Multivariate consistency of resting-state fMRI connectivity maps acquired on a single individual over 2.5 years, 13 sites and 3 vendors

AmanPreet Badhwar, MSc, PhD[1,2], Yannik Collin-Verreault, BSc[1], Pierre Orban, PhD[3,2], Sebastian Urchs, MSc[1,4], Isabelle Chouinard, MRT[5], Jacob Vogel, BSc[4], Olivier Potvin, PhD[5], Simon Duchesne, PhD[5,6], Pierre Bellec, PhD[1,2]

[1] Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Canada; [2] Université de Montréal, Montréal, Canada; [3] Centre de recherche de l'Institut universitaire en santé mentale de Montréal, Montréal, Canada; [4] McGill University, Montréal, Canada; [5]Centre CERVO, Quebec City Mental Health Institute, Quebec, Canada; [6] Department of Radiology, Faculty of Medicine, Université Laval, Quebec, Canada

Corresponding Author

Dr. AmanPreet Badhwar
Centre de Recherche, Institut Universitaire de Gériatrie de Montréal
Université de Montréal
Montréal, QC, Canada H3W 1W5
Tel: 514-340-3540 ext. 3367
Fax: 514-340-2802
Email: amanpreet.badhwar@criugm.qc.ca

Running title: Multivariate consistency of long-term multisite resting-state fMRI

Abstract

Studies using resting-state functional magnetic resonance imaging (rsfMRI) are increasingly collecting data at multiple sites in order to speed up recruitment or increase sample size. Multisite studies potentially introduce systematic biases in connectivity measures across sites, which may negatively impact the detection of clinical effects. Long-term multisite biases (i.e. over several years) are still poorly understood. The main objective of this study was to assess the long-term consistency of rsfMRI multisite connectivity measures derived from the Canadian Dementia Imaging Protocol (CDIP, www.cdip-pcid.ca).

Nine to ten minutes of functional BOLD images were acquired from an adult cognitively healthy volunteer scanned repeatedly at 13 Canadian sites on three scanner makes (General Electric, Philips and Siemens) over the course of 2.5 years. RsfMRI connectivity maps were extracted for each session in seven canonical functional networks. The reliability (spatial Pearson's correlation) of maps was about 0.6, with moderate effects (up to 0.2) of scanner makes and sites. The time elapsed between scans had a negligible effect on the consistency of connectivity maps. To assess the utility of such measures in machine learning models, we pooled the long-term longitudinal data with a single-site, short-term (1 month) data sample acquired on 26 subjects (10 scans per subject), called HNU1. Using randomly selected pairs of scans from each subject, we quantified the ability of a data-driven unsupervised cluster analysis to match the two scans. In this "fingerprinting" experiment, we found that scans from the Canadian subject could be matched with high accuracy (>85% for some networks), and fell in the range of accuracies observed for the HNU1 subjects.

Overall, these results support the feasibility of multivariate, machine learning analysis of rsfMRI measures in a multisite study that extends for several years, even with fairly short (approximately ten minutes) time series.

