similarities of spatial clusters and connectivity between the two TR runs, we next examined reproducibility of PPI effects by calculating Dice coefficients of thresholded statistical maps or PPI matrices between the two TR runs (**Figure 9**). For the voxel-wise analysis of both LMOG and RMOG seeds, when varying  $t$  threshold, the non-deconvolution method showed higher level overlap compared with the deconvolution method (**Figure 9A**). When thresholding statistical maps with matched number of surviving voxels, a similar pattern could still be observed that the non-deconvolution method produced larger overlaps than the deconvolution method (**Figure 9B**). For the ROI-wise analysis, however, Dice coefficients were at similar level between two PPI methods at most  $t$  and percentile thresholds. But at very high  $t$  threshold or percentile thresholds, the deconvolution method seemed to produce larger overlaps (higher Dice coefficients; **Figures 9C,D**).

## Reliability of PPI Effects

We calculated ICC between the two TR runs to reflect reliability of PPI effects. The voxel-wise maps of ICC showed that there were typically low reliability in both methods and ROIs, even in the regions that showed consistent negative PPI effects (**Figure S1**). We then plotted the histograms of ICCs in voxels from the whole brain (gray lines) and within regions that showed significant PPI effects (red lines; **Figures 10A–D**). It turns out that the distributions of ICCs within significant regions are only slightly different from the distributions of correlations in the whole brain, with means around 0.07. The distributions of ICCs were not different between deconvolution and non-deconvolution methods. Similar distributions of ICCs were also found for the ROI-wise analysis (**Figures 10E,F**, and **Figure S2**). We found five PPI effects that were consistently significant in both TR runs and methods. And the ICCs for the five effects were also small and close to zero.





**FIGURE 9 |** Dice coefficients of thresholded negative PPI effects between the two TR runs as functions of t threshold and percentile threshold for the voxel-wise analysis (upper panels) and ROI-wise analysis (lower panels). **(A)** Dice coefficients of negative PPI effects from the voxel-wise analysis between the two TR runs using the t threshold. **(B)** Dice coefficients of negative PPI effects from the voxel-wise analysis between the two TR runs using the percentile threshold. **(C)** Dice coefficients of negative PPI effects from the ROI-wise analysis between the two TR runs using the t threshold. **(D)** Dice coefficients of negative PPI effects from the ROI-wise analysis between the two TR runs using the percentile threshold. The lowest t used for calculating overlap is 1.7, which approximately corresponds to  $p < 0.05$ . The lowest percentile is 80, which is approximately corresponds to the largest proportions of voxels at  $p < 0.05$ .

## Measurement Error

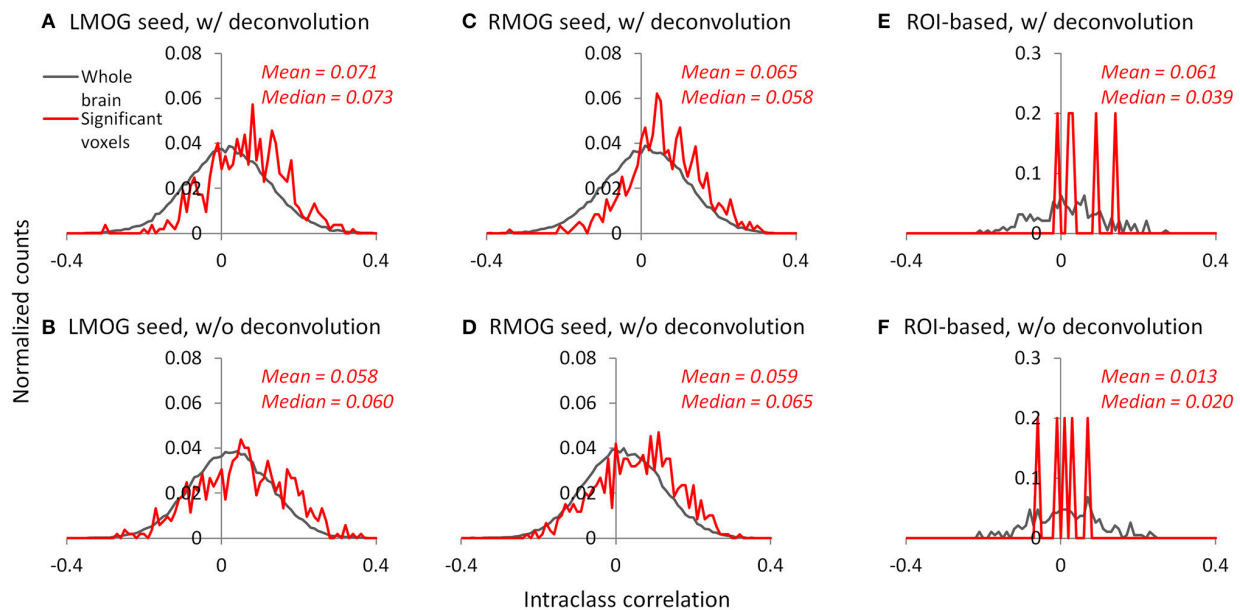
We calculated coefficients of variation (Bland and Altman, 1996) on task activations and PPI effects to reflect measurement error (Figure 11). The variations of activation in the LMOG and RMOG were about 70% of the mean activation, while the variation of activation in the thalamus was about 270% of the mean activation (Figure 11A). In contrast, the variations of PPI effects through the 27 ROIs were about 500% of the mean effects for both the LMOG and RMOG seeds (Figures 11B,C), which indicated much larger variation of PPI effects compared with activations. The deconvolution and non-deconvolution methods had similar level of coefficients of variations. But when directly comparing the two methods, there was a trend that the non-deconvolution method had smaller coefficients of variation than the deconvolution method in most of the ROIs (Figures 11D,E).

## Miscellaneous Analysis

To gain further insight to the cases of deconvolution failure, we calculated correlations of PPI terms between deconvolution and

non-deconvolution methods for the LMOG and RMOG seeds (Figure 12A). In both TR runs, the distributions of correlations centered approximately on 0.7, and there were outliers whose correlations were only 0.2 or 0.3. This is in contrast with the simulation results (Figure 1B, 40 s cycle), where the correlations were around 0.9.

We identified the worst case in Figure 12A (black arrow indicated), and deconvolved and reconvolved it with the HRF using SPM's method (Figure 12B). The raw and reconvolved signals look dramatically different, with the reconvolved signal resembling a smoothed version of the original signal. Smoothness is indeed the case for the SPM version of deconvolution (Gitelman et al., 2003), because it utilizes regularization to suppress high frequency components of cosine basis functions those were used to approximate the neuronal level physiological variable. To directly illustrate this point, we performed fast Fourier transformation on the time series of the RMOG for all the subjects on the raw, deconvolved, and reconvolved time series for the two TR runs (Figure 13). It could be seen that



**FIGURE 10 |** Histograms (normalized) of intraclass correlations of PPI effects between the two TR runs across the whole brain (gray lines) and in statistically significant voxels or effects (red lines). (A–C) Show the results of the left middle occipital (LMOG) seed, right middle occipital (RMOG) seed, and ROI-based analysis using the deconvolution method, respectively. (D–F) Show the results of the LMOG seed, RMOG seed, and ROI-based analysis using the non-deconvolution method, respectively. The significant voxels or effects were determined using intersection of the two runs and two methods each thresholded at  $p < 0.01$ .

after deconvolution, high frequency components have been suppressed in both TR runs. Particularly, there is a black line that shows higher power between frequencies of 0.2 to 0.4 Hz in the raw data plot of 645 ms TR run, which coincides to be the outlier observed in **Figure 12**. The high frequency component was suppressed, so that the reconvolved signal looks smooth.

## DISCUSSION

By analyzing two separate runs of visual checkerboard task from a large sample ( $n = 138$ ), the current study first replicated previously reported negative PPI effects between visual cortex and widespread brain regions, and then showed negative PPI effects among visual areas centered in the bilateral fusiform gyrus. By comparing results from two separate runs, we showed that group averaged effects were largely reproducible; however, the inter-subject reliabilities of the PPI effects were typically low. By comparing the deconvolution and non-deconvolution PPI methods, we demonstrated that the results by the two methods were in general very similar, but the non-deconvolution produced larger statistical effects and spatial extents. The non-deconvolution method may reduce inter-subject variations and increase overlaps of results between the two runs in some circumstances compared with the deconvolution method.

## Functional Connectivity during Checkerboard Stimulation

The voxel-wise analysis of the LMOG and RMOG seeds replicated our previous results which only analyzed a sub-set of 26 subjects (Di et al., 2015, 2017b). In our previous work (Di et al.,

2017b) we could only identify significant PPI effects using the RMOG seed, while the current study demonstrated similar PPI effects from both the LMOG and RMOG seeds. Furthermore, we illustrated that the spatial extent of regions that showed reduced connectivity with the MOG seed could be much larger and extended to other brain regions such as the insula and bilateral sensorimotor cortex. This further suggests a higher extent of functional segregation between the visual cortex and other brain systems during such a simple visual stimulation task compared with the fixation. The current study also extended previous study by analyzing task modulated connectivity effects among cytoarchitecturally defined visual areas. Reduced functional connectivity was observed among many visual areas, with the bilateral FG1 as hub regions. FG1 is the most posterior portion of the fusiform gyrus, which just laid anterior to the occipital cortex (Caspers et al., 2013). It is thought a transition zone between lower retinotopic visual areas and higher category specific brain areas, and integrates information from different retinotopic visual areas to higher category specific brain areas (Caspers et al., 2014). Therefore, it is reasonable to see that the FG1 showed reduced functional connectivity with many lower visual areas in the checkerboard condition, because the simple stimuli cannot form a meaningful percept of a specific category.

The thalamus is a critical subcortical structure in the brain, which not only relay sensory information to the cortex, but also thought to mediate corticocortical communications (Guillery and Sherman, 2002; Saalmann and Kastner, 2011). The PPI analysis of the thalamus, however, did not show consistent effects in different TR runs or different methods. It may be because that the visual thalamus is small in size compared with cortical visual

areas, and the signals in the thalamus are not reliable enough. The current results do suggest some reduced connectivity between the visual thalamus to the primary visual cortex, and increased connectivity between the visual thalamus to the anterior portion of the thalamus, basal ganglia, and insula. However, the results are weak and unreliable, especially considering that the current analysis had included 138 subjects.

## Reproducibility and Reliability of PPI Effects

To our knowledge, the current study is the first one to evaluate reproducibility and reliability on PPI effects. The current analysis did not only reproduce the results reported previously (Di et al., 2017b), but also examined the reproducibility between two runs. Although the two runs were scanned using different parameters, most importantly the temporal and spatial resolutions, the patterns of PPI effects turned out to be quite similar between the two runs. The run with 645 ms TR seemed to generate larger spatial extent in the voxel-wise analysis and more statistically significant results in the ROI-wise analysis. This is consistent with our prediction, because there are more time points in the 645 ms TR run than in the 1,400 ms TR run, which could yield higher statistical power. We do notice that in some scenarios, i.e., voxel-wise analysis with deconvolution, the PPI results in 645 ms TR run had smaller effect size and spatial extent, which might be due to failure of deconvolution.

On the other hand, the results indicated that inter-subject reliabilities are typically low (around 0.07) no matter which PPI method was used. The low reliability should be compared with those of simple task activations, which showed reasonably high reliability regardless of the scan length. The reliability of PPI effects in the current analysis are also much lower than previous reported test–retest reliabilities on task activations (Raemaekers et al., 2007; Plichta et al., 2012) and resting-state functional connectivity (Zuo et al., 2010b; Guo et al., 2012). Of course the short scan lengths could be one factor that explains the low reliability of PPI effects. But it should be also emphasized that the reliability of higher order interaction effects (i.e., the PPI) should be much lower than the main effects of task activations and task-free functional connectivity. A scan length that is sufficient for obtaining reliable task activations may not be necessarily enough to yield reliable task modulated connectivity estimates. This factor should be taken into account when designing studies on task based connectivity.

## Deconvolution and PPI

The PPI results using both the deconvolution and non-deconvolution methods are in general very similar. This is consistent with the simulation showing that the PPI term calculated from the convolution then multiplication method is very similar to the hypothetical PPI term with a known neural activity in a block-designed task. When comparing the differences of PPI results with these two methods, the non-deconvolution method seems to be able to generate larger statistical effects and greater spatial extents or number of significant effects. The non-deconvolution method also increased the Dice coefficients of thresholded PPI maps between the two TR

runs. However, the Dice coefficients of thresholded PPI matrices between the two TR runs are quite similar between the two PPI methods, and the deconvolution method may be even benefiting at higher thresholds. These results highlighted the uncertainty of deconvolution method in PPI analysis.

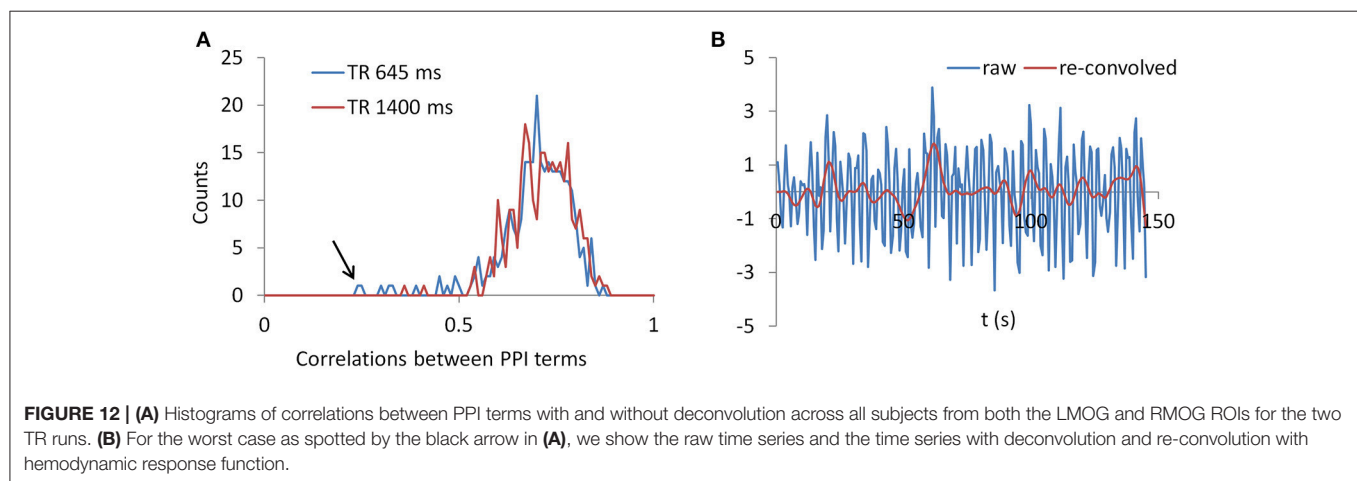
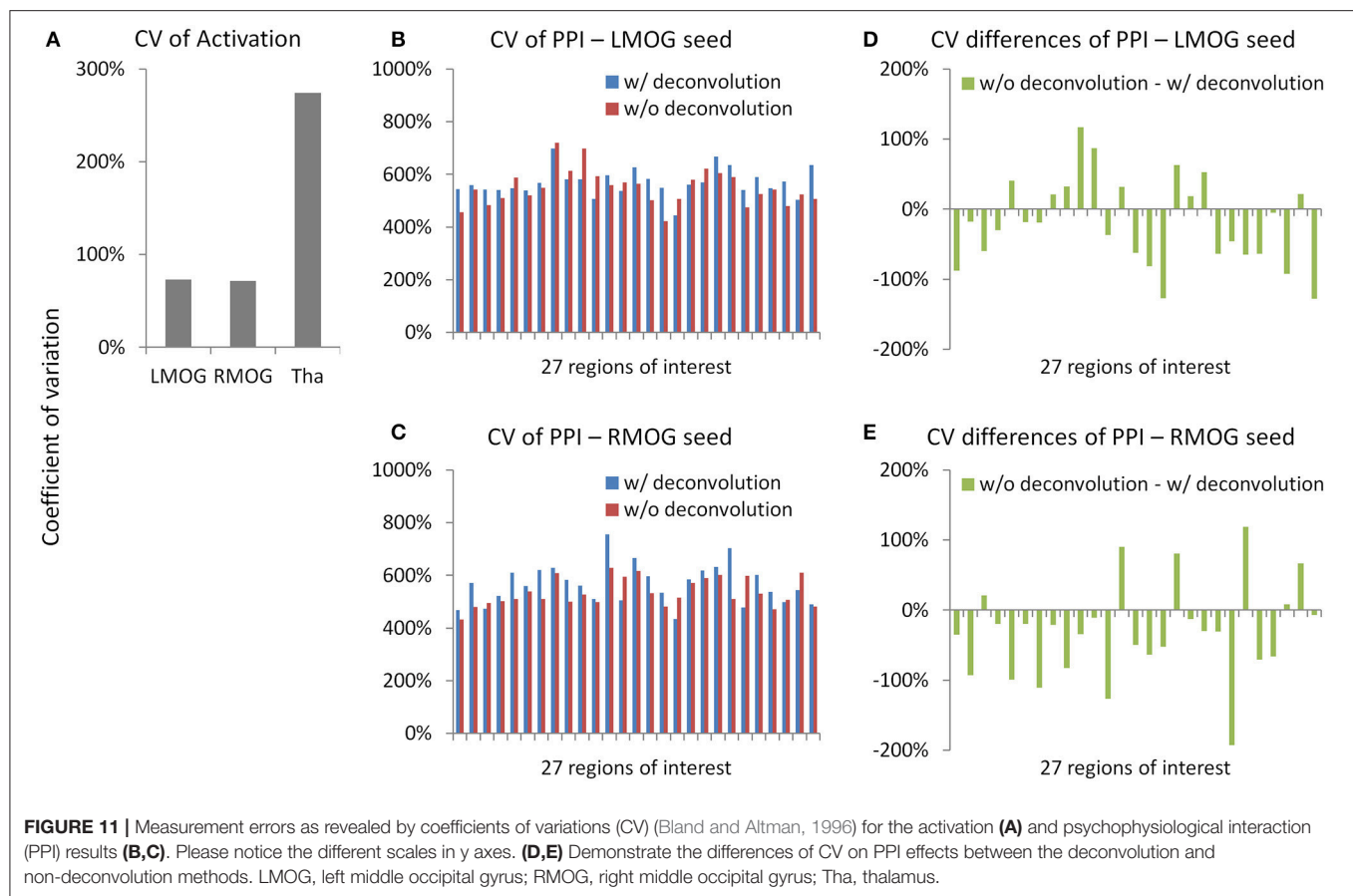
We have shown that the correlations of PPI terms between deconvolution and non-deconvolution methods may have outliers whose correlations were only 0.2 or 0.3 (Figure 12), which is in contrast with the simulation results (Figure 1B). The lower correlations of PPI terms from empirical data compared with the simulations imply that there might be some uncountable variations introduced during the deconvolution/convolution of real fMRI data. Indeed, deconvolution is rather a practical problem to recover underlying signals from some recorded measures, than a simple mathematical problem as depicted in Equation (2). In the practical context, measurement noises need to be taken into account in the deconvolution model. For fMRI, the goal of deconvolution is to recover neuronal activities from observed BOLD signals, where there are plenty of noises during MRI recording. The deconvolution should be expressed as follows with an additional error term:

$$x_{\text{Physio}} = z_{\text{Physio}} * \text{hrf} + \varepsilon \quad (5)$$

In this circumstance, some noises would be removed during deconvolution so that a signal deconvolved and convolved back with a HRF will no longer be the same as the original signal. The noise characteristics and regularization methods for recovering  $z_{\text{Physio}}$  become critical to the success of deconvolution.

As have been shown in Figure 13, SPM's deconvolution method explicitly suppresses high frequency components with the intention that the hemodynamic response is slow therefore high frequency components may represent noises. But this may overly smooth the data and remove useful information in higher frequency bands, thus making PPI results with the deconvolution method less sensitive than those with the direct PPI method. This problem may be more severe for short TR data, because there are more high frequency components in the data. On the other hand, high frequency signals in BOLD have been increasingly recognized as functionally meaningful (Chen and Glover, 2015; Gohel and Biswal, 2015; Lewis et al., 2016), and high frequency components may be critical for connectivity dynamics. Given that multiband imaging technique has made fMRI sampling rate much faster, proper treatment of high frequency signals may be critical in deconvolution of fMRI signals and connectivity analysis in general.

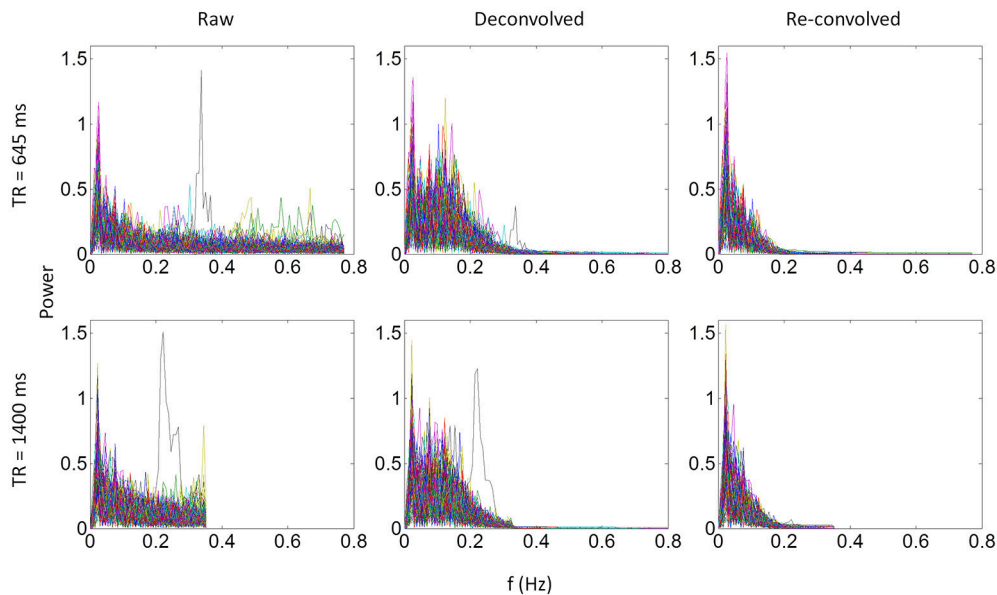
Given the facts that the two PPI methods can generate similar results for the current block-designed task and the non-deconvolution method may increase statistical power, we lean toward a conclusion that the non-deconvolution PPI method may be a better choice for a block-designed task. This is in line with the recommendation by FSL (O'Reilly et al., 2012). Of course, deconvolution is still necessary for an event-related task design, because the PPI terms calculated from the convolution then multiplication method are dramatically different from those calculated from the multiplication then convolution method (Figure 1). It's also worth mentioning that



it has been suggested that the beta series method (Rissman et al., 2004) might be an alternative method for event-related designed data (Cisler et al., 2014). Lastly, there are indeed many variety of deconvolution methods (Makni et al., 2008; Havlicek et al., 2011; Wu et al., 2013), and some of the methods may be more suitable for fMRI signals and PPI analysis. Systematic comparisons between these different methods are needed in the future.

The current analyses are mostly based on empirical fMRI data. One limitation of empirical analysis is that there is no known ground truth to compare with. Simulation may be an alternative way to approach the question. However, development of biological realistic models for task modulated connectivity is still challenging, so that the deconvolution problem is difficult to study using simulations at the current stage. In addition, the similarities and differences between PPI results of the





**FIGURE 13 |** Power spectrums of time series from the right middle occipital gyrus seed for each of the 138 subjects for the 645 ms run (**upper panels**) and 1,400 ms run (**lower panels**). Each line in a plot represents one subject. Left, middle, and right panels show the power spectrum of the raw, deconvolved, and re-convolved time series, respectively.

deconvolution and non-deconvolution methods depend on the variability of hemodynamic response in real fMRI data, which cannot be simply derived from simulations. Therefore, we believe that the current empirical analysis is suitable for the question of deconvolution.

## Practical Implications on PPI Analysis

The current study analyzed data from a simple task design with one task condition and one baseline condition. In real fMRI experiments, however, there are usually more than two conditions. To deal with multiple conditions, it was recommended that each task condition is modeled separately with respect to all other conditions (McLaren et al., 2012). In such “generalized PPI” framework, each experimental condition is modeled as the same way as the checkerboard condition in the current study. It is reasonable to conclude that the similarities of PPI results with and without deconvolution could be generalized to experiments with more than two conditions.

Task related functional connectivity as measured by PPI analysis is typically much smaller, in terms of effect size, reproducibility, and reliability, than simple task activations, and has much larger measurement error. To ensure enough statistical power and reliability, a larger sample size than typical activation studies and enough scan length for each subject are necessary. The design for an fMRI task needs to consider scan length as a critical factor, if the goal of the study is to examine task related connectivity. To date, it is still largely unknown how long a scan is needed for reliability capture task related connectivity. We can only get some insights from resting-state connectivity research, where large scale test-retest datasets are available (Biswal et al., 2010; Zuo et al., 2014). In resting-state literature, it has been suggested that at least 5 min of scan is

needed for reliability estimate functional connectivity (Van Dijk et al., 2010; Birn et al., 2013). Then at least 5 min of scan length for a single task condition is needed for task based fMRI. If the PPI effects are going to be compared between two experimental conditions, which is usually the case for a well-designed cognitive neuroimaging study, the required scan length would be much longer. Of course, direct examinations of the effect of scan length on task related connectivity estimates are still needed in future research.

The PPI method takes advantages of the dynamic aspect of the BOLD signals. Therefore, it's preferable to adopt faster sampling rate to capture temporal dynamics, which may in turn lead to sacrifice of other aspects of the signals, e.g., spatial resolution. The current results support the idea that shorter TR may be beneficial for PPI analysis. Of course, faster sampling rate could be accomplished by new developments of MRI techniques such as, multi-band acquisition (Feinberg and Yacoub, 2012). However, the current results also suggested some pitfalls of using short TR data. The currently used HRF models and deconvolution method may be not quite suitable for fast TR data, so that the PPI method with deconvolution may fail in some cases in short TR data. More work is still needed to validate and optimize models on high speed fMRI data. Of course, high spatial resolution has its own advantage on mapping small brain structures such as the thalamus. So that the considerations of temporal and spatial resolutions may also need to take into account the spatial scales of the regions that are studied.

## CONCLUSION

We demonstrated that the deconvolution and non-deconvolution PPI methods generated similar results on a

simple block-designed task. The deconvolution method may be beneficial in terms of statistical power and reproducibility. Taken together, deconvolution may be not necessary for PPI analysis for block-designed fMRI data. When using a large sample, group mean PPI effects are reproducible; however, inter-subject reliabilities of the PPI effects are quite limited. Systematic evaluations on scan length and reliability may be necessary before studying inter-subject differences or group differences of PPI effects.

## ETHICS STATEMENT

This study involves re-analysis of open-access fMRI dataset. We did not use any personal identifiable information in the current analysis.

## REFERENCES

- Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., and Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable? *Neuroimage* 11, 66–84. doi: 10.1006/nimg.1999.0516
- Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., et al. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83, 550–558. doi: 10.1016/j.neuroimage.2013.05.099
- Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107
- Bland, J. M., and Altman, D. G. (1996). Statistics notes: measurement error proportional to the mean. *BMJ* 313:106. doi: 10.1136/bmj.313.7049.106
- Caspers, J., Zilles, K., Amunts, K., Laird, A. R., Fox, P. T., and Eickhoff, S. B. (2014). Functional characterization and differential coactivation patterns of two cytoarchitectonic visual areas on the human posterior fusiform gyrus. *Hum. Brain Mapp.* 35, 2754–2767. doi: 10.1002/hbm.22364
- Caspers, J., Zilles, K., Eickhoff, S. B., Schleicher, A., Mohlberg, H., and Amunts, K. (2013). Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Struct. Funct.* 218, 511–526. doi: 10.1007/s00429-012-0411-8
- Chen, J. E., and Glover, G. H. (2015). BOLD fractional contribution to resting-state functional connectivity above 0.1Hz. *Neuroimage* 107, 207–218. doi: 10.1016/j.neuroimage.2014.12.012
- Cisler, J. M., Bush, K., and Steele, J. S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. *Neuroimage* 84, 1042–1052. doi: 10.1016/j.neuroimage.2013.09.018
- Di, X., and Biswal, B. B. (2015). Characterizations of resting-state modulatory interactions in the human brain. *J. Neurophysiol.* 114, 2785–2796. doi: 10.1152/jn.00893.2014
- Di, X., Fu, Z., Chan, S. C., Hung, Y. S., Biswal, B. B., and Zhang, Z. (2015). Task-related functional connectivity dynamics in a block-designed visual experiment. *Front. Hum. Neurosci.* 9:543. doi: 10.3389/fnhum.2015.00543
- Di, X., Huang, J., and Biswal, B. B. (2017a). Task modulated brain connectivity of the amygdala: a meta-analysis of psychophysiological interactions. *Brain Struct. Funct.* 222, 619–634. doi: 10.1007/s00429-016-1239-4
- Di, X., Reynolds, R. C., and Biswal, B. B. (2017b). Imperfect (de)convolution may introduce spurious psychophysiological interactions and how to avoid it. *Hum. Brain Mapp.* 38, 1723–1740. doi: 10.1002/hbm.23413
- Dubois, J., and Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends Cogn. Sci.* 20, 425–443. doi: 10.1016/j.tics.2016.03.014
- Feinberg, D. A., and Yacoub, E. (2012). The rapid development of high speed, resolution and precision in fMRI. *Neuroimage* 62, 720–725. doi: 10.1016/j.neuroimage.2012.01.049

## AUTHOR CONTRIBUTIONS

XD and BB conceived the research idea. XD performed data analysis. XD and BB wrote the manuscript.

