NeuroImage 61 (2012) 1471–1483

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

One-year test–retest reliability of intrinsic connectivity network fMRI in older adults

Christine C. Guo [a], Florian Kurth [b], Juan Zhou [a], Emeran A. Mayer [b], Simon B. Eickhoff [c,d], Joel H. Kramer [a], William W. Seeley [a,*]

[a] Memory and Aging Center, UCSF Department of Neurology, University of California San Francisco, CA, USA
[b] Oppenheimer Family Center for Neurobiology of Stress, Departments of Medicine, Physiology and Psychiatry, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
[c] Institute for Neuroscience and Medicine (INM-1), Research Center Julich, Julich, Germany
[d] Institute of Clinical Neuroscience and Medical Psychology, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

ARTICLE INFO

Article history:
Accepted 6 March 2012
Available online 14 March 2012

ABSTRACT

"Resting-state" or task-free fMRI can assess intrinsic connectivity network (ICN) integrity in health and disease, suggesting a potential for use of these methods as disease-monitoring biomarkers. Numerous analytical options are available, including model-driven ROI-based correlation analysis and model-free, independent component analysis (ICA). High test–retest reliability will be a necessary feature of a successful ICN biomarker, yet available reliability data remains limited. Here, we examined ICN fMRI test–retest reliability in 24 healthy older subjects scanned roughly one year apart. We focused on the salience network, a disease-relevant ICN not previously subjected to reliability analysis, as well as the default mode network. Most ICN analytical methods proved reliable (intraclass coefficients > 0.4) and were further improved by wavelet analysis. Seed-based ROI correlation analysis showed high scan-wise reliability, whereas graph theoretical analysis and temporal concatenation group ICA proved most reliable at the individual unit-wise level (voxels, ROIs). Including global signal regression in ROI-based correlation analyses reduced reliability. Our study provides a direct comparison between the most commonly used ICN fMRI methods and potential guidelines for measuring intrinsic connectivity in aging control and patient populations over time.

Published by Elsevier Inc.

Introduction

Large-scale distributed neural networks organize healthy brain function and represent the selective targets of neurodegenerative illness (Seeley et al., 2009). "Resting-state" or task-free fMRI provides a new tool for examining intrinsic connectivity network (ICN) integrity and, potentially, for following networks over time (Achard and Bullmore, 2007; Damoiseaux et al., 2006; Fox and Raichle, 2007; Fox et al., 2005; Harrison et al., 2008). Across neuropsychiatric diseases, more sensitive and reliable longitudinal imaging biomarkers would increase the efficiency of drug discovery pipelines, most of which continue to rely on long-term clinical outcome measures. Emerging data suggest that task-free fMRI represents a promising method for tracking longitudinal change (Bai et al., 2011; Park et al., 2011), but rigorous evaluation is needed about the test–retest reliability of ICN measures.

A few recent studies have examined ICN fMRI reliability in healthy young subjects (Braun et al., 2011; Schwarz and McGonigle, 2011; Shehzad et al., 2009; Wang et al., 2011; Zuo et al., 2010). Collectively, these studies covered diverse analytical strategies, including the seed-based region-of-interest (ROI) approach, ROI correlation matrices, graph theoretical analyses, and independent component analysis (ICA). Across these strategies, ICN fMRI showed moderate to good reliability (intraclass correlation coefficients (ICCs) > 0.4) in healthy young adults re-scanned after one year. Some of these initial studies were conducted on the same publicly available dataset but examined different ICNs, making it difficult to compare results across methods and studies. To support neurodegenerative disease research, ICN fMRI reliability needs to be established in healthy older individuals, and lingering questions regarding optimal methods need to be resolved.

Here, we examined test–retest reliability of ICN fMRI in 24 healthy older adults scanned approximately one year apart. To provide a direct comparison between several related analytical strategies, we focused primarily on one ICN, the salience network (Seeley et al., 2007, 2009). This network is anchored by anterior cingulate and anterior insular cortices, which feature intrinsic connectivity to each other and to a host of subcortical, limbic, and autonomic controls sites whose co-activation may help to represent the emotional significance (i.e. salience) of ambient conditions (Seeley et al., 2007). This ICN has been widely replicated in the ICN literature and pertains to several neurological and psychiatric illnesses, including frontotemporal dementia, autism, and schizophrenia (Seeley et al., 2009; Uddin and Menon, 2009; White et al., 2010; Wiech et al., 2010; Zhou et al., 2010), but has not been the subject of previous reliability analyses. To enhance the generalizability of the salience network findings, we further examined reliability of the default mode

* Corresponding author at: Box 1207, San Francisco, CA 94143-1207, USA.
E-mail address: wseeley@memory.ucsf.edu (W.W. Seeley).

1053-8119/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.neuroimage.2012.03.027

network (DMN), a widely studied ICN associated with aging and Alzheimer's disease (Greicius et al., 2004; Zhou et al., 2010). The findings provide a valuable guide for ICN analysis in aging populations.

Materials and methods

Subjects

Subjects were recruited through the Hillblom Healthy Aging Network at the University of California, San Francisco (UCSF) Memory and Aging Center. A phone screen followed by a comprehensive neuropsychological assessment and a neurological exam was administered, and subjects were deemed eligible if they had a Clinical Dementia Rating (CDR) Scale total score of 0, a Mini Mental State Examination (MMSE) of 28 or higher, no significant history of neurological disease or structural lesion on MRI, and were psychoactive medication-free. Then, 24 (13 females) subjects were selected based on age (range 60–80) and the availability of two MRI scans roughly a year apart, each with a consensus diagnosis of cognitively normal within 180 days of each scan. Scans for each subject were separated by an average of 13 (s.d. 3) months. The mean age at the first scan was 66.7 (s.d. 6.4) years.

Image acquisition

All structural and functional images were acquired at the UCSF Neuroscience Imaging Center, on a 3 T Siemens Tim Trio scanner equipped with a 12-channel receiver head coil. A volumetric magnetization prepared rapid gradient echo (MP-RAGE) sequence was used to obtain T1-weighted images of the entire brain (TR/TE/TI = 2300/2.98/900 ms, flip angle of 9°, a bandwidth of 240 Hz/pixel, sagittal orientation with a FOV = 256 × 240 mm and 160 slices, voxel size = 1 mm^3). Task-free fMRI scans were obtained using 36 axial slices (3 mm thick with a gap of 0.6 mm) parallel to the plane connecting the anterior and posterior commissures and covering the whole brain using a T2*-weighted gradient echo–echo planar sequence (TR/TE = 2000/27 ms, flip angle 80°; FOV = 230 × 230 mm; matrix size: 92 × 92; in-plane voxel size: 2.5 × 2.5 mm). All subjects underwent 8 min of scanning (240 images) after being instructed only to remain awake with their eyes closed.

Image processing and analysis

After discarding the first 16 s to allow for magnetic field stabilization, functional images were realigned and unwarped, slice-time corrected, co-registered, spatially normalized to standard space and smoothed with a 4 mm full-width at half-maximum Gaussian kernel using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). Unwarping was performed to reduce artifacts due to movement-by-deformation interactions. Co-registration was performed between the mean T2* images and the subject's own T1-weighted image, and normalization was carried out by calculating the warping parameters between the subject's T1-weighted image and the Montreal Neurological Institute T1 image template and applying them to all functional images in the sequence. Subsequently, the functional images were re-sampled at a voxel size of 2 mm^3. These preprocessed images were then used for seed-based ROI, matrix, graph theoretical, and ICA analyses, as described in the following sections.

Motion parameters calculated during realignment were used to separate the 24 subjects into two subgroups. The maximum head motion in 3D space for each brain volume was computed as the root-sum-square of the three parameters for translation ($\sqrt{x^2+y^2+z^2}$) and rotation ($\sqrt{r_1{}^2+r_2{}^2+r_3{}^2}$) movements, respectively. Group 'mvmt<3' included the 20 subjects (mean ± s.d. age, 66.0 ± 6.6 years; 11 females) whose maximum head motion was less than 3 mm in translation and less than 3° in rotation during both scans, and group 'mvmt<2' included the 15 subjects (mean ± s.d. age, 65.1 ± 5.6 years; 7 females) whose maximum head motion was less than 2 mm in translation and 2° in rotation during both scans. Reliability analyses were performed separately on 'all subjects', 'mvmt<3' and 'mvmt<2' groups to assess the effects of motion on the reliability of ICN fMRI measures. We chose this grouping strategy in part to simulate what a researcher might do in practice by setting a movement threshold for subject exclusion.

Test–retest reliability

The reliability of each ICN fMRI measure was quantified by calculating the intra-class coefficient (ICC) across these measures derived from the two scans (McGraw, 1996; Shrout and Fleiss, 1979). A one-way ANOVA was applied to the measures of the two scan sessions across subjects, to calculate between-subject mean square (MS_p) and mean square error (MS_e). ICC values were then calculated as:

$$ICC = \frac{MSp - MSe}{MSp + (d-1)MSe}$$

where d = the number of observations per subject.

This form of the ICC measures the absolute agreement between the measures of the two different scan sessions and has been used in previous reliability analyses of fMRI data (Braun et al., 2011; Schwarz and McGonigle, 2011; Shehzad et al., 2009; Telesford et al., 2010; Wang et al., 2011; Zuo et al., 2010). For the present study, test–retest reliability was characterized as excellent (ICC > 0.8), good (ICC 0.6–0.79), moderate (ICC 0.4–0.59), fair (ICC 0.2–0.39) or poor (ICC < 0.2).

ICN fMRI measures

We compared the test–retest reliability of commonly used ICN analytical strategies, including three model-driven approaches, the seed-based ROI approach, ROI correlation matrix, and graph theoretical analyses, and two model-free approaches, ICA with template-matching and temporal concatenation group ICA with back reconstruction (Table 1). To make meaningful comparisons between methods, we assessed the reliability of these approaches in measuring functional connectivity within the same network — the salience network, an ICN anchored by bilateral anterior insular and anterior cingulate cortex with robust connectivity to subcortical and limbic structures (Seeley et al., 2007). We further assessed the reliability of seed-based ROI and ICA approaches within the DMN.

Seed-based ROI approaches

Previous studies identified the right frontoinsula as an effective ROI seed for deriving the salience network (Seeley et al., 2007;

Table 1
Terminology used to describe reliability for each analytical strategy.

	Scan-wise	Unit-wise
Seed-based	ROI connectivity map (mean of all gray matter voxels' z scores)	Voxel (z score)
ROI matrix	Matrix (mean of all pairs' z scores)	ROI pair (z score)
Graph	Graph (mean of each nodes' graph metrics: K/C/B)	ROI node (graph metric: K/C/B)
ICA	ICA component (mean of all gray matter voxels' z scores)	Voxel (ICA component z score)
TC-GICA	ICA component (mean of all gray matter voxels' z scores)	Voxel (ICA component z score)

Sridharan et al., 2008). Since our purpose was to assess reliability of seed-based ROI approaches more generally, we used four clusters within the anterior insula as salience network ROI seeds to ensure that our findings would not be limited to one particular ROI seed. These four anterior insula clusters were drawn from an activation likelihood estimate meta-analysis of task-based fMRI studies that activated the insula and represent foci related to cognitive and social-emotional paradigms. The cognitive clusters are located in the dorsal anterior insula (one per hemisphere), whereas the social-emotional clusters are positioned in the ventral anterior insula (also one per hemisphere — for further details, see Kurth et al., 2010).

The four anterior insula clusters were used as seeds in four separate seed-based ROI intrinsic connectivity analyses, following previous methods (Seeley et al., 2007). Briefly, temporal filtering was performed on each brain voxel using a band-pass filter (0.0083/s<f<0.15/s) to reduce the effect of low-frequency drift and high-frequency noise (Lowe et al., 1998). Then, the average voxel-wise time series from each ROI was detrended and used as a covariate of interest in a whole-brain, linear regression, statistical parametric analysis. This procedure generated a statistical parametric map from each scan session, where each voxel was scored based on its spontaneous BOLD signal correlation with the seed ROI used in the analysis (henceforth referred to as "connectivity"). Additionally, we manipulated several frequently used noise reduction strategies, including regression of white matter, CSF, non-brain (voxels that are not gray matter, white matter or CSF) and global signal time courses and motion parameters, to determine the impact of these nuisance regressors on reliability.

For every ICN derivation approach, we assessed reliability at the scan-wise and individual unit-wise levels. Specifically for the seed-based ROI approach, the scan-wise level was defined as the average connectivity across all gray matter voxels to the seed ROI, whereas the unit-wise level was defined as the connectivity of each individual voxel to the seed ROI. Therefore, reliability was calculated from the baseline and follow-up scans using (1) the mean connectivity score across all gray matter voxels, generating a single ICC representing the scan-wise level reliability and (2) the connectivity score of each gray matter voxel, generating ICCs for all voxels, which were then averaged to produce a single mean ICC representing the unit-wise level reliability (Table 1).

Previous studies suggest that voxels strongly connected to the ROI seed show higher reliability (Shehzad et al., 2009). To reassess this issue, we ranked all gray matter voxels based on their connectivity strength as determined through group-level analyses. For each of the four ROI seeds, statistical parametric maps from all scans (24 subjects × 2 scans = 48 scans) were entered into a second-level, random-effects analysis (Poline et al., 1997) to generate a group-level connectivity map. Age and gender were included as nuisance regressors. The values in the resulting group images were used to determine the connectivity strength for each voxel. Voxels within the top 25th percentile for connectivity strength were then used to assess the reliability of the most strongly connected voxels.

To examine reliability of DMN-related network measures, we investigated the seed-based ROI approach by creating a 4-mm radius spherical ROI in left posterior cingulate cortex (MNI coordinates: −6, −58, 28), the same region and coordinates used to derive the DMN in a previous reliability study of young adults (Shehzad et al., 2009).

ROI correlation matrix

A second major ROI-based approach employed in ICN fMRI analyses involves the construction of region-, cluster-, or voxel-level pairwise matrices, which represent the connectivity of each "unit" to every other unit in the matrix. As part of a separate ongoing study, we used the four connectivity maps derived from the four anterior insula ROIs described in the previous section to generate a group of "connectivity clusters". These four maps were entered into a full-factorial analysis, to identify brain areas with connectivity only to the dorsal anterior insula clusters, only to the ventral anterior insula clusters, or to both cluster pairs. The resulting 68 connectivity clusters, because they represent data-driven, connectivity-based contours, may offer advantages over standard spherical ROIs or landmark-based parcellation units often used in matrix-based analyses (Shirer et al., 2011). Connectivity was calculated between each ROI and the other 67 ROIs, generating 2278 unique ROI pair-wise connectivity measures.

As in the seed-based ROI approach, band-pass filtering (0.0083/s<f<0.15/s) was applied to the BOLD signal time courses for each voxel during preprocessing. Then, the mean time series from each ROI was extracted and detrended. Pearson (partial) correlation between the time series of each ROI pair was computed in Matlab, controlling for one or several of the confounding signals, including white matter, CSF, non-brain and global signal time courses and motion parameters. The Pearson correlation coefficients were then converted to z scores with Fisher's transformation for reliability analysis.

Again, we assessed test–retest reliability of the ROI connectivity matrix at the scan-wise and individual unit-wise levels. Here, the scan-wise level was defined as the average connectivity across all 2278 ROI pairs within the matrix, whereas the unit-wise level was defined as the connectivity of each individual ROI pair. Therefore, reliability was calculated from the baseline and follow-up scans using (1) the mean connectivity score across the matrix of ROI pairs, generating a single ICC representing the scan-wise level reliability and (2) the connectivity score of each ROI pair, generating 2278 ICCs, which were then averaged to produce a single mean ICC representing the unit-wise level reliability (Table 1).

To test whether ROI pairs with stronger connections were more reliable, we ranked the matrix ROI pairs based on their connectivity strength. A group-level connectivity matrix was determined by averaging ROI connectivity matrices from all scans (24 subjects × 2 scans = 48 scans). The values in the resulting group-level matrix were used to determine the ROI pairs in the top 25th percentile for connectivity strength, which were then used to assess the reliability of the most strongly connected ROI pairs. To visualize the reproducibility of ROI matrix at the group level, two connectivity matrices were derived for baseline and follow-up by averaging ROI connectivity matrices across the baseline and follow-up scans of all subjects (Fig. 3A).

Wavelet transformation of ROI time series

Wavelet transformation produces a time-scale decomposition that partitions the total energy of a signal into different scale components, corresponding to certain frequency ranges (see Bullmore et al., 2004 for a review). Wavelet analysis is particularly well suited to analysis of signals that have fractal scaling or 1/f properties, as is typical of fMRI BOLD timeseries at rest (Maxim et al., 2005). Here, we performed wavelet analysis on the 68 salience network ROI clusters used to construct the ICN matrix described in the previous section, using the WMTSA Wavelet Toolkit for MATLAB (http://www.atmos.washington.edu/~wmtsa/). We applied a maximum overlap discrete wavelet transform (MODWT) with Haar scaling filter to each ROI timeseries during each scan session to obtain the contributing signal in the following four frequency components: scale 1 (0.15 to 0.25 Hz), scale 2 (0.08 to 0.15 Hz), scale 3 (0.03 to 0.08 Hz) and scale 4 (0.01 to 0.03 Hz, Percival and Walden, 2000). In other words, each ROI timeseries was transformed to four series of wavelet coefficients at the four scales. Then, Pearson (partial) correlation was computed between the wavelet coefficients of each ROI pair at each scale. Thus, each ROI pair's connectivity was measured at each of the four scales and thus four wavelet correlation matrices were

generated corresponding to the four scales. Additionally, wavelet coefficients of one or several of the confounding signals, including white matter, CSF, non-brain and global signal time courses and motion parameters, were removed in the partial correlation analyses. The Pearson correlation coefficients were then converted to z scores with Fisher's transformation for reliability analysis.

Test–retest reliability analysis and group-level analysis were performed on the wavelet correlation matrices in the same manner as the original ROI matrices, separately at the four scales. Since wavelet-transformed ROI matrix analysis was most reliable at scale 3, as expected based on the frequency spectrum associated with ICNs, only the reliability results from scale 3 are presented in the main text.

Graph theoretical analysis

Graph theoretical analyses were performed using the adjacency matrices generated from the ROI matrix approach as described in previous sections, by applying a series of thresholds (from 0 to 1 in increments of 0.05) to determine the presence of connections between ROIs. In these adjacency matrices, each of the 68 ROIs represented one network node, and the connectivity between each pair of ROIs represented an edge in the network. Both binary and weighted adjacency matrices were analyzed for reliability. To generate binary matrices, edges with suprathreshold correlation coefficients were considered connected and assigned values of 1, whereas edges with subthreshold coefficients were considered not connected and assigned values of 0. To generate weighted matrices, each suprathreshold edge retained its correlation coefficient as its edge weight, whereas subthreshold edges below threshold were assigned values of 0.

The following graph metrics were calculated for each node in the salience network matrix: degree (K), clustering coefficient (C), and betweenness centrality (B), using the Brain Connectivity Toolbox (Rubinov and Sporns, 2010) and matlab_bgl within the Matlab environment. Degree measures the connectivity of each node, calculated as the sum of number/weight of links connected to each node. Clustering coefficient measures local neighborhood connectivity, calculated as the fraction of a node's neighbors that are neighbors of each other. Betweenness centrality measures node centrality, calculated as the fraction of all shortest paths in the network that contain a given node.

These graph metrics were selected because 1) they assign value to each individual node, allowing us to determine reliability at both the scan-wise and unit-wise levels and 2) they can be applied to both weighted and binary matrices, allowing us to compare reliability between these matrix types across different thresholds.

We assessed test–retest reliability for each graph metric at the scan-wise and individual unit-wise levels. Here, the scan-wise level was defined as the average graph metric across all 68 nodes within the graph, whereas the unit-wise level was defined as the graph metric for the individual node. Therefore, reliability was calculated from the baseline and follow-up scans using (1) the mean graph metrics across the graph, generating a single ICC representing the scan-wise level reliability and (2) the graph metric of each node, generating 68 ICCs, which were then averaged to produce a single mean ICC representing the unit-wise level reliability (Table 1).

Independent component analysis (ICA) with template matching

Spatial probabilistic ICA was used to isolate ICN maps following previous methods (Beckmann and Smith, 2004; Zhou et al., 2010). Briefly, preprocessed images from each scan were concatenated into one 4D file and entered into FSL 4.0 Melodic ICA software (http://www.fmrib.ox.ac.uk/fsl/ index.html). In separate subanalyses, we allowed the program to determine the number of components for each data set automatically (estimated component number ranged from 27 to 70, median 36) or fixed the component number to 20, as used in a previous study of reliability with this method (Zuo et al., 2010). Next, we used an automated template matching procedure to obtain subject-specific best-fit ICN maps for the salience network (Seeley et al., 2007, 2009). Goodness-of-fit was calculated by comparing each component from each subject to binarized group ICA maps of the salience network built from 15 healthy young subjects (ages 19–40; mean age, 26.5 years; nine females, all right-handed) from a separate dataset (Habas et al., 2009). Within the selected ICA component for the salience network, each voxel was given a z-score, which reflects the degree to which its timeseries correlates with the overall ICA component timeseries.

For the ICA with template matching approach, we assessed test–retest reliability at the scan-wise and individual unit-wise levels. Here, the scan-wise level was defined as the average z-score across all gray matter voxels within the whole-brain map, whereas the unit-wise level was defined as the individual voxel's z-score. Therefore, reliability was calculated from the baseline and follow-up scans using (1) the mean z-score across all gray matter voxels, generating a single ICC representing the scan-wise level reliability and (2) the z-score of each gray matter voxel, generating ICCs for all voxels, which were then averaged to produce a single mean ICC representing the unit-level reliability (Table 1).

To assess whether voxels more significantly belonging to the ICA-derived salience network component had higher reliability, we ranked all gray matter voxels based on their connectivity (z-score), as determined from a group-level analysis in which the selected ICA components for each scan were entered into second-level, random-effects analyses (Poline et al., 1997) to generate the salience network at the group level (24 subjects $\times$ 2 scans $=$ 48 scans). Age and gender were included as nuisance regressors. The resulting group images were used to determine the voxels in the top 25th percentile for connectivity within the ICA-derived salience network component.

ICA components for DMN were identified using a mirror approach to that used for the salience network. That is, components were matched to visually selected templates derived from a separate dataset (Habas et al., 2009) and analyzed using otherwise identical methods.

Temporal concatenation group ICA

Temporal concatenation group ICA with back reconstruction (TC-GICA) was performed using the Group ICA of fMRI Toolbox (GIFT), implemented in Matlab. Briefly, TC-GICA involves three stages: data reduction, ICA, and back reconstruction (Calhoun et al., 2004). First, principal components analysis (PCA) was used to reduce the data dimensionality for each subject. After each subject's functional data were reduced, the data were concatenated and entered into a second data reduction step using PCA. Second, the reduced, group-concatenated data were entered into the ICA algorithm to calculate spatially independent components. The number of components were either estimated by the program or set to 20, following the previous reliability study to examine ICA methods (Zuo et al., 2010). The estimated component number was 32, 31, 31 for all, mvmt$<$3, mvmt$<$2 groups, respectively. Third, back reconstruction was performed using GICA3, which previous studies have suggested may improve accuracy of estimating subject-specific spatial maps and time courses over the original GICA and spatio-temporal regression (dual-regression) (Erhardt et al., 2010). The group-level ICA components corresponding to salience network and DMN were selected by visual inspection and confirmed using the same template-matching procedure employed for individual ICA.

We again assessed test–retest reliability at the scan-wise and unit-wise levels using each voxel's TC-GICA z-scores, in similar manner as

described for ICA with template matching (Table 1). Then, to assess whether voxels that more significantly belonged to the ICA network showed higher reliability, we ranked all gray matter voxels based on the group-level significance scores produced by TC-GICA to identify the voxels in the top 25th percentile of z-scores.

Statistical tests

Statistical tests were carried out in Graphpad Prism 5 (http://graphpad.com/prism/Prism.htm). To test the significance of nuisance regressors' impact on scan-wise reliability for the seed-based ROI approach, paired t-tests (d.f. = 2) were used to compare scan-wise ICC values with and without the given nuisance regressor across the ROI analyses for each of the four seed ROIs. To test the significance of nuisance regressors' impact on unit-wise reliability for the seed-based ROI approach, paired t-tests were used to compare the mean gray matter voxel-wise ICC values with and without the given nuisance regressor across the four ROI analyses.

Results

Twenty-four older controls underwent task-free fMRI at two time points, 300–400 days apart. Salience network intrinsic connectivity was determined using three model-driven approaches, the seed-based ROI, ROI correlation matrix, and graph theoretical analyses, and two model-free approaches, ICA with template-matching and temporal concatenation group ICA with back reconstruction (TC-GICA). Test–retest reliability was assessed at the scan-wise level, using summary measures of network connectivity from each scan session, or at the individual unit level, using connectivity measures of each individual unit (voxel, node-pair, or node), which were then averaged to reflect reliability across the individual units of analysis (Table 1).

Seed-based ROI analysis provides good scan-wise reliability

We used four anterior insula ROIs (Kurth et al., 2010) to generate four intrinsic connectivity maps closely related to the salience

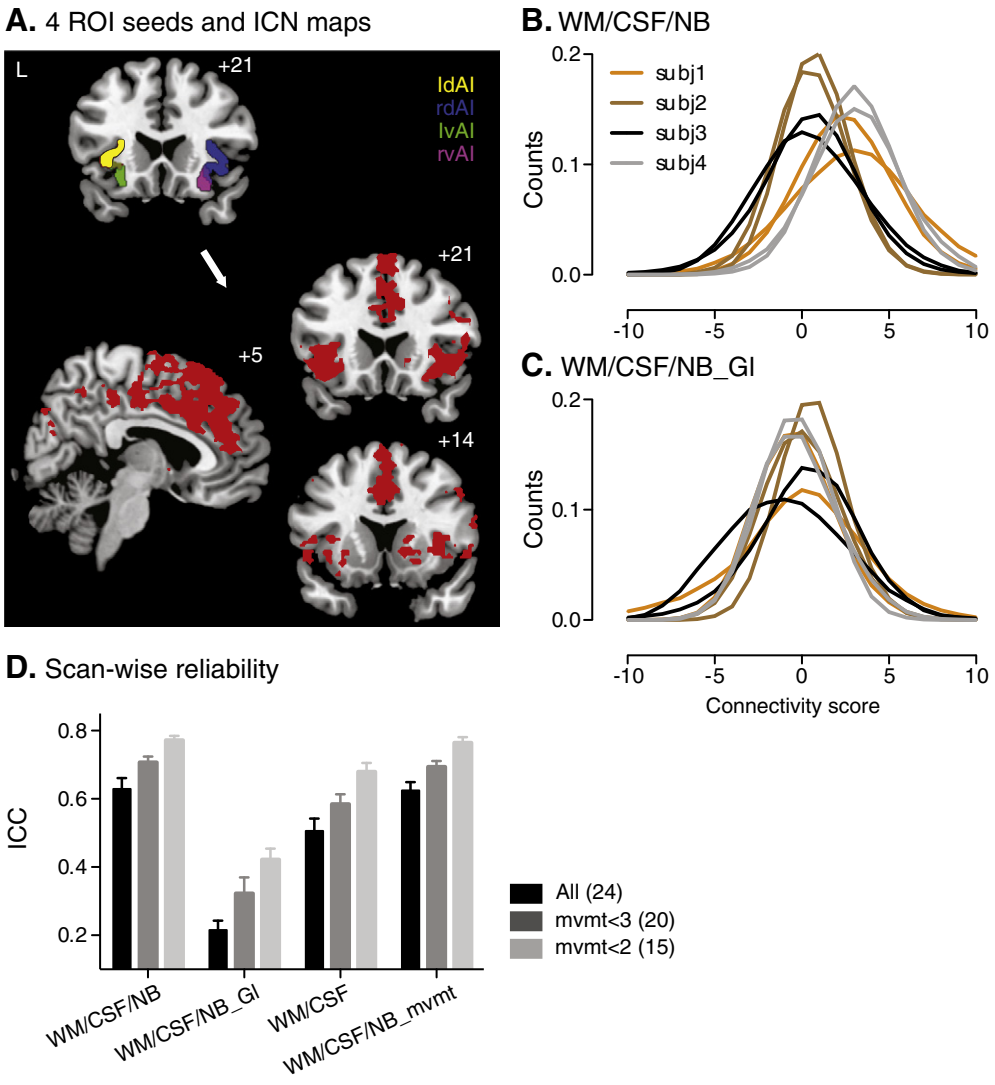

Fig. 1. ICNs generated by seed-based ROI analysis and their scan-wise reliability. A. The four anterior insular (AI) clusters used as ROI seeds (upper left image; left dorsal AI, ldAI — yellow; right dorsal AI, rdAI — blue; left ventral AI, lvAI — green; right ventral AI, rvAI — magenta) and the overlap of their corresponding group-level functional connectivity maps, threshold at $t > 7$, cluster size > 300 voxels (lower images). B. Distributions of gray matter voxels' connectivity scores from the ROI analysis using right ventral AI as the seed. White matter, CSF and non-brain signals were included as nuisance regressors. For visual clarity, data from four representative subjects are shown. Lines of the same color represent data from the two scans of the same subjects. C. Same as B, except the global signal was included as a nuisance regressor, in addition to white matter, CSF and non-brain. Data are from the same four subjects as in B. D. Mean scan-wise ICC across the four ROI analyses for all subjects (black), mvmt < 3 (gray) and mvmt < 2 (light gray) groups. Number of subjects in each group is indicated in parentheses. Error bars signify s.e.m. across the four ROI analyses. Nuisance regressors included in the analyses are indicated along the x-axis. WM, white matter; NB, non-brain; GI, Global signal; mvmt, movement parameters.

network. These maps showed overlapping ICNs across key network regions as reported previously, including bilateral anterior insula, anterior cingulate cortex, striatum and additional frontal and temporal areas (Fig. 1A, red color map representing the overlap of the four connectivity maps at the group level).

To assess the scan-wise reliability of salience network connectivity, we calculated the mean connectivity score across all gray matter voxels to the four ROI seeds for each scan and used these mean connectivity measures to compute the scan-wise ICC across the two scans. Using the ROI analysis with the right ventral anterior insula ROI seed as an example, the distribution of voxel connectivity scores was right-shifted toward positive values (Fig. 1B). More importantly, these connectivity distributions varied more between subjects than they did within subjects and between scans. All four seed-based ROI analyses showed good reliability (Fig. 1D), with a mean ICC = 0.63 (s.e.m. = 0.06), suggesting that seed-based ROI approaches were reliable at the overall network level.

Subject motion during scanning had a robust and graded impact on reliability when using the seed-based ROI approach; this effect was consistent across ROI analyses using all four ROI seeds. Reliability was higher for scans with less motion — ICC reached nearly 0.8 for group 'mvmt<2'compared with 0.6 in the combined group that included subjects with greater head motion (Fig. 1D).

Unit-wise reliability of seed-based ROI analysis: improved reliability among voxels with higher connectivity

Test–retest reliability of the seed-based ROI analyses varied widely at the voxel level, with a distribution that spanned from poor to good (Fig. 2A, black). ICC was calculated for each voxel based on its connectivity to the seed ROI at baseline and follow-up. These unit (voxel)-wise ICC distributions showed modal values in the fair range for all four anterior insula seeds, with mean ICCs of 0.3–0.4 (Fig. 2A, black) across the four seeds, when WM, CSF and non-brain signals were removed as nuisance regressors. Across the four ROI analyses, roughly 35% of the voxels showed moderate reliability or better (ICC > 0.4).

Voxels with stronger connectivity to the ROI seeds showed greater reliability. For each of the four ROI seeds, gray matter voxels were ranked based on connectivity strength. Voxels in the top 25th percentile for connectivity showed ICCs roughly 10% higher than the overall voxel-wise ICC average (Fig. 2B).

Similar to seed-based ROI scan-wise reliability, subject motion greatly reduced unit-wise reliability (Fig. 2B).

Effects of global signal and other nuisance regressors on reliability

Including the global signal as a regressor altered the distribution of functional connectivity computed with the seed-based ROI approach and had a striking impact on scan-wise and unit-wise reliability. Consistent with previous studies (Braun et al., 2011; Murphy et al., 2009; Schwarz and McGonigle, 2011), we found that removing the global signal introduced negative correlations and reduced positive correlations (compare Figs. 1B,C). Comparing the distributions of connectivity strength with and without global signal as a nuisance regressor, including global signal regressor created a more symmetric distribution centered at 0 (Fig. 1C). More importantly, including the global signal significantly reduced both scan-wise and unit-wise ICCs, by more than 50% (Figs. 1D and 2B; $p<0.0001$, paired t test).

Other regressors in the model also significantly influenced reliability, yet to a much lesser extent (Figs. 1D and 2B; $p<0.05$, paired t test). Seed-based ROI analysis was most reliable when including WM, CSF and non-brain signal but not the global signal regressors (Figs. 1D and 2B). On the other hand, although subjects with greater head movement showed lower reliability, including movement parameters as nuisance regressors did not significantly improve reliability nor did it mitigate reliability differences seen among groups with more or less head movement (Fig. 1D, $p = 0.67$; Fig. 2B, $p = 0.26$; paired t test).

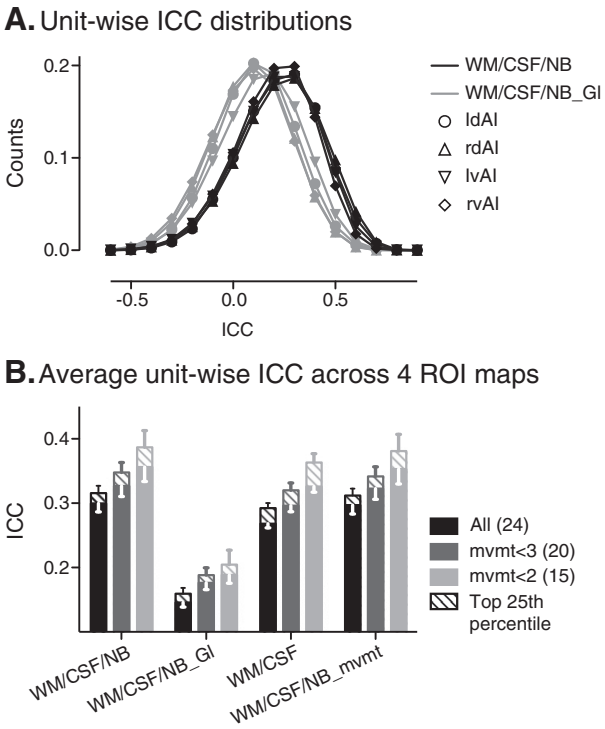

Fig. 2. Unit-wise reliability of seed-based ROI analysis. A. Distribution of ICC values of all gray matter voxels from each of the four ROI analyses, with white matter, CSF and non-brain nuisance regressors (black) and with global signal regressor added (gray). B. Mean unit-wise ICC across the four ROI analyses for all subjects (black), mvmt<3 (gray) and mvmt<2 (light gray) groups. Number of subjects in each group is indicated in parentheses. Error bars signify s.e.m. of the mean ICC values across the four ROI analyses. Nuisance regressors included in the analyses are indicated along the x-axis. Mean unit-wise ICCs of voxels in the top 25th percentile for connectivity are shown as striped bars. WM, white matter; NB, non-brain; GI, global signal; mvmt, movement parameters.

ROI correlation matrix analysis showed moderate scan-wise reliability

While the seed-based ROI approach examines connectivity of each voxel to a single ROI, ROI correlation matrix analysis allows researchers to explore connectivity between all regions within a network. We evaluated the reliability of an ROI correlation matrix composed of 68 salience network ROIs, derived from the same four ROI seeds used in the seed-based ROI analyses. As was seen for the single seed-based analyses, in ROI matrix analyses, the highest reliability was achieved by including WM, CSF and non-brain signal regressors and omitting the global signal and movement parameters. Therefore, for simplicity, all subsequent results are described and illustrated with those settings applied.

Overall, scan (matrix)-wise analysis showed moderate-to-good test–retest reliability. For each ROI pair, we computed the partial correlation coefficient (Pearson's r) between their BOLD signal time courses (controlling for nuisance covariates). The qualitative stability of these correlation matrices was evident at the group level, with baseline and follow-up scans showing convergent group-averaged matrices (Fig. 3A). The Pearson's r distributions were right-shifted toward positive values (Fig. 3B). Fig. 3B shows these correlation distributions for four representative subjects' baseline and follow-up scans, which again exhibit greater between-subjects than within-subjects variance. ROI matrix analysis showed moderate to good scan-wise reliability (ICC = 0.49, Fig. 3C), which is lower than the

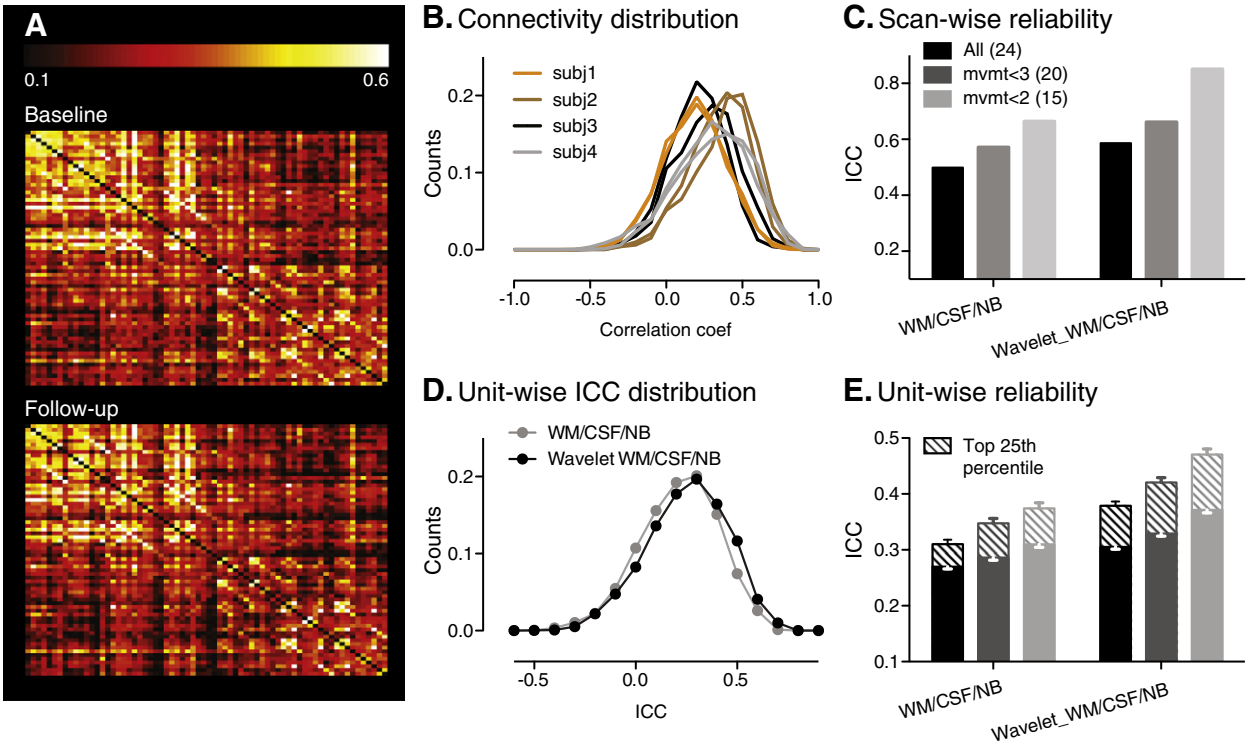

Fig. 3. Reliability of ROI matrix analysis. A. Group-level ROI connectivity matrix averaged from all subjects' baseline (top) and follow-up (bottom) scans. B. Distribution of correlation coefficients between all ROI pairs in the salience network matrix. For visual clarity, data are shown for four representative subjects. C. Scan-wise ICCs from ROI matrix analysis on all subjects, mvmt<3 and mvmt<2 groups, without and with wavelet transformation at scale 3. D. Distribution of ROI matrix unit-wise ICCs without (gray) and with (black) wavelet transformation at scale 3. E. Mean unit-wise ICCs of ROI matrix analysis on all subjects (black), mvmt<3 (gray) and mvmt<2 (light gray) groups, without and with wavelet transformation at scale 3. Error bars signify s.e.m. across the ROI pairs examined. Mean unit-wise ICCs of voxels in the top 25th percentile for connectivity are shown as striped bars. WM, white matter; NB, non-brain; wavelet, wavelet transformation at scale 3.

good to nearly excellent scan-wise reliability seen with the seed-based ROI approach (Fig. 1D).

We assessed unit-wise reliability by calculating ICC for each ROI pair. This approach revealed a wide ICC distribution across pairs in the matrix (Fig. 3D, gray). The distributions of unit (pair)-wise ICC were skewed toward positive values, with mean ICC of 0.2–0.3 (Fig. 3E, left columns). Around 25% ROI pairs had moderate reliability (ICC>0.4). ROI pairs with stronger connectivity showed greater reliability. Unit-wise ICCs of the top 25th percentile for connectivity showed 15% higher ICCs than the overall average (Fig. 3E, left columns).

Wavelet analysis increased ROI matrix reliability

ROI matrix reliability improved substantially when ROI timeseries correlations were computed after wavelet transformation. We applied maximal overlap discrete wavelet transform (MODWT) to the BOLD time courses of the 68 matrix ROIs to extract frequency-dependent correlation matrices. First, ROI time courses were transformed to wavelet coefficients at four scales, capturing signals within the corresponding four frequency domains. Then, partial correlation coefficients were computed between matrix ROIs' wavelet coefficients to produce four wavelet matrices corresponding to the four frequency domains (see Materials and methods for details). We found that the salience network matrix was most reliable at scale 3 (0.03 to 0.08 Hz), with 10–20% reduction in ICC measures at other scales, consistent with previous reports that neural intrinsic connectivity is represented by low frequency BOLD signal fluctuations (<0.1 Hz). Since our study focuses on reliability rather than wavelet analysis per se, only results from wavelet transformation at scale 3 are presented here.

Scan (matrix)-wise ICC scores were consistently higher after wavelet transformation across the three levels of subject motion (Figs. 3C,D). In particular, scan-wise reliability for subjects with less than 2 mm of motion amplitude was excellent (ICC = 0.84; Fig. 3D). Furthermore, wavelet transformation improved unit (pair)-wise reliability, improving the ICCs by around 20% across groups (Fig. 3E).

Graph theoretical measures showed good reliability at the scan-wise and individual unit-wise levels

We further examined the reliability of graph theoretical measures calculated from the salience network matrix. To derive graphs and graph metrics, we hard-thresholded the salience network ROI correlation matrix (from 0 to 1 in increments of 0.05) to generate binary and weighted adjacency matrices (see Materials and methods for details). Each ROI represented a node in the network, and the connectivity between each ROI pair represented an edge. Here, we present reliability analysis for weighted and unweighted degree and clustering coefficient, two of the most common graph metrics examined in fMRI analyses. Degree and clustering coefficient were the most reliable and representative in our reliability analysis (Fig. 4). Betweenness centrality, in contrast, had poor reliability (Supplementary Fig. 1).

We assessed scan-wise reliability by calculating graph-wise ICCs for each graph metric. We found moderate ICCs for unweighted and weighted graphs, which were stable (~0.5) across a range of connectivity thresholds from 0.05 to 0.45, although clustering coefficients showed slightly lower reliability with more stringent matrix thresholding (Fig. 4A). At even higher thresholds (sparser adjacency matrices), ICCs decreased in general, consistent with a recent analysis of graph metric reliability in young controls (Braun et al., 2011). Alternative thresholding methods, such as soft-thresholding and proportional-thresholding, did not substantially impact reliability.

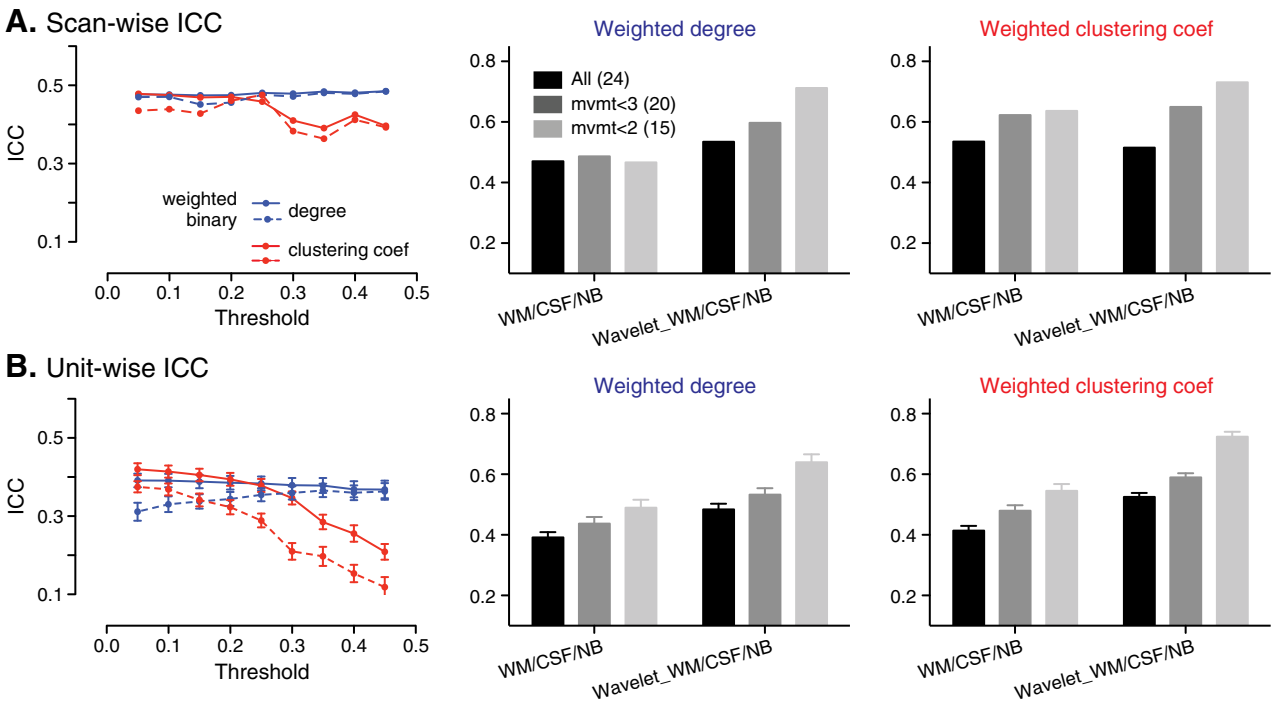

Fig. 4. Reliability of graph theoretical analysis. A. left panel, scan-wise ICCs across different thresholds, for degree (blue) and clustering coefficient (red) from weighted (solid line) and binary (dashed line) adjacency matrices. Scan-wise ICCs of degree (middle panel) and clustering coefficient (right panel) from the weighted matrix thresholded at $r > 0.05$, for all subjects (black), mvmt < 3 (gray) and mvmt < 2 (light gray) groups, without and with wavelet transformation at scale 3. B. left panel, mean unit-wise ICCs across different thresholds, for degree (blue) and clustering coefficient (red) from weighted (solid line) and binary (dashed line) adjacency matrices. Mean unit-wise ICCs for degree (middle panel) and clustering coefficient (right panel) from weighted matrix threshold at $r > 0.05$, for all subjects (black), mvmt < 3 (gray), and mvmt < 2 (light gray) groups, without and with wavelet transformation at scale 3. Error bars signify s.e.m. across nodes. WM, white matter; NB, non-brain; wavelet, wavelet transformation at scale 3.

Thus, in the following section, we present reliability results from weighted matrices thresholded at $r > 0.05$.

Unit (node)-wise graph metric reliability proved superior to unit (voxel or ROI pair)-wise reliability derived from the seed-based ROI or ROI matrix approaches, with graph metrics showing roughly 30% higher ICCs (Fig. 4B; compare with Figs. 2B and 3D). Clustering coefficients derived from weighted matrices provided the most reliable graph metric at the node level, particularly at less stringent hard thresholds (Fig. 4B, solid red line). As seen for the ROI matrix correlation analyses, reliability of graph metrics was improved by wavelet transformation of the nodal timeseries. In the low subject motion group, wavelet transformed matrices (scale 3) produced degree and clustering coefficient ICCs reflecting good unit-wise reliability (ICC > 0.6, Fig. 4B, right panel).

Temporal concatenation group ICA improved reliability over ICA with template matching

We further examined reliability of a widely used model-free ICN fMRI approach, spatial ICA, implemented using the template-matching approach and the more recently developed temporal concatenation group ICA with back reconstruction. With both methods, each subject's final salience network map consists of voxel-wise z-scores, which represent the degree to which that voxel belongs to the overall ICA component.

The template-matching ICA approach showed moderate scan-wise reliability (ICC ~ 0.4), but poor unit (voxel)-wise reliability, whether component number was estimated by the software or fixed to 20 (ICC ~ 0.1, Figs. 5C,D). Although, the group-level salience network maps at baseline and follow-up showed remarkably similar spatial patterns (Fig. 5A, lower panel), examination of the individual subject-level data revealed that the best-fit components selected were often dissimilar, as illustrated by one randomly selected subject (Fig. 5A, upper panel). As a result, unit-wise ICC was low (~0.1, Fig. 5D). Previous studies suggest that target components can be reproducibly chosen by trained investigators using visual inspection (Damoiseaux et al., 2006). We therefore repeated the reliability analysis on components selected visually, but no improvement in reliability was realized, suggesting that the poor reliability of ICA was not due to template-matching algorithm alone.

Temporal concatenation group ICA with back reconstruction (TC-GICA) greatly improved both scan-wise and unit-wise reliability (Figs. 5C,D), consistent with a previous study in young controls (Zuo et al., 2010). TC-GICA extracted much more reliable salience network components than template-matching ICA, particularly when the component number was fixed to 20. Scan-wise reliability approached the good range (ICC ~ 0.6) and unit-wise ICC was in the fair range (ICC ~ 0.3). Furthermore, unit-wise reliability improved by ~60%, into the moderate range, for voxels within the top 25th percentile for component z-scores (Fig. 5D). Moreover, ICCs were similar across the range of subject motion, suggesting that ICA reliability was less sensitive to movement artifacts compared to other analytical approaches, perhaps because movement-related noise is separated from neural signal by the ICA algorithm (Figs. 5C,D).

Test–retest reliability of DMN showed similar properties

Previous test–retest reliability studies of task-free fMRI in healthy young adults revealed qualitatively similar findings to those presented here despite examining different ICNs and even whole-brain connectivity (Braun et al., 2011; Schwarz and McGonigle, 2011; Shehzad et al., 2009; Telesford et al., 2010; Wang et al., 2011; Zuo et al., 2010). Nonetheless, to support generalization of the present salience network findings to other ICNs, we repeated the seed-based ROI and ICA reliability analyses for the DMN. For the seed-based ROI approach, including global signal removal produced the lowest

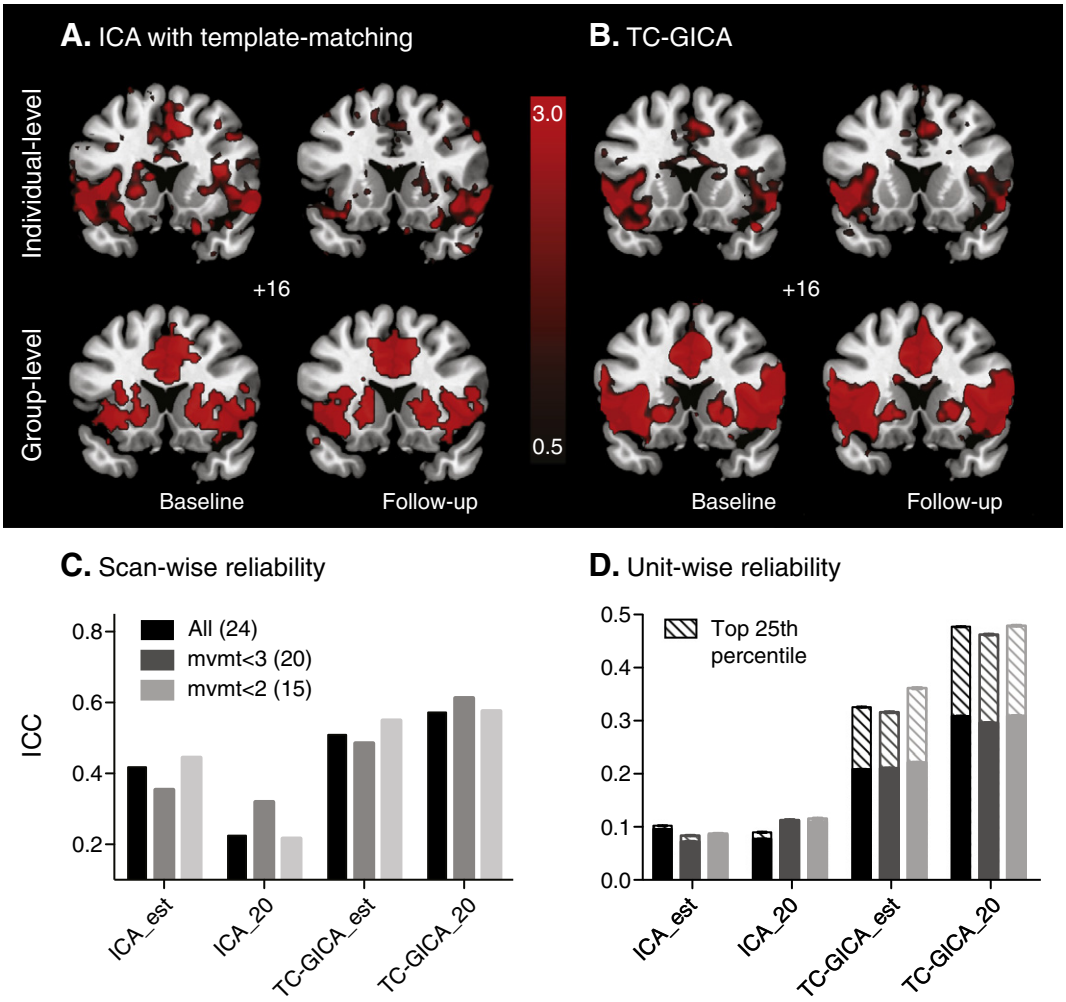

Fig. 5. Reliability of ICA. A. Salience network components identified using ICA with template-matching of a randomly selected subject's baseline (left) and follow up (right) scans (upper panel), and the group-level maps averaged from all subjects' baseline and follow-up scans (lower panel). Individual maps are shown at $z > 1.5$. Group level maps are shown at $p < 0.01$ uncorrected height and $p < 0.01$ corrected extended thresholds B. Individual-level (of the same subject as in A; upper panel) and group-level (lower panel) salience network components identified using TC-GICA, for baseline (left) and follow-up (right) scans. Maps are shown at $z > 1.5$. C. Scan-wise ICCs using individual ICA with estimated (ICA_est) or fixed (ICA_20) component number, or TC-GICA with estimated (TC-GICA_est) or fixed (TC-GICA_20) component number for all subjects (black), mvmt<3 (gray) and mvmt<2 (light gray) groups. Number of subjects in each group is indicated in parentheses. ICA methods used are indicated along the x-axis. D. Mean unit-wise ICCs using individual ICA with estimated (ICA_est) or fixed (ICA_20) component number, TC-GICA with estimated (TC-GICA_est) or fixed (TC-GICA_20) component number for all subjects (black), mvmt<3 (gray), and mvmt<2 (light gray) groups. Error bars signify s.e.m. across all voxels. Mean unit-wise ICCs of voxels in the top 25th percentile for connectivity are shown as striped bars.

reliability, whereas reliability was highest when averaged non-brain signal regressors were included, and subjects with greater head motion showed lower reliability (Fig. 6A). For the ICA approach, TC-GICA showed higher reliability than subject-level ICA with template-matching, although subject-level ICA was more reliable for deriving the DMN than for deriving the salience network (Figs. 6B and 5D). Both ROI and ICA approaches showed higher scan-wise reliability than unit-wise reliability (Fig. 6) for the DMN, consistent with the salience network findings (Figs. 1–5).

Discussion

This study provides a comprehensive investigation of the test–retest reliability of task-free intrinsic connectivity fMRI in healthy older adults. We performed a systematic analysis and direct comparison of commonly employed ICN analytical strategies and preprocessing manipulations. To constrain scope and complement previous studies, we focused on the salience network, an ICN of broad clinical and neuroscientific relevance for which reliability has not been examined, as well as DMN, in light of its relevance to aging research and its examination in previous reliability studies. In general, we found that model-driven methods based on ROI BOLD signal correlations provided the highest scan-wise reliability, whereas graph theoretical measures provided the highest unit-wise reliability. Within the model-free methods, temporal concatenation group ICA improved reliability over subject-level ICA approach. Single summary measures for each subject, derived at the scan-wise level, showed greater reliability than the averaged reliability calculated across individual units, such as voxels or ROIs. Inclusion of global signal regressors and subjects with greater head motion worsened reliability, whereas inclusion of averaged non-brain signal regressors and wavelet transformation of functional timeseries improved reliability. These findings provide a foundation for developing ICN fMRI as a longitudinal biomarker for diseases of aging.

Reliability of ICN-fMRI in older controls is comparable to younger controls

In general, our results on long-term test–retest reliability in older controls are comparable to previous findings in young controls, despite substantial differences in subjects' age and the ICNs examined

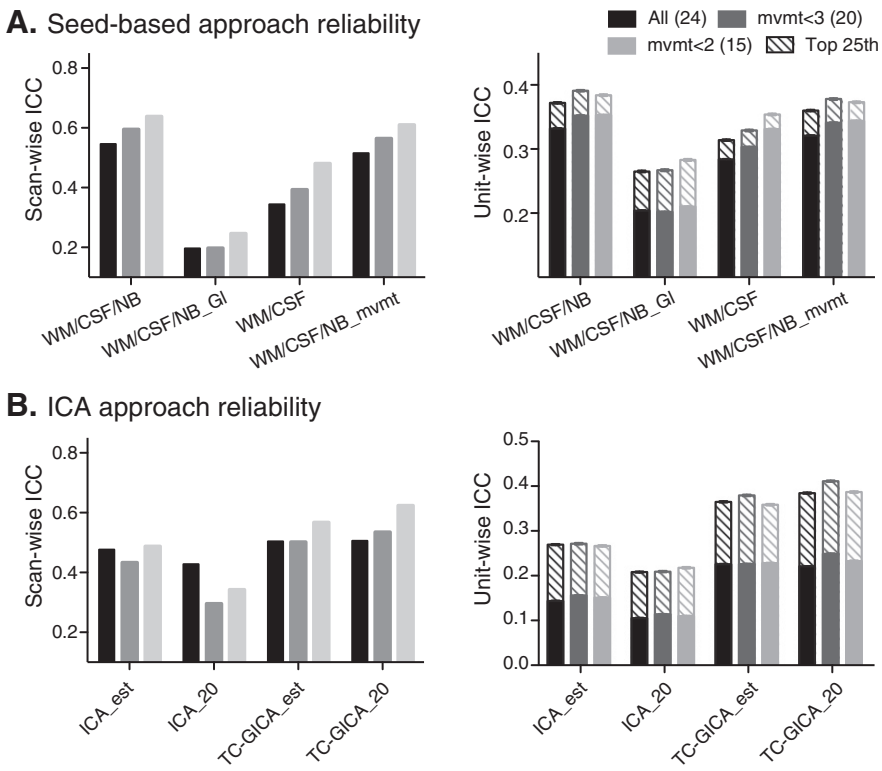

Fig. 6. Reliability of functional connectivity measures within the DMN. A. Scan-wise ICC (left) and mean unit-wise (right) of the ROI analysis for all subjects (black), mvmt<3 (gray) and mvmt<2 (light gray) groups. Mean unit-wise ICCs of voxels in the top 25th percentile for connectivity are shown as striped bars. Number of subjects in each group is indicated in parentheses. Nuisance regressors included in the analyses are indicated along the x-axis. WM, white matter; NB, non-brain; GI, global signal; mvmt, movement parameters. B. Scan-wise ICCs (left) and mean unit-wise ICCs (right) using individual ICA with estimated (ICA_est) or fixed (ICA_20) component number, TC-GICA with estimated (TC-GICA_est) or fixed (TC-GICA_20) component number for all subjects (black), mvmt<3 (gray), and mvmt<2 (light gray) groups. Error bars signify s.e.m. across all voxels. Mean unit-wise ICCs of voxels in the top 25th percentile for connectivity are shown as striped bars.

(Table 2, but see below for discussion on Wang et al., 2011). Our findings extend previous work, however, by providing reliability estimates for the same networks on the same subjects using a diverse array of methods. Furthermore, the salience network and DMN revealed similar reliability findings, suggesting that our results are likely to generalize to other ICNs. Accordingly, these data may serve as a guide for researchers seeking to compare methods for their specific research purpose.

Effects of global signal and other regressors on reliability

Preprocessing of task-free fMRI data is designed to remove potential sources of noise, usually by entering nuisance covariates thought to capture such noise in the general linear model. However, there remains a lack of consensus regarding the optimal choice of covariates. At one extreme, some studies removed no potentially confounding signals (Nakamura et al., 2009; van den Heuvel et al., 2008; Wang et al., 2010), whereas others removed WM, CSF, and global signals along with time courses of movement parameters (Hayasaka and Laurienti, 2010; Shehzad et al., 2009). Most studies, including our own (Seeley et al., 2009), have removed some subset of these confounding signals (Achard and Bullmore, 2007; Liu et al., 2008; Mumford et al., 2010; Wang et al., 2009). Here, we compared three combinations of nuisance covariates and found that the highest reliability was achieved by removing WM, CSF and non-brain signals but not the global signal. Including movement parameters had little impact.

Global signal regression in task-free fMRI has been debated on other grounds (Birn et al., 2006; Chang and Glover, 2010; Fox et al., 2009; Murphy et al., 2009; Schölvinck et al., 2010; Vincent et al., 2006). Arguments in favor of global signal correction emphasize the value of removing non-neural noise, such as respiration-induced fluctuations. The procedure, however, markedly shifts the distribution of correlation among brain voxels, reducing positive correlations and inducing anti-correlations (Fig. 1C; Murphy et al., 2009; Schwarz and McGonigle, 2011; Braun et al., 2011). Here, global signal correction had a profound effect on both scan-wise and unit-wise reliability of correlation-based analyses (Figs. 1D and 2B), as reported in some previous analyses (Braun et al., 2011; Schwarz and McGonigle, 2011). This observation may help reconcile discrepant findings in previous reliability analyses. For example, two studies analyzed the reliability of graph theoretical analysis on the same whole brain network (based on parcellation units derived from the AAL atlas) (Schwarz and McGonigle, 2011; Wang et al., 2011), but one study (Wang et al., 2011) reported much lower ICCs than the other (Schwarz and McGonigle, 2011) and our study. This discrepancy could be explained by the removal of global signal in one study (Wang et al., 2011) but not in the other (Schwarz and McGonigle, 2011). The impact of global signal regression on reliability may reflect the recentering of each subject's connectivity distribution about a mean of zero, which diminishes between-subject variance (Figs. 1B, C), thus lowering ICC. However, two previous studies (Braun et al., 2011; Schwarz and McGonigle, 2011) reported that global signal removal had less impact or even increased ICCs when applied to some graph metrics (e.g., global efficiency). They attributed this observation to the compromised left tail characteristics of the networks after global signal regression.

Motion-related artifact remains a major concern for ICN fMRI data, especially in children or aging or diseased subjects who may struggle to achieve complete stillness during the scan (Power et al., 2011; Van Dijk et al., 2012). We found that head motion had a major negative impact on reliability, in particular for analytical strategies based on correlations between ROIs. The impact of motion was not rescued by

Table 2
Subject demographics and reliability results summarized for the present older control analyses and for the young controls assessed in previous studies. Reliability results listed here are those assessed on all subjects in both the present and previous studies.

Dataset	UCSF older control	NYU young control
Age at 1st scan, yrs (s.d.)	66.7 (6.4)	20.5 (8.4)
Gender, M/F	11/13	11/15
Scan interval, mos (s.d.)	13 (3)	11 (5)
Education, yrs (s.d.)	17.5 (1.8)	N.A.
Network evaluated	Salience	Multiple
Seed-based ROI approach		
Scan-wise ICC	0.54–0.68	N.A.
Unit-wise ICC	0.27–0.33	0.13–0.45[a,b]
ROI matrix analysis		
Scan-wise ICC	0.49	0.50[c]
Unit-wise ICC	0.26	0.22–0.23[b]
ROI matrix analysis (wavelet)		
Scan-wise ICC	0.59	N.A.
Unit-wise ICC	0.32	N.A.
Weighted clustering coefficient		
Scan-wise ICC	0.54	~0.5[c,d]; ~0.2[e]
Unit-wise ICC	0.41	~0.2[e]
Weighted clustering coefficient (wavelet)		
Scan-wise ICC	0.52	N.A.
Unit-wise ICC	0.53	N.A.
TC-GICA		
Scan-wise ICC	0.58	N.A.
Unit-wise ICC	0.31	N.A.
	0.62[f]	0.45–0.65[f,g]

Note: Values listed here are means (except where specified). N.A. = not assessed. ICC values for older controls were computed with white matter, CSF and non-brain signals as the nuisance regressors and averaged across all units (voxels, ROI pairs and nodes), unless noted otherwise.

[a] Seeded separately at posterior cingulate cortex, supplementary motor area, and the inferior parietal sulcus and ICC averaged across significant voxels (positive and negative).

[b] Mutiple networks; White matter, CSF, global signals and movement parameters removed (Shehzad et al., 2009);

[c] Whole brain network based on ROIs defined by automated anatomical labeling (AAL) altas. White matter, CSF signals and movement parameters removed (Schwarz and McGonigle, 2011).

[d] Whole brain network based on ROIs defined by AAL. White matter, CSF signals and movement parameters removed (Braun et al., 2011).

[e] Whole brain network based on ROIs defined by AAL, Harvard-Oxford atlas and meta-analysis. Global signal and movement parameters removed (Wang et al., 2011).

[f] Modal ICCs of voxels in the identified components.

[g] Components associated with sensory, motor, higher order cognitive function and the default mode network (Zuo et al., 2010).

including motion parameters as nuisance regressors, consistent with other studies (Power et al., 2011). In one reliability study of whole brain connectivity in a young control dataset, the authors found that removing motion signals improved reliability (Schwarz and McGonigle, 2011). These disparate findings could reflect differing magnitudes of head motion, the different networks investigated, or differences in the methods used to remove motion signals. In the previous study (Schwarz and McGonigle, 2011), motion parameters were deconvolved from each gray matter voxel during image preprocessing, whereas here motion parameters were included in the model as regressors.

Our study compared test–retest reliability across task-free fMRI analytical strategies. Reliability represents just one key methodological attribute, however, with validity and sensitivity being the other important goals that may or may not be enhanced by more reliable methods. One might question, for example, whether retaining the global signal inflates reliability by including systematic sources of physiological noise that persist between scan sessions (Guijt et al., 2007). Such noise would represent a reproducible but invalid (non-target) signal. This possibility could be addressed in future studies that acquire physiological variables (heart rate, respiratory rate and depth) during fMRI scanning.

Summary ICN measures showed the greatest reliability

For both seed-based ROI and ROI correlation matrix approaches, scan-wise reliability was considerably higher than unit-wise reliability. Moreover, graph theoretical measures produced the best unit-wise reliability, based on metrics calculated from individual nodes. Overall, our results suggest that summary measures that reflect network connectivity as a whole provide the highest reliability, falling within the good to excellent range. This observation is consistent with reliability studies on young controls and with studies of test–retest reliability over shorter (days) intervals (Table 2). For ROI correlation matrices, scan-wise reliability (Schwarz and McGonigle, 2011) was more than twice as high as the average unit (ROI pair)-wise reliability. As for short-term reliability, graph theoretical analysis was reported to have high reliability in older subjects (ICCs > 0.75, Telesford et al., 2010). On the other hand, even in the young controls, ROI matrix pairs showed only moderate to good short-term reliability (Schwarz and McGonigle, 2011). These studies differed somewhat, however, in the data preprocessing steps and ICNs analyzed, limiting the inferences that can be reached through direct comparisons. In addition, (Braun et al., 2011) found first-order graph metrics (measures directly derived from the adjacency matrix) generally had lower reliability then second-order (measures derived from two or more first order metrics).

The greater robustness of scan-wise measures might not be surprising, considering that these measures incorporate a greater body of the acquired data, but it is encouraging, since single-value (scan-wise) measures are more attractive as longitudinal biomarkers than multiple measures (unit-wise) derived for each subject. The challenge is to identify the most informative (e.g., disease-targeted) network, whose scan-wise measures may capture the biological phenomenon of interest. On the other hand, a small number of particular voxels and ROI pairs within the salience network provided higher unit-wise reliability than the scan-wise measures, suggesting that future studies aiming to capture longitudinal change may benefit from exploring both scan-wise and unit-wise measures.

Reliability of ROI correlation-based methods was improved by wavelet transformation

ROI correlation matrices and graph metrics have most commonly been studied using Pearson's correlation between ROIs' BOLD signal timeseries. Several recent studies, however, have applied wavelet analysis to generate matrices and graph metrics from the correlations between ROI-based wavelet coefficients (Achard and Bullmore, 2007; Achard et al., 2006; Meunier et al., 2009; Supekar et al., 2008). These studies did not, however, compare the effectiveness of wavelet analysis to alternative methods directly. The present study provided a quantitative comparison between ICN fMRI reliability with and without wavelet transformation of the ROI timeseries and found that test–retest reliability was greatly improved after applying MODWT wavelet transformation. This observation suggests that wavelet transformation may constructively augment standard image analysis strategies, especially for longitudinal studies.

Reliability of subject-level ICA with template matching vs. temporal concatenation group ICA

In older controls, we replicated a previous study in young controls, which suggested that temporal concatenation group ICA was more reliable than standard ICA with template-matching based on individual subject scans (Zuo et al., 2010). Several studies on age-related diseases, including some of our own, have employed the ICA template-matching approach, combining components across subjects for

group-level analysis (Greicius et al., 2004; Mohammadi et al., 2009; Seeley et al., 2009; Zhou et al., 2010). Although template-matching ICA approaches may suffice for between-group comparisons, producing biologically plausible effects, the present data argue against the use of this method for longitudinal within-subjects ICN analysis, particularly for ICNs other than the DMN.

The higher reliability provided by TC-GICA is encouraging, yet the application of this method to diseases of aging requires further validation. The first step in the method is to concatenate the task-free fMRI data from all subjects in a study into a single timeseries. While this approach may be straightforward for examining ICNs in healthy subjects, it could create problems for disease studies by incorporating diseased networks into the dataset used to define the components analyzed. When the goal is to detect deviations of a patient group from normal, using patients to define components may reduce the sensitivity of this method. This issue merits further study.

Future directions

ICN fMRI provides a simple, non-invasive, inexpensive technique, and these characteristics make it an attractive potential diagnostic or disease-monitoring biomarker. Sensitivity as a diagnostic biomaker has been explored in several diseases (Fox and Greicius, 2010). Application of task-free fMRI as a longitudinal biomarker will require both reliability and sensitivity to longitudinal change. In other words, the ideal measures should be not only proven stable over time in the absence of disease but also highly attuned to longitudinal decline or improvement. Ultimately, the best measures will allow detection of clinical meaningful benefits over short intervals in few subjects; this goal may require trade-offs between reliability and sensitivity. Future studies on aging and neurodegenerative populations are needed to explore the sensitivity of available and new analytical methods to longitudinal changes.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2012.03.027.

Acknowledgments

The authors thank Efstathios Gennatas, Paul Keselman, Lara Stables, William Irwin, for technical assistance with data collection and storage. This work was supported by a grant from the Larry L. Hillblom Foundation to J.H.K. Finally, we thank our study participants for their generous contributions to aging research.

References

Achard, S., Bullmore, E., 2007. Efficiency and cost of economical brain functional networks. PLoS Comput. Biol. 3, e17.
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. J. Neurosci. 26, 63–72.
Bai, F., Xie, C., Watson, D.R., Shi, Y., Yuan, Y., Wang, Y., Yue, C., Tang, Y., Wu, D., Zhang, Z., 2011. Aberrant hippocampal subregion networks associated with the classifications of aMCI subjects: a longitudinal resting-state study. PLoS One 6 (12).
Beckmann, C., Smith, S., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. on Medical Imaging 23 (2), 137–152.
Birn, R.M., Diamond, J.B., Smith, M.A., Bandettini, P.A., 2006. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. Neuroimage 31, 1536–1548.
Braun, U., Plichta, M.M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., Mier, D., Mohnke, S., Heinz, A., Erk, S., et al., 2011. Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. Neuroimage 59, 1404–1412.
Bullmore, E., Fadili, J., Maxim, V., Sendur, L., Whitcher, B., Suckling, J., Brammer, M., Breakspear, M., 2004. Wavelets and functional magnetic resonance imaging of the human brain. Neuroimage 23 (Suppl. 1), S234–S249.
Calhoun, V.D., Pekar, J.J., Pearlson, G.D., 2004. Alcohol intoxication effects on simulated driving: exploring alcohol-dose effects on brain activation using functional MRI. Neuropsychopharmacology 29, 2097–2107.
Chang, C., Glover, G.H., 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage 50, 81–98.
Damoiseaux, J.S., Rombouts, S.A., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. U.S.A. 103, 13848–13853.
Erhardt, E.B., Rachakonda, S., Bedrick, E.J., Allen, E.A., Adali, T., Calhoun, V.D., 2010. Comparison of multi-subject ICA methods for analysis of fMRI data. Hum. Brain Mapp. 32, 2075–2095.
Fox, M.D., Greicius, M., 2010. Clinical applications of resting state functional connectivity. Front. Syst. Neurosci. 4, 19.
Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8, 700–711.
Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., 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.
Fox, M.D., Zhang, D., Snyder, A.Z., Raichle, M.E., 2009. The global signal and observed anticorrelated resting state brain networks. J. Neurophysiol. 101, 3270–3283.
Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc. Natl. Acad. Sci. U.S.A. 101, 4637–4642.
Guijt, A., Sluiter, J., Frings-Dresen, M., 2007. Test–retest reliability of heart rate variability and respiration rate at rest and during light physical activity in normal subjects. Arch. Med. Res. 38, 113–120.
Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C.F., Menon, V., Greicius, M.D., 2009. Distinct cerebellar contributions to intrinsic connectivity networks. J. Neurosci. 29, 8586–8594.
Harrison, B.J., Pujol, J., Ortiz, H., Fornito, A., Pantelis, C., Yücel, M., 2008. Modulation of brain resting-state networks by sad mood induction. PLoS One 3, e1794.
Hayasaka, S., Laurienti, P.J., 2010. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. Neuroimage 50, 499–508.
Kurth, F., Eickhoff, S.B., Schleicher, A., Hoemke, L., Zilles, K., Amunts, K., 2010. Cytoarchitecture and probabilistic maps of the human posterior insular cortex. Cereb. Cortex 20, 1448–1461.
Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Yu, C., Liu, H., Liu, Z., Jiang, T., 2008. Disrupted small-world networks in schizophrenia. Brain 131, 945–961.
Lowe, M.J., Mock, B.J., Sorenson, J.A., 1998. Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. Neuroimage 7, 119–132.
Maxim, V., Sendur, L., Fadili, J., Suckling, J., Gould, R., Howard, R., Bullmore, E., 2005. Fractional Gaussian noise, functional MRI and Alzheimer's disease. Neuroimage 25, 141–158.
McGraw, K.O., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1, 30–46.
Meunier, D., Achard, S., Morcom, A., Bullmore, E., 2009. Age-related changes in modular organization of human brain functional networks. Neuroimage 44, 715–723.
Mohammadi, B., Kollewe, K., Samii, A., Krampfl, K., Dengler, R., Münte, T.F., 2009. Changes of resting state brain networks in amyotrophic lateral sclerosis. Exp. Neurol. 217, 147–153.
Mumford, J.A., Horvath, S., Oldham, M.C., Langfelder, P., Geschwind, D.H., Poldrack, R.A., 2010. Detecting network modules in fMRI time series: a weighted network analysis approach. Neuroimage 52, 1465–1476.
Murphy, K., Birn, R.M., Handwerker, D.A., Jones, T.B., Bandettini, P.A., 2009. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage 44, 893–905.
Nakamura, T., Hillary, F.G., Biswal, B.B., 2009. Resting network plasticity following brain injury. PLoS One 4, e8220.
Park, C., Chang, W.H., Ohn, S.H., Kim, S.T., Bang, O.Y., Pascual-Leone, A., Kim, Y.H., 2011. Longitudinal changes of resting-state functional connectivity during motor recovery after stroke. Stroke 42, 1357–1362.
Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge.
Poline, J.B., Worsley, K.J., Evans, A.C., Friston, K.J., 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage 5, 83–96.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2011. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154.
Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059–1069.
Schölvinck, M.L., Maier, A., Ye, F.Q., Duyn, J.H., Leopold, D.A., 2010. Neural basis of global resting-state fMRI activity. Proc. Natl. Acad. Sci. U.S.A. 107, 10238–10243.
Schwarz, A.J., McGonigle, J., 2011. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. Neuroimage 55, 1132–1146.
Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27, 2349–2356.
Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., Greicius, M.D., 2009. Neurodegenerative diseases target large-scale human brain networks. Neuron 62, 42–52.
Shehzad, Z., Kelly, A.M., Reiss, P.T., Gee, D.G., Gotimer, K., Uddin, L.Q., Lee, S.H., Margulies, D.S., Roy, A.K., Biswal, B.B., et al., 2009. The resting brain: unconstrained yet reliable. Cereb. Cortex 19, 2209–2229.
Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2011. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb. Cortex 22, 158–165.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.
Sridharan, D., Levitin, D.J., Menon, V., 2008. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. Proc. Natl. Acad. Sci. U.S.A. 105, 12569–12574.

Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D., 2008. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput. Biol. 4, e1000100.

Telesford, Q.K., Morgan, A.R., Hayasaka, S., Simpson, S.L., Barret, W., Kraft, R.A., Mozolic, J.L., Laurienti, P.J., 2010. Reproducibility of graph metrics in fMRI networks. Front. Neuroinform. 4, 117.

Uddin, L.Q., Menon, V., 2009. The anterior insula in autism: under-connected and under-examined. Neurosci. Biobehav. Rev. 33 (8), 1198–1203.

van den Heuvel, M.P., Stam, C.J., Boersma, M., Hulshoff Pol, H.E., 2008. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. Neuroimage 43, 528–539.

Van Dijk, K.R., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 59, 431–438.

Vincent, J.L., Snyder, A.Z., Fox, M.D., Shannon, B.J., Andrews, J.R., Raichle, M.E., Buckner, R.L., 2006. Coherent spontaneous activity identifies a hippocampal–parietal memory network. J. Neurophysiol. 96, 3517–3531.

Wang, J., Wang, L., Zang, Y., Yang, H., Tang, H., Gong, Q., Chen, Z., Zhu, C., He, Y., 2009. Parcellation-dependent small-world brain functional networks: a resting-state fMRI study. Hum. Brain Mapp. 30, 1511–1523.

Wang, L., Li, Y., Metzak, P., He, Y., Woodward, T.S., 2010. Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition. Neuroimage 50, 862–872.

Wang, J.H., Zuo, X.N., Gohel, S., Milham, M.P., Biswal, B.B., 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.

White, T.P., Joseph, V., Francis, S.T., Liddle, P.F., 2010. Aberrant salience network (bilateral insula and anterior cingulate cortex) connectivity during information processing in schizophrenia. Schizophr. Res. 123, 105–115.

Wiech, K., Lin, C., Brodersen, K.H., Bingel, U., Ploner, M., Tracey, I., 2010. Anterior insula integrates information about salience into perceptual decisions about pain. J. Neurosci. 30 (48), 16324–16331.

Zhou, J., Greicius, M.D., Gennatas, E.D., Growdon, M.E., Jang, J.Y., Rabinovici, G.D., Kramer, J.H., Weiner, M., Miller, B.L., Seeley, W.W., 2010. Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. Brain 133, 1352–1367.

Zuo, X.N., Kelly, C., Adelstein, J.S., Klein, D.F., Castellanos, F.X., Milham, M.P., 2010. Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach. Neuroimage 49, 2163–2177.