Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: A multicentric resting-state fMRI study

Article *in* NeuroImage · July 2015

DOI: 10.1016/j.neuroimage.2015.07.010 · Source: PubMed

45 authors, including:

Jorge Jovicich
Università degli Studi di Trento
172 PUBLICATIONS **5,917** CITATIONS

SEE PROFILE

Moira Marizzoni
IRCCS Centro San Giovanni di Dio, Fatebenefratelli, Brescia
101 PUBLICATIONS **1,022** CITATIONS

SEE PROFILE

Rocco Marchitelli
French Institute of Health and Medical Research
21 PUBLICATIONS **141** CITATIONS

SEE PROFILE

Roser Sala-Llonch
University of Barcelona
77 PUBLICATIONS **3,092** CITATIONS

SEE PROFILE

Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: A multicentric resting-state fMRI study

Jorge Jovicich, Ludovico Minati, Moira Marizzoni, Rocco Marchitelli, Roser Sala-Llonch, David Bartrés-Faz, Jennifer Arnold, Jens Benninghoff, Ute Fiedler, Luca Roccatagliata, Agnese Picco, Flavio Nobili, Oliver Blin, Stephanie Bombois, Renaud Lopes, Régis Bordet, Julien Sein, Jean-Philippe Ranjeva, Mira Didic, Hélène Gros-Dagnac, Pierre Payoux, Giada Zoccatelli, Franco Alessandrini, Alberto Beltramello, Núria Bargalló, Antonio Ferretti, Massimo Caulo, Marco Aiello, Carlo Cavaliere, Andrea Soricelli, Lucilla Parnetti, Roberto Tarducci, Piero Floridi, Magda Tsolaki, Manos Constantinidis, Antonios Drevelegas, Paolo Maria Rossini, Camillo Marra, Peter Schönknecht, Tilman Hensch, Karl-Titus Hoffmann, Joost P. Kuijer, Pieter Jelle Visser, Philip Scheltens, Giovanni B. Frisoni

PII: S1053-8119(15)00619-9
DOI: doi: 10.1016/j.neuroimage.2015.07.010
Reference: YNIMG 12399

To appear in: *NeuroImage*

Received date: 7 February 2015
Accepted date: 3 July 2015

Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: a multicentric resting-state fMRI study

Jorge Jovicich a,*,1, Ludovico Minati a,ai,1, Moira Marizzoni b, Rocco Marchitelli a, Roser Sala-Llonch c, David Bartrés-Faz c, Jennifer Arnold e, Jens Benninghoff e, Ute Fiedler e, Luca Roccatagliata f,g, Agnese Picco h, Flavio Nobili h, Oliver Blin i, Stephanie Bombois j, Renaud Lopes k, Régis Bordet ah, Julien Sein l, Jean-Philippe Ranjeva l, Mira Didic m,n, Hélène Gros-Dagnac o,p, Pierre Payoux o,p, Giada Zoccatelli q, Franco Alessandrini q, Alberto Beltramello q, Núria Bargalló r, Antonio Ferretti s,t, Massimo Caulo s,t, Marco Aiello u, Carlo Cavaliere u, Andrea Soricelli u,v, , Lucilla Parnetti w, Roberto Tarducci x, Piero Floridi ag, Magda Tsolaki y, Manos Constantinidis z, Antonios Drevelegas z,aa, Paolo Maria Rossini ab,ac, Camillo Marra ad, Peter Schönknecht am, Tilman Hensch am, Karl-Titus Hoffmann ae, Joost P. Kuijer ah, Pieter Jelle Visser al, Philip Scheltens ak, Giovanni B. Frisoni b,af, The PharmaCog Consortium

a Center for Mind/Brain Sciences (CIMEC), University of Trento, Rovereto, Italy

b LENITEM Laboratory of Epidemiology, Neuroimaging, & Telemedicine — IRCCS San Giovanni di Dio-FBF, Brescia, Italy

c Department of Psychiatry and Clinical Psychobiology, Universitat de Barcelona and IDIBAPS, Barcelona, Spain

e LVR-Clinic for Psychiatry and Psychotherapy, Institutes and Clinics of the University Duisburg-Essen, Essen, Germany

f Department of Neuroradiology, IRCSS San Martino University Hospital and IST, Genoa, Italy

g Department of Health Sciences, University of Genoa, Genoa, Italy

h Department of Neuroscience, Ophthalmology, Genetics and Mother–Child Health (DINOGMI), University of Genoa, Genoa, Italy

i Pharmacology, Assistance Publique — Hôpitaux de Marseille, Aix-Marseille University — CNRS, UMR 7289, Marseille, France

j Department of Neurology, EA1046, Lille University, Lille, France

k Department of Neuroradiology, EA1046, Lille University, Lille, France

l CRMBM–CEMEREM, UMR 7339, Aix Marseille Université — CNRS, Marseille, France

m APHM, CHU Timone, Service de Neurologie et Neuropsychologie, Marseille, France

n Aix Marseille Université, Inserm, INS UMR_S 1106, 13005, Marseille, France

o INSERM, Imagerie cérébrale et handicaps neurologiques, UMR 825, Toulouse, France

p Université de Toulouse, UPS, Imagerie cérébrale et handicaps neurologiques, UMR 825, CHU Purpan, Place du Dr Baylac, Toulouse Cedex 9, France

q Department of Neuroradiology, General Hospital, Verona, Italy

r Department of Neuroradiology and Magnetic Resonance Image core Facility, Hospital Clínic de Barcelona, IDIBAPS, Barcelona, Spain

s Department of Neuroscience Imaging and Clinical Sciences, University “G. d'Annunzio” of Chieti, Italy

t Institute for Advanced Biomedical Technologies (ITAB), University “G. d'Annunzio” of Chieti, Italy

u IRCCS SDN, Naples, Italy

v University of Naples Parthenope, Naples, Italy

w Section of Neurology, Centre for Memory Disturbances, University of Perugia, Perugia, Italy

x Medical Physics Unit, Perugia General Hospital, Perugia, Italy

y 3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece

z Interbalkan Medical Center of Thessaloniki, Thessaloniki, Greece

aa Department of Radiology, Aristotle University of Thessaloniki, Thessaloniki, Greece

ab Dept. Geriatrics, Neuroscience & Orthopaedics, Catholic University, Policlinic Gemelli, Rome, Italy

ac IRCSS S.Raffaele Pisana, Rome, Italy

ad Center for Neuropsychological Research, Catholic University, Rome, Italy

ae Department of Neuroradiology, University Hospital Leipzig, Leipzig, Germany

af Memory Clinic and LANVIE, Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland

ag Neuroradiology Unit, Perugia General Hospital, Perugia, Italy

ah Department of Pharmacology, EA1046, Lille University, Lille, France

ai Scientific Department, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

aj Department of Psychiatry and Neuropsychology, Alzheimer Center Limburg, University of Maastricht, Maastricht, The Netherlands

ak Alzheimer Centre and Department of Neurology, Vrije Universiteit University Medical Center, Amsterdam, The Netherlands

al Deptartment of Physics and Medical Technology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands

am Department of Psychiatry, University Hospital Leipzig, Leipzig, Germany

Keywords (max 6):
Default Mode Network,
Reproducibility,
Functional connectivity,
Multi-center,
Multi-site MRI

* Corresponding author:
Jorge Jovicich
Assistant Professor
Center for Mind/Brain Sciences, University of Trento, Italy
Phone: +39-0461-88 3064
Fax: +39-0461-88 3066
Email: jorge.jovicich@unitn.it

[1]Authors contributed equally to this work.

Abstract

To date, limited data are available regarding the inter-site consistency of test-retest reproducibility of functional connectivity measurements, in particular with regards to integrity of the default mode network (DMN) in elderly participants. We implemented a harmonized resting-state fMRI protocol on 13 clinical scanners at 3.0 Tesla using vendor-provided sequences. Each site scanned a group of 5 healthy elderly participants twice, at least a week apart. We evaluated inter-site differences and test-retest reproducibility of both temporal signal-to-noise ratio (tSNR) and functional connectivity measurements derived from: i) seed-based analysis (SBA) with seed in the posterior cingulate cortex (PCC), ii) group independent component analysis (ICA) separately for each site (site ICA), and iii) consortium ICA, with group ICA across the whole consortium. Despite protocol harmonization, significant and quantitatively important inter-site differences remained in the tSNR of resting-state fMRI data; these were plausibly driven by hardware and pulse sequence differences across scanners which could not be harmonized. Nevertheless, the tSNR test-retest reproducibility in the consortium was high (ICC=0.81). The DMN was consistently extracted across all sites and analysis methods. While significant inter-site differences in connectivity scores were found, there were no differences in the associated test-retest error. Overall, ICA measurements were more reliable than PCC-SBA, with site ICA showing higher reproducibility than consortium ICA. Across the DMN nodes, the PCC yielded the most reliable measurements (≈4% test-retest error, ICC=0.85), the medial frontal cortex the least reliable (≈12%, ICC=0.82) and the lateral parietal cortices were in between (site ICA). Altogether these findings support usage of harmonized multisite studies of resting-state functional connectivity to characterize longitudinal effects in studies that assess disease progression and treatment response.

1 Introduction

Functional connectivity, i.e. resting-state activity synchronization, among the constituent nodes of the Default Mode Network (DMN) (Gusnard and Raichle, 2001, Greicius et al., 2003, Fox and Raichle, 2007, Buckner et al., 2008) is sensitive to normal ageing and neuropsychiatric disease (Bassett and Bullmore, 2009, Rosazza and Minati, 2011, Anticevic et al., 2012, Damoiseaux, 2012, Castellanos et al., 2013, Pievani et al., 2014). Longitudinal assessment of DMN connectivity is therefore of interest as a potential biomarker of disease prediction/progression and treatment response (Persson et al., 2014). Despite the associated technical and logistical challenges, multicenter longitudinal studies are particularly attractive as they allow the acquisition of large datasets over diverse populations while distributing load across consortium participants (Van Horn and Toga, 2009).