## ACKNOWLEDGMENTS

This research was supported by National Institute of Health grants R01 AG032088 and R01 DA038895.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2017.00573/full#supplementary-material>

- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6, 218–229. doi: 10.1006/nimg.1997.0291
- Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.* 35, 346–355. doi: 10.1002/mrm.1910350312
- Gitelman, D. R., Penny, W. D., Ashburner, J., and Friston, K. J. (2003). Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* 19, 200–207. doi: 10.1016/S1053-8119(03)00058-2
- Gohel, S. R., and Biswal, B. B. (2015). Functional integration between brain regions at rest occurs in multiple-frequency bands. *Brain Connect.* 5, 23–34. doi: 10.1089/brain.2013.0210
- Guillery, R. W., and Sherman, S. M. (2002). Thalamic relay functions and their role in corticocortical communication. *Neuron* 33, 163–175. doi: 10.1016/S0896-6273(01)00582-7
- Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., et al. (2012). One-year test–retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. doi: 10.1016/j.neuroimage.2012.03.027
- Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651. doi: 10.1016/j.neuroimage.2003.11.029
- Havlicek, M., Friston, K. J., Jan, J., Brazdil, M., and Calhoun, V. D. (2011). Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. *Neuroimage* 56, 2109–2128. doi: 10.1016/j.neuroimage.2011.03.005
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Kujovic, M., Zilles, K., Malikovic, A., Schleicher, A., Mohlberg, H., Rottschy, C., et al. (2013). Cytoarchitectonic mapping of the human dorsal extrastriate cortex. *Brain Struct. Funct.* 218, 157–172. doi: 10.1007/s00429-012-0390-9
- Lewis, L. D., Setsompop, K., Rosen, B. R., and Polimeni, J. R. (2016). Fast fMRI can detect oscillatory neural activity in humans. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6679–E6685. doi: 10.1073/pnas.1608117113
- Makni, S., Beckmann, C., Smith, S., and Woolrich, M. (2008). Bayesian deconvolution fMRI data using bilinear dynamical systems. *Neuroimage* 42, 1381–1396. doi: 10.1016/j.neuroimage.2008.05.052
- Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S. B., Wils, M., et al. (2006). Cytoarchitectonic analysis of the human extrastriate cortex in the region of V5/MT+: a probabilistic, stereotaxic map of area hOc5. *Cereb. Cortex* 17, 562–574. doi: 10.1093/cercor/bhj181

- McLaren, D. G., Ries, M. L., Xu, G., and Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* 61, 1277–1286. doi: 10.1016/j.neuroimage.2012.03.068
- Nooner, K. B., Colcombe, S. J., Tobe, R. H., Mennes, M., Benedict, M. M., Moreno, A. L., et al. (2012). The NKI-rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front. Neurosci.* 6:152. doi: 10.3389/fnins.2012.00152
- O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., and Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci.* 7:nss055. doi: 10.1093/scan/nss055
- Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., et al. (2012). Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. *Neuroimage* 60, 1746–1758. doi: 10.1016/j.neuroimage.2012.01.129
- Raemaekers, M., Vink, M., Zandbelt, B., van Wezel, R. J. A., Kahn, R. S., and Ramsey, N. F. (2007). Test–retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage* 36, 532–542. doi: 10.1016/j.neuroimage.2007.03.061
- Rissman, J., Gazzaley, A., and D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage* 23, 752–763. doi: 10.1016/j.neuroimage.2004.06.035
- Roebroeck, A., Formisano, E., and Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58, 296–302. doi: 10.1016/j.neuroimage.2009.09.036
- Rombouts, S. A., Barkhof, F., Hoogenraad, F. G., Sprenger, M., and Scheltens, P. (1998). Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. *Magn. Reson. Imaging* 16, 105–113. doi: 10.1016/S0730-725X(97)00253-1
- Rottschy, C., Eickhoff, S. B., Schleicher, A., Mohlberg, H., Kujovic, M., Zilles, K., et al. (2007). Ventral visual cortex in humans: cytoarchitectonic mapping of two extrastriate areas. *Hum. Brain Mapp.* 28, 1045–1059. doi: 10.1002/hbm.20348
- Saalmann, Y. B., and Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron* 71, 209–223. doi: 10.1016/j.neuron.2011.06.027
- Taylor, P. A., Gohel, S., Di, X., Walter, M., and Biswal, B. B. (2012). Functional covariance networks: obtaining resting-state networks from intersubject variability. *Brain Connect.* 2, 203–217. doi: 10.1089/brain.2012.0095
- Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009
- Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* 4, 274–290. doi: 10.1111/j.1745-6924.2009.01125.x
- Wu, G.-R., Liao, W., Stramaglia, S., Ding, J.-R., Chen, H., and Marinazzo, D. (2013). A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Med. Image Anal.* 17, 365–374. doi: 10.1016/j.media.2013.01.003
- Yuan, R., Di, X., Taylor, P. A., Gohel, S., Tsai, Y.-H., and Biswal, B. B. (2016). Functional topography of the thalamocortical system in human. *Brain Struct. Funct.* 221, 1971–1984. doi: 10.1007/s00429-015-1018-7
- Zuo, X.-N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* 1:140049. doi: 10.1038/sdata.2014.49
- Zuo, X.-N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., et al. (2010a). The oscillating brain: complex and reliable. *Neuroimage* 49, 1432–1445. doi: 10.1016/j.neuroimage.2009.09.037
- Zuo, X.-N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010b). Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach. *Neuroimage* 49, 2163–2177. doi: 10.1016/j.neuroimage.2009.10.080

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor is currently editing a Research Topic with one of the authors, BB, and confirms the absence of any other collaboration.

Copyright © 2017 Di and Biswal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Brain Network Dynamics Adhere to a Power Law

Dardo G. Tomasi<sup>1\*</sup>, Ehsan Shokri-Kojori<sup>1</sup> and Nora D. Volkow<sup>1,2</sup>

<sup>1</sup> National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD, USA, <sup>2</sup> National Institute on Drug Abuse, Bethesda, MD, USA

## OPEN ACCESS

### Edited by:

Xi-Nian Zuo,  
Chinese Academy of Sciences (CAS),  
China

### Reviewed by:

Xin Di,  
New Jersey Institute of Technology,  
USA  
Maarten Mennes,  
Radboud University Nijmegen,  
Netherlands

### \*Correspondence:

Dardo G. Tomasi  
tomasidg@mail.nih.gov

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 13 November 2016

**Accepted:** 31 January 2017

**Published:** 14 February 2017

### Citation:

Tomasi DG, Shokri-Kojori E and  
Volkow ND (2017) Brain Network  
Dynamics Adhere to a Power Law.  
Front. Neurosci. 11:72.  
doi: 10.3389/fnins.2017.00072

The temporal dynamics of complex networks such as the Internet are characterized by a power scaling between the temporal mean and dispersion of signals at each network node. Here we tested the hypothesis that the temporal dynamics of the brain networks are characterized by a similar power law. This realization could be useful to assess the effects of randomness and external modulators on the brain network dynamics. Simulated data using a well-established random diffusion model allowed us to predict that the temporal dispersion of the amplitude of low frequency fluctuations (ALFF) and that of the local functional connectivity density (lFCD) scale with their temporal means. We tested this hypothesis in open-access resting-state functional magnetic resonance imaging datasets from 66 healthy subjects. A robust power law emerged from the temporal dynamics of ALFF and lFCD metrics, which was insensitive to the methods used for the computation of the metrics. The scaling exponents (ALFF:  $0.8 \pm 0.1$ ; lFCD:  $1.1 \pm 0.1$ ; mean  $\pm$  SD) decreased with age and varied significantly across brain regions; multimodal cortical areas exhibited lower scaling exponents, consistent with a stronger influence of external inputs, than limbic and subcortical regions, which exhibited higher scaling exponents, consistent with a stronger influence of internal randomness. Findings are consistent with the notion that external inputs govern neuronal communication in the brain and that their relative influence differs between brain regions. Further studies will assess the potential of this metric as biomarker to characterize neuropathology.

**Keywords:** FCDM, ALFF, lFCD, functional connectivity (FC), graph theory analysis, brain networks, Taylor's law, numerical simulations

## INTRODUCTION

During resting-state functional magnetic resonance imaging (rsfMRI) (Biswal et al., 1995) the human brain sequentially engages in a series of diverse free-streaming subject-driven mental states supported by different brain networks (Mason et al., 2007; Doucet et al., 2012; Shirer et al., 2012; Liu and Duyn, 2013; Yang et al., 2014). These complex and time-varying functional operations require a dynamic brain network topology to support the context-dependent coordination of neuronal populations (Allen et al., 2014; Zalesky et al., 2014) and its characterization and measurement could facilitate development of clinical biomarkers in neurology and psychiatry (Hutchison et al., 2013). Thus, the temporal dynamics of the human brain connectome (Chang and Glover, 2010; Sakoğlu et al., 2010) provides a new metric of brain function to assess healthy and disease conditions (Calhoun et al., 2014). However, our lack of understanding of the principles governing network dynamics may preclude the interpretation of the observed dynamics, which increases



the within-subjects variability of the functional connectivity metrics (Tomasi et al., 2016a,b). A better understanding of how the collective behavior of neuronal communities contributes to the observable dynamics is crucial for the interpretation of the dynamics of functional connectivity.

Previous studies have shown that temporal mean  $\langle S_i \rangle$ , and dispersion,  $\sigma_i$ , of the activity at a given node are related through a power law across network nodes (Argollo de Menezes and Barabasi, 2004)

$$\sigma_i = a \langle S_i \rangle^b, \quad (1)$$

where the scaling exponent,  $b$ , is a property of the network. Based on theoretical grounds and independent from the topology of the network,  $b$  equals either  $\frac{1}{2}$  or 1, which reflect a competition between the system's internal collective dynamics and temporal changes in the external environment (Argollo de Menezes and Barabasi, 2004). Specifically, in the absence of external modulation,  $b = \frac{1}{2}$ , but when external driving forces become dominant,  $b = 1$ . For instance, whereas the network of internet routers is characterized by  $b = \frac{1}{2}$ , the network of highways and the World Wide Web are characterized by  $b = 1$ . However, empirical evidence from ecology, where (1) describes the spatiotemporal variability of natural populations, supports the existence of intermediate  $b$ -values (Taylor, 1961) suggesting that meaningful temporal dynamic require  $\frac{1}{2} < b < 1$ .

Inasmuch as brain networks have scale-free (Barabasi and Albert, 1999; Eguíluz et al., 2005) and small-world (Watts and Strogatz, 1998) properties exhibited by complex networks we hypothesized that the mean and  $\sigma$  of FC properties such as ALFF, the amplitude of the low frequency fluctuations (Yang et al., 2007) or *l*FC<sub>D</sub>, the local degree of connectivity (Tomasi and Volkow, 2010) would reveal the characteristic power scaling properties exhibited by other complex networks. Specifically, we hypothesized that the mean and  $\sigma$  would be related by the power law (1) and that different brain networks would exhibit different scaling exponents reflecting differential balance between internal randomness (random firing) and external inputs (non-random firing). We selected functional connectivity (ALFF and *l*FC<sub>D</sub>) metrics rather than raw signals because the mean and dispersion values of the BOLD-fMRI signals are not expected to be in agreement with Equation (1).

## METHODS

To interpret the observed power scaling law (1), we study a simple dynamical model based on random diffusion. Using this model and functional connectivity information extracted from rfMRI datasets, we assessed the validity of Equation (1) in the context of brain functional connectivity. However, since direct application of Equation (1) to the mean and dispersion values of the raw fMRI time series is meaningless (the MRI signal mainly reflects tissue properties such as water density and T1 and T2 relaxation rates, which do not change as a function of time; the BOLD signal is zero-mean by definition), we simulated the temporal dynamics of ALFF and *l*FC<sub>D</sub>.

## Model

Similar to previous studies (Argollo de Menezes and Barabasi, 2004), to model the signal  $S(t)$  we simulated the random diffusion of  $W$  walkers (messages) on a network of  $N$  nodes described by its adjacency matrix,  $A_{ij}$ . Each walker was placed at a randomly chosen network node from which it departed randomly along one of the edges of that node in the next time step. This diffusion process was independently repeated 1,200 times and we recorded the number of incoming visits by various walkers at each network node to compute the time-varying signal at each node,  $S_i(t)$ . Temporal fluctuations in  $W$  were used to simulate externally induced modulations in  $S_i(t)$ , which for random networks and scale-free networks results in  $b = 1$  exponent in (1) (Argollo de Menezes and Barabasi, 2004). Thus we varied the number of walkers as a function of time as:  $W(t) = W + \xi(t)$ , where  $\xi(t)$  was a uniformly distributed random variable in the interval  $[-\Delta W, \Delta W]$ , with  $\Delta W = k \cdot 10^3$  and  $k = 0, 1, 2, \dots, 9$ , and  $W = 10^4$ .

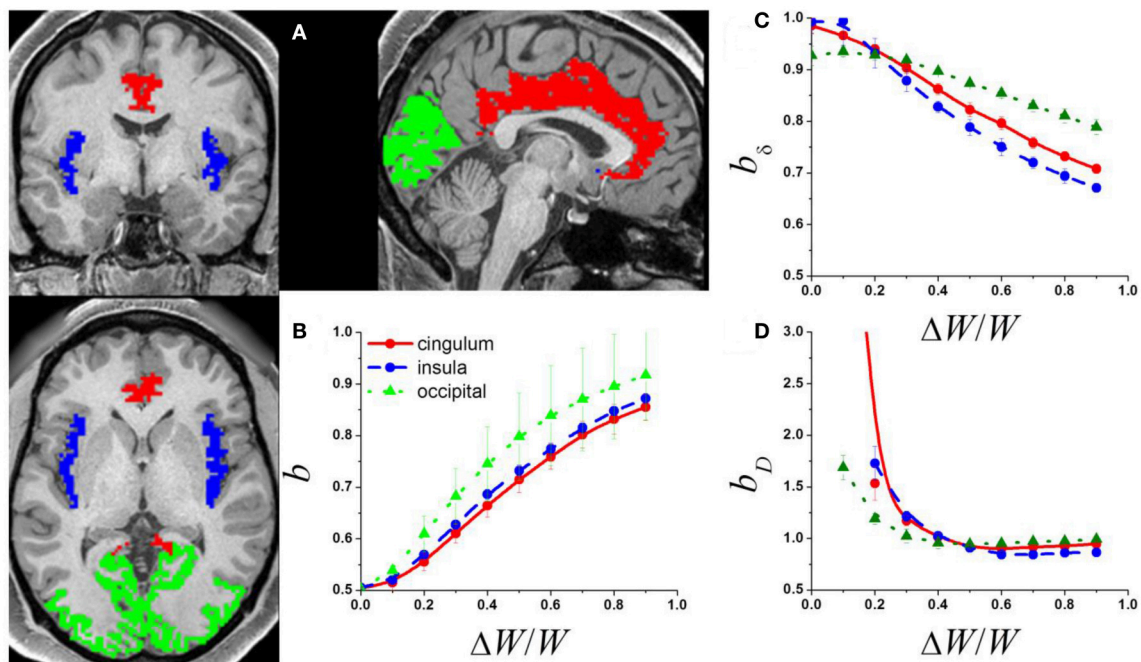
## Simulations

The FreeSurfer gray matter parcellations (wmparc.2.nii) for 7 randomly selected MRI datasets were used to determine imaging voxels in the occipital, cingulate and insular networks (Figure 1A). The occipital network comprised bilateral cuneus, lateral occipital, lingual, and pericalcarine cortices (number of nodes/voxels,  $N = 8,200 \pm 600$ ). The cingulate network comprised bilateral rostral anterior, caudal anterior, isthmus, and posterior cingulate ( $N = 3,100 \pm 400$ ). The insular network comprised the bilateral insula ( $N = 2,300 \pm 100$ ). The Pearson correlation was used to compute correlation matrices reflecting the functional connectivity between voxels within each network for each subject. A correlation threshold  $R = 0.2$  ( $p < 0.05$ ) was used to compute the corresponding binary adjacency matrices. We implemented the diffusion model described above (Argollo de Menezes and Barabasi, 2004). We assumed the signal is proportional to the rate of incoming messages at each node as a function of time, which was simulated using 1,200 steps.

To simulate the dynamics in the amplitude of the signal fluctuations,  $\delta_i$ , at each node we segmented the  $S_i(t)$  data (1,200 time points) into 23 epochs (window length: 100 time points; window shift: 50 time points) using a popular rectangular sliding window approach (Chang and Glover, 2010). The temporal standard deviation of  $S_i(t)$  during each epoch was used to estimate  $\delta_i$ . Degree,  $D_i$ , the number of links connected to a network node (Rubinov and Sporns, 2010), was computed for each of the 23 epochs of the synthetic  $S_i(t)$  data using a correlation threshold,  $R > 0.5$  ( $p < 10^{-7}$ ). The linear model  $\log(\sigma_X) = \log(a) + b \log(X)$  with 2 freely adjustable parameters:  $\log(a)$  and  $b$ , was used to fit the power law (1) to the temporal mean and dispersion values of the dynamic  $\delta$  and  $D$  metrics ( $X$ ).

## Datasets

To test the predictions of the random diffusion model we analyzed rfMRI datasets drawn from the Human Connectome Project (HCP; <http://www.humanconnectome.org/>). No experimental activity with any involvement of human subjects



**FIGURE 1 | (A)** Exemplary single-subject structural data showing occipital (green) cingulate (red) and insular (blue) gray matter networks used to compute the adjacency matrix from the corresponding rfMRI datasets. The adjacency matrices of these networks and a random diffusion model were used to produce simulated signal fluctuations,  $S_i$ , with variable relative external modulation ( $\Delta W/W$ ) at each network node. The scaling exponent,  $b$ , was obtained by linear fitting the temporal mean and dispersion values of ( $S_i$ ) in a log-log plot. **(B)** Average  $b$  across network nodes and 7 subjects as a function of  $\Delta W/W$  for the 3 different networks. Scaling exponent as a function of  $\Delta W/W$  for the 3 different networks for: **(C)** the amplitude of the signal fluctuations,  $b_\delta$ ; and **(D)** the degree of the functional connectivity,  $b_D$  (see Methods).

took place at the author's institutions. The 66 participants (age:  $30 \pm 3$  years; 32 females; Subject IDs: 100408, 103515, 103818, 105115, 105216, 106319, 110411, 118730, 118932, 119833, 120212, 122317, 123117, 125525, 127933, 128632, 129028, 130013, 131924, 133625, 133827, 133928, 134324, 136833, 137128, 138231, 138534, 140824, 142828, 143325, 144226, 149337, 149539, 150423, 151526, 153429, 156637, 158540, 159239, 159340, 160123, 161731, 162329, 163129, 165840, 167743, 172332, 178950, 182739, 191437, 192439, 192540, 194140, 197550, 199150, 199251, 200614, 201111, 210617, 217429, 249947, 250427, 255639, 304020, 307127, 329440) of the WU-Minn HCP Q1 data release included in this study provided written informed consent according to procedures approved by the IRB at Washington University in St. Louis.

Resting-state (eyes open) functional images were acquired using a gradient-echo-planar (EPI) sequence with multiband factor 8, TR 720 ms, TE 33.1 ms, flip angle  $52^\circ$ ,  $104 \times 90$  matrix size, 72 slices, 2 mm isotropic voxels, and 1200 timepoints (Smith et al., 2013; Uğurbil et al., 2013). Scans were repeated twice using different phase encoding directions (LR and RL) on each of two imaging sessions (REST1 and REST2) collected on different days. The “minimal preprocessing” datasets, which include gradient distortion correction, rigid-body realignment, field-map processing, spatial normalization to the stereotactic space of the Montreal Neurological Institute (MNI), high pass filtering (1/2,000 Hz frequency cutoff) (Glasser et al., 2013), independent

component analysis-based denoising (Salimi-Khorshidi et al., 2014), and brain masking were used in this study.

## Preprocessing

Framewise displacements, FD, computed for every time point from head translations and rotations using a radius of  $r = 50$  mm (Power et al., 2012) did not differ between MRI sessions or phase encoding directions across subjects ( $p > 0.2$ , paired  $t$ -test;  $\langle \text{FD} \rangle = 0.176 \pm 0.05$  mm). Scrubbing was not implemented to preserve the frequency spectra used for the computation of ALFF. Multilinear regression of head translations and rotations were used to minimize motion related fluctuations in the MRI signals (Tomasi and Volkow, 2010). Standard 0.01–0.08 Hz band-pass filtering was used to minimize physiologic noise of high frequency components.

## Dynamic ALFF and fCD

The average of the power spectrum's square root in the 0.01–0.08 Hz low frequency bandwidth was used to compute the ALFF (Yang et al., 2007). Functional connectivity density mapping was used to compute the fCD (Tomasi and Volkow, 2010) at three different thresholds  $R > 0.3$ , 0.4 and 0.5. A sliding window approach (Chang and Glover, 2010) with two different window lengths (72s and 144s) and two different window shapes (rectangular and Hamming) was used to compute dynamic ALFF and fCD maps with 2-mm isotropic

resolution at two different temporal resolutions (36s and 72s). The window shift was set as half of the window length.

## Region-of-Interest (ROI) Analysis

To test the power law (1) we contrasted scaling factors for the simulated signal fluctuations ( $\delta$ ) and degree ( $D$ ) against those for ALFF and  $I_{FCD}$ . Since  $I_{FCD}$  has high sensitivity and specificity for gray matter (Tomasi et al., 2016c), the FC metrics were averaged within the anatomical gray matter regions-of-interest for each individual to minimize confounds arising from the variability of the folding patterns of cortical gray matter. Specifically, the FreeSurfer gray matter parcellations (wmparc.2.nii) were used as ROIs to compute the averages of the temporal mean and dispersion values of  $ALFF(t)$  and  $I_{FCD}(t)$  within 34 cortical and 9 subcortical gray matter regions in each brain hemisphere. A probabilistic atlas for each of the gray matter parcellations was developed by averaging each of the gray matter parcellations across subjects independently, and used to display

ROI results (i.e.,  $b_{ALFF}$  or  $b_{I_{FCD}}$ ) in the MNI stereotactic space (Figures 2A, 3A).

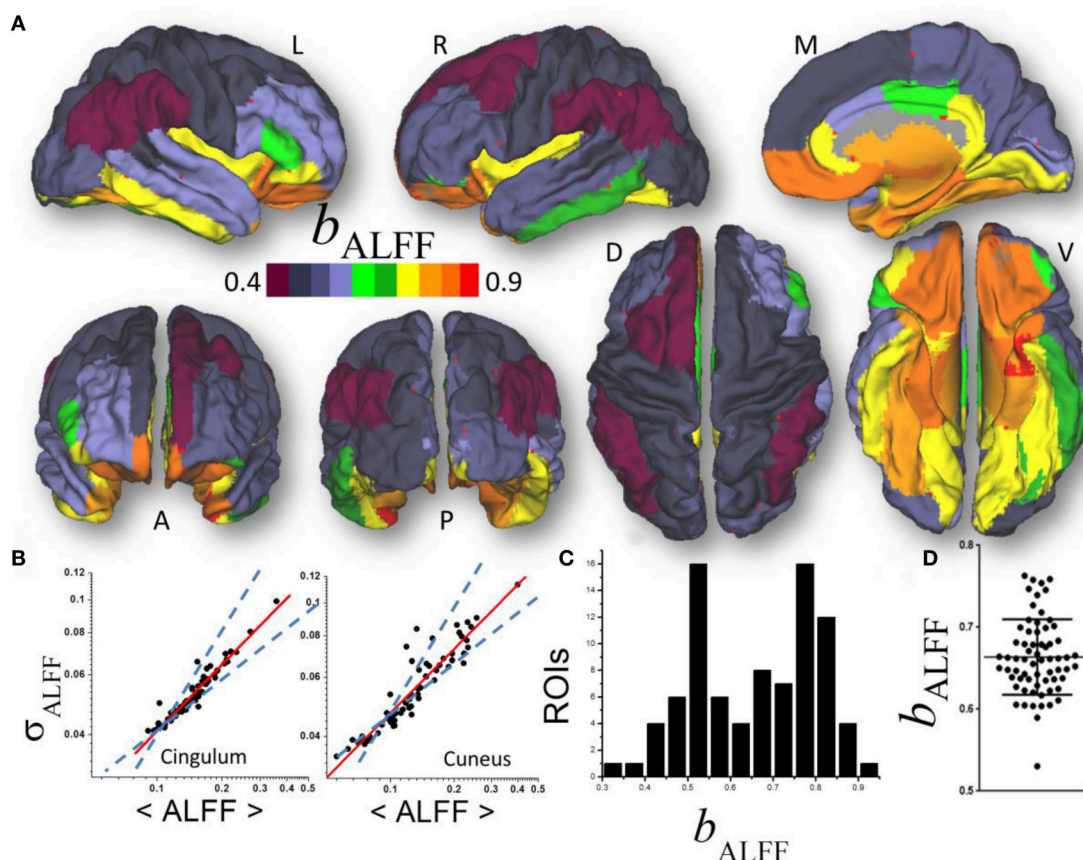
## Statistical Methods

The linear model  $\log(\sigma) = \log(a) + b \log(X)$  with 2 free adjustable parameters:  $\log(a)$  and  $b$ , was used to fit the power law (1) to the temporal mean and dispersion values of the dynamic ALFF and  $I_{FCD}$  metrics ( $X$ ). Paired  $t$ -test was used to assess within subjects differences in  $b_{ALFF}$  and  $b_{I_{FCD}}$  as a function of session, phase encoding direction, correlation threshold, and window length and shape. Two samples  $t$ -test and Pearson correlation were used to assess gender and aging effects on  $b_{ALFF}$  and  $b_{I_{FCD}}$ .

## RESULTS

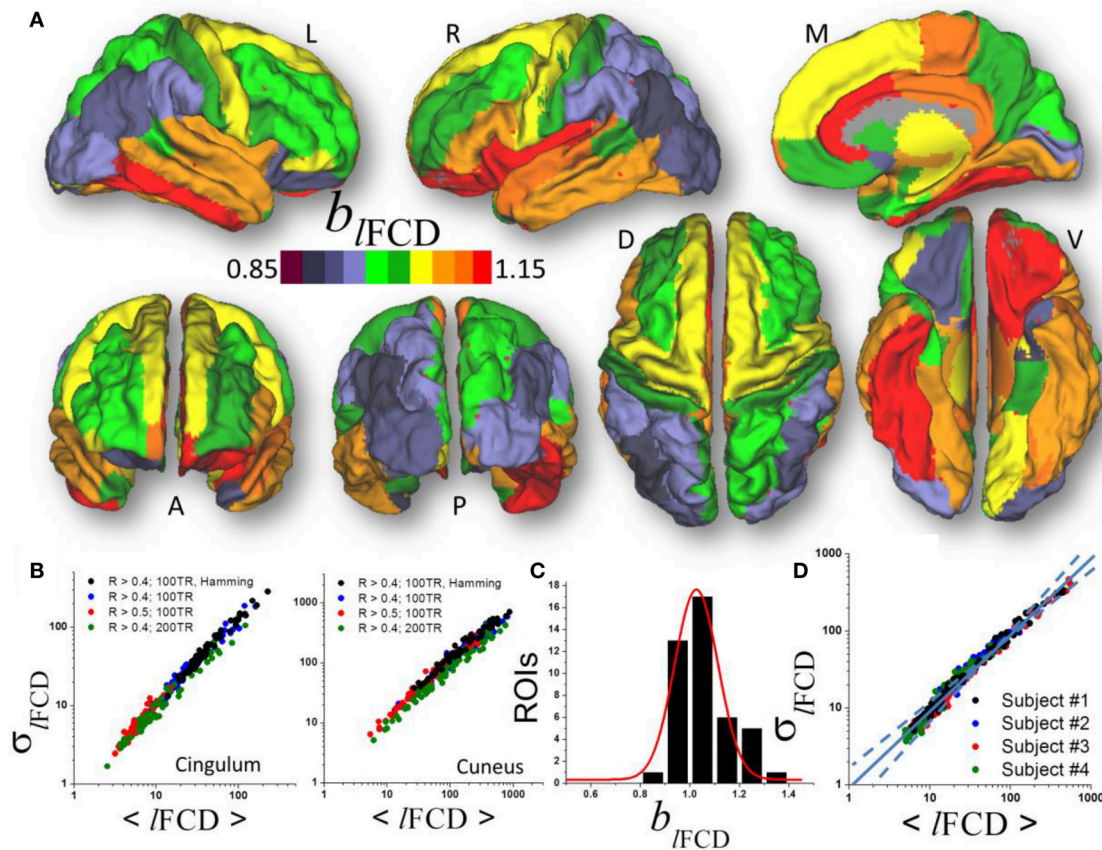
### Simulations

The power law (1) fitted well ( $R^2 > 0.8$ ) the temporal mean and standard deviation values of  $S_i(t)$  across nodes. The  $b$  exponents



**FIGURE 2 | (A)** Average scaling exponent ( $b_{ALFF}$ ) for the temporal dynamics of the amplitude of low frequency fluctuations (ALFF) computed across nodes independently for each of the individual anatomical ROIs, superimposed on left (L), right (R), dorsal (D), medial (M), ventral (V) anterior (A), and posterior (P) views of the cerebral surface of the PALS\_B12 template. **(B)** Scatter plots showing the good agreement across 66 subjects (dots) between the power law (1) (red line) and the dynamics of ALFF which is characterized by its temporal mean,  $\langle ALFF \rangle$ , and dispersion,  $\sigma_{ALFF}$ . Dashed lines are the upper ( $b_\delta = 1$ ; pure randomness) and lower ( $b_\delta = 1/2$ ; pure external modulation) limits for the scaling exponent. **(C)** Frequency count histogram reflecting the probability distribution of  $b_{ALFF}$  across cortical and subcortical gray matter ROIs. **(D)** Scatter plot demonstrating the variability of  $b_{ALFF}$  across 66 young adults.





