

Published in final edited form as:

Neuroimage. 2010 February 1; 49(3): 2163–2177. doi:10.1016/j.neuroimage.2009.10.080.

Reliable Intrinsic Connectivity Networks: Test-Retest Evaluation Using ICA and Dual Regression Approach

Xi-Nian Zuo^a, Clare Kelly^a, Jonathan S. Adelstein^a, Donald F. Klein^{a,b,c}, F. Xavier Castellanos^{a,b}, and Michael P. Milham^a

^aPhyllis Green and Randolph Cōwen Institute for Pediatric Neuroscience, New York University Child Study Center, New York, USA

^bNathan Kline Institute for Psychiatric Research, Orangeburg, New York, USA

^cDepartment of Psychiatry, Professor Emeritus, College of Physicians and Surgeons, Columbia University, New York, USA

Abstract

Functional connectivity analyses of resting-state fMRI data are rapidly emerging as highly efficient and powerful tools for in vivo mapping of functional networks in the brain, referred to as intrinsic connectivity networks (ICNs). Despite a burgeoning literature, researchers continue to struggle with the challenge of defining computationally efficient and reliable approaches for identifying and characterizing ICNs. Independent component analysis (ICA) has emerged as a powerful tool for exploring ICNs in both healthy and clinical populations. In particular, temporal concatenation group ICA (TC-GICA) coupled with a back-reconstruction step produces participant-level resting state functional connectivity (RSFC) maps for each group-level component. The present work systematically evaluated the test-retest reliability of TC-GICA derived RSFC measures over the short-term (< 45 minutes) and long-term (5 – 16 months). Additionally, to investigate the degree to which the components revealed by TC-GICA are detectable via single-session ICA, we investigated the reproducibility of TC-GICA findings. First, we found moderate-to-high short- and long-term test-retest reliability for ICNs derived by combining TC-GICA and dual regression. Exceptions to this finding were limited to physiological- and imaging-related artifacts. Second, our reproducibility analyses revealed notable limitations for template matching procedures to accurately detect TC-GICA based components at the individual scan level. Third, we found that TC-GICA component's reliability and reproducibility ranks are highly consistent. In summary, TC-GICA combined with dual regression is an effective and reliable approach to exploratory analyses of resting state fMRI data.

Keywords

test-retest reliability; intrinsic connectivity network; ICA; dual regression; resting state

1. Introduction

Functional connectivity of resting-state fMRI data is rapidly emerging as a highly efficient and powerful tools for in vivo mapping of neural circuitry in the human brain. Essentially, resting state functional connectivity (RSFC) approaches detect coherent patterns of low-frequency (<

Corresponding author xinian.zuo@nyumc.org (Xi-Nian Zuo), amclarekelly@gmail.com (Clare Kelly), jsadelstein@gmail.com (Jonathan S. Adelstein), donalddk737@aol.com (Donald F. Klein), castef01@nyumc.org (F. Xavier Castellanos), michael.milham@nyumc.org (Michael P. Milham).

0.1 Hz) fluctuations in the resting state BOLD signal, referred to as “intrinsic connectivity networks” (ICNs). Despite a burgeoning literature demonstrating the utility of RSFC approaches (Fox and Raichle, 2007), researchers continue to struggle with defining computationally efficient and reliable approaches for identifying and characterizing ICNs.

Inspired by a seminal demonstration of RSFC within the motor system (Biswal et al., 1995), seed-based correlation represents the dominant approach for studying RSFC. This method detects ICNs by identifying voxels whose timeseries significantly correlate with the mean timeseries of voxels within an a priori seed region of interest (ROI). In addition to generating highly detailed maps of complex functional systems (Margulies et al., 2007; Zhang et al., 2008; Di Martino et al., 2008; Roy et al., 2009; Nioche et al., 2009), seed-based correlation analyses are commonly used to identify abnormalities in ICNs related to psychopathology (Greicius, 2008; Broyd et al., 2009), as well as relationships between RSFC measures and individual differences in behavioral (Kelly et al., 2008; Fox et al., 2007) and trait characteristics (Di Martino et al., 2009). Overall, this mass-univariate approach has proven to be a powerful, efficient, and reliable tool for neuroimaging (Shehzad et al., 2009).

Seed-based analytic approaches are not without limitation, however. Most notably, the ICNs derived with seed-based correlation are highly dependent upon choice of a seed ROI. Further, the seed ROI timeseries includes indeterminate noise, in contrast to the idealized task design matrix used in task-based analyses. While several preprocessing strategies commonly are employed to remove noise (e.g., regressing out the global signal), such corrections can affect the data (e.g., inducing artifactual negative correlations) (Murphy et al., 2009). Finally, as a general linear model (GLM)-based univariate method, seed-based correlation disregards relationships among multiple voxels (i.e., it only analyzes the relationship between the seed timeseries and one other voxel at a time).

In light of these issues, neuroimaging researchers have adopted a multivariate signal processing method known as independent component analysis (ICA) to explore the spatial-temporal properties of resting state fMRI data (Kiviniemi, 2003; van de Ven et al., 2004; Beckmann et al., 2005). In theory, without any explicit a priori knowledge, ICA aims to separate spatially (spatial ICA: sICA) or temporally (temporal ICA: tICA) independent patterns from their linearly mixed BOLD signals via maximization of mutual independence among components (Stone et al., 1999). Generally, sICA is the more appropriate choice for analyzing resting state fMRI data given the small number of time points included in most fMRI datasets. Thus, we also used sICA in this paper. For simplicity, in the remainder of this paper, we refer to sICA as ICA.

ICA offers several potential advantages over seed-based correlation. First, as noted above, ICA is a multivariate, data-driven approach. It thus requires no a priori hypothesis or model of brain activity. Second, by taking account of multiple simultaneous voxel-voxel relationships, ICA detects interacting networks of regions, rather than the single region-dominant (i.e., seed ROI) networks produced by seed-based correlation. Finally, ICA is capable of extracting noise (e.g., scanner, physiological and motion artifacts) from the desired dataset. ICA-based denoising is fully data-driven, automatic (Thomas et al., 2002; Perlberg et al., 2007; Tohka et al., 2008) and relatively unaffected by different temporal sampling rates (DeLuca et al., 2006), thus avoiding the need for a priori specification of noise that arises in seed-based correlation.

Despite its promise, the unconstrained nature of ICA raises a significant challenge for group-based analyses. If a separate ICA is conducted for each participant in a study, a researcher must first match the components between participants before being able to carry out group analyses (Wang and Peterson, 2008) – a task that is complicated by the unconstrained component order in ICA (Hyvarinen and Oja, 2000). Different numbers of components also can be extracted for

different individuals, further complicating the challenge. Accordingly, some have advocated carrying out a single ICA analysis across an entire group of participants (see (Guo and Pagnoni, 2008) and (Calhoun et al., 2009) for reviews on various group ICA methods). This can be accomplished in a relatively efficient manner by carrying out an ICA on a single large dataset by temporally concatenating all individual datasets (Calhoun et al., 2001; Beckmann et al., 2005). This method is referred to as temporal concatenation group ICA (TC-GICA). Based upon group-level components identified by TC-GICA, individual participant components can be generated using approaches such as principal component analysis (PCA) back-projection (Calhoun et al., 2001), or more recently developed GLM dual regression approaches (Filippini et al., 2009; Beckmann et al., 2009). The functionally relevant components of all group-level components are referred to as ICNs.

The potential utility of group-based ICA approaches has been demonstrated by an increasing number of studies examining clinical populations. Clinical samples with Alzheimer's disease or dementia (Greicius et al., 2004; Seeley et al., 2009; Rombouts et al., 2009), schizophrenia (Jafri et al., 2008), depression (Greicius et al., 2007), epilepsy (Zhang et al., 2009), Huntington's disease (Wolf et al., 2008), and amyotrophic lateral sclerosis (Mohammadi et al., 2009) have demonstrated altered ICA-derived ICNs. Despite this promising start, it is important to note that several factors may reduce the reliability of ICA approaches for mapping RSFC. First, the resting state is inherently unconstrained and thus subject to variation related to a participant's state. Second, factors unrelated to the participant can introduce significant variability (e.g., inter-session differences in scanner performance). Third, no ideal means exist for assessing the optimal number of components. Finally, the initial values for ICA estimation are random, which introduces further variability into the findings (Himberg et al., 2004; Yang et al., 2008). In considering these factors, in order to provide biomarkers based on neural circuitry, the test-retest (TRT) reliability of ICA-derived ICNs must first be assessed.

Several groups have addressed the reliability or consistency of ICA. Damoiseaux et al. (2006) demonstrated 10 qualitatively consistent ICNs across subjects by applying tensor ICA to resting state fMRI data from two imaging sessions separated by 5 – 14 days. They found that (1) spatial patterns of the 10 ICNs from the two sessions were quite consistent by visual inspection, and (2) regions exhibiting the highest RSFC within a component tended to exhibit the least variation across 100 surrogate datasets created via bootstrapping. Recently, using TC-GICA with three initial PCA data reductions, Chen et al. (2008) demonstrated the consistency of ICA-derived ICNs across 5 sessions within 16 days. Two more recent studies also showed high reproducibility and inter-rater selection reliability of an ICA-derived “default mode” network (Meindl et al., 2009; Franco et al., 2009). The above four studies focused only on the spatial consistency of ICNs across occasions within approximately 2 weeks, and they did not quantitatively assess both the voxel-wise intra- and inter-session TRT reliability of each group-level component.

In the current paper, we address the short-term (< 1 hour) and long-term (> 5 months) TRT of group-level components. We conducted these analyses with a previously published resting state fMRI dataset of 26 participants scanned three times on two different occasions (Shehzad et al., 2009; Zuo et al., 2010). Specifically, we: (a) used the TC-GICA approach to derive group-level components across all participants and sessions, (b) used dual regression to back-reconstruct each group-level component for each of the 3 scans at the individual participant-level, and (c) calculated voxel-wise intra- and inter-session intraclass correlation (ICC) on the basis of these individual-level back-reconstructed components. Based on prior work (Shehzad et al., 2009), we expected to find moderate-to-high TRT reliability for the various ICNs detected except for those components reflecting physiological or scanner-related noise.

While the present work relied upon the group-based TC-GICA approach for detecting components, alternative ICA-based analytic strategies exist. In particular, many studies have employed individual participant based ICA approaches, combined with a template matching for a group analysis (Greicius et al., 2004, 2007; Seeley et al., 2009; Mohammadi et al., 2009). To date, no study has systematically examined the issue of how easily components identified by group ICA can be detected using individual ICA. Thus, a secondary goal of the present work was to assess the reproducibility of components detected by TC-GICA at the participant-level. Specifically, the reproducibility analysis measured the degree to which components identified using group ICA were present in the results of single-participant ICA. For each participant, we carried out an individual ICA for each of the 3 scans. Then, for each group-level component, a commonly employed template matching procedure (Garrrity et al., 2007) was applied to identify the best matching component in each of the three ICA analyses. These best-matched individual-level components were used to evaluate the reproducibility.

Finally, we ranked the group-level components according to their TRT reliability and reproducibility. We hypothesized that functionally relevant ICNs would show high ranks in both TRT reliability and reproducibility. In contrast, we predicted that the components corresponding to various sources of physiological noise and scanner artifact would exhibit both low TRT reliability and reproducibility.

2. Materials and Methods

2.1. Participants

Twenty-six participants (mean age 20.5 ± 4.8 years, 11 males) were scanned three times. The data were used in our earlier studies to examine the TRT reliability of seed-based correlation measures (Shehzad et al., 2009; Zuo et al., 2010). The participants had no history of psychiatric or neurological illness, as confirmed by clinical assessment. Informed consent was obtained prior to participation. Data were collected according to protocols approved by the institutional review boards of New York University (NYU) and the NYU School of Medicine.