The sensitivity of longitudinal studies is often limited by between-sessions test-retest reproducibility of the parameter(s) of interest (Atkinson et al., 2001, Castellanos et al., 2013). As recently reviewed, several factors can affect the test-retest reproducibility of DMN connectivity measurements at a single-site level, including demographics, psychophysiological state, scanner hardware, pulse sequence settings, data preprocessing and analysis methods. Nevertheless, single-site studies have indicated that the between-sessions test-retest reproducibility of the DMN is fair, and that DMN functional connectivity measurements may therefore deserve consideration as a functional biomarker in longitudinal studies (Zuo and Xing, 2014). However, the reproducibility from single sites using different MRI systems, different acquisition protocol details and different analysis methods cannot necessarily be extrapolated to the reproducibility that may be found in a consortium using a harmonized acquisition and analysis protocol.

In fact, until very recently, limited multisite resting-state fMRI data have been available, making it difficult to evaluate the consistency of test-retest reproducibility of DMN connectivity.

This is an important shortcoming, because heterogeneous reproducibility can bias and severely limit the power of multisite longitudinal investigations. The Consortium for Reliability and Reproducibility (CoRR: http://fcon_1000.projects.nitrc.org/indi/CoRR/html/index.html) is a very recent effort which aims at addressing these limitations by creating and maintaining a public repository for resting state fMRI reproducibility data (Zuo et al., 2014).

Comparisons between identical 3.0 T scanners conducted on healthy participants have not revealed significant differences in temporal signal-to-noise ratio (tSNR), in the default mode and attention networks (Huang et al., 2012), nor in graph–based connectivity parameters (Braun et al., 2012). Unfortunately, such studies do not reflect the fact that multi-site investigations, almost invariably involve multiple scanner configurations (models and vendors) having heterogeneous hardware performance (number of channels, RF noise factor, gradient strength, etc.) and software settings (pulse sequence design, reconstruction and filtering parameters, etc.).

There also remains some controversy around which data analysis method is preferable to measure DMN connectivity in multisite settings. Since its inception in seminal work demonstrating intrinsic functional connectivity in the resting brain (Biswal et al., 1995), seed-based analysis (SBA) has remained a popular choice. The precuneus and posterior cingulate cortex (PCC) play a pivotal role in DMN connectivity, and as such, consideration of the blood-oxygen level-dependent signal (BOLD) average time-course from the PCC robustly characterizes the DMN at a single subject level (Andrews-Hanna et al., 2007, Buckner et al., 2008, Fransson and Marrelec, 2008). An alternative method not involving anatomical priors is independent component analysis (ICA) (Calhoun et al., 2001, ta et al., 2005). While this method is arguably more robust than SBA to physiological and movement-related noise, the choice of the number of spatial components is not trivial and may entail a trade-off between avoiding splitting the DMN over multiple components and avoiding merging of unrelated networks. Diverse implementations of ICA are available and give comparable results in single-site studies conducted mostly on healthy young participants (Shehzad et al., 2009, Meindl et al., 2010, Van Dijk et al., 2010, Zuo et al., 2010, Li et al., 2012), but to our knowledge no

data are available regarding the test-retest reproducibility of ICA-derived DMN measurements in multisite studies of elderly subjects.

Predicated on the above, we set out to: i) implement a harmonized international multi-site 3.0 Tesla MRI data acquisition protocol for resting-state fMRI (13 sites in 6 European countries, covering 3 common scanner vendors and 8 different scanner models), ii) acquire across-session test-retest data (at least one week apart) on healthy elderly participants (5 per site), and iii) evaluate the between-session reproducibility of tSNR and DMN functional connectivity measured using ICA and SBA. For ICA, group analysis was performed both at the single-site level (separate decomposition and back-reconstruction for each site) and at the consortium level (pooling all sites together).

2 Materials and Methods

Participant demographics, study design and data preparation steps have been described in recent morphometry (Jovicich et al., 2013) and diffusion (Jovicich et al., 2014) studies from the PharmaCog project, but are repeated here following applicable updates. The test-retest raw data from this study are publicly available (https://neugrid4you.eu/).

2.1 Participants

Thirteen sites across Italy (Verona, Genoa, Rome, Chieti, Perugia and Naples), Spain (Barcelona), France (Marseille, Lille, and Toulouse), Germany (Essen, Leipzig), Greece (Thessaloniki) and the Netherlands (Amsterdam) provided imaging data. Each site recruited 5 participants in the age range 50-80 years, who underwent two imaging sessions 7-60 days apart at the same site. This short test-retest interval minimized potential biological changes, allowing us to specifically address the reproducibility error inherent in the measurement techniques. Participant demographics and test-retest intervals are reported in Table 1. All participants had no history of

psychiatric, neurological or systemic disease, were Caucasian and provided written informed consent following procedures approved by the local institutional review board of the institution where scanning was performed. Detailed inclusion and exclusion criteria are described elsewhere (Jovicich et al., 2014).

2.2 Data acquisition

Scanner vendors, models and software versions are listed in Table 1. While each session involved a range of structural imaging sequences, for this study we only utilized the volumetric T1 series (Jovicich et al., 2013) and the resting state echo-planar imaging acquisitions. For EPI acquisitions, implemented with manufacturer-provided single-shot sequences, the following parameters could be set identically at all sites: nominal voxel size 3 x 3 x 3 mm^3, TE = 30 ms, TR = 2.7 s, $\alpha = 85^{\circ}$ (Ernst angle), bicommissural orientation with interleaved slice order (equidistant on Philips, default interleaved on GE and Siemens, see Table 1 for unexpected variations to this prescription), 0.45 mm slice gap, 40 slices, 200 volumes, no parallel imaging. The TR was set to the smallest common value attainable across all scanners. Acquisition time was 9 min, a duration known from previous work to yield reproducible connectivity results (Van Dijk et al., 2010, Birn et al., 2013, Liao et al., 2013, Zuo et al., 2013). Participants were instructed to relax, keep their eyes closed and try to avoid engaging into any thinking.

Some acquisition parameters including head RF coil design, pulse sequence and fat suppression method were impossible to standardize due to inherent system differences; these parameters were thus determined separately for each scanner and are reported in Table 1. Images were reconstructed and exported disabling any additional user-controllable filtering steps, and multi-channel images were combined using the sum-of-squares method.

Before acquiring brain data, each site ran a baseline test using fBIRN agar phantoms (Friedman and Glover, 2006a), distributed centrally and kept at each acquisition site (Glover et al.,

2012). All sites were requested to perform 5 acquisitions at 1 week intervals, each of which consisted of two repetitions of the same rsfMRI protocol used for the participants: the first to warm up gradient coils, and the second for actual measurement.

Data were anonymized and stored as previously described (Jovicich et al., 2013, Jovicich et al., 2014).

2.3 Phantom tSNR analysis

Scanner stability metrics were derived from those proposed by the fBIRN Consortium (Friedman and Glover, 2006a, Glover et al., 2012). For brevity here we focus on phantom tSNR, defined as the voxel-wise functional image intensity mean along the linearly detrended time course (first 4 volumes eliminated to allow for steady state equilibrium) divided by the temporal standard deviation, finally averaged across voxels on a region-of-interest positioned centrally (20x20 pixels, central phantom slice). This represents a well-accepted measure of temporal stability having direct relevance to resting-state fMRI analyses (Parrish et al., 2000, LaBar et al., 2001). For the purpose of correlation with brain data, tSNR measurements from all valid phantom sessions were combined. Phantom tSNR test-retest reproducibility was estimated as a mean of all possible pairwise combinations of measurements per site.

2.4 Brain data preprocessing and tSNR measurement

Data were preprocessed according to the pipeline detailed in Figure 1, which involved SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) running under Matlab R2012a (The MathWorks, Inc., Natick MA, USA) and code developed in-house. These steps consisted of slice-timing correction (according to site-specific assumed order, as detailed in Table 1), rigid-body realignment and

removal of movement-susceptibility interactions, subtraction of baseline fluctuations via fitting a fourth-order polynomial, low-pass filtering using a second-order Butterworth filter having f_{3dB}=0.09 Hz, removal of covariance with the 6 first-order head movement vectors (translations and rotations) and with average white matter and cerebrospinal fluid signals derived from individual tissue masks derived from the structural scan. Nuisance regressors were temporally filtered as above. Head movement magnitude was quantified as median frame-to-frame displacement.

To avoid interpolation artifacts, voxel-wise human brain tSNR maps were calculated in native space on the filtered time-series and thereafter warped to MNI space alongside the EPI volumes by applying a non-linear transform determined from the average EPI volume and SPM8's built-in EPI template (resolution 2x2x2 mm^3). The average whole-brain tSNR over the individual gray matter mask for each scan was calculated. Prior to further analyses, the normalized EPI volumes were spatially smoothed through a Gaussian kernel having 8 mm full width at half-maximum (Figure 1).

2.5 Seed-based and independent component analysis (SBA, ICA)

Seed-based analysis (SBA, flowchart in Figure 1) was performed as a fixed-effect group analysis based on the average time-course calculated from a non-spherical seed located in precuneus and posterior cingulate cortex (PCC). For consistency with ICA, the same temporal preprocessing step was applied, i.e. mean removal per time-point, and spatial maps were transformed into z-scores (Figure 1). SBA was performed separately considering two seed masks: 1) one with MNI centroid coordinates [0, -50, 28] mm, volume 27 ml and derived from an independent dataset of healthy participants (Rosazza and Minati, 2011, Rosazza et al., 2012), 2) the other with MNI centroid coordinates [0, -60, 28] mm, volume 27 ml and derived from the aggregate DMN component generated by consortium ICA10, thresholded at z>6.6 to yield a matched seed volume. The two seed masks are shown overlapped in Supplementary Figure 1. While the atlas

mask avoided issues of circularity and was derived from a homogeneous sample, the ICA mask represented a better match to our data in terms of participant age and scanner field strength.

Independent component analysis (ICA, flowchart in Figure 1) was performed using GroupICAT/GIFT V.3.0a, separately extracting 10 and 20 independent components (Calhoun et al., 2001). Decomposition in a larger number of components was not attempted as results with 20 components were significantly worse than those with 10 components (see Results section), and an even larger number of components would have resulted in substantial DMN splitting as detailed elsewhere (Abou-Elseoud et al., 2010, Zuo et al., 2010). ICA was computed in two ways: 1) separately for each site, i.e. combining test and retest sessions for 5 participants (thereafter: site ICA), and 2) combining all 65 participants and sessions in a single analysis (consortium ICA). In both cases, the spatial maps used were transformed to z-scores. The group DMN was automatically identified based on maximization of number of significantly correlated voxels at $z>4$ on the aggregate component in the four main DMN constituent regions; the result of automatic identification always agreed with expert operator inspection. Single-subject and session DMN were obtained by back-reconstruction using the GICA algorithm (Calhoun et al., 2001).