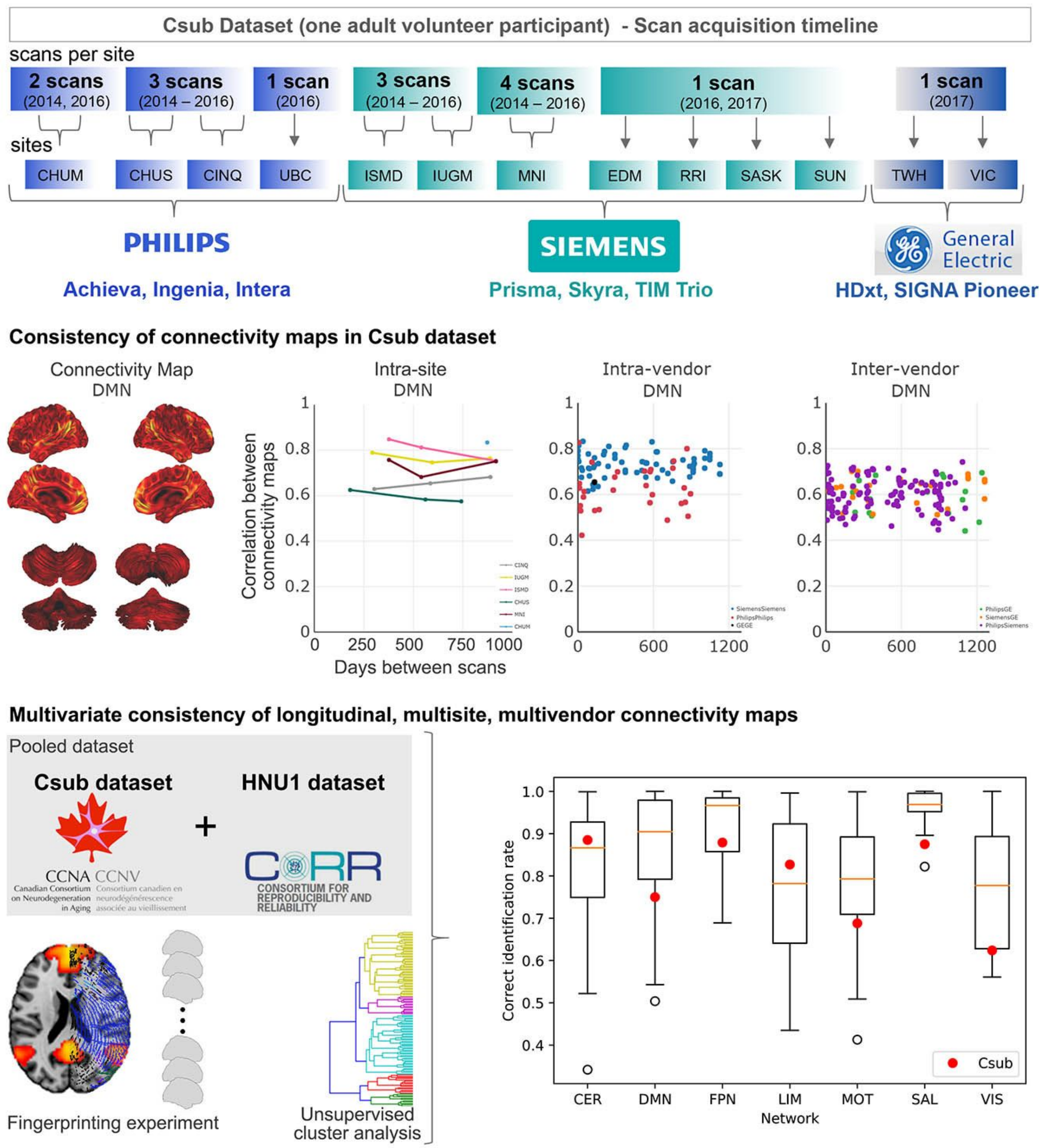
Keywords

resting-state fMRI, consistency, multisite, longitudinal, fingerprinting

HIGHLIGHTS

- Assessed the long-term (2.5 yrs) consistency of multisite rsfMRI connectivity maps
- Time elapsed between scans had negligible effect on consistency of connectivity maps
- Consistency of intra-subject long-term multisite maps was greater than inter-subject
- Individual connectivity fingerprint is preserved across sites, vendors and time

Graphical Abstract



1. INTRODUCTION

Paradigm-free ("resting-state") functional MRI (rsfMRI) can be used to detect spatially distributed functional connectivity networks in health and their alterations in disease (Badhwar et al. 2017; Matthews and Hampshire 2016). Neuroimaging phenotypes, however, typically exhibit considerable heterogeneity between patients (Dong et al. 2017; Drysdale et al. 2017) and large datasets are needed to achieve sufficient statistical power for reliable detection (Button et al. 2013). Such large patient cohorts frequently surpass the recruitment capacity of single clinical centres. A number of initiatives have pooled multisite data on normal or patient cohorts, such as attention deficit hyperactivity disorder (Brown et al. 2012), autism spectrum disorder (Di Martino et al. 2017; Nielsen et al. 2013), diabetes (Saggar et al. 2017), depression (Drysdale et al. 2017), schizophrenia (Skåtun et al. 2017; Cheng et al. 2015), Alzheimer's disease (Alzheimer's Disease Neuroimaging Initiative[1]), population imaging genetics (UK Biobank[2]) and normal brain development (Adolescent Brain Cognitive Development Study or ABCD[3]). However, it is still unclear to what degree the use of multiple scanners introduces additional variance in neuroimaging measures, especially for studies that will collect data for several years. Here we report on the multisite fMRI protocol of a large multisite initiative, lead by the Canadian Consortium for Neurodegeneration in Aging (CCNA[4]), which is recruiting 1600 individuals on the spectrum of age-related dementias

[1] http://adni.loni.usc.edu/about/
[2] http://www.ukbiobank.ac.uk/
[3] https://addictionresearch.nih.gov/abcd-study
[4] http://ccna-ccnv.ca/

over the course of 4 years, as well as 660 cognitively normal individuals, at over 30 sites. The CCNA initiative relies on harmonized acquisition parameters set forth in the *Canadian Dementia Imaging Protocol* (CDIP[5]), implementing a series of site qualification, quality control, and assurance procedures.

