PLOS ONE

Test-Retest Reliability of Graph Metrics in Functional Brain Networks: A Resting-State fNIRS Study

Haijing Niu[1], Zhen Li[1], Xuhong Liao[2,3], Jinhui Wang[2], Tengda Zhao[1], Ni Shu[1], Xiaohu Zhao[4]*, Yong He[1]*

1 State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China, 2 Center for Cognition and Brain Disorders, Hangzhou Normal University, Hangzhou, China, 3 Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China, 4 Imaging Department, Shanghai TongJi Hospital, TongJi University, Shanghai, China

Abstract

Recent research has demonstrated the feasibility of combining functional near-infrared spectroscopy (fNIRS) and graph theory approaches to explore the topological attributes of human brain networks. However, the test-retest (TRT) reliability of the application of graph metrics to these networks remains to be elucidated. Here, we used resting-state fNIRS and a graph-theoretical approach to systematically address TRT reliability as it applies to various features of human brain networks, including functional connectivity, global network metrics and regional nodal centrality metrics. Eighteen subjects participated in two resting-state fNIRS scan sessions held ~20 min apart. Functional brain networks were constructed for each subject by computing temporal correlations on three types of hemoglobin concentration information (HbO, HbR, and HbT). This was followed by a graph-theoretical analysis, and then an intraclass correlation coefficient (ICC) was further applied to quantify the TRT reliability of each network metric. We observed that a large proportion of resting-state functional connections (~90%) exhibited good reliability ($0.6 < ICC < 0.74$). For global and nodal measures, reliability was generally threshold-sensitive and varied among both network metrics and hemoglobin concentration signals. Specifically, the majority of global metrics exhibited fair to excellent reliability, with notably higher ICC values for the clustering coefficient (HbO: 0.76; HbR: 0.78; HbT: 0.53) and global efficiency (HbO: 0.76; HbR: 0.70; HbT: 0.78). Similarly, both nodal degree and efficiency measures also showed fair to excellent reliability across nodes (degree: 0.52~0.84; efficiency: 0.50~0.84); reliability was concordant across HbO, HbR and HbT and was significantly higher than that of nodal betweenness (0.28~0.68). Together, our results suggest that most graph-theoretical network metrics derived from fNIRS are TRT reliable and can be used effectively for brain network research. This study also provides important guidance on the choice of network metrics of interest for future applied research in developmental and clinical neuroscience.

Competing Interests: Dr. YH is an Editor of PLOS ONE journal. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: yong.he@bnu.edu.cn (YH); xhzhao999@gmail.com (XHZ)

Introduction

The human brain is a highly complex system that can be represented as a structurally or functionally interconnected network that assures rapid segregation and integration of information processing. Considerable progress has recently been made in describing the topological organization of human whole-brain networks using neuroimaging data and graph-theoretical approaches [1,2,3,4]. A wealth of research has demonstrated that human structural or functional brain networks can be constructed using a variety of non-invasive neuroimaging techniques, including structural MRI [5], diffusion MRI [6,7,8], blood oxygenation level dependent functional MRI (BOLD-fMRI) [9,10,11], and functional near-infrared spectroscopy (fNIRS) [12]. Graph theory analysis has further revealed that these brain networks exhibit many non-random topological properties, such as small-world efficiency, modularity and highly connected hubs. These quantitative topological properties have been found to change with

normal development, aging and disease conditions (for reviews, see [10,13,14,15,16]).

Compared to traditional functional imaging techniques such as BOLD-fMRI, fNIRS is a newly developed but promising neuroimaging technology that uses light in the near-infrared spectrum (670–900 nm) to noninvasively monitor cortical hemodynamic variations induced by neural activation [17,18]. The primary advantages of fNIRS over other neuroimaging techniques include portability, effective cost, ecological validity and higher temporal resolution. One of the main contributions of fNIRS research is the detection of resting-state functional connectivity (RSFC), which characterizes the synchronization of spontaneous neural activity between spatially remote brain regions [19,20,21,22,23,24,25]. Very recently, we proposed resting-state fNIRS (R-fNIRS) and graph-theoretical approaches to explore the topological architecture of human whole-brain functional networks [12]. In the R-fNIRS based network analysis framework, the channels are considered vertices, and RSFCs between

channels are considered edges. Using this approach, we previously demonstrated that R-fNIRS can effectively characterize the topological attributes of human brain networks, such as small-world efficiency, modularity and highly connected hubs [12]. The results of this study were also consistent with recent findings from BOLD-fMRI research, suggesting the feasibility and validity of combining R-fNIRS and graph theory analysis to identify the functional properties of human brain networks. However, it still remains largely unknown whether the topological brain network measures derived from R-fNIRS data are repeatable or test-retest (TRT) reliable. If these metrics can be consistently replicated, they have the potential to be adopted as biomarkers in further cognitive neuroscience and clinical research. This could prove important in longitudinal studies of brain changes and in research using repeated measurements in the context of normal development or even pharmacological treatment. Consequently, it is an important and necessary task to explore the TRT reliability of functional brain networks derived from human R-fNIRS data.

In the present study, we aimed to provide a comprehensive assessment of the TRT reliability of graph metrics for R-fNIRS brain networks. To this end, we collected human R-fNIRS data in a group of healthy young adults over two sessions with a 20-minute interval between sessions. For each participant, we constructed whole-brain functional networks by computing temporal correlations between the time series of pairs of channels based on hemoglobin concentration. We further analyzed three sets of network properties (functional connectivity, global network metrics and regional nodal metrics) using graph-theoretical approaches [12]. Finally, intraclass correlation coefficients (ICC) were calculated to assess the TRT reliability of these network metrics.

Materials and Methods

Subjects and protocol

Twenty-one healthy right-handed subjects (mean age: 24.5 years; range: 21–27 years; 17 male) participated in this study. All subjects were college students recruited from Beijing Normal University. The study was approved by the Review Board at the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, and written informed consent was obtained from each subject prior to the experiment. Each participant underwent two eleven-minute sessions of resting-state fNIRS scanning separated by an interval of twenty minutes. Participants were instructed to remain still during scanning and to keep their eyes closed without falling asleep. During the rest interval, participants were instructed to sit still while wearing the probe holder, but they were allowed slight body and head motion. The signal quality from all measurement channels will be separately checked before the performance of the second session for each participant. If the signal at one or several channel(s) showed low signal-to-noise ratio, the whole probe holder or the specific source-detector pair determining the defective measurement channel will be readjusted.

Data acquisition and preprocessing

The data were acquired using a continuous wave (CW) near-infrared optical imaging system (CW6, TechEn Inc., MA, USA) equipped with 12 laser sources and 24 detectors designed to cover almost the entire head, including the frontal, temporal, parietal, and occipital lobes. The probe arrays allowed for 46 different measurement channels, with 3.2 cm of source-detector separation (Fig. 1). The positioning of the probes was based on the international 10–20 system, and the positions were measured at each scanning session to establish consistent placement. Specifically, two detectors around channels 17 and 18 were placed in T3 and T4, and the midpoint of the two detectors was localized in Cz. The system generated two wavelengths of near-infrared light (690 and 830 nm) and measured changes in the concentrations of oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) based on the modified Beer-Lambert law [26]. The sum of the HbO and HbR signals was defined as the total hemoglobin (HbT) concentration. Considering that the spontaneous fluctuations of each concentration data usually include various non-neural signals (e.g., respiratory and cardiac noises), we thus adopted a band-pass filtering with cutoff frequencies of 0.009 and 0.08 Hz to reduce the effects of these noises [23,27,28], which was similar to previous resting-state fMRI study [29]. Three subjects were excluded due to poor contact between the probe and scalp, leaving 18 subjects on which further analysis was performed. To obtain a stable signal, the first minute of each hemodynamic time series was discarded.
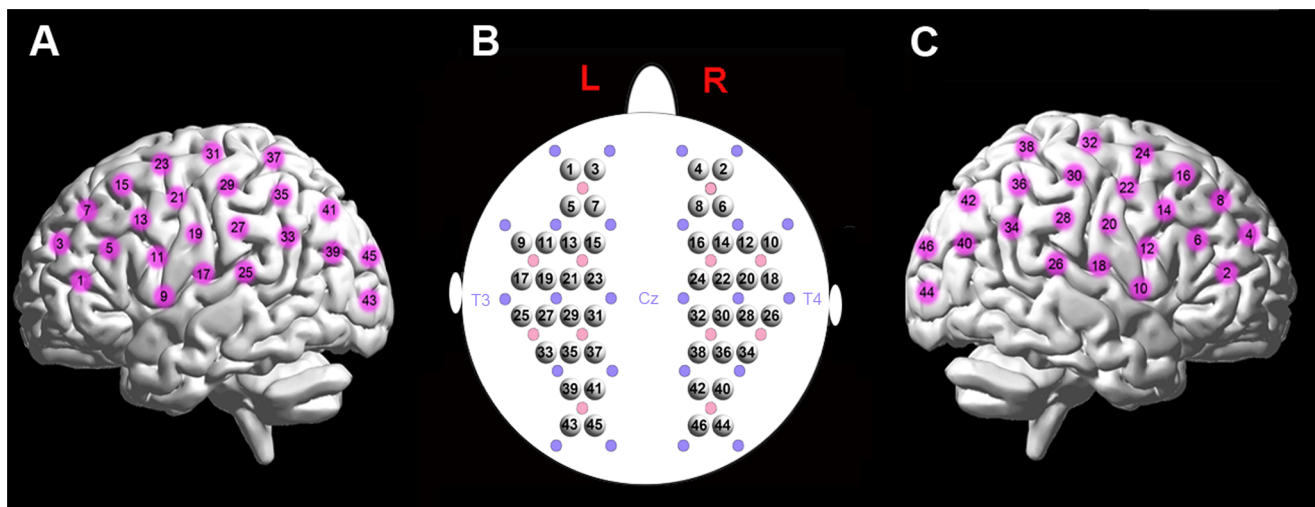

Figure 1. Schematic of fNIRS channel localization. (A) and (C) display the left and right view of measurement channels on an anatomical brain image. (B) The arrangement of the whole-head 46 measurement channels on a brain template. Note that the red and purple solid circles represent sources and detectors, respectively.
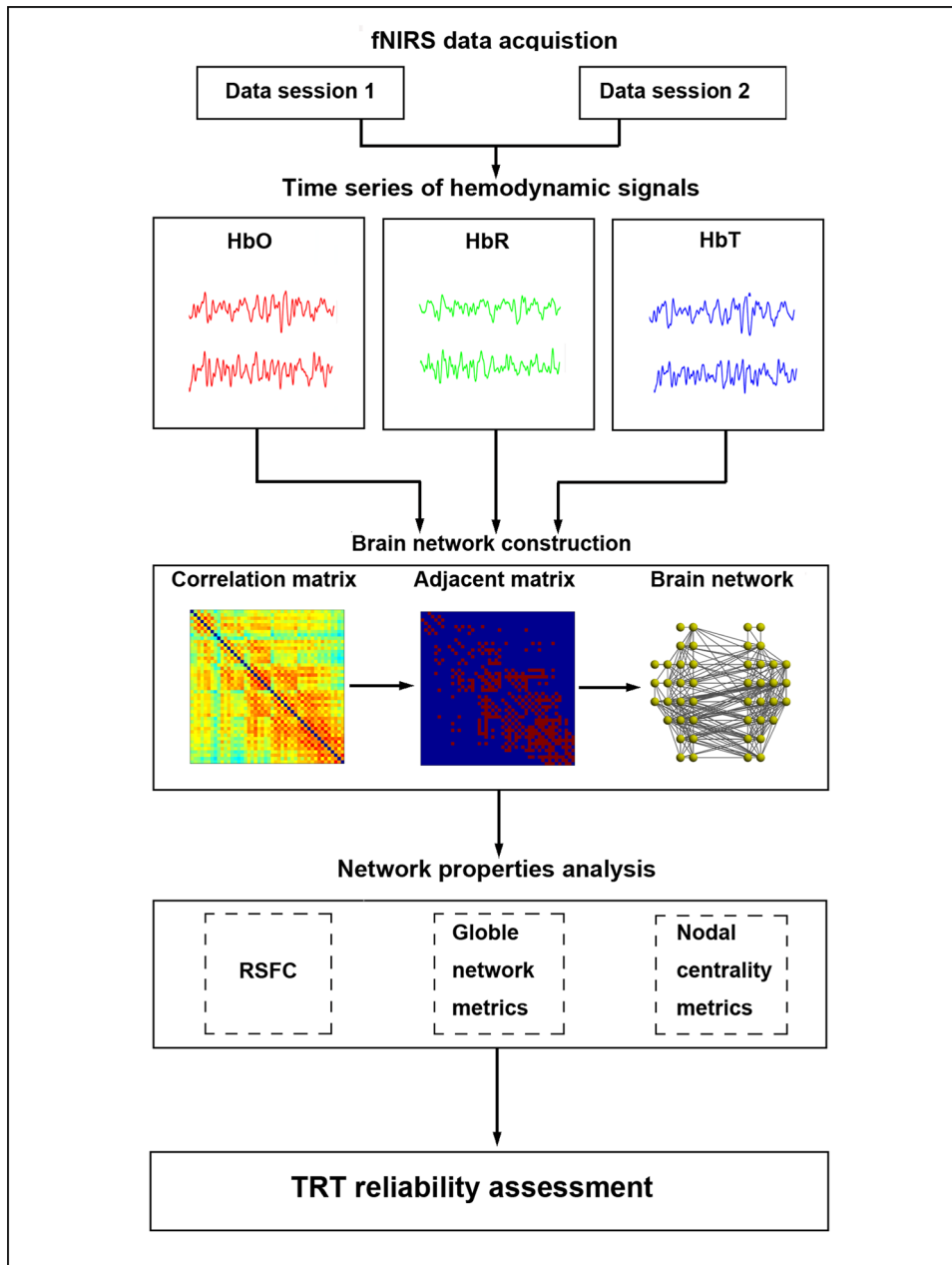doi:10.1371/journal.pone.0072425.g001

Figure 2. Demonstration of R-fNIRS brain network reliability analysis.
doi:10.1371/journal.pone.0072425.g002

The sampling rate for the optical signal was set to 25 Hz, which resulted in 15,000 sample points for each 10-min dataset for each subject.