The aggregate DMN component from consortium ICA thresholded at $z>4$ was furthermore used as a mask for all connectivity strength measurements in the main nodes: medial prefrontal cortex (MFC), precuneus and posterior cingulate cortex (PCC), left and right parietal cortex (LPC and RPC, respectively). This choice was motivated by the need to obtain high-quality regions of interest representative of the cohort under study. Usage of data from another study for this purpose would have resulted in less reliable measurements due to the lower topographical overlap following different acquisition parameters. Besides, determining the regions for measuring in this manner did not result in circularity since the measurements were performed at the level of single participants and sessions, while the regions were defined only once, based on the aggregated group data.

For each participant and session, SBA yielded three variables, namely an average *z*-score in LPC, RPC and MFC; site and consortium ICA each yielded four variables, namely average *z*-score in PCC, LPC, RPC and MFC.

2. 6 Statistical analysis

Inter-site tSNR differences were separately tested for phantom and brain data using the Kruskall-Wallis test. Inter-site differences in DMN functional connectivity were similarly evaluated, separately for the four nodes (PCC, MFC, RPC and LPC) and three analysis methods (site ICA, consortium ICA and SBA). These comparisons were performed on test-session data. The test-retest reproducibility of these variables, expressed as absolute percent difference between test and retest, was also compared.

The consortium-level reproducibility of each variable was quantified via the intra-class correlation coefficient (ICC) for the degree of absolute agreement (Shrout and Fleiss, 1979) following rank-order data transformation (McGraw and Wong, 1996). Further, to assess voxel-wise spatial test-retest reproducibility the congruence between test-retest DMN spatial distribution was measured using the Jaccard similarity coefficient, setting $z>2$ as voxel thresholds for SBA and ICA.

All correlations were evaluated non-parametrically using Spearman's rank correlation coefficient and the significance level was set to $p<0.05$ for all tests.

3 Results

The interval between test and retest scans ranged between 7-69 days, with a median of 14 days. The following protocol deviations were detected: 1) Site 01 scanned one participant with

incorrect TR (2.3 s) in both test and retest sessions, and one participant with different TRs between sessions (test: 2.3 s, retest: 2.7 s); 2) Site 11 scanned two participants (4 and 5) with inconsistent voxel size across (test: 3.5x3.5x4.0, retest: 3x3x3 mm^3); 3) Site 13 consistently used an in-plane voxel size of 3.37x3.37 mm^2 and sequential ascending slice order; 4) Sites 7 and 8 performed interleaved ascending acquisitions with a pitch of 6 six slices (instead of 2). We took a conservative approach and did not reject the above datasets, with consideration to the fact that such protocol deviations are common in multicentric studies and may only bias the results towards worse reproducibility.

Given that site ICA with 20 components yielded substantially lower connectivity and reproducibility scores than site ICA with 10 components (Table 2), for brevity we did not take that analysis further and only report findings from ICA with 10 components. With this setting, the DMN was reliably extracted: beyond the four main nodes only small and weaker clusters of correlated activity were observed bilaterally in the medial and lateral temporal lobe, and in the middle frontal gyrus (Supplementary Figure 2). Similarly, given that PCC SBA with the seed mask derived from the consortium ICA yielded slightly worse results compared to using the seed from the previous study (Table 3), that analysis was not taken further; corresponding spatial maps are, however, visible in Supplementary Figure 3.

3.1 Phantom and brain tSNR, head movement

Significant differences in phantom tSNR were detected across sites (Figure 2a; $\chi^2=70$, $p<0.001$), following a pattern clearly related to the number of receive coils and fat suppression method (Table 1). Across sites equipped with Siemens and Philips scanners, the tSNR was lowest where a birdcage coil was used (Site 01), comparable for sites using 8 or 12 receive channels and fat saturation pulses, and highest at the site equipped with a 20-channel receive coil (Site 05). The three GE sites (all using 8-channel coils) were associated with higher phantom tSNR, plausibly due

to usage of spectral-spatial RF pulses for water-only excitation; these pulses produce broader slice profiles and therefore yield slightly larger effective voxel volumes, as well as usage of a Fermi filter on the raw *k*-space data, reducing temporal noise (Glover et al., 2012).The corresponding phantom test-retest tSNR error was also significantly heterogeneous across MRI sites, but the inter-site differences followed a different distribution (Figure 2b; χ^2=91, p<0.001), The reason for this variability across sites in phantom tSNR reproducibility is unclear. Factors such as improper consistency of phantom positioning, phantom temperature instabilities or improper gradient pre-heating may have played a role.

Across sites, brain GM tSNR values were in the range 58-218, median 118, with significant inter-site differences (Figure 2c; χ^2=34, p=0.001) which were, at the level of site medians, positively correlated to phantom tSNR (Figure 2e; ρ=0.69, p=0.01), suggesting that they were driven primarily by acquisition settings rather than participant characteristics (Figure 3). The test-retest reproducibility error of brain tSNR was not significantly different across sites (Figure 2d; p=0.1), with a pooled site median 8%, a site median range 2-13%, and high reproducibility, with ICC=0.81 (C.I. 0.70-0.88).

Head movement, quantified as median frame-to-frame displacement, was constrained, having test session site median 0.07 mm, site median range 0.03-0.11 mm (Supplementary Figure 4); there was a significant difference in test-session movement across sites (χ^2=22, p=0.04), but test-retest movement reproducibility was similar (p=0.1).

3.2 Inter-site consistency of DMN functional connectivity measures

Compared to the substantial inter-site brain tSNR variability, DMN maps and associated regional functional connectivity measures derived from ICA and SBA appeared relatively consistent (Figures 3 and 4); similarly, site and consortium maps, as well as the two SBA maps

derived from the atlas and ICA seeds, appeared strongly consistent (Supplementary Figures 2 and 3).

For site ICA, there were significant inter-site effects in test-session ICA *z*-score for PCC (χ^2=41, p<0.001), MFC (χ^2=40, p<0.001), RPC (χ^2=35, p=0.001) and LPC (χ^2=46, p<0.001). These differences were substantially reduced for consortium ICA, which abolished MRI site effects for PCC and MFC (p=0.2 and 0.08 respectively) and yielded weaker effects for LPC (χ^2=37, p<0.001) and RPC (χ^2=23, p=0.03). While significant, these differences were overall quantitatively much smaller than those observed for tSNR. There were no significant MRI site effects for SBA z-scores for MFC or RPC (p=0.1 and 0.3 respectively), with only a weak effect for LPC (χ^2=25, p=0.01), which however was primarily due to larger intra-site variation than ICA, rather than small inter-site differences. Notably, site ICA, consortium ICA and SBA yielded regional connectivity measurements that were strongly correlated to one another (Supplementary Figure 5).

3.3 Inter-site consistency of reproducibility of DMN functional connectivity measures

Test-retest reproducibility, expressed as absolute percent magnitude error, was not different across sites for any of the measures under consideration (Figure 5; p≥0.07 for site ICA, p≥0.1 for consortium ICA and p≥0.06 for SBA). Accordingly, we did not find significant correlations between test-retest reproducibility and brain tSNR except for SBA z-score in LPR (ρ=-0.33, p=0.006).

However, Wilcoxon signed-rank tests revealed significant differences among data analysis methods, wherein SBA was associated with larger test-retest error (p≤0.001) and no consistent differences were found between site- and consortium ICA (Table 4). Across the four DMN nodes under consideration, the PCC was associated with best measurement reproducibility (≈4% test-retest error with both ICA methods).

Further analysis based on ICC measurement on pooled data from all sites confirmed these findings, with higher ICC for site ICA than SBA and intermediate results for consortium ICA (Table 5).

Spatial test-retest reproducibility assessed with the Jaccard index between test and retest also confirmed different performance of the three methods, with median 0.62, 0.60 and 0.43 for site ICA, consortium ICA and SBA respectively; SBA performed significantly worse than ICA ($p<0.001$), without a significant difference between the two ICA implementations ($p=0.8$). The Jaccard index was positively correlated with tSNR for site ($\rho=0.41$, $p<0.001$) and consortium ICA ($\rho=0.37$, $p=0.003$), and for SBA ($\rho=0.29$, $p=0.02$).

3.4. Other resting state network components

We additionally considered four components beyond the DMN which were identifiable on the consortium ICA10 spatial maps: left fronto-parietal, bilateral temporo-parietal, bilateral sensory-motor and visual/occipital (Figure 6). For these components, the mean z-score was calculated over all regions included in a mask generated by thresholding the aggregate component at $z>4$. For all of them except the last, significant inter-site differences were found ($p\leq0.01$), however the reproducibility was always consistent across sites (Table 6).

4 Discussion

The main study findings are as follows: 1) Despite careful harmonization of the fMRI acquisition protocol, strong phantom and brain tSNR differences remain across sites; 2) Regardless of inhomogeneous tSNR, the DMN is always detected, albeit with some significant differences in regional functional connectivity metrics across sites; 3) Regardless of the inhomogeneous tSNR, the relative test-retest reproducibility error of regional functional connectivity metrics is rather

consistent across sites; 4) Across the four main DMN nodes, the PCC shows strongest connectivity and lowest test-retest reproducibility error; 5) ICA appears to yield more reliable DMN connectivity measurements relative to SBA.

Temporal SNR: variable due to MRI system differences but reproducible

Across sites the median tSNR in phantom series was ≈200, in agreement with previous reports (Friedman and Glover, 2006b). Inter-site differences were primarily driven by the number of receive coils and fat suppression method. All sites using Siemens and Philips equipment used fat saturation pulses, and across these sites the effect of number of receive channels was well-evident: lowest tSNR for birdcage coil, similar tSNR for 8- and 12-channel coils, highest tSNR for 20 channels. The GE-equipped sites used spectral water-only excitation pulses, which knowingly yield broader slice profiles and consequently larger effective voxel size (Glover et al., 2012); the tSNR was therefore higher than for Philips and Siemens systems equipped with comparable 8-channel receive coils and having similar acquisition parameters.