**FIGURE 3 | (A)** Average scaling exponent ( $b_{I/FCD}$ ) for the temporal dynamics of the local functional connectivity density ( $I/FCD$ ) computed across nodes independently for each of the individual anatomical ROIs superimposed on left (L), right (R), dorsal (D), medial (M), ventral (V) anterior (A), and posterior (P) views of the cerebral surface of the PALS\_B12 template. **(B)** Scatter plots across 66 subjects showing the robustness of the power law (1) that reflects the dynamics of  $I/FCD$  to changes in correlation thresholds, sliding window lengths and shapes. **(C)** Frequency count histogram reflecting the probability distribution of  $b_{I/FCD}$  across cortical and subcortical gray matter ROIs. **(D)** Scatter plot demonstrating the moderate differences power law (computed across 86 ROIs) in four typical subjects.

increased monotonically with  $\Delta W$ , which is consistent with the notion that internal randomness (diffusion) and external modulation ( $\Delta W$ ) proportionally alter  $S_i(t)$  in the network (Argollo de Menezes and Barabasi, 2004). Thus,  $\Delta W$  contributed to the temporal variability of the signal at each network node, gradually increasing  $b$  from  $\frac{1}{2}$  to 1 in all three brain networks (Figure 1B) as it occurs in other complex networks. Thus, if its magnitude is significant ( $\Delta W \sim \frac{1}{2}(W)$ ), the external modulation can dominate the dynamics of  $S_i(t)$ .

The mean and dispersion values of  $\delta_i$  computed across epochs were also in good agreement with the power law (1). Our simulations suggest that,  $b_\delta \sim 1$ , when internal randomness dominates over the external modulations (Figure 1C). However,  $b_\delta$  decreased with the amplitude of the external modulation and was constrained in the interval  $[0.5, 1]$ . Similarly, the mean and dispersion values of  $D_i$  computed across epochs were in good agreement with the power law (1). Our simulations suggest that  $b_D \sim 1$  when the external modulation dominates over the internal randomness, but  $b_D$  increases significantly above 1 when the relative weight of the external modulation decreases (Figure 1D). The power law failed

to fit the data when internal randomness dominated over the external modulation ( $\Delta W/W > \frac{1}{2}$ ) suggesting lack of association between the mean and dispersion values of  $D$  in this regime.

## Amplitude of Fluctuations

A linear fit of whole-brain average and dispersion values of ALFF on a log-log plot computed across nodes demonstrated good agreement between Equation (1) and the dynamic amplitude of the signal fluctuations in each of the individual ROI ( $b_{ALFF} = 0.66 \pm 0.16$ , mean  $\pm$  standard deviation;  $28 < t$ -score  $< 294$ ;  $P < 1E-37$ ; Figure 2A). Consistent findings emerged from average and dispersion values within anatomical regions, independently for each of the 86 gray matter ROIs ( $R^2 > 0.8$ ; Figures 2B,C). The average scaling exponent was not different for subcortical (cerebellum, thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbens and ventral diencephalon;  $b_{ALFF} = 0.78 \pm 0.04$ , mean  $\pm$  standard error) than for cortical ( $b_{ALFF} = 0.81 \pm 0.07$ ) regions ( $p > 0.4$ ), independent of session, phase encoding direction (LR vs. RL), sliding window length (72s vs. 144s) and shape (rectangular vs. Hamming).



There were no significant differences in  $b_{\text{ALFF}}$  across subcortical regions. However,  $b_{\text{ALFF}}$  varied significantly across cortical regions. Specifically, the scaling exponent was higher for limbic (cingulum, orbitofrontal, parahippocampal and entorhinal) and visual (lingual, fusiform and pericalcarine) areas, the temporal and insular cortices and pars orbitalis ( $b_{\text{ALFF}} = 0.89 \pm 0.02$ ) than for occipital (cuneus, lateral occipital), parietal (inferior, superior, precuneus, postcentral), language (opercularis, triangularis, supramarginal) and prefrontal (paracentral, precentral, rostral, middle and superior frontal) areas ( $b_{\text{ALFF}} = 0.72 \pm 0.03$ ;  $p < 10^{-9}$ ; **Figure 2B**). The scaling exponent had normal distribution (center  $b_{\text{ALFF}} = 0.80$ ; width = 0.16) across the 86 gray matter ROIs ( $R^2 = 0.999$ , Gaussian fit; **Figure 2C**).

Significant between-subjects variability in the scaling exponent emerged from the data when we fitted Equation (1) to the mean and dispersion values of ALFF across the 43 ROIs, independently for each individual ( $b_{\text{ALFF}} = 0.66 \pm 0.05$ ; **Figure 2D**) and with similar robustness ( $R^2 > 0.96$ ). The scaling exponent slightly decreased with age (slope =  $-0.03/\text{decade}$ ;  $R = -0.234$ ;  $p = 0.03$ , one-tailed). However, there were no significant gender differences ( $p > 0.77$ ; two-tailed two-sample  $t$ -test) in  $b_{\text{ALFF}}$ .

## Local Degree

Similar to ALFF, a linear fit of whole-brain average and dispersion values of  $I_{\text{FCD}}$  on a log-log plot computed across nodes demonstrated good agreement between Equation (1) and the local degree of brain functional connectivity in each of the individual ROI ( $b_{I_{\text{FCD}}} = 1.05 \pm 0.17$ , mean  $\pm$  standard deviation;  $43 < t\text{-score} < 179$ ;  $P < 3E-49$ ; **Figure 3A**). Average and dispersion values within anatomical regions showed consistent findings with those from the whole-brain analysis ( $R^2 > 0.8$ ; **Figure 3B**). The average scaling exponent was higher for subcortical ( $b_{I_{\text{FCD}}} = 1.23 \pm 0.09$ , mean  $\pm$  standard error) than for cortical ( $b_{I_{\text{FCD}}} = 1.06 \pm 0.10$ ) regions ( $p < 10^{-3}$ , two-tailed two-sample  $t$ -test), independent of the correlation threshold used in the computation of the  $I_{\text{FCD}}$  ( $R > 0.3$ , 0.4, or 0.5), session, phase encoding direction, window length (72s vs. 144s) and shape.

For  $I_{\text{FCD}}$ , the scaling exponent was higher for limbic (cingulum, orbitofrontal, parahippocampal, and entorhinal), language (opercularis, orbitalis, triangularis), temporal (inferior, middle superior), and frontal (paracentral, superior and pole), insula and fusiform gyrus ( $b_{I_{\text{FCD}}} = 1.13 \pm 0.07$ ) than for occipital (cuneus, lateral occipital, lingual and pericalcarine), parietal (inferior, superior, precuneus, supramarginal, paracentral, postcentral), prefrontal (precentral, rostral, middle, and superior) and temporal (entorhinal temporal pole, transverse) areas ( $b_{I_{\text{FCD}}} = 0.98 \pm 0.06$ ;  $p < 10^{-6}$ ; **Figure 3B**). Across the 86 gray matter ROIs the scaling exponent had a right-skewed distribution with peak at  $b_{I_{\text{FCD}}} = 1.03$  and width = 0.17 ( $R^2 = 0.95$ , Gaussian fit; **Figure 3C**). Fitting mean and dispersion values of  $I_{\text{FCD}}$  across the 43 gray matter ROIs, independently for each subject, revealed modest between-subjects variability in the scaling exponent ( $b_{I_{\text{FCD}}} = 1.09 \pm 0.06$ ; **Figure 3D**), and  $b_{I_{\text{FCD}}}$  did not show significant age or gender differences ( $p > 0.23$ ).

In visual areas (pericalcarine, lateral occipital, and cuneus) the standardized scaling factors  $b_z$  were lower than average and were significantly lower for  $I_{\text{FCD}}$  than for ALFF ( $p < 0.0005$ ,  $t$ -test; **Figure 4**, left). In prefrontal regions (middle Frontal, superior frontal, precentral, paracentral, pars opercularis, and caudal anterior cingulate), the lower than average  $b_z$  was lower for ALFF than for  $I_{\text{FCD}}$  ( $p < 0.001$ ). In frontal and temporal poles, entorhinal and lingual cortex showed the higher than average  $b_z$  was higher for ALFF than for  $I_{\text{FCD}}$  ( $p < 0.0001$ ; **Figure 4** right). In anterior (rostral) and posterior (isthmus) cingulate, fusiform gyrus and subcortical regions (hippocampus, thalamus and cerebellum) the higher than average  $b_z$  was higher for  $I_{\text{FCD}}$  than for ALFF ( $p < 0.0006$ ).

## Effect of Bandpass Filtering

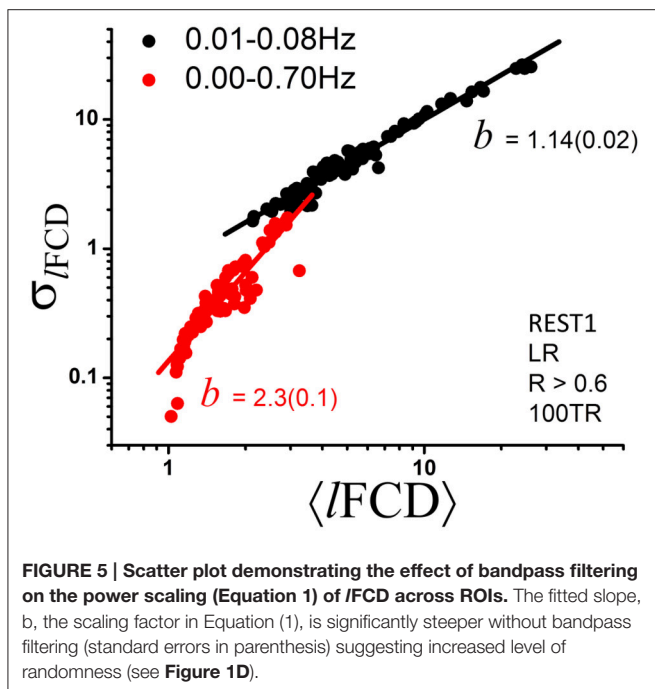
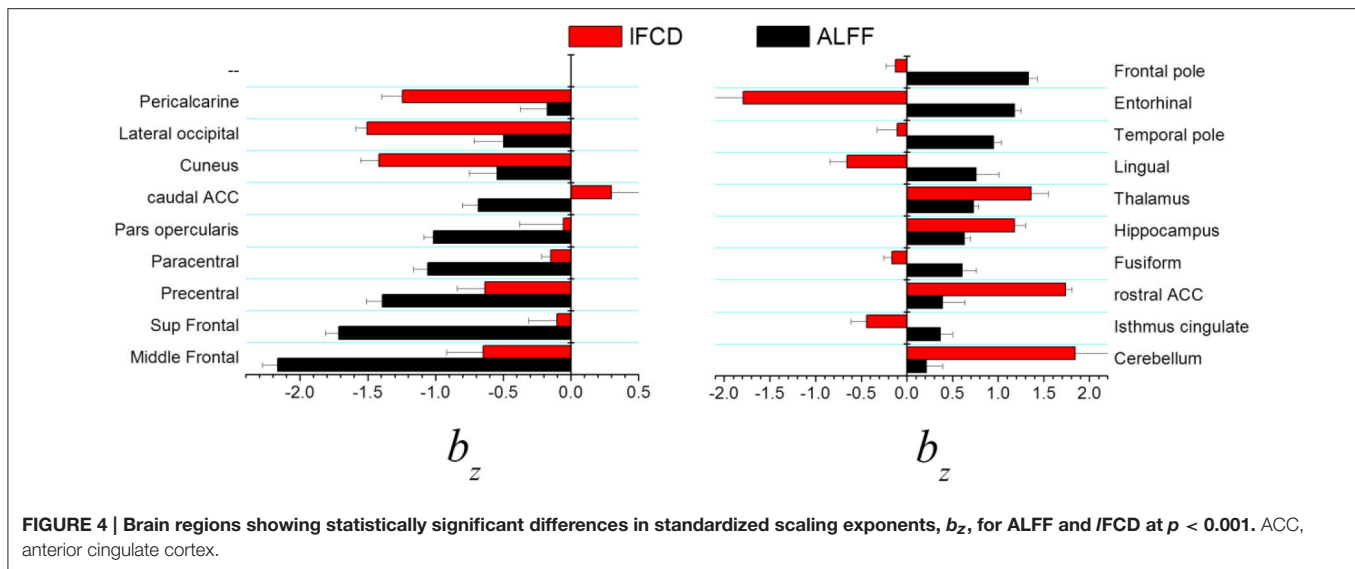
Given that frequency information may be of interest and that the ICA-FIX denoising procedure can remove a significant fraction of the physiological noise of respiratory origin (Salimi-Khorshidi et al., 2014), we also computed dynamic  $I_{\text{FCD}}$  measures without 0.01–0.08 Hz bandpass filtering to assess the effect of higher frequencies on the power scaling law (Equation 1). Without bandpass filtering the scaling exponent  $b$  of the dynamic  $I_{\text{FCD}}$  metrics was significantly larger than with bandpass filtering ( $p < 0.0001$ ; **Figure 5**), and the agreement between the data and Equation (1) was significantly reduced [ $R^2 = 0.82$  (without) and 0.96 (with bandpass filtering)].

## DISCUSSION

Here we show for the first time that the mean and the dispersion values of dynamic FC metrics such as ALFF or  $I_{\text{FCD}}$  are linked by a power law (1). This characteristic of complex networks such as rivers and highways networks, the Internet and the World Wide Web (Argollo de Menezes and Barabasi, 2004), and many biological systems (Taylor, 1961), reflects the competition between the system's internal collective dynamics and changes in the external environment. This strongly suggests that the dynamics of the FC metrics embeds important functional information, a possibility previously highlighted (Hutchison et al., 2013; Calhoun et al., 2014; Rashid et al., 2014; Hutchison and Morton, 2015), which could help in the development of biomarkers of brain function.

Our simulations were based on a random diffusion model previously proposed by Argollo de Menezes and Barabasi to explain the power scaling between the mean and the dispersion of the signals observed in natural and technological networks (Argollo de Menezes and Barabasi, 2004). Whereas the approach by Argollo de Menezes and Barabasi was based on random and scale-free networks (Barabasi and Albert, 1999; Barabási, 2009), the present approach was based on real FC networks directly extracted from *in vivo* resting fMRI data. The scaling exponents for the brain in the present work are consistent with those obtained previously in random and scale-free networks (Argollo de Menezes and Barabasi, 2004).

Here we extended the random diffusion model in order to simulate the amplitude of spontaneous signal fluctuation and the degree of connectivity. Our simulations suggest that under



pure randomness (i.e., without external driving forces,  $\Delta W = 0$ ) the mean and the dispersion values of the amplitude of signal fluctuations and degree are associated by power laws with scaling exponents  $b_\delta = 1$  and  $b_D > 1$ , respectively. However, under the influence of dynamic external modulations ( $\Delta W/W \sim 1$ ),  $b_\delta < 1$  and  $b_D = 1$  characterize the dynamic behavior of the signal fluctuations and degree. The analysis of variability of resting-state fMRI datasets from the HCP database shows a range of scaling exponents for ALFF ( $0.5 < b_{\text{ALFF}} < 1$ ) and for IFCF ( $1 < b_{\text{IFCF}} < 1$ ), which is consistent with the presence of dynamic external modulations of brain activity ( $0.5 < b_\delta < 1$ ) and the corresponding degree ( $1 < b_D$ ). Overall, our findings are also

consistent with the existence of dynamic modulations of brain activity that may reflect orchestrated dynamic neural processing (Yu et al., 2012; Allen et al., 2014; Gonzalez-Castillo et al., 2015).

This is the first study to document differences in scaling exponents between brain regions. Multimodal association areas (opercularis, triangularis, rostral, middle and superior frontal, precentral and paracentral, inferior and superior parietal and precuneus), somatosensory (supramarginal, postcentral) and visual (cuneus, lateral occipital) unimodal association areas showed low scaling exponent both for ALFF ( $b_{\text{ALFF}} \sim 0.7$ ) and for IFCF ( $b_{\text{IFCF}} \sim 1$ ). These findings suggest that the dynamics of the FC metrics was driven by external inputs ( $\Delta W/W > 1/2$ ) rather than by internal random processes ( $\Delta W/W < 0.5$ ; Figures 1C,D), which is also consistent with the existence of dynamic modulations of resting brain activity (Yu et al., 2012; Allen et al., 2014; Gonzalez-Castillo et al., 2015). The multimodal cortex is highly interconnected with higher-order association areas involved in cognition and motor planning (Goldman-Rakic, 1988). Thus dynamic engagement of functional connectivity hubs in multimodal and unimodal association cortices may explain the low scaling exponent in these regions. On the other hand, limbic and subcortical regions exhibited relatively higher scaling exponents ( $b_{\text{ALFF}} \sim 0.8$  and  $b_{\text{IFCF}} \sim 1.2$ ) suggesting a stronger influence of internal randomness in the resting dynamics of the FC metrics in these regions.

We identify regional differences in the influence of internal randomness for different FC metrics. The direct comparison of standardized measures suggests a weaker influence of randomness in visual areas for IFCF than for ALFF and in prefrontal areas for ALFF than for IFCF, and a stronger influence of randomness in subcortical and limbic regions for IFCF than for ALFF. ALFF and IFCF reflect different network properties. Whereas ALFF is proportional to the BOLD signal fluctuations that reflect neuronal communication (Logothetis et al., 2001), the synchronous fluctuations of local communities measured

by  $l_{FCD}$  reflects the local degree of connectivity (Tomasi and Volkow, 2010).

The scaling exponent for ALFF, and to a lesser extent for  $l_{FCD}$ , showed significant variability ( $\Delta b_{ALFF} = 12\%$ ;  $\Delta b_{l_{FCD}} = 9\%$ ) across subjects suggesting that the dynamics of the  $b$  has potential as a biomarker for psychiatry and neurology. To illustrate the potential of this metric here we show that even in a relatively small sample (66 subjects) with narrow age range (22–35 years),  $b_{ALFF}$  is sensitive to aging effects, consistent with previous studies in large samples ( $\sim 1000$  subjects) with wide age range (17–82 years) that documented age-related decreases in FC (Biswal et al., 2010; Tomasi and Volkow, 2012).

The scaling exponent for  $l_{FCD}$  increased significantly above 1 when frequencies other than those in the 0.01–0.08 Hz band were not removed from the data. At the same time, the agreement with a power scaling was reduced when Equation (1) was fitted to the data without bandpass filtering. This likely reflects the introduction of additional randomness and is consistent with increased noise level and lack of additional information at higher frequencies than those in the 0.01–0.08 Hz band.

The brain normally operates under certain level of randomness that is important for multiple operation including perception and decision-making. The relevance of internal neuronal noise has been most extensively studied for visual perception (Brascamp et al., 2006; Kim et al., 2006). Theoretical studies have also shown that randomness may influence behavioral responses when there are multiple routes to action and suggested that noise generated by random firing rates of neurons can be used to predict a decision (Rolls, 2012). Since limbic and subcortical regions support automatic, implicit decision making (Floresco et al., 2008; Mitchell, 2015) the higher scaling exponents in these regions suggests an important role of randomness in implicit decision making processes. The sensitivity to randomness of  $b$  could be useful for studying psychiatric disorders such as autism, which is associated with increased randomness of endogenous brain oscillations (Lai et al., 2010).

## Study Limitations

Note that  $b = \frac{1}{2}$  emerges either from diffusion or from flow models, independently of the number of steps in the diffusion model, and from random networks as well as from scale-free networks. This indicates that  $b = \frac{1}{2}$  is not a particular property of the random diffusion model, but it is shared by several dynamic processes (Argollo de Menezes and Barabasi, 2004). Our computational resources did not allow demanding whole brain network simulations at 2-mm isotropic resolution

( $\sim 10^5$  nodes/voxels). Thus, our simulations suggesting that when internal randomness dominates over the external modulations ( $\Delta W/W \sim 0$ )  $b_8 \sim 1$  and  $b_D > 1$ , but when external modulations dominate over internal randomness ( $\Delta W/W \sim 1$ )  $b_8 \sim 0.5$  and  $b_D \sim 1$  are limited to the 3 exemplary networks in this work. However, it is likely that they apply also to the whole brain. Instrumental noise likely resulted in overestimations of intrinsic randomness in subcortical regions for which the 32 channel RF coil used by the HCP has low sensitivity. Since the theoretical model was developed across network nodes, the interpretation of the power law across ROIs and subjects could be considered controversial. Our empirical evidence, however, suggests that the temporal mean and standard deviation values of dynamic functional connectivity metrics also adhere to a power law computed across ROIs or subjects, which are consistent with the power law computed across nodes (i.e., across nodes of each individual ROI,  $b_{l_{FCD}} = 1.05 \pm 0.17$  mean  $\pm$  standard deviation; across the 86 gray matter, ROIs  $b_{l_{FCD}} = 1.03 \pm 0.17$ ; across 66 subjects,  $b_{l_{FCD}} = 0.98 \pm 0.16$ ). This suggests similar effects of randomness and external modulators on power scaling factors computed across network nodes, ROIs or subjects.

Dynamic  $l_{FCD}$  is restricted to the local functional connectivity cluster. We did not assess the dynamics of global functional connectivity density (gFCD) because at high spatiotemporal resolution gFCD is extremely demanding and beyond our computational resources. However, this is not a strong limitation because previous studies have shown that the  $l_{FCD}$  and gFCD metrics are proportional to one another (Tomasi and Volkow, 2011).

## AUTHOR CONTRIBUTIONS

DT designed the study, carried the analyses and wrote the manuscript. ES developed imaging tools and wrote the manuscript. NV wrote the manuscript.

## ACKNOWLEDGMENTS

Data was provided by the Human Connectome Project (HCP), WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research and by the McDonnell Center for Systems Neuroscience at Washington University. This work was accomplished with support from the National Institute on Alcohol Abuse and Alcoholism (Y1AA-3009).

## REFERENCES

- Allen, E., Damaraju, E., Plis, S., Erhardt, E., Eichele, T., and Calhoun, V. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Argollo de Menezes, M., and Barabasi, A. (2004). Fluctuations in network dynamics. *Phys. Rev. Lett.* 92:028701. doi: 10.1103/PhysRevLett.92.028701
- Barabási, A. (2009). Scale-free networks: a decade and beyond. *Science* 325, 412–413. doi: 10.1126/science.1173299
- Barabasi, A., and Albert, R. (1999). Emergence of scaling in random networks. *Science* 286, 509–512. doi: 10.1126/science.286.5439.509
- Biswal, B., Mennes, M., Zuo, X., Gohel, S., Kelly, C., Smith, S., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107

- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Brascamp, J., Van Ee, R., Noest, A., Jacobs, R., and van den Berg, A. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *J. Vis.* 6, 1244–1256. doi: 10.1167/6.11.8
- Calhoun, V., Miller, R., Pearson, G., and Adali, T. (2014). The chroconnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015
- Chang, C., and Glover, G. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50, 81–98. doi: 10.1016/j.neuroimage.2009.12.011
- Doucet, G., Naveau, M., Petit, L., Zago, L., Crivello, F., Jobard, G., et al. (2012). Patterns of hemodynamic low-frequency oscillations in the brain are modulated by the nature of free thought during rest. *Neuroimage* 59, 3194–3200. doi: 10.1016/j.neuroimage.2011.11.059
- Eguíluz, V., Chialvo, D., Cecchi, G., Baliki, M., and Apkarian, A. (2005). Scale-free brain functional networks. *Phys. Rev. Lett.* 94:018102. doi: 10.1103/PhysRevLett.94.018102
- Floresco, S., St. Onge, J., Ghods-Sharifi, S., and Winstanley, C. (2008). Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. *Cogn. Affect. Behav. Neurosci.* 8, 375–389. doi: 10.3758/CABN.8.4.375
- Glasser, M., Sotiropoulos, S., Wilson, J., Coalson, T., Fischl, B., Andersson, J., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127
- Goldman-Rakic, P. (1988). Topography of cognition: parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci.* 11, 137–156. doi: 10.1146/annurev.ne.11.030188.001033
- Gonzalez-Castillo, J., Hoy, C., Handwerker, D., Robinson, M., Buchanan, L., Saad, Z., et al. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8762–8767. doi: 10.1073/pnas.1501242112
- Hutchison, R., and Morton, J. (2015). Tracking the brain's functional coupling dynamics over development. *J. Neurosci.* 35, 6849–6859. doi: 10.1523/JNEUROSCI.4638-14.2015
- Hutchison, R., Womelsdorf, T., Allen, E., Bandettini, P., Calhoun, V., Corbetta, M., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378. doi: 10.1016/j.neuroimage.2013.05.079
- Kim, Y., Grabowecky, M., and Suzuki, S. (2006). Stochastic resonance in binocular rivalry. *Vis. Res.* 46, 392–406. doi: 10.1016/j.visres.2005.08.009
- Lai, M., Lombardo, M., Chakrabarti, B., Sadek, S., Pasco, G., Wheelwright, S., et al. (2010). A shift to randomness of brain oscillations in people with autism. *Biol. Psychiatry* 68, 1092–1099. doi: 10.1016/j.biopsych.2010.06.027
- Liu, X., and Duyn, J. (2013). Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4392–4397. doi: 10.1073/pnas.1216856110
- Logothetis, N., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. doi: 10.1038/35084005
- Mason, M., Norton, M., Van Horn, J., Wegner, D., Grafton, S., and Macrae, C. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295
- Mitchell, A. (2015). The mediodorsal thalamus as a higher order thalamic relay nucleus important for learning and decision-making. *Neurosci. Biobehav. Rev.* 54, 76–88. doi: 10.1016/j.neubiorev.2015.03.001
- Power, J., Barnes, K., Snyder, A., Schlaggar, B., and Petersen, S. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018
- Rashid, B., Damaraju, E., Pearson, G., and Calhoun, V. (2014). Dynamic connectivity states estimated from resting fMRI Identify differences among Schizophrenia, bipolar disorder, and healthy control subjects. *Front. Hum. Neurosci.* 8:897. doi: 10.3389/fnhum.2014.00897
- Rolls, E. (2012). Willed action, free will, and the stochastic neurodynamics of decision-making. *Front. Integr. Neurosci.* 6:68. doi: 10.3389/fnint.2012.00068
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Sakoglu, Ü., Pearson, G. D., Kiehl, K. A., Wang, Y. M., Michael, A. M., and Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *MAGMA* 23, 351–366. doi: 10.1007/s10334-010-0197-8
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C., Glasser, M., Griffanti, L., and Smith, S. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468. doi: 10.1016/j.neuroimage.2013.11.046
- Shirer, W., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb. Cortex* 22, 158–165. doi: 10.1093/cercor/bhr099
- Smith, S., Beckmann, C., Andersson, J., Auerbach, E., Bijsterbosch, J., Douaud, G., et al. (2013). Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168. doi: 10.1016/j.neuroimage.2013.05.039
- Taylor, L. (1961). Aggregation, variance and the mean. *Nature* 189, 732–735. doi: 10.1038/189732a0
- Tomasi, D., Shokri-Kojori, E., and Volkow, N. (2016a). Temporal changes in local functional connectivity density reflect the temporal variability of the amplitude of low frequency fluctuations in gray matter. *PLoS ONE* 11:e0154407. doi: 10.1371/journal.pone.0154407
- Tomasi, D., Shokri-Kojori, E., and Volkow, N. (2016b). Temporal evolution of brain functional connectivity metrics: could 7 min of rest be enough? *Cereb. Cortex* doi: 10.1093/cercor/bhw1227. [Epub ahead of print].
- Tomasi, D., Shokri-Kojori, E., and Volkow, N. D. (2016c). High-resolution functional connectivity density: hub locations, sensitivity, specificity, reproducibility, and reliability. *Cereb. Cortex* 26, 3249–3259. doi: 10.1093/cercor/bhv171
- Tomasi, D., and Volkow, N. (2010). Functional connectivity density mapping. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9885–9890. doi: 10.1073/pnas.1001414107
- Tomasi, D., and Volkow, N. (2011). Functional connectivity hubs in the human brain. *Neuroimage* 57, 908–917. doi: 10.1016/j.neuroimage.2011.05.024
- Tomasi, D., and Volkow, N. (2012). Aging and functional brain networks. *Mol. Psychiatry* 17, 549–558. doi: 10.1038/mp.2011.81
- Uğurbil, K., Xu, J., Auerbach, E., Moeller, S., Vu, A., Duarte-Carvajalino, J., et al. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. *Neuroimage* 80, 80–104. doi: 10.1016/j.neuroimage.2013.05.012
- Watts, D., and Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442. doi: 10.1038/30918
- Yang, H., Long, X., Yang, Y., Yan, H., Zhu, C., Zhou, X., et al. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *Neuroimage* 36, 144–152. doi: 10.1016/j.neuroimage.2007.01.054
- Yang, Z., Craddock, R. C., Margulies, D. S., Yan, C. G., and Milham, M. P. (2014). Common intrinsic connectivity states among posteromedial cortex subdivisions: insights from analysis of temporal dynamics. *Neuroimage* 93(Pt 1), 124–137. doi: 10.1016/j.neuroimage.2014.02.014
- Yu, Q., Allen, E., Sui, J., Arbabshirani, M., Pearson, G., and Calhoun, V. (2012). Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. *Curr. Top. Med. Chem.* 12, 2415–2425. doi: 10.2174/156802612805289890
- Zalesky, A., Fornito, A., Cocchi, L., Gollo, L., and Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10341–10346. doi: 10.1073/pnas.1400181111