2.2. Imaging methods

Three resting-state scans were obtained for each participant using a Siemens Allegra 3.0 Tesla scanner. Each scan consisted of 197 contiguous EPI functional volumes (TR = 2000 ms; TE = 25 ms; flip angle = 90° , 39 slices, matrix = 64×64 ; FOV = 192 mm; acquisition voxel size = $3 \times 3 \times 3$ mm). Scans 2 and 3 were conducted in a single scan session, 45 minutes apart, and were 5-16 months (mean 11 ± 4) after scan 1. All individuals were asked to relax and remain still with their eyes open during the scan. For spatial normalization and localization, a high-resolution T1-weighted magnetization prepared gradient echo sequence was also obtained (MPRAGE, TR = 2500 ms; TE = 4.35 ms; TI = 900 ms; flip angle = 8° ; 176 slices, FOV = 256mm).

2.3. Data preprocessing

Data preliminary processing was carried out using FMRIB Software Library (FSL: version 4.1): 1) slice timing correction for interleaved acquisitions using Sinc interpolation with a Hanning windowing kernel; 2) head motion correction via a robust and accurate intra-modal volume linear registration ($_{MCFLIRT}$) (Jenkinson et al., 2002); 3) spatial smoothing with a 6mm FWHM Gaussian kernel; 4) 4D intensity normalization of all volumes by the same factor; 5) spatial normalization via estimating a B-spline basis nonlinear transformation from an individual functional space to MNI152 standard brain space ($_{FLIRT+FNIRT}$) (Andersson et al., 2007). After preprocessing, we excluded one participant from our subsequent analyses due to large amount of head motion in one of the three scans. The final data for this study thus consisted

of 75 scans from 25 participants. These data are fully available for downloading via http://www.nitrc.org/projects/nyu_trt.

2.4. Overall analytic strategy

Figure 1A illustrates the overall analytic strategy:

- **TRT reliability: component maps.** First, we employed TC-GICA to generate **group-level components** across all 25 participants and 3 scans/participant (total = 75 scans). Next, for each group-level component, we (1) used dual regression (DR) to back-reconstruct the corresponding individual scan component (i.e., individual-level DR component) for each of the 75 scans, and (2) calculated voxel-wise intra-session (scan 2 vs. scan 3) and inter-session (scan 1 vs. scan 2/3) TRT reliability based upon the individual-level DR components.
- **Reproducibility: component maps.** We investigated the extent to which group-level components are detectable when ICA is carried out at the individual scan level. This was accomplished by 1) carrying out a single scan ICA for each of the 75 individual scans (i.e., individual-level ICA), and then 2) for each group-level component, determining the best match between a template component and those identified in each of the 75 individual-level ICA runs (i.e., individual-level components). We explored the potential impact of the specific strategy employed for generating template components by repeating our analyses with two different template component sets: 1) the group-level TC-GICA component set (i.e., TC-GICA template matching) and 2) the individual-level DR components set (i.e., DR template matching).
- **Rank ordering: component maps.** We rank-ordered the group-level components based upon their TRT reliability and reproducibility.
- **TRT reliability: estimation of the number of components.** For each of five estimation methods, we assessed intra- and inter-session TRT reliability for the number of components estimated for an individual scan.

2.5. Independent component analysis

2.5.1. Probabilistic ICA—We briefly review probabilistic ICA and relevant properties that are important for applications in resting state fMRI studies (Beckmann and Smith, 2004). The ICA decomposes a resting state fMRI dataset into a linear mixture of spatially independent components (ICs) plus Gaussian noise. It has been implemented as the multivariate exploratory linear optimized decomposition into independent components (MELODIC) toolbox in FSL. It has the following mathematical representation:

$$\mathbf{X} = \mathbf{AS} + \mathbf{E}. \quad (1)$$

In (1),

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1(N-1)} & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2(N-1)} & x_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{T(N-1)} & x_{TN} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1P} \\ a_{21} & \cdots & a_{2P} \\ \vdots & \ddots & \vdots \\ a_{T1} & \cdots & a_{TP} \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1(N-1)} & s_{1N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{P1} & s_{P2} & \cdots & s_{P(N-1)} & s_{PN} \end{bmatrix}, \mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1(N-1)} & e_{1N} \\ e_{21} & e_{22} & \cdots & e_{2(N-1)} & e_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ e_{T1} & e_{T2} & \cdots & e_{T(N-1)} & e_{TN} \end{bmatrix}.$$

For resting state fMRI data, T is the number of time points, N is the number of voxels, and P is the number of ICs. These matrices can be further summarized and explained as follows: $\mathbf{X} = [x_1, x_2, \dots, x_T]^t$, in which each row vector represents a resting state BOLD volume at a time point t ; $\mathbf{S} = [s_1, s_2, \dots, s_P]^t$, in which the row vectors represent each IC; $\mathbf{A} = [a_1, a_2, \dots, a_T]^t$, in which each row includes the contributions of all P ICs to the resting state fMRI BOLD volume at the corresponding time point (thus each column constitutes a timeseries); $\mathbf{E} = [e_1, e_2, \dots, e_T]^t$ denotes the background multivariate Gaussian noise. Accordingly, this method decomposes a resting state fMRI BOLD volume into several spatially independent volumes (i.e., ICs) and relevant timeseries. Of note, explicit modeling of the background Gaussian noise effectively can reduce the noise-induced asymptotic bias of the ICA estimation (Beckmann et al., 2005; Cordes and Nandy, 2007). For a given number of ICs, the ICs can be solved by a maximum likelihood estimation using the FastICA algorithm (Hyvarinen and Oja, 2000; Beckmann and Smith, 2004). In the current work, these algorithms are not our focus and therefore their details are not explained here.

2.5.2. Group-level components—TC-GICA was used to generate group-level components across all participants and sessions (Beckmann et al., 2005). This approach consists of 3 fundamental steps: 1) Estimation of a mean covariance matrix: all 75 individual fMRI datasets are spatially concatenated in MNI152 standard space and used to estimate the mean covariance matrix; 2) PCA reduction of individual datasets: for a given number of ICs, the mean covariance matrix spans a common subspace for all fMRI data. All individual fMRI data were projected into this common subspace to reduce the individual fMRI data; and 3) Probabilistic ICA on temporally concatenated data: all reduced individual data were temporally concatenated and fed into the ICA algorithm(1). This procedure produces the final group-level components. Specifically, it is formularized as the following:

$$\begin{bmatrix} \mathbf{X}_{11}^* \\ \mathbf{X}_{21}^* \\ \mathbf{X}_{31}^* \\ \mathbf{X}_{12}^* \\ \mathbf{X}_{22}^* \\ \mathbf{X}_{32}^* \\ \vdots \\ \mathbf{X}_{1,25}^* \\ \mathbf{X}_{2,25}^* \\ \mathbf{X}_{3,25}^* \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{21} \\ \mathbf{A}_{31} \\ \mathbf{A}_{12} \\ \mathbf{A}_{22} \\ \mathbf{A}_{32} \\ \vdots \\ \mathbf{A}_{1,25} \\ \mathbf{A}_{2,25} \\ \mathbf{A}_{3,25} \end{bmatrix} \mathbf{S} + \begin{bmatrix} \mathbf{E}_{11} \\ \mathbf{E}_{21} \\ \mathbf{E}_{31} \\ \mathbf{E}_{12} \\ \mathbf{E}_{22} \\ \mathbf{E}_{32} \\ \vdots \\ \mathbf{E}_{1,25} \\ \mathbf{E}_{2,25} \\ \mathbf{E}_{3,25} \end{bmatrix}. \quad (2)$$

Here, \mathbf{X}_{ij}^* is the dimension-reduced fMRI data from the i -th scan of the j -th participant and \mathbf{A}_{ij} , \mathbf{E}_{ij} are the relevant mixing matrix and the background noise matrix ($1 \leq i \leq 3$, $1 \leq j \leq 25$);

S includes the ICs shared by all 75 individual scans (i.e., group-level components). All components are standardized into Z-score maps by dividing the relevant component weight by the standard deviation of the background noise. Thus, the group-level component measures not only the raw component but also a signal-to-noise ratio (SNR). Finally, a spatial mixture model is applied to the Z-score map to infer whether the voxels were significantly modulated by the associated timeseries ($p > 0.5$). In the current work, to examine large-scale spatial networks, we fixed the number of components to 20 and performed a 20-component MELODIC (Smith et al., 2009). In addition, we also used MELODIC to automatically estimate the number of components, and conducted a second MELODIC using the estimated number of components (42). Of note, we based the TC-GICA analysis on the entire dataset from the 25 participants across the three scans (a total of 75 scans). This approach was adopted to: 1) provide a realistic equivalent of population-based studies that tend to base their group-level analyses on the large dataset (controls and patients) rather than a single subset (Filippini et al., 2009; Rombouts et al., 2009; Jafri et al., 2008; Damoiseaux et al., 2008; Wolf et al., 2008; Calhoun et al., 2004), and 2) obtain the best possible ICA estimation, independent of any single session's noise.

2.6. Dual regression approach

In order to assess the TRT reliability of each group-level component, we used the dual regression approach to build individual-level DR components (Filippini et al., 2009; Beckmann et al., 2009). This method is based on the following GLM dual regression equations:

$$\mathbf{X}_{ij}^t = \mathbf{S}^t (\mathbf{A}_{ij}^{(1)})^t + (\mathbf{E}_{ij}^{(1)})^t, \quad \mathbf{X}_{ij} = \mathbf{A}_{ij}^{(1)} \mathbf{S}_{ij} + \mathbf{E}_{ij}^{(2)}, \quad 1 \leq i \leq 3; 1 \leq j \leq 25. \quad (3)$$

In equation (3), \mathbf{X}_{ij} represents the fMRI data from the i -th scan of the j -th participant. As in a previous study (Calhoun et al., 2004), the first part of equation (3) uses **unthresholded** group-level components **S** as the spatial predictors of the individual fMRI volumes and results in the regression matrix $\mathbf{A}_{ij}^{(1)}$ containing the relevant individual regression weights in the time domain (i.e., timeseries). These timeseries were then used as the temporal predictors for the individual fMRI timeseries in the second regression equation. The resulting regression matrix \mathbf{S}_{ij} contains regression weights for each of the components in the spatial domain, which serve as our measure of functional connectivity (i.e., the individual-level DR components). These individual-level DR components were subsequently used to evaluate the TRT reliability of the group-level components.

Supplementary analyses examined whether or not TRT reliability is impacted by the addition of variance normalization on the temporal predictors prior to the second-level regression. Variance normalization is commonly used to remove the potential impact of amplitude information on regression results, placing greater emphasis on the shape of the timeseries instead. The comparison of TRT reliability between the two approaches is presented in the supplementary materials: Part 2 - Impact of Several Factors on TRT Reliability.

2.7. Test-retest reliability

As in our prior work (Shehzad et al., 2009; Zuo et al., 2010), we used intra-class correlation coefficients (ICC) to assess TRT reliability. Multiple variants of ICC exist, each with different advantages and limitations (Shrout and Fleiss, 1979). The specific form used here is:

$$ICC_c = \frac{MS_p - MS_e}{MS_p + (d - 1) MS_e}. \quad (4)$$

In this equation, MS_p is the between-participants mean square and MS_e is the error mean square. The relevant statistical computation and interpretation can be found in Appendix A. Of note, this ICC_c calculation does not penalize for systematic errors (e.g., differences in ICs between scans). Its usage is recommended for TRT situations in which systematic errors due to learning effects or fatigue effects may occur (e.g., over the duration of the scan session), which are not directly related to reliability (Rousson et al., 2002; Weir, 2005). We have also analyzed our data using a form that does penalize for systematic error (the form as noted in Appendix A) and supplied the results of TRT reliability analyses as Figure S4 (inter-session ICC_u maps) and Figure S5 (relevant rank ordering using ICC_u).