Brain tSNR comparisons across studies are challenging due to influence of a multitude of factors, including MRI hardware (number of channels, RF noise factor etc.), pulse sequence settings, pre-processing pipeline and brain areas selected for measurement (Triantafyllou et al., 2005, Bellgowan et al., 2006, Triantafyllou et al., 2006, Triantafyllou et al., 2011). Here, we quantified tSNR at the end of the preprocessing pipeline and averaged it over the whole-brain gray matter mask; this yielded median tSNR≈120, which varied significantly across sites. The variability potentially resulted from the superposition of MRI system and random participant differences, exacerbated by the relatively small number of participants scanned at each site. To disentangle the two aspects, we correlated the brain and phantom measures and found a moderate positive correlation, which suggested that even in-vivo, inter-site tSNR differences are primarily driven by

MRI system parameters. Notably, across sites the tSNR reproducibility for brain data was good (median error 7%, ICC=0.81), in agreement with previous single-site reports (Huang et al., 2012).

Functional connectivity strength: relatively consistent despite heterogeneous tSNR

All analysis methods (PCC-SBA, site ICA and consortium ICA) revealed synchronized activity in the key DMN constituent regions (Raichle et al., 2001, Greicius et al., 2003, Buckner et al., 2008). While there were significant inter-site differences in regional connectivity measurements, particularly for site ICA and SBA, these appeared relatively constrained compared to the much larger tSNR discrepancy; furthermore, these differences were attenuated and in some cases not significant for consortium ICA.

In agreement with previous work, we found that although SBA- and ICA-derived estimates of regional connectivity are quantitatively different, they are strongly correlated (Van Dijk et al., 2010). These findings are in agreement with those of the largest multi-site study of resting state fMRI data, the 1000 Functional Connectomes Project (Biswal et al., 2010), wherein no harmonization was attempted (Biswal et al., 2010).

Test-retest reproducibility: consistent across sites but ICA is more reliable than SBA

Two previous studies have evaluated between-session reproducibility of DMN functional connectivity over scans performed several months apart and analyzed using PCC-SBA (Shehzad et al., 2009) and ICA (Zuo et al., 2010) on the same data (3.0T, 26 healthy young participants). While ICA and SBA results were not directly compared, in line with our findings, PCC measurements yielded highest test-retest reproducibility and the ICC was in the range 0.45-0.65. One difference with (Zuo et al., 2010) is that they used 20 ICA components, which caused DMN splitting and consequent ambiguity in DMN identification; here, decomposition in 10 components was found to consistently group the main DMN nodes in a single component, which was highly consistent across

sites and allowed reliable automatic detection (Rosazza and Minati, 2011); notably, this choice may not be ideal if the purpose is to reliably extract multiple components.

Our findings are also in line with those of Meindl et al. (2010), who investigated the test-retest reproducibility of DMN activation patterns (3.0T, 18 healthy young participants) using ICA and concluded that PCC measurements yield the highest inter-session agreement. Another study (Li et al., 2012) directly compared the test-retest reproducibility of SBA and ICA on the same dataset (3.0T, 32 healthy young participants, two scans 2 months apart). Compared to our findings, they reported similar ICC values for SBA but lower reproducibility for ICA; this difference is ascribed to preprocessing and analysis settings, given that Li et al. (2012) extracted the DMN with single-subject ICA concatenation (two sessions). Our Jaccard spatial overlap measurements are in line with those of a previous DMN reproducibility study (3.0 T, 6 healthy young participants), wherein moderate reproducibility, namely 45% spatial overlap, was reported (Van Dijk et al., 2010).

To the authors' knowledge, this is the first report suggesting that the test-retest reproducibility of ICA is superior to that of PCC-SBA. We tentatively interpret this finding as a consequence of the fact that ICA, being a data-driven method, is particularly effective at removing signal sources which are unrelated to DMN activity and negatively bias the test-retest reproducibility, such as variations in systemic physiological state (Beckmann et al., 2005). However any comparison between ICA and SBA techniques should be interpreted with caution, given that they are inherently different and affected by distinct factors. We further found that performing group ICA at site level yielded similar percent absolute error reproducibility but higher ICC than consortium-level analysis. The superiority of SBA scores obtained using a seed mask from a different, previous study is unexpected but is ascribed to the incomplete overlap between the two seed masks under consideration (Supplementary Figure 1); it is possible that in our consortium ICA-level analysis, the substantial heterogeneity among sites impaired the extraction of an accurate spatial map compared with the homogeneous sample (Rosazza and Minati, 2011).

Limitations and future directions

General study design limitations have already been discussed in recent morphometry (Jovicich et al., 2013) and diffusion (Jovicich et al., 2014) investigations from the same consortium. In particular, limitations with regards to the fact that different participants were studied at each site and that the number of participants (five) per site was rather low, making it difficult to disentangle scanner-driven inter-site differences and sampling effects. Another limitation is that the test-retest repeatability was estimated from two sessions only, which could lead to variability underestimation.

Recent studies on young healthy participants have demonstrated that DMN connectivity is sensitive to circadian rhythm, with gradually reduced synchronization from morning to afternoon (Blautzik et al., 2013, Hodkinson et al., 2014). This suggests that test-retest reproducibility could be higher following standardization of acquisition time, particularly in the morning when connectivity is strongest. We did not find a significant effect of acquisition time, but this may be due to limited power as only 10 participants underwent test and retest acquisitions at different times of the day (morning/afternoon).

The standardization of acquisition parameters unavoidably came with a cost, in that vendor-specific optimized techniques could not be applied. In particular, the fastest common acquisition rate allowing full brain coverage (TR=2.7s) was relatively low compared to recent studies which have shown that shorter TRs (<1s) offer improved sensitivity and specificity of functional connectivity characterization (Feinberg et al., 2010, Smith et al., 2012, Wang et al., 2013, Kalcher et al., 2014). At present, multiband acquisition sequences delivering such short TRs are not yet widely available, particularly at sites without vendor research agreements, hence multisite studies are difficult to realize with these techniques.

Due to its relevance to Alzheimer's disease, this study focused on assessing the reproducibility of DMN functional connectivity measurements (Buckner, 2013, Pievani et al., 2014); while it appears plausible that the results may generalize to other networks, this should be confirmed in future studies undertaken using this freely available dataset or other data. Our

preliminary data on other networks beyond the DMN suggest that this is the case. Further, we restricted analyses to SBA and ICA as these are the most common data analytic techniques in use to date for clinical studies, but graph-theoretical measurements of network architecture are also gaining ground, and may offer improved sensitivity to pathological change (Friston, 2011, van den Heuvel and Sporns, 2013, Sporns, 2014, Stam, 2014); future multisite studies should therefore also investigate the test-retest reproducibility of graph-derived metrics.

5 Conclusions

The test-retest reproducibility of DMN functional connectivity as measured by ICA and PCC-SBA is consistent across sites despite highly heterogeneous tSNR due to hardware and pulse sequence differences; furthermore, site-by-site and consortium ICA give more reliable measurements of DMN functional connectivity than SBA. These findings support consideration of resting-state fMRI as a functional biomarker in multicentric longitudinal studies.

6 Acknowledgements

PharmaCog is funded by the EU-FP7 for the Innovative Medicine Initiative (grant no. 115009). Authors are grateful to all members and collaborators of the PharmaCog project, and particularly to L. Venturi, G. Borsci, T. Günther and A. Monnet for their contribution in the start-up phase. LM was funded by Scienze Mente-Cervello (Rovereto, Italy).

Conflicts of interest

The authors have no conflicts of interests to declare.

Figure captions

Figure 1. Data preprocessing pipeline. See "Brain data pre-processing and tSNR measurement" for details.

Figure 2. Comparison of phantom and whole-brain gray-matter tSNR: inter-site differences in phantom tSNR (a), corresponding test-retest percent phantom tSNR reproducibility (b), and inter-site differences in whole-brain tSNR averaged over each subject's gray-matter mask (c), corresponding test-retest percent whole-brain gray-matter tSNR reproducibility (d), and correlation between phantom and brain tSNR at the level of site medians (e). Central mark: median, edges: 25%/75%, whiskers: range, isolated points: outliers.

Figure 3. Statistical parametric maps for tSNR (top), PCC-SBA (middle), and site ICA with 10 components (bottom). Group average maps are presented, and for PCC-SBA and site ICA a threshold of $z>1$ was applied for visualization purposes. Regions-of-interest for connectivity measurement were obtained by thresholding at $z>4$ the aggregate DMN component from consortium ICA (not shown for brevity, see Supplementary Figures 2 and 3). Central mark: median, edges: 25%/75%, whiskers: range, isolated points: outliers.

Figure 4. Regional functional connectivity measures for the DMN nodes (PCC: precuneus and posterior cingulate cortex; LPC: left parietal cortex; RPC: right parietal cortex; and MFC: medial frontal cortex) as determined using site ICA, consortium ICA and PCC-SBA. Central mark: median, edges: 25%/75%, whiskers: range, isolated points: outliers.

Figure 5. Test-retest reproducibility error, expressed as absolute percent difference, of regional functional connectivity measures for the DMN nodes (PCC: precuneus and posterior cingulate cortex; LPC: left parietal cortex; RPC: right parietal cortex; and MFC: medial frontal cortex), as determined using site ICA, consortium ICA and PCC-SBA. Central mark: median, edges: 25%/75%, whiskers: range, isolated points: outliers.

Figure 6. Neural activity components identified on the consortium ICA10: a) default-mode network, b) left fronto-parietal, c) bilateral temporo-parietal, d) bilateral sensory-motor and e) visual/occipital. Bar charts show corresponding mean z-scores (calculated over the entire components, masked thresholding the aggregate component at z>4) and test-retest error.

References

Abou-Elseoud A, Starck T, Remes J, Nikkinen J, Tervonen O, Kiviniemi V (2010) The effect of model order selection in group PICA. Hum Brain Mapp 31:1207-1216.

Andrews-Hanna JR, Snyder AZ, Vincent JL, Lustig C, Head D, Raichle ME, Buckner RL (2007) Disruption of large-scale brain systems in advanced aging. Neuron 56:924-935.