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Tomasi, Shokri-Kojori and Volkow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Inconsistency in Abnormal Brain Activity across Cohorts of ADHD-200 in Children with Attention Deficit Hyperactivity Disorder

Jian-Bao Wang<sup>1,2,3</sup>, Li-Jun Zheng<sup>1,2,3</sup>, Qing-Jiu Cao<sup>4</sup>, Yu-Feng Wang<sup>4</sup>, Li Sun<sup>4</sup>, Yu-Feng Zang<sup>1,2,3\*</sup> and Hang Zhang<sup>5\*</sup>

<sup>1</sup> Center for Cognition and Brain Disorders and the Affiliated Hospital, Hangzhou Normal University, Hangzhou, China, <sup>2</sup> Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China, <sup>3</sup> Institutes of Psychological Sciences, College of Education, Hangzhou Normal University, Hangzhou, China, <sup>4</sup> Institute of Mental Health, The Sixth Hospital, Peking University, Beijing, China, <sup>5</sup> Paul C. Lauterbur Research Centers for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

## OPEN ACCESS

### Edited by:

Bharat B. Biswal,  
University of Medicine and Dentistry of  
New Jersey, United States

### Reviewed by:

Mitul Ashok Mehta,  
King's College London, United  
Kingdom  
Mingrui Xia,  
Beijing Normal University, China

### \*Correspondence:

Yu-Feng Zang  
zangyf@gmail.com  
Hang Zhang  
kevinhangbnu@foxmail.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 11 January 2017

**Accepted:** 19 May 2017

**Published:** 06 June 2017

### Citation:

Wang J-B, Zheng L-J, Cao Q-J,  
Wang Y-F, Sun L, Zang Y-F and  
Zhang H (2017) Inconsistency in  
Abnormal Brain Activity across  
Cohorts of ADHD-200 in Children with  
Attention Deficit Hyperactivity  
Disorder. *Front. Neurosci.* 11:320.  
doi: 10.3389/fnins.2017.00320

Many papers have shown results from the multi-site dataset of resting-state fMRI (rs-fMRI) in attention deficit hyperactivity disorder (ADHD), a data-sharing project named ADHD-200. However, few studies have illustrated that to what extent the pooled findings were consistent across cohorts. The present study analyzed three voxel-wise whole-brain metrics, i.e., amplitude of low-frequency fluctuation (ALFF), regional homogeneity (ReHo), and degree centrality (DC) based on the pooled dataset as well as individual cohort of ADHD-200. In addition to the conventional frequency band of 0.01–0.08 Hz, sub-frequency bands of 0–0.01, 0.01–0.027, 0.027–0.073, 0.073–0.198, and 0.198–0.25 Hz, were assessed. While the pooled dataset showed abnormal activity in some brain regions, e.g., the bilateral sensorimotor cortices, bilateral cerebellum, and the bilateral lingual gyrus, these results were highly inconsistent across cohorts, even across the three cohorts from the same research center. The standardized effect size was rather small. These findings suggested a high heterogeneity of spontaneous brain activity in ADHD. Future studies based on multi-site large-sample dataset should be performed on pooled data and single cohort data, respectively and the effect size must be shown.

**Keywords:** attention deficit hyperactivity disorder, resting state fMRI, multi-site dataset, ADHD-200, voxel-wise whole-brain analysis

## INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders in children (Polanczyk et al., 2015). It is a highly heterogeneous disease, involving multiple deficits and multiple neural pathways (Castellanos et al., 2006; Bush, 2010). The complicated pathophysiology of ADHD has been widely investigated through task and resting-state functional magnetic resonance imaging (fMRI) studies. Task-state fMRI studies commonly employed various task paradigms, e.g., Go/No Go (Schulz et al., 2004; Newman et al., 2015), Eriksen Flanker Task (Vaidya et al., 2005; Vasic et al., 2014). These tasks are complicated, and various paradigms did not exhibit consistent results (Cortese et al., 2012). In contrast, resting-state fMRI (rs-fMRI) is easy to be implemented and provides a consistent approach for clinical investigations. Thus, more and more researchers perform rs-fMRI studies on brain disorders, including ADHD.

ADHD-200, as one of the most widely used multi-site MRI dataset of brain disorders, has attracted considerable attention from the ADHD research community. This dataset released by ADHD-200 consortium contains ten independent cohorts from eight different sites (ADHD-200-Consortium, 2012). These cohorts provide rs-fMRI and anatomical MRI data of both ADHD and typically developing children (TDC), about 776 participants in total. ADHD-200 facilitated the investigation of the neural basis of ADHD, and about 30 studies based on this dataset have been published according to PubMed (e.g., Tomasi and Volkow, 2012; Elton et al., 2014; Sripada et al., 2014; Carmona et al., 2015).

Most studies on ADHD-200 pooled data of cohorts and explored the abnormal brain activity for ADHD. Increasing number of these studies were reported in recent years. For example, Mills et al. (2012) pooled data of Brown University (BU), Peking University (PKU), Kennedy Krieger Institute (KKI), and New York University (NYU) together and observed increased connection between the medial and anterior dorsal thalamus and the basal ganglia in ADHD (Mills et al., 2012). Pooling data of PKU, NYU together, Zhang et al. (2014) found affected brain regions in ADHD mainly located in the orbito-frontal cortex, inferior/superior frontal gyrus, anterior cingulate gyrus, and calcarine cortex (Zhang et al., 2014). Pooling cohorts together facilitated the establishment of a large sample size and tended to provide very positive results. However, to what extent the pooled results are consistent across individual cohorts remains unknown. To the best of our knowledge, only one study on ADHD-200 dataset answered this question (Cai et al., 2015). They found that ADHD group of cohorts NYU, PKU, and OHSU consistently showed decreased network-interaction among the salience network (SN), central executive network (CEN), and default mode network (DMN). Notably, the network analysis could not indicate the exact aberrant brain regions for ADHD, and it remains unclear whether findings of the local brain regions for ADHD are consistent across cohorts or not.

The present study aimed to examine the consistency of abnormal local brain regions across cohorts of ADHD-200. Specifically, we analyzed three voxel-wise whole-brain metrics, i.e., amplitude of low-frequency fluctuation (ALFF) (Zang et al., 2007), regional homogeneity (ReHo) (Zang et al., 2004), and degree centrality (DC) (Buckner et al., 2009). Importantly, the analytic processes of these kinds of methods are very similar across studies, and hence facilitate the coordinate-based meta-analysis (CB-meta) which helps to find regions of consistent activity across fMRI studies (Bartra et al., 2013; Herz et al., 2014; Iwabuchi et al., 2015). Analysis of these metrics is often performed at the frequency band of 0.01–0.08 Hz which has been widely used in rs-fMRI studies. In addition to this conventional band, rs-fMRI signals at some sub-frequency bands can also be modulated by different resting state (e.g., eyes closed and eyes open; Yuan et al., 2014) as well as by disease (e.g., chronic pain; Malinen et al., 2010; Otti et al., 2013). These sub-frequency bands, i.e., Slow-6 (<0.01 Hz; Lv et al., 2013; Zhang et al., 2015a), Slow-5 (0.01–0.027 Hz), Slow-4 (0.027–0.073 Hz; Zuo et al., 2010; Han et al., 2011; Zhang et al., 2013), Slow-3 (0.073–0.198 Hz), and Slow-2 (0.198–0.25 Hz; Wang et al., 2015), were also investigated

in the present study in order to obtain more information through the frequency-dependent characteristic.

## METHODS AND MATERIALS

### Subjects and Data Acquisition

The data we used in this study is publicly available from the ADHD-200 Consortium ([http://fcon\\_1000.projects.nitrc.org/indi/adhd200/](http://fcon_1000.projects.nitrc.org/indi/adhd200/)). The ADHD-200 dataset contains both functional and anatomical MRI data contributed by eight institutions. Each cohort was approved by the research ethics review boards of each institution. Signed informed consent was obtained from all participants or their legal guardian before participation.

We first selected the data cohorts according to the following criteria: (1) Including both ADHD and TDC groups. So the data from the BU, University of Pittsburgh and, Washington University were excluded; (2) Employing the same TR with <2,000 ms across the cohort. According to this criterions, data from NeuroImage (TR = 1,960 ms), KKI (TR = 2,500 ms), and OHSU (TR = 2,500 ms) were excluded. Then, the NYU, PKU1, PKU2, and PKU3 cohorts were included in our research. The PKU2 and PKU3 cohorts only had male subjects, so the female subjects in NYU and PKU1 cohorts were excluded to remove potential confounding effect of gender to the consistency across cohort. Left-handedness subjects were also excluded for each cohort. After case-by-case matching age between ADHD and TDC, 58 subjects from NYU, 30 from PKU1, 56 from PKU2, and 38 from PKU3 were included in the current study. Demographic information was summarized in **Table 1**. Flow-chart of data exclusion was shown in **Figure 1**.

Psychostimulant medications were withheld at least 24 h prior to scanning. The inclusion and exclusion criteria and more detailed demographic characteristics of the participants of the four cohorts can be seen in the [http://fcon\\_1000.projects.nitrc.org/indi/adhd200/](http://fcon_1000.projects.nitrc.org/indi/adhd200/). The rs-fMRI data of the four cohorts were from three scanners, with TR of 2 s for all. PKU1 and PKU2 used the same scanner but scanning parameters were slightly different. The detailed parameters were listed in the Supplementary Table 1.

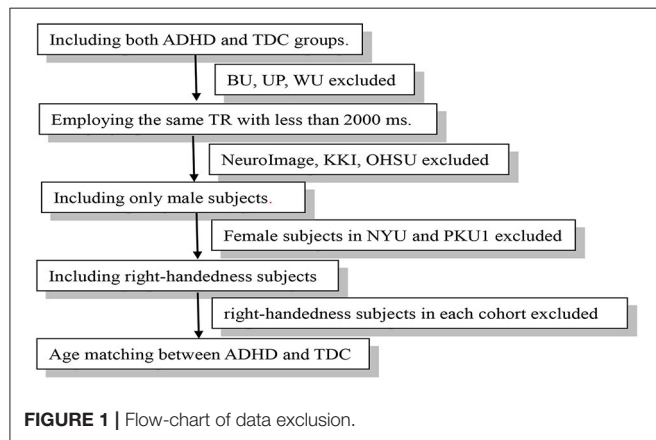
### Data Preprocessing

Functional images of each subject were preprocessed by using Data Processing Assistant for Resting-State fMRI (DPARSF) (Chao-Gan and Yu-Feng, 2010) which is based on Statistical Parametric Mapping (SPM8) (<http://www.fil.ion.ucl.ac.uk/spm>) and Resting-State fMRI Data Analysis Toolkit (Song et al., 2011). Preprocessing was performed as follows: removal of the first ten volumes to avoid signal instability and to get subjects adapted to the scanning noise. Then, the number of time point is 170 at least (NYU), so the first 170 volumes were included for individuals in PKU1, PKU2, and PKU3 considering the comparability across cohorts (Molloy et al., 2014; Carmona et al., 2015). Slice timing correction, image realignment to correct head motion were followed. After individual structural images were segmented after co-registered to functional images, functional images were spatial normalized to Montreal Neurological Institute (MNI) space at 3 mm isotropic voxel resolution applying the unified segmentation

**TABLE 1** | Demographic information of each cohort in the current study.

	NYU		PKU1		PKU2		PKU3	
	ADHD	TDC	ADHD	TDC	ADHD	TDC	ADHD	TDC
N	29	29	15	15	28	28	19	19
Gender (male)	29	29	15	15	28	28	19	19
Age (years)	12.1 ± 2.9	12.2 ± 2.8	11.2 ± 2.3	11.6 ± 1.5	12.7 ± 1.7	11.7 ± 1.8	13.2 ± 1.3	13.3 ± 1.0
IQ	106.1 ± 16.0	115.3 ± 14.3	101.7 ± 12.4	123.0 ± 14.2	111.5 ± 12.7	121.6 ± 12.2	102.7 ± 10.4	111.7 ± 12.7
Subtype (C/I/H)	19/10/0	–	9/6/0	–	16/12/0	–	12/7/0	–

Data are presented as mean ± SD. C, ADHD –Combined; I, ADHD –Inattentive; H, ADHD –Hyperactive/Impulsive.



parameters. The linear trend, head motion parameter measured by Friston-24 model, white matter (WM), and cerebrospinal fluid (CSF) signals were further regressed out as nuisance covariates. Then, three voxel-wise whole-brain analytic methods, i.e., ALFF, ReHo, and DC, were further used to analyze these preprocessed data.

## ALFF Calculation

ALFF is the amplitude of low frequency fluctuations of the blood oxygen level dependent (BOLD) signal of every single voxel (Zuo et al., 2010). ALFF calculation was the same as the procedure in Zang et al. (2007). After preprocessing, the 4D rs-fMRI data of each participant was spatially smoothed with a 6 mm FWHM Gaussian kernel and then, the linear trend was removed from the time course of each voxel. Then, ALFF was calculated for the conventional low frequency band (0.01–0.08 Hz) as well as five sub-bands, i.e., Slow-6 (0–0.01 Hz), Slow-5 (0.01–0.027 Hz), Slow-4 (0.027–0.073 Hz), Slow-3 (0.073–0.198 Hz), and Slow-2 (0.198–0.25 Hz).

## ReHo Calculation

ReHo is a voxel-wise measure of the local synchronization of the time courses of nearest neighboring voxels (usually 27 voxels). It was calculated by using Kendall's coefficient of concordance (KCC) as follows:

$$W = \frac{\sum (R_i)^2 - n(\bar{R})^2}{\frac{1}{12}K^2(n^2 - n)} \quad (1)$$

where  $W$  is the KCC among given voxels, ranged from 0 to 1;  $R_i$  is the sum rank of the  $i$ th time point;  $\bar{R} = ((n+1)K/2)$  is the mean of  $R_i$ 's;  $K$  is the number of time courses within a measured cluster (27 in the current study); and  $n$  is the number of ranks. After the removing of linear trend, the time course of each voxel, band-pass filtering was performed for six sub-bands as in ALFF analysis. ReHo was then calculated for each sub-band. The spatial smoothing (FWHM = 6 mm) was performed after ReHo calculation as did in previous studies (Zang et al., 2004).

## DC Calculation

Degree centrality (DC) represents the node characteristic of large-scale brain intrinsic connectivity networks by capturing the relationship with the entire brain network in the voxel level (Zuo et al., 2012). We used weighted DC since it provides a more precise centrality characterization of functional brain networks than binary version (Cole et al., 2010). Specifically, after preprocessing, the linear trend of the time course of each voxel was removed, and then band-pass filtering was performed for six sub-bands as in ALFF analysis. The Pearson correlation was performed between the time course of each voxel with that of every other voxel in the entire brain (Buckner et al., 2009). The correlation coefficients with  $r > 0.2$  were summed up for each voxel and then a weighted DC was obtained for each voxel. 0.2 was used as threshold to eliminate counting voxels that had low temporal correlation and it has been proved that different threshold selections did not qualitatively change the results (Buckner et al., 2009). As did in ReHo calculation, spatial smoothing may introduce possible artificial local correlations, we performed spatial smoothing (FWHM = 6 mm) after DC calculation as did elsewhere as follows (Zuo et al., 2012):

$$D = \sum a_{ij}$$

$$\text{Where } j = 1 \dots N, i \neq j, a_{ij} = \begin{cases} 0, & a_{ij} < 0.2 \\ a_{ij}, & a_{ij} \geq 0.2 \end{cases} \quad (2)$$

Negative correlation was removed according to previous fMRI studies (Liao et al., 2013; Li et al., 2015). It was not calculated separately because the physiological basis of the negative correlations was ambiguous (Fox et al., 2009; Murphy et al., 2009).

ALFF measures the amplitude of time series fluctuation at each voxel (Zang et al., 2007), ReHo depicts the local synchronization of the time series of neighboring voxels (Zang

et al., 2004), and DC represents the large-scale brain intrinsic connectivity in the voxel level (Buckner et al., 2009). Thus, the three measures of fMRI probe into the brain activity from different aspects.

## Statistical Analysis

ALFF, ReHo, and DC maps of each frequency band were compared between the groups of children with ADHD and TDC. Two-sample *t*-tests were performed on the pooled data and each cohort, respectively. The full scale IQ and mean framewise displacement (FD) were included as nuisance covariates (Jenkinson et al., 2002; Yan et al., 2013), and cohort was further taken as a covariate for the *t*-tests on the pooled data. For each cohort, the statistical analyses were performed in study-specific functional volume masks including only voxels (in MNI152 standard space) present in at least 80% of the participants and then intersect with gray-matter mask to reduce non-cortical noise. The mask of the pooled data is the intersection of cohorts' masks. The results were corrected for multiple comparisons with a combined threshold of single voxel's  $p < 0.05$  and cluster size  $> 139, 144, 136, 129$ , and  $129$  voxels for the cohorts and pooled data, corresponding to corrected  $p < 0.05$  determined by Monte Carlo simulation and the mask of each cohort. The AlphaSim estimation was performed by DPABI V2.3 (<http://rfmri.org/dpabi>; Yan et al., 2016). At the same time, to reduce the possibility of false negative results and, hence, a more lenient threshold ( $p < 0.05$ , cluster size  $> 10$  voxels) was also used for each cohort.

We also performed the analyses of standardized effect size (SES) of each measurement based on Cohen's *d* which is calculated as the equation as follows (Cohen, 2013):

$$\text{Cohen's } d = \frac{\bar{X}_{ADHD} - \bar{X}_{TDC}}{S_{ALL}},$$

$$S_{ALL} = \sqrt{\frac{(n_{ADHD} - 1)S_{ADHD}^2 + (n_{TDC} - 1)S_{TDC}^2}{n_{ADHD} + n_{TDC} - 2}} \quad (3)$$

According to equation of independent two-sample *t*-test as follows:

$$t = \frac{\bar{X}_{ADHD} - \bar{X}_{TDC}}{\sqrt{\frac{(n_{ADHD} - 1)S_{ADHD}^2 + (n_{TDC} - 1)S_{TDC}^2}{n_{ADHD} + n_{TDC} - 2} \left( \frac{1}{n_{ADHD}} + \frac{1}{n_{TDC}} \right)}} \quad (4)$$

The relationship of Cohen's *d*- and *t*-value can be obtained as follows:

$$\text{Cohen's } d = t \sqrt{\frac{n_{ADHD} + n_{TDC}}{n_{ADHD} \cdot n_{TDC}}} \quad (5)$$

According to Equation (5), we transformed *t* maps into SES map for each cohort and pooled data. Then a combined threshold  $SES > 0.30$  and cluster size  $> 129$  voxels was used which corresponded to a combination threshold of  $t > 1.974$  ( $p < 0.05$ ) of the pooled data. The same threshold was applied to the SES maps of each cohort. SES of 0.30 corresponded to  $t = 1.141$ ,

0.822, 1.124, and 0.926 ( $p = 0.26, 0.43, 0.27$ , and  $0.36$ ) for NYU, PKU1, PKU2, and PKU3, respectively.

To view the consistency of results, the thresholded *t*-maps and SES-maps were binarized and overlapped among the four cohorts. Further, in order to view how consistent the results of individual cohorts are with the pooled results, the overlapped map of cohorts was further overlapped with the binary map of the pooled data. The number of overlapped voxels across 4 and 3 cohorts was quantified using Dice overlap coefficient (Dice, 1944; Burunat et al., 2016) where the voxel number of intersection was divided by the total voxel number of all the cohorts.

## RESULTS

### Results of Pooled Data in Conventional Frequency Band

The abnormal brain regions in the conventional low frequency band (0.01–0.08 Hz) for children with ADHD of the pooled data were shown in **Figure 2** and **Table 2**. Children with ADHD had increased ALFF and DC in the bilateral lingual gyrus (**Figures 2A,C**). ReHo and DC were decreased in the bilateral cerebellum. In addition, the three methods detected some method-specific abnormality such as the bilateral paracentral lobule (**Figure 2B**) and the left insula (**Figure 2C**).

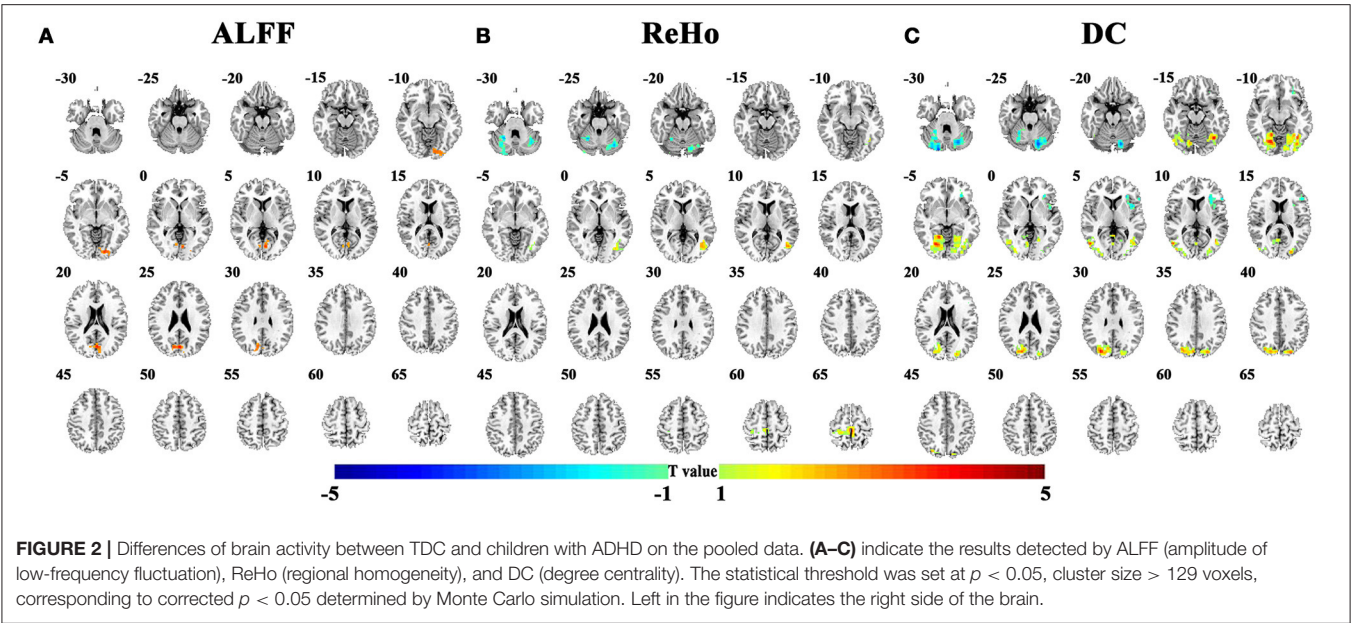
### Consistency across Cohorts in Conventional Frequency Band

The abnormal brain activity in the conventional frequency band (0.01–0.08 Hz) was identified for each cohort, and the overlapped results across the four cohorts were shown in **Figure 3** (See details of each cohort in Supplementary Figures 1–6). Only a few voxels showed overlapped abnormality from three or four cohorts by any method (ALFF, ReHo, or DC). Using DC, we observed 6 voxels overlapped from NYU, PKU2, and PKU3 in the left inferior occipital gyrus and fusiform gyrus. Even if taking the overlapped abnormality from 2 cohorts into consideration, only a few clusters were overlapped, e.g., the cerebellum by ReHo as well by DC (**Figures 3B,C**), the bilateral cuneus (**Figure 3C**) by DC.

The overlapped results of the pooled data and individual cohorts were shown in **Figure 4**. Some clusters detected in individual cohorts could not be observed in the results of pooled data, e.g., in the cuneus for ReHo (purple marked in **Figure 4B**) and thalamus for DC (purple marked in **Figure 4C**). Although, some clusters could be identified as the overlapped regions from two cohorts, they were not be observed in the pooled data, such as the right cerebellum for ReHo (yellow marked in **Figure 4B**).

The overlapped SES maps of each cohort and pooled data for ALFF, ReHo, and DC were shown in **Figure 5** (with a combined threshold of  $SES > 0.3$ ). Some clusters showing overlaps from more than 3 cohorts could be also shown in the pooled data (red marked). These clusters included the bilateral cerebellum for ReHo and DC (**Figures 5B,C**), right calcarine for ALFF and DC (**Figures 5A,C**) and the bilateral paracentral lobule for ReHo (**Figure 5B**). However, if the SES threshold was set at 0.5, these





**FIGURE 2 |** Differences of brain activity between TDC and children with ADHD on the pooled data. (A–C) indicate the results detected by ALFF (amplitude of low-frequency fluctuation), ReHo (regional homogeneity), and DC (degree centrality). The statistical threshold was set at  $p < 0.05$ , cluster size  $> 129$  voxels, corresponding to corrected  $p < 0.05$  determined by Monte Carlo simulation. Left in the figure indicates the right side of the brain.

**TABLE 2 |** Differences between TDC and ADHD on pooled data.

Method	Region	L/R	BA	Peak MNI coordinates			<i>t</i>	Number of voxels
				<i>x</i>	<i>y</i>	<i>z</i>		
ALFF	Lingual gyrus/Cuneus	L/R	18	0	−75	24	3.59	236
ReHo	Cerebellum	L/R	–	33	−69	−36	−3.70	537
	Paracentral lobule/Postcentral gyrus	L/R	3/4/6	−3	−33	69	4.15	187
	Mid. temporal gyrus	L	19	−6	51	−6	3.34	332
DC	Cerebellum	L/R	–	−18	−75	−24	−3.96	744
	Mid. occipital/Lingual gyrus	R	18	21	−72	−9	4.28	1,054
	Mid. occipital/Lingual gyrus	L	18	−39	−63	−12	4.30	635
	Insula	L	13	−30	30	9	−3.15	202

ALFF, amplitude of low-frequency fluctuation; ReHo, regional homogeneity; DC, degree centrality; Mid., middle; L, left; R, right; BA, Brodmann’s area.

clusters showed no overlap (Figure 6). The overlapped SES maps across the 4 cohorts and the SES maps of the pooled data were shown in Supplementary Figure 7.