Network construction

In a previous study, we developed a procedure to construct whole-brain functional networks based on R-fNIRS data [12]. Briefly, RSFC was calculated with Pearson correlation analysis between the time series of every pair of nodes, where nodes were measurement channels. This resulted in a {46×46} correlation matrix for each participant at each session. The correlation matrix was then thresholded into a binary matrix that described the topological organization of the functional networks. As in our previous study [12], we chose sparsity (S) as a threshold measure,

and sparsity was defined as the number of actual connections divided by the maximum possible number of connections in the network, ranging from 0.01 to 0.99 at an interval of 0.01.

Network analysis

We explored two sets of topological measures for the brain networks: 1) global network metrics: small-world properties (clustering coefficient C_p, characteristic path length L_p, normalized clustering coefficient γ, normalized characteristic path length λ, and small-world σ), efficiency parameters (local efficiency E_{loc} and global efficiency E_{glob}), modularity Q, hierarchy β, and assortativity r; and 2) nodal characteristics: nodal degree, nodal efficiency, and nodal betweenness. For each subject at each sparsity level, we computed the entire network metrics based on an in-house
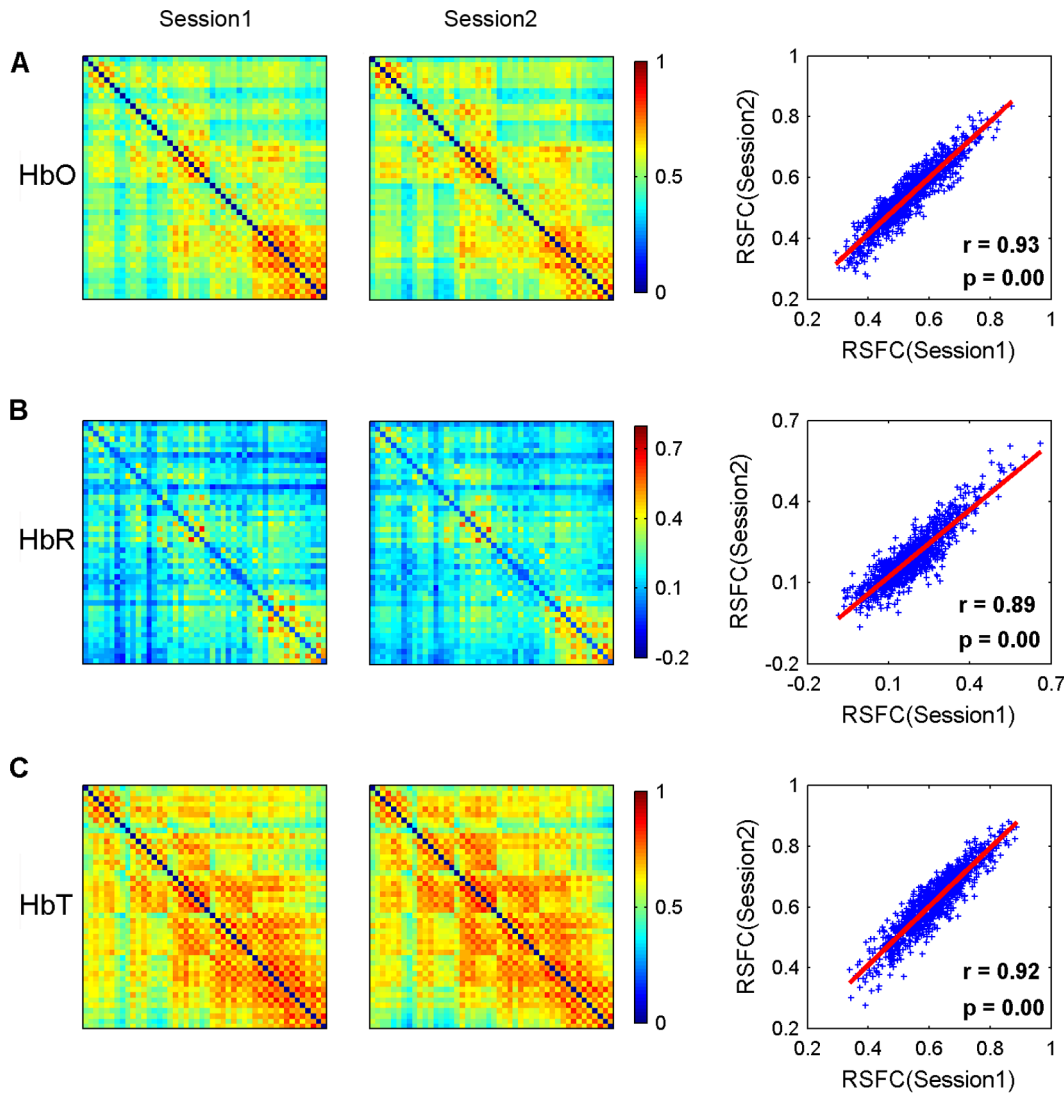
Figure 3. Spatial similarity of RSFC maps. Group-level RSFC maps for session 1 and session 2 and their Pearson correlation are displayed in the first to third columns. Figures (A) to (C) correspond to the RSFC data derived from HbO, HbR, and HbT, respectively. High similarity between sessions was observed in both the qualitative visual inspection and quantitative correlational analysis.
doi:10.1371/journal.pone.0072425.g003

GRETNA package. These metrics are described in further detail below.

Small-world. Small-world measures of a network (clustering coefficient, C_p, and characteristic path length, L_p) were originally proposed by Watts and Strogatz (1998). Briefly, C_p is the average of clustering coefficients over all nodes in a network, where the clustering coefficient of a node v is defined as the number of existing connections among the neighbors of the node divided by all their possible connections. C_p quantifies the extent of local cliquishness of a network [30,31]. L_p is defined as the average of the shortest path lengths (i.e., the minimum number of edges that link any two nodes of the network) between any pair of nodes in the network. It quantifies the capability of parallel information propagation within a network [32]. Mathematically, a real network would be considered small-world if it meets the following requirements: $\gamma = C_p{}^{real}/C_p{}^{rand} > 1$ and $\lambda = L_p{}^{real}/L_p{}^{rand} \approx 1$ [31], where $C_p{}^{rand}$ and $L_p{}^{rand}$ are the mean clustering coefficient and characteristic path length of matched random networks that preserve the same number of nodes, edges, and degree distribution as the real network [33]. The two

metrics can also be summarized into a simple quantitative small-worldness index σ: $\sigma = \gamma/\lambda > 1$ [34].

Efficiency. The global efficiency of a network is defined as the average inverse shortest path length. Correspondingly, local efficiency is defined as the efficiency of the local sub-group of a node v that comprises only the direct neighbors of node v. Global and local efficiency measure the ability of a network to transmit information at the global and local level, respectively [32].

Modularity. The modularity measure Q is defined by the following equation [35]:

$$Q = \sum_{m=1}^{N} \left[\frac{l_m}{L} - \left(\frac{d_m}{2L} \right)^2 \right],$$

(1)

where N is the number of modules, L is the total number of links in the network, l_m is the total number of links between nodes in module m, and d_m is the sum of the degrees of the nodes in module

Table 1. Pearson correlations at individual-level RSFC maps between sessions.

Subject ID	HbO		HbR		HbT	
	r	p	r	p	r	p
1	0.64	2.64×10^{-121}	0.40	2.34×10^{-41}	0.76	3.60×10^{-192}
2	0.66	2.69×10^{-129}	0.62	5.54×10^{-110}	0.66	5.67×10^{-132}
3	0.56	3.83×10^{-86}	0.59	1.81×10^{-96}	0.56	2.83×10^{-88}
4	0.75	1.49×10^{-184}	0.63	1.16×10^{-117}	0.66	1.04×10^{-130}
5	0.76	2.72×10^{-198}	0.69	3.22×10^{-145}	0.60	8.93×10^{-104}
6	0.75	7.12×10^{-188}	0.78	1.12×10^{-210}	0.74	6.87×10^{-183}
7	0.54	3.86×10^{-79}	0.77	1.87×10^{-206}	0.72	7.23×10^{-263}
8	0.76	6.81×10^{-198}	0.53	3.35×10^{-75}	0.79	9.78×10^{-221}
9	0.85	4.04×10^{-294}	0.88	0.00	0.74	3.36×10^{-182}
10	0.82	1.30×10^{-253}	0.78	3.31×10^{-208}	0.90	0.00
11	0.69	1.64×10^{-147}	0.61	1.86×10^{-106}	0.74	1.40×10^{-176}
12	0.66	1.51×10^{-132}	0.63	2.81×10^{-117}	0.62	1.15×10^{-109}
13	0.69	9.30×10^{-147}	0.52	1.08×10^{-72}	0.52	9.43×10^{-72}
14	0.73	2.17×10^{-173}	0.75	3.82×10^{-186}	0.72	8.98×10^{-166}
15	0.57	1.01×10^{-88}	0.55	1.33×10^{-81}	0.55	1.84×10^{-84}
16	0.65	5.15×10^{-127}	0.78	1.41×10^{-214}	0.75	7.91×10^{-189}
17	0.86	4.53×10^{-297}	0.90	0.00	0.81	1.64×10^{-238}
18	0.67	2.42×10^{-137}	0.56	4.73×10^{-88}	0.66	1.86×10^{-128}
mean	0.70		0.67		0.69	
std	0.09		0.13		0.10	

m. It has been shown that connections are usually denser within modules than between them.

Hierarchy. In a hierarchical network, nodes with a high degree show a lower clustering coefficient, while nodes with a low degree show a higher clustering coefficient, reflecting the top-down organization of the network. This ratio can be quantified by [36]:

$$C_p \approx k^{-\beta} . \qquad (2)$$

The coefficient β was estimated by fitting a linear regression line to the plot of $log(C_p)$ versus $log(k)$.

Assortativity. Assortativity is a measure of how often nodes are linked to other nodes with a similar degree. Accordingly, it is defined as the correlation between the degree of a node and the mean degree of its direct neighbors [37].

Nodal degree, efficiency, and betweenness. The degree of a node v is defined as the number of edges linked directly to the node. The efficiency of a node v is the inverse of the harmonic mean of the minimum path length between the node v and all other nodes in the network [30,38]. Nodal efficiency is a measure of local connectivity. The betweenness of a node v is defined as the number of shortest paths between any two nodes that run through node v [39]. Nodal betweenness quantifies how important a node is within a network. It is defined as:

$$N_{bc}(v) = \sum_{m \neq v \neq n} \frac{\delta_{mn}(v)}{\delta_{mn}} \qquad (3)$$

where δ_{mn} is the total number of shortest paths from node m to node n, and $\delta_{mn}(v)$ is the number of shortest paths from node m to

node n that pass through node v. A node with a high N_{bc} indicates a greater impact of the node on information flow across the whole network.

Test-retest reliability analysis

The intraclass correlation coefficient was used to evaluate the test–retest reliability of R-fNIRS brain networks [40]. To comprehensively investigate the test–retest reliability of the graph theory-derived R-fNIRS network, we analyzed the reliability of R-fNIRS network properties from three perspectives: resting-state functional connectivity (RSFC), global network metrics and regional nodal centrality metrics. For each metric, the individual RSFC or network parameters values of all subjects in both two sessions were first merged into one 18×2 matrix, with rows corresponding to participants and columns corresponding to repeated measurements in two sessions. Notably, for RSFC, the 18×2 matrix represents the connection strength between any two nodes across two sessions; for network metrics, each 18×2 matrix represents the global or nodal parameter values across two sessions. Using a one-way ANOVA on each matrix, with random participant effects, we split the total variance of each network property into between-participant (MS_b) and within-participant (MS_w) variance. Then, the ICC values were calculated according to the following equation [40]:

$$ICC = \frac{MS_b - MS_w}{MS_b + (k-1)MS_w} \qquad (4)$$

where k is the number of repeated observations per participant, MS_b is the between-participant variance and MS_w is the within-participant variance. All reliability indices discussed above were evaluated according to the criteria proposed by Sampat and Winer et al. [41,42], where an ICC value from 0 to 0.25 indicates poor reliability; 0.25 to 0.4 low reliability; 0.4 to 0.6 fair reliability; 0.6 to 0.75 good reliability; and 0.75 to 1.0 excellent reliability. Notably, for one metric, if the within-participant variance is larger than the between-participant variance, it will lead to a negative ICC. Theoretically, such a situation is impossible [43], and the reasons for negative values is even currently unclear [44]. Thus, the negative ICC will be set to zero (not reliable), as generally suggested [24,45]. Because we constructed the network over a continuous sparsity threshold range, the ICC value is a function of the threshold. To further conduct a threshold-independent reliability assessment, we also calculated the integral or area under the curve (AUC) of each network metric [46] to be used as the corresponding ICC scalar of network evaluation. Figure 2 displays a flowchart of the reliability analysis of the R-fNIRS brain network.