Anticevic A, Cole MW, Murray JD, Corlett PR, Wang XJ, Krystal JH (2012) The role of default network deactivation in cognition and disease. Trends Cogn Sci 16:584-592.

Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, Oates JA, Peck CC, Schooley RT, Spilker BA, Woodcock J, Zeger SL, Grp BDW (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol Ther 69:89-95.

Bassett DS, Bullmore ET (2009) Human brain networks in health and disease. Curr Opin Neurol 22:340-347.

Beckmann CF, DeLuca M, Devlin JT, Smith SM (2005) Investigations into resting-state connectivity using independent component analysis. Philos T Roy Soc B 360:1001-1013.

Bellgowan PSF, Bandettini PA, van Gelderen P, Martin A, Bodurka J (2006) Improved BOLD detection in the medial temporal region using parallel imaging and voxel volume reduction. Neuroimage 29:1244-1251.

Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V (2013) The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage 83:550-558.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar Mri. Magnet Reson Med 34:537-541.

Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM, Beckmann CF, Adelstein JS, Buckner RL, Colcombe S, Dogonowski AM, Ernst M, Fair D, Hampson M, Hoptman MJ,

Hyde JS, Kiviniemi VJ, Kotter R, Li SJ, Lin CP, Lowe MJ, Mackay C, Madden DJ, Madsen KH, Margulies DS, Mayberg HS, McMahon K, Monk CS, Mostofsky SH, Nagel BJ, Pekar JJ, Peltier SJ, Petersen SE, Riedl V, Rombouts SARB, Rypma B, Schlaggar BL, Schmidt S, Seidler RD, Siegle GJ, Sorg C, Teng GJ, Veijola J, Villringer A, Walter M, Wang LH, Weng XC, Whitfield-Gabrieli S, Williamson P, Windischberger C, Zang YF, Zhang HY, Castellanos FX, Milham MP (2010) Toward discovery science of human brain function. P Natl Acad Sci USA 107:4734-4739.

Blautzik J, Vetter C, Peres I, Gutyrchik E, Keeser D, Berman A, Kirsch V, Mueller S, Poppel E, Reiser M, Roenneberg T, Meindl T (2013) Classifying fMRI-derived resting-state connectivity patterns according to their daily rhythmicity. Neuroimage 71:298-306.

Braun U, Plichta MM, Esslinger C, Sauer C, Haddad L, Grimm O, Mier D, Mohnke S, Heinz A, Erk S, Walter H, Seiferth N, Kirsch P, Meyer-Lindenberg A (2012) Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. Neuroimage 59:1404-1412.

Buckner RL (2013) The brain's default network: origins and implications for the study of psychosis. Dialogues Clin Neurosci 15:351-358.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci 1124:1-38.

Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001) Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human Brain Mapping 13:43-53.

Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP (2013) Clinical applications of the functional connectome. Neuroimage 80:527-540.

Damoiseaux JS (2012) Resting-state fMRI as a biomarker for Alzheimer's disease? Alzheimers Res Ther 4:8.

Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5:e15710.

Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8:700-711.

Fransson P, Marrelec G (2008) The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: Evidence from a partial correlation network analysis. Neuroimage 42:1178-1184.

Friedman L, Glover GH (2006a) Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. Neuroimage 33:471-481.

Friedman L, Glover GH (2006b) Report on a multicenter fMRI quality assurance protocol. J Magn Reson Imaging 23:827-839.

Friston KJ (2011) Functional and effective connectivity: a review. Brain Connect 1:13-36.

Glover GH, Mueller BA, Turner JA, van Erp TG, Liu TT, Greve DN, Voyvodic JT, Rasmussen J, Brown GG, Keator DB, Calhoun VD, Lee HJ, Ford JM, Mathalon DH, Diaz M, O'Leary DS, Gadde S, Preda A, Lim KO, Wible CG, Stern HS, Belger A, McCarthy G, Ozyurt B, Potkin SG (2012) Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. J Magn Reson Imaging 36:39-54.

Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. P Natl Acad Sci USA 100:253-258.

Gusnard DA, Raichle ME (2001) Searching for a baseline: Functional imaging and the resting human brain. Nat Rev Neurosci 2:685-694.

Hodkinson DJ, O'Daly O, Zunszain PA, Pariante CM, Lazurenko V, Zelaya FO, Howard MA, Williams SCR (2014) Circadian and homeostatic modulation of functional connectivity and

regional cerebral blood flow in humans under normal entrained conditions. J Cerebr Blood F Met 34:1493-1499.

Huang L, Wang X, Baliki MN, Wang L, Apkarian AV, Parrish TB (2012) Reproducibility of structural, resting-state BOLD and DTI data between identical scanners. PLoS One 7:e47684.

Jaccard P (1912) The distribution of flora in the alpine zone. New Phytologist 11:37–50.

Jovicich J, Marizzoni M, Bosch B, Bartres-Faz D, Arnold J, Benninghoff J, Wiltfang J, Roccatagliata L, Picco A, Nobili F, Blin O, Bombois S, Lopes R, Bordet R, Chanoine V, Ranjeva JP, Didic M, Gros-Dagnac H, Payoux P, Zoccatelli G, Alessandrini F, Beltramello A, Bargallo N, Ferretti A, Caulo M, Aiello M, Ragucci M, Soricelli A, Salvadori N, Tarducci R, Floridi P, Tsolaki M, Constantinidis M, Drevelegas A, Rossini PM, Marra C, Otto J, Reiss-Zimmermann M, Hoffmann KT, Galluzzi S, Frisoni GB (2014) Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. Neuroimage 101:390-403.

Jovicich J, Marizzoni M, Sala-Llonch R, Bosch B, Bartres-Faz D, Arnold J, Benninghoff J, Wiltfang J, Roccatagliata L, Nobili F, Hensch T, Trankner A, Schonknecht P, Leroy M, Lopes R, Bordet R, Chanoine V, Ranjeva JP, Didic M, Gros-Dagnac H, Payoux P, Zoccatelli G, Alessandrini F, Beltramello A, Bargallo N, Blin O, Frisoni GB (2013) Brain morphometry reproducibility in multi-center 3T MRI studies: a comparison of cross-sectional and longitudinal segmentations. Neuroimage 83:472-484.

Kalcher K, Boubela RN, Huf W, Bartova L, Kronnerwetter C, Derntl B, Pezawas L, Filzmoser P, Nasel C, Moser E (2014) The Spectral Diversity of Resting-State Fluctuations in the Human Brain. PLoS One 9.

LaBar KS, Gitelman DR, Mesulam MM, Parrish TB (2001) Impact of signal-to-noise on functional MRI of the human amygdala. Neuroreport 12:3461-3464.

Li ZJ, Kadivar A, Pluta J, Dunlop J, Wang Z (2012) Test-retest stability analysis of resting brain activity revealed by blood oxygen level-dependent functional MRI. Journal of Magnetic Resonance Imaging 36:344-354.

Liao XH, Xia MR, Xu T, Dai ZJ, Cao XY, Niu HJ, Zuo XN, Zang YF, He Y (2013) Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. Neuroimage 83:969-982.

McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30-46.

Meindl T, Teipel S, Elmouden R, Mueller S, Koch W, Dietrich O, Coates U, Reiser M, Glaser C (2010) Test-Retest Reproducibility of the Default-Mode Network in Healthy Individuals. Human Brain Mapping 31:237-246.

Parrish TB, Gitelman DR, LaBar KS, Mesulam MM (2000) Impact of signal-to-noise on functional MRI. Magn Reson Med 44:925-932.

Persson J, Pudas S, Nilsson LG, Nyberg L (2014) Longitudinal assessment of default-mode brain function in aging. Neurobiol Aging 35:2107-2117.

Pievani M, Filippini N, van den Heuvel MP, Cappa SF, Frisoni GB (2014) Brain connectivity in neurodegenerative diseases--from phenotype to proteinopathy. Nat Rev Neurol 10:620-633.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. P Natl Acad Sci USA 98:676-682.

Rosazza C, Minati L (2011) Resting-state brain networks: literature review and clinical applications. Neurol Sci 32:773-785.

Rosazza C, Minati L, Ghielmetti F, Mandelli ML, Bruzzone MG (2012) Functional Connectivity during Resting-State Functional MR Imaging: Study of the Correspondence between Independent Component Analysis and Region-of-Interest-Based Methods. Am J Neuroradiol 33:180-187.

Shehzad Z, Kelly AMC, Reiss PT, Gee DG, Gotimer K, Uddin LQ, Lee SH, Margulies DS, Roy AK, Biswal BB, Petkova E, Castellanos FX, Milham MP (2009) The Resting Brain: Unconstrained yet Reliable. Cereb Cortex 19:2209-2229.

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420-428.

Smith SM, Miller KL, Moeller S, Xu JQ, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, Van Essen DC, Feinberg DA, Yacoub ES, Ugurbil K (2012) Temporally-independent functional modes of spontaneous brain activity. P Natl Acad Sci USA 109:3131-3136.

Spearman C (1904) The proof and measurement of association between two things. Int J Epidemiol 39:1137-1150.

Sporns O (2014) Contributions and challenges for network models in cognitive neuroscience. Nat Neurosci 17:652-660.

Stam CJ (2014) Modern network science of neurological disorders. Nat Rev Neurosci 15:683-695.

Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL (2005) Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. Neuroimage 26:243-250.

Triantafyllou C, Hoge RD, Wald LL (2006) Effect of spatial smoothing on physiological noise in high-resolution fMRI. Neuroimage 32:551-557.

Triantafyllou C, Polimeni JR, Wald LL (2011) Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. Neuroimage 55:597-606.

van den Heuvel MP, Sporns O (2013) Network hubs in the human brain. Trends in Cognitive Sciences 17:683-696.

Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL (2010) Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. Journal of Neurophysiology 103:297-321.

Van Horn JD, Toga AW (2009) Multisite neuroimaging trials. Curr Opin Neurol 22:370-378.

Wang X, Jiao Y, Tang T, Wang H, Lu Z (2013) Investigating Univariate Temporal Patterns for Intrinsic Connectivity Networks Based on Complexity and Low-Frequency Oscillation: A Test-Retest Reliability Study. Neuroscience 254:404-426.

Zuo XN, Kelly C, Adelstein JS, Klein DF, Castellanos FX, Milham MP (2010) Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach. Neuroimage 49:2163-2177.