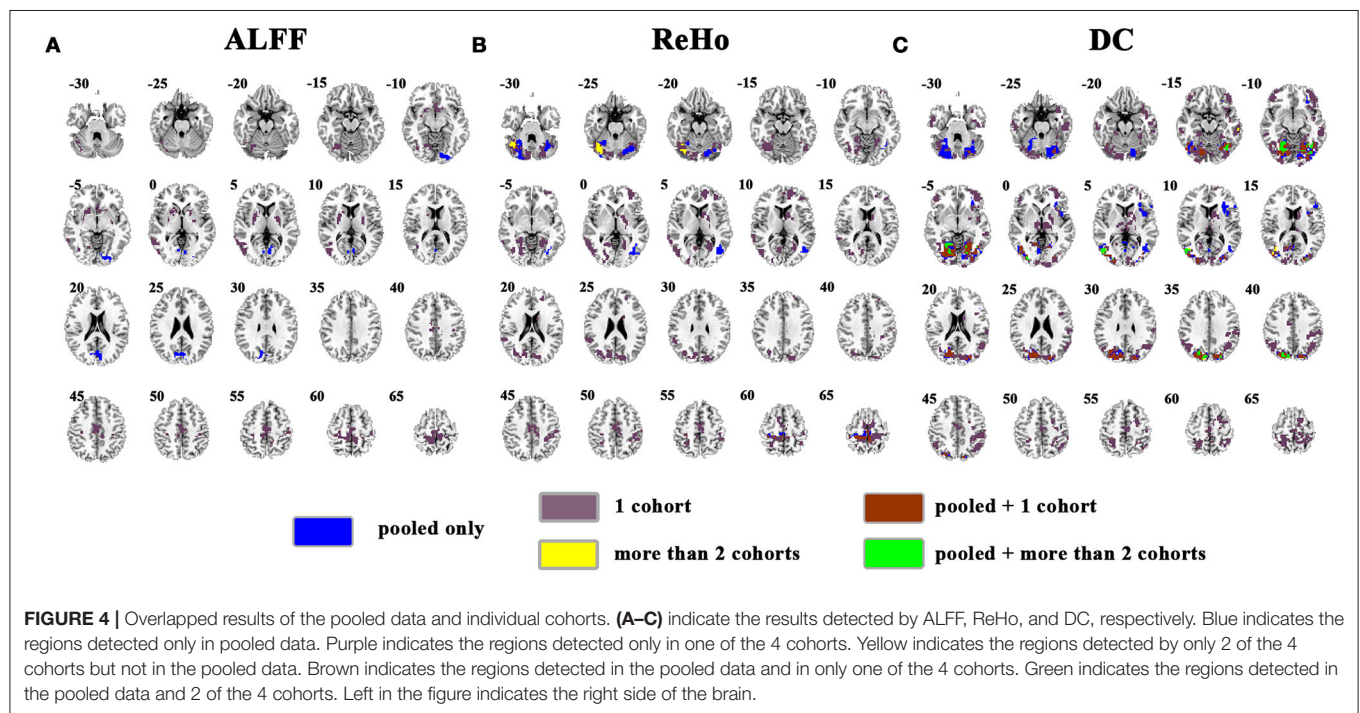
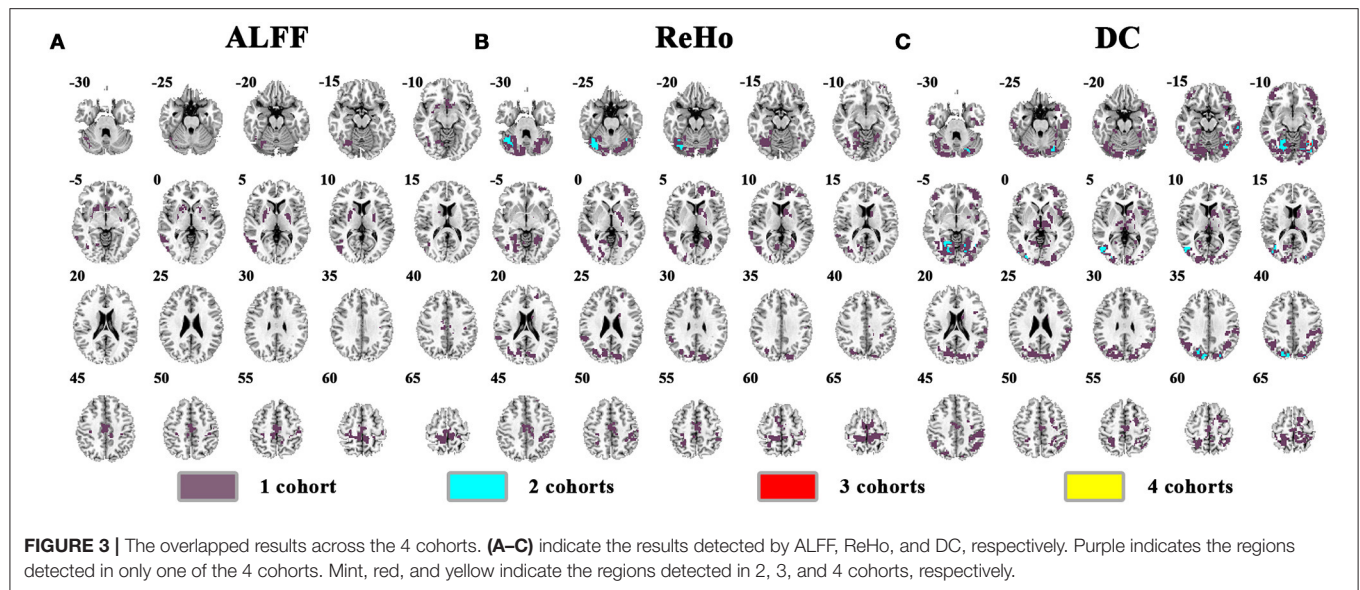
Consistency across Cohorts in Sub-Frequency Bands

After investigation in conventional low frequency band (0.01–0.08 Hz) as shown above, overlapped results across cohort were further examined in several sub-frequency bands including Slow-6/5/4/3/2. Furthermore, to reduce the possibility of false negative results and, a more lenient threshold ( $p < 0.05$ , cluster size  $> 10$  voxels) was also applied for each cohort. There is no voxel overlapped by all the cohorts. The number of the overlapped voxels was not more than 12 across three cohorts, and the highest Dice overlap coefficient is only 0.0131 (Table 3). In each sub-frequency band, most overlapped clusters were also observed from 2 cohorts (see details in Supplementary Figures 8–10).

DISCUSSION

The present study examined the consistency of abnormal local brain activity across cohorts of ADHD-200. We applied three voxel-wise whole brain analytic methods (ALFF, ReHo, and DC), strict and lenient statistical thresholds, and conventional frequency band (0.01–0.08 Hz) and sub-frequency bands (Slow/2/3/4/5/6) in the analysis process. Results from these analyses indicated that the abnormal local brain activity across cohorts of ADHD-200 was inconsistent.

The data of all four cohorts were first pooled together in the present study, as the general process way of the studies using ADHD-200 (Sato et al., 2013; Zhang et al., 2014). The abnormal brain activity for ADHD was identified in the clusters, such as the bilateral sensorimotor cortices and the bilateral lingual gyrus. Our further analysis showed that these results from pooled data were not consistent across cohorts. Most of the clusters identified for pooled data could not be observed in the results for individual

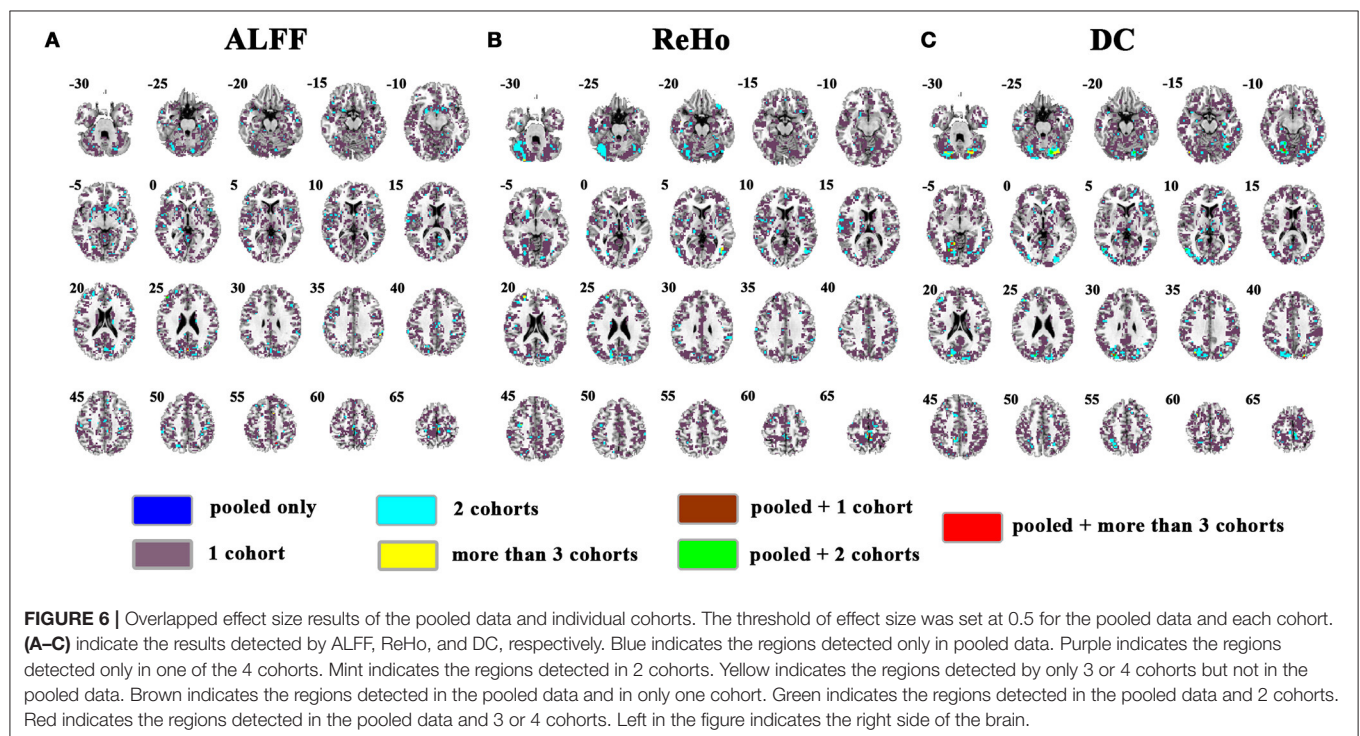
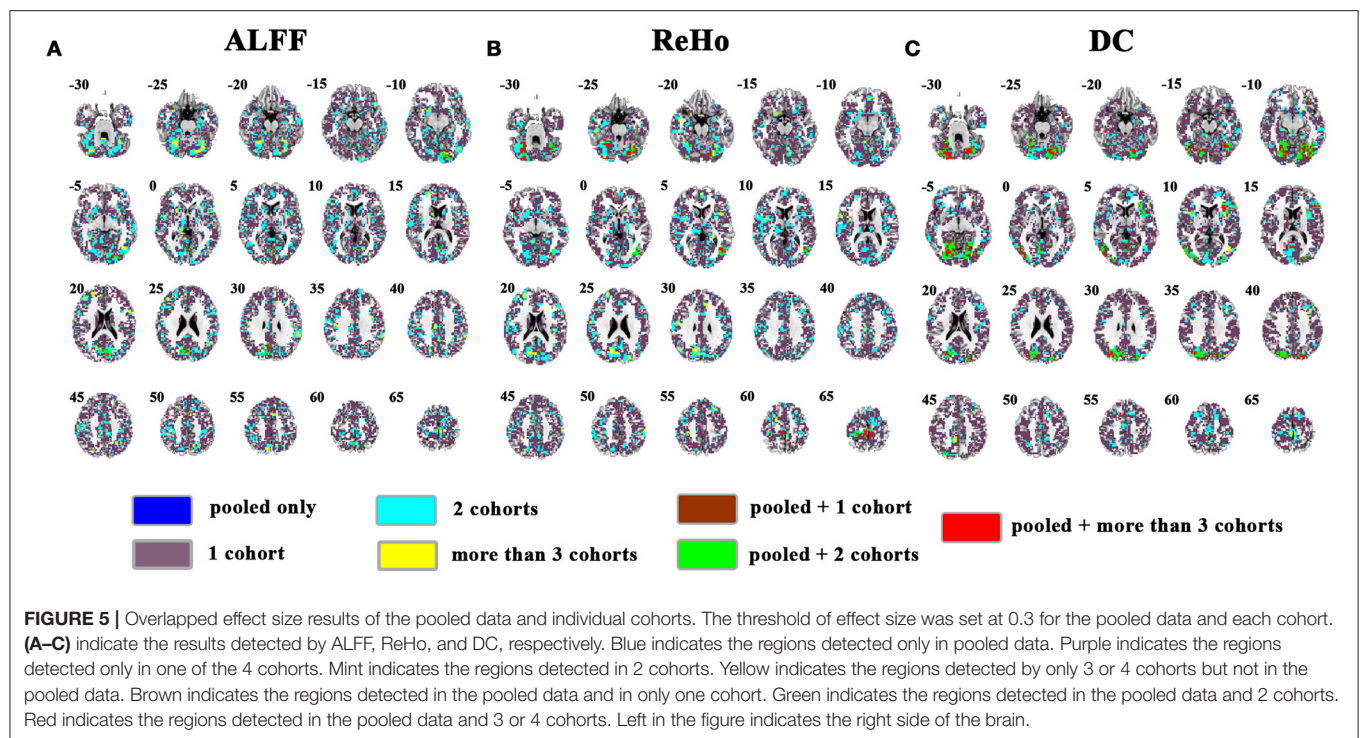


cohort. This finding was further supported by the analyses of SES. The overlapped regions did not reach a medium (0.5) level. Thus, the results of directly pooled data from different cohorts do not mean consistent results among the cohorts included, and the SES of the results should be examined in the future studies of large sample dataset. Future studies derived from multi-site large-sample dataset should not only present the statistical result of a pooled data, but also present the results of each cohort of both *t*-map and SES.

Moreover, all examined cohorts did not exhibit overlapped clusters, suggesting a high heterogeneity of ADHD. We noticed

a recent finding that detected the consistent abnormality across cohorts of ADHD-200 (Cai et al., 2015). Using the resource allocation index (RAI) (a measure of network interactions across the SN, CEN, and DMN), Cai et al. found RAI was significantly lower in children with ADHD than in control subjects and the results were reproducible across three independent cohorts. While abnormality of network interaction may reveal the complexity of spontaneous brain activity in ADHD, it could not illustrate which brain region is abnormal. From the perspective of clinical practice, analytic methods for precise localization of the abnormality in a whole-brain voxel-based





way should be emphasized. Whole-brain voxel-based analysis facilitates coordinate-based meta-analysis (CB-meta) which can help to define precise localization of abnormal spontaneous brain activity by quantitatively aggregating independent results reported in a standard coordinate space (Eickhoff et al., 2009)

and further help to guide intervention therapies, such as deep brain stimulation and transcranial magnetic stimulation (Zang et al., 2015). Thus, the present study used three whole-brain voxel-based measurements, i.e., ALFF, ReHo, and DC. These measurements are widely employed in rs-fMRI studies to access

**TABLE 3 |** Clusters which were the overlap for three/four cohorts and contained maximal number of voxels.

Method	Number of overlapped cohorts	Region	L/R	BA	Number of overlapped voxels	Dice
<b>CONVENTIONAL BAND (0.01–0.08 Hz)</b>						
ALFF	4			None		
	3			None		
ReHo	4			None		
	3	Paracentral lobule	R	4	12	0.0041
DC	4			None		
	3	Cerebellum	R	–	12	0.0096
<b>SLOW-6 (0–0.01 Hz)</b>						
ALFF	4			None		
	3			None		
ReHo	4			None		
	3	Med. frontal cortex	L	11	1	0.0004
DC	4			None		
	3			None		
<b>SLOW-5 (0.01–0.027 Hz)</b>						
ALFF	4			None		
	3			None		
ReHo	4			None		
	3	Paracentral lobule	L/R	4	3	0.0016
DC	4			None		
	3	Paracentral lobule	L/R	4	7	0.0025
<b>SLOW-4 (0.027–0.073 Hz)</b>						
ALFF	4			None		
	3			None		
ReHo	4			None		
	3	Cerebellum	R	–	6	0.0026
DC	4			None		
	3	Mid. occipital gyrus	L	19	11	0.0131
<b>SLOW-3 (0.073–0.198 Hz)</b>						
ALFF	4			None		
	3	Paracentral lobule	L	4	1	0.0005
ReHo	4			None		
	3	Mid. frontal gyrus	L	8	2	0.0013
DC	4			None		
	3	Supplementary motor area	R	6	1	0.0004
<b>SLOW-2 (0.198–0.25 Hz)</b>						
ALFF	4			None		
	3			None		
ReHo	4			None		
	3	Sup. frontal gyrus	R	6	2	0.001
DC	4			None		
	3			None		

Sup., superior; Mid., Middle; Med., Medial; L, left; R, right. The threshold was  $p < 0.05$  and cluster size  $> 10$  voxels for each cohort.

local brain activity from different aspects. Here we applied these three measurements to explore the consistent local abnormality of children with ADHD across cohorts. Nevertheless, consistent results across cohorts were not identified through any one of the three measurements.

The present study not only focused on the conventional frequency band but also stressed several sub-frequency bands. Frequency-dependent investigation provides us a new prospect to investigate the physiological mechanism of the brain activity. A recent rs-fMRI study reported some frequency-dependent abnormalities for children with ADHD (Yu et al., 2015). For example, in the orbital frontal cortex (OFC), the frequency bands of slow-3 and slow-2 contributed more to the differences than did the slow-5 and slow-4 bands. We found that the detected differences between ADHD and TDC are different according to different frequency bands. For example, compared with TDC, children with ADHD had decreased DC in the left inferior parietal gyrus only in slow-3 but others frequency bands and decreased DC in the bilateral putamen/thalamus only in slow-4 but others frequency bands (Supplementary Figure 11). Previous studies often consider the Slow-6 ( $<0.01$  Hz) as signal drift, and it was usually discarded from further analysis. However, our recent publications on finger force feedback task have challenged this issue. ReHo and ALFF of basal ganglia in Slow-6 showed difference between real and sham feedback conditions, and the ALFF in Slow-6 was related to finger force (Zhang et al., 2015a,b). Moreover, ReHo difference between ADHD and TDC in Slow-6 was detected in previous study (Yu et al., 2015). Thus, the Slow-6 was involved in our analysis and differences can be detected. However, the results couldn't be detected in any other cohorts.

Several limitations exist in the present study. First, we could not explore the contribution of different subtype to the inconsistency in ADHD neuroimaging findings because of the small sample size for statistical analysis. For example, PKU1 only included 6 inattention and 9 combined subjects. Their contribution should be explored on a large sample dataset in the future. Second, we only used three whole brain voxel-based measurements to evaluate the consistency across cohorts. Thus, our observations were restricted to these measurements. Investigations with more whole-brain voxel-based measurements will be helpful.

## CONCLUSIONS

Data-sharing projects like ADHD-200 provide large sample analysis. But pooled data itself is not enough. The current study used three whole-brain voxel-based analytic methods, i.e., ALFF, ReHo, and DC not only on the pooled data but also on each individual cohort. We found that the findings based on the pooled data of ADHD-200 were inconsistent across the individual cohorts. Even in a more lenient threshold, this inconsistency could be observed. Such inconsistency could be found not only in the conventional low frequency-band (0.01–0.08 Hz) but also in a few sub-frequency band of Slow-2/3/4/5/6. These results support the view that ADHD is a highly heterogeneous disorder. Future studies should try more efforts



on exploring more consistent findings of rs-fMRI data of ADHD. Data sharing could benefit improving the reproducibility of neuroimage studies, and we suggest that analysis based on multi-site large-sample dataset should be performed on pooled data and single cohort, respectively.

## AUTHOR CONTRIBUTIONS

YZ, HZ, and QC conceived and designed the experiment. JW and HZ performed the data analysis. YZ, LZ, LS, and YW provided advice on the analysis and interpretation of the results. JW, HZ, and YZ wrote the paper.

## FUNDING

This work is supported by grants from the National Key Basic Research Program of China [973 Program, 2014CB846104], the National Natural Science Foundation of China [81271652,

81520108016, 31471084, 81401481, 81371496, and 81471382], the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning [CNLYB1508] and Key Laboratory for Magnetic Resonance and Multimodality Imaging of Guangdong Province [2014B030301013]. YZ is partly supported by “Qian Jiang Distinguished Professor” program.

## ACKNOWLEDGMENTS

The authors acknowledge the contribution of ADHD-200 consortium organizers for sharing the raw data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2017.00320/full#supplementary-material>

## REFERENCES

- ADHD-200-Consortium, T. (2012). The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Front. Syst. Neurosci.* 6:62. doi: 10.3389/fnsys.2012.00062
- Bartra, O., McGuire, J. T., and Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427. doi: 10.1016/j.neuroimage.2013.02.063
- Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., et al. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J. Neurosci.* 29, 1860–1873. doi: 10.1523/JNEUROSCI.5062-08.2009
- Burunat, I., Toivainen, P., Alluri, V., Bogert, B., Ristaniemi, T., Sams, M., et al. (2016). The reliability of continuous brain responses during naturalistic listening to music. *Neuroimage* 124, 224–231. doi: 10.1016/j.neuroimage.2015.09.005
- Bush, G. (2010). Attention-deficit/hyperactivity disorder and attention networks. *Neuropsychopharmacology* 35, 278–300. doi: 10.1038/npp.2009.120
- Cai, W., Chen, T., Szegletes, L., Supekar, K., and Menon, V. (2015). Aberrant cross-brain network interaction in children with attention-deficit/hyperactivity disorder and its relation to attention deficits: a multisite and cross-site replication study. *Biol. Psychiatry*. doi: 10.1016/j.biopsych.2015.10.017. [Epub ahead of print].
- Carmona, S., Hoekzema, E., Castellanos, F. X., Garcia-Garcia, D., Lage-Castellanos, A., Van Dijk, K. R., et al. (2015). Sensation-to-cognition cortical streams in attention-deficit/hyperactivity disorder. *Hum. Brain Mapp.* 36, 2544–2557. doi: 10.1002/hbm.22790
- Castellanos, F. X., Sonuga-Barke, E. J. S., Milham, M. P., and Tannock, R. (2006). Characterizing cognition in ADHD: beyond executive dysfunction. *Trends Cogn. Sci.* 10, 117–123. doi: 10.1016/j.tics.2006.01.011
- Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: a MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge.
- Cole, M. W., Pathak, S., and Schneider, W. (2010). Identifying the brain's most globally connected regions. *Neuroimage* 49, 3132–3148. doi: 10.1016/j.neuroimage.2009.11.001
- Cortese, S., Kelly, C., Chabernaud, C., Proal, E., Di Martino, A., Milham, M. P., et al. (2012). Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. *Am. J. Psychiatry* 169, 1038–1055. doi: 10.1176/appi.ajp.2012.11101521
- Dice, L. R. (1944). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302. doi: 10.2307/1932409
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., and Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* 30, 2907–2926. doi: 10.1002/hbm.20718
- Elton, A., Alcauter, S., and Gao, W. (2014). Network connectivity abnormality profile supports a categorical-dimensional hybrid model of ADHD. *Hum. Brain Mapp.* 35, 4531–4543. doi: 10.1002/hbm.22492
- Fox, M. D., Zhang D., Snyder, A. Z., and Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *J. Neurophysiol.* 101, 3270–3283. doi: 10.1152/jn.90777.2008
- Han, Y., Wang, J., Zhao, Z., Min, B., Lu, J., Li, K., et al. (2011). Frequency-dependent changes in the amplitude of low-frequency fluctuations in amnesic mild cognitive impairment: a resting-state fMRI study. *Neuroimage* 55, 287–295. doi: 10.1016/j.neuroimage.2010.11.059
- Herz, D. M., Eickhoff, S. B., Lokkegaard, A., and Siebner, H. R. (2014). Functional neuroimaging of motor control in Parkinson's disease: a meta-analysis. *Hum. Brain Mapp.* 35, 3227–3237. doi: 10.1002/hbm.22397
- Iwabuchi, S. J., Krishnadas, R., Li, C., Auer, D. P., Radua, J., and Palaniyappan, L. (2015). Localized connectivity in depression: a meta-analysis of resting state functional imaging studies. *Neurosci. Biobehav. Rev.* 51, 77–86. doi: 10.1016/j.neubiorev.2015.01.006
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132
- Li, S., Ma, X., Huang, R., Li, M., Tian, J., Wen, H., et al. (2015). Abnormal degree centrality in neurologically asymptomatic patients with end-stage renal disease: a resting-state fMRI study. *Clin Neurophysiol.* 127, 602–609. doi: 10.1016/j.clinph.2015.06.022
- Liao, X.-H., Xia, M.-R., Xu, T., Dai, Z.-J., Cao, X.-Y., Niu, H.-J., et al. (2013). Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. *Neuroimage* 83, 969–982. doi: 10.1016/j.neuroimage.2013.07.058
- Lv, Y., Margulies, D. S., Villringer, A., and Zang, Y.-F. (2013). Effects of finger tapping frequency on regional homogeneity of sensorimotor cortex. *PLoS ONE* 8:e64115. doi: 10.1371/journal.pone.0064115
- Malinen, S., Vartiainen, N., Hlushchuk, Y., Koskinen, M., Ramkumar, P., Forss, N., et al. (2010). Aberrant temporal and spatial brain activity during rest in patients with chronic pain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6493–6497. doi: 10.1073/pnas.1001504107

- Mills, K. L., Bathula, D., Dias, T. G., Iyer, S. P., Fenesy, M. C., Musser, E. D., et al. (2012). Altered cortico-striatal-thalamic connectivity in relation to spatial working memory capacity in children with ADHD. *Front. Psychiatry* 3:2. doi: 10.3389/fpsy.2012.00002
- Molloy, E. K., Meyerand, M. E., and Birn, R. M. (2014). The influence of spatial resolution and smoothing on the detectability of resting-state and task fMRI. *Neuroimage* 86, 221–230. doi: 10.1016/j.neuroimage.2013.09.001
- Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., and Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* 44, 893–905. doi: 10.1016/j.neuroimage.2008.09.036
- Newman, E., Jernigan, T. L., Lisdahl, K. M., Tamm, L., Tapert, S. F., Potkin, S. G., et al. (2015). Go/No Go task performance predicts cortical thickness in the caudal inferior frontal gyrus in young adults with and without ADHD. *Brain Imaging Behav.* 10, 880–892. doi: 10.1007/s11682-015-9453-x
- Otti, A., Guendel, H., Wohlschlaeger, A., Zimmer, C., and Noll-Hussong, M. (2013). Frequency shifts in the anterior default mode network and the salience network in chronic pain disorder. *BMC Psychiatry* 13:84. doi: 10.1186/1471-244X-13-84
- Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., and Rohde, L. A. (2015). Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J. Child Psychol. Psychiatry* 56, 345–365. doi: 10.1111/jcpp.12381
- Sato, J. R., Takahashi, D. Y., Hoexter, M. Q., Massier, K. B., and Fujita, A. (2013). Measuring network's entropy in ADHD: a new approach to investigate neuropsychiatric disorders. *Neuroimage* 77, 44–51. doi: 10.1016/j.neuroimage.2013.03.035
- Schulz, K. P., Fan, J., Tang, C. Y., Newcorn, J. H., Buchsbaum, M. S., Cheung, A. M., et al. (2004). Response inhibition in adolescents diagnosed with attention deficit hyperactivity disorder during childhood: an event-related fMRI study. *Am. J. Psychiatry* 161, 1650–1657. doi: 10.1176/appi.ajp.161.9.1650
- Song, X. W., Dong, Z. Y., Long, X. Y., Li, S. F., Zuo, X. N., Zhu, C. Z., et al. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS ONE* 6:e25031. doi: 10.1371/journal.pone.0025031
- Sripada, C., Kessler, D., Fang, Y., Welsh, R. C., Prem Kumar, K., and Angstadt, M. (2014). Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder. *Hum. Brain Mapp.* 35, 4693–4705. doi: 10.1002/hbm.22504
- Tomasi, D., and Volkow, N. D. (2012). Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. *Biol. Psychiatry* 71, 443–450. doi: 10.1016/j.biopsych.2011.11.003
- Vaidya, C. J., Bunge, S. A., Dudukovic, N. M., Zalecki, C. A., Elliott, G. R., and Gabrieli, J. D. E. (2005). Altered neural substrates of cognitive control in childhood ADHD: evidence from functional magnetic resonance imaging. *Am. J. Psychiatry* 162, 1605–1613. doi: 10.1176/appi.ajp.162.9.1605
- Vasic, N., Plichta, M. M., Wolf, R. C., Fallgatter, A. J., Susic-Vasic, Z., and Gron, G. (2014). Reduced neural error signaling in left inferior prefrontal cortex in young adults with ADHD. *J. Atten. Disord.* 18, 659–670. doi: 10.1177/1087054712446172
- Wang, J., Zhang, Z., Ji, G. J., Xu, Q., Huang, Y., Wang, Z., et al. (2015). Frequency-specific alterations of local synchronization in idiopathic generalized epilepsy. *Medicine* 94:e1374. doi: 10.1097/md.0000000000001374
- Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., et al. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage* 76, 183–201. doi: 10.1016/j.neuroimage.2013.03.004
- Yan, C. G., Wang, X. D., Zuo, X. N., and Zang, Y. F. (2016). DPABI: data processing & analysis for (resting-state) brain imaging. 14, 339–351. doi: 10.1007/s12021-016-9299-4
- Yu, X., Yuan, B., Cao, Q., An, L., Wang, P., Vance, A., et al. (2015). Frequency-specific abnormalities in regional homogeneity among children with attention deficit hyperactivity disorder: a resting-state fMRI study. *Sci. Bull.* 61, 682–692. doi: 10.1007/s11434-015-0823-y
- Yuan, B. K., Wang, J., Zang, Y. F., and Liu, D. Q. (2014). Amplitude differences in high-frequency fMRI signals between eyes open and eyes closed resting states. *Front. Hum. Neurosci.* 8:503. doi: 10.3389/fnhum.2014.00503
- Zang, Y., Jiang, T., Lu, Y., He, Y., and Tian, L. (2004). Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22, 394–400. doi: 10.1016/j.neuroimage.2003.12.030
- Zang, Y. F., He, Y., Zhu, C. Z., Cao, Q. J., Sui, M. Q., Liang, M., et al. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev.* 29, 83–91. doi: 10.1016/j.braindev.2006.07.002
- Zang, Y. F., Zuo, X. N., Milham, M., and Hallett, M. (2015). Toward a meta-analytic synthesis of the resting-state fMRI literature for clinical populations. *Biomed Res. Int.* 2015:435265. doi: 10.1155/2015/435265
- Zhang, H., Gao, Z. Z., and Zang, Y. F. (2015a). An fMRI study of local synchronization in different subfrequency bands during the continuous feedback of finger force. *Biomed Res. Int.* 2015:273126. doi: 10.1155/2015/273126
- Zhang, H., Zhang, L., and Zang, Y. (2015b). Fluctuation amplitude and local synchronization of brain activity in the ultra-low frequency band: an fMRI investigation of continuous feedback of finger force. *Brain Res.* 1629, 104–112. doi: 10.1016/j.brainres.2015.10.023
- Zhang, J., Kendrick, K. M., Lu, G., and Feng, J. (2014). The fault lies on the other side: altered brain functional connectivity in psychiatric disorders is mainly caused by counterpart regions in the opposite hemisphere. *Cereb. Cortex* 25, 3475–3486. doi: 10.1093/cercor/bhu173
- Zhang, J., Wei, L., Hu, X., Zhang, Y., Zhou, D., Li, C., et al. (2013). Specific frequency band of amplitude low-frequency fluctuation predicts Parkinson's disease. *Behav. Brain Res.* 252, 18–23. doi: 10.1016/j.bbr.2013.05.039
- Zuo, X.-N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., et al. (2010). The oscillating brain: complex and reliable. *Neuroimage* 49, 1432–1445. doi: 10.1016/j.neuroimage.2009.09.037
- Zuo, X. N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., et al. (2012). Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875. doi: 10.1093/cercor/bhr269