Denoising using independent component analysis

A general problem with R-fNIRS recordings is that the acquired signals often include various different types of signal noise and artifacts (e.g., instrumental noise, motion-induced artifacts and physiological noises). Independent component analysis (ICA) provides an important technique to reduce the effects of these noise sources [25,47,48]. In this study we re-analyzed our R-fNIRS data and assessed the TRT reliability of functional brain networks after denoising based on a temporal ICA analysis. The analysis procedure we used was similar to that in previous R-fNIRS studies [24,25] and was performed here by using a publicly available software, *FastICA v2.5* (www.cis.hut.fi/ projects/ica/fastica/). Specifically, the parameter settings of the *FastICA* algorithm [49] included: number of iteration steps = 10000, step length = 0.00001, approach = "deflation", initial
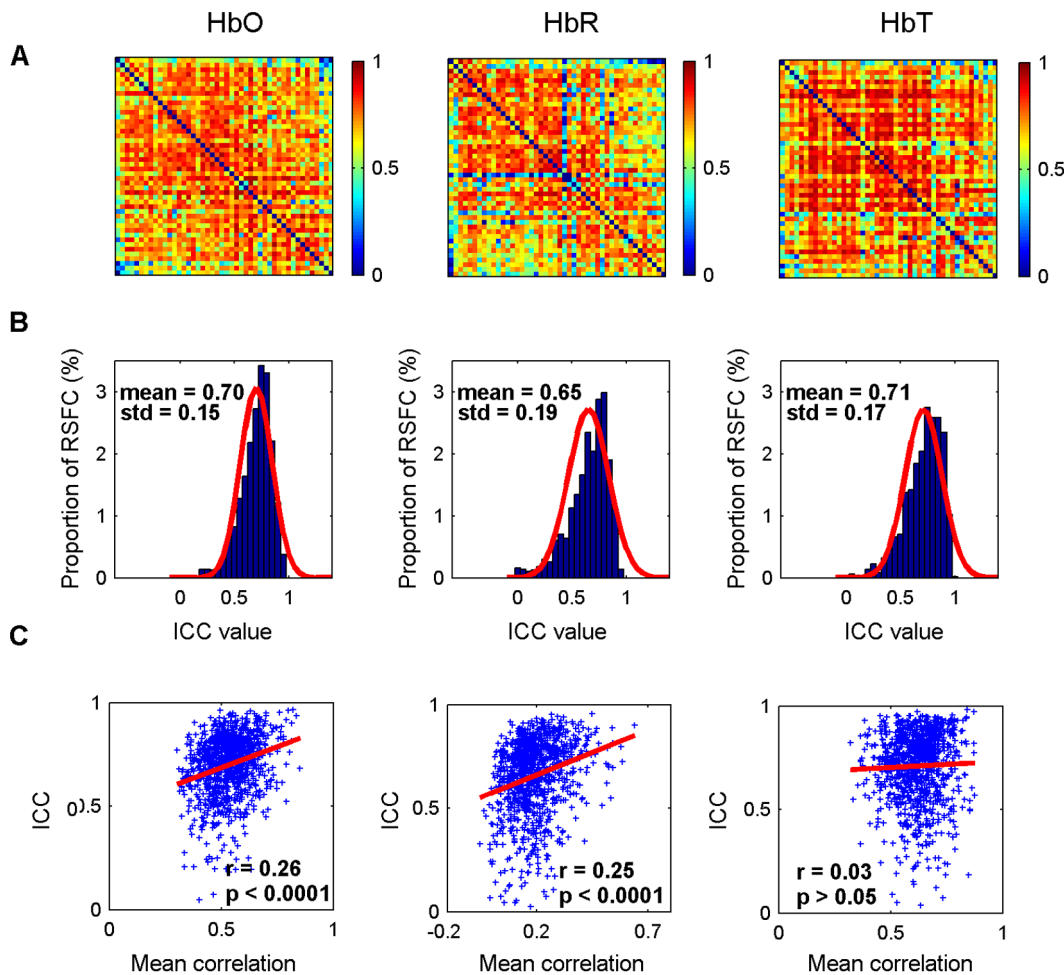
Figure 4. Reliability analysis of RSFC maps. The first to third columns correspond to the data derived from HbO, HbR, and HbT, respectively. (A, B) The TRT reliability of RSFC maps and their corresponding reliability distributions. The reliability displays approximately normal configuration for all 1035 (i.e., 46×45/2) connections. The connections exhibit good reliability across HbO (mean *ICC* values 0.70), HbR (0.65) and HbT (0.71). (C) The relationship between RSFC strength and reliability as assessed by scatterplots. Each dot represents the group-level RSFC strength and the corresponding *ICC* value at the same connections. The trend lines were obtained by a linear least-squares fit method. Significant ($p<0.05$) positive correlations were found for HbO and HbR signals, suggesting stronger RSFC leads to higher reliability for both these signals.
doi:10.1371/journal.pone.0072425.g004

value = "random", nonlinearity = "skew", fine-tune = "on", and stabilization = "on". The ICA analysis was separately performed on the raw data of HbO, HbR and HbT with the following procedures: extracting steady hemoglobin concentration signals (e.g. removing the initial time points from total data length), reducing the dimensionality of the data with principal component analysis (PCA), performing ICA analysis on the reduced dimensional data, identifying noise components, removing noise from the measured data, and calculating "real" neural activity signals. *FastICA* can automatically adjust the results of the analysis required to compensate for these preprocessing steps. The number of retained principal components (PCs) was determined according to the minimum number of PCs that retained more than 99% of data variance [50]. Meanwhile, the number of independent components through ICA decomposition was set to be equal to the number of PCs. After ICA analysis, the components related to noise and artifacts were identified from three aspects: temporal profiles, spatial maps and power spectra. A component would be considered noise if it met one of the following conditions [25]: *i*) the corresponding temporal profile included sudden jumps, slowly-varied U or inverted U-shaped spike, or numerous inter-current

quick spikes; *ii*) the dominant frequency of power spectra of the component was outside the range of 0.01~0.1Hz; *iii*) the spatial map of the component presented a global and spatially dispersive pattern. Notably, the map could be approximately related to the interference in superficial layers of head [48]. After identifying these types of noise components, each concentration signal of HbO, HbR, and HbT was reconstructed with a particular component eliminated from the original R-fNIRS time course by replacing zero in the corresponding column of mixing matrix [48]. Finally, we reconstructed the brain networks and then computed their topological properties and the *ICCs* as described above.

Results

TRT reliability of RSFC maps

The group-level RSFC maps of two R-fNIRS scans (session 1 and session 2) and the linear relationship between them are presented in Fig. 3A–C. In general, the spatial patterns of the RSFC maps exhibited high similarity between sessions across HbO, HbR, and HbT, and this can be observed in both the
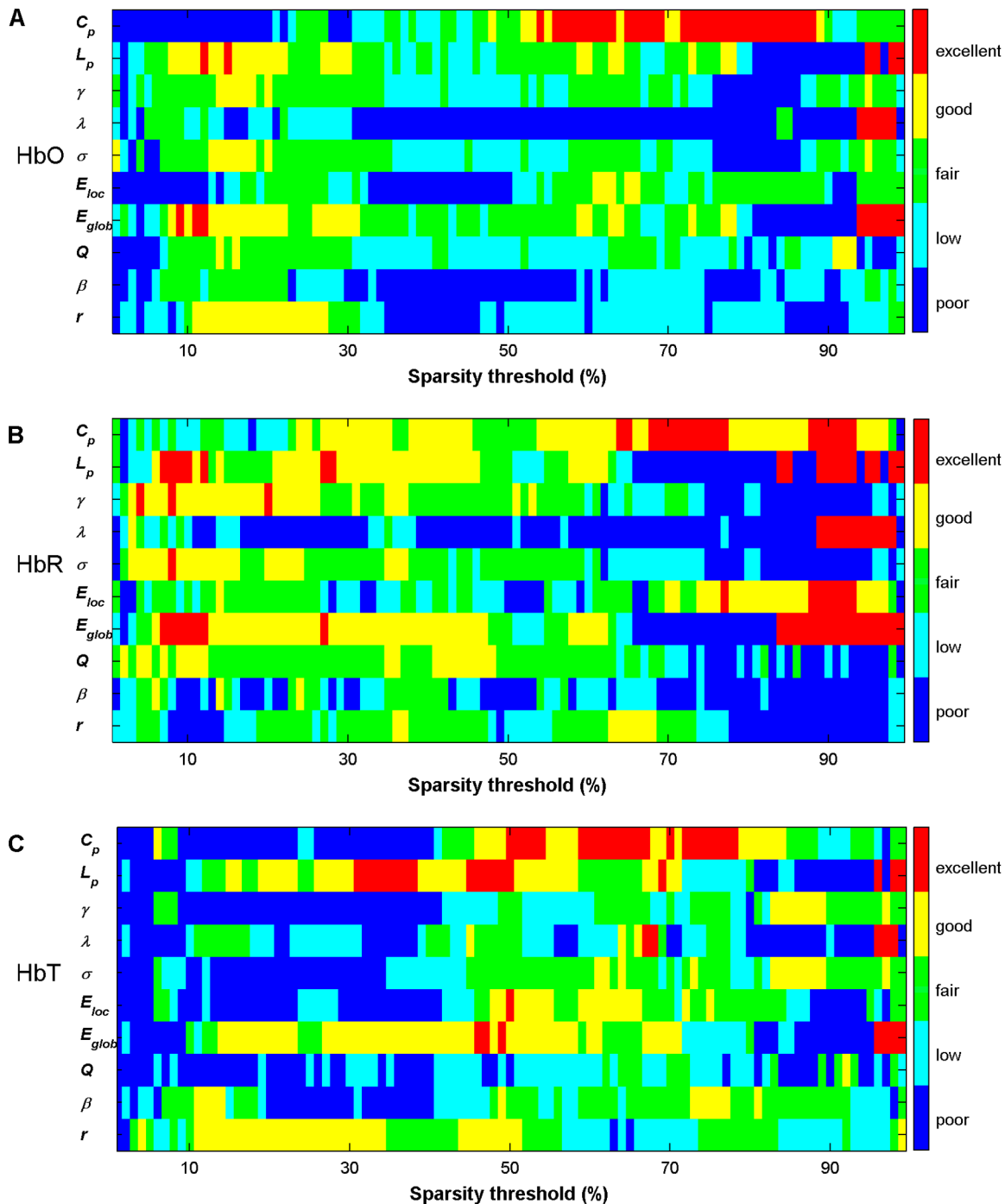
Figure 5. TRT reliability of global network metrics as a function of sparsity threshold. (A–C) The global metric reliability was derived from HbO, HbR, and HbT, respectively. Five colors correspond to five different reliability grades. The red, yellow, green, cyan, and blue colors represent excellent ($0.75 < ICC < 1$), good ($0.6 < ICC < 0.75$), fair ($0.4 < ICC < 0.6$), low ($0.25 < ICC < 0.4$), and poor ($ICC < 0.25$) reliability of global network metrics, respectively. C_p, L_p, γ, λ, and σ denote the clustering coefficient, characteristic path length, normalized clustering coefficient, normalized characteristic path length, and small-world, respectively. E_{loc} and E_{glob} denote local efficiency and global efficiency, respectively. Q, β, and r denote modularity, hierarchy, and assortativity, respectively.
doi:10.1371/journal.pone.0072425.g005

qualitative visual inspection and quantitative correlation analysis (HbO: $r = 0.93$, $p < 0.0001$; HbR: $r = 0.89$, $p < 0.0001$; HbT: $r = 0.92$, $p < 0.0001$). The individual RSFC correlation between sessions was also examined independently, and these results are summarized in Table 1. The RSFC maps between sessions also manifested high similarities ($p < 0.001$) for each participant and

each hemoglobin concentration signal; the average correlation coefficients across subjects were 0.70 ± 0.09 (mean and standard deviation) for HbO, 0.67 ± 0.13 for HbR, and 0.69 ± 0.10 for HbT.

The consistency of these measures was further confirmed by an analysis of the reliability of the RSFC maps. Figure 4 shows the TRT reliability for the RSFC maps and their corresponding

Table 2. Significant differences in the global network metric (across subjects) between sessions revealed by a paired t-test.