In the present work, we assessed TRT reliability as follows: 1) TC-GICA was first carried out across all participants and sessions (total 75 scans) to identify group-level components; 2) dual regression was used to reconstruct the individual-level DR components; and 3) at each voxel, intra-session TRT reliability was calculated as the ICC_c between scan 2 and scan 3. Inter-session TRT reliability was computed as the ICC_c between scan 1 and the mean of scans 2 and 3. Given the close temporal proximity of scans 2 and 3 (45 minutes apart) relative to scan 1 (5 – 16 moths), we averaged scans 2 and 3 together for our estimation of the long-term TRT reliability, rather than using a single scan from the second session (e.g., scan 2 or scan 3) (Zuo et al., 2010). This decision was made in order to obtain the best possible estimate for the second session data, and thus long-term TRT reliability.

2.8. Reproducibility of group-level components via individual-level ICA

We examined the extent to which group-level components detected via TC-GICA are reproducible via individual-level ICA. First, given the number of ICs, we performed an individual probabilistic ICA for each of the 75 scans (25 participants \times 3 scans). Then, for each of the group-level components, we identified the closest match in each of the 75 individual-level ICA runs using a template matching procedure based upon the spatial correlation in equation (5). We selected Pearson spatial correlation as our matching metric because of its frequent usage in the literature (Garritty et al., 2007; Esposito et al., 2006; Harrison et al., 2008).

$$r_k^{ij} = \frac{\sum_{n=1}^N \left(IC_k^{\text{tmp}}(n) - \bar{IC}_k^{\text{tmp}} \right) \left(IC_k^{ij}(n) - \bar{IC}_k^{ij} \right)}{\left[\sum_{n=1}^N \left(IC_k^{\text{tmp}}(n) - \bar{IC}_k^{\text{tmp}} \right)^2 \sum_{n=1}^N \left(IC_k^{ij}(n) - \bar{IC}_k^{ij} \right)^2 \right]^{1/2}} \quad (5)$$

In (5), r_k^{ij} measures the degree to which a given individual-level component from the individual-level ICA on the j -th ($j = 1, 2, \dots, 25$) participant's i -th ($i = 1, 2, 3$) scan (i.e., IC_k^{ij}), matches the k -th template component IC_k^{tmp} ($k=1, \dots, P$). Before we calculated the matching score, we found the voxel with the maximum absolute Z-score, and multiplied the volume by the sign of the Z-score. This ensures that the voxel with the greatest absolute Z-score is positive. The matching score for each best-matched individual-level component can be used to measure the reproducibility of corresponding group-level component at the individual-level. For an easy understanding of such a template matching procedure, the process is illustrated in Figure 1B.

Of note, while using the template matching score for group-based analysis, a well-considered study must perform a very careful visual inspection of the matching results (e.g. Harrison et al., 2008). In the current study, in considering the issue of the stability of template matching, we repeated our template-matching procedure using individual-level DR component as the

template (i.e., DR template matching) for a given group-level IC, rather than the overall group-level component map (i.e., TC-GICA template matching). The overlap ratio of individual-level components between the two template matching strategies was used to measure the stability of the template matching. Specifically, for each group-level component, the ratio of overlap of all individual-level components across the two template matching strategies was computed as the agreement between the two template matching procedures as defined in equation (6).

$$R_k = \frac{\#\{1 \leq m \leq 75: IC_k^{\text{TC-GICA}}(m) = IC_k^{\text{DR}}(m)\}}{75} \quad (6)$$

where $IC_k^{\text{TC-GICA}}(m)$ is the m -th individual-level component matched by the k -th TC-GICA template, and $IC_k^{\text{DR}}(m)$ is the m -th individual-level component matched by the k -th DR template of the m -th scan. $\#$ is the number of agreements between the two template matching procedures.

2.9. Rank ordering group-level components

We used two ordering methods to rank the group-level components. First, for each group-level component, a histogram of the ICC values within the group-level component mask (i.e., a binary version of the thresholded group-level component) was used to demonstrate the ranges of TRT reliability across different components. We then took the first most frequent ICC value $ICC^{(k,1)}$ and the second most frequent ICC value $ICC^{(k,2)}$ from each ICC histogram. After this feature selection step, each TRT reliability map can be simplified to a bivariate sample dataset $(ICC^{(k,1)}, ICC^{(k,2)})$ where $k = 1, 2, \dots, P$. These TRT reliability data then can be ranked using a method for sorting multivariate observations based on a minimal spanning tree (see Appendix B). The second index for ranking group-level components was the mean Pearson's correlation coefficient between the individual-level DR template and the best matched individual-level component (i.e., the reproducibility).

2.10. Reliable algorithms for estimating the number of components

For each individual resting state fMRI dataset, we applied five algorithms to estimate the number of components automatically: Akaike information criterion (Akaike, 1973), Bayesian information criterion (Schwarz, 1978), minimum description length (Rissanen, 1978), probabilistic PCA (Minka, 2001) and entropy rate enhanced minimum description length (Li et al., 2007). The first four methods were performed using MELODIC in FSL. The last method was carried out in a Group ICA fMRI Toolbox (GIFT; version 1.3f). Finally the ICC_c was applied to evaluate the TRT reliability of each method.

3. Results

We carried out 20-component and 42-component TC-GICA. All TRT reliability and reproducibility analyses were performed equally for both ICA decompositions. In this section, we only present the results of the 20-component ICA decomposition. The findings of the 42-component decomposition are summarized in the supplementary materials: Part 1 - Reliability and Reproducibility for 42-Component MELODIC Analysis Combining Dual Regression.

3.1. Group-level components

Figure 2 depicts the 20 component maps generated by carrying out a lower order (number of components fixed to 20) TC-GICA across all participants and sessions. The component ordering for the figure was based on the relative ranking of the percentage of variance explained by each component. The components observed were highly consistent with those previously

reported (van de Ven et al., 2004; Beckmann et al., 2005; DeLuca et al., 2006; Damoiseaux et al., 2006; Smith et al., 2009; Kiviniemi et al., 2009).

We sorted the 20 components into two broad classes - functionally relevant components (i.e., ICNs) and scanner/physiological artifactual components - based on visual inspection of each component's spatial profile (e.g., biological plausibility, comparability to patterns previously reported in ICA-based studies) and timeseries-based power spectrum profile (e.g., whether or not signals < 0.1Hz were prominent). We noted 4 components that appeared to be associated with artifactual sources: cerebrospinal fluid (IC01), white matter (IC03), head motion (IC05), and large vessels (IC16). These four components accounted for 39.4% of the total variance in the resting state fMRI data. Several functionally relevant components consistent with prior reports were also revealed in our results. Two components (IC04 and IC15) are involved in vision. IC09 combines visual and motor regions including the occipital pole, superior parietal cortex and precentral gyrus. IC13 includes brain regions such as the primary motor cortex and primary and association auditory cortices. Several components include regions related to various high-order brain functions: fronto-parietal networks corresponding to cognition and language functions (IC07 and IC19), medial-frontal including anterior cingulate and paracingulate associated with executive control (IC08) and three “default mode” networks (IC10, IC12 and IC14). We found six other components that are rarely reported or investigated systematically corresponding to the cerebellum (IC11 and IC18), a motor-striatal component (IC02), a ventromedial prefrontal component (IC17), a brainstem component (IC06), and a temporal-lobe component (IC20). Of note, we found several components that exhibit anti-correlation relationships between regions (IC04, IC08, IC14 and IC15). In particular, the executive and attentional network (IC08) and the “default mode” network (IC14) demonstrated prominent anti-correlation relationships (Figure S1).

We detected the classic “default mode” network, although in the form of three components that we interpret as sub-networks. The first is a medial-prefrontal subsystem (IC12), the second is a posterior cingulate/precuneus subsystem (IC10), and the third is a temporal subsystem (IC14). These three subsystems mainly overlap in the posterior cingulate cortex and medial prefrontal cortex (Figure S2). As we discuss below, the existence of three overlapping but differentiable sub-networks may account for some of the variations in the specific spatial distributions or functional specialization of the “default mode” network reported across ICA studies (Buckner et al., 2008; Harrison et al., 2008).

3.2. Test-retest reliability

For each group-level component, the voxel-wise intra-session (Figure 3) and inter-session (Figure 4) TRT reliability maps were generated and summarized by calculating the first most frequent ICC value (i.e., mode; see Figure 5) in their histograms (Figure S3). Consistent with our predictions, most regions within noise components 1, 3 and 5 showed relatively low TRT reliability (modal ICC: 0.2 – 0.45) both within and between sessions. Noise component 16 (large vessels) exhibited moderate reliability within-session (modal ICC: 0.525), but lower between-session reliability (modal ICC: 0.375). This suggests that cardiac factors may be more stable within-session than over the longer term.

Also consistent with our predictions, components commonly associated with sensory, motor, higher order cognitive function and the “default-mode” network all exhibited moderate to high TRT reliability both within (modal ICC: 0.5 – 0.65) and between (modal ICC: 0.45 – 0.65) sessions. The six remaining components showed relatively low reliability (modal within-session ICC: 0.275 – 0.4; between-session ICC: 0.2 – 0.375), with the exception of a cerebellar component (modal within-session ICC: 0.575; modal between-session ICC: 0.5).

3.3. Reproducibility

Our reproducibility findings indicated marked variability across the 20 components with respect to their reproducibility at the individual scan level, regardless of the template matching approach employed (see Figure 6). Greatest reproducibility was noted for IC04 (i.e., medial visual network), whose mean matching score was 36% (for DR template matching) and 62% (for TC-GICA template matching) greater than the mean of the remaining components' mean matching scores. The lowest reproducibility was noted for artifact components IC01, IC03 and IC20, as well as IC18 (cerebellum). The mean match score between the template and individual-level components was 34% greater across components when we employed the DR template matching, rather than TC-GICA template matching (paired t-tests: $p < 10^{-11}$). However, the components selected by the two template matching approaches were frequently contradictory, except for IC04, which had the highest reproducibility. Figure 6 indicates the number of disagreements on the matching results between the two template matching procedures. This finding is particularly worrisome since the most commonly employed approach for template matching in the ICA literature is based on an overall group template matching (Mantini et al., 2009; Greicius et al., 2004, 2007; Damoiseaux et al., 2008).

Finally, we used a random effects model to construct a group statistical map for each of the 20 group-level components, using the best matched individual-level components across the 75 scans, for each of the two template matching procedures. A relatively high degree of concordance was noted between the group statistical maps produced using the two matching procedures, as well as with the TC-GICA group-level components (see Figure 7). This suggests that while template matching can produce numerous misidentifications of individual-level components, these misidentifications do not appreciably affect the resulting group statistical maps. This is a source of concern, as the concordance of group statistical maps can give the erroneous impression that template matching is effective, when this is frequently not the case.

3.4. Ranking and ordering group-level components

Based on the histogram of the TRT reliability for each group-level component, we generated a minimal spanning tree to rank components. Figure 8 shows the group-level components ranked with the nodes denoting the minimal spanning tree. The fronto-parietal (IC19 and IC07), visual (IC04), "default mode" (IC10 and IC14), auditory-motor (IC13), executive control (IC08) and one cerebellum component (IC11) demonstrated the highest ranks for both intra- and inter-session TRT reliability. In contrast, the low reliability ranks corresponded to those noise or artifact components associated with white matter, head motion, cerebrospinal fluid, and presumed imaging artifacts. Additionally, the other cerebellum component (IC18), the motor-striatal component (IC02), the ventromedial prefrontal component (IC17), and the temporal-lobe component (IC20) all ranked low in TRT reliability. This lack of stability across scan sessions might be related to unknown imaging artifacts or other sources of variability.

The group-level components were also ranked by their reproducibility. Again, functionally relevant components (i.e., ICNs) ranked highly in reproducibility, while components apparently dominated by physiological noise and scanner artifact were less reproducible (Figure 9).

3.5. Estimated number of components

To assess component number estimation algorithms, we applied five estimation approaches to approximate the number of components in the resting state fMRI signals. Figure 10 illustrates the bar graphs of the numbers of components across the five algorithms. The probabilistic PCA-based Laplacian algorithm produced a mean of around 20 components (Scan1: 23.0 ± 12.3 ; Scan2: 19.4 ± 3.5 ; Scan3: 20.8 ± 3.8). Built-in methods in *MELODIC* such as Akaike information criterion, Bayesian information criterion, and minimum description length produce smaller

estimations than the probabilistic PCA method. The entropy rate enhanced minimum description length method generates similar estimates to the minimum description length method. The intra- and inter-session ICC values were calculated (Table S1). All five algorithms exhibit high short-term reliability (ICC > 0.74) and moderate to poor long-term reliability. Probabilistic PCA had the lowest intra-session and inter-session TRT reliability.

4. Discussion

The present work evaluated the reliability and reproducibility of TC-GICA-derived measures of RSFC, yielding four main results. First, we found moderate-to-high short- and long-term TRT reliability for RSFC measures derived by combining TC-GICA and dual regression. Exceptions to this finding were limited to components corresponding to physiological- and imaging-related artifacts. Second, we found that the degree to which components detected using TC-GICA were reproducible at the level of a single-scan ICA varied across components but not sessions, with the medial visual component having the highest reproducibility. Importantly, our analyses revealed the potential for worrisome levels of inaccuracy in template matching procedures. This inaccuracy potential appeared to affect most components, but could easily be overlooked. Third, we found the reliability- and reproducibility-based rank orderings of the TC-GICA components were highly consistent, with functionally relevant ICNs ranking notably higher than noise components. Finally, we demonstrated that component estimation algorithms tend to exhibit high intra-session test-retest reliability, but low to moderate inter-session reliability, with poor concordance among algorithms.

4.1. The test-retest reliability and reproducibility

The greater reliability of ICNs compared to noise components provides support for the notion that stable trait differences exist among individuals with respect to their ICN characteristics (Shehzad et al., 2009; Buckner et al., 2009; Smith et al., 2009; Damoiseaux et al., 2006). This statement is not inferring that there is no dynamic range for ICN characteristics – quite the contrary. Numerous studies have documented the impact of factors such as task demands (Esposito et al., 2006; Fransson, 2006; Hampson et al., 2006; Harrison et al., 2008; Kelly et al., 2008; Yan et al., 2009), arousal (Boly et al., 2008; Fukunaga et al., 2006), and medication status (Kelly et al., 2009b) on ICN properties. As such, we posit that ICNs in each individual (whether it be the spatial extent or strength of functional connectivity) exist at equilibrium, around which transient changes in their properties may occur in response to both external influences and state-related factors. Of note, several papers have argued that development (Fair et al., 2009; Kelly et al., 2009a) and learning (Albert et al., 2009) can produce long-term changes in these properties. It is worth noting that the most reliable ICNs were those related to fronto-parietal regions supporting attentional selection and encoding (IC19) (Milham et al., 2001), primary visual perception (IC15), and “default mode” function (IC14) (Gusnard et al., 2001). The stability of these ICNs may reflect the sustained reliance of the brain on these networks for moment-to-moment functions. Alternatively, it may suggest these networks have a more consistent response to the “resting state” fMRI environment.

Our reproducibility analyses raised significant concern about the usage of individual scan based ICA approaches and template matching. The two approaches to identifying group-level components at the individual scan-level exhibited a high degree of discordance with respect to the component selected as the best fit for an individual. Yet, across all scans, each of the template matching approaches managed to capture the gestalt of the targeted group-level component. This should caution against the suggested use of matching scores as a metric of ICN integrity, and consequently, as potentially promising clinical biomarkers (Greicius et al., 2004). In this regard, dual regression based approaches clearly engender a higher degree of confidence.

4.2. Sub-networks of the “default mode” network

In the present work, the “default mode” network was divided into three sub-networks: the medial prefrontal sub-network, the posterior cingulate/precuneus sub-network and the medial temporal lobe sub-network. Each of the three sub-networks exhibited high test-retest reliability. This component-splitting phenomenon previously was thought to be a result of overestimating the model order, but recently was reported to reflect functional segregation or hierarchy within the “default mode” network (Buckner et al., 2008; Kiviniemi et al., 2009; Smith et al., 2009; Uddin et al., 2009; Harrison et al., 2008). Intriguingly, the three sub-networks we observed are well matched with the anatomical and functional theory of the “default mode” network proposed by Buckner et al. (2008). For example, the three sub-networks overlapped in the posterior cingulate cortex and medial prefrontal cortex, which are posited to be critical to interactions between the medial prefrontal subsystem and the medial temporal lobe subsystem (Buckner et al., 2008).

4.3. Anti-correlations in ICA-derived ICNs

Our TC-GICA analyses revealed multiple ICNs that were accompanied by anti-phasic or “anti-correlated” functional networks. Since Fox et al. (2005) systematically demonstrated the anti-correlation between the “default mode” network and task-positive network, the anti-correlation has been suggested to reflect either a competitive mechanism or segregation of processing between functional systems (Fox et al., 2005; Kelly et al., 2008). Recently, concerns have arisen regarding the potential artifactual nature of findings of negative connectivity, resulting from global corrections procedures commonly employed in conjunction with seed-based approaches (Murphy et al., 2009; Fox et al., 2009; Weissenbacher et al., 2009). Here, TC-GICA generated a pattern of negative connectivity between the “default mode” network and an anterior cingulate-based network involving regions supporting executive control and attentional function (Long et al., 2008). Several recent theoretical models support the existence of anti-correlations (Honey et al., 2007; Steyn-Ross et al., 2009; Deco et al., 2009), suggesting that negative connectivity may be a true organizational characteristic of the resting state brain, with neural origins (Popa et al., 2009). Of note, our results suggest less prominent negative connectivity between the “default mode” network and the dorsal visual attention stream (IC09), than prior work using seed-based approaches (Fox et al., 2006). It may be beneficial for future studies to investigate such inconsistencies.

4.4. Limitations and future directions

Despite its successes, the present work has several limitations worth noting. First, usage of random initial values is a well-documented source of variability for ICA. Recent work by Himberg et al. (2004) has suggested that a potential solution to this problem is to carry out a series of ICA analyses using different initial values, and then combining the results via hierarchical clustering. The focus of the present work was on assessing the reliability and reproducibility for a single ICA run, rather than the aggregate result. In part, this decision was made to reflect the more common approach currently employed in the literature. However, future work may directly quantify the impact of random initial values on reliability and reproducibility, and in addition to look at how solutions such as that of Himberg et al. (2004) mediate it.

Second, our examination of TRT reliability focused on the combination of TC-GICA and dual regression, though does not speak to either step individually. In the first stage, dual regression can be applied to any template brain map, not just those derived from ICA. As such, future work may focus on the dual regression step specifically, using template images derived by alternative approaches, such as those recently developed by Smith et al. (2009) using the BrainMap database. Similarly, in the second stage, dual regression can use step-wise regression or univariate regression, not only multiple regression. Future work may examine the impact of

different second-level regression strategies on TRT reliability. Finally, it would be worthwhile to explore the effects of ICA analyses using alternative approaches for obtaining individual RSFC maps following group analyses (e.g., PCA back-projection) (Calhoun et al., 2001). The test-retest data used in this study are publicly available via http://www.nitrc.org/projects/nyu_trt to facilitate the above or other examinations in future.

Third, the present work relied upon the template matching using a commonly employed measure of goodness of fit the spatial correlation coefficient. Unfortunately, to date, there is no standard approach to determining what an adequate match score is, and misclassification using this template matching approach is clearly possible for components with a high degree of similarity (e.g., “default mode” network components: IC10, IC12 and IC14 in Figure S2). In fact, our results actually showed that the template matching procedure could be highly affected by the template selection strategy. Alternative strategies exist to calculate the template matching fitness (Greicius et al., 2004, 2007), though they likely face the same issues we demonstrated. Future work may focus on the determination of an optimal matching procedure.

4.5. Conclusions

In summary, temporal concatenation group ICA (TC-GICA), combined with dual regression is an effective approach for exploratory analyses of resting state fMRI data, exhibiting moderate to high test-retest reliability for intrinsic connectivity networks. Additionally, while many components detected via TC-GICA are reproducible at the individual scan level, our findings raise cautions with respect to group-based analyses that rely on individual scan-based ICA approaches due to the challenges of component matching.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported, in part, by grants from NIMH (R01MH081218), and the Stavros S. Niarchos Foundation to F.X.C., and from the Leon Levy Foundation to M.P.M.; and by gifts from Linda and Richard Schaps, and Jill and Bob Smith to F.X.C. We desire to express our thanks to Dr. Bharat B. Biswal for helpful advice and Dr. Christian F. Beckmann for invaluable discussions on the dual regression method. Dr. Maarten Mennes helped to improve the readability of the manuscript.

Appendix

A. Computation of Intraclass Correlation Coefficients

To calculate the TRT reliability of RSFC for each voxel in each group-level component, for each voxel, we consider a random sample of n subjects with d repeated measurements of a continuous variable Y characterizing the RSFC. We denote Y_{ij} as the j -th measurement made on the i -th subject (for $i = 1, \dots, n$ and $j = 1, \dots, d$). In the current situation, Y_{ij} denotes the RSFC of a component from the i -th participant's j -th measuring occasions. We apply a voxel-wise single-factor ANOVA on each component as the following decomposition of Y_{ij} :

$$Y_{ij} = \mu + p_i + t_j + e_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq d \quad (7)$$

where μ is a fixed parameter and p_i , t_j , e_{ij} are independent random effects normally distributed with mean 0 and variances σ_p^2 , σ_t^2 and σ_e^2 . The term p_i is the participant effect, whereas t_j is the systematic error (i.e., the scanning occasion effect) and e_{ij} is the measurement error. One way

to assess reliability of variable Y may use the concept of intraclass correlation (ICC) defined as

$$\rho_u = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_t^2 + \sigma_e^2}. \quad (8)$$

Obviously, the ICC has the desired property to characterize the test-retest reliability, i.e., becoming smaller when either σ_t^2 or σ_e^2 become larger. Also, if one does not consider systematic error when assessing reliability, one may use the following equation:

$$\rho_c = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}. \quad (9)$$

Solving the ANOVA model (7), we can estimate the sample version of ρ_u and ρ_c . Briefly, from data Y_{ij} , we compute three sums of squares:

$$\begin{aligned} SS_p &= d \sum_{i=1}^n (\bar{Y}_{i.} - \bar{Y}_{..})^2 \\ SS_t &= n \sum_{j=1}^d (\bar{Y}_{.j} - \bar{Y}_{..})^2 \\ SS_e &= \sum_{i=1}^n \sum_{j=1}^d (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 \\ \bar{Y}_{i.} &= \frac{\sum_{j=1}^d Y_{ij}}{d}, \quad \bar{Y}_{.j} = \frac{\sum_{i=1}^n Y_{ij}}{n}, \quad \bar{Y}_{..} = \frac{\sum_{i=1}^n \sum_{j=1}^d Y_{ij}}{nd}. \end{aligned}$$

The expectations of $MS_p = SS_p/(n-1)$, $MS_t = SS_t/(d-1)$ and $MS_e = SS_e/((n-1)(d-1))$ are $d\sigma_p^2 + \sigma_e^2$, $n\sigma_t^2 + \sigma_e^2$ and σ_e^2 , respectively. We therefore obtain unbiased estimates of σ_p^2 , σ_t^2 and σ_e^2 as

$$\begin{aligned} \hat{\sigma}_p^2 &= \frac{MS_p - MS_e}{d} \\ \hat{\sigma}_t^2 &= \frac{MS_t - MS_e}{n} \\ \hat{\sigma}_e^2 &= MS_e. \end{aligned}$$

Then an estimate of ρ_u is

$$ICC_u = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_t^2 + \hat{\sigma}_e^2} = \frac{MS_p - MS_e}{MS_p + (d-1)MS_e + (d/n)(MS_t - MS_e)} \quad (10)$$

while an estimate of ρ_c is

$$ICC_c = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2} = \frac{MS_p - MS_e}{MS_p + (d-1)MS_e}. \quad (11)$$

B. Ranking TRT Reliability Maps Using Minimal Spanning Trees

Using the bivariate feature $(ICC_c^{(k,1)}, ICC_c^{(k,2)})$, $k = 1, 2, \dots, P$ of each ICC histogram with 80 bins, we can generate a fully connected graph $G = (V, E)$ with P nodes. A minimal spanning tree (MST) of the graph G is a tree subgraph that contains all nodes and whose edges have the least total distance. We use the Kruskal algorithm for forming the MST of G .

Algorithm 1 (Kruskal (1956)). Formation of a minimal spanning tree T from a connected graph with P nodes and with edge distances in E :

- 0 Set $T = \emptyset$ (the empty set), and $k = 0$.
- 1 Choose the edge e from E with the shortest distance.
- 2 Remove e from E .
- 3 If e does not create a cycle in T , then add e to T and set $k = k + 1$
- 4 if $k < P - 1$, go to step 1.

Based on the MST of G , we rank these nodes (i.e., components in the current study) by defining a starting node at an endpoint of a tree diameter and then proceed through the tree in such a way as to visit any shallow subtrees at a given node before proceeding to the deeper subtrees. In our situation, we choose to begin on the right (i.e., large TRT reliability) of the tree (Figure 8) and use the ordering rule to get the final ordering for these group-level components.

References

- Akaike, H. Information theory and an extension of the maximum likelihood principle.. Second International Symposium on Information Theory; 1973. p. 267-281.
- Albert NB, Robertson EM, Miall RC. The resting human brain and motor learning. *Curr Biol* 2009;19(12):1023–7. [PubMed: 19427210]
- Andersson, JLR.; Jenkinson, M.; Smith, S. Non-linear registration, aka spatial normalisation. 2007. Tech. Rep. TR07JA2, Oxford Centre for Functional MRI of the Brain.
- Beckmann, C.; Mackay, C.; Filippini, N.; Smith, S. Group comparison of resting-state fmri data using multi-subject ica and dual regression.. 15th Annual Meeting of Organization for Human Brain Mapping; 2009. poster 441 SU-AM
- Beckmann CF, DeLuca M, Devlin JT, Smith SM. Investigations into resting-state connectivity using independent component analysis. *Philos Trans R Soc Lond, B, Biol Sci* 2005;360(1457):1001–13. [PubMed: 16087444]
- Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging* 2004;23(2):137–52. [PubMed: 14964560]
- Biswal B, Yetkin F, Haughton V, Hyde J. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med* 1995;34(4):537–541. [PubMed: 8524021]
- Boly M, Phillips C, Tshibanda L, Vanhaudenhuyse A, Schabus M, Dang-Vu TT, Moonen G, Hustinx R, Maquet P, Laureys S. Intrinsic brain activity in altered states of consciousness: how conscious is the default mode of brain function? *Ann N Y Acad Sci* 2008;1129:119–29. [PubMed: 18591474]
- Broyd S, Demanuele C, Debener S, Helps S, James C, Sonuga-Barke E. Default-mode brain dysfunction in mental disorders: A systematic review. *Neuroscience & Biobehav. Rev* 2009;33(3):279–96.
- Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 2008;1124:1–38. [PubMed: 18400922]
- Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, Andrews-Hanna JR, Sperling RA, Johnson KA. Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to alzheimer's disease. *Journal of Neuroscience* 2009;29(6):1860–1873. [PubMed: 19211893]

- Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional mri data using independent component analysis. *Hum. Brain Mapp* 2001;14(3):140–51. [PubMed: 11559959]
- Calhoun VD, Pekar JJ, Pearlson GD. Alcohol intoxication effects on simulated driving: exploring alcohol-dose effects on brain activation using functional MRI. *Neuropsychopharmacology* 2004;29(11):2097–17. [PubMed: 15316570]
- Calhoun VD, Liu J, Adali T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 2009;45(1 Suppl):S163–72. [PubMed: 19059344]
- Chen S, Ross T, Zhan W, Myers C, Chuang K, Heishman S, Stein E, Yang Y. Group independent component analysis reveals consistent resting-state networks across multiple sessions. *Brain Research* 2008;1239(6):141–151. [PubMed: 18789314]
- Cordes D, Nandy R. Independent component analysis in the presence of noise in fmri. *Magnetic Resonance Imaging* 2007;25(9):1237–48. [PubMed: 17509787]
- Damoiseaux JS, Beckmann CF, Arigita EJS, Barkhof F, Scheltens P, Stam CJ, Smith SM, Rombouts SARB. Reduced resting-state brain activity in the “default network” in normal aging. *Cereb Cortex* 2008;18(8):1856–64. [PubMed: 18063564]
- Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci USA* 2006;103(37):13848–53. [PubMed: 16945915]
- Deco G, Jirsa V, McIntosh AR, Sporns O, Kötter R. Key role of coupling, delay, and noise in resting brain fluctuations. *Proc Natl Acad Sci USA* 2006;106(25):10302–7. [PubMed: 19497858]
- DeLuca M, Beckmann CF, Stefano ND, Matthews PM, Smith SM. fmri resting state networks define distinct modes of long-distance interactions in the human brain. *NeuroImage* 2006;29(4):1359–67. [PubMed: 16260155]
- Di Martino A, Scheres A, Margulies DS, Kelly AMC, Uddin LQ, Shehzad Z, Biswal B, Walters JR, Castellanos FX, Milham MP. Functional connectivity of human striatum: a resting state fmri study. *Cereb Cortex* 2008;18(12):2735–47. [PubMed: 18400794]
- Di Martino A, Shehzad Z, Kelly C, Roy AK, Gee DG, Uddin LQ, Gotimer K, Klein DF, Castellanos FX, Milham MP. Relationship between cingulo-insular functional connectivity and autistic traits in neurotypical adults. *The Am J of Psychiatry* 2009;166(8):891–9.
- Esposito F, Bertolino A, Scarabino T, Latorre V, Blasi G, Popolizio T, Tedeschi G, Cirillo S, Goebel R, Salle FD. Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Res Bull* 2006;70(4-6):263–9. [PubMed: 17027761]
- Fair DA, Cohen AL, Power JD, Dosenbach NUF, Church JA, Miezin FM, Schlaggar BL, Petersen SE. Functional brain networks develop from a “local to distributed” organization. *PLoS Computational Biology* 2009;5(5):e1000381. [PubMed: 19412534]
- Filippini N, Macintosh B, Hough M, Goodwin G, Frisoni G, Smith S, Matthews P, Beckmann C, Mackay C. Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. *Proc Natl Acad Sci USA* 2009;106(17):7209–7214. [PubMed: 19357304]
- Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci* 2007;8(9):700–11. [PubMed: 17704812]
- Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc Natl Acad Sci USA* 2006;103(26):10046–51. [PubMed: 16788060]
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Essen DCV, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA* 2005;102(27):9673–8. [PubMed: 15976020]
- Fox MD, Snyder AZ, Vincent JL, Raichle ME. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron* 2007;56(1):171–84. [PubMed: 17920023]
- Fox MD, Zhang D, Snyder AZ, Raichle ME. The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology* 2009;1–47.
- Franco A, Pritchard A, Calhoun V, Mayer A. Interrater and inter-method reliability of default mode network selection. *Hum. Brain Mapp* 2009;30(7):2293–303. [PubMed: 19206103]

- Fransson P. How default is the default mode of brain function? further evidence from intrinsic bold signal fluctuations. *Neuropsychologia* Jan;2006 44(14):2836–45. [PubMed: 16879844]
- Fukunaga M, Horovitz S, Vangelder P, Dezwart J, Jansma J, Ikonomidou V, Chu R, Deckers R, Leopold D, Duyn J. Large-amplitude, spatially correlated fluctuations in bold fmri signals during extended rest and early sleep stages. *Magnetic Resonance Imaging* 2006;24(8):979–992. [PubMed: 16997067]
- Garrity AG, Pearlson GD, McKiernan K, Lloyd D, Kiehl KA, Calhoun VD. Aberrant “default mode” functional connectivity in schizophrenia. *American journal of psychiatry* 2007;164(3):450–7. [PubMed: 17329470]
- Greicius M. Resting-state functional connectivity in neuropsychiatric disorders. *Curr Opin Neurol* 2008;21(4):424–30. [PubMed: 18607202]
- Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H, Reiss AL, Schlaggar AF. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological Psychiatry* 2007;62(5):429–37. [PubMed: 17210143]
- Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proc Natl Acad Sci USA* 2004;101(13):4637–42. [PubMed: 15070770]
- Guo Y, Pagnoni G. A unified framework for group independent component analysis for multi-subject fmri data. *NeuroImage* 2008;42(3):1078–93. [PubMed: 18650105]
- Gusnard DA, Raichle ME, Raichle ME. Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci* 2001;2(10):685–94. [PubMed: 11584306]
- Hampson M, Driesen NR, Skudlarski P, Gore JC, Constable RT. Brain connectivity related to working memory performance. *Journal of Neuroscience* 2006;26(51):13338–43. [PubMed: 17182784]
- Harrison BJ, Pujol J, López-Solà M, Hernández-Ribas R, Deus J, Ortiz H, Soriano-Mas C, Yücel M, Pantelis C, Cardoner N. Consistency and functional specialization in the default mode brain network. *Proc Natl Acad Sci USA* 2008;105(28):9781–6. [PubMed: 18621692]
- Himberg J, Hyvärinen A, Esposito F. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage* 2004;22(3):1214–22. [PubMed: 15219593]
- Honey CJ, Kötter R, Breakspear M, Sporns O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci USA* 2007;104(24):10240–5. [PubMed: 17548818]
- Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Networks* 2000;13(4-5):411–430. [PubMed: 10946390]
- Jafri MJ, Pearlson GD, Stevens M, Calhoun VD. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. *NeuroImage* 2008;39(4):1666–81. [PubMed: 18082428]
- Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 2002;17(2):825–41. [PubMed: 12377157]
- Kelly AC, Martino AD, Uddin LQ, Shehzad Z, Gee DG, Reiss PT, Margulies DS, Castellanos FX, Milham MP. Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cerebral Cortex* 2009a;19(3):640–653. [PubMed: 18653667]
- Kelly AMC, Uddin LQ, Biswal BB, Castellanos FX, Milham MP. Competition between functional brain networks mediates behavioral variability. *NeuroImage* 2008;39(1):527–37. [PubMed: 17919929]
- Kelly C, de Zubicaray G, Martino AD, Copland DA, Reiss PT, Klein DF, Castellanos FX, Milham MP, McMahon K. L-dopa modulates functional connectivity in striatal cognitive and motor networks: a double-blind placebo-controlled study. *Journal of Neuroscience* 2009b;29(22):7364–78. [PubMed: 19494158]
- Kiviniemi V. Independent component analysis of nondeterministic fmri signal sources. *NeuroImage* 2003;19(2):253–260. [PubMed: 12814576]
- Kiviniemi V, Starck T, Remes J, Long X, Nikkinen J, Haapea M, Veijola J, Moilanen I, Isohanni M, Zang Y, Tervonen O. Functional segmentation of the brain cortex using high model order group pica. *Hum. Brain Mapp* 2009;30(12):3865–86. [PubMed: 19507160]

- Kruskal JB. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* 1956;7(1):48–50.
- Li YO, Adali T, Calhoun VD. Estimating the number of independent components for functional magnetic resonance imaging data. *Hum. Brain Mapp* 2007;28(11):1251–66. [PubMed: 17274023]
- Long XY, Zuo XN, Kiviniemi V, Yang Y, Zou QH, Zhu CZ, Jiang TZ, Yang H, Gong QY, Wang L, Li KC, Xie S, Zang YF. Default mode network as revealed with multiple methods for resting-state functional mri analysis. *Journal of Neuroscience Methods* 2008;171(2):349–55. [PubMed: 18486233]
- Mantini D, Caulo M, Ferretti A, Romani GL, Tartaro A. Noxious Somatosensory Stimulation Affects the Default Mode of Brain Function: Evidence from Functional MR Imaging. *Radiology* 2009;253(3):797–804. [PubMed: 19789220]
- Margulies DS, Kelly AMC, Uddin LQ, Biswal BB, Castellanos FX, Milham MP. Mapping the functional connectivity of anterior cingulate cortex. *NeuroImage* 2007;37(2):579–88. [PubMed: 17604651]
- Meindl T, Teipel S, Elmouden R, Mueller S, Koch W, Dietrich O, Coates U, Reiser M, Glaser C. Test-retest reproducibility of the default-mode network in healthy individuals. *Hum. Brain Mapp*. 2009 Electronic publication ahead of print.
- Milham MP, Banich MT, Webb A, Barad V, Cohen NJ, Wszalek T, Kramer AF. The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Brain research Cognitive brain research* 2001;12(3):467–73. [PubMed: 11689307]
- Minka, T. Automatic choice of dimensionality for PCA. *Massachusetts Inst. Technol.; Cambridge*: 2000. Tech. Rep. 514
- Mohammadi B, Kollewé K, Samii A, Krampfl K, Dengler R, Münte T. Changes of resting state brain networks in amyotrophic lateral sclerosis. *Exp Neurol* 2009;217(1):147–153. [PubMed: 19416664]
- Murphy K, Birn R, Handwerker D, Jones T, Bandettini P. The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *NeuroImage* 2009;44(3):893–905. [PubMed: 18976716]
- Nioche C, Cabanis EA, Habas C. Functional connectivity of the human red nucleus in the brain resting state at 3t. *AJNR American journal of neuroradiology* 2009;30(2):396–403. [PubMed: 19022864]
- Perlberg V, Bellec P, Anton J-L, Péligrini-Issac M, Doyon J, Benali H. CORSICA: correction of structured noise in fmri by automatic identification of ica components. *Magnetic Resonance Imaging* 2007;25(1):35–46. [PubMed: 17222713]
- Popa D, Popescu AT, Paré D. Contrasting activity profile of two distributed cortical networks as a function of attentional demands. *J Neurosci* 2009;29(4):1191–201. [PubMed: 19176827]
- Rissanen J. Modeling by shortest data description. *Automatica* 1978;14(5):465–471.
- Rombouts SARB, Damoiseaux JS, Goekoop R, Barkhof F, Scheltens P, Smith SM, Beckmann CF. Model-free group analysis shows altered bold fmri networks in dementia. *Hum. Brain Mapp* 2009;30(1):256–66. [PubMed: 18041738]
- Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statist. Med* 2002;21(22):3431–46.
- Roy AK, Shehzad Z, Margulies DS, Kelly AMC, Uddin LQ, Gotimer K, Biswal BB, Castellanos FX, Milham MP. Functional connectivity of the human amygdala using resting state fmri. *NeuroImage* 2009;45(2):614–626. [PubMed: 19110061]
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978;6(2):461–464.
- Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. *Neuron* 2009;62(1):42–52. [PubMed: 19376066]
- Shehzad Z, Kelly A, Reiss P, Gee D, Gotimer K, Uddin L, Lee S, Margulies D, Roy A, Biswal B, Petkova E, Castellanos F, Milham M. The resting brain: Unconstrained yet reliable. *Cereb Cortex* 2009;19(10):2209–29. [PubMed: 19221144]
- Shrout P, Fleiss J. Intraclass correlations - uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420–428. [PubMed: 18839484]
- Smith S, Fox P, Miller K, Glahn D, Fox P, Mackay C, Filippini N, Watkins K, Toro R, Laird A, Beckmann C. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci USA* 2009;106(31):13040–5. [PubMed: 19620724]

- Steyn-Ross ML, Steyn-Ross DA, Wilson MT, Sleigh JW. Modeling brain activation patterns for the default and cognitive states. *NeuroImage* 2009;45(2):298–311. [PubMed: 19121401]
- Stone, JV.; Porrill, J.; Buchel, C.; Friston, K. Spatial, temporal, and spatiotemporal independent component analysis of fMRI data.. *Proceedings of the 18th Leeds Statistical Research Workshop on Spatial-Temporal Modeling and Its Applications*; 1999. p. 2328
- Thomas CG, Harshman RA, Menon RS. Noise reduction in bold-based fmri using component analysis. *NeuroImage* 2002;17(3):1521–37. [PubMed: 12414291]
- Tohka J, Foerde K, Aron AR, Tom SM, Toga AW, Poldrack RA. Automatic independent component labeling for artifact removal in fmri. *NeuroImage* 2008;39(3):1227–45. [PubMed: 18042495]
- Uddin LQ, Kelly AC, Biswal BB, Castellanos FX, Milham MP. Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. *Hum. Brain Mapp* Feb;2009 30(2): 625–637. [PubMed: 18219617]
- van de Ven VG, Formisano E, Prvulovic D, Roeder CH, Linden DEJ. Functional connectivity as revealed by spatial independent component analysis of fmri measurements during rest. *Hum. Brain Mapp* 2004;22(3):165–78. [PubMed: 15195284]
- Wang Z, Peterson BS. Partner-matching for the automated identification of reproducible ica components from fmri datasets: Algorithm and validation. *Hum. Brain Mapp* 2008;29(8):875–893. [PubMed: 18058813]
- Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C. Correlations and anticorrelations in resting-state functional connectivity mri: a quantitative comparison of preprocessing strategies. *NeuroImage* 2009;47(4):1408–16. [PubMed: 19442749]
- Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19(1):231–40. [PubMed: 15705040]
- Wolf RC, Sambataro F, Vasic N, Schönfeldt-Lecuona C, Ecker D, Landwehrmeyer B. Aberrant connectivity of lateral prefrontal networks in presymptomatic huntington's disease. *Exp Neurol* 2008;213(1):137–44. [PubMed: 18588876]
- Yan C, Liu D, He Y, Zou Q, Zhu C, Zuo X, Long X, Zang Y. Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. *PLoS ONE* 2009;4(5):e5743. [PubMed: 19492040]
- Yang Z, LaConte S, Weng X, Hu X. Ranking and averaging independent component analysis by reproducibility (RAICAR). *Hum. Brain Mapp* 2008;29(6):711–25. [PubMed: 17598162]
- Zhang D, Snyder AZ, Fox MD, Sansbury MW, Shimony JS, Raichle ME. Intrinsic functional relations between human cerebral cortex and thalamus. *Journal of Neurophysiology* 2008;100(4):1740–1748. [PubMed: 18701759]
- Zhang Z, Lu G, Zhong Y, Tan Q, Yang Z, Liao W, Chen Z, Shi J, Liu Y. Impaired attention network in temporal lobe epilepsy: A resting fmri study. *Neuroscience Letters* 2009;458(3):97–101. [PubMed: 19393717]
- Zuo XN, Di Martino A, Kelly C, Shehzad ZE, Gee DG, Klein DF, Castellanos FX, Biswal BB, Milham MP. The oscillating brain: Complex and reliable. *Neuroimage* 2010;49(2):1432–1445. [PubMed: 19782143]

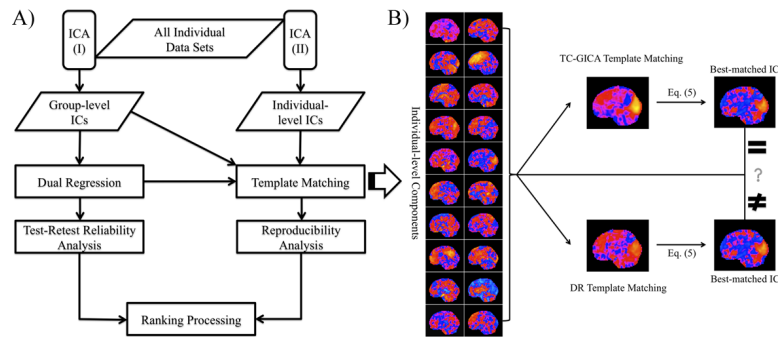


Figure 1. Overall Analytic Strategy

A) All individual datasets were analyzed by two ICA procedures: (I) temporal concatenation group ICA (TC-GICA) and (II) individual-level ICA. TC-GICA first generates the group-level components, and then reconstructs the individual-level independent components (ICs) using a dual regression (DR) approach (i.e., individual-level DR components) for the test-retest reliability analysis. Individual-level ICA produces the individual-level components for all individual datasets and combines a template matching procedure to examine the reproducibility. Measures of reliability and reproducibility are used to rank-order the group-level components. B) The template matching procedure uses two template strategies: group-level components or individual-level DR components. Given a group-level component of interests, for each of the three scans of a participant, the best-matched individual-level component is selected from among their 20 individual-level components based on the matching process outlined above (shown here with only 1 scan dataset from a participant). Firstly, across all brain voxels, the spatial correlation coefficient of Z-score between an individual-level component and a template component is calculated using Eq (5). The template component can be the group-level component (i.e., TC-GICA template matching) or the corresponding individual-level DR component (i.e., DR template matching). Secondly, the component with the highest matching score is then selected as the best-matched individual-level component. The final matched individual-level components from TC-GICA template matching and DR template matching could be the same or different.

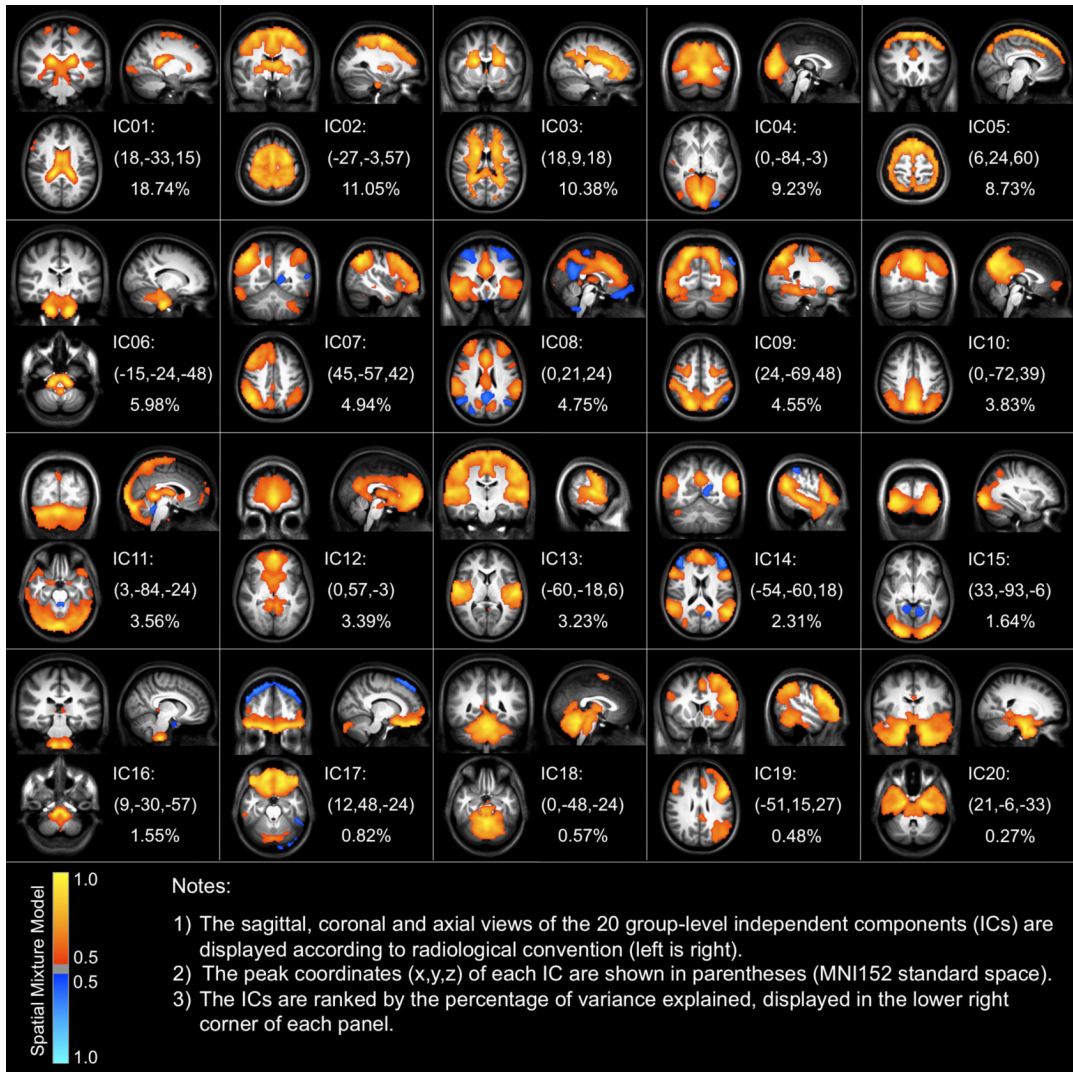


Figure 2. Group-level Components Revealed by the Temporal Concatenation Group Independent Component Analysis

The sagittal, coronal and axial views of the 20 group-level independent components (ICs) identified using TC-GICA are displayed according to radiological convention (left is right). The Z-statistic map relating to each IC is thresholded by a spatial mixture model ($p > 0.5$) and a cluster size ($> 80 \text{ mm}^3$). The peak coordinates (x, y, z) of each IC are shown in parentheses (MNI152 standard space).

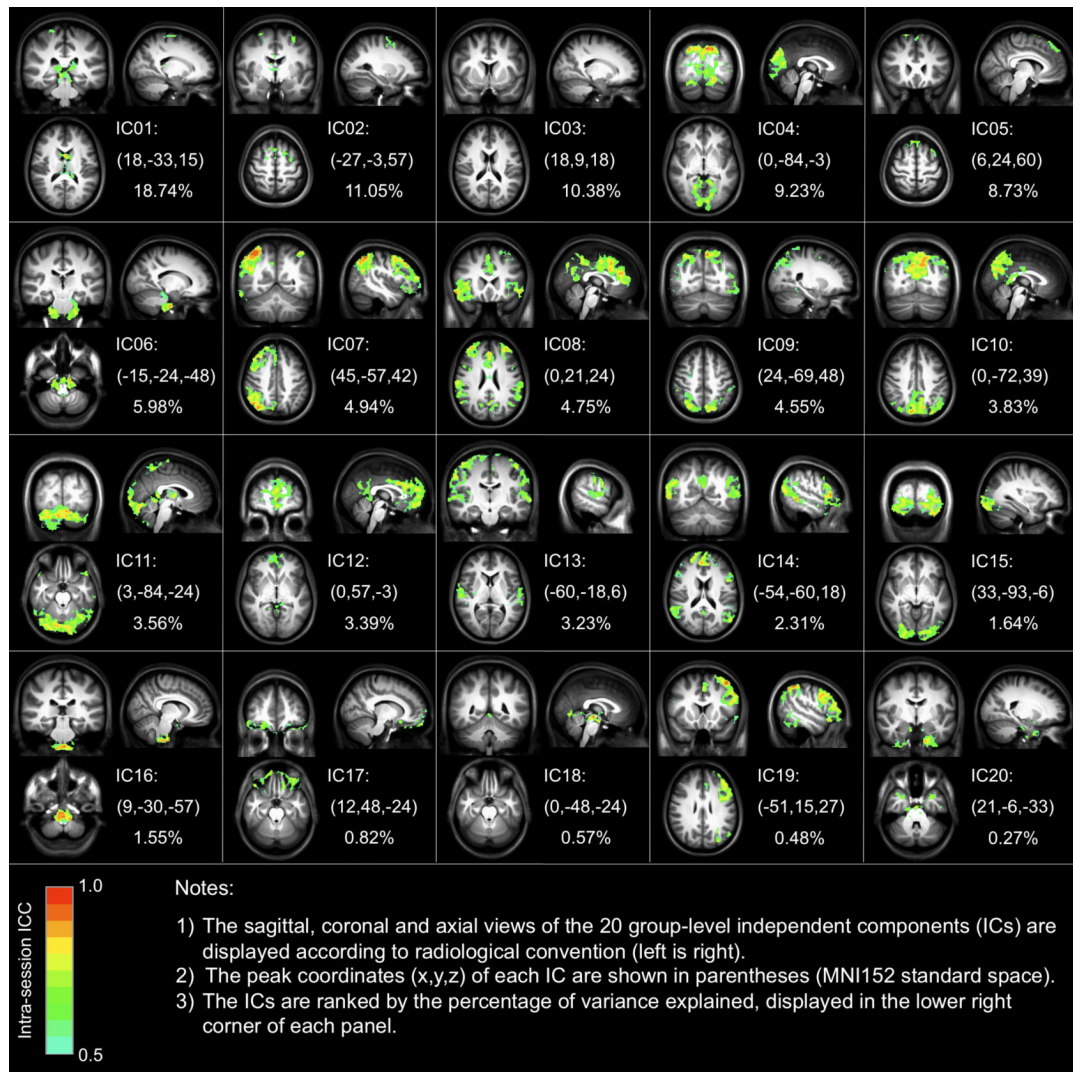


Figure 3. Short-term Intra-session Test-Retest Reliability Maps for Group-level Components

This figure depicts the voxel-wise short-term or intra-session (i.e., < 45 min) reliability of ICA-derived resting state functional connectivity. The sagittal, coronal and axial views of the reliability maps for each group-level component are displayed in radiological convention. The ICC map for each IC is thresholded at ICC > 0.5, with a minimum cluster size of 80 mm³. The peak coordinates (x, y, z) of each IC are shown in parentheses (MNI152 standard space).

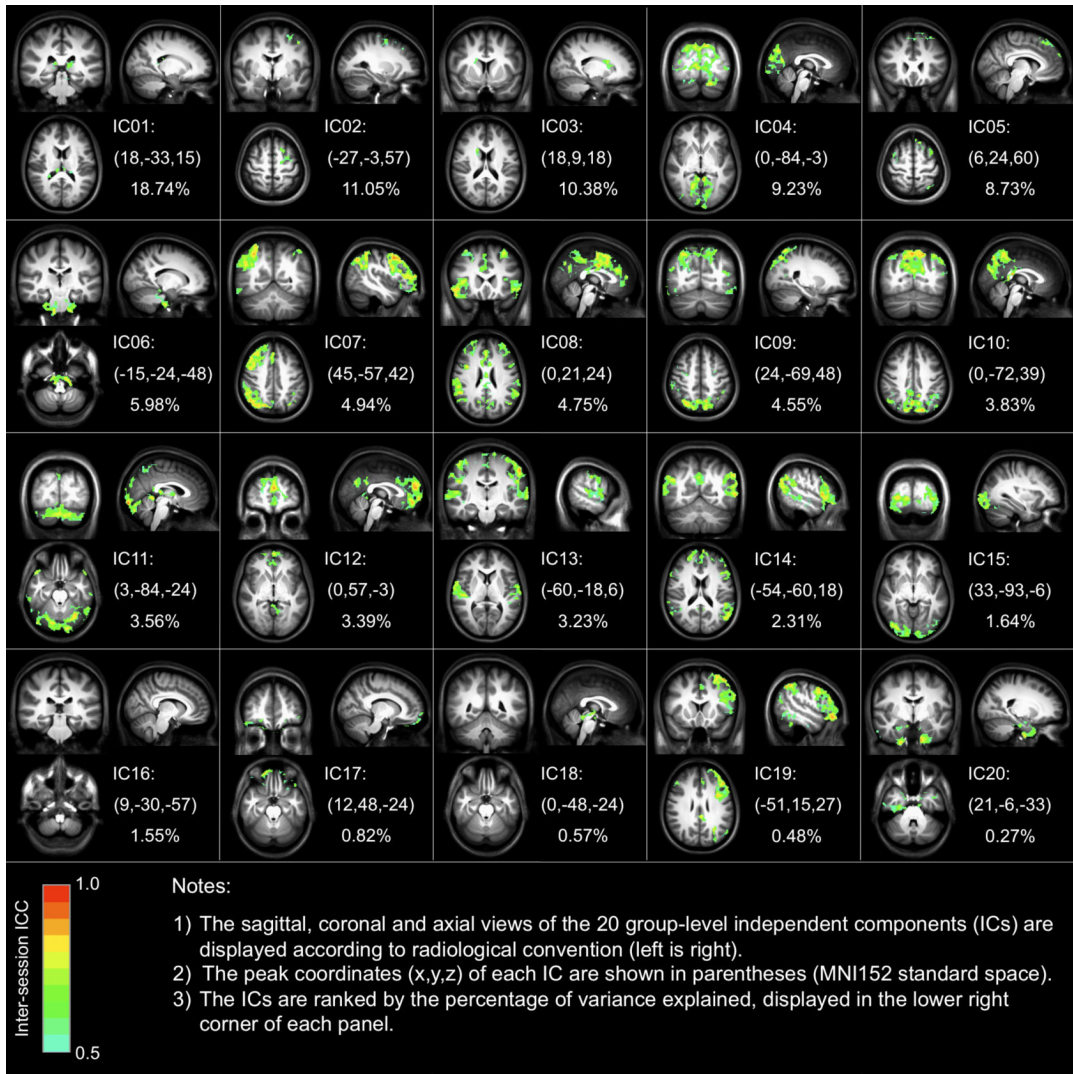


Figure 4. Long-term Inter-session Test-Retest Reliability Maps for Group-level Components
This figure depicts the voxel-wise long-term or inter-session (i.e., > 5 months) reliability of ICA-derived resting state functional connectivity. The sagittal, coronal and axial views of the reliability maps for each group-level component are displayed in radiological convention. The ICC map for each IC is thresholded at ICC > 0.5, with a minimum cluster size of 80 mm³. The peak coordinates (x, y, z) of each IC are shown in parentheses (MNI152 standard space). The ICs are ranked by the percentage of variances explained; displayed in the lower right corner of each panel.

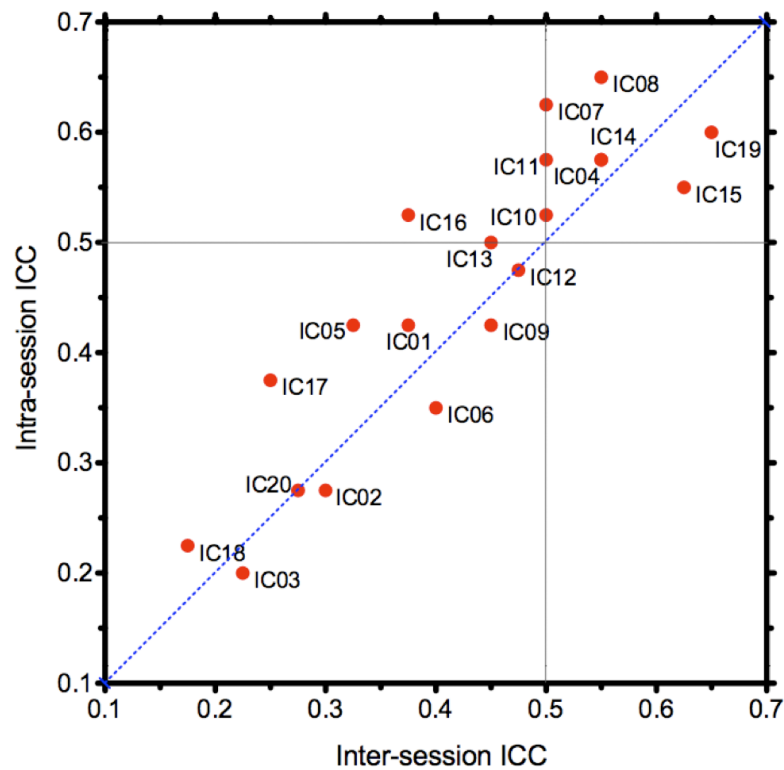


Figure 5. The Most Frequent Values in the ICC Histogram for Each Group-level Component

This scatter plot depicts the most frequent values appearing in both inter-session (x -axis) and intra-session (y -axis) ICC histograms for each group-level independent component (red dots). The histogram consists of 40 bins extending from 0 to 1. Two thin black lines are drawn to display the critical value ($ICC = 0.5$) for both intra-session and inter-session test-retest reliability, and the blue dash line characterizes the positions with equal intra- and inter-session reliability.

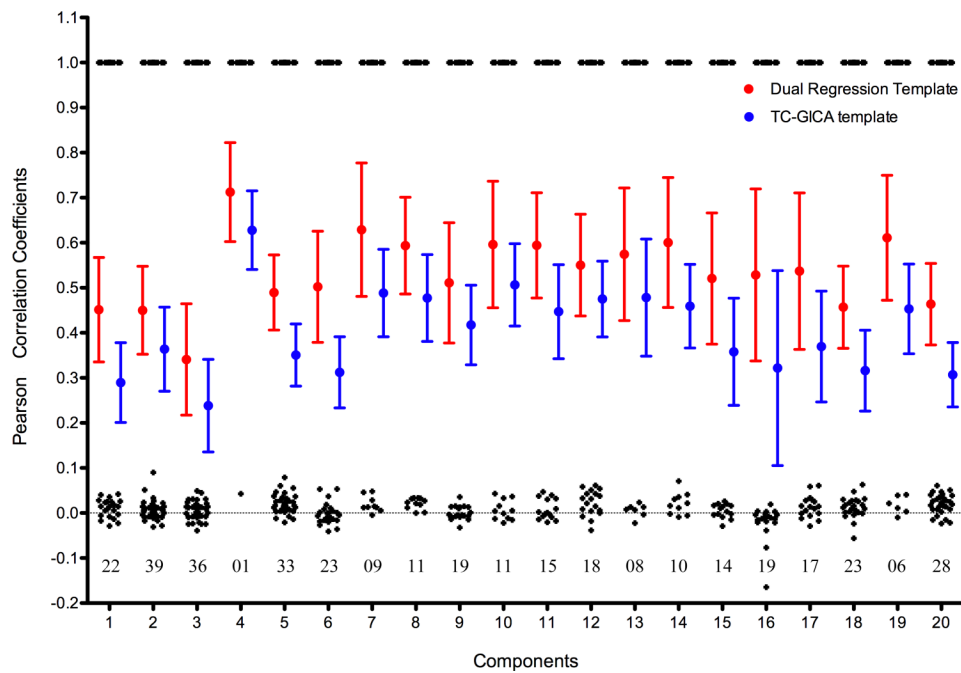


Figure 6. Individual Best Matching Scores As a Reproducibility Measure

This figure depicts the individual reproducibility measurements for each group-level component (IC). All individual best matching scores are depicted as an error bar graph for each group-level component. Two template strategies were adopted for calculating the individual best matching scores: individual dual-regression derived templates (red) and group-level temporal concatenation group ICA (TC-GICA) derived templates (blue). Meanwhile, for each group-level IC, the consistency of matched results between the two template strategies is calculated as the spatial correlation coefficient (black) between individual dual-regression matched results and individual TC-GICA matched results. The numbers below these black points (i.e., around the line $y = 0$) represent the number of disagreements between the two template matching strategies.

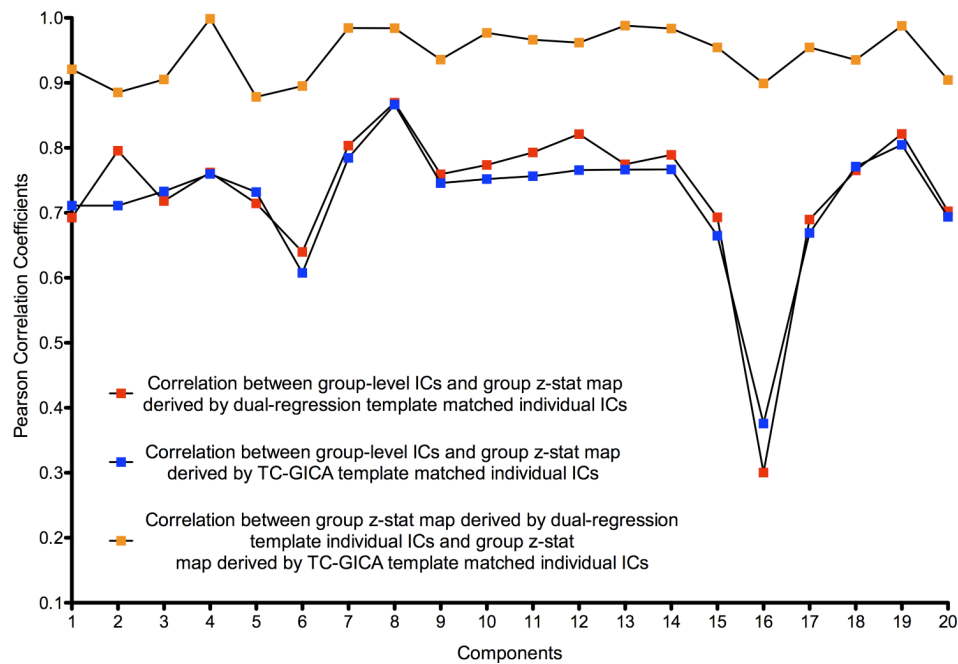


Figure 7. Consistency of Group-level Maps Based on Template Matching Results

This figure demonstrates the consistency of template matching results from two different template strategies at the group-level: (1) the spatial correlation between a TC-GICA derived group-level component and a group z-statistic map derived by a random effect group analysis on individual dual-regression template-matched components (red squares), and (2) the spatial correlation between a group-level component and a group z-statistic map derived by a random effect group analysis on individual TC-GICA template-matched components (blue squares). Also shown is the spatial correlation between the two group z-statistic maps based on the individual components using the two template strategies (orange squares).

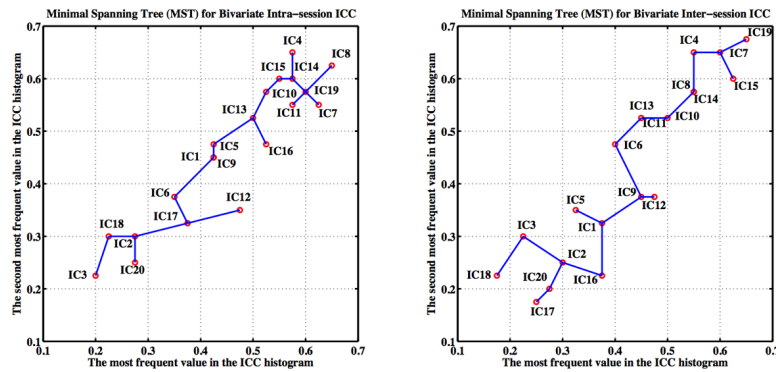


Figure 8. Ranking and Ordering the Test-Retest Reliability Maps Using Their Bivariate Minimal Spanning Tree Representations

For both intra-session (left) and inter-session (right) reliability, based on the bivariate feature (i.e., red circles) of each ICC histogram with 80 bins, a fully connected graph can be built. A minimal spanning tree (MST) of this graph is a tree sub-graph that contains all nodes and whose edges have the least total distance. The components are ranked by walking along the MST of the corresponding graph constrained with a rule. The rule defines a starting node at an endpoint of a tree diameter and then to proceed through the tree in such a way as to visit any shallow sub-trees at a given node before proceeding to the deeper sub-trees. In our situation, we choose to begin on the right (i.e., large test-retest reliability) of the MST and use the ordering rule to get the final ranks for these group-level components.

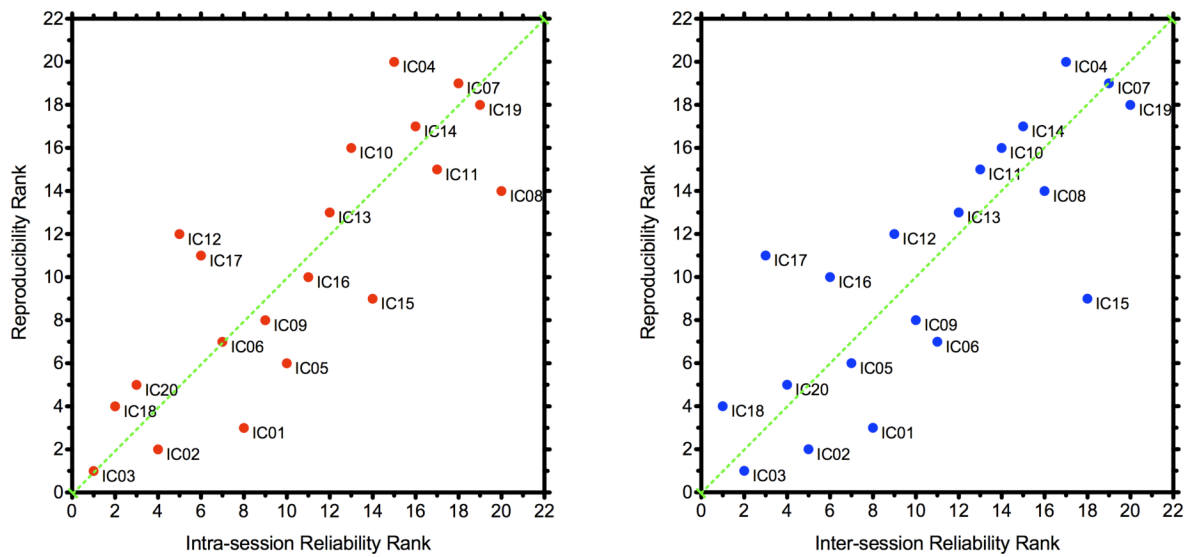


Figure 9. Test-Retest Reliability Ranks and Reproducibility Ranks of Each Group-level Component

These two scatter plots plot the reliability and reproducibility ranks for each group-level component. Left panel: intra-session reliability ranks versus reproducibility ranks; Right panel: inter-session reliability ranks versus reproducibility ranks.

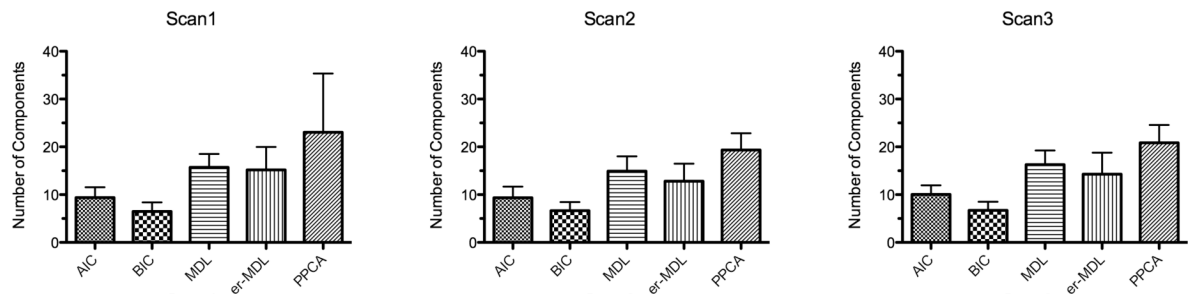


Figure 10. Number of Components Estimated by Five Automatic Estimation Algorithms across Three Scans

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MDL: Minimum Description Length; er-MDL: Entropy Rate enhanced Minimum Description Length; PPCA: Probabilistic Principal Component Analysis.