The main objectives of the present study were to assess the inter-site and longitudinal consistency of rsfMRI measures derived from a single traveling Canadian subject (Csub) scanned repeatedly at several CCNA sites using CDIP. An additional objective was to assess whether the inter-site variance would interfere with a simple machine learning task. We concentrated on fingerprinting (Finn et al. 2015), i.e. identifying paired scans from the same subject in a large multisubject dataset. For this purpose, we pooled the Csub scans with a public dataset featuring multiple retest scans per individual.

The impact of multisite acquisition on rsfMRI connectivity has recently gained attention in a series of studies. Using retrospective rsfMRI data, Yan and colleagues first demonstrated the existence of systematic variations in resting-state connectivity across 18 sites, by contrasting average connectivity patterns of independent groups composed of (mostly) young healthy subjects (Yan et al. 2013). Dansereau et al. further extended this analysis on a subset of eight sites with 3T scanners, showing that average group resting-state network maps could be consistently observed using parcel-based functional connectomes (Dansereau et al. 2017). The authors also

[5] www.cdip-pcid.ca

reported widespread site effects, present across all resting-state networks, with some sites associated with larger bias than others.

One major limitation of both analyses (Dansereau et al. 2017; Yan et al. 2013) was the reliance on retrospective data, which was a mix of different acquisition parameters (e.g. voxel size, repetition time) as well as scanner make and field strength, all of which may exaggerate the amplitude of site differences. This limitation was addressed by Jovicich and colleagues (Jovicich et al. 2016), who investigated rsfMRI data collected across 13 sites using a harmonized acquisition protocol at 3T on three scanner platforms: Siemens Medical Systems (Siemens), Philips Healthcare (Philips) and General Electric Healthcare (GE). Even with a harmonized protocol, the authors observed significant differences across sites using cross-sectional human volunteer data comprised of independent groups of five participants scanned at each site. This result may partly reflect significant inter-site differences in temporal signal-to-noise ratio (tSNR) maps, observed both on geometric phantoms and volunteer data. This study also scanned each cohort twice over two weeks (median), and demonstrated that retest reliability of connectivity maps was comparable across sites for the major resting-state networks.

A second limitation shared by the multisite rsfMRI studies reviewed thus far was that different participants were recruited at each site, thereby keeping open the possibility that site effects simply reflected differences in participant characteristics. Only a single cohort experiment can unambiguously capture inter-site differences, with the same individual(s) being scanned repeatedly at each site. Two recent studies

implemented such an approach. First, Noble and colleagues (Noble et al. 2017) acquired rsfMRI data on eight participants in two scan sessions separated by 24 hours, and repeated this experiment at eight different sites (all 3T; Siemens and GE scanners) with the same participant cohort and a harmonized protocol. They found that inter-site differences in connectivity measures were substantially explained by the variability of individual within-site measures, especially for short (5 min) acquisition times (Noble et al. 2017). This conclusion applied to individual region-to-region connectivity, yet multivariate reliability measures from whole-brain connectivity maps were more reliable both within and across sites. In a second, independent study, An and colleagues (An et al. 2017) acquired rsfMRI data on 10 traveling participants in two sessions separated by 30 mins on three scanners, also using a harmonized protocol. Unlike Noble and colleagues, they collected data on Philips scanners in addition to Siemens and GE, but only had one scanner per vendor. Short-term reliability was shown to be better on GE, relative to Siemens and Philips scanners. Unlike Jovicich and colleagues (Jovicich et al. 2016), the authors did not find differences in tSNR ratio across scanner vendors, and there was good reliability of whole brain connectivity maps between scanner vendors.

A major question left open in the literature is how multisite acquisitions impact rsfMRI over the long periods of time (years) needed to complete enrolment in large studies such as CCNA. To address this question, we wanted to move beyond traditional measures of consistency for repeated measures (such as intra-class correlation (Fleiss and Cohen 1973)) because fMRI connectivity maps are high dimensional, multivariate measures, and their primary use case in many instances (e.g. CCNA), will be to serve

as features for machine learning prognostic models. Categorical guidelines for interpretation of consistency measures may not translate well in this multivariate predictive context (Koo and Li 2016; Cicchetti and Sparrow 1981). Moreover, since functional connectivity maps can act as a 'fingerprint' for accurately identifying subjects within a large group (Finn et al. 2015), we selected the accuracy of fingerprinting as a benchmark to assess the consistency of longitudinal fMRI scans. The specific aims and hypotheses of this study were as follows:

a) Evaluate the effect of scanning site, scanner vendor and time elapsed between scans (up to 2.5 years) on the consistency of connectivity maps. Based on the previous literature reviewed above, we hypothesized moderate vendor and site effects. We also hypothesized only a small effect of time: although age effects are detectable in adults, 2 years remain within the error margin of age prediction based on fMRI connectivity (Li, Satterthwaite, and Fan 2018).

b) Contrast intra-subject consistency of connectivity maps in a multisite, longitudinal data against intra-subject and inter-subject consistency for a short-term, single site data. Our hypothesis was that site effects would reduce intra-subject consistency (inter-site), but that it would remain higher than inter-subject consistency (intra-site). The rationale for this hypothesis is that the inter-subject differences in brain connectivity are large compared to longitudinal intra-subject differences (Gratton et al. 2018).

c) Evaluate whether the identity of a subject can be reliably identified in the context of multisite, long-term longitudinal data, when pooled with within-site, short-term

longitudinal data (fingerprint experiment). In the absence of prior literature, we did not have a specific hypothesis for this aim.

We implemented three experiments to address these aims. We first tested the effect of time, vendor and site using a linear regression analysis of network connectivity maps generated from fMRI data collected on Csub, scanned over 2.5 years at 13 sites using CDIP implemented on one of 3 scanner vendors (Sections 3.1 and 3.2, aim a). We then compared the intra-subject consistency in the Csub data with both intra-subject and inter-subject consistency in a public sample[6] released as part of the Consortium On Resting-state Reproducibility (CORR) (Zuo et al. 2014) and comprised of 30 healthy adults scanned 10 times each over one month (Section 3.3, aim b). Finally, we performed a fingerprinting experiment using scans from the pooled dataset (Section 3.4, aim c), i.e. attempting to match the identity of participants based on pairs of resting-state connectivity maps.

2. METHODS

2.1 Canadian subject dataset (Csub)

All brain imaging data were acquired from a volunteer Csub: a healthy male with no history of (a) psychiatric and/or neurological illnesses; (b) psychoactive drug usage; or (c) contraindications to MRI. Csub was 42 years old at the start of data collection (2014). In total, the participant underwent 25 scanning sessions at 13 CCNA imaging sites; using scanners from three manufacturers (Philips, Siemens and GE), see Table 1.

[6] http://fcon_1000.projects.nitrc.org/indi/CoRR/html/hnu_1.html

Center	Abbr.	Scanner Vendor	2014	2015	2016	2017
Centre Hospitalier de l'Université de Montréal	CHUM	Philips, Achieva	x		x	
Centre Hospitalier Universitaire de Sherbrooke	CHUS	Philips, Ingenia	x	x	x	
Le Consortium d'Imagerie en Neurosciences et Santé Mentale de Québec	CINQ	Philips, Achieva	x	x	x	
University of British Columbia	UBC	Philips, Intera			x	
Institut Universitaire en Santé Mentale Douglas	ISMD	Siemens, Magnetom TIM Trio	x	x	x	
Institut Universitaire de Gériatrie de Montréal, Montréal	IUGM	Siemens, Magnetom TIM Trio	x	x	x	
Montreal Neurological Institute	MNI	Siemens, Magnetom TIM Trio	x	xx	x	
Peter S. Allen MR Research Centre	EDM	Siemens, Prisma			x	
Robarts Research Institute	RRI	Siemens, Prisma				x
Royal University Hospital	SASK	Siemens, Skyra				x
Sunnybrook Health Sciences Center/Sunnybrook Research Institute	SUN	Siemens, Prisma				x
Toronto Western Hospital	TWH	GE, HDxt				x
West Coast Medical Imaging	VIC	GE, SIGNA Pioneer				x