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Wang, Zheng, Cao, Wang, Sun, Zang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Investigating Brain Connectomic Alterations in Autism Using the Reproducibility of Independent Components Derived from Resting State Functional MRI Data

Mohammed A. Syed<sup>1</sup>, Zhi Yang<sup>2,3</sup>, Xiaoping P. Hu<sup>4</sup> and Gopikrishna Deshpande<sup>5,6,7\*</sup>

<sup>1</sup> Computer Science and Software Engineering Department, Auburn University, Auburn, AL, United States, <sup>2</sup> Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China, <sup>3</sup> Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai, China, <sup>4</sup> The Department of Bioengineering, University of California, Riverside, Riverside, CA, United States, <sup>5</sup> The Department of Electrical and Computer Engineering, AU MRI Research Center, Auburn University, Auburn, AL, United States, <sup>6</sup> The Department of Psychology, Auburn University, Auburn, AL, United States, <sup>7</sup> The Alabama Advanced Imaging Consortium at Auburn University, University of Alabama Birmingham, Auburn, AL, United States

## OPEN ACCESS

### Edited by:

Bharat B. Biswal,  
University of Medicine and Dentistry of  
New Jersey, United States

### Reviewed by:

Xu Lei,  
Southwest University, China  
Rui Li,  
Institute of Psychology (CAS), China

### \*Correspondence:

Gopikrishna Deshpande  
gopi@auburn.edu

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 01 January 2017

**Accepted:** 31 July 2017

**Published:** 08 September 2017

### Citation:

Syed MA, Yang Z, Hu XP and  
Deshpande G (2017) Investigating  
Brain Connectomic Alterations in  
Autism Using the Reproducibility of  
Independent Components Derived  
from Resting State Functional MRI  
Data. *Front. Neurosci.* 11:459.  
doi: 10.3389/fnins.2017.00459

**Significance:** Autism is a developmental disorder that is currently diagnosed using behavioral tests which can be subjective. Consequently, objective non-invasive imaging biomarkers of Autism are being actively researched. The common theme emerging from previous functional magnetic resonance imaging (fMRI) studies is that Autism is characterized by alterations of fMRI-derived functional connections in certain brain networks which may provide a biomarker for objective diagnosis. However, identification of individuals with Autism solely based on these measures has not been reliable, especially when larger sample sizes are taken into consideration.

**Objective:** We surmise that metrics derived from Autism subjects may not be highly reproducible within this group leading to poor generalizability. We hypothesize that functional brain networks that are most reproducible within Autism and healthy Control groups separately, but not when the two groups are merged, may possess the ability to distinguish effectively between the groups.

**Methods:** In this study, we propose a “discover-confirm” scheme based upon the assessment of reproducibility of independent components obtained from resting state fMRI (discover) followed by a clustering analysis of these components to evaluate their ability to discriminate between groups in an unsupervised way (confirm).

**Results:** We obtained cluster purity ranging from 0.695 to 0.971 in a data set of 799 subjects acquired from multiple sites, depending on how reproducible the corresponding components were in each group.

**Conclusion:** The proposed method was able to characterize reproducibility of brain networks in Autism and could potentially be deployed in other mental disorders as well.

**Keywords:** autism, fMRI, independent components, reproducibility, clustering

## INTRODUCTION

Autism Spectrum Disorder (ASD) is characterized as a developmental disability leading to significant social, communication and behavioral challenges (American Psychiatric Association, 2013). In 2010, an estimate from the Autism and Developmental Disabilities Monitoring (ADDMM) Network involving 11 sites revealed that 14.7 per 1,000 or 1 in 68 children aged 8 years were affected by this disorder (Wingate et al., 2012; Baio, 2014). In addition, this study discovered that one in 54 males and one in 252 females in the ADDMM communities had Autism. These disorders have been found to be very heritable (Muhle et al., 2004). In addition, approximately 18.7% of infants with at least one older sibling with Autism developed this disorder (Ozonoff et al., 2011). Given the societal implications of Autism, early diagnosis and intervention has become paramount. However, Autism is currently diagnosed using behavioral tests which can be subjective. Consequently, objective non-invasive biomarkers of Autism are being actively researched.

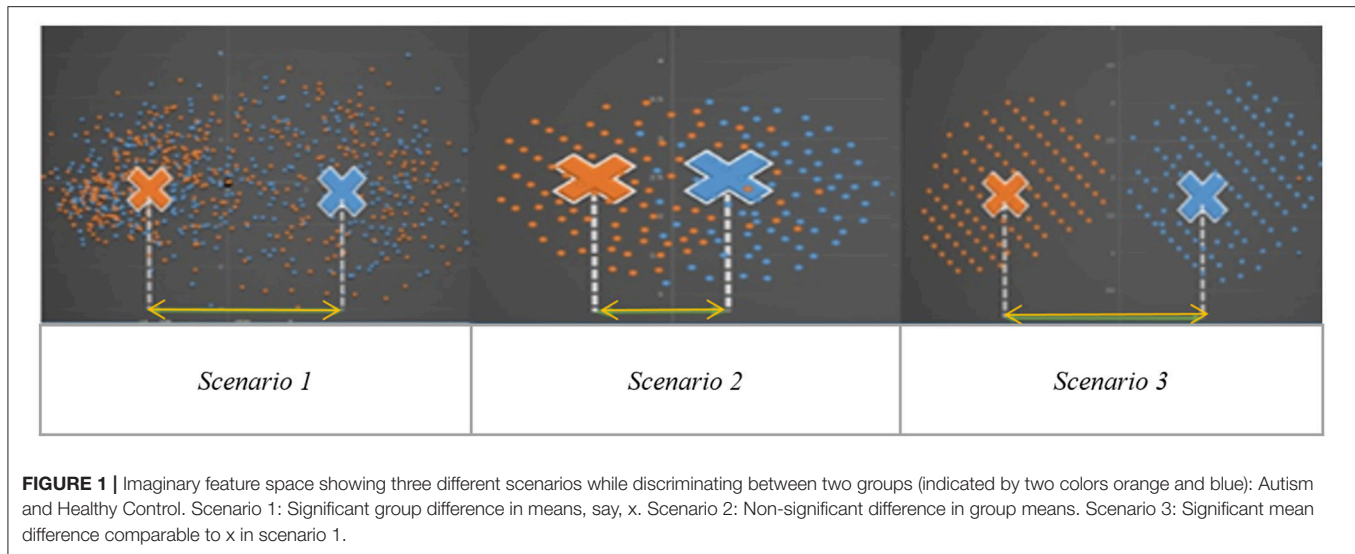
In order to find objective biomarkers of Autism, studies have used information from brain imaging techniques such as structural Magnetic Resonance Imaging (MRI). Ecker et al. (2010) used a multiparameter classification approach involving a support vector machine (SVM) to characterize the structural pattern of gray matter anatomy in adults with ASD and examined a set of five morphological parameters such as volumetric and geometric features at each spatial location on the cortical surface to discriminate between people with ASD and controls. Jiao et al. (2010) built diagnostic models for ASD based upon regional thickness measurements extracted from surface-based morphometry (SBM) and compared these models to diagnostic models based on volumetric morphometry using four machine learning techniques: support vector machines (SVM), multilayer perceptrons (MLPs), functional trees (FTs), and logistic model trees (LMTs). Voxel-based morphometry along with a multivariate pattern analysis approach was used by Uddin et al. (2011) to determine multiple brain regions showing atypical structural organization in children with Autism. Calderoni et al. (2012) examined whole brain volumes of female subjects with ASD using mass-univariate and pattern classification approaches. Sato et al. (2013) extracted individual subject features from inter-regional thickness correlations based on structural MRI which were later used in a machine learning framework to obtain subject level prediction of severity scores based upon neurobiological criteria rather than behavioral information. Libero et al. (2015) examined multiple brain imaging modalities to investigate the neural architecture in the same set of subjects using techniques such as decision tree classification analysis. Functional (as opposed to structural) MRI has been used in several studies on Autism as well. The feasibility of a functional MRI connectivity diagnostic assay for Autism was investigated by Anderson et al. (2011) after obtaining pairwise functional connectivity measurements from a lattice of 7,266 regions of interest covering the entire gray matter and using a single resting state blood oxygen level-dependent scan of 8 min for classification in each subject. Coutanche et al. (2011) used data from an fMRI study of the neural

basis for face processing in subjects with ASD to illustrate that multi-voxel pattern analysis (MVPA) may provide a sensitive functional biomarker of clinical symptom severity. Wang et al. (2012) used a multi-scale clustering methodology known as "data cloud geometry" to extract functional connectivity patterns from fMRI for the recognition of ASD subjects by applying it to correlation matrices of 106 regions of interest (ROIs) in subjects with ASD and controls. Deshpande et al. (2013) used supervised machine learning and fMRI to show alterations in causal connectivity in the brain could serve as a potential non-invasive neuroimaging signature for Autism. Nielsen et al. (2013) also used pairwise functional connectivity measurements from a lattice of 7,266 regions of interest covering the gray matter for 964 subjects to conclude that multisite classification based on functional connectivity derived from resting state fMRI of Autism performed better than chance using a simple leave-one-out classifier. Maximo et al. (2013) used regional homogeneity and local density approaches at different spatial scales and examined local connectivity in ASD, while Supekar et al. (2013) showed hyper-connectivity in a sample of relatively younger Autistic kids using resting state fMRI. The common theme emerging from the studies mentioned above is that Autism is characterized by altered functional connectivity in certain brain networks and that characterizing this appropriately using MRI-based methods may provide a biomarker for objective diagnosis.

Independent Component Analysis (ICA) is a blind source separation technique which is commonly employed for extracting brain networks involving spatially distributed regions with similar/correlated temporal activity (Bell and Sejnowski, 1995), especially in the baseline resting state. Consequently, it has been applied to investigate altered brain networks in Autism using fMRI. Specifically, Von von dem et al. (2013) employed ICA to demonstrate that individuals with Autism had reduced functional connectivity within the Default Mode Network (DMN), an important resting state brain network (Greicius et al., 2003). Assaf et al. (2010) studied the role of altered functional connectivity of the default mode sub-networks in ASDs using short resting fMRI scans and ICA. In spite of these studies showing reduced connectivity in certain brain networks in Autism, identification of individuals with Autism solely based on these measures has not been reliable, especially in samples of large sizes (Nielsen et al., 2013). We surmise that one major factor contributing to this state of affairs may be that metrics derived from Autism and/or Control subjects may not be highly reproducible within their respective group. Consequently, such metrics have poor generalizability, leading to lower cluster purities. Therefore, in this paper, we hypothesize that functional brain networks which are most reproducible separately within Autism and healthy Control groups, but not reproducible when both groups are merged, may possess the ability to effectively discriminate between the groups. The basis for this hypothesis is illustrated in **Figure 1** which shows an imagined feature space where we want to discriminate between the two groups (Autism and healthy Control). Please note that **Figure 1** has not been drawn to scale and is an illustrative schematic.

Scenario-1 corresponds to the situation wherein the two groups have significantly different means (say,  $x$ ) in the feature





space. However, within each group, the features have poor reproducibility (i.e., they are more scattered in the feature space), likely due to the heterogeneity of the disorder. Therefore, even if the group means are statistically separated, such features will give poor cluster purity. Scenario-2 is a situation where there is no significant difference between means, but the features are reproducible in the combined group (i.e., Autism + Control group), i.e., they are less scattered in the feature space even when both groups are combined. These two scenarios indicate that features which are highly reproducible separately in each group but are not reproducible in the combined (Autism + Control) group are likely to provide purer clusters while discriminating between the Autism and Control groups. In the third scenario, the features are not only statistically separated between the groups (with the difference between the group means comparable to “ $x$ ” in Scenario-1), but also reproducible within each group, i.e., less scattered in feature space within each group. In order to test our hypothesis, traditional ICA-based characterization of the functional brain needs to be modified such that reproducibility information is considered while choosing independent components. Therefore, we propose a methodology involving assessment of reproducibility of independent components, followed by clustering analysis of such components for evaluating their discriminability between groups in an unsupervised way. Accordingly, we applied a recently introduced algorithm, “generalized Ranking and Averaging Independent Component Analysis by Reproducibility” (gRAICAR, <https://github.com/yangzhi-psy/gRAICAR>) (Yang et al., 2012), which can provide independent components that are highly reproducible within a given group of subjects. This technique is an extension of a framework previously developed for single subject analysis called Ranking and averaging independent component analysis by reproducibility—RAICAR (Yang et al., 2008) and has been successfully used in a number of applications (Yang et al., 2014a,b). In this work, gRAICAR was applied to Autism Brain Imaging Data Exchange (ABIDE) data (Di Martino et al., 2014) to estimate the independent components which are most

reproducible, in Autism and Control groups, respectively, but not reproducible in the combined group. We input the spatial maps of such independent components into a k-means clustering algorithm and determined the purity of each cluster with respect to the *a priori* clinical diagnosis received by subjects.

## MATERIALS AND METHODS

### Composition of the Subject Sample

We utilized resting-state functional magnetic resonance imaging (R-fMRI) data from 799 individuals provided by Autism Brain Imaging Data Exchange (ABIDE). The data we used had 392 individuals with Autism spectrum disorders and 407 age- and sex-matched typical controls (TCs). These data came from 13 different imaging sites and included 700 male and 99 female subjects (Table 1) between 7 and 64 years of age. Data were fully anonymized wherein all 18 HIPAA (Health Insurance Portability and Accountability)-protected health information identifiers were removed. Data contributions were based on studies approved by the local Institutional Review Boards. Detailed information regarding the imaging data sets and associated phenotypic protocols can be found at [http://fcon\\_1000.projects.nitrc.org/indi/abide](http://fcon_1000.projects.nitrc.org/indi/abide). Data acquisition parameters and individual site details are also available on this web site.

### Pre-processing

We first converted the data downloaded from ABIDE database, which was in DICOM format, to Neuroimaging Informatics Technology Initiative (NIfTI) format. In order to complete the first step, we used dcm2nii software which is freely available at <http://www.mccauslandcenter.sc.edu/mricro/mricron/dcm2nii.html>.

In the next step, we used a combination of Data Processing Assistant for Resting-State fMRI (Yan and Zang, 2010; DPARSF, <http://www.restfmri.net>), which is a plug-in software based on Statistical Parametric Mapping or SPM (<http://www.fil.ion.ucl.ac.uk/spm>), and uses functionality from Resting-State fMRI Data Analysis Toolkit (REST 1.7) (Song et al., 2011), both of which

**TABLE 1** | Institute names used in our study from ABIDE data and subject distribution by diagnosis code, Autism and Control.

ID	Institute name	Autism	Control
1	California Institute of Technology	19	19
2	Kennedy Krieger Institute	22	33
3	University of Leuven	29	35
4	Ludwig Maximilians University Munich	24	33
5	Oregon Health and Science University	12	14
6	University of Pittsburgh	30	27
7	Social Brain Lab UMC Groningen NIN	15	15
8	San Diego State University	14	21
9	Stanford University	17	18
10	Trinity College Dublin	24	25
11	University of California Los Angeles	62	47
12	University of Michigan	68	77
13	University of Utah	56	43

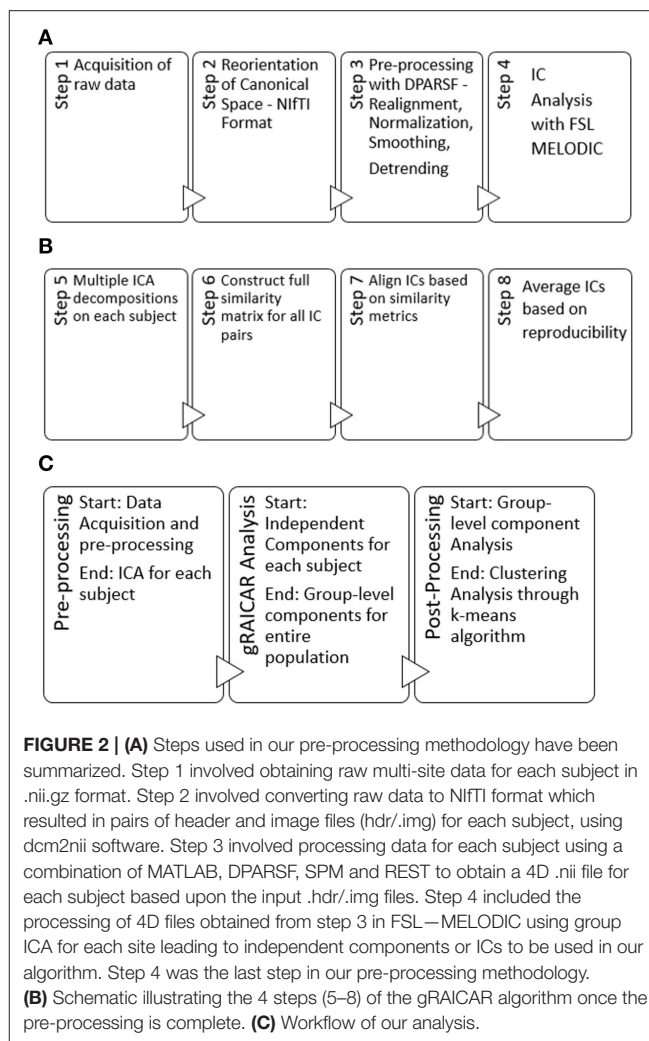
run on MATLAB. DPARSF was used to perform realignment of 3D brain volumes at each instant relative to the initial volume using 6-parameter rigid body registration, normalization to MNI (Montreal Neurological Institute) template using nonlinear warping, spatial smoothing using a Gaussian kernel with full width at half maximum of  $4\text{ mm} \times 4\text{ mm} \times 4\text{ mm}$ , de-trending using linear polynomial and temporal band-pass filtering using the frequency range of 0.01–0.1 Hz.

Four Dimensional NIFTI-1 format images (<http://nifti.nimh.nih.gov/nifti-1>) from the pre-processing described above were then used in FMRIB Software Library v5.0 (Woolrich et al., 2009; Jenkinson et al., 2012) (FSL by Analysis Group, FMRIB, Oxford, UK) to obtain a set of independent components for each subject using Multivariate Exploratory Linear Optimized Decomposition into Independent Components (MELODIC) algorithm (Beckman and Smith, 2004; Beckman et al., 2005). FSL provides analysis tools for fMRI, MRI and DTI brain imaging data, including ICA for decomposing single or multiple 4D data sets into linearly independent spatial components. More information on MELODIC is available at <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC>. We used the MELODIC analysis tool to perform standard 2D spatial ICA on each subject resulting in time courses (one per component) in the mixing matrix and spatial maps (one per component). The number of components for each subject was determined by MELODIC through automatic dimensionality estimation. We saved MELODIC results for each subject and used them in the algorithm we describe in the following section, for finding reproducible independent components.

## gRAICAR Algorithm

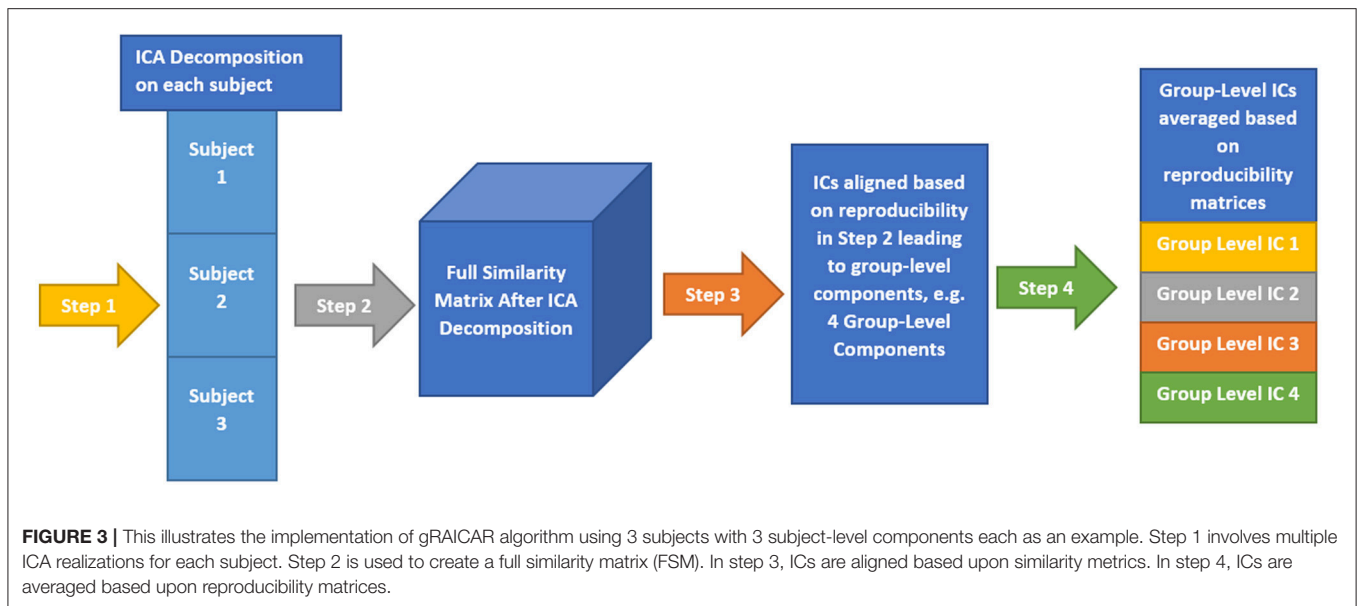
The dataset from subject  $s$  ( $s=1,2,\dots,n$ ) can be represented as a  $t_s \times v_s$  matrix,  $M_s$ , where  $t_s$  represents the number of time points and  $v_s$  the number of voxels. The data matrix,  $M_s$ , can be decomposed into  $c_s$  independent components (ICs) in spatial domain  $s_s(c_s \times v_s$  matrix,  $M_c$ ) and their corresponding mixing time courses  $a_s(t_s \times c_s$  matrix,  $M_a$ ).

Here, we provide a brief overview of gRAICAR and the readers are referred to the original paper by Yang et al. (2012) for a more comprehensive description. The algorithm contains four



**FIGURE 2** | (A) Steps used in our pre-processing methodology have been summarized. Step 1 involved obtaining raw multi-site data for each subject in .nii.gz format. Step 2 involved converting raw data to Nifti format which resulted in pairs of header and image files (hdr/.img) for each subject, using dcm2nii software. Step 3 involved processing data for each subject using a combination of MATLAB, DPARSF, SPM and REST to obtain a 4D .nii file for each subject based upon the input .hdr/.img files. Step 4 included the processing of 4D files obtained from step 3 in FSL—MELODIC using group ICA for each site leading to independent components or ICs to be used in our algorithm. Step 4 was the last step in our pre-processing methodology. (B) Schematic illustrating the 4 steps (5–8) of the gRAICAR algorithm once the pre-processing is complete. (C) Workflow of our analysis.

stages of processing as summarized in **Figure 2**. (1) The first step involved performing ICA decomposition  $d$  times for each subject using random initial values leading to  $d \times n$  realizations where  $n$  is the number of subjects. We refer to these realizations as  $REs$ . In our study,  $RE_{ij}$  refers to the ICs from  $j$ th realization of subject  $i$ . (2) In its second step, a full similarity matrix (FSM) that had relational measures between all  $REs$  was constructed. Similarity between two  $REs$  in this algorithm was quantified by using normalized mutual information or  $NMI$ . (3) In the third step,  $REs$  that were found to be highly reproducible across subjects and ICA realizations were extracted and aligned. Two related  $REs$  were considered as individual-level components with the same underlying group-level component or an aligned component (AC). For each AC, the algorithm generated a  $dn \times dn$  reproducibility matrix,  $M_R$ , within which  $NMIs$  between all pairs of  $REs$  pertaining to the AC were collected. (4) In the fourth step, we aligned ACs to obtain group-level component maps and examined the inter-subject consistency. While **Figure 2B** illustrates the algorithm in general terms, we demonstrate the specific implementation of this algorithm for an example of three subjects in **Figure 3**.



We applied the gRAICAR algorithm separately to Autism, Control and Combined groups. The first step involved performing ICA decomposition  $d$  (~5,000 for this study) times for each subject using random initial values leading to  $d \times n$  realizations where  $n$  is the number of subjects. Specifically,  $d \times n$  or  $d \times 392$  realizations of ICs for Autism group,  $d \times 407$  realizations in the TC group and  $d \times 799$  realizations in the combined group were obtained. These ICA realizations are named  $REs$  and  $RE_{ij}$  is used to denote the set of ICs from the  $j$ th realization of the  $i$ th subject.

In the second stage of gRAICAR, we constructed a full similarity matrix (**FSM**). This matrix has relational measures between all  $REs$ . Block structure of the **FSM** represents subject blocks (**SBs**) that in turn represent subject-wise relationships. Elements within these blocks can determine similarity between  $REs$  from the same subject or pairs of  $REs$  from different subjects depending on the location of the block. In these **SBs**, there are  $d \times d$  realization blocks (**RBs**) providing pair-wise similarity between  $REs$  from different ICA realizations. This similarity between two **RBs** was quantified by using normalized mutual information or **NMI** (Pluim et al., 2003). **NMI** is one if two variables are identical and zero if they are statistically independent, revealing higher order statistical similarity as opposed to second order similarity expressed by correlation or covariance (Maes et al., 1997). **NMI** between each IC pair were computed using mutual information using the algorithm proposed by Kraskov et al. (2004). **NMIs** were further standardized within an **RB**, resulting in standardized **NMI** (**SNMI**). In order to demonstrate the technical underpinnings and rationale for this stage, **Figure 4** presents the block structure of **FSM** with 3 artificial subjects and two ICA realizations each (i.e.,  $d = 2$ ,  $n = 3$ ).

Blocks marked with solid lines are called Subject blocks or **SBs**. Off diagonal **SBs** indicate the similarity between pairs of  $REs$  from different subjects while the ones on the diagonal reflect that from the same subject. The **RB** is represented as a  $c_i \times c_m$  matrix,  $M_{RB}$ ,

with  $R_{ij-mk}$  reflecting the similarity between  $RE_{ij}$  and  $RE_{mk}$  ( $i, m: 1, 2, \dots, n, j, k: 1, 2, \dots, d$ ). **NMI** as mentioned above for two  $REs$  can now be calculated as:

$$R_{ij-mk}[y, z] = NMI(RE_y \in RE_{ij}, RE_z \in RE_{mk}) = \frac{H(RE_y) + H(RE_z)}{H(RE_y, RE_z)} - 1 \quad (1)$$

where  $H(RE_y)$ ,  $H(RE_z)$ ,  $H(RE_y, RE_z)$  represent entropies of random variables,  $RE_y$ , and  $RE_z$ , and the mutual entropy between them ( $1 \leq y \leq C_i, 1 \leq z \leq C_m$ ). The **NMIs** were further normalized in the alignment procedure using,

$$\mathcal{R}_{ij-mk}[y, z] = \frac{R_{ij-mk}[y, z] - \text{mean}(R_{ij-mk}[y, *] \cup R_{ij-mk}[* , z])}{\text{std}(R_{ij-mk}[y, *] \cup R_{ij-mk}[* , z])} \quad (2)$$

In Equation (2),  $*$  represents all **NMI** values in row  $y$  or column  $z$  of the **RB**. This standardized **NMI** or **SNMI** can be used to calculate the specificity of individual similarity values associated with a given  $RE$  within the **RB**. The diagonal **RBs** are normally set to zero since they represent identity matrices and are therefore not of interest.