Global network metrics	HbO		HbR		HbT	
	t	*p*	*t*	*p*	*t*	*p*
C_p	−0.12	0.91	0.68	0.51	0.82	0.42
L_p	−0.05	0.96	0.24	0.82	−1.13	0.27
γ	1.80	0.09	0.53	0.60	0.36	0.72
λ	−1.28	0.22	−0.84	0.42	0.08	0.94
σ	1.99	0.06	0.43	0.67	0.43	0.67
E_{loc}	0.06	0.95	0.52	0.61	1.15	0.26
E_{glob}	−0.15	0.88	−0.29	0.78	0.52	0.61
Q	0.65	0.53	−0.26	0.80	0.88	0.39
β	1.31	0.21	2.44	0.03	1.37	0.19
r	−0.25	0.81	−2.63	0.02	−0.07	0.94

The AUC of each global metric was used as the statistical variable.
doi:10.1371/journal.pone.0072425.t002

reliability distributions. The average *ICC* values across all 1035 (i.e., 46×45/2) connections were 0.70±0.15 for HbO, 0.65±0.19 for HbR, and 0.71±0.17 for HbT. These *ICC* values also displayed an approximately normal distribution for all 1035 connections at each concentration signal (Fig. 4B). Notably, most of the connections for each concentration signal also exhibited fair to excellent reliability. For HbO, 985 of 1035 (95%) functional connections exhibited fair to excellent reliability (fair: 214, 20.7%; good: 402, 38.8%; excellent: 369, 35.7%). Similarly, for HbR and HbT, 902 of 1035 (87%) and 948 of 1035 (92%) functional connections exhibited fair to excellent reliability (HbR: fair: 244, 23.6.7%; good: 333, 32.2%; excellent: 328, 31.7%. HbT: fair: 229, 22.1%; good: 321, 31%; excellent: 398, 38.5%). Furthermore, a Pearson's correlation analysis revealed significant positive correlations between RSFC strength and reliability for HbO ($r = 0.26$, $p < 0.0001$) and HbR ($r = 0.25$, $p < 0.0001$) signals, whereas no significant correlations were found for HbT ($r = 0.03$, $p > 0.05$) (Fig. 4C).
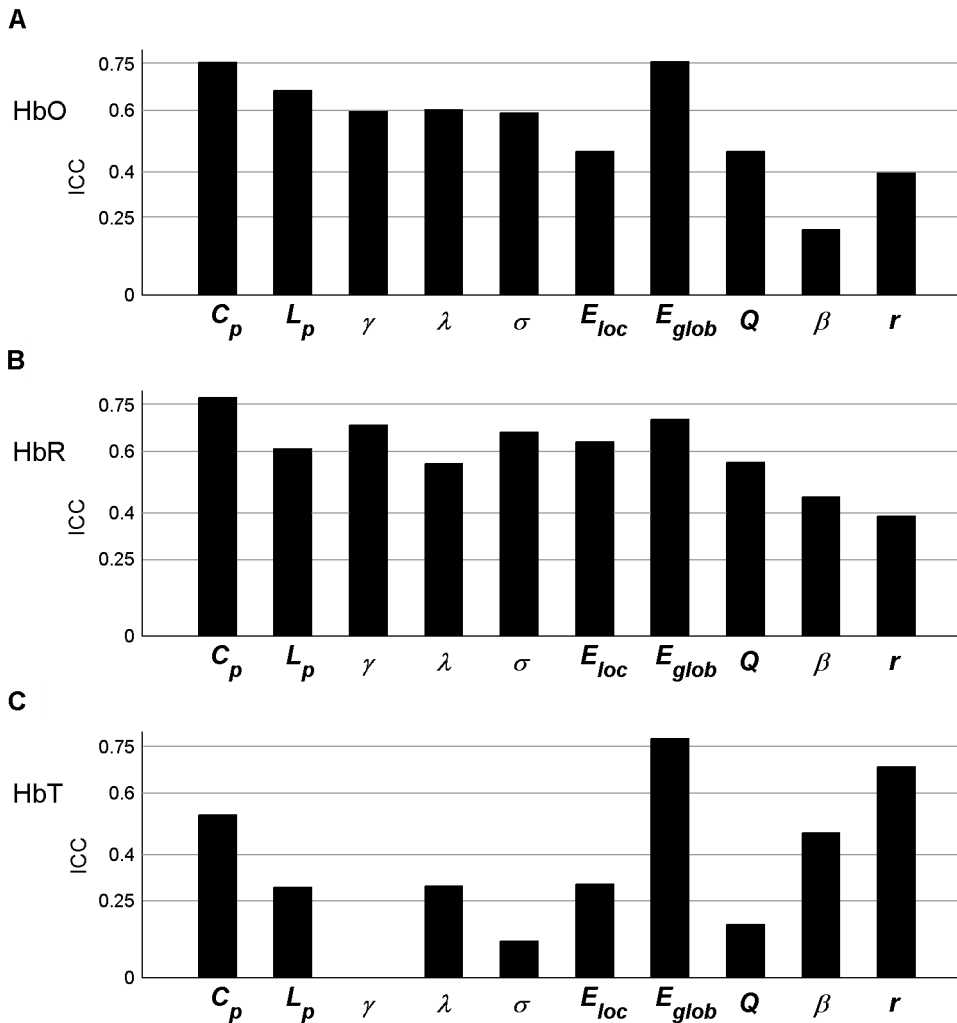
Figure 6. Threshold-independent reliability analysis of global network metrics. The areas under the curves (AUCs) of each global metric were used to provide threshold-independent reliability evaluation. (A–C) The global metric reliability was derived from HbO, HbR, and HbT, respectively. C_p, L_p, γ, λ, and σ denote the clustering coefficient, characteristic path length, normalized clustering coefficient, normalized characteristic path length, and small-world, respectively. E_{loc} and E_{glob} denote local efficiency and global efficiency, respectively. Q, β, and r denote modularity, hierarchy, and assortativity, respectively.
doi:10.1371/journal.pone.0072425.g006

TRT reliability of global network metrics

Figure 5 shows the TRT reliability of 10 global network metrics against the increasing sparsity thresholds, from 0.01 to 0.99, stepped by 0.01. In general, the reliability of the network metrics manifested considerable variation across the sparsity threshold values. The reliability also displayed moderate variation across different network metrics at the same thresholds, irrespective of the HbO, HbR, and HbT signals. The reliability of most of the network metrics was maintained at a fair to excellent level over a broad threshold range for the three hemoglobin signals. For example, global efficiency E_{glob} exhibited fair to excellent reliability in the threshold ranges of 0.1 to 0.7 for each hemoglobin signal.

Additionally, the AUC was calculated to test for significant differences in each metric across subjects between sessions. The results indicated that almost all parameters (except β in HbR) were not significantly different (paired t-tests, $p > 0.05$, Table 2), suggesting a high similarity across global network measures between sessions at the individual level. The AUC of each metric was also employed to calculate the reliability measure. The results of the reliability analyses are shown in Fig. 6A–C for HbO, HbR, and HbT, respectively. For both HbO and HbR signals, almost all network metrics consistently exhibited fair to excellent reliability (except β in HbO and r in HbR showing poor and low reliability, respectively). In contrast, four of ten metrics showed fair to excellent reliability for HbT (Fig. 6C). Numerically, the ICC values for HbR were the highest ($ICC = 0.60$, averaged across all metrics), followed by HbO (mean $ICC = 0.55$); HbT had the lowest relative ICC values (mean $ICC = 0.36$). Further statistical analysis using a one-way repeated measures ANOVA revealed that the reliability of all network metrics from the three different hemoglobin concentration signals exhibited significant differences [F(2, 9)

$= 5.20$; $p = 0.017$] and that HbR generally displayed the best relative reliability (paired t-tests, $p = 0.03$). Viewed from a single network metric, the clustering coefficient C_p and global efficiency measure E_{glob} displayed relatively better reliability, with an average reliability level of excellent ($ICC > 0.7$) across the HbO, HbR, and HbT signals.

TRT reliability of nodal metrics

Figure 7 shows the nodal reliability of the degree, efficiency and betweenness measures as a function of the sparsity threshold values. Similar to global network metrics, the reliability of the nodal metrics also changes with sparsity and varies between metrics. Visually, betweenness exhibited the worst reliability across the range of thresholds, while the reliabilities of degree and efficiency were higher and comparable to one another for the HbO, HbR, and HbT signals. Furthermore, nodal reliability was estimated based on a threshold-independent metric, the AUC of each nodal metric. Figure 8 shows the reliability distribution of the total 46 nodes on one brain image. The nodes with fair to excellent reliability held ~90% (41/46) of the total nodes and were distributed widely throughout the whole brain (Figs. 8A and 8B), suggesting that nodal reliability is not related to the spatial position of the nodes in brain regions.

We further performed a two-factor repeated measures ANOVA to investigate the effects of different nodal measures (degree, efficiency and betweenness) and different concentration signals (HbO, HbR and HbT) on nodal reliability. The AUC-based ICC values were used for this statistical analysis. The results showed a significant main effect of network measures [$F (2, 45) = 63.47$, $p < 0.0001$] (Fig. 9A) and a significant main effect of concentration signals [$F (2, 45) = 5.54$, $p = 0.0054$] (Fig. 9B). There was also a
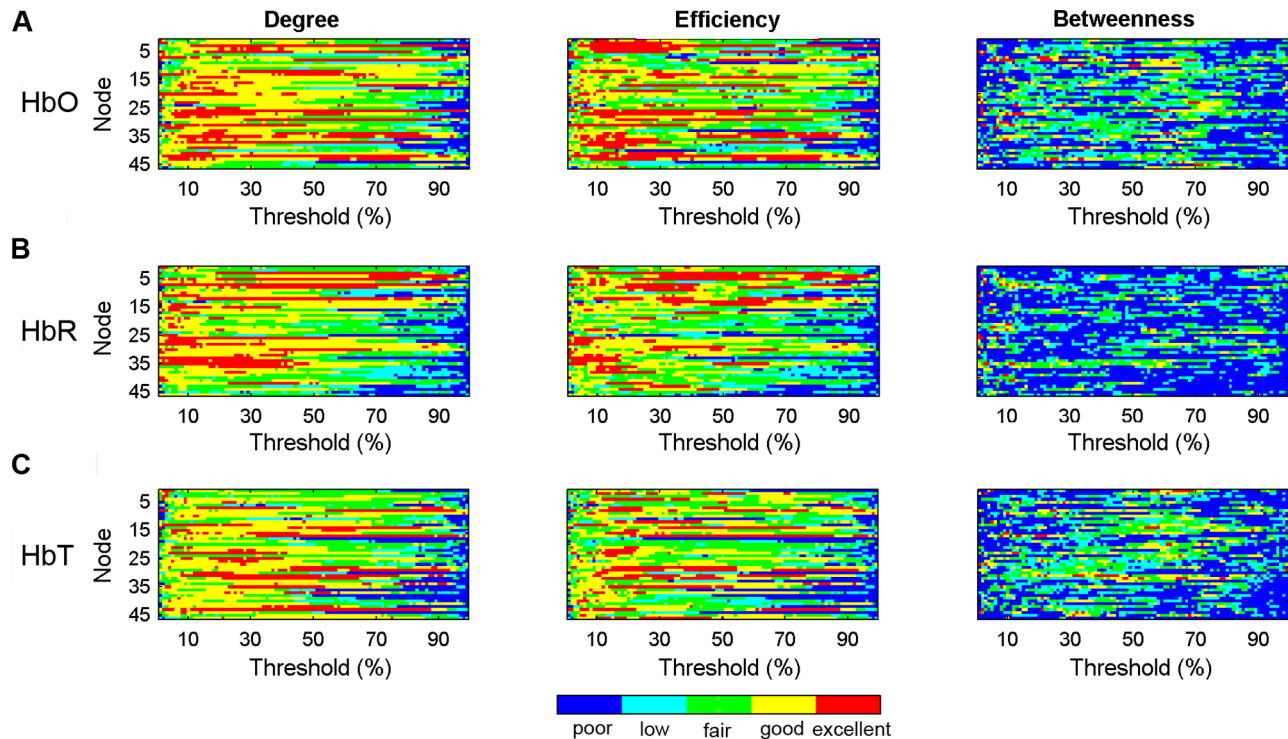