Zuo XN, Xing XX (2014) Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. Neurosci Biobehav Rev 45:100-118.

Zuo XN, Xu T, Jiang L, Yang Z, Cao XY, He Y, Zang YF, Castellanos FX, Milham MP (2013) Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. Neuroimage 65:374-386.

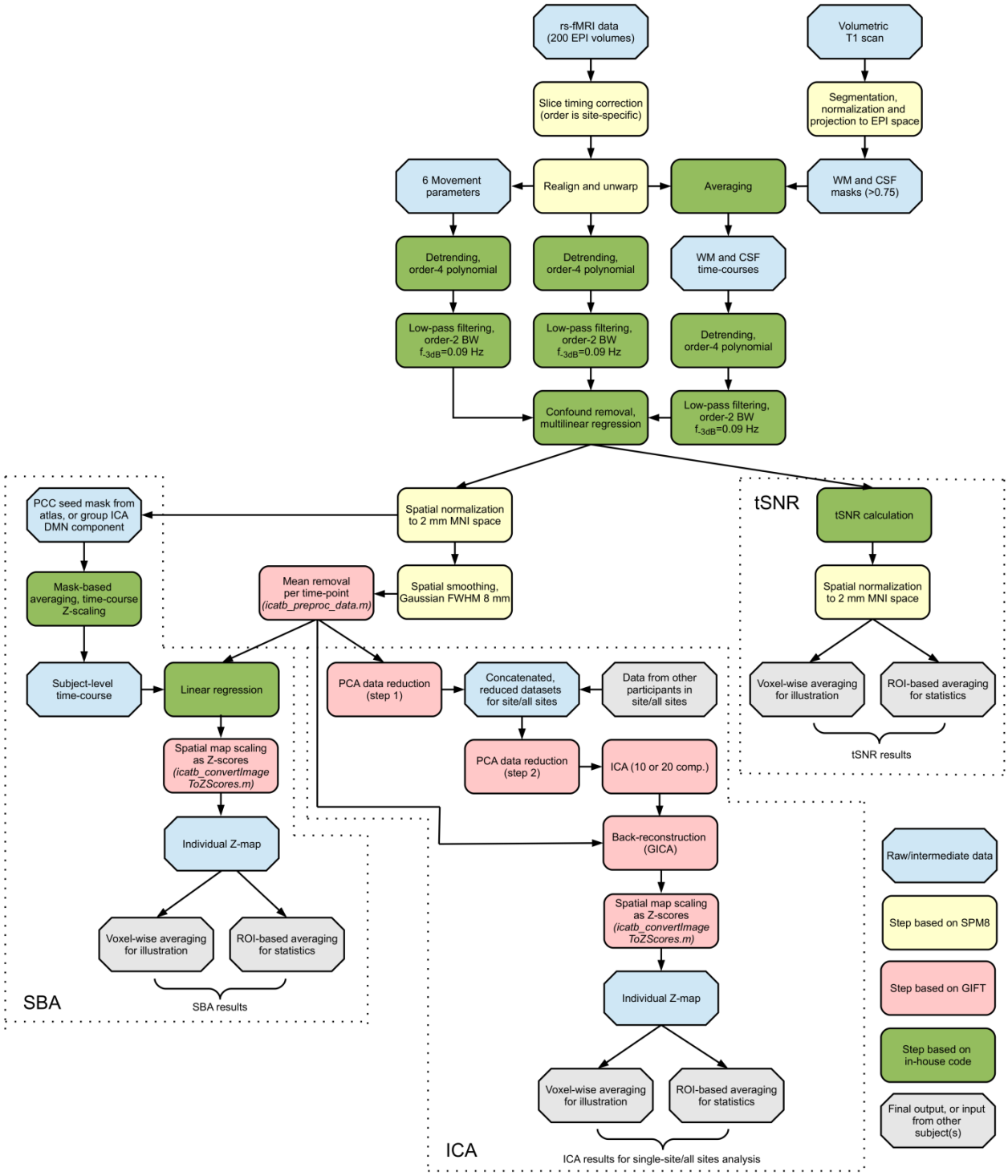

Figure 1

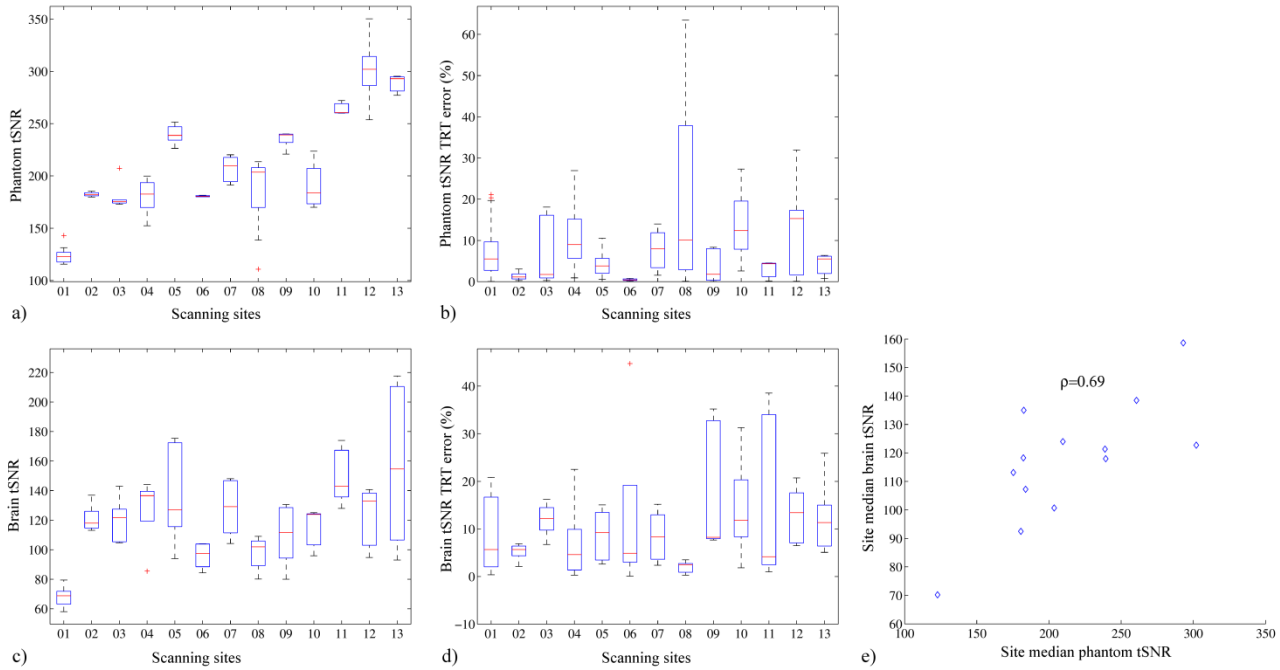

Figure 2

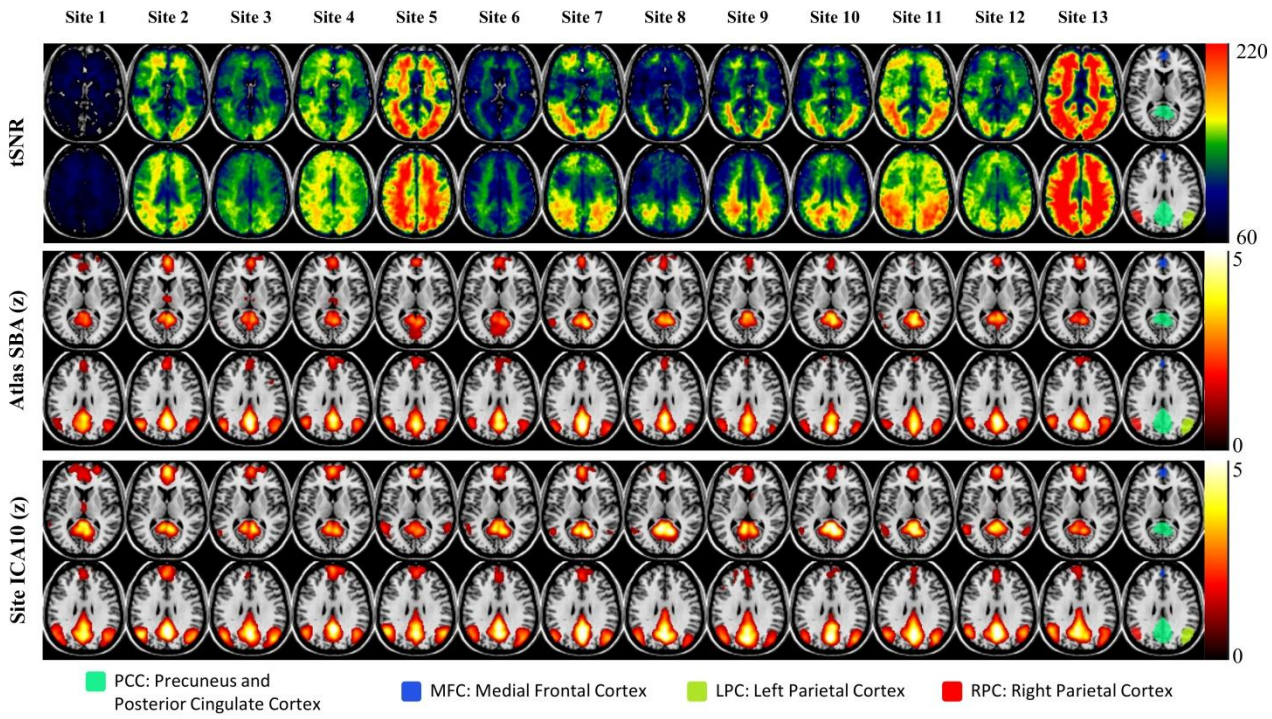

Figure 3

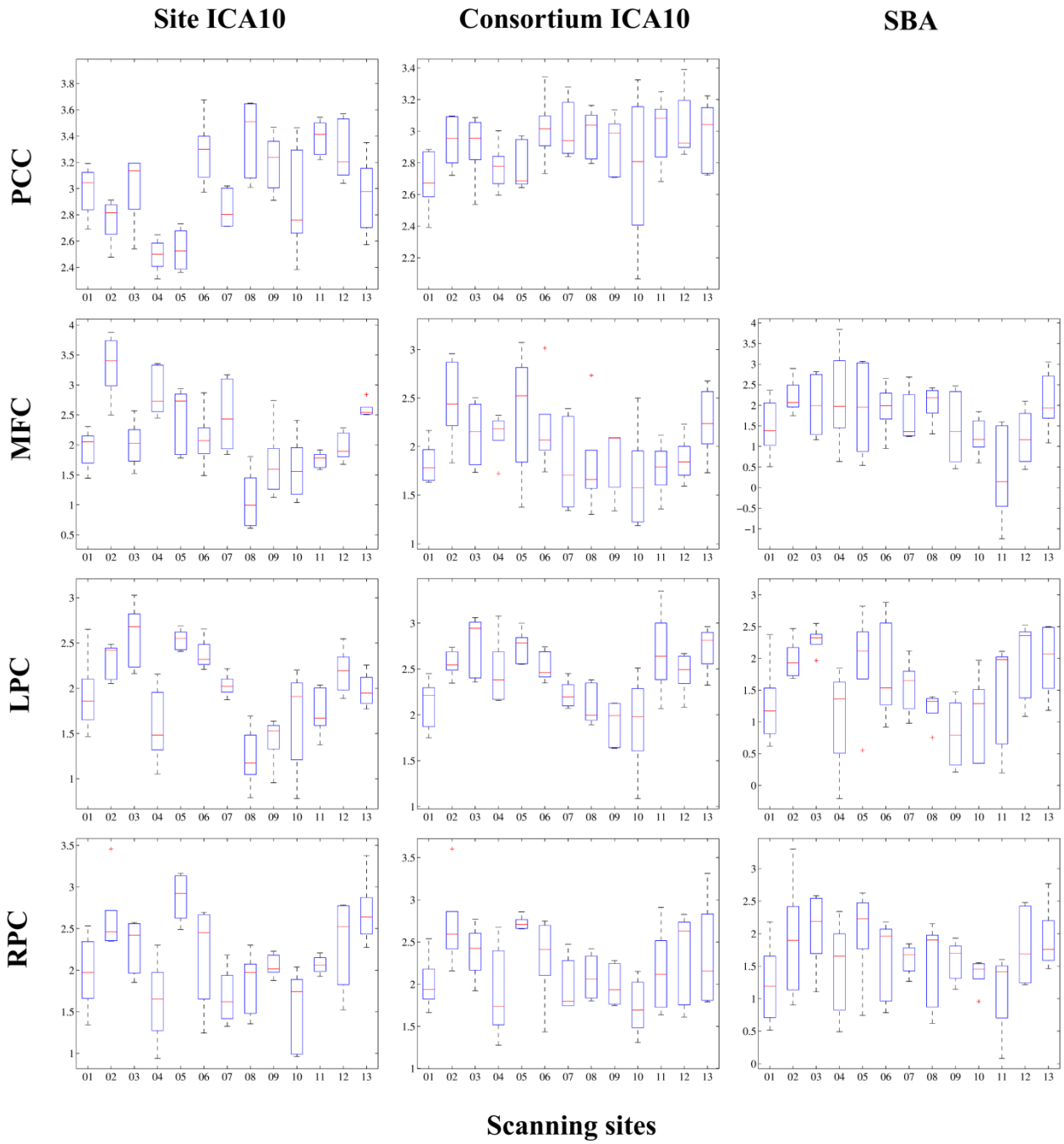

Figure 4

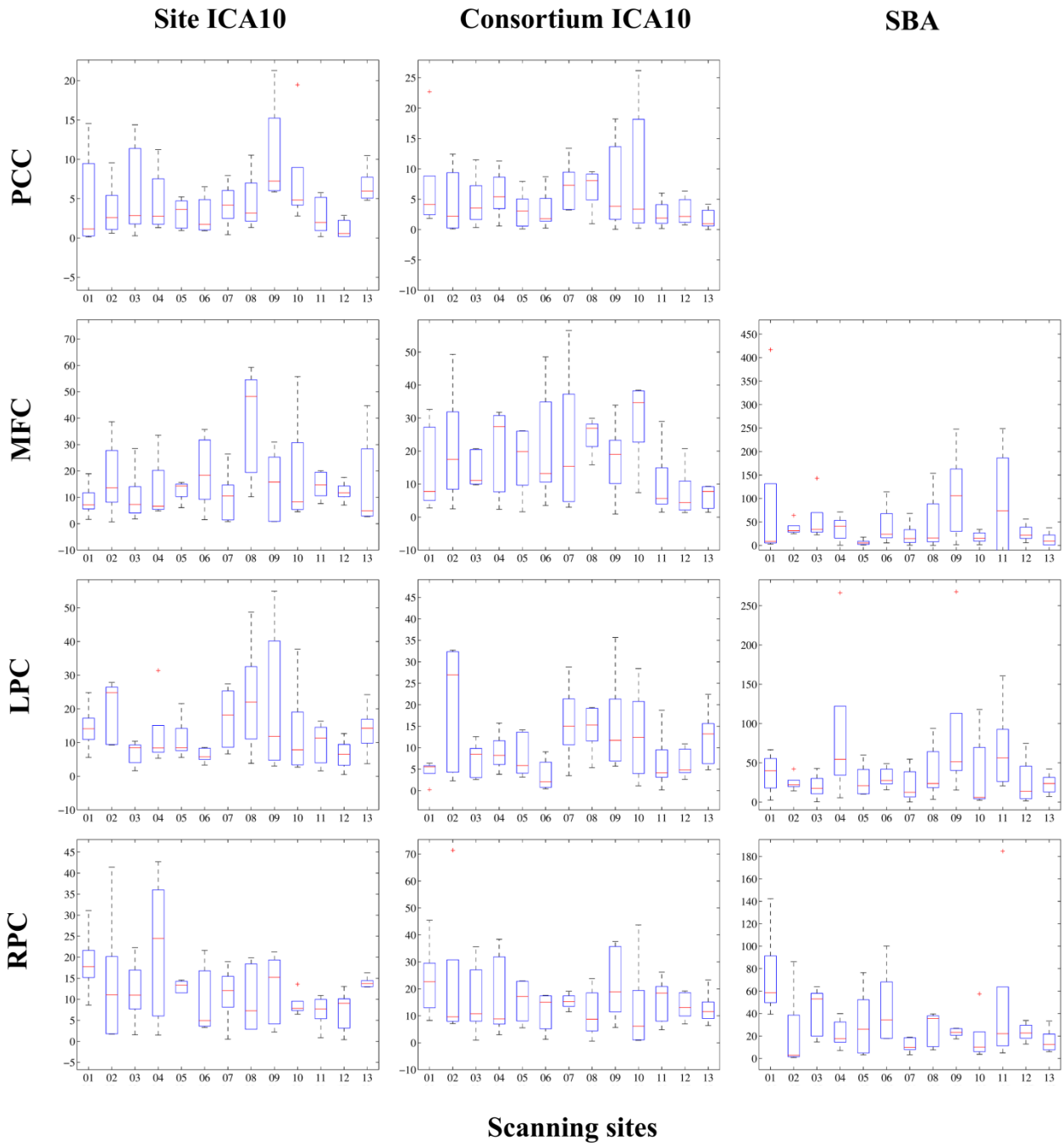

Figure 5

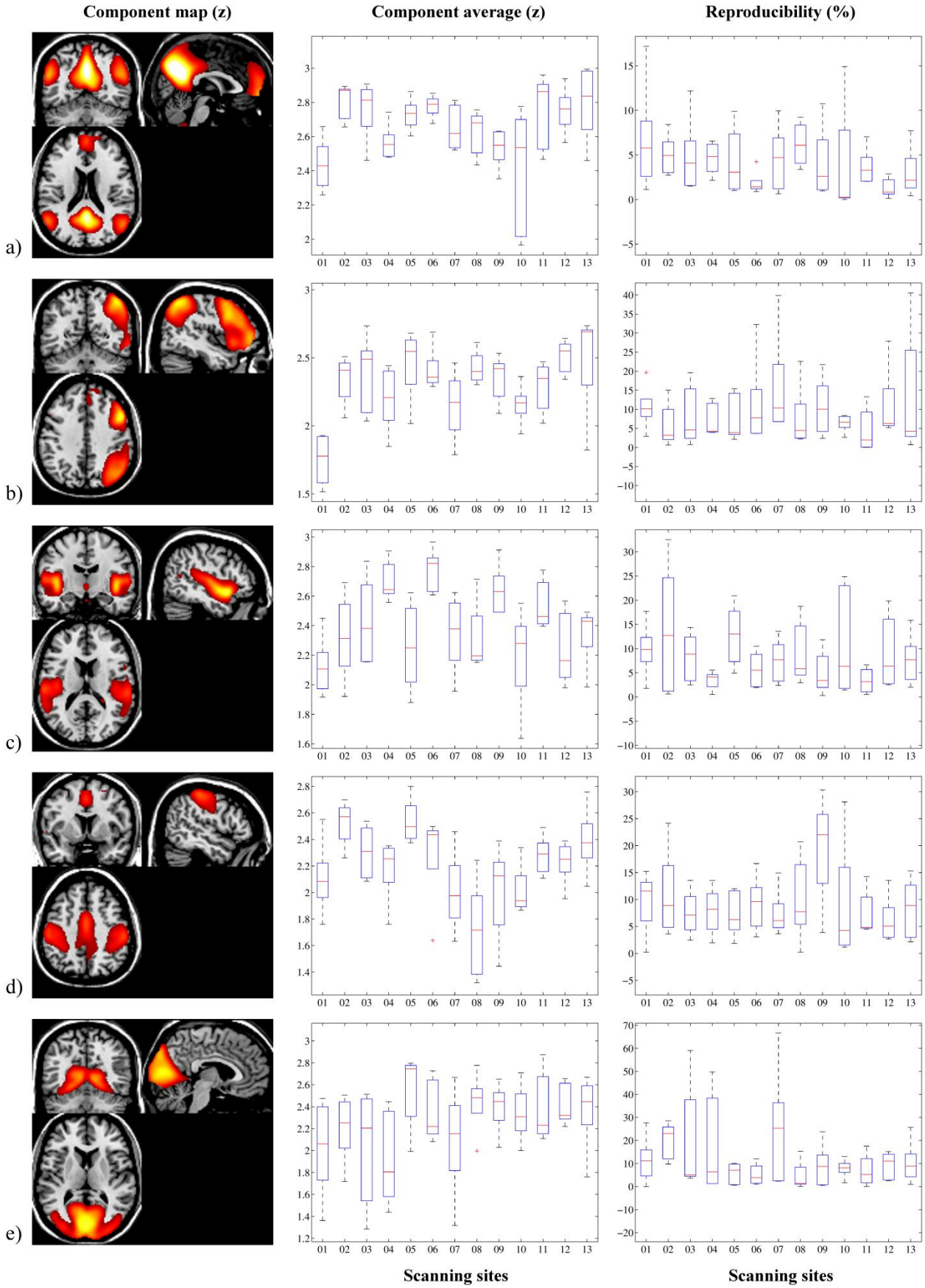

Figure 6

Table 1. Summary of demographic data, MR system specifications and EPI acquisition parameters across sites.

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12	Site 13
MRI site location	Verona	Barcelona	Leipzig	Marseille	Essen	Naples	Lille	Toulouse	Chieti	Perugia	Genoa	Thessaloniki	Amsterdam
Participant age: mean±stdev, (range) years	67.8±9.9 (26)	74.6±2.7 (6)	62.8±2.6 (6)	66.0±8.3 (20)	52.4±1.5 (3)	59.0±3.5 (9)	64.2±5.3 (13)	59.2±4.5 (12)	68.8±4.3 (11)	60.8±10.3(24)	58.2±2.2 (5)	56.6±5.5 (5)	62.8±8.2 (21)
Test-Retest Time interval (days)	28 ± 23	10 ± 3	13 ± 3	23 ± 22	11 ± 5	19±15	15 ± 11	14 ± 10	11±5	10±4	24 ± 17	32±8	11 ± 6
Sex, (females/N)	2/5(40%)	5/5(100%)	3/5(60%)	4/5(80%)	2/5(40%)	2/5 (40%)	3/5(60%)	3/5 (60%)	5/5 (100%)	3/5 (60%)	2/5 (40%)	3/5 (60%)	3/5 (60%)
Scanner make and model	Siemens Allegra	Siemens TrioTim	Siemens TrioTim	Siemens Verio	Siemens Skyra	Siemens Biograph mMR	Philips Achieva	Philips Achieva	Philips Achieva	Philips Achieva	GE HDxt	GE HDxt	GE Discovery MR750
MR system software version	VA25A	B17	B17	B17	D11	B18P	3.2.2	3.2.2	3.2.2	3.2.2	15 M4A	15 M4A	DV22.0[1] DV23.1
TX / RX coil	Birdcage	Body / 8-chan.	Body / 8-chan.	Body / 12-chan.	Body / 20-chan.	Body / 12-chan.	Body / 8-chan.	Body / 8-chan.	Body / 8-chan.	Body / 8-chan.	Body / 8-chan.	Body / 8-chan.	Body / 8-chan
Fat suppression	Fat Sat.	Fat Sat.	Fat Sat.	Fat Sat.	Fat Sat.	Fat Sat.	SPIR	SPIR	SPIR	SPIR	Water only	Water only	Water only
In-plane matrix	64x64	72x72	72x72	72x72	72x72	72x72	80x80	80x80	80x80	80x80	64x64	64x64	64x64
Slice order	2,4…40, 1,3…39	2,4…40, 1,3…39	2,4…40, 1,3…39	2,4…40, 1,3…39	2,4…40, 1,3…39	2,4…40, 1,3…39	1,7…37, 2,8…38 etc.	1,7…37, 2,8…38 etc.	1,3…39, 2,4…40	1,3…39, 2,4…40	1,3…39, 2,4…40	1,3…39, 2,4…40	1…40

[1]: A software upgrade took place during the study at site 13. DV22 was used for the first participant and DV23.1 for the rest.

Table 2. Comparison of default-mode network (DMN) functional connectivity measures extracted based on 10 vs. 20 ICA components, when combining data acquired at each single site separately (5 subjects, 2 sessions per analysis). Site ICA10 consistently yielded higher connectivity scores associated with lower test retest reproducibility (TRT) error. Superscript "*" denotes statistical significance of the Wilcoxon signed rank test; "n.s." denotes absence of a significant effect. PCC: precuneus and posterior cingulate cortex, LPC: left parietal cortex, RPC: right parietal cortex, and MFC: medial frontal cortex.

Measure		Site ICA10 (median)	Site ICA20 (median)	Rank-sum paired comparison	
				p-value	z-score
z-score	PCC	3.0	2.1	<0.001*	-6.8
	MFC	2.1	1.6	0.01*	-2.5
	LPC	2.0	1.7	<0.001*	-4.0
	RPC	2.1	1.7	<0.001*	-5.5
z-score TRT Reproducibility (median error)	PCC	4%	12%	<0.001*	-6.0
	MFC	12%	20%	0.001*	-3.2
	LPC	9%	13%	0.007*	-2.7
	RPC	12%	21%	<0.001*	-4.4
Spatial TRT reproducibility (Jaccard index)		0.62	0.47	<0.001*	-7.0

Table 3. Comparison of default-mode network (DMN) functional connectivity measures extracted by seed-based analysis (SBA) based on average time-course 1) from the posterior cingulate cortex (PCC) as derived from a previous study (40 healthy, young participants, 1.5 T field strength, threshold $t>2$, volume 27 ml), see text for description) or 2) from the PCC cluster yielded by GICA in Consortium ICA10 (threshold $z>6.6$, volume 28 ml). Corresponding masks are shown in Supplementary Figure 2. Limited differences were observed, but usage of the seed mask derived from the previous study yielded, for some regions, higher z-scores and lower test retest reproducibility (TRT) error. Superscript "*" denotes statistical significance of the Wilcoxon signed PCC rank test; "n.s." denotes absence of a significant effect. PCC: precuneus and posterior cingulate cortex, LPC: left parietal cortex, RPC: right parietal cortex, and MFC: medial frontal cortex.

DMN metrics		PCC SBA seed from previous study (median)	PCC SBA seed from Consortium ICA10 (median)	Rank sum paired comparison (previous study vs. GICA seed)	
				p-value	Z-score
Z-score	MFC	1.8	1.5	<0.001*	-6.2
	LPC	1.6	1.4	<0.001*	-3.5
	RPC	1.7	1.7	0.5	n.s.
Z-score TRT reproducibility (median error)	MFC	22%	32%	<0.001*	-4.2
	LPC	26%	34%	0.004*	-2.9
	RPC	22%	27%	0.1	n.s.
Spatial TRT reproducibility (Jaccard index)		0.43	0.43	0.9	n.s.

Table 4. Functional connectivity test-retest (TRT) reproducibility error estimated across the whole consortium at each one of the main default-mode network (DMN) nodes and for each analysis method used: site ICA10 (10 independent components, ICA performed at each site separately for 5 subjects, 2 sessions), consortium ICA10 (10 independent components, ICA performed once for all sites, 65 subjects, 2 sessions) and PCC SBA using a seed mask derived from a previous study (see text for details). Reproducibility errors of SBA were significantly higher than those obtained by either ICA method. Superscript "*" denotes statistical significance of the Wilcoxon signed rank test; "n.s." denotes absence of a significant effect. PCC: precuneus and posterior cingulate cortex, LPC: left parietal cortex, RPC: right parietal cortex, and MFC: medial frontal cortex.

DMN Node	Median TRT error in functional connectivity			Rank paired differences (p-value)	
	Site ICA10	Consortium ICA10	PCC SBA (previous study seed)	PCC SBA (previous study seed)	Site vs. Consortium ICA10
PCC	4%	3%	--.	n.s.	--
MFC	12%	13%	22%	n.s.	<0.001*
LPC	9%	7%	26%	0.04*	<0.001*
RPC	12%	13%	22%	0.008*	<0.001*

Table 5. Test-retest reproducibility of functional connectivity estimated from ICC analysis for the main default-mode network (DMN) nodes and each analysis method used: site ICA10, consortium ICA10 and PCC SBA using a seed from a previous study. PCC: precuneus and posterior cingulate cortex, LPC: left parietal cortex, RPC: right parietal cortex, and MFC: medial frontal cortex.

DMN Nodes	Rank ICC with upper and lower bound confidence intervals		
	Site ICA10	Consortium ICA10	PCC SBA (seed from previous study)
PCC	0.85 (C.I. 0.76-0.90)	0.73 (C.I. 0.60-0.83)	--
MFC	0.82 (C.I. 0.72-0.88)	0.67 (C.I. 0.51-0.79)	0.66 (C.I. 0.50-0.78)
LPC	0.84 (C.I. 0.74-0.90)	0.75 (C.I. 0.63-0.84)	0.59 (C.I. 0.40-0.73)
RPC	0.82 (C.I. 0.72-0.89)	0.61 (C.I. 0.43-0.74)	0.54 (C.I. 0.34-0.69)

Table 6. Comparison of ICA components beyond the default-mode network (DMN). See Figure 6 for associated charts and spatial maps. Superscript "*" denotes statistical significance of the Kruskal-Wallis test; "reprod." denotes reproducibility; "n.s." denotes absence of a significant effect.

ICA component	Mean z-score (entire component)	Inter-site effect (mean z-score)	z-score (TRT reprod. median error)	Inter-site effect (reprod.)
Default-mode network (DMN)	2.7	χ^2=26, p=0.01*	3%	n.s.
Left fronto-parietal (FP)	2.4	χ^2=26, p=0.01*	7%	n.s.
Temporo-parietal (TP)	2.4	χ^2=30, p=0.002*	6%	n.s.
Sensori-motor (SM)	2.3	χ^2=30, p=0.003*	8%	n.s.
Visual (VS)	2.3	n.s.	8%	n.s.

Highlights (5 max, 85 characters with spaces per highlight)

- We implement a multi-site 3T MRI protocol for resting state fMRI in 13 sites
- We acquire across-session test-retest (TRT) data on 64 healthy elderly participants
- Despite harmonization strong phantom and brain tSNR differences remain across sites
- TRT error of regional DMN functional connectivity is consistent across sites
- ICA yields more reliable DMN connectivity measurements relative to SBA