We then extracted highly reproducible  $REs$  across subjects and ICA realizations and aligned them in the third stage of the gRAICAR algorithm. In order to do so, the algorithm searched all **SNMI** entries within **SBs** reflecting the similarity between pairs of  $REs$  from different subjects to determine a global maximum. Two  $REs$  that were found to be related were seen as individual-level components but with the same underlying group-level component, also known as an aligned component (**AC**). These **RBs** were then searched to locate the local maxima within them as they indicate possible locations of the aligned component

	SUBJECT 1	SUBJECT 2	SUBJECT 3
SUBJECT 1	RE 11-11 RE 11 - 12 RE 12-11 RE 12 - 12	RE 11-21 RE 11 - 22 RE 12-21 RE 12 - 22	RE 11-31 RE 11 - 32 RE 12-31 RE 12 - 32
SUBJECT 2	RE 21-11 RE 21 - 12 RE 22-11 RE 22 - 12	RE 21-21 RE 21 - 22 RE 22-21 RE 22 - 22	RE 21-31 RE 21 - 32 RE 22-31 RE 22 - 32
SUBJECT 3	RE 31-11 RE 31 - 12 RE 32-11 RE 32 - 12	RE 31-21 RE 31 - 22 RE 32-21 RE 32 - 22	RE 31-31 RE 31 - 32 RE 32-31 RE 32 - 32

**FIGURE 4 |** Block structure of full similarity matrix (**FSM**) based upon our example with 3 subjects and 2 sets of subject-level components (RE, a set of ICs produced after two ICA decompositions). In this example, each block shows similarity between 2 pairs of REs from the same subject or two different subjects. Also, RE 11-31 indicates inter-subject similarity between RE 1 of subject 1 and RE 1 of subject 3. The first digit in 11 is an index to the subject whereas the second is an index to the RE. Solid colored, on the diagonal blocks indicate similarity within the components of a subject or intra-subject similarity whereas off diagonal ones indicate that between the components of two different subjects.

in different ICA realizations and subjects. Rows and columns containing these maxima were eliminated from the **FSM** when all **RCs** associated with the aligned component were located. This procedure was repeated until  $c_{max} = \max(c \mid 1 \leq f \leq n)$  **ACs** had been discovered, where  $c$  is the number of ICs and  $f$  ( $\leq n$ ) is the number of subjects.

A  $dn \times dn$  reproducibility matrix,  $M_{rep}$ , for each **AC** was then generated by collecting **NMIs** between all pairs of **REs** related to the **AC**. **NMIs** were used to provide a more straightforward interpretation of similarity. A maximum of one **RC** was selected per **AC** in each ICA realization to form its reproducibility matrix. Information contained within the reproducibility matrix was then divided into two metrics: inter-subject consistency and intra-subject reliability. Inter-subject consistency in this case was defined as the mean of all **NMIs** within inter-subject blocks. For a given **AC**, its consistency between subjects  $i$  and  $m$  can be calculated as:

$$\alpha_{im} = \frac{\text{mean}(\mathcal{R}_{i'-m'})}{\frac{\sum_{j=1}^D \sum_{k=1}^D \mathcal{R}_{ij-mk}[y(i,j), z(m,k)]}{K^2}}, \quad 1 \leq i, m \leq N, i \neq m \quad (3)$$

Equation (3) is representative of the mean **NMI** within the inter-subject block “ $i-m$ ” in the reproducibility matrix, as it averages all the **NMI** values located at the intersection between realization  $j$  of subject  $i$  and realization  $k$  of subject  $m$ .

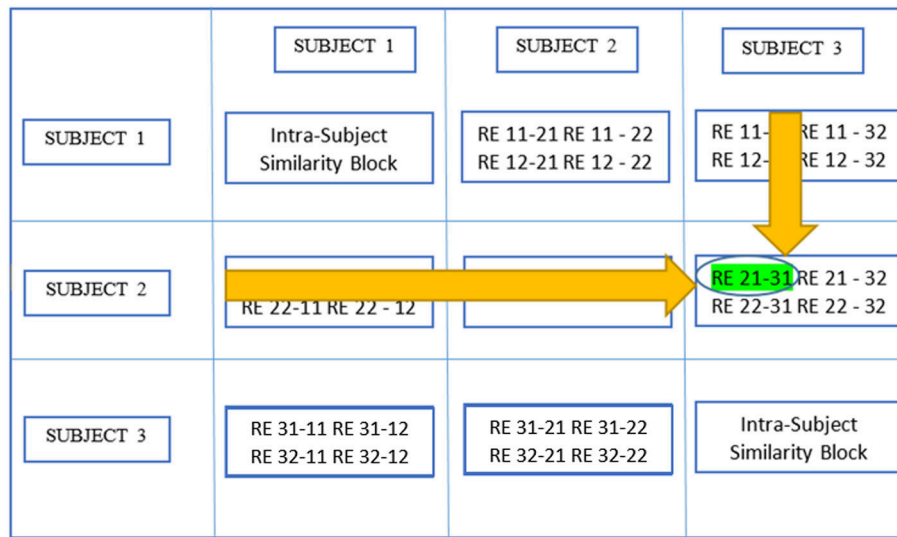
**Figure 5**, which is a continuation of **Figure 4**, provides a demonstration of the third stage summarizing higher level stage

description earlier using the same scenario as in the previous figure. The large circle mark represents a global maximum which is calculated by searching all **SNMIs** within the off-diagonal **SBs**. This enables compatibility with larger variations across subjects than within subjects. In this case, the global maximum was located at  $R_{ij-mk}[y, z]$  which represents the  $y$ th row and the  $z$ th column of the **RB**. The two related **REs**,  $RE_{y(i,j)}$  and  $RE_{z(m,k)}$ , are treated as individual-level components with the same underlying group-level component or **AC**.  $RE_{y(i,j)}$  represents the  $y$ th component of the  $j$ th realization of the  $i$ th subject.

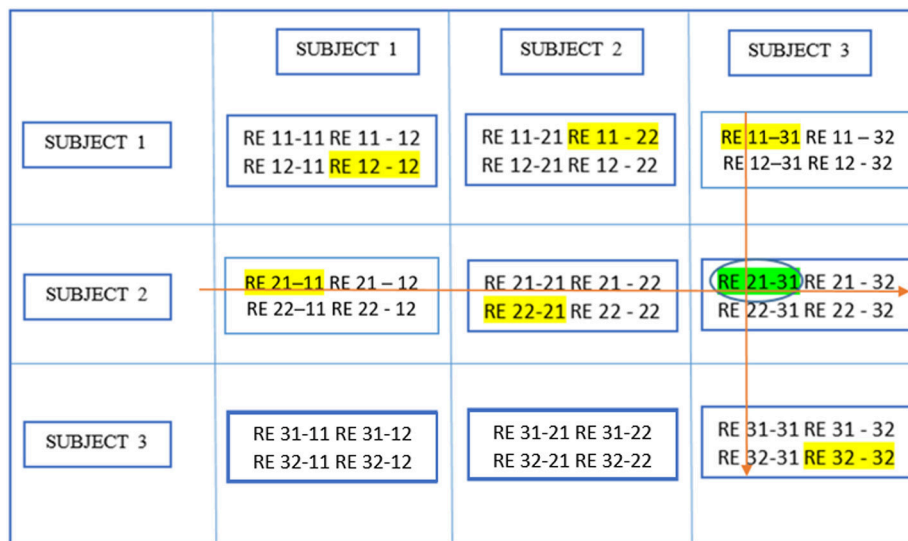
**Figure 6** demonstrates the next step which is to locate local maxima within these **RBs** by searching the  $y$ th rows of **RBs**  $\mathcal{R}_{ij-..}$  or all **RBs** containing  $RE_{y(i,j)}$ , and the  $z$ th columns of **RBs**  $\mathcal{R}_{..-mk}$  or all **RBs** containing  $RE_{z(m,k)}$ . This leads to the identification of the aligned component in different ICA realizations and subjects,  $[y, v_1]$  or **RE 21-11** in this example and  $[u_1, z]$  or **RE 11-31** where  $u_1$  and  $v_1$  are the relevant **RE** positions in individual **RBs** reflecting the largest similarity with  $RE_{y(i,j)}$  and  $RE_{z(m,k)}$  respectively. In this case,  $u_1 = v_1$  means  $[y, v_1]$  and  $[u_1, z]$  pick up the same **RE** and the resulting component is thought of as pertaining to the aligned component determined by  $RE_{y(i,j)}$  and  $RE_{z(m,k)}$ . If  $u_1$  and  $v_1$  are not equal, either  $u_1$  or  $v_1$  is picked based upon a voting procedure to determine the proximity of one or the other to more of those **REs** probed as the  $u_1 = v_1$  case.

In the fourth and final stage of **gRAICAR**, we estimated **AC** maps and corresponding mixing time courses by using weighted averages of their related **REs**. To compute the weighted average of the **REs**, the first step is to define a subject load





**FIGURE 5** | Determine global maximum. In this figure, the global maximum has a circle around it with highlighting.



**FIGURE 6** | Determine block-wise maxima with highlighting in each block. Global maximum in this figure has a circle around it in addition and has been shown in an earlier figure as well.

on inter-subject consistency representing the contribution or inter-subject centrality of a given subject to a given AC as follows:

$$\tau_i = \frac{1}{N-1} \sum_{m=1, m \neq i}^N \alpha_{im}, 1 \leq i \leq N \quad (4)$$

In (4),  $\alpha_{im}$  refers to inter-subject consistency metric between subjects  $i$  and  $m$ . This equation can also be phrased as the inter-subject centrality of a subject in a given AC is the mean of the inter-subject consistency metrics between this subject and all others. The spatial maps and mixing time courses of an AC can

be computed by combining this subject load on inter-subject consistency and the intra-subject reliability, as follows:

$$gIC_n = \frac{\sum_{i=1}^N [\beta_i \tau_i \sum_{j=1}^D RE_{p(i,j,n)}]}{K \sum_{i=1}^N \beta_i \tau_i}, 1 \leq n \leq c_{max} \quad (5)$$

$RE_{p(i,j,n)}$  represents the RE or the spatial map of the IC as identified in the  $j$ th realization of the  $i$ th subject.  $p$  indexes the location of the REs and can vary with different realizations and

subjects. The weights are different for each AC, computed by the AC specific reproducibility matrix.

ACs that were consistent across subjects were then statistically detected. The significance of cross-subject consistency of the resulting AC was explored using a two-step methodology. A non-parametric test was applied to select the AC consistent across all subjects. One RE from each ICA realization in the FSM was randomly sampled with replacement and the mean of inter-subject consistency metrics was computed after a non-participating subject was artificially generated. The aforementioned approach was very similar to the enhancement to the original RAICAR algorithm proposed by Yang et al. (2008). The aforementioned procedure was repeated 500 times. Resultant means of the inter-subject consistency metrics were combined to produce a null-distribution of inter-subject consistency. The 95th percentile, corresponding to a significance of  $p = 0.05$ , of the null distribution provided a threshold at this point. ACs with mean inter-subject consistency metrics greater than the aforementioned threshold value were regarded as common ACs across subjects. Null distributions of the subject loads on inter-subject consistency and intra-subject reliability for each one of the aforementioned ACs were generated by randomly assigning REs with replacement in the reproducibility matrix to artificially generated subjects. Corresponding to a significance value of  $p = 0.05$ , thresholds for the aforementioned metrics were then determined at 95th percentile of the corresponding null distributions. At this point, subjects above both of these threshold levels were considered to be representative of the AC under consideration. The main tasks pertaining to the fourth stage of gRAICAR algorithm were to estimate AC maps and corresponding time courses after weighted averaging their related REs, statistically detect ACs that were consistent across all subjects, and construct a graph for each AC for the characterization of relationships among subjects from an inter-subject consistency perspective.

## Clustering Analysis

K-means algorithm has been previously used in fMRI analysis in several studies (Liu et al., 2012; Zhang and Li, 2012; Allen et al., 2014). We used this algorithm to examine the level of separation between Autism and TC groups for the ICs which were reproducible within each group, but not reproducible in the combined group. Clustering was unsupervised without using *a priori* subject groupings. We determined cluster purity per cluster as shown below:

$$Purity = \frac{1}{N} \sum_{i=1}^k \max_j |c_i \cap t_j| \quad (6)$$

In (6),  $N$  represents the number of data points or subjects,  $k$  the number of clusters,  $c_i$  the cluster in our analysis, and  $t_j$  the classification with maximum count for cluster  $c_i$ .

Equation (7) shows our approach to determine sensitivity values where  $SEN$  represents sensitivity,  $TP$  true positive, and  $FN$  true negative.

$$SEN = \frac{TP}{TP + FN} \quad (7)$$

Equation (8) shows our approach to determine specificity values where  $SPC$  represents specificity,  $TN$  true negative, and  $FP$  false positive.

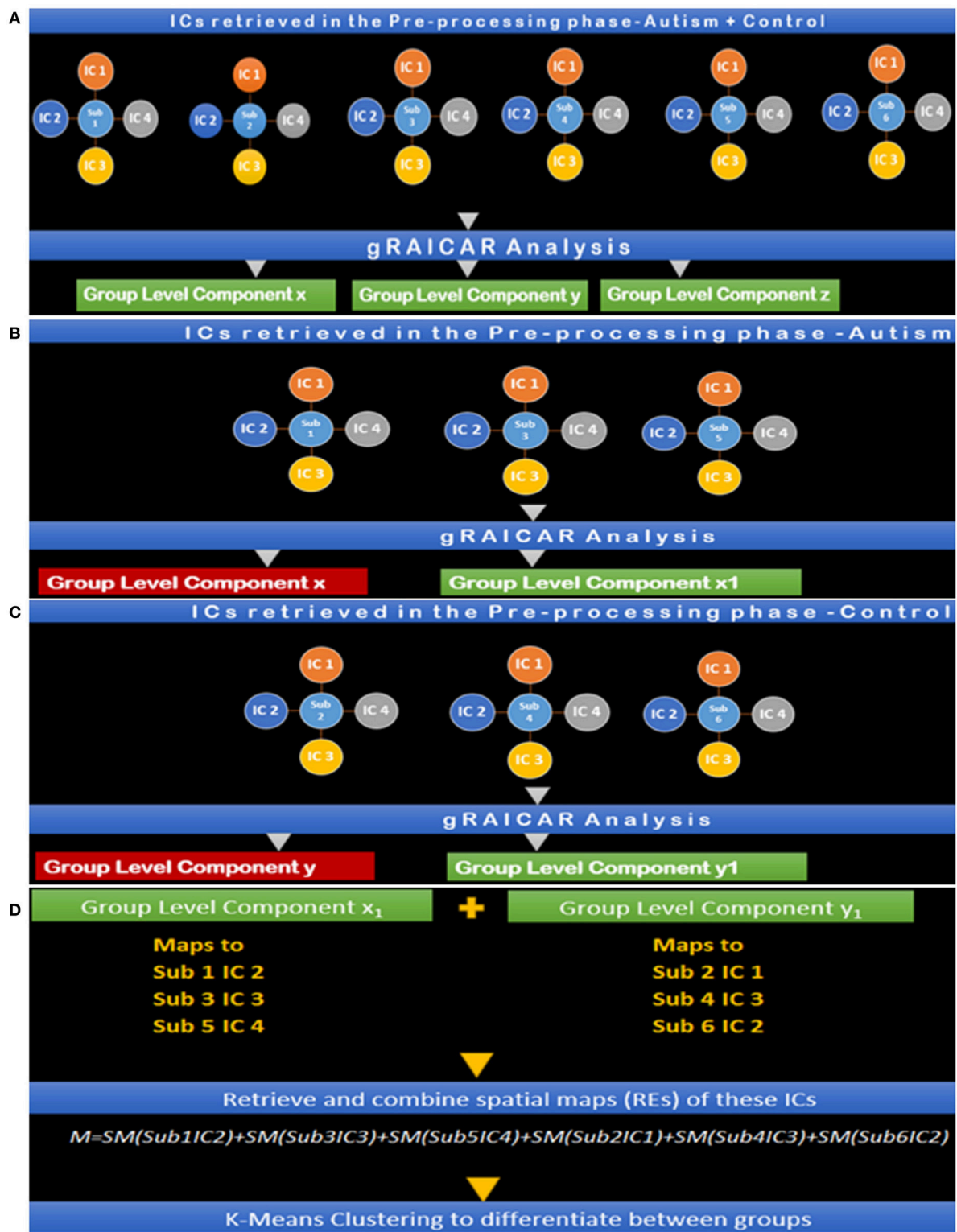
$$SPC = \frac{TN}{TN + FP} \quad (8)$$

## Analysis Workflow

This section presents our implementation and workflow. For technical details and the rationale behind every step, we have included a technical discussion in earlier sections of this paper. We applied the gRAICAR algorithm thrice: first on the Autism group, second on the Control group, and then on the combined group. For the Autism and Control groups, we had 54 and 49 group-level components, respectively. For the combined (Autism + Control) group, we had 54 group-level components. We then examined these group-level components using criteria presented above describing the steps of gRAICAR algorithm and inter-subject consistency in (3). These criteria gave us 11 group-level components in the Autism group and 3 in the Control group. For all subjects, we accessed post-MELODIC analysis results and retrieved spatial maps associated with the ICs corresponding to each selected group-level component. MELODIC analysis was a part of data pre-processing and described earlier in this paper. We then processed these spatial maps in MATLAB wherein the spatial map associated with the IC index of the current subject was retrieved and singleton dimensions were removed. The resulting array was reshaped using MATLAB's reshape function (<http://www.mathworks.com/help/matlab/ref/reshape.html>), thus giving us an  $m \times n$  matrix where  $m$  is 1 for the current subject and  $n$  is  $61 \times 73 \times 61$  ( $=271,633$ ) which was the size of each spatial map associated with the current IC index. After all subjects were processed, we had a  $392 \times 271,633$  matrix for the Autism group and  $407 \times 271,633$  matrix for the Control group. Suppose the resulting matrix for Autism is  $A$  while that for TC is  $C$ . We then combined  $A$  and  $C$  giving us a  $799 \times 271,633$  matrix. We applied the k-means algorithm using this matrix to examine how subjects were clustered based on their spatial maps without *a priori* groupings. The aforementioned process was repeated for all permutations of group-level components selected based on pairing a component from the Autism group with one from the Control group resulting in 33 k-means clustering analysis (Autism Group:  $11 \times$  Control Group: 3). We had set up the algorithm to partition the data set into two clusters since we had two subject groups, Autism and Control. For each of these clustering permutations, the purity of clusters was identified based on how many subjects were correctly (or wrongly) clustered along with other subjects with the same diagnosis.

From the above analysis, the pair with maximum cluster purity was identified. Let the corresponding components be  $A^x$  and  $C^x$  in Autism and Control groups, respectively. The component in Controls with maximum spatial correlation with  $A^x$  (say,  $C_a^x$ ) and the component in Autism with the maximum spatial correlation with  $C^x$  (say,  $A_c^x$ ), were identified using the following approach:

$$\lambda = \max \sum_{i=1}^k cov(\omega, \omega_i) \quad (9)$$



**FIGURE 7 | (A)** Step I: gRAICAR Analysis on Combined (Autism + Control) Group producing group level components x, y, and z. **(B)** Step II: gRAICAR Analysis on Autism Group Only, producing group-level components, x and x<sub>1</sub>. x is discarded since it was produced in step I. **(C)** Step III: gRAICAR Analysis on Control Group Only, producing group-level components, y and y<sub>1</sub>. y is discarded since it was produced in step I. **(D)** Step IV: Group level components from Steps II and III were combined by retrieving spatial maps corresponding to ICs these group-level components represented for each subject within each group as shown in the figure. Group level components reproducible in the combined group also found in individual analysis (steps II and III) were excluded.

$\lambda$  represents maximum covariance in (9), between group level component,  $\omega$ , from the group being analyzed and  $\omega_i$ , that from the opposite group with  $k$  being the total number of group level components in the opposite group.

Two more k-means clustering analysis were performed by pairing  $A^x$  with  $C_a^x$ , and  $C^x$  with  $A_c^x$ . This analysis was carried out to ascertain whether the reproducible components in each group, when paired with the corresponding component with similar spatial distribution in the other group, can effectively discriminate between the groups. The entire analysis pipeline is illustrated in **Figures 7A–D**.

Steps I–IV presented in **Figure 7** illustrate the concepts and summarize the processing by gRAICAR algorithm and k-means clustering analysis. For demonstration purposes, 6 artificial subjects, 3 from the Autism group (denoted by A) and the other 3 from Control group (denoted by C) are shown. Each example subject is assigned 4 ICs as shown. This is an arbitrary number illustrating the concept and the number of ICs was not constant in actual processing. In actual processing, 799 subjects were used with a variable number of ICs. (I) This step shows gRAICAR processing on all subjects in the combined group (Autism + Control groups) and the resulting list of group-level components, 3 in this case:  $x$ ,  $y$ , and  $z$ . (II) This step shows gRAICAR analysis on Autism group only from the example and the list of group-level components obtained as in step I. Component  $x$  was found in step I as well and is discarded after visual examination. (III) This step shows gRAICAR analysis on Control group only and the list of group-level components obtained as in steps I and II. Component  $y$  was found in step I hence discarded. (IV) In this step, we completed multiple tasks. We combined the group-level components  $x_1$  and  $y_1$  by mapping these to individual ICs for each subject. We then retrieved spatial maps for each IC representing a subject under the group-level component and linearly combined them using MATLAB creating a matrix we called “ $M$ .” Finally, we used k-means clustering algorithm in MATLAB using  $M$  to investigate the separation of components between groups.

Once the clustering was complete, we constructed an inter-subject Euclidean distance matrix within both Autism and Control groups using spatial maps associated with each subject for component pairings ( $C^x, A^x$ ), ( $A_c^x, C^x$ ), and ( $C_a^x, A^x$ ). A self-organizing map or SOM analyzes input vectors in the input space and learns, in an unsupervised manner, to classify them accordingly (Kohonen, 1988, 2001). The result includes a low-dimensional (one- or two-) discretized representation of the input space of the training samples referred to as a map.

Neighboring neurons in SOMs learn recognizing neighboring sections of the input space which leads them to not only learn the distribution but the topology of the training vectors used as input. These neurons are arranged in physical positions based upon a topology function and distances between them are calculated using a distance function.

Adjacent neurons in the topology generally are close in the input space as well. In our study, we used SOMs to visualize the reproducibility and separation of the subjects in feature-space in addition to the numerical values given by k-means. High dimensionality in k-means was scaled using SOMs for optimal visualization. We obtained individual spatial maps for  $A^x$  and  $C^x$

and stacked them into a matrix. We then used this matrix as input to a  $5 \times 5$  SOM for visualization as described earlier. This process was repeated for ( $A^x, C_a^x$ ) and ( $A_c^x, C^x$ ).

## RESULTS

Let us first examine the most reproducible group-level components within each group. We found 54 group-level components within the Autism group and 49 such components within the Control group. By combining selected components as described earlier, the range of cluster purity was 0.69–0.97 using unsupervised k-means clustering over all permutations of 11 group-level components from Autism and 3 from Controls. The average purity value was 0.89 with a standard deviation of 0.06.

**Figures 8, 9** show the spatial maps of  $A^x$  and  $C^x$ , respectively. The highest cluster purity value was 0.97 obtained by combining these two group-level components. **Figure 10** presents a map of pie charts based on a  $5 \times 5$  SOM to visualize the reproducibility and separation of the two groups using  $A^x$  and  $C^x$  as described earlier. Each pie chart represents the number of subjects from a given group, Autism or Control. As an example, a solid red chart represents all subjects from the Autism group whereas a solid blue all from the Control group.

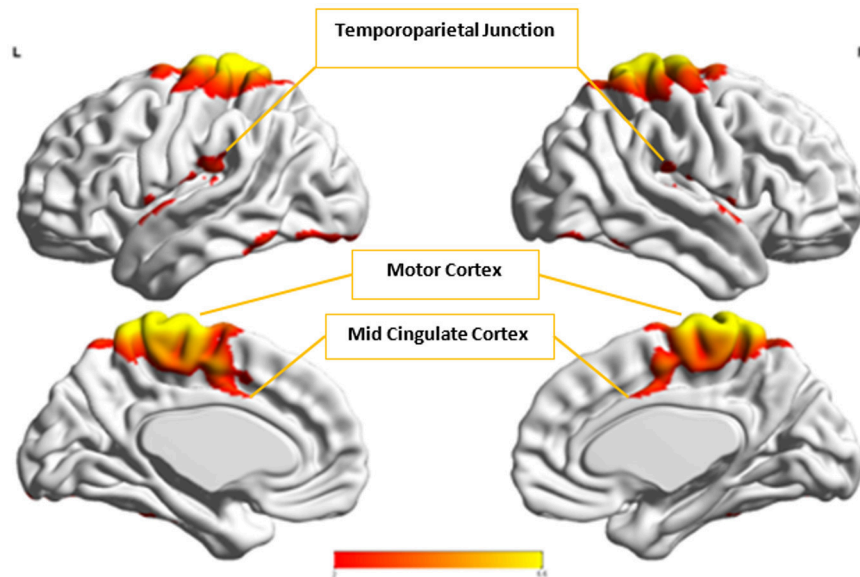
In the next step, we ascertained which group-level components in the opposite group had the highest spatial correlation to  $A^x$  and  $C^x$ , using all 54 components from Autism and 49 from Control groups depending upon the comparison being carried out.  $A^x$  was found to have the highest spatial correlation value of 0.29 ( $p < 0.001$ ) with  $C_a^x$  in the Control group while  $C^x$  had the highest spatial correlation value of 0.63 ( $p < 0.001$ ) with  $A_c^x$  in the Autism group. We then combined  $C^x$  with  $A_c^x$  and subjected them to k-means clustering analysis. This produced cluster purity of 0.895 with a sensitivity of 0.893 and a specificity of 0.897. Similarly, we combined  $A^x$  with  $C_a^x$  as described earlier and completed k-means clustering analysis. This resulted in a cluster purity of 0.607 with a sensitivity of 0.43 and specificity of 0.77. **Figures 11, 12** show the spatial profiles of  $A_c^x$  and  $C_a^x$ , respectively.

We also used pie charts to visualize the reproducibility and separation of subjects using  $5 \times 5$  SOMs for these combinations:  $A^x + C_a^x$  and  $C^x + A_c^x$ . These visualizations are presented in **Figures 13, 14**. In both cases, a dotted line represents the approximate separation observed between the two groups. Numbers on each pie chart represent the neuron in the SOM. It can be observed that the purity of individual pie charts drops when using spatial equivalents (**Figures 13, 14**) as compared to the most reproducible components in each group (**Figure 10**). This is a reflection of higher purity and separation in **Figure 10** (97.1%) to lower purity values and hence lower separation in **Figures 13, 14** (0.607 and 0.895 respectively).

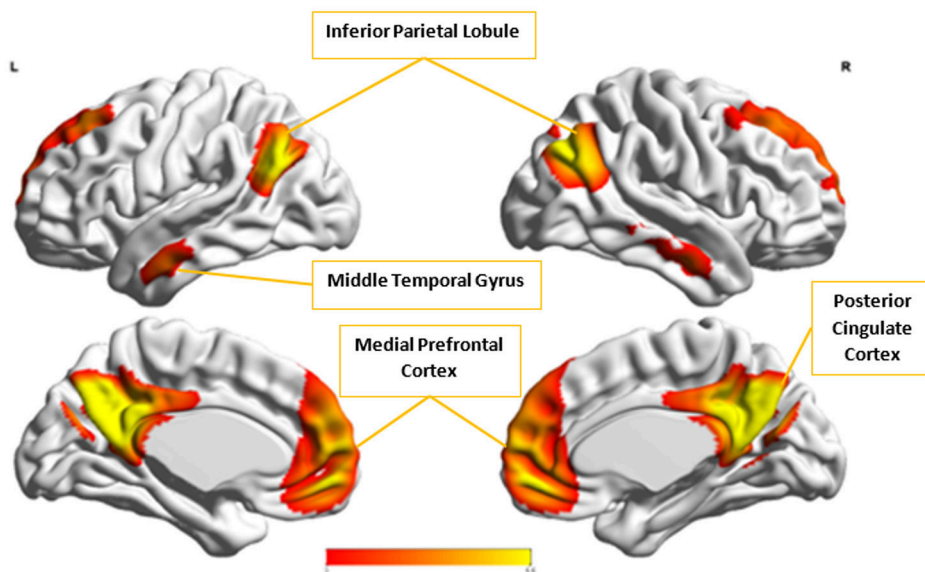
## DISCUSSION

We used a discover-confirm scheme wherein during the “discover” phase, we used gRAICAR to retrieve reproducible components in each group and during the “confirm” phase, we used unsupervised clustering to determine the separation





**FIGURE 8 |** The group-level component,  $A^X$ , from Autism group that produced the highest cluster purity value of 0.971 when combined with another group-level component from the Control group,  $C^X$ .

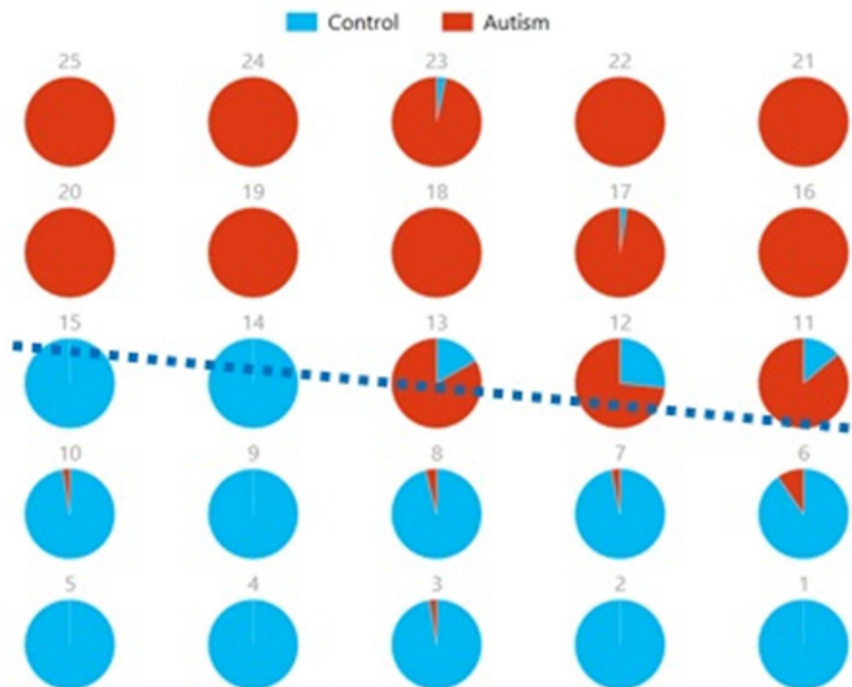


**FIGURE 9 |** The group-level component,  $C^X$ , from Control group that produced the highest cluster purity value of 0.971 when combined with another group-level component,  $A^X$ , from the Autism group. These regions represent the default mode network (DMN).

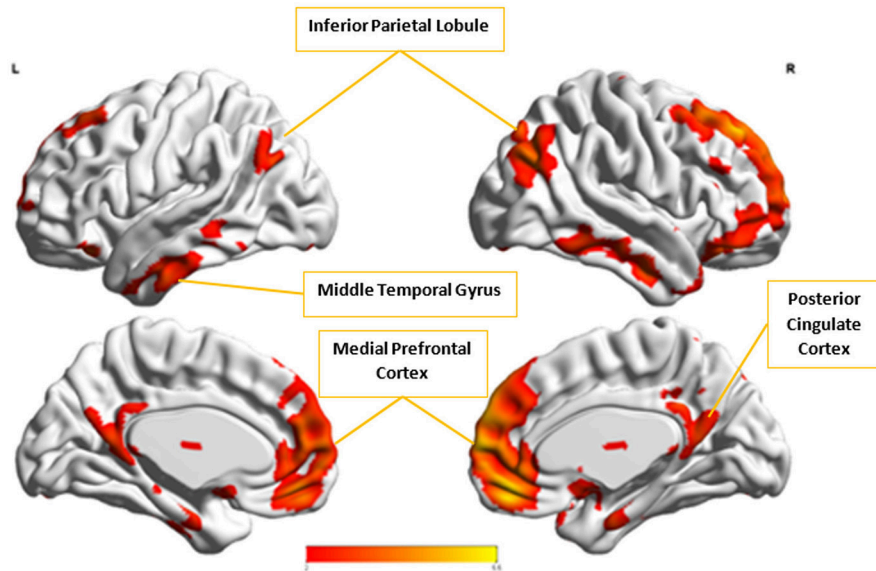
between groups based on the reproducible components in each group. Further, the separation was visualized using self-organizing maps or SOMs. This is a novel methodological framework for investigating discriminative features between diagnostic groups as opposed to performing group-wise statistical tests or supervised classification.

Even though multiple studies have shown altered fMRI-based connectivity in certain brain networks in Autism using machine learning techniques, identifying individuals with Autism based

on these measures has not been reliable especially in larger sized samples (Anderson et al., 2011; Plitt et al., 2015). We hypothesized that functional brain networks which are most reproducible separately within Autism and Control groups, but not reproducible when analyzing both groups as merged, may lead to effective discrimination between the groups. We tested the above hypothesis by finding the most reproducible ICA components (which represent brain networks) first in the merged and then in separate Autism and Control groups. Our



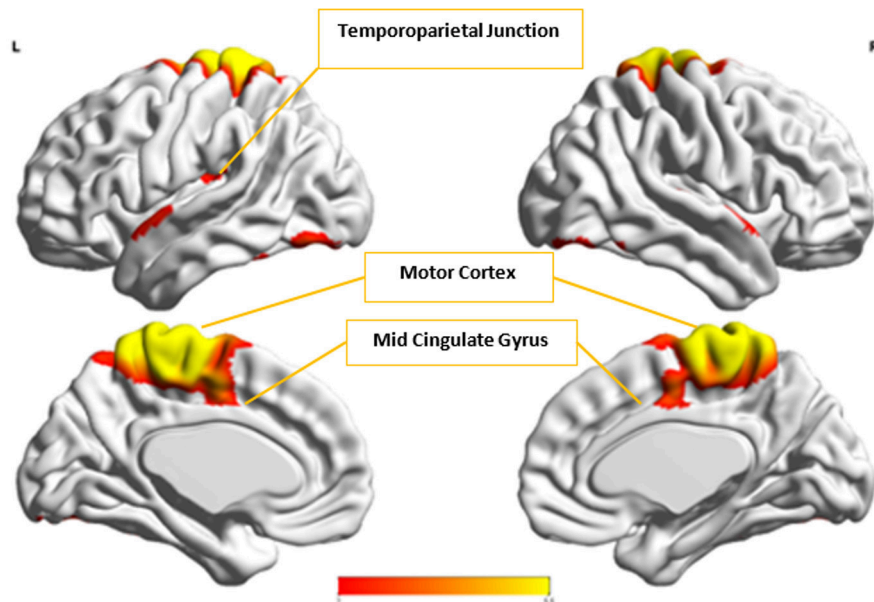
**FIGURE 10 |** Pie chart visualizations based on a  $5 \times 5$  SOM for  $A_C^X$  and  $C_C^X$  showing Autism and Control groups by neuron. Each pie chart corresponds to a neuron, represented by the number on each chart, in the SOM map for these components. This map indicates group separation approximately in the middle with Autism group populating the upper while Control the lower half of the SOM, represented by the dotted blue line.



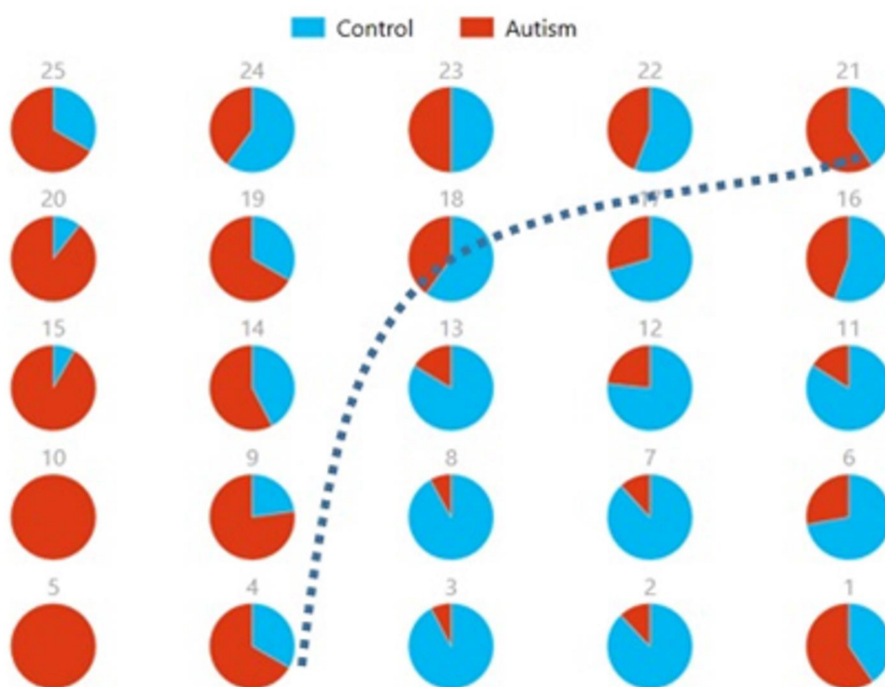
**FIGURE 11 |** This figure represents the spatial map for  $A_C^X$ , the group-level component in the Autism group with the highest spatial correlation with  $C_C^X$ . These regions represent the default mode network (DMN).

results, shown in the previous section, indeed support the above hypothesis. SOM visualizations provided along with spatial maps of the group-level components give further insight into the reproducibility of certain brain networks as well as their differences between groups based on our proposition.

The overall cluster purity we obtained from our multisite fMRI data set, obtained by averaging the results obtained from the three scenarios was 0.824 with a sensitivity of 0.77 and specificity of 0.87. Previous studies using the same data set, but supervised classification methods instead of unsupervised



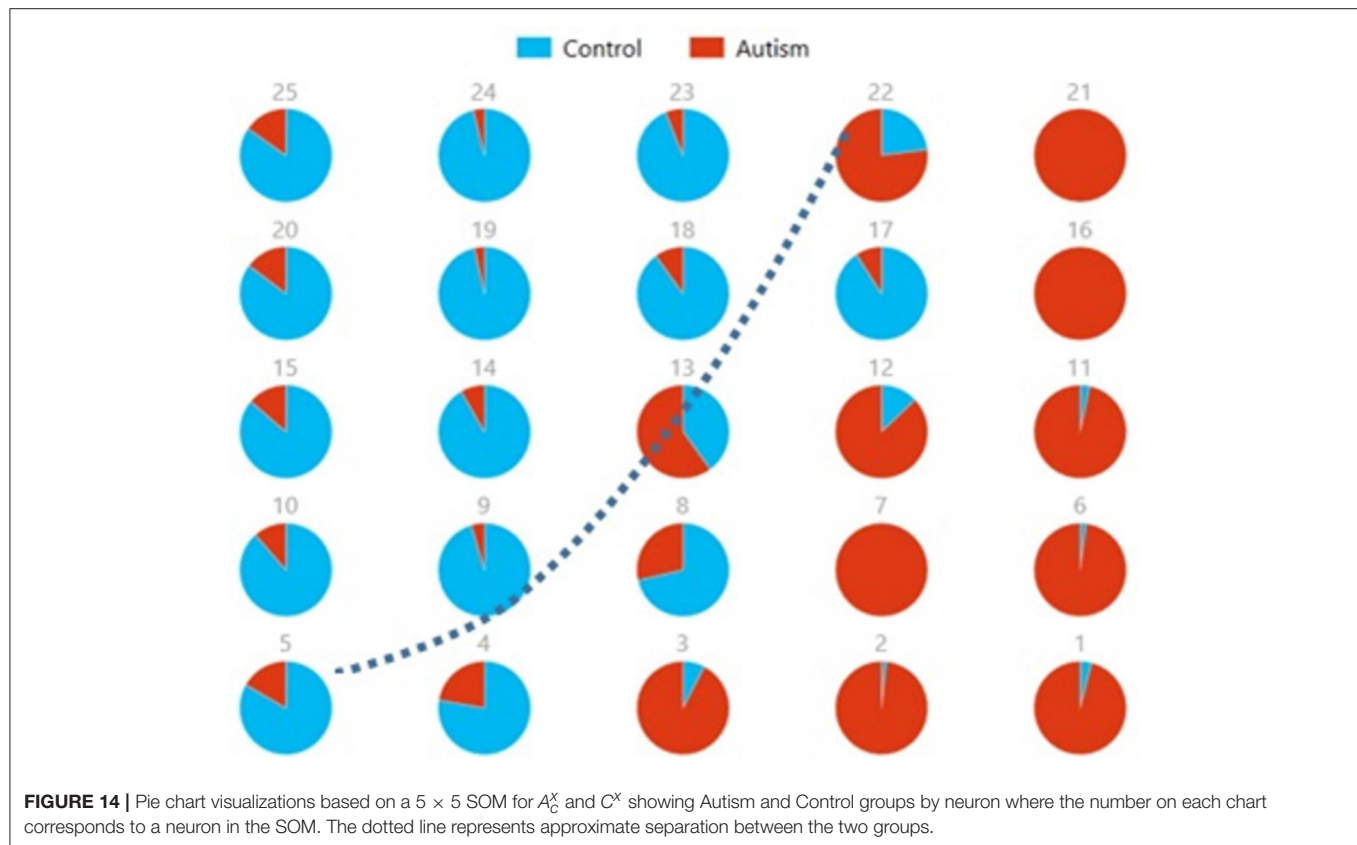
**FIGURE 12 |** This figure represents the spatial map for  $C_a^X$ , the group-level component in the Control group with the highest spatial correlation with  $A^X$ .



**FIGURE 13 |** Pie chart visualizations based on a  $5 \times 5$  SOM for  $A^X$  and  $C_a^X$  showing Autism and Control groups by neuron where the number on each chart corresponds to a neuron in the SOM. The dotted line represents approximate separation between the two groups.

clustering methods, have reported classification accuracies between 0.6 and 0.8 depending on whether they used a larger or smaller sub-sample of the ABIDE database (Anderson et al., 2011; Nielsen et al., 2013). Given the fact that the methods used here are

different from the previous studies mentioned above, it would not be fair to directly compare our cluster purity with theirs. Instead, we would like to make the point that characterizing reproducibility of brain networks in different groups as well as the



**FIGURE 14 |** Pie chart visualizations based on a  $5 \times 5$  SOM for  $A_c^x$  and  $C_a^x$  showing Autism and Control groups by neuron where the number on each chart corresponds to a neuron in the SOM. The dotted line represents approximate separation between the two groups.

merged sample is a novel idea which may hold promise, especially in the context of disorders such as Autism. This is because the most discriminative features identified via the proposed method are more likely to be generalizable to a larger sample given the reproducibility constraint.

$C_x$  and  $A_c^x$ , which provided highest discriminability between the groups, represent the default mode network (DMN) in Control and Autism groups, respectively. The DMN in Autism appears less prominent and incohesive. Decreased functional connectivity in default mode subnetworks contributes to core deficits observed in ASD patients (Assaf et al., 2010) whereas activity was reduced in the autism group in the ventral medial prefrontal cortex/ventral anterior cingulate cortex (Kennedy and Courchesne, 2008). Visuospatial working memory deficiency within the DMN was discovered in adolescents with ASD (Chien et al., 2016) and the regions of DMN functional connectivity in the bilateral inferior parietal lobule and posterior cingulate cortex were found to be smaller in ASD patients (Yasuhiro et al., 2016). On the other hand,  $A^x$  and  $C_a^x$  represent regions of the motor network, mid cingulate cortex and temporal-parietal junction. Even though these regions have been implicated in autism (Chiu et al., 2008; Chantiluke et al., 2014; Kestemont et al., 2014; Nebel et al., 2014), it was not as discriminatory as the DMN. To summarize, our methodology first discovered highly reproducible components separately in Autism and Control groups pointing to functional networks described in this section. These components or functional networks they pointed to

from both groups, when combined and analyzed in clustering analysis as described, provided high cluster purities, hence the ability to distinguish between the two groups. Functional networks discovered by applying our methodology separately in groups confirm earlier findings on alterations involving these networks in Autism. Results obtained from analyzing these networks support our hypothesis that functional networks highly reproducible separately in groups lead to higher cluster purities and discriminability.

## Limitations and Future Directions

Despite the fact that the ABIDE database provides invaluable means to analyze multisite resting state fMRI data sets with significant statistical power, there are certain inherent limitations to this data set. Site to site variability in acquisition parameters, subject populations, scanner performance, and research protocols may all be confounding factors when it comes to the sensitivity for detecting abnormalities (Nielsen et al., 2013). It could be argued that the analysis of individual site data sets separately may provide a higher cluster purity. However, such results may be less easily translatable to the clinic because inter-site variability is something any potential clinical method will have to cope with. Both groups in ABIDE, Autism and healthy Control, appeared to have subjects with average to above-average range of IQ in addition to variation in diagnostic subtypes (Asperger's and PDD-NOS) across sites. A broader range of IQ levels need to be included in further studies since R-fMRI studies



allow the inclusion of individuals with lower IQ than task based studies. In addition, not all sites spanned childhood to middle adulthood but further studies can include a deeper examination of the development of brain providing insight into developmental dynamics of Autism (Di Martino et al., 2014).

We used a novel analysis framework involving gRAICAR as described earlier (Yang et al., 2012). Despite its robustness, there are several limitations including computational and physical memory costs. We were able to mitigate computational and physical memory concerns by using parallel processing and cloud computing. gRAICAR further provides the ability to parallelize one of the processing stages hence reducing the computational time and increasing efficiency. We had used gRAICAR code in a UNIX/MATLAB environment. Also in the absence of a threshold in gRAICAR to determine the existence of a relationship, the RCs are forced to align with a group-level component even if there is low similarity with other RCs. In future studies, it would be interesting to investigate how gRAICAR performs in site-level analytics within Autism and ABIDE data sets. Our methodology can also be expanded to other neurological disorders to determine the utility of this algorithm in future studies.

## REFERENCES

- Allen, E., Damaraju, E., Plis, S., Erhardt, E., Eichele, T., and Calhoun, V. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, Arlington, VA: American Psychiatric Publishing.
- Anderson, J., Nielsen, J., Froehlich, A., DuBray, M., Druzgal, T., Cariello, A., et al. (2011). Functional connectivity magnetic resonance imaging classification of Autism. *Brain* 134, 3742–3754. doi: 10.1093/brain/awr263
- Assaf, M., Jagannathan, K., Calhoun, V., Miller, L., Stevens, M., Sahl, R., et al. (2010). Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage* 53, 247–256. doi: 10.1016/j.neuroimage.2010.05.067
- Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 Sites, United States, 2010. *Morbid. Mort. Weekly Rep. (MMWR)* 63, 1–21. Available online at: <https://www.cdc.gov/mmwr/pdf/ss/ss6302.pdf>
- Beckman, C., DeLuca, M., Devlin, J., and Smith, S. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc.* 360, 1001–1013. doi: 10.1098/rstb.2005.1634
- Beckman, C., and Smith, S. (2004). Probabilistic independent component analysis. *IEEE Trans. Med. Imaging* 23, 137–152. doi: 10.1109/TMI.2003.822821
- Bell, A., and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129
- Calderoni, S., Retico, A., Biagi, L., Tancredi, R., Muratori, F., and Tosetti, M. (2012). Female children with autism spectrum disorder: an insight from mass-univariate and pattern classification analyses. *Neuroimage* 59, 1013–1022. doi: 10.1016/j.neuroimage.2011.08.070
- Chantiluke, K., Barrett, N., Giampietro, V., Santosh, P., Brammer, M., Simmons, A., et al. (2014). Inverse fluoxetine effects on inhibitory brain activation in non-comorbid boys with ADHD and with ASD. *Psychopharmacology* 232, 2071–2082. doi: 10.1007/s00213-014-3837-2
- Chien, H. Y., Gau, S., Isaac, S., and Tseng, W. Y. (2016). Deficient visuospatial working memory functions and neural correlates of the default-mode network

## AUTHOR CONTRIBUTIONS

MS: Main author responsible for data acquisition, preparation, methodology implementation, post-processing, visualizations, documenting results and paper writing work. ZY: Introduced and published the methodology implemented in the paper, research and editing consultant. XH: Methodology contributor, editing consultant. GD: Principal Investigator and project scientist, principal editor.

## ACKNOWLEDGMENTS

The authors would like to thank National Science Foundation—NSF (Grant # 0966278) for funding this study. ZY is funded by the National Science Foundation of China (81270023, 81571756, PI: ZY), Foundation of Beijing Key Laboratory of Mental Disorders (2014JSJB03, PI: ZY), Beijing Nova Program for Science and Technology (XXJH2015B079, PI: ZY), The Outstanding Young Investigator Award of Institute of Psychology, Chinese Academy of Sciences (Y4CX062008, PI: ZY) and Shanghai Key Laboratory of Psychotic Disorders (13dz2260500, start-up fund to ZY). XH is supported in part by NIH (DA033393).

- in adolescents with autism spectrum disorder. *Autism Res.* 9, 1058–1072. doi: 10.1002/aur.1607
- Chiu, P., Kayali, M., Kishida, K., Tomlin, D., Klinger, L., Klinger, M., et al. (2008). Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron* 57, 463–473. doi: 10.1016/j.neuron.2007.12.020
- Coutanche, M., Thompson-Schill, S., and Schultz, R. (2011). Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. *Neuroimage* 57, 113–123. doi: 10.1016/j.neuroimage.2011.04.016
- Deshpande, G., Libero, L., Sreenivasan, K., Deshpande, H., and Kana, R. (2013). Identification of neural connectivity signatures of autism using machine learning. *Front. Hum. Neurosci.* 7:670. doi: 10.3389/fnhum.2013.00670
- Di Martino, A., Yan, C., Li, Q., Denio, E., Castellanos, F., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78
- Ecker, C., Marquand, A., Mourão-Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., et al. (2010). Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *J. Neurosci.* 30, 10612–10623. doi: 10.1523/JNEUROSCI.5413-09.2010
- Greicius, M., Krasnow, B., Reiss, A., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100
- Jenkinson, M., Beckmann, C., Behrens, T., Woolrich, M., and Smith, S. (2012). the FMRIB Software Library (FSL). *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Jiao, Y., Chen, R., Ke, X., Chu, K., Lu, Z., and Herskovits, E. H. (2010). Predictive models of autism spectrum disorder based on brain regional cortical thickness. *Neuroimage* 50, 589–599. doi: 10.1016/j.neuroimage.2009.12.047
- Kennedy, D., and Courchesne, E. (2008). Functional abnormalities of the default network during self- and other-reflection in autism. *Soc. Cogn. Affect. Neurosci.* 3, 177–190. doi: 10.1093/scan/nsn011
- Kestemont, J., Ma, N., Baetens, K., Clément, N., Van Overwalle, F., and Vandekerckhove, M. (2014). Neural correlates of attributing causes to the self,

- another person and the situation. *Soc. Cogn. Affect. Neurosci.* 10, 114–121. doi: 10.1093/scan/nsu030
- Kohonen, T. (1988). *Self-Organization and Associative Memory* 8. Heidelberg: Springer-Verlag.
- Kohonen, T. (2001). *Self-Organizing Maps, 3rd Edn.* Heidelberg: Springer-Verlag.
- Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. Stat. Nonl. Soft Matter Phys.* 69:066138. doi: 10.1103/PhysRevE.69.066138
- Libero, L., DeRamus, T., Lahti, A., Deshpande, G., and Kana, R. (2015). Multimodal neuroimaging based classification of Autism Spectrum Disorder using anatomical, neurochemical and white matter correlates. *Cortex* 66, 46–59. doi: 10.1016/j.cortex.2015.02.008
- Liu, W., Awate, S., and Fletcher, P. (2012). “Group analysis of resting-state fMRI by hierarchical markov random fields,” in *Medical Image Computing and Computer-Assisted Intervention - Lecturer Notes in Computer Science* (Nice). doi: 10.1007/978-3-642-33454-2\_24
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Img* 16, 187–198. doi: 10.1109/42.563664
- Maximo, J., Keown, C., Nair, A., and Müller, R. (2013). Approaches to local connectivity in autism using resting state functional connectivity MRI. *Front. Hum. Neurosci.* 7:605. doi: 10.3389/fnhum.2013.00605
- Muhle, R., Trentacoste, S. V., and Rapin, I. (2004). The genetics of Autism. *Pediatrics* 113 e472–86. doi: 10.1542/peds.113.5.e472
- Nebel, M., Eloyan, A., Barber, A., and Mostofsky, S. (2014). Precentral gyrus functional connectivity signatures of autism. *Front. Syst. Neurosci.* 8: 80. doi: 10.3389/fnsys.2014.00080
- Nielsen, J., Zielinski, B., Fletcher, P., Alexander, A., Lange, N., Bigler, E., et al. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Front. Hum. Neurosci.* 7:599. doi: 10.3389/fnhum.2013.00599
- Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., et al. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 128, e488–e495. doi: 10.1542/peds.2010-2825
- Plitt, M., Kelly, A., and Martina, A. (2015). Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *Neuroimage Clin.* 7, 359–366. doi: 10.1016/j.nicl.2014.12.013
- Pluim, J., Maintz, J., and Viergever, M. (2003). Mutual-information-based registration of medical images: a survey. *EEE Trans. Med. Imaging* 22, 986–1004. doi: 10.1109/TMI.2003.815867
- Sato, J. R., Hoexter, M. Q., Oliveira, P. P., Brammer, M. J., MRC AIMS Consortium, Murphy, C., et al. (2013). Inter-regional cortical thickness correlations are associated with autistic symptoms: a machine-learning approach. *J. Psychiatr. Res.* 47, 453–459. doi: 10.1016/j.jpsychires.2012.11.017
- Song, X., Dong, Z., Long, X., Li, S., Zuo, X., Zhu, C., et al. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS ONE* 6:e25031. doi: 10.1371/journal.pone.0025031
- Supekar, K., Uddin, L., Khouzam, A., Phillips, J., Gaillard, W., Kenworthy, L., et al. (2013). Brain hyper-connectivity in children with autism and its links to social deficits. *Cell Reports* 5, 738–747. doi: 10.1016/j.celrep.2013.10.001
- Uddin, L. Q., Menon, V., Young, C. B., Ryali, S., Chen, T., Khouzam, A., et al. (2011). Multivariate searchlight classification of structural magnetic resonance imaging in children and adolescents with autism. *Biol. Psychiatry* 70, 833–841. doi: 10.1016/j.biopsych.2011.07.014
- von dem, H. E., Stoyanova, R., Baron-Cohen, S., and Calder, A. (2013). Reduced functional connectivity within and between ‘social’ resting state networks in autism spectrum conditions. *Soc. Cogn. Affect. Neurosci.* 8, 694–701. doi: 10.1093/scan/nss053
- Wang, H., Chen, C., and Fushing, H. (2012). Extracting multiscale pattern information of fMRI based functional brain connectivity with application on classification of autism spectrum disorders. *PLoS ONE* 7:e45502. doi: 10.1371/journal.pone.0045502
- Wingate, M., Mulvihill, B., Kirby, R. S., Pettygrove, S., Cunniff, C., Meaney, F., et al. (2012). Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *Morbid. Mort. Weekly Rep.* 61, 1–19. Available online at: <https://www.cdc.gov/mmwr/pdf/ss/ss6103.pdf>
- Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45(Suppl.), S173–S186. doi: 10.1016/j.neuroimage.2008.10.055
- Yan, C., and Zang, Y. (2010). DPARSF: A MATLAB Toolbox for “Pipeline” data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Yang, Z., Chang, C., Xu, T., Jiang, L., Handwerker, D., Castellanos, F., et al. (2014a). Connectivity trajectory across lifespan differentiates the precuneus from the default network. *NeuroImage* 89, 45–56. doi: 10.1016/j.neuroimage.2013.10.039
- Yang, Z., LaConte, S., Weng, X., and Hu, X. (2008). Ranking and averaging independent component analysis by reproducibility (RAICAR). *Hum. Brain Mapp.* 29, 711–725. doi: 10.1002/hbm.20432
- Yang, Z., Xu, Y., Xu, T., Hoy, C., Handwerker, D., Chen, G., et al. (2014b). Brain network informed subject community detection in early-onset schizophrenia. *Sci. Rep.* 4:5549. doi: 10.1038/srep05549
- Yang, Z., Zuo, X., Wang, P., Li, Z., LaConte, S., Bandettini, P., et al. (2012). Generalized RAICAR: discover homogeneous subject (sub)groups by reproducibility of their intrinsic connectivity networks. *Neuroimage* 63, 403–414. doi: 10.1016/j.neuroimage.2012.06.060
- Yasuhiro, F., Masafumi, H., Hideki, O., Kenji, M., Hiromichi, I., and Takashi, I. (2016). Default mode network abnormalities in children with autism spectrum disorder detected by resting-state functional magnetic resonance imaging. *J. Med. Invest.* 63, 204–208. doi: 10.2152/jmi.63.204
- Zhang, S., and Li, C. (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *NeuroImage* 59, 3548–3562. doi: 10.1016/j.neuroimage.2011.11.023

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Syed, Yang, Hu and Deshpande. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership