Developmental Cognitive Neuroscience 15 (2015) 83–93

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: http://www.elsevier.com/locate/dcn

Short-term test–retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder

Krishna Somandepalli[a], Clare Kelly[a,1], Philip T. Reiss[b,c], Xi-Nian Zuo[d], R.C. Craddock[e,f], Chao-Gan Yan[a,c,d], Eva Petkova[b,c], F.X. Castellanos[a,c], Michael P. Milham[e,f], Adriana Di Martino[a,*]

[a] Phyllis Green and Randolph Cowen Institute for Pediatric Neuroscience, The Child Study Center at NYU Langone Medical Center, 1 Park Avenue, New York, NY 10016, USA
[b] Division of Biostatistics, The Child Study Center at NYU Langone Medical Center, 1 Park Avenue, New York, NY 10016, USA
[c] Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA
[d] Key Laboratory of Behavioral Science and Magnetic Resonance Imaging Research Center, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[e] Center for the Developing Brain, Child Mind Institute, 445 Park Avenue, New York, NY 10022, USA
[f] Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA

ARTICLE INFO

Article history:
Received 2 March 2015
Received in revised form 7 August 2015
Accepted 9 August 2015
Available online 11 August 2015

Keywords:
Test–retest reliability
Intraclass correlation coefficient
ADHD
Resting state fMRI
Image intraclass correlation coefficient (I2C2)

ABSTRACT

To date, only one study has examined test–retest reliability of resting state fMRI (R-fMRI) in children, none in clinical developing groups. Here, we assessed short-term test–retest reliability in a sample of 46 children (11–17.9 years) with attention-deficit/hyperactivity disorder (ADHD) and 57 typically developing children (TDC). Our primary test–retest reliability measure was the intraclass correlation coefficient (ICC), quantified for a range of R-fMRI metrics. We aimed to (1) survey reliability within and across diagnostic groups, and (2) compare voxel-wise ICC between groups. We found moderate-to-high ICC across all children and within groups, with higher-order functional networks showing greater ICC. Nearly all R-fMRI metrics exhibited significantly higher ICC in TDC than in children with ADHD for one or more regions. In particular, posterior cingulate and ventral precuneus exhibited group differences in ICC across multiple measures. In the context of overall moderate-to-high test–retest reliability in children, regional differences in ICC related to diagnostic groups likely reflect the underlying pathophysiology for ADHD. Our currently limited understanding of the factors contributing to inter- and intra-subject variability in ADHD underscores the need for large initiatives aimed at examining their impact on test–retest reliability in both clinical and developing populations.

1. Introduction

The identification of biomarkers for neurodevelopmental disorders, a high priority for functional connectomics (Castellanos et al., 2013; Di Martino et al., 2014), depends on the development of measures that yield consistent results when repeated over time, i.e., their test–retest reliability must be adequate. A growing literature has worked to establish the test–retest reliability of common resting state functional magnetic resonance imaging (R-fMRI) measures (for review, see Zuo and Xing, 2014). Initial results have been encouraging, showing moderate-to-high short- and long-term test–retest reliability for an array of R-fMRI metrics, including: seed-based functional connectivity (e.g., Shehzad et al., 2009), amplitude of low-frequency fluctuations (ALFF; e.g., Zuo et al., 2010a), independent component analysis (ICA) based-indices (e.g., Thomason et al., 2011; Zuo et al., 2010b), regional homogeneity (ReHo) (Zuo et al., 2013) and voxel-mirrored homotopic connectivity (VMHC; Zuo et al., 2010c). These studies focused almost exclusively on neurotypical adults; only one study specifically examined test–retest reliability in children (Thomason et al., 2011). That study demonstrated high consistency of connectivity networks identified using ICA in typically developing children

* Corresponding author.
E-mail address: dimara01@nyumc.org (A. Di Martino).
[1] Present address: Trinity College Institute of Neuroscience, Trinity College, Dublin, Ireland.

http://dx.doi.org/10.1016/j.dcn.2015.08.003

(TDC). However, questions remain about the generalizability of these findings for a broader array of commonly examined R-fMRI metrics and for children with clinical conditions.

Here, we systematically quantified test–retest reliability of a range of R-fMRI metrics in clinical and nonclinical developing participants by leveraging a convenience sample of children with and without Attention-Deficit/Hyperactivity Disorder (ADHD) who completed two scans in the same session (~25 min apart). We focused on R-fMRI measures previously shown to be sensitive to brain development and increasingly investigated in neuropsychiatric disorders (e.g., Di Martino et al., 2014; Collin and van den Heuvel, 2013; Craddock et al., 2013; Dennis and Thompson, 2014; Fox and Greicius, 2010; Hagmann et al., 2012; Uddin et al., 2010). Specifically, we examined VMHC (which characterizes interhemispheric interactions; Zuo et al., 2010c); ReHo (local connectivity; Zang et al., 2004), ALFF (regional variability of the BOLD signal; Zang et al., 2007) and its normalized variant (fALFF; Zou et al., 2008). Finally, based on consistent findings of altered default network integrity in neurodevelopmental disorders, including ADHD (Castellanos et al., 2008; Posner et al., 2014), we examined posterior cingulate cortex (PCC) functional connectivity using seed based correlations.

As summarized in Table 1, our primary aim was to quantify the test–retest reliability of a range of R-fMRI metrics in school-age children with and without ADHD. Several measures of test–retest reliability are available and have been used for R-fMRI (e.g., Thomason et al., 2011; Shehzad et al., 2009; Shou et al., 2013; Zuo et al., 2010a). In recognition of evidence of regional variation in test–retest reliability of R-fMRI metrics (see review, Zuo and Xing, 2014) the regional effects of ADHD on intrinsic brain organization (Posner et al., 2014), and of its widespread use, we selected voxelwise intraclass correlation coefficient (ICC; Shrout and Fleiss, 1979) as our primary test–retest reliability measure. In addition to ICC, we also surveyed two other test–retest performance measures as they have been used in other imaging studies (e.g., Shehzad et al., 2009; Thomason et al., 2011; Zuo et al., 2010a), and provide complementary information to ICC. These include Kendall's Coefficient of Concordance (KCC; Kendall and Smith, 1939) – and image intraclass correlation coefficient (I2C2; Shou et al., 2013). KCC is the non-parametric counterpart of ICC assessing voxel-wise consistency between scans; I2C2 is a global measure of reliability that generalizes ICC to volumetric imaging data.

Secondarily, we aimed to directly compare voxel-wise ICC between ADHD and TDC. To ensure that any differences in reliability observed between the two groups could be attributed to diagnostic status, as opposed to commonly observed differences in scanner head-motion, we ensured that the two groups were matched on head motion (e.g., mean frame-wise displacement; Jenkinson et al., 2002).

2. Materials and methods

2.1. Sample

We selected 103 children (aged 12.1 ± 3.1 years) from a larger sample of 179 children (97 TDC and 82 ADHD) aged between 8 and 18 years old who completed two resting-state scans. Selection was based on meeting our criteria for imaging quality control (see Supplementary text) leading to a sample of 57 TDC and 46 children with ADHD for our analyses. The selected and excluded subjects did not differ significantly in severity of ADHD symptoms, age, nor IQ (see Supplementary text). Data from 44 TDC and 12 children with ADHD were included in one or more previous reports (Koyama et al., 2011; Yang et al., 2014; Chabernaud et al., 2012; Di Martino et al., 2013).

Presence or absence of ADHD and Axis-I psychiatric comorbidity (according to the Diagnostic and Statistical Manual of Mental Disorders, 4th ed., text rev.; DSM-IV-TR; American Psychiatric Association, 2000) was reached by trained clinicians based on parent and child interviews using the Schedule for Affective Disorders and Schizophrenia for School-age Children – Present and Lifetime Version (KSADS-PL; Kaufman et al., 1997), information from prior available records, and direct observation during testing. Children with ADHD were included regardless of their psychiatric comorbidity except for psychotic disorders, major depression, bipolar and conduct disorder. Inclusion as TDC required absence of any DSM-IV-TR Axis I diagnosis based on the same assessment protocol and no history of treatment with psychoactive medications. Absence of known neurological or genetic syndromes was required for all participants. Handedness was evaluated with the Edinburgh Handedness Inventory (Oldfield, 1971) and parents provided information about ethnicity/race and socioeconomic status. The groups did not differ significantly in ethnicity/race, age, sex, socioeconomic status, or handedness (see Table 2).

Of the 46 children with ADHD, 22 met diagnostic criteria for combined type (ADHD-C), 21 for ADHD predominantly inattentive type (ADHD-I) and 3 for ADHD-not otherwise specified (ADHD-NOS). Psychiatric comorbidity with other DSM-IV-TR Axis-I disorders was present in 8 (17%) of the children with ADHD (see Table 2). Clinicians also recorded history of psychotropic medication use per parent report. Thirty-three of the children with ADHD were naïve to psychoactive medications; of the remaining 13 children with ADHD, one was off stimulants for more than one year prior to the scan. The remaining 12 children were currently being treated with stimulants. All withdrew stimulant ~24 h prior to the scans, except for one child who was last given an immediate release stimulant four hours prior to the scan. The study procedures were approved by the New York University (NYU) and the NYU School of Medicine Institutional Review Boards, and all parents and children provided written informed consent/assent.

2.2. Data acquisition

As detailed elsewhere (e.g., Koyama et al., 2011), imaging data were acquired using a Siemens Allegra 3T at the NYU Center for Brain Imaging. A T1-weighted image (MPRAGE, TR = 2530 ms; TE = 3.25 ms; TI = 1100 ms; flip angle = 7°; 128 slices; FOV = 256 mm; voxel-size = 1 mm × 1.3 mm × 1.3 mm) and two 6-min resting state scans (multi-echo echo planar imaging (EPI) sequence reconstructed as conventional single-shot EPI; 180 time points; TR = 2000 ms; effective TE = 33 ms; flip angle = 90°; 33 slices; voxel-size = 3 mm × 3 mm × 4 mm) were acquired during the same scan session (<1 h apart). To improve functional-to-anatomical co-registration and to 'unwarp' geometrical distortions created by magnetic field inhomogeneities (Jezzard, 2012), we acquired a calibration scan including a field map prior to the first EPI scan. Because we selected data collected across ongoing studies, acquisition conditions varied slightly. Specifically, in most cases both scans were collected while participants kept their eyes open, but in a few, children were instructed to close their eyes for one of the two scans (see Table 2). As described below, eye status was accounted for in our analyses. Additionally, while all children completed a DTI scan immediately before the second rest scan, for some children this was preceded by a task fMRI sequence (see Table 2); groups did not differ significantly in MRI protocols. Finally, the time interval between Scan 1 and Scan 2 varied across subjects; slight but significant diagnostic group differences were accounted for in group analyses, as described below.

Table 1
Summary of Analyses and Results.

Objective	Methods	Results
1. *Survey short-term (i.e., intra-session) test–retest reliability* in children across and within diagnostic groups for the following R-fMRI metrics: f/ALFF, ReHo, VMHC, PCC-iFC.	1. *Primary measure*: Intraclass Correlation Coefficient (ICC) using Linear Mixed Effects model. 2. *Secondary complementary measures*: 1. Image ICC (I2C2) 2. Kendall's Coefficient of Concordance (KCC)	1. Moderate (ICC > 0.4) to high (ICC > 0.7) across all R-fMRI metrics. 2. Large-scale regional variations existed across all R-fMRI metrics with higher- order cortical networks showing higher reliability. 3. Moderate to high I2C2 (>0.4) for f/ALFF and ReHo. Low to moderate I2C2 (0.4 > I2C2 > 0.2) for VMHC and PCC-iFC. 4. High KCC across all R-fMRI metrics.
2. *Compare voxel-wise test–retest reliability between children with ADHD and TDC* for each of the R-fMRI examined (i.e., f/ALFF, ReHo, VMHC, PCC-iFC)	1. Modified Fisher *Z*-transform for voxel-wise ICC	1. Significant ICC differences (TDC > ADHD) for all measures except VMHC, albeit varying in spatial extent and magnitude.

2.3. Data preprocessing

Data preprocessing was carried out using the Configurable Pipeline for the Analysis of Connectomes (C-PAC version 0.3.4, http://fcp-indi.github.com), an open-source, Nipype-based, automated processing pipeline with an efficient interface to software packages such as FSL, AFNI and ANTs. C-PAC was customized to include a workflow for performing field map correction on EPI data. Preprocessing for individual R-fMRI data consisted of (1) slice timing correction for interleaved acquisitions, (2) 3D motion correction (realignment using 3 translational and 3 rotational parameters), (3) unwarping geometrical distortions using the B_0 field map created from the calibration scan using FUGUE (Jenkinson, 2003), (4) mean-based intensity normalization, (5) linear and quadratic de-trending, (6) nuisance regression (see below), (7) temporal band-pass filtering (0.01–0.1 Hz, except for ALFF and its fractional variant, fALFF),

Table 2
Characteristics of the Sample.

	TDC ($N = 57$)		ADHD ($N = 46$)		Group differences		
	Mean	SD	Mean	SD	*t*-Statistic[a]	df	*p*
Age	12.5	3.0	11.4	3.1	1.81	94	0.073
Full IQ[b]	112	14	107	15	1.96	90	0.053
Verbal IQ	112	13	107	14	2.27	87	0.045
Performance IQ	109	13	105	15	1.73	89	0.088
Handedness score	0.6	0.3	0.6	0.3	0.01	97	0.988
CPRS[c] DSM-IV Total T score	45.0	5.38	70.3	9.4	16.03	67	<0.0001
	N (%)		*N* (%)		X^2	df	
Male	29 (51)		31 (67)		2.22	1	0.137
SES[d] (Class 4 or 5)	39 (68)		35 (76)		0.41	1	0.522
Race							
Caucasian	27 (47)		23 (50)		0.16	2	0.923
African-American	17 (30)		14 (30)				
Other	13 (23)		9 (20)				
Medication treatment							
Medication Naïve	–		33		–		
Not naïve but off medication(s)	–		1		–		
Current stimulant treatment	–		12		–		
Comorbidity							
DBD-NOS + GAD	–		2		–		
ODD[e]	–		3		–		
Speech language impairment[f]	–		2		–		
Encopresis	–		1		–		
	Mean	SD	Mean	SD	*t*-Statistic[a]	df	*p*
Time between scans (min)	22.9	2.6	24.8	5.4	2.18	61	0.033
Mean FD (Jenkinson et al., 2002)[g]	0.05	0.02	0.06	0.02	1.89	67	0.064
	N (%)		*N* (%)		X^2	df	
EFT between rest scans[h]	20 (35)		11 (21)		1.03	1	0.311
Exited scanner between rest scans	4 (7)		9 (19)		2.58	1	0.108
Eyes open for both rest scans	33 (58)		24 (52)		0.15	1	0.703

[a] Unpaired Welch corrected *t*-test.
[b] Intelligence was estimated with the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) for all children but one who was evaluated with the Kaufman Brief Intelligence Test (KBIT; Kaufman and Kaufman, 2004).
[c] Conners Parent Rating Scale (CPRS; Conners, 1997).
[d] SES measures with the Hollingshead Four-Factor Index of Socioeconomic Status (Hollingshead, 1975).
[e] Two children with disruptive behavior not otherwise specified (DBD-NOS) and generalized anxiety disorder (GAD), two of the three children with oppositional defiant disorder (ODD) also had additional comorbidity, one with anxiety-NOS and the other with enuresis.
[f] One of these children also presented with dyslexia and dysgraphia.
[g] Mean FD averaged across Scans 1 and Scan 2 per Jenkinson et al. (2002).
[h] EFT: Eriksen Flanker Task.

(8) registration (see below), and (9) spatial smoothing using a 6 mm Gaussian kernel at full-width half maximum.

2.3.1. Nuisance regression

Consistent with prior studies (Satterthwaite et al., 2012; Yan et al., 2013a), to control for the effects of head motion and to reduce the influence of signals of no interest, we regressed the preprocessed data on 24 parameters (Friston et al., 1996) obtained from the motion correction procedure (6 head motion parameters, their values from one time point before, and the squared values of these 12 items), and on the mean time courses obtained from white matter (WM) and cerebrospinal fluid (CSF), using subject-specific masks with tissue type probability threshold of 0.6 for both WM and CSF. To account for any residual additive noise (Yan et al., 2013b), we included individual subject means for the given R-fMRI metric as a nuisance covariate. This step was performed at the group-level (see below).

For comparison with other common nuisance regression approaches, secondary analyses used two other subject-level nuisance signal regression strategies: (1) CompCor: five principal components derived from WM and CSF were included in the model (Behzadi et al., 2007) and (2) global signal regression (GSR): signals from WM, CSF, and a whole-brain mask were regressed. Both of these alternative regression approaches also included the 24 Friston motion parameters.

2.3.2. Registration

For each participant, we performed a pairwise registration of their two EPI scans, to calculate a "midway" point between them (Reuter et al., 2012). Each EPI was then registered to this "midway" functional volume, and linear registration of the "midway" volume to the subject's high-resolution structural image was performed using FSL FLIRT with boundary-based registration (Greve and Fischl, 2009; Jenkinson et al., 2002; Jenkinson and Smith, 2001). This functional-to-anatomical co-registration was improved by intermediate registration to a low-resolution image obtained from the calibration scan, followed by B_0 unwarping (see Supplementary Material and Supplementary Figure 1). The resulting images were then transformed into MNI152 (Montreal Neurological Institute) space at 2 mm^3 resolution using ANTs (Avants et al., 2010). For VMHC analyses, functional data in MNI space were registered to a symmetric template obtained as in Zuo et al. (2010c).

2.4. R-fMRI metrics

Amplitude of low frequency fluctuations (ALFF; Zang et al., 2007) and *Fractional ALFF* (fALFF; Zou et al., 2008): ALFF is the standard deviation of a band-pass filtered (here, 0.01–0.1 Hz) voxel time series and fALFF is the ratio of ALFF to the standard deviation of the unfiltered time series.

Regional homogeneity (ReHo) is a measure of local coherence that is calculated from the Kendall's Coefficient of Concordance (Zang et al., 2004) between the time series of a given voxel and those of its nearest neighbors (here 26).

Voxel-mirrored homotopic connectivity (VMHC) is the Pearson correlation between the time series of each voxel and its symmetrical inter-hemispheric counterpart (Anderson et al., 2011; Zuo et al., 2010c).

PCC intrinsic functional connectivity (iFC): We extracted the average time series for the posterior cingulate cortex (PCC: −8, −56, 26; 8 mm diameter sphere) (Andrews-Hanna et al., 2010) and then calculated the Pearson correlation coefficient between this PCC time course and that of every other brain voxel.

2.5. Intraclass correlation coefficient (ICC) within and between groups

For each R-fMRI metric we computed ICC (Shrout and Fleiss, 1979) across all 103 children, as well as within ADHD and TDC groups, separately. Specifically, we used an "ICC-type variance ratio" defined by Shrout and Fleiss (pg. 423; 1979) used in previous studies (e.g., Shehzad et al., 2009; Zuo et al., 2010a). We then tested for group differences in voxel-wise ICC between ADHD and TDC. Consistent with prior work we focused on ICC > 0.4, which is considered to reflect the moderate to high range of test–retest reliability (Landis and Koch, 1977; Zuo et al., 2013).

2.5.1. Within-group ICC

To quantify inter- and intra-individual variability, we employed a linear mixed effects (LME) model (Hoffman and State, 2010; Laird and Ware, 1982; State, 2010). To calculate ICC at each voxel for a given continuous fMRI measure λ, consider a sample of n subjects with k repeated measurements each. Letting λ_{ij} denote the voxelwise metric for the ith participant measured at jth time (for $i = 1,\ldots n; j = 1,\ldots k$), we use the following two-level LME to decompose λ_{ij} at each voxel:

$$\lambda_{ij} = \lambda_i + \beta_1 \text{mean FD}_{ij} + \beta_2 \text{eye status}_{ij} + \beta_3 \text{subject mean}_{ij} + e_{ij}, \text{ with}$$
$$\lambda_i = \mu + p_i + \beta_3 \text{age}_i + \beta_4 \text{sex}_i + \beta_5 \Delta t_i \quad (1)$$

where μ is the group average, p_i is the random effect of the ith participant and e_{ij} is an error term; these are independent and normally distributed with mean 0 and variances σ_p^2 and σ_e^2 that are to be estimated. β_1–β_5 denote the effects of covariates: both intra-individual variables, namely mean frame-wise displacement (mean FD; Jenkinson et al., 2002), eye-status (i.e., eyes closed vs. open at scan), individual subject means for each R-fMRI measure (subject mean), and the inter-individual variables age, sex, and time interval between Scan 1 and Scan 2 (Δt). The ICC of λ is computed, at each voxel, as:

$$\text{ICC}(\lambda) = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2}, \quad (2)$$

where the variance component estimates $\hat{\sigma}_p^2$ and $\hat{\sigma}_e^2$ are derived by restricted maximum likelihood (ReML) as in Zuo et al. (2013).

2.5.2. Between-group ICC differences

To assess whether reliability differs significantly between ADHD and TDC, we used a test procedure based on Fisher's z-transformation proposed by Konishi and Gupta (1987; page 231; formula 3.7). The voxelwise test statistic to compare the two groups is given by

$$Z_{\text{TDC-ADHD}} = \frac{Z(\text{ICC}_{\text{TDC}}) - Z(\text{ICC}_{\text{ADHD}})}{\sqrt{1/[N_{\text{TDC}} - d - 2] + 1/[N_{\text{ADHD}} - d - 2]}}$$
$$= \frac{1/2\left[\log[(1+\rho_{\text{TDC}})/(1-\rho_{\text{TDC}})] - \log[(1+\rho_{\text{ADHD}})/(1-\rho_{\text{ADHD}})]\right]}{\sqrt{1/[N_{\text{TDC}} - d - 2] + 1/[N_{\text{ADHD}} - d - 2]}} \quad (3)$$

where ρ refers to the ICC for the group indicated by the subscript, N_{ADHD} and N_{TDC} are the two groups' sample sizes, and d corresponds to the number of covariates (here $d = 6$) in the LME model. The resulting voxel-wise maps were whole-brain corrected using Gaussian random field theory, thresholded at $Z > 2.3$ voxel-wise and at $p < 0.05$ cluster-wise.[2]

[2] The approximate standard deviation (SD) of a Z-transformed Pearson correlation is given by $1/(\sqrt{N-3})$. Similarly, the SD for a partial correlation with d

Table 3
Peak voxel and center of gravity (COG) coordinates for between group ICC differences.

R-fMRI metric	# voxels	Cluster index	Peak *Z* coordinates			COG coordinates			Labels
			X	*Y*	*Z*	*X*	*Y*	*Z*	
ALFF	13,200	1	39	61	37	44.6	50.8	38.5	Paracingulate gyrus, cingulate gyrus (posterior division), R thalamus, R putamen, brain stem
fALFF	279	1	45	45	65	45.8	43.8	64.2	Postcentral gyrus, cingulate gyrus (posterior division)
ReHo	1383	1	70	42	56	48.5	41.3	52.4	Supramarginal gyrus, superior temporal gyrus
	6371	2	18	55	36	20.5	44.8	47.2	Pre- and post-central gyrus
PCC-iFC	250	1	53	37	52	45.8	35.1	48.7	Paracingulate gyrus, superior frontal gyrus
	599	2	44	85	46	46.0	86.2	47.3	Precuneus, cingulate gyrus (posterior division)

Note: Center of Gravity (COG) and peak coordinates obtained using FSL easythresh function in MNI152.2 mm^3 space; labels obtained using FSL function atlas query with Harvard-Oxford Cortical and Subcortical atlases.

Finally, for consistency with other studies (e.g., Shehzad et al., 2009; Thomason et al., 2011; Zuo et al., 2010a), we surveyed complementary measures of test-rest reliability including (1) image intraclass correlation coefficient (I2C2; Shou et al., 2013), a global measure of reliability that generalizes the classic ICC for volumetric imaging data; and (2) a non-parametric measure, Kendall's Coefficient of Concordance (KCC; Kendall and Smith, 1939) which assesses the voxel-wise consistency of a R-fMRI metric between scans. As above, we computed these metrics across all subjects, and within the TDC and ADHD groups separately (maps for each ADHD subtype – Inattentive (ADHD-I) and Combined (ADHD-C) – were similar to each other; data not shown).

3. Results

3.1. Within-groups ICC

As summarized in Table 1, both across all subjects and within the two diagnostic groups (i.e., ADHD and TDC), moderate-to-high voxelwise ICC were obtained for cortical and subcortical regions, albeit with varying magnitude across R-fMRI metrics and brain areas (see Fig. 1 for maps and distribution of ICC magnitudes). We observed high ICC (i.e., 0.7 and above) for ALFF and ReHo and moderate-to-high ICC (i.e., 0.4–0.8) for fALFF and VMHC. As observed by Zuo and Xing (2014), ReHo showed high ICC both with respect to magnitude and spatial extent, and ICC of ALFF was substantially greater than that of fALFF. As observed in healthy adults (Shehzad et al., 2009; Yan et al., 2013a), PCC-iFC ICC was highest in regions exhibiting significant correlations with PCC, but across the whole brain ranged from fair to moderate (i.e., 0.2–0.6). As shown in Supplementary Figure 2, these results were consistent across preprocessing approaches.

To summarize ICC in terms of large-scale functional cortical networks, we computed the Fisher's *Z*-transformed ICC within six of the seven cortical networks described by Yeo et al. (2010). Adequate coverage of ventral medial orbitofrontal cortex was not available for all participants due to susceptibility artifacts of these ventral regions; accordingly, the limbic network was excluded from this analysis. This step was conducted for all R-fMRI measures except PCC-iFC, which largely overlaps with the Default network.

We compared the distribution of voxel-wise ICC values among functional networks using the rank-based test of Kruskal (1952). The result was significant ($p < 0.0001$) for all R-fMRI measures, indicating that reliability differed appreciably among the six functional networks. Fig. 2 (in which higher rank means higher ICC) shows that the Default network exhibited the highest reliability overall (ranked first for fALFF, ReHo and VMHC; fourth for ALFF), while the Visual network was the least reliable across all measures. In general, networks supporting higher order functions (Default, Ventral and Dorsal Attention and Frontoparietal Control network) were more reliable than networks supporting perceptual and somatomotor function, for all measures except ALFF.

3.2. Other measures of reliability

Secondary analyses on other test–retest performance measures included voxel-wise consistency with KCC (Fig. 3). Global measures of reliability of the brain space indexed by I2C2 (Fig. 4) showed a pattern of results similar to ICC.

3.3. Between-group ICC differences

Significant differences in ICC between ADHD and TDC groups were observed for all R-fMRI metrics except VMHC (Fig. 5). No R-fMRI metric showed regions with significantly greater ICC for ADHD than TDC, while we observed focal clusters with greater ICC in TDC compared to ADHD. Although the spatial location of group differences varied across measures, they appear most consistently in aspects of ventral precuneus and PCC (see Fig. 5 and Supplementary Figure 3). Peak coordinates of clusters differing between groups are shown in Table 3. As shown in Supplementary Figure 3, these results were consistent across preprocessing approaches except for ALFF and fALFF, suggesting that these measures are more susceptible to variation in preprocessing (Zuo et al., 2014).

To facilitate the interpretation of ICC diagnostic group differences, and illustrate the scale, range and distribution of these variance components across the whole brain, at each voxel, we extracted the variance components (i.e., between- and within-subject variance) used in the estimation of ICC for each diagnostic group separately. As shown in Supplementary Figure 4, both the range and scale of these components were greater in individuals with ADHD, for all measures and particularly so for ALFF.

4. Discussion

We quantified short-term test–retest reliability of a range of R-fMRI metrics in a convenience sample of 46 children with ADHD and 57 TDC matched for "micro" levels of head motion. We observed

covariates is given by $1/(\sqrt{N-d-3})$ as per Levy and Narula (1978); Testing hypotheses concerning partial correlations – some methods and discussion. Int. Stat. Rev. 46, 215–218. It follows that for ICC of two measurements, the approximate SD for *Z*-transformed ICC is given by $1/(\sqrt{N-2})$ (Fisher, 1958, Chapter 7; Konishi and Gupta, 1987). Thus heuristically, if ICC is derived from a linear mixed model with *d* fixed effects, it is reasonable to approximate the SD as $1/(\sqrt{N-d-2})$.

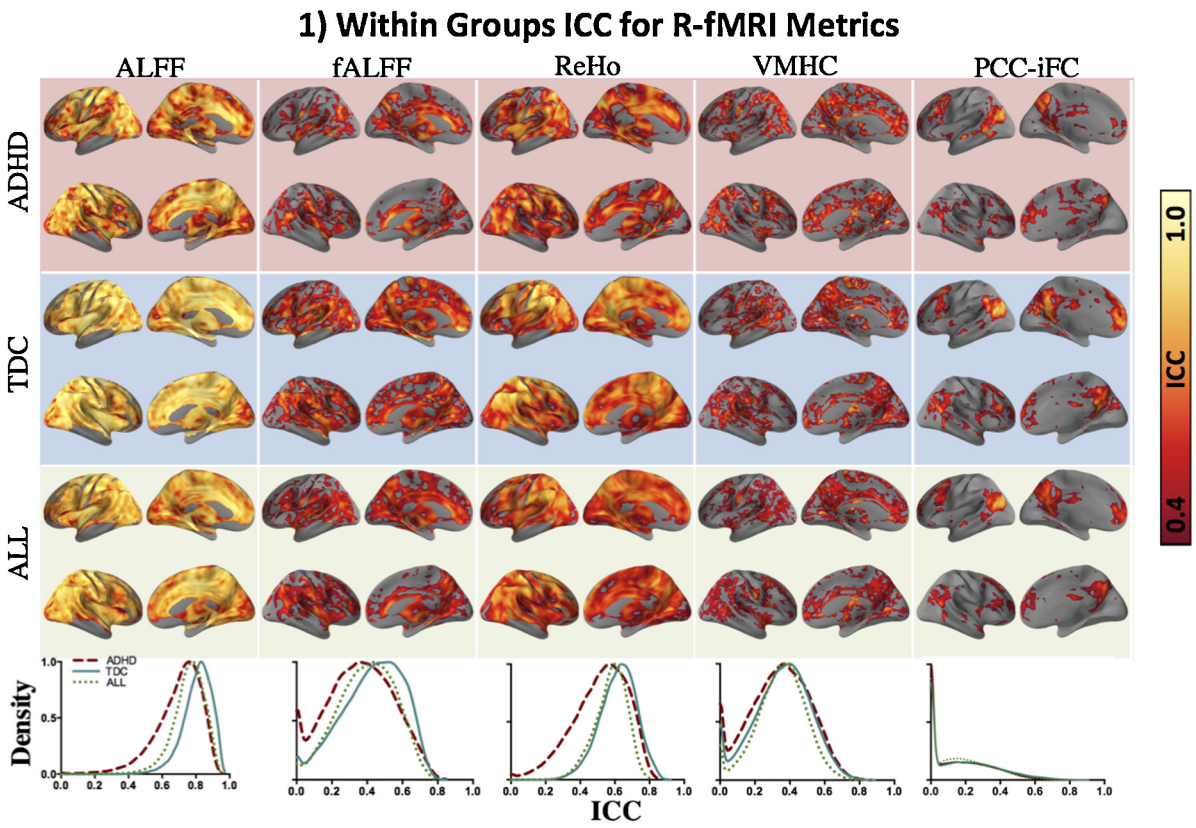

Fig. 1. Within groups ICC for R-fMRI metrics. Surface maps show spatial distribution of voxel-wise intraclass correlation coefficients (ICC) within the ADHD and TDC groups and across all subjects (ALL), from the top to the bottom row, respectively. The color bar indicates fair (ICC = 0.4; dark red) to perfect reliability (ICC = 1.0; white). The kernel density estimate plots in the lower-most row illustrate the distribution of ICC for ADHD, TDC and ALL. The peaks of the density curves indicate the most frequently observed ICC magnitude (*x*-axes) for a given R-fMRI metric. For all figures, ICC within gray matter only is reported (mask created using the MNI152 gray matter tissue prior included with FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL), thresholded at 25% tissue-type probability). As the density of ICC for PCC-iFC was computed for the whole gray matter, a large portion of area under the curve lies below ICC = 0.2, however the corresponding surface maps show ICC > 0.4 for regions known to be significantly correlated with PCC (i.e., default network). For VMHC, although there is only one ICC value for each pair of homotopic voxels, results are projected onto both hemispheres to minimize confusion regarding the laterality of the results. Surface maps were generated using the pysurfer package in python (http://pysurfer.github.io/index.html) and density plots were generated using Gaussian kernel density estimation, available through the scipy package in python (http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html). ADHD: attention deficit hyperactivity disorder; TDC: typically developing children; ALFF: amplitude of low frequency fluctuations; fALFF: fractional ALFF; ReHo; regional homogeneity; VMHC: voxel-mirrored homotopic connectivity; PCC-iFC: intrinsic functional connectivity (iFC) of the posterior cingulate cortex (PCC; seed coordinates: −8, −56, 26, 8 mm diameter sphere).

moderate-to-high ICC across diagnostic groups for the R-fMRI metrics ALFF, fALFF, ReHo, VMHC, and PCC-iFC. These findings, together with similar results from secondary analyses on other measures of test–retest performance (i.e., I2C2, KCC), suggest that the R-fMRI metrics we tested are generally reliable within the same session in children with or without ADHD. Nevertheless, test–retest reliability across R-fMRI measures exhibited regional variation. Higher-order cortical networks such as the Default network, Frontoparietal Control (Control) network, Dorsal Attention (DorsAttn) and Ventral Attention (VentAttn) networks showed the highest test–retest reliability (ICC), followed by the Sensory/motor network and Visual network. Regions also varied in terms of diagnostic differences in voxel-wise ICC. All R-fMRI metrics except VMHC exhibited significantly greater ICC in TDC than children with ADHD, though spatial extent and magnitude varied among R-fMRI metrics. Below, we discuss these findings in more detail.

4.1. Test–retest performance in children

Our findings of moderate-to-high test–retest performance in children were consistent with a prior report both at short- (within session) and long-term (2–3 years) intervals in ~15 TDC (Thomason et al., 2011) which found high consistency across six sensory and cognitive cortical ICA networks. Here, we extended the investigation of test–retest performance to a range of other R-fMRI metrics that are increasingly used to study intrinsic properties of brain architecture in children and adults (Anderson et al., 2011; Craddock et al., 2013; Di Martino et al., 2014; Zuo et al., 2010c). Across measures, the pattern of findings was consistent with Thomason et al.'s results in terms of both the magnitude of test–retest reliability observed and the spatial cortical network distribution.

Across R-fMRI metrics, our findings in children were also qualitatively comparable to results from previous studies conducted in adults over both short and long intervals (Shehzad et al., 2009; Zuo et al., 2010a,c, 2013). However, age has been found to affect test–retest reliability, at least in adulthood (Song et al., 2012). As such, a quantitative investigation of ICC across the life span for a range of R-fMRI metrics is necessary – surveying short-and long-term test–retest reliability across the life span will allow the age-related dependence of test-rest reliability to be quantified.

In addition to ICC we also surveyed two other test–retest performance measures as they have been used in other imaging studies (e.g., Thomason et al., 2011; Shehzad et al., 2009; Shou et al., 2013) and provide complementary information to ICC. They all yielded moderate-to-high test–retest reliability. We note, however, that each of these measures captures different aspects of test–retest performance. For both voxel-wise ICC and KCC, one can generate a spatial distribution of scores that can be informative or overwhelming, depending on the specific question being asked. In contrast I2C2 is a global summary measure. As such, I2C2 is less sensitive

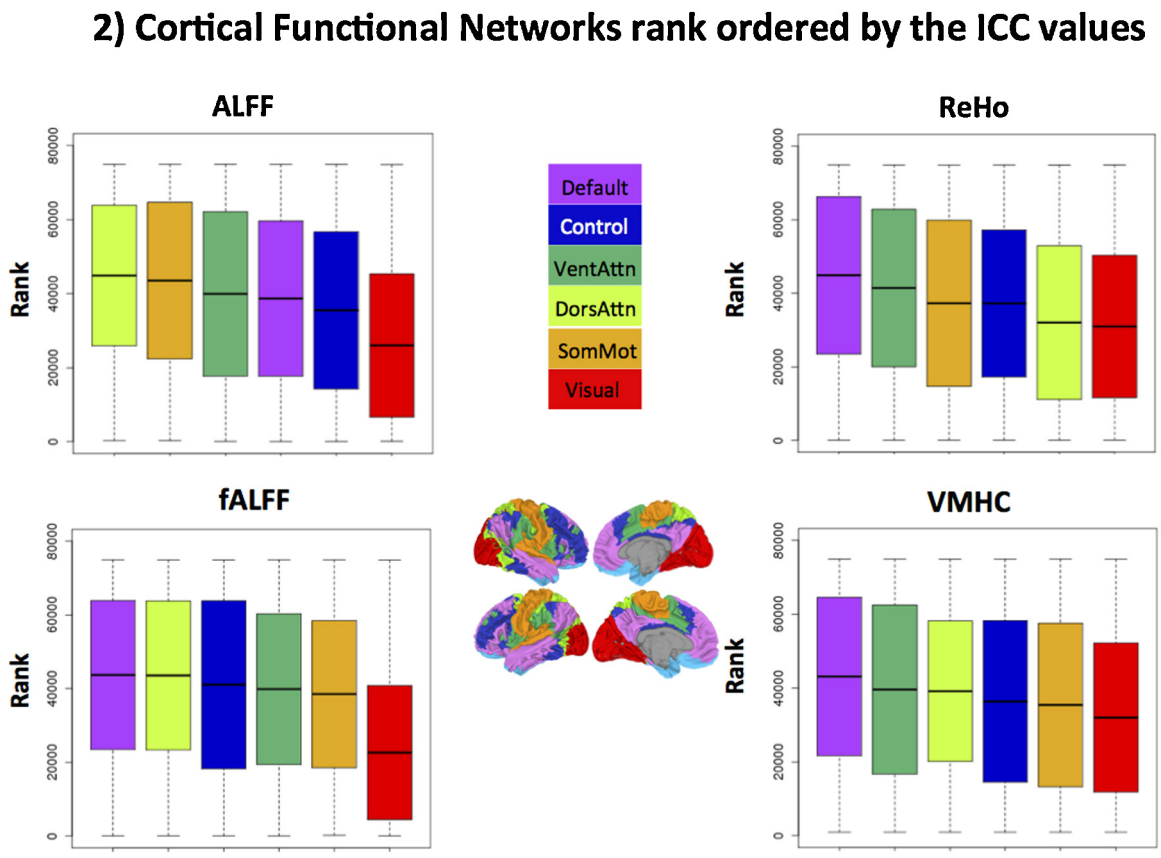

Fig. 2. Cortical functional networks rank ordered by their ICC values. Boxplots showing the distribution of voxel-wise *Z* scored ICC values of R-fMRI metrics across all 103 subjects in 6 of the 7 cortical functional parcellations defined by Yeo et al. (2010). From right to left, default network (purple), frontoparietal control network (Control; blue), ventral attention network (VentAttn, dark green), dorsal attention network (DorsAttn; light green), somatomotor (SomMot) and visual networks are the units used here. The limbic network (light blue in the surface map) in the original set of networks determined by Yeo et al. (2010) was excluded due to poor coverage in that region. The solid horizontal black line in the boxplot indicates the mean value of ICC rank in a given network and the solid box indicates the standard deviation for the distribution of ranks. The upper and lower whiskers represent the minimum and maximum ranks assigned, respectively. ALFF: amplitude of low frequency fluctuations; fALFF: fractional ALFF; ReHo: regional homogeneity; VMHC: voxel mirrored homotopic connectivity.

to specific regional information; however, it takes inter-individual variation into account and, since it provides a single measure of reliability for an entire image volume, it simplifies the comparison of different processing pipelines and imaging acquisition protocols on reliability. In summary, no single test–retest performance measure is superior to the others, as also illustrated by our findings.

4.2. Test–retest voxel-wise ICC in clinical populations

Only two previous studies in adults have examined test–retest reliability in clinical populations. In both cases, long-term reliability (2.5–16 months between scans) was assessed. Turner et al. (2012) found moderate-to-high test–retest reliability of ALFF in both adults with schizophrenia and healthy comparisons, though quantitative between-group comparisons were not conducted. Another study (Blautzik et al., 2013) reported lower test–retest reliability of ICA in elderly individuals with mild amnestic cognitive impairment relative to elderly controls, although again, quantitative between-group comparisons were not performed. Our study is the first to examine test–retest reliability in children with a developmental disorder (ADHD) and to quantitatively compare diagnostic groups. In contrast to the aforementioned studies, our examination of short-term test–retest reliability revealed regions with significantly greater ICC in TDC relative to ADHD, though not all R-fMRI metrics exhibited the same magnitude of group differences in reliability. For instance, VMHC showed no group differences in short-term reliability, and fALFF exhibited minimal group differences. Across the remaining R-fMRI measures (i.e., PCC-iFC, ReHo, and ALFF) the PCC and ventral precuneus stood out as consistently exhibiting diagnostic group differences in test–retest reliability.

We note that while head motion during scanning can affect test–retest reliability (Yan et al., 2013a; Zuo et al., 2014) and is likely to affect diagnostic group differences in reliability, in this study ADHD and TDC groups were matched for motion, which was generally low and controlled both at the individual and the group levels. We also exclude variation in other nuisance signals as a driving factor for ICC groups differences, because groups did not differ on mean voxel-wise SNR, nor on the global correlation matrix computed per Saad et al. (2013) (data not shown). As such, it is unlikely that our findings of ICC group differences reflected motion or other currently known artifacts. Rather, in interpreting our results, we reiterate that ICC depends on both inter- and intra-subject variability. Their underlying factors may differ between diagnostic groups and may be anchored in the physiopathology of ADHD, which to date remains unclarified. ADHD is highly heterogeneous in both its clinical and biological presentations, likely increasing inter-subject variability. Consistent with this notion, we noted higher between-subject variability for ADHD across R-fMRI measures. In considering the source(s) of such variability, qualitative comparisons of ICC profiles across ADHD DSM-IV-TR subtypes (i.e., ADHD-I and ADHD-C) revealed virtually no differences (data not shown), albeit in smaller subsamples. As acknowledged by recent studies calling for a biological redefinition of ADHD subtypes (e.g., Karalunas et al., 2014),

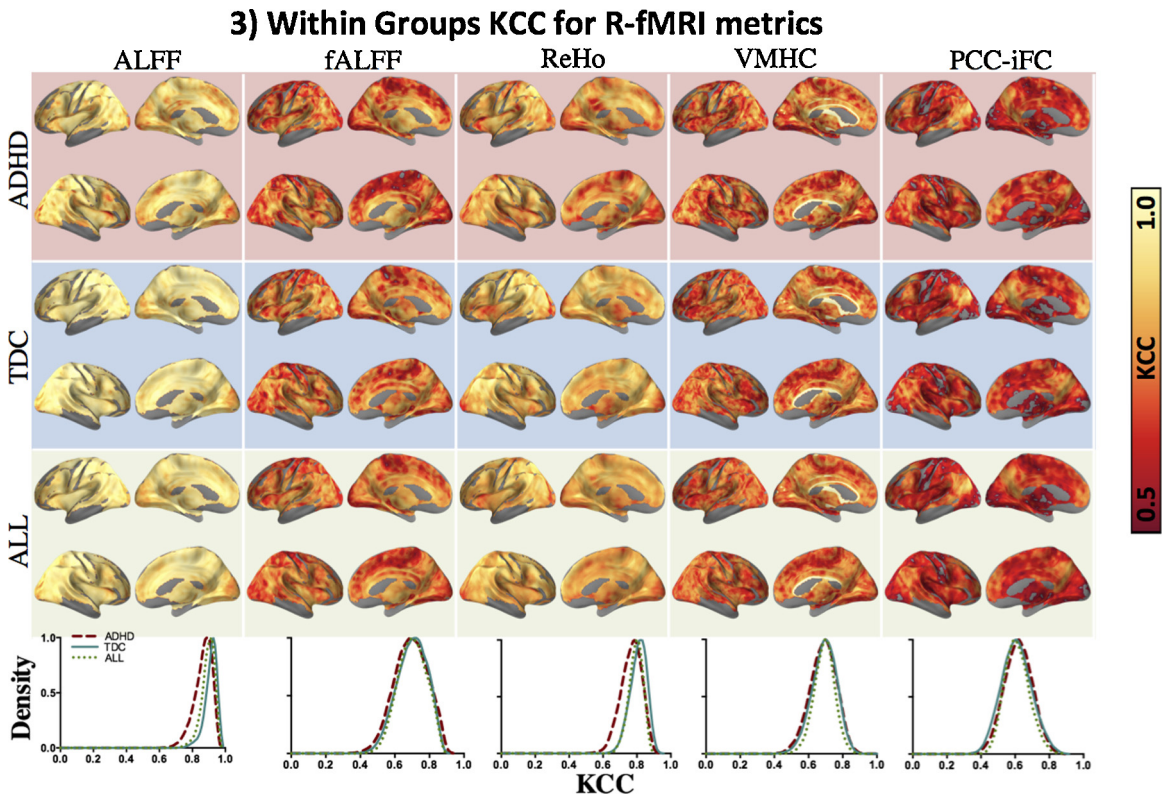

Fig. 3. KCC for different R-fMRI metrics. Spatial distribution of voxel-wise Kendall's Coefficient of Concordance (KCC) within the ADHD and TDC groups and across all subjects (ALL) from top to the bottom row. Color bar shows moderate (KCC = 0.5; dark red) to perfect concordance (KCC = 1.0; white). Lowermost row shows kernel density estimate plots illustrating the distribution of KCC, the peaks of the density curves indicate the frequently observed KCC value for a corresponding R-fMRI metric. For all figures, KCC only within gray matter was considered. For VMHC, although there is only one KCC value for each pair of homotopic voxels, results are projected onto both hemispheres to minimize confusion regarding laterality of results. Surface maps generated using pysurfer package in python (http://pysurfer.github.io/index.html) and density plots generated using Gaussian kernel density estimation available through scipy package in python (http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html).

other sources of heterogeneity need to be identified and explained if reliable biomarkers are to be attained.

Accompanying elevated between-subject variability in ADHD, we also observed elevated *intra*-subject variance in ADHD (also contributing to the lower ICC in ADHD vs. TDC). This could be related to a prominent characteristic of ADHD: increased intra-subject variability observed across a wide range of behaviors (Castellanos et al., 2005; Kofler et al., 2013). Abnormal regulation of network temporal dynamics (e.g., Allen et al., 2012) could be responsible for the ADHD-related increases in intra-subject variability. If such

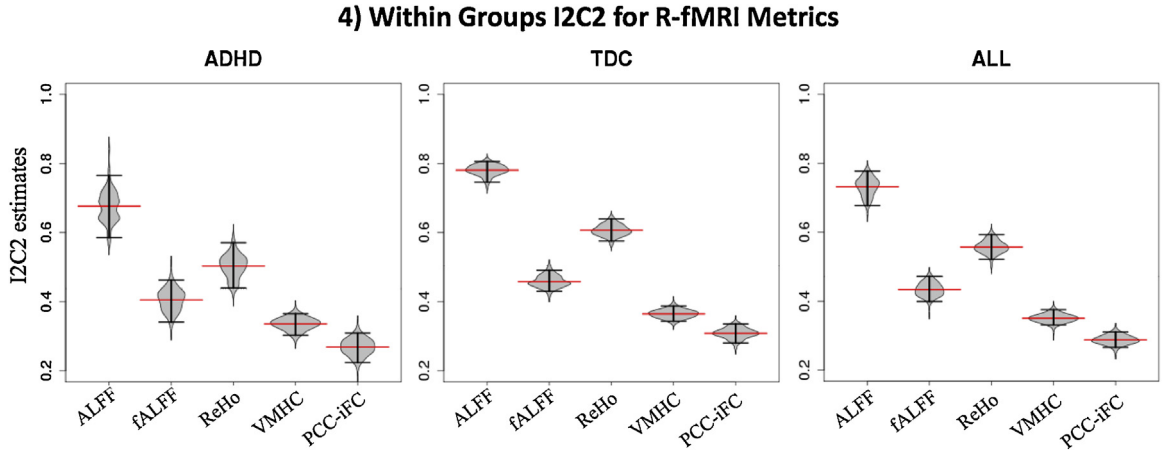

Fig. 4. Within groups I2C2 for R-fMRI metrics. The estimates of image intraclass correlation coefficients (I2C2) for different R-fMRI metrics for ADHD, TDC groups and across all subjects are represented by the red lines across each beanplot. The beanplots show the distribution of the variance in the I2C2 estimator (horizontal black lines indicate 2.5% and 97.5% confidence intervals). I2C2 estimation was constrained to a gray matter mask (threshold = 25% tissue-type probability). I2C2 estimations were performed in R using the software provided by Shou et al. (2013; http://www.smart-stats.org/wiki/image-intra-class-correlation-coefficient-i2c2). ADHD: attention deficit hyperactivity disorder; TDC: typically developing children; ALFF: amplitude of low frequency fluctuations; fALFF: fractional ALFF; ReHo; regional homogeneity; VMHC: voxel mirror homotopic connectivity; PCC-iFC: intrinsic functional connectivity (iFC) using seed based correlation with posterior cingulate cortex (PCC; seed coordinates: −8, −56, 26, 8 mm diameter sphere).

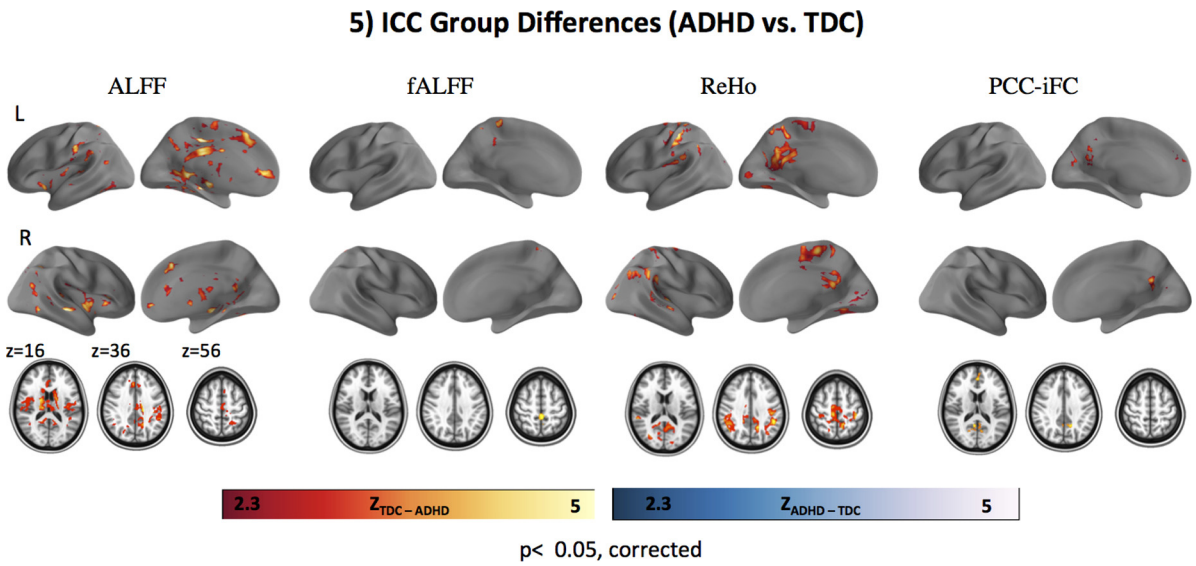

Fig. 5. ICC group differences (ADHD vs. TDC). Results of voxel-wise comparison of intraclass correlation coefficients (ICC) between diagnostic groups (ADHD vs. TDC) i.e., $Z_{TDC-ADHD}$ indicates significantly higher ICC in TDC compared to ADHD (red to yellow color bar). $Z_{ADHD-TDC}$ indicate significantly higher ICC in ADHD vs. TDC (blue to white color bar): no such areas were observed for any R-fMRI measures. No significant group differences in ICC were observed for VMHC. For all analyses, Gaussian random field theory was employed (minimum $Z > 2.3$; cluster significance: $p < .05$, corrected). Surface maps were generated using the pysurfer package in python (http://pysurfer.github.io/index.html) and axial maps (z 16, 36, 56) were generated using Analysis of Functional NeuroImages software (AFNI; http://afni.nimh.nih.gov/afni). ADHD: attention deficit hyperactivity disorder; TDC: typically developing children; ALFF: amplitude of low frequency fluctuations; fALFF: fractional ALFF; ReHo; regional homogeneity; VMHC: voxel-mirrored homotopic connectivity; PCC-iFC: intrinsic functional connectivity (iFC) of the posterior cingulate cortex (PCC; seed coordinates: −8, −56, 26, 8 mm diameter sphere).

dynamics differ between ADHD and TDC, we would expect measures of test–retest reliability to be affected. In this regard we note that PCC and precuneus exhibit substantial variation in iFC temporal dynamics in healthy control adults (Yang et al., 2014). Since PCC and ventral precuneus exhibited consistent diagnostic group differences in test–retest reliability (ICC), future studies should attempt to disentangle the influences of inter- and intra-subject factors. This will allow for a better understanding of the role of PCC/precuneus functioning in ADHD and its impact on test–retest reliability.

4.3. Limitations

Results should be interpreted in light of study limitations. First, we only examined short-term test–retest reliability, which reflects the "best case" and thus represents the expected ceiling. Typical test–retest designs involve intervals of days to months (Zuo et al., 2014). Longer intervals can be of interest for studies of interventions, although they also entail potential confounds of experiential effects or developmental changes. Of note, the Nathan Kline Institute-Rockland Sample initiatives (Nooner et al., 2012) are making openly available test–retest scans (1–4 weeks apart) that are being collected as part of an ongoing longitudinal study examining connectomes from ages 6 to 21 (see http://fcon_1000.projects.nitrc.org/indi/enhanced/ for details).

Because we used a convenience sample, we could not control for several data acquisition factors that may affect test–retest reliability. The sources of variation in R-fMRI measures may be extrinsic (e.g., experiment-related) and intrinsic (e.g., subject-related).

The former include factors such as between-scan intervals, which slightly differed between groups, and eye-status (open, closed; e.g., Patriat et al., 2013). Nevertheless, both were controlled as covariates in analyses. Other extrinsic factors such as time of day, whether a task was performed between rest scans, seasonal variation, and intrinsic factors such as female menstrual cycles or satiety could not be considered. Another possible source of intrinsic variability may be related to the reported changes of iFC over time (within a scan session e.g., Allen et al., 2012; Yang et al., 2014). No other studies have assessed the impact of variability in R-fMRI measures as a function of mental states to date and our study was not designed to do so as only ~6 min of R-fMRI data were collected. The impact of these sources of variability on test–retest reliability is unknown. Such factors should be considered in future studies.

Simultaneous physiological recordings (cardiac and respiratory rates) were not available for all our subjects. Accordingly, we did not directly correct for physiological noise, which could affect ICC measurements (e.g., Birn et al., 2014; Gotts et al., 2012). This is most relevant for ALFF, and those findings should be interpreted with caution. Specifically, ALFF is sensitive to artifactual signals arising from physiological noise aliased into the low-frequency range (i.e., 0.001–0.1 Hz) (e.g., Birn et al., 2006; Chang and Glover, 2009). Because fALFF is more effective in minimizing artifacts associated with physiological signals (Zuo et al., 2010a) we recommend either using this metric as the preferred index of the amplitude of low frequency oscillations or using it along with ALFF. As shown by Yan et al. (2013b), preprocessing steps aimed at controlling for physiological and other noise could affect ICC values. We tested for this by using three different preprocessing approaches and found a similar pattern of results, suggesting that preprocessing did not substantially affect test–retest reliability of between-group differences. Finally, in addition to quantifying the reliability of R-fMRI indices, future studies should assess the validity and accuracy of R-fMRI measures by integrating R-fMRI with electrophysiological recordings (e.g., Keller et al., 2011, 2013).

5. Conclusions

Within the same scan session, test–retest reliability in both children with ADHD and TDC is moderate to high for a range of R-fMRI measures. Although we detected regional differences in test–retest reliability between diagnostic groups, these were relatively circumscribed and varied across measures. While our results are encouraging, current limited understanding of the contributions of inter- and intra-subject variability to test–retest reliability underscores the need for large test–retest initiatives such as the

Consortium for Reliability and Repeatability (CoRR; http://fcon_1000.projects.nitrc.org/indi/CoRR/html/).

Funding sources

This work was supported by grants from National Institute of Mental Health (K23MH087770 to ADM; R01MH081218 to FXC; 5U01MH099059 to MPM); from the National Institute of Child Health and Human Development (R01HD065282 to FXC), the Stavros Niarchos Foundation (FXC and MPM); the Leon Levy Foundation (MPM, ADM, and CK); as well as the Major Joint Fund for International Cooperation and Exchange of the National Natural Science Foundation (81220108014), the Natural Science Foundation of China (81171409), the Chinese Academy of Sciences Key Research Program (CAS: KSZD-EW-TZ-002) and the support of the "CAS Hundred Talents" program to XNZ. No funding sources contributed to preparing this manuscript.

Conflict of interest

None declared.

Acknowledgments

The authors are grateful to the children and parents who made this research possible. The authors also wish to thank the research staff of the Phyllis Green and Randolph Cowen Institute for Pediatric Neuroscience for help in participant recruitment, assessment, data collection and data entry, as well as Ms. Hallie Brown for editorial suggestions of an earlier version of the manuscript. We also thank the staff of NYU Center for Brain Imaging, Mr. Keith Sanzenbach, Dr. Pablo Velasco and Dr. Edward Vessel for their support. Many of the datasets included in this manuscript were deposited, as fully anonymized data, in the ADHD200 database (http://fcon_1000.projects.nitrc.org/indi/adhd200), and/or the Autism Brain Imaging Data Exchange repository (ABIDE; http://fcon_1000.projects.nitrc.org/indi/abide/data) and/or the National Database for Autism Research (NDAR; http://ndar.nih.gov/).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dcn.2015.08.003.

References

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D., 2012. Tracking whole-brain connectivity dynamics in the resting state. Cereb. Cortex 24, 663–676.

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Text Revision. American Psychiatric Association, Washington, DC.

Anderson, J.S., Druzgal, T.J., Froehlich, A., Du Bray, M.B., Lange, N., Alexander, A.L., Abildskov, T., Nielsen, J.A., Cariello, A.N., Cooperrider, J.R., Bigler, E.D., Lainhart, J.E., 2011. Decreased interhemispheric functional connectivity in autism. Cereb. Cortex 21, 1134–1146.

Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., Poulin, R., Buckner, R.L., 2010. Functional-anatomic fractionation of the brain's default network. Neuron 65, 550–562.

Avants, B.B., Tustison, N.J., Song, G., Gee, J.C., 2010. ANTs, Advanced Normalization Tools. http://www.picsl.upenn.edu/ANTS/

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101.

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.

Birn, R., Cornejo, M.D., Molloy, E.K., Patriat, R., Meier, T., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., 2014. The influence of physiological noise correction on test–retest reliability of resting-state functional connectivity. Brain Connect. 4, 511–522.

Blautzik, J., Keeser, D., Berman, A., Paolini, M., Kirsch, V., Mueller, S., Coates, U., Reiser, M., Teipel, S.J., Meindl, T., 2013. Long-term test–retest reliability of resting-state networks in healthy elderly subjects and mild cognitive impairment patients. J. Alzheimer's Dis. 34, 741–754.

Castellanos, F.X., Sonuga-Barke, E.J.S., Scheres, A., Di Martino, A., Hyde, C., Walters, J.R., 2005. Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. Biol. Psychiatry 57, 1416–1423.

Castellanos, F.X., Margulies, D.S., Kelly, A.M.C., Uddin, L.Q., Ghaffari, M., Kirsch, A., Shaw, D., Shehzad, Z., Di Martino, A., Biswal, B., Sonuga-Barke, E.J.S., Rotrosen, J., Adler, L.A., Milham, M.P., 2008. Cingulate-precuneus interactions: a new locus of dysfunction in adult attention-deficit/hyperactivity disorder. Biol. Psychiatry 63, 332–337.

Castellanos, F.X., Di Martino, A., Craddock, R.C., Mehta, A.D., Milham, M.P., 2013. Clinical applications of the functional connectome. Neuroimage 80, 527–540.

Chang, C., Glover, G.H., 2009. Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. Neuroimage 47, 1448–1459.

Chabernaud, C., Mennes, M., Kelly, C., Nooner, K., Di Martino, A., Castellanos, F.X., Milham, M.P., 2012. Dimensional brain–behavior relationships in children with attention-deficit/hyperactivity disorder. Biol. Psychiatry 71, 434–442.

Collin, G., van den Heuvel, M.P., 2013. The ontogeny of the human connectome: development and dynamic changes of brain connectivity across the life span. Neuroscientist 19, 616–628.

Conners, C.K., 1997. Conners' Rating Scales-Revised Technical Manual. Multi-Health Systems, Inc., North Tonawanda, NY.

Craddock, R.C., Jbabdi, S., Yan, C.G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P., 2013. Imaging human connectomes at the macroscale. Nat. Med. 10, 524–539.

Dennis, E.L., Thompson, P.M., 2014. Reprint of: mapping connectivity in the developing brain. Int. J. Dev. Neurosci. 32, 41–57.

Di Martino, A., Zuo, X.N., Kelly, C., Grzadzinski, R., Mennes, M., Schvarcz, A., Rodman, J., Lord, C., Castellanos, F.X., Milham, M.P., 2013. Shared and distinct intrinsic functional network centrality in autism and attention-deficit/hyperactivity disorder. Biol. Psychiatry 74, 623–632.

Di Martino, A., Fair, D.A., Kelly, C., Satterthwaite, T.D., Castellanos, F.X., Thomason, M.E., Craddock, R.C., Luna, B., Leventhal, B.L., Zuo, X.N., Milham, M.P., 2014. Unraveling the miswired connectome: a developmental perspective. Neuron 83, 1335–1353.

Fisher, R.A., 1958. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Fox, M.D., Greicius, M., 2010. Clinical applications of resting state functional connectivity. Front. Syst. Neurosci. 4, 19.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355.

Gotts, S.J., Simmons, W.K., Milbury, L.A., Wallace, G.L., Cox, R.W., Martin, A., 2012. Fractionation of social brain circuits in autism spectrum disorders. Brain 135, 2711–2725.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72.

Hagmann, P., Grant, P.E., Fair, D.A., 2012. MR connectomics: a conceptual framework for studying the developing brain. Front. Syst. Neurosci. 6, 43.

Hoffman, E.J., State, M.W., 2010. Progress in cytogenetics: implications for child psychopathology. J. Am. Acad. Child Adolesc. Psychiatry 49, 736–751.

Hollingshead, A.B., 1975. Four Factor Index of Social Status, New Haven.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.

Jenkinson, M., 2003. Fast, automated, N-dimensional phase-unwrapping algorithm. Magn. Reson. Med. 49, 193–197.

Jezzard, P., 2012. Correction of geometric distortion in fMRI data. Neuroimage 62, 648–651.

Karalunas, S.L., Fair, D., Musser, E.D., Aykes, K., Iver, S.P., Nigg, J.T., 2014. Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: toward biologically based nosologic criteria. JAMA Psychiatry 71, 1015–1024.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., Ryan, N., 1997. The schedule for affective disorders and schizophrenia for school aged children: present and lifetime version (K-SADS-PL): initial reliability and validity data. J. Am. Acad. Child Adolesc. Psychiatry 36, 980–988.

Kaufman, A.S., Kaufman, N.L., 2004. Kaufman Brief Intelligence Test, 2nd ed. Pearson, Inc., Bloomington, MN.

Kendall, M.G., Smith, B.B., 1939. The problem of m rankings. Ann. Math. Stat. 10, 275–287.

Keller, C.J., Bickel, S., Entz, L., Ulbert, I., Milham, M.P., Kelly, C., Mehta, A.D., 2011. Intrinsic functional architecture predicts electrically evoked responses in the human brain. Proc. Natl. Acad. Sci. U. S. A. 108, 10308–10313.

Keller, C.J., Bickel, S., Honey, C.J., Groppe, D.M., Entz, L., Craddock, R.C., Lado, F.A., Kelly, C., Milham, M., Mehta, A.D., 2013. Neurophysiological investigation of spontaneous correlated and anticorrelated fluctuations of the BOLD signal. J. Neurosci. 33, 6333–6342.

Kofler, M.J., Rapport, M.D., Sarver, D.E., Raiker, J.S., Orban, S.A., Friedman, L.M., Kolomeyer, E.G., 2013. Reaction time variability in ADHD: a meta-analytic review of 319 studies. Clin. Psychol. Rev. 33, 795–811.

Konishi, S., Gupta, A.K., 1987. Inferences about interclass and intraclass correlations from familial data. Biostatistics 38, 225–233.

Koyama, M.S., Di Martino, A., Zuo, X.N., Kelly, C., Mennes, M., Jutagir, D.R., Castellanos, F.X., Milham, M.P., 2011. Resting-state functional connectivity indexes reading competence in children and adults. J. Neurosci. 31, 8617–8624.
Kruskal, W., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 583–621.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Levy, K.J., Narula, S.C., 1978. Testing hypotheses concerning partial correlations – some methods and discussion. Int. Stat. Rev. 46, 215–218.
Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., Sikka, S., Gutman, D., Bangaru, S., Schlachter, R.T., Kamiel, S.M., Anwar, A.R., Hinz, C.M., Kaplan, M.S., Rachlin, A.B., Adelsberg, S., Cheung, B., Khanuja, R., Yan, C., Craddock, R.C., Calhoun, V., Courtney, W., King, M., Wood, D., Cox, C.L., Kelly, A.M., Di Martino, A., Petkova, E., Reiss, P.T., Duan, N., Thomsen, D., Biswal, B., Coffey, B., Hoptman, M.J., Javitt, D.C., Pomara, N., Sidtis, J.J., Koplewicz, H.S., Castellanos, F.X., Leventhal, B.L., Milham, M.P., 2012. The NKI-Rockland Sample: a model for accelerating the pace of discovery science in psychiatry. Front. Neurosci. 6, 152.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Patriat, R., Molloy, E.K., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., Birn, R.M., 2013. The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. Neuroimage 78, 463–473.
Posner, J., Park, C., Wang, Z., 2014. Connecting the dots: a review of resting connectivity MRI studies in attention-deficit/hyperactivity disorder. Neuropsychol. Rev. 24, 3–15.
Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.
Saad, Z.S., Reynolds, R.C., Jo, H.J., Gotts, S.J., Chen, G., Martin, A., Cox, R.W., 2013. Correcting brain-wide correlation differences in resting-state FMRI. Brain Connect. 3, 339–352.
Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632.
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., Petkova, E., Castellanos, F.X., Milham, M.P., 2009. The resting brain: unconstrained yet reliable. Cereb. Cortex 19, 2209–2229.
Shou, H., Eloyan, A., Lee, S., Zipunnikov, V., Crainiceanu, A.N., Nebel, N.B., Caffo, B., Lindquist, M.A., Crainiceanu, C.M., 2013. Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (I2C2). Cognit. Affect. Behav. Neurosci. 13, 714–724.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.
Song, J., Desphande, A.S., Meier, T.B., Tudorascu, D.L., Vergun, S., Nair, V.A., Biswal, B.B., Meyerand, M.E., Birn, R.M., Bellec, P., Prabhakaran, V., 2012. Age-related differences in test–retest reliability in resting-state brain functional connectivity. PLoS ONE 7, e49847.
State, M.W., 2010. The genetics of child psychiatric disorders: focus on autism and Tourette syndrome. Neuron 68, 254–269.
Thomason, M.E., Dennis, E.L., Joshi, A.A., Joshi, S.H., Dinov, I.D., Chang, C., Henry, M.L., Johnson, R.F., Thompson, P.M., Toga, A.W., Glover, G.H., Van Horn, J.D., Gotlib, I.H., 2011. Resting-state fMRI can reliably map neural networks in children. Neuroimage 55, 165–175.
Turner, J.A., Chen, H., Mathalon, D.H., Allen, E.A., Mayer, A.R., Abbott, C.C., Calhoun, V.D., Bustillo, J., 2012. Reliability of the amplitude of low-frequency fluctuations in resting state fMRI in chronic schizophrenia. Psychiatry Res. 201, 253–255.
Uddin, L.Q., Supekar, K., Menon, V., 2010. Typical and atypical development of functional human brain networks: insights from resting-state FMRI. Front. Syst. Neurosci. 4, 21.
Wechsler, D., 1999. Wechsler Abbreviated Scale of Intelligence (WASI). The Psychological Corporation, San Antonio, TX.
Yan, C.G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R.C., Di Martino, A., Li, Q., Zuo, X.N., Castellanos, F.X., Milham, M.P., 2013a. A comprehensive assessment of regional variation in the impact of micromovement head motion on functional connectomics. Neuroimage 76C, 183–201.
Yan, C.G., Craddock, R.C., Zuo, X.N., Zang, Y.F., Milham, M.P., 2013b. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. Neuroimage 80, 246–262.
Yang, Z., Craddock, R.C., Margulies, D.S., Yan, C.G., Milham, M.P., 2014. Common intrinsic connectivity states among posteromedial cortex subdivisions: insights from analysis of temporal dynamics. Neuroimage 93, 124–137.
Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2010. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165.
Zang, Y., Jiang, T., Lu, Y., He, Y., Tian, L., 2004. Regional homogeneity approach to fMRI data analysis. Neuroimage 22, 394–400.
Zang, Y.F., Yong, H., Chao-Zhe, Z., Qing-Jiu, C., Man-Qiu, S., Meng, L., Li-Xia, T., Tian-Zi, J., Wang, Y.F., 2007. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 29, 83–91.
Zou, Q.H., Zhu, C.Z., Yang, Y., Zuo, X.N., Long, X.Y., Cao, Q.J., Wang, Y.F., Zang, Y.F., 2008. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. J. Neurosci. Methods 172, 137–141.
Zuo, X.N., Di Martino, A., Kelly, C., Shehzad, Z.E., Gee, D.G., Klein, D.F., Castellanos, F.X., Biswal, B.B., Milham, M.P., 2010a. The oscillating brain: complex and reliable. Neuroimage 49, 1432–1445.
Zuo, X.N., Kelly, C., Adelstein, J.S., Klein, D.F., Castellanos, F.X., Milham, M.P., 2010b. Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach. Neuroimage 49, 2163–2177.
Zuo, X.N., Kelly, C., Di Martino, A., Mennes, M., Margulies, D.S., Bangaru, S., Grzadzinski, R., Evans, A., Zang, Y., Castellanos, F.X., Milham, M.P., 2010c. Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. J. Neurosci. 30, 15034–15043.
Zuo, X.N., Xu, T., Jiang, L., Yang, Z., Cao, X.Y., He, Y., Zang, Y.F., Castellanos, F.X., Milham, M.P., 2013. Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. Neuroimage 65, 374–386.
Zuo, X.N., Xing, X.X., 2014. Test–retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. Neurosci. Biobehav. Rev. 45C, 100–118.
Zuo, X.N., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blautzik, J., Breitner, J.C., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S.J., Courtney, W., Craddock, R.C., Di Martino, A., Dong, H.M., Fu, X., Gong, Q., Gorgolewski, K.J., Han, Y., He, Y., He, Y., Ho, E., Holmes, A., Hou, X.H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S.M., Lainhart, J., Lei, X., Li, H.J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y., Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D.S., Mayer, A.R., Meindl, T., Meyerand, M.E., Nan, W., Nielsen, J.A., O'Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villringer, A., Wang, H., Wang, K., Wei, D., Wei, G.X., Weng, X.C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zhen, Z., Zhou, Y., Zhu, X.T., Milham, M.P., 2014. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci. Data 1, 140049.