Figure 7. TRT reliability of nodal centrality metrics as a function of sparsity threshold. (A–C) The nodal metric reliability was derived from HbO, HbR, and HbT, respectively. The five colors correspond to five different reliability grades: red, yellow, green, cyan, and blue represent excellent (0.75< *ICC* <1), good (0.6< *ICC* <0.75), fair (0.4< *ICC* <0.6), low (0.25< *ICC* <0.4), and poor (*ICC*<0.25) reliability of the nodal centrality metrics, respectively.
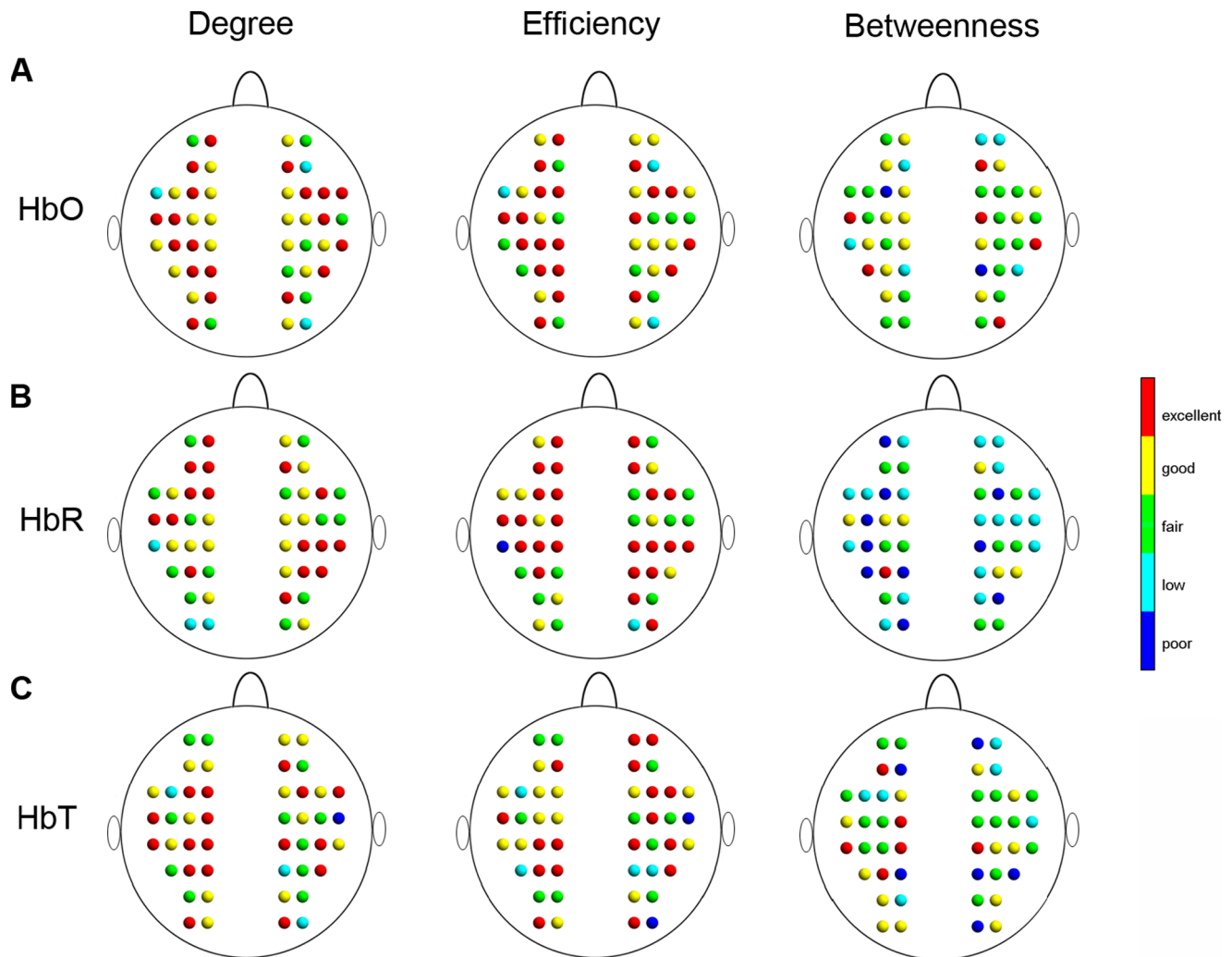doi:10.1371/journal.pone.0072425.g007

Figure 8. Threshold-independent reliability analysis of nodal centrality metrics. The areas under the curves (AUCs) of each nodal metric were used to provide threshold-independent reliability evaluation. (A–C) The nodal reliability was derived from HbO, HbR, and HbT, respectively. Different colors in the nodes correspond to different reliability grades: red, yellow, green, cyan, and blue colors represent excellent (0.75< ICC <1), good (0.6< ICC <0.75), fair (0.4< ICC <0.6), low (0.25< ICC <0.4), and poor (ICC<0.25) reliability of the nodal centrality metrics, respectively.
doi:10.1371/journal.pone.0072425.g008

significant interaction between network measures and concentration signals [F (2, 45) = 11.79, p<0.0001], indicating that the reliability of nodal network measures was not consistent across all concentration signals. Post hoc comparisons (paired t-tests) showed that the reliability of degree and efficiency was concordant and was higher than that of betweenness (p<0.0001) for HbO, HbR, and HbT. The high reliability of degree and efficiency did not significantly differ among these three concentration signals (p > 0.05), suggesting the robustness of degree and efficiency for varying types of hemoglobin information.

TRT reliability of network metrics after denoising using ICA

All the results of network analysis after denoising using ICA are presented in Figs. 10–11, Figs. S1, S2, S3, S4, S5 in Supporting Information S1, and Tables S1–S2 in Supporting Information S1. In general, we found that the main results remained few changes as compared to those without ICA analysis. For example, the ICC values of RSFC maps with ICA denosing were 0.63±0.20 for HbO, 0.68±0.20 for HbR, and 0.71±0.21 for HbT, which was

highly compatible with the results without ICA analysis (HbO: 0.70±0.15; HbR: 0.65±0.19; HbT: 0.71±0.17) (Figs. S1–S2 in Supporting Information S1). Meanwhile, at the individual level, the RSFC maps still remain a high correlation between two sessions after denoising using ICA (Table S1 in Supporting Information S1). For global and nodal network metrics, the reliability was threshold-sensitive and varied among both network metrics and hemoglobin concentration signals (Figs. 10–11 and Figs. S3, S4, S5 in Supporting Information S1), which was highly consistent with the results without ICA denosing. Specifically, the ICC values for HbR were the highest (ICC = 0.60, averaged across all metrics), followed by HbO (mean ICC = 0.48) and HbT (mean ICC = 0.40). Moreover, the global efficiency E_{glob} still displayed the highest reliability (the average ICC = 0.83 across the HbO, HbR, and HbT signals) among all 10 global metrics (Fig. 10 and Fig. S3 in Supporting Information S1) and both the nodal degree and efficiency centralities showed higher ICC values than the betweenness centrality (p<0.0001) for all three hemoglobin concentration signals (Fig. 11 and Figs. S4–S5 in Supporting Information S1). There were no significant differences in the ICA-
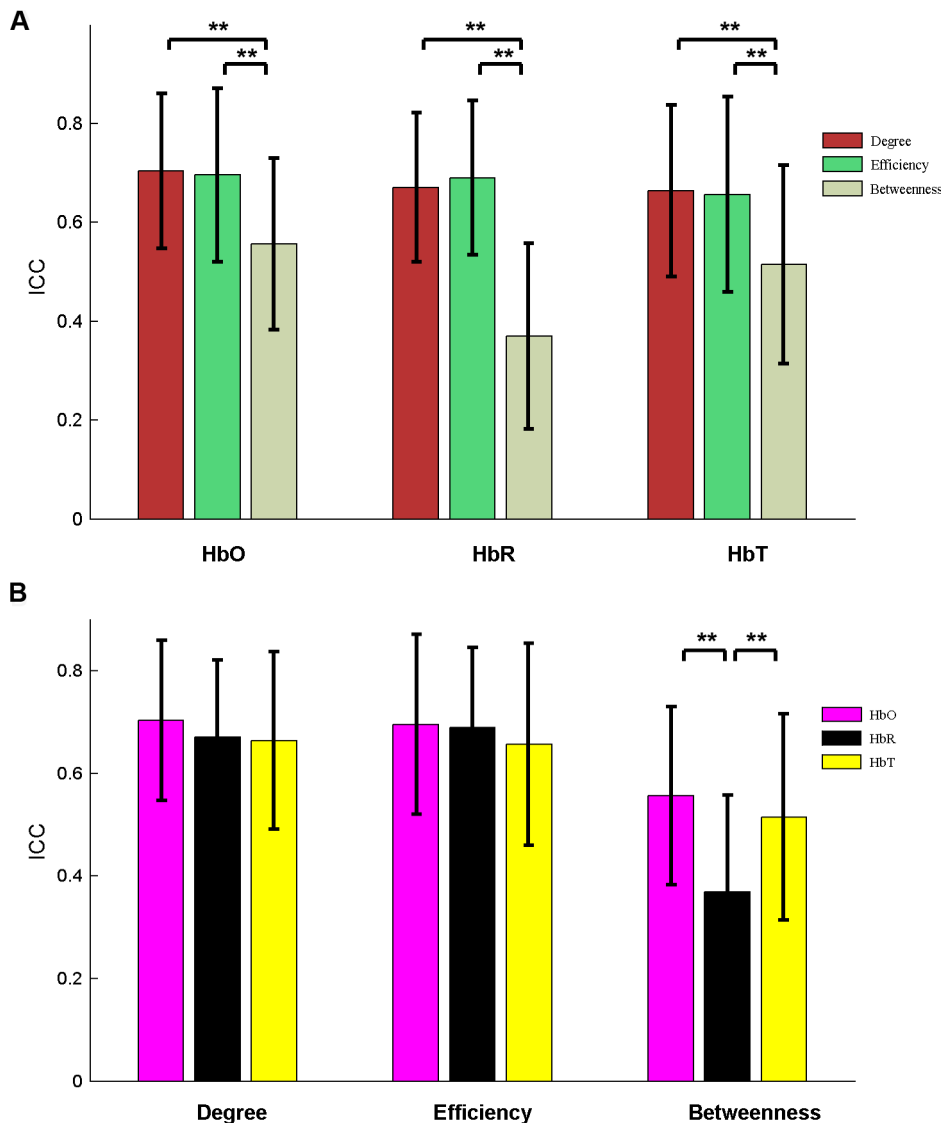
Figure 9. Significant differences (paired t-test) in nodal metric reliability. (A) Reliability for three nodal centrality metrics (degree, efficiency, and betweenness). (B) Reliability for three concentration signals (HbO, HbR, and HbT). Note that the reliability of nodal degree and efficiency is concordant across HbO, HbR, and HbT and is also significantly ($p<0.0001$) higher than that of nodal betweenness. Error bars correspond to the standard deviation of the mean across the total nodes. The double asterisk indicates $p<0.0001$.
doi:10.1371/journal.pone.0072425.g009

derived global network metrics between sessions (Table S2 in Supporting Information S1). Notably, we also observed small variations on the reliability magnitude of several network metrics with ICA-based denoising. For example, the ICC values of E_{loc} and E_{glob} slightly increased for HbO, HbR, and HbT signals, whereas the ICC values of C_p and L_p slightly decreased for HbO and HbT signals, as compared to the results without ICA denosing (Fig. 6 vs. Fig. 10).

Discussion

Previously, we showed that R-fNIRS can be used in combination with graph theory methodology to reveal topological properties of human functional brain networks [12]. In the current study, we further addressed the TRT reliability of these topological properties. A comprehensive assessment of network reliability was carried out on three sets of network measures

(RSFC, global network metrics and regional nodal metrics). These network metrics were calculated based on two sets of noise-reduction data using frequency-based and ICA-based denoising approaches, respectively, and the results consistently demonstrated that the R-fNIRS brain networks possessed good reliability on RSFC and fair to excellent reliability on most global and nodal network metrics.

TRT reliability of RSFC maps

Based on a correlational analysis and a reliability assessment, we demonstrated that the fNIRS-based RSFC maps have high similarity (Pearson correlation: $r=0.91\pm0.03$, averaged across three concentration signals) and high reliability ($ICC=0.69\pm0.03$, averaged across three concentration signals) across sessions. Our results are compatible with the findings of previous RSFC reliability studies based on resting-state fMRI [51,52] and R-fNIRS [24] data. For example, Shehzad and colleagues [51] found
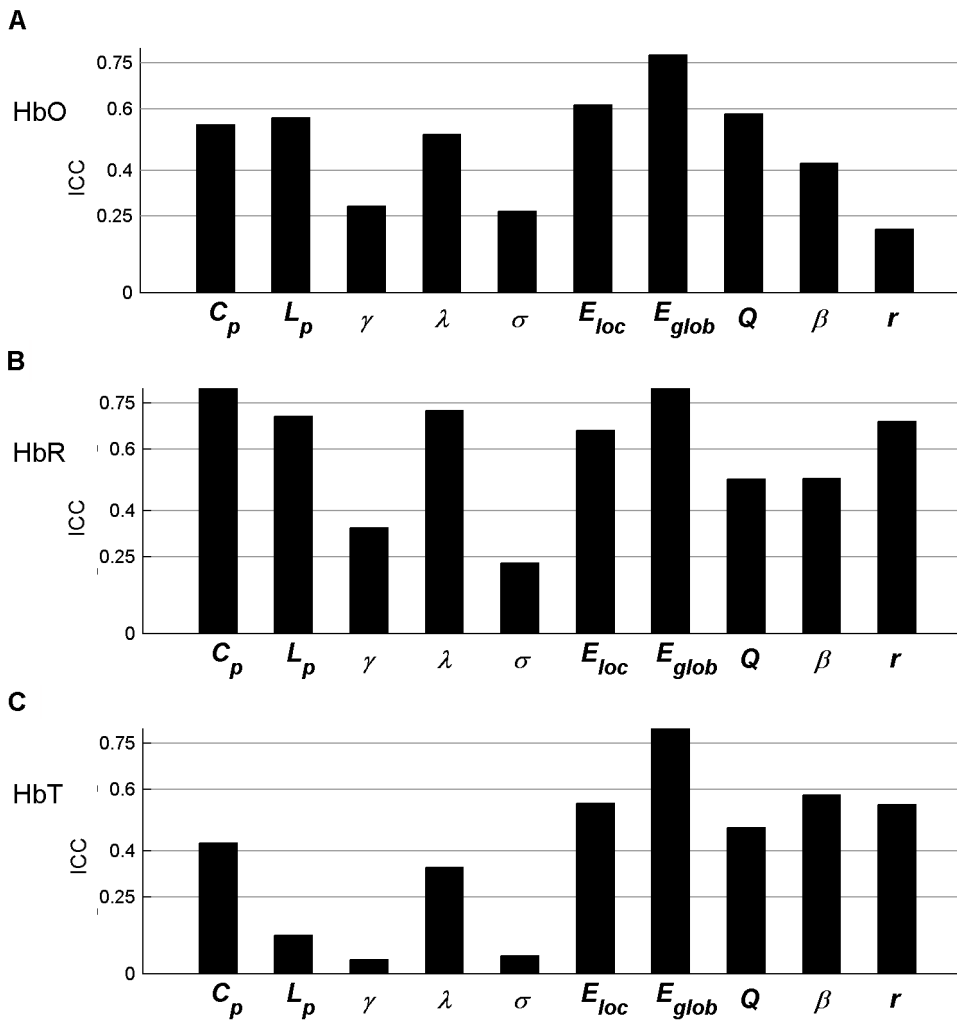
Figure 10. Threshold-independent reliability analysis of ICA- derived global network metrics. The areas under the curves (AUCs) of each global metric were used to provide threshold-independent reliability evaluation. (A–C) The global metric reliability was derived from HbO, HbR, and HbT, respectively. C_p, L_p, γ, λ, and σ denote the clustering coefficient, characteristic path length, normalized clustering coefficient, normalized characteristic path length, and small-world, respectively. E_{loc} and E_{glob} denote local efficiency and global efficiency, respectively. Q, β, and r denote modularity, hierarchy, and assortativity, respectively.
doi:10.1371/journal.pone.0072425.g010

that the fMRI-derived RSFC maps showed remarkable spatial consistency and modest to good reliability across different scanning sessions for both region-of-interest and voxel-wise analyses. The high reliability of RSFC maps were also further confirmed by Wang et al. [52] at a whole-brain level. Using R-fNIRS data, Zhang et al. [24] demonstrated that the RSFC of the sensorimotor regions was TRT reliable at both individual and group levels within and between sessions. In the present study we showed the high reliability of whole-brain RSFC maps based on the R-fNIRS data. Collectively, these accumulated evidences demonstrated that the measure of RSFC is reliable and trustworthy to measure the brain's intrinsic functional architecture regardless of the imaging techniques of fNIRS and fMRI. Additionally, we found positive correlation between RSFC reliability and connectivity strength for HbO and/or HbR. This finding suggests that the reliability of functional connectivity for these two concentration signals is partly determined by their connection strength.

TRT reliability of global network metrics

In recent years, TRT reliability of global network metrics has been examined using different imaging modalities, including MEG and fMRI techniques [52,53,54,55,56]. These research findings provide solid foundation for potential applied studies on longitudinal changes, measurements of normal development, or drug treatment-related effects. In the present study, we investigated the TRT reliability of global metrics of R-fNIRS brain networks from two perspectives. First, we examined the effect of varying threshold values on network reliability and found that the reliability of all global metrics is related to those selected threshold values. The threshold-sensitive characteristics of global network metrics are considered as typical concerns in brain network study, and have also been demonstrated in previous fMRI studies [52,53,54]. For example, Wang and colleagues [52] pointed out that the reliability of global network metrics varied from poor to excellent grades which depended on the selected threshold values. These observations suggest that it is extremely important to select an appropriate threshold for obtaining reliable results in the study of brain networks. However, it needs to notice that there is currently no
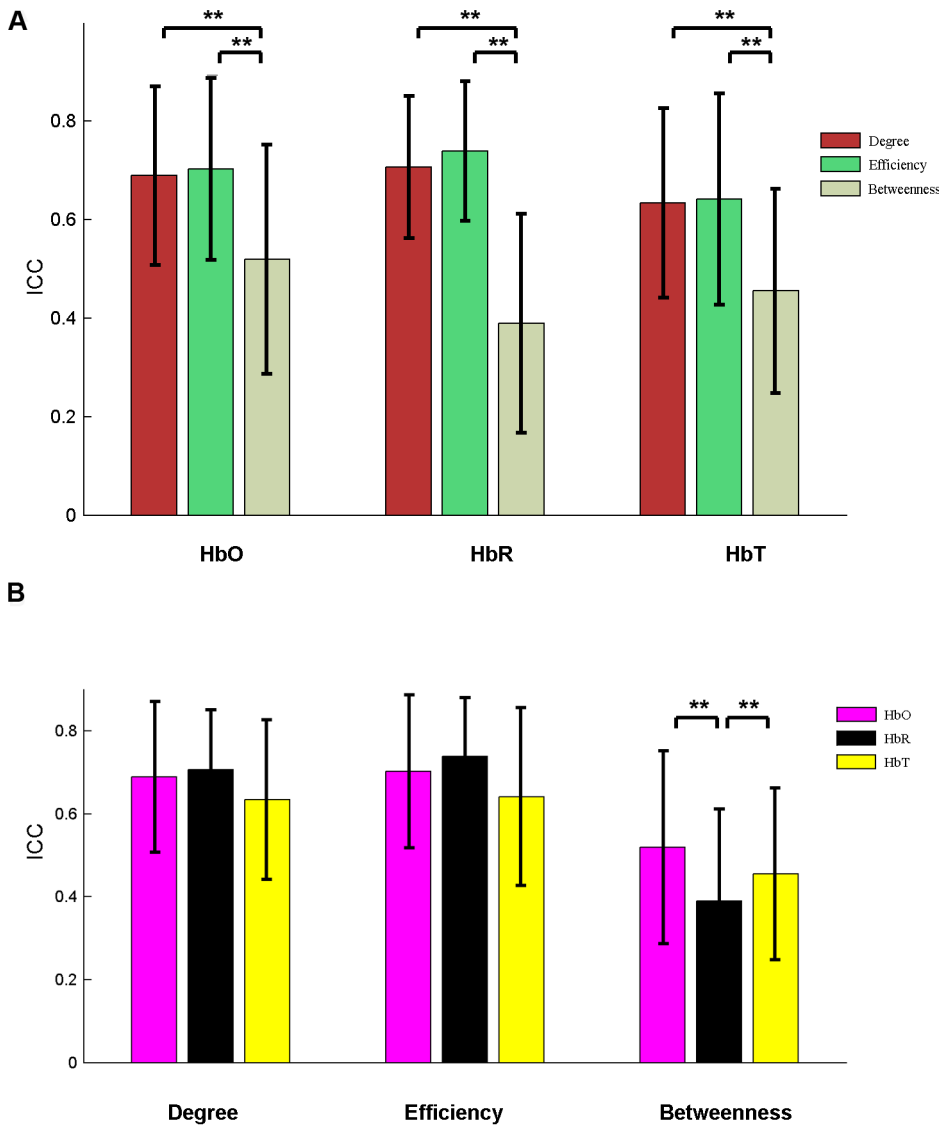
Figure 11. Significant differences (paired t-test) in ICA-derived nodal metric reliability. (A) Reliability for three nodal centrality metrics (degree, efficiency, and betweenness). (B) Reliability for three concentration signals (HbO, HbR, and HbT). Each of these three concentration signals was denoised using ICA. Note that the reliability of nodal degree and efficiency is concordant across HbO, HbR, and HbT and is also significantly ($p<0.0001$) higher than that of nodal betweenness. Error bars correspond to the standard deviation of the mean across the total nodes. The double asterisk indicates $p<0.0001$.
doi:10.1371/journal.pone.0072425.g011

accepted standard in the literature for determining the suitable threshold values applied to different network studies. Generally, a compromise method is to investigate the network over a continuous threshold range from the sparsest connection to the densest connection.

Second, we examined the TRT reliability of global network metrics by excluding the impact of thresholds and found that most global network metrics exhibited overall high TRT reliability. Of note, our observation is consistent with a previous report from Telesford et al., [56], who demonstrated the TRT reliability of graph metrics of functional human brain network based on a set of BOLD-fMRI data and also investigated an overall high TRT reliability on some global network metrics such as the clustering coefficient C_p, path length L_p, and global efficiency E_{glob}. However, it is worth noting that these findings are inconsistent with those shown in two previous BOLD-fMRI studies demonstrating low

[52] and modest [53] TRT reliability for the global network metrics, respectively. For example, Wang et al. [52] found that most of the global network metrics displayed overall low TRT reliability, especially for the clustering coefficient C_p that uniformly showed poor reliability ($ICC<0.25$) regardless of which type of network (binarized or weighted) adopted for that study. By contrast, Braun et al. [53] found that the reliability of global network metrics was moderate ($0.41< ICC <0.59$) but highly depended on the processing strategies such as the selection of frequency-band range and the use of global signal regression. For the discrepancies among these studies, we speculate that it could be attributable to the different imaging modalities, scanning intervals, network size, population size, and the node and edge definitions applied in these studies. Moreover, the resting-state imaging data captures the information of complex integration among various brain regions, and the integration always exhibits a
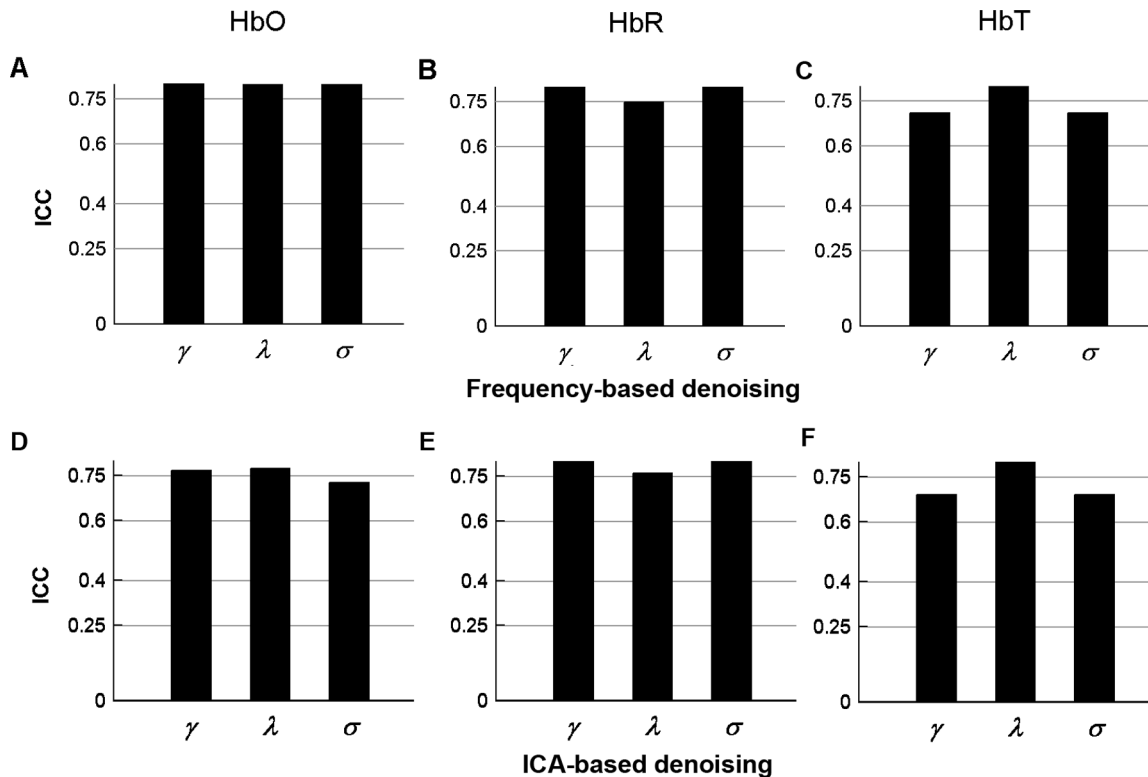
Figure 12. Reliability analysis of global network metrics derived from the correlation matrix randomization method. The areas under the curves (AUCs) of each global metric were used to provide threshold-independent reliability evaluation. (A–C) show the frequency-derived (i.e. band-pass filter) global metric reliability and (D–F) show the ICA-derived global metric reliability. γ, λ, and σ denote the normalized clustering coefficient, normalized characteristic path length, and small-world, respectively.
doi:10.1371/journal.pone.0072425.g012

dynamic, time-varying fashion on the order of second or minute [57]. In the future, it would be interesting to compare TRT reliability of graphic metrics of brain functional networks derived from the technique of simultaneous acquisition of different modalities (e.g., fNIRS vs. BOLD-fMRI). In addition, it is worth noting that there also exist several global network metrics exhibiting low reliability across different fNIRS signals (i.e., HbO, HbR, and HbT), such as the characteristic path length Lp (Fig.10C), hierarchy (Fig. 6A), assorsitivity (Fig. 10A), and modularity parameter Q (Fig.10C). The possible reasons for the low reliability global-metrics could be 1) that these network metrics depend on the fine structure of observed networks, while the fine structure could be sensitive to the subtle fluctuations of functional connections between two sessions; 2) that these network metrics characterizing aspects of intrinsic brain function that can be less robustly estimated using fNIRS signal(s) [53]; 3) that these network metrics manifested a low between-subject variance in differentiating subjects or a high within-subject variance across scanning sessions [52]. Moreover, we also found that the normalized metrics of normalized clustering coefficient γ (Figs. 6C and 10C), normalized characteristic path length λ (Figs. 6C and 10C) and small-world scale σ (Figs. 6C, 10B, and 10C) also exhibited low reliability across HbO, HbR and/or HbT signals. We speculated that the low reliability of these three network metrics might be related to the randomization approaches which were used to generate the matched random networks. To examine this point, we recalculated these three global metrics (normalized clustering coefficient, normalized characteristic path length, and small-worldness) using the correlation matrix randomization method

that preserves the transitive network structures [58]. Likewise, the reliability level of these three metrics was also quantified by *ICC*. We found that the reliability of these three metrics was largely improved (Fig. 12) as compared to the results obtained from that topology randomization method (Fig. 6 and Fig. 10). The improvement of reliability could be attributable to decreased within-subject variance due to the conservative estimation of global metrics induced by the correlation matrix randomization algorithm. This observation also suggested that it would be important to choose appropriate randomization methods for achieving reliable results of brain network analysis.

TRT reliability of nodal metrics

Similar to the findings in global network measures, we also observed that nodal reliability was threshold-sensitive for local nodal metrics. This finding further highlights the importance of threshold selection for achieving reliable results. Additionally, regarding to the reliability assessment of nodal centrality metrics, a previous fMRI study [52] has made a systematic investigation on 6 nodal metrics (including the nodal degree, efficiency, betweenness, cluster coefficient, participant coefficient, and normalized participant coefficient), and found that the nodal degree had the highest TRT reliability. In the current study, we found that in addition to the nodal degree, the nodal efficiency also exhibited fairly high reliability, and both were significantly higher than that reliability of betweenness regardless of the HbO, HbR or HbT signals. More importantly, the reliability of both nodal degree and nodal efficiency exhibited no significant differences across the different hemoglobin concentration signals, suggesting that both these

metrics are robust to different hemodynamic contrasts. Similarly, in the previous BOLD-fMRI study [52], the authors also found that the reliability of nodal network properties was minimally affected by many factors such as scanning time interval, network membership (networks excluding or including negative correlations) and network type (binarized or weighted networks). These findings indicated that the reliability of nodal properties (e.g., nodal degree and efficiency) was more robust and least affected, as compared to that of global network metrics. However, we also noticed that among the nodal centrality metrics, the nodal betweenness was the most sensitive to spatial nodal configuration and most unreliable across different concentration contrasts, which indicated that the nodal betweenness metric should be used with caution if no additional effort was made to improve its reliability. This observation from fNIRS data was also proved to be consistent with the previous BOLD-fMRI results [52].

Comparisons of graph metric reliability among the three hemoglobin concentrations

In the current study, we found that the three hemoglobin concentrations (HbO, HbR, and HbT) exhibited different reliability profiles for graph-based network metrics: the HbR showed the best global reliability (mean $ICC = 0.604$), while the HbT showed the worst reliability ($ICC < 0.4$ for most global metrics). In our previous study [12], we also demonstrated that although all the three types of concentration signals can be used to characterize topological organization of intrinsic functional brain architecture, the quantitative network parameters differed significantly. The discrepancies among different concentration signals results may be attributable to their different nature in hemodynamic responses to transient neural activity and/or the difference in signal-to-noise ratio in the R-fNIRS measurements [23,59]. While all of these three hemoglobin concentration signals are computed from the same raw R-fNIRS data, each signal reflects different hemodynamic response during neuronal activity. Under normal circumstances, the HbO signal reflects regional cerebral blood flow changes and the HbT signal reflects regional cerebral blood volume changes within the imaging field. HbR is generated through oxygen utilization in cerebral tissue. During fNIRS measurement, the signals recording the regional changes in cerebral blood flow and cerebral blood volume were generally aliased into various systemic physiological noises. The noise components were further propagated into different concentration signals during the calculation of hemoglobin concentration. It has been demonstrated that the HbR signal was generally minimally aliased into various noise components compared to the signals of HbO and HbT [60,61], which, therefore, could bring about higher reliability for HbR signal. By contrast, both the HbO and HbT signals were more affected by changes in physiological conditions such as heart rate, blood pressure, blood flow, and breathing cycle [48,60,62,63], which, to some extent, could reduce the reliability of the resultant network metrics.

In spite of the overall low reliability of global network metrics for HbT signal, we also noticed that several global metrics (e.g., hierarchy β and assortativity r) displayed fairly good reliability whereas they showed relatively lower reliability for the HbO and/or HbR signal (Fig. 6 and Fig. 10). This implies that the TRT reliability of some network metrics is specific to a certain concentration signal. Consequently, these findings highlight that fNIRS researchers should choose the optimum hemoglobin concentration contrast with caution for achieving reliable network investigation. Beyond the specificity of the high reliability of network properties on hemoglobin concentration, we also found that several network parameters (e.g., clustering coefficient C_p and

global efficiency measure E_{glob}) exhibited excellent reliability on average across the three concentration signals. This suggests that C_p and E_{glob} could be utilized as two reliable biomarkers of functional brain architecture. In addition to these global metrics, we also found two nodal metrics: nodal degree and efficiency, which had mean ICC values across three concentration signals larger than 0.65, displayed not only the best reliability among all network measures but also the most robust performance among all network measures but also the most robust performance among the different hemoglobin concentration signals. These reliable and robust features suggest these metrics as reliable candidates to reveal the topological organization of intrinsic functional networks in future fNIRS studies in the human brain.

Noise reduction in R-fNIRS data analysis

A common problem in R-fNIRS studies is that the spontaneous low-frequency fluctuations are contaminated by various different types of noise and artifacts such as instrumental noise, motion-induced artifacts, and physiological interferences which often occurs in the superficial layers of the head and in the brain tissue [62,64,65,66]. These noises could have different sources and possess various frequency distributions of spectra. A general way to reduce the effect of these kinds of noise is the utilization of a band-pass filter with cut-off frequency from 0.009 to 0.08 Hz that represents the frequency range of hemodynamic signals thought to emanate from spontaneous neural activity [23,27]. Using such a noise-reduction strategy, several R-fNIRS studies have explored intrinsic RSFC patterns in multiple brain networks involved in the sensorimotor [20,23,25], auditory [20], visual [23,25], and language systems [67] as well as the whole-brain functional network [19]. The results from these studies are proved to be compatible with each other [20,23,25] and with previous fMRI investigations [27,68,69]. It suggests that the low-frequency spontaneous fluctuations measured in R-fNIRS signals may reflect spontaneous brain activity in spite of various confounding origins (e.g., systematic interferences in the superficial layers from skin and the scalp of the head). Nonetheless, the superficial interferences have been considered an important noise source in R-fNIRS data [70]. Recently, several denoising methods have been proposed for separating the noise from neuronal signals, which includes short separation measurement [71,72], independent component analysis [48], and adaptive filter method [73,74]. These methods are playing crucial roles in reducing the noise effects on the signals of spontaneous neural activity. Among them, both the short-distance measurement and the adaptive filter methods have a common experimental consideration that they need a design of additional short channels to record the superficial signal. However, a recent study has pointed out that the location of short separation measurement strongly impacts the performance of noise reduction in fNIRS [75]. Thus, it would be important and necessary to consider the number and the location of short-separation channels while adopting these two methods. In contrast, the ICA is a pure data-driven approach and does not require specific considerations of experimental design. Due to the capability of blind source separation, ICA can separate various noise and artifacts from measurement data, which have been widely used in both task-based and resting-state fMRI studies. In this study, we used the ICA approach for denoising and found that the reliability of many network metrics was not improved after the denoising processing. One reason could be attributable to the intrinsic instability of these network metrics [53], as estimated using fNIRS signal(s). Meantime, the characteristic of blind source separation of ICA makes the separated source components lack of explicit physiological explanation, which could lead to the difficulties of

completely identifying and removing the noise components from the measured signals. Therefore, future studies are needed to test whether TRT reliability of brain functional networks could be improved by utilizing recording respiration and heart rates while collecting R-fNIRS data or other noise-reduction approaches (e.g., short-separation measurement).

Further considerations

Several issues and limitations of the present study should be further addressed. First, the present study focused mainly on the evaluation of short-term reliability of R-fNIRS brain network metrics. Of note, one previous BOLD-fMRI study [52] found that the scanning interval had an influence on the network reliability and that long-term scans provided a higher TRT reliability than short-term scans on some network metrics such as normalized characteristic path length λ. Therefore, it would be our interests to examine the long-term TRT reliability in future fNIRS brain network studies. Second, several potential factors may affect the assessment of reliability of R-fNIRS brain networks but have not been elucidated in the present study, which includes: variation in fNIRS measurement environment (e.g., natural light and dark conditions) [76], variation in probe configuration across different imaging systems [77], variation in probe placement across participants and across scanning sessions[78], and potential change in participants' mental states during resting-state scanning [51]. In the future, the effects of these factors need to be systematically and comprehensively investigated. Third, we limited our examination of TRT reliability to widely used network metrics (10 global and 3 nodal characteristics). However, there are other network metrics, such as synchronization and vulnerability, whose TRT reliability should be evaluated in the future. Finally, there are different preprocessing strategies in fNIRS data analysis. Previous BOLD-fMRI studies have demonstrated that several factors (e.g., global signal removal, frequency bands and correlation metrics) might affect the network reliability [52,55]. Future studies will be required to investigate the influence of different preprocessing techniques on network reliability within the context of R-fNIRS.

Conclusions

We comprehensively evaluated the TRT reliability of graph metrics of R-fNIRS brain networks. Our results demonstrate that the R-fNIRS brain networks show good TRT reliability on RSFC and fair to excellent reliability on most global metrics and two nodal network metrics (nodal degree and efficiency). These findings suggest that the topological metrics with high reliability in R-fNIRS brain networks are potential to be employed in future studies of developmental and clinical neuroscience. Future research should also focus on the investigation and improvement of those network metrics that display low reliability.

Supporting Information

Supporting Information S1 Figure S1, Spatial similarity of ICA-derived RSFC maps. Group-level RSFC maps for session 1 and session 2 and their Pearson correlation are displayed in the first to third columns. Figures (A) to (C) correspond to the RSFC

data derived from HbO, HbR, and HbT, respectively. High similarity between sessions was observed in both the qualitative visual inspection and quantitative correlational analysis. Figure S2, Reliability analysis of ICA-derived RSFC maps. The first to third columns correspond to the data derived from HbO, HbR, and HbT, respectively. (A, B) The TRT reliability of RSFC maps and their corresponding reliability distributions. The reliability displays approximately normal configuration for all 1035 (i.e., $46 \times 45/2$) connections. The connections exhibit good reliability across HbO (mean ICC values 0.63), HbR (0.68) and HbT (0.65). (C) The relationship between RSFC strength and reliability as assessed by scatterplots. Each dot represents the group-level RSFC strength and the corresponding ICC value at the same connections. The trend lines were obtained by a linear least-squares fit method. Significant ($p<0.05$) positive correlations were found for HbO signals, suggesting stronger RSFC leads to higher reliability for this signal. Figure S3, TRT reliability of ICA-derived global network metrics as a function of sparsity threshold. (A–C) The global metric reliability was derived from HbO, HbR, and HbT, respectively. Five colors correspond to five different reliability grades. The red, yellow, green, cyan, and blue colors represent excellent ($0.75< ICC <1$), good ($0.6< ICC <0.75$), fair ($0.4< ICC <0.6$), low ($0.25< ICC <0.4$), and poor ($ICC<0.25$) reliability of global network metrics, respectively. C_p, L_p, γ, λ, and σ denote the clustering coefficient, characteristic path length, normalized clustering coefficient, normalized characteristic path length, and small-world, respectively. E_{loc} and E_{glob} denote local efficiency and global efficiency, respectively. Q, β, and r denote modularity, hierarchy, and assortativity, respectively. Figure S4, TRT reliability of ICA-derived nodal centrality metrics as a function of sparsity threshold. (A–C) The nodal metric reliability was derived from HbO, HbR, and HbT, respectively. The five colors correspond to five different reliability grades: red, yellow, green, cyan, and blue represent excellent ($0.75< ICC <1$), good ($0.6< ICC <0.75$), fair ($0.4< ICC <0.6$), low ($0.25< ICC <0.4$), and poor ($ICC<0.25$) reliability of the nodal centrality metrics, respectively. Figure S5, Threshold-independent reliability analysis of ICA-derived nodal centrality metrics. The areas under the curves (AUCs) of each nodal metric were used to provide threshold-independent reliability evaluation. (A–C) The nodal reliability was derived from HbO, HbR, and HbT, respectively. Different colors in the nodes correspond to different reliability grades: red, yellow, green, cyan, and blue colors represent excellent ($0.75< ICC <1$), good ($0.6< ICC <0.75$), fair ($0.4< ICC <0.6$), low ($0.25< ICC <0.4$), and poor ($ICC<0.25$) reliability of the nodal centrality metrics, respectively. Table S1, Pearson correlations at individual-level ICA-derived RSFC maps between sessions. Table S2, Statistical comparisons of the ICA-derived global network metrics (across subjects) between sessions.
(DOC)

Author Contributions

Conceived and designed the experiments: HJN YH. Performed the experiments: HJN TDZ. Analyzed the data: ZL HJN. Contributed reagents/materials/analysis tools: XHL JHW NS XHZ YH. Wrote the paper: HJN.

References

1. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10: 186–198.
2. Bullmore E, Sporns O (2012) The economy of brain network organization. Nat Rev Neurosci 13: 336–349.
3. He Y, Evans A (2010) Graph theoretical modeling of brain connectivity. Current Opinion in Neurology 23: 341–350
4. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. NeuroImage 52: 1059–1069.

5. He Y, Chen ZJ, Evans AC (2007) Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI. Cerebral Cortex 17: 2407–2419.

6. Gong G, He Y, Concha L, Lebel C, Gross DW, et al. (2009) Mapping Anatomical Connectivity Patterns of Human Cerebral Cortex Using In Vivo Diffusion Tensor Imaging Tractography. Cerebral Cortex 19: 524–536.

7. Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey C, et al. (2008) Mapping the structural core of human cerebral cortex. PLoS biology 6: e159.

8. Hagmann P, Kurant M, Gigandet X, Thiran P, Wedeen VJ, et al. (2007) Mapping Human Whole-Brain Structural Networks with Diffusion MRI. PLoS ONE 2: e597.

9. Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E (2006) A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs. The Journal of Neuroscience 26: 63–72.

10. He Y, Dagher A, Chen Z, Charil A, Zijdenbos A, et al. (2009) Impaired small-world efficiency in structural cortical networks in multiple sclerosis associated with white matter lesion load. Brain 132: 3366–3379.

11. Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, et al. (2005) Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain. Cerebral Cortex 15: 1332–1342.

12. Niu H, Wang J, Zhao T, Shu N, He Y (2012) Revealing Topological Organization of Human Brain Functional Networks with Resting-State Functional near Infrared Spectroscopy. PLoS ONE 7: e45771.

13. Bassett DS, Bullmore ET, Meyer-Lindenberg A, Apud JA, Weinberger DR, et al. (2009) Cognitive fitness of cost-efficient brain functional networks. Proceedings of the National Academy of Sciences 106: 11747–11752.

14. Pievani M, de Haan W, Wu T, Seeley WW, Frisoni GB (2011) Functional network disruption in the degenerative dementias. The Lancet Neurology 10: 829–843.

15. Wang J, Zuo X, He Y (2010) Graph-based network analysis of resting-state functional MRI. Frontiers in Systems Neuroscience 4.

16. Xia M, He Y (2011) Magnetic Resonance Imaging and Graph Theoretical Analysis of Complex Brain Networks in Neuropsychiatric Disorders. Brain connectivity 1: 349–365.

17. Jobsis F (1977) Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198: 1264–1267.

18. Villringer A, Chance B (1997) Non-invasive optical spectroscopy and imaging of human brain function. Trends in Neurosciences 20: 435–442.

19. Homae F, Watanabe H, Otobe T, Nakano T, Go T, et al. (2010) Development of Global Cortical Networks in Early Infancy. The Journal of Neuroscience 30: 4877–4882.

20. Lu C-M, Zhang Y-J, Biswal BB, Zang Y-F, Peng D-L, et al. (2010) Use of fNIRS to assess resting state functional connectivity. Journal of Neuroscience Methods 186: 242–249.

21. Mesquita RC, Franceschini MA, Boas DA (2010) Resting state functional connectivity of the whole head with near-infrared spectroscopy. Biomed Opt Express 1: 324–336.

22. White BR, Liao SM, Ferradal SL, Inder TE, Culver JP (2011) Bedside optical imaging of occipital resting-state functional connectivity in neonates. Neuro-Image.

23. White BR, Snyder AZ, Cohen AL, Petersen SE, Raichle ME, et al. (2009) Resting-state functional connectivity in the human brain revealed with diffuse optical tomography. NeuroImage 47: 148–156.

24. Zhang H, Duan L, Zhang Y-J, Lu C-M, Liu H, et al. (2011) Test–retest assessment of independent component analysis-derived resting-state functional connectivity based on functional near-infrared spectroscopy. NeuroImage 55: 607–615.

25. Zhang H, Zhang Y-J, Lu C-M, Ma S-Y, Zang Y-F, et al. (2010) Functional connectivity as revealed by independent component analysis of resting-state fNIRS measurements. NeuroImage 51: 1150–1161.

26. Cope M, Delpy D (1988) System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. Medical and Biological Engineering and Computing 26: 289–294.

27. Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic Resonance in Medicine 34: 537–541.

28. Sasai S, Homae F, Watanabe H, Sasaki AT, Tanabe HC, et al. (2012) A NIRS–fMRI study of resting state network. NeuroImage 63: 179–193.

29. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences of the United States of America 102: 9673–9678.

30. Latora V, Marchiori M (2001) Efficient Behavior of Small-World Networks. Physical Review Letters 87: 198701.

31. Watts DJ, Strogatz SH (1998) Collective dynamics of ʾsmall-world/ʾ networks. Nature 393: 440–442.

32. Latora V, Marchiori M (2003) Economic small-world behavior in weighted networks. The European Physical Journal B – Condensed Matter and Complex Systems 32: 249–263.

33. Maslov S, Sneppen K (2002) Specificity and Stability in Topology of Protein Networks. Science 296: 910–913.

34. Humphries MD, Gurney K, Prescott TJ (2006) The brainstem reticular formation is a small-world, not scale-free, network. Proceedings of the Royal Society B: Biological Sciences 273: 503–511.

35. Newman MEJ (2006) Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103: 8577–8582.

36. Ravasz E, Barabási A-L (2003) Hierarchical organization in complex networks. Physical Review E 67: 026112.

37. Newman MEJ (2002) Assortative Mixing in Networks. Physical Review Letters 89: 208701.

38. Achard S, Bullmore E (2007) Efficiency and Cost of Economical Brain Functional Networks. PLoS Comput Biol 3: e17.

39. Linton CF (1977) A Set of Measures of Centrality Based on Betweenness. Sociometry 40: 35–41.

40. Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin 86: 420–428.

41. Sampat MP, Whitman GJ, Stephens TW, Broemeling LD, Heger NA, et al. (2006) The reliability of measuring physical characteristics of spiculated masses on mammography. British Journal of Radiology 79: S134–S140.

42. Winer BJ (1962) Statistical principles in experimental design. New York, NY, US: McGraw-Hill Book Company. x, 672 p.

43. Rousson V, Gasser T, Seifert B (2002) Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in medicine 21: 3431–3446.

44. Müller R, Büttner P (1994) A critical discussion of intraclass correlation coefficients. Statistics in medicine 13: 2465–2476.

45. Kong J, Gollub RL, Webb JM, Kong J-T, Vangel MG, et al. (2007) Test–retest study of fMRI signal change evoked by electroacupuncture stimulation. NeuroImage 34: 1171–1181.

46. He Y, Wang J, Wang L, Chen ZJ, Yan C, et al. (2009) Uncovering Intrinsic Modular Organization of Spontaneous Brain Activity in Humans. PLoS ONE 4: e5226.

47. Stone JV (2002) Independent component analysis: an introduction. Trends in Cognitive Sciences 6: 59–64.

48. Kohno S, Miyai I, Seiyama A, Oda I, Ishikawa A, et al. (2007) Removal of the skin blood flow artifact in functional near-infrared spectroscopic imaging data through independent component analysis. Journal of Biomedical Optics 12: 062111–062111.

49. Hyvarinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. Neural Networks, IEEE Transactions on 10: 626–634.

50. van de Ven VG, Formisano E, Prvulovic D, Roeder CH, Linden DEJ (2004) Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Human Brain Mapping 22: 165–178.

51. Shehzad Z, Kelly AMC, Reiss PT, Gee DG, Gotimer K, et al. (2009) The Resting Brain: Unconstrained yet Reliable. Cerebral Cortex 19: 2209–2229.

52. Wang J-H, Zuo X-N, Gohel S, Milham MP, Biswal BB, et al. (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.

53. Braun U, Plichta MM, Esslinger C, Sauer C, Haddad L, et al. (2012) Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. NeuroImage 59: 1404–1412.

54. Deuker L, Bullmore ET, Smith M, Christensen S, Nathan PJ, et al. (2009) Reproducibility of graph metrics of human brain functional networks. Neuro-Image 47: 1460–1468.

55. Liang X, Wang J, Yan C, Shu N, Xu K, et al. (2012) Effects of Different Correlation Metrics and Preprocessing Factors on Small-World Brain Functional Networks: A Resting-State Functional MRI Study. PLoS ONE 7: e32766.

56. Telesford QK, Morgan AR, Hayasaka S, Simpson SL, Barret W, et al. (2010) Reproducibility of graph metrics in fMRI networks. Frontiers in NEUROIN-FORMATICS 4: 1–10.

57. Chu CJ, Kramer MA, Pathmanathan J, Bianchi MT, Westover MB, et al. (2012) Emergence of Stable Functional Networks in Long-Term Human Electroen-cephalography. The Journal of Neuroscience 32: 2703–2713.

58. Zalesky A, Fornito A, Bullmore E (2012) On the use of correlation as a measure of network connectivity. NeuroImage 60: 2096–2106.

59. Gagnon L, Yücel MA, Dehaes M, Cooper RJ, Perdue KL, et al. (2012) Quantification of the cortical contribution to the NIRS signal over the motor cortex using concurrent NIRS-fMRI measurements. NeuroImage 59: 3933–3940.

60. Kirilina E, Jelzow A, Heine A, Niessing M, Wabnitz H, et al. (2012) The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. NeuroImage 61: 70–81.

61. Strangman G, Culver JP, Thompson JH, Boas DA (2002) A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. NeuroImage 17: 719–731.

62. Boas DA, Dale AM, Franceschini MA (2004) Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. NeuroImage 23 Suppl 1: S275–288.

63. Toronov V, Franceschini MA, Filiaci M, Fantini S, Wolf M, et al. (2000) Near-infrared study of fluctuations in cerebral hemodynamics during rest and motor stimulation: temporal analysis and spatial mapping. Medical physics 27: 801–815.

64. Diamond SG, Perdue KL, Boas DA (2009) A cerebrovascular response model for functional neuroimaging including dynamic cerebral autoregulation. Mathematical Biosciences 220: 102–117.

65. Obrig H, Neufang M, Wenzel R, Kohl M, Steinbrink J, et al. (2000) Spontaneous low frequency oscillations of cerebral hemodynamics and metabolism in human adults. NeuroImage 12: 623–639.

66. Payne SJ, Selb J, Boas DA (2009) Effects of Autoregulation and CO2 Reactivity on Cerebral Oxygen Transport. Annals of Biomedical Engineering 37: 2288–2298.

67. Zhang Y-J, Lu C-M, Biswal BB, Zang Y-F, Peng D-L, et al. (2010) Detecting resting-state functional connectivity in the language system using functional near-infrared spectroscopy. Journal of Biomedical Optics 15: 047003–047008.

68. Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proceedings of the National Academy of Sciences 100: 253–258.

69. Lowe MJ, Mock BJ, Sorenson JA (1998) Functional Connectivity in Single and Multislice Echoplanar Imaging Using Resting-State Fluctuations. NeuroImage 7: 119–132.

70. Cooper RJ, Gagnon L, Goldenholz DM, Boas DA, Greve DN (2012) The utility of near-infrared spectroscopy in the regression of low-frequency physiological noise from functional magnetic resonance imaging data. NeuroImage 59: 3128–3138.

71. Gagnon L, Perdue K, Greve DN, Goldenholz D, Kaskhedikar G, et al. (2011) Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. NeuroImage 56: 1362–1371.

72. Saager RB, Telleri NL, Berger AJ (2011) Two-detector Corrected Near Infrared Spectroscopy (C-NIRS) detects hemodynamic activation responses more robustly than single-detector NIRS. NeuroImage 55: 1679–1685.

73. Zhang Q, Brown EN, Strangman GE (2007) Adaptive filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study. Journal of Biomedical Optics 12: 044014–044014.

74. Zhang Q, Strangman GE, Ganis G (2009) Adaptive filtering to reduce global interference in non-invasive NIRS measures of brain activation: How well and when does it work? NeuroImage 45: 788–794.

75. Gagnon L, Cooper RJ, Yücel MA, Perdue KL, Greve DN, et al. (2012) Short separation channel location impacts the performance of short channel regression in NIRS. NeuroImage 59: 2518–2528.

76. Leff DR, Orihuela-Espina F, Elwell CE, Athanasiou T, Delpy DT, et al. (2011) Assessment of the cerebral cortex during motor task behaviours in adults: A systematic review of functional near infrared spectroscopy (fNIRS) studies. NeuroImage 54: 2922–2936.

77. Niu H, Khadka S, Tian F, Lin Z-J, Lu C, et al. (2011) Resting-state functional connectivity assessed with two diffuse optical tomography systems. Journal of Biomedical Optics 16: 046006–046005.

78. Hoshi Y (2007) Functional near-infrared spectroscopy: current status and future prospects. Journal of Biomedical Optics 12: 062106–062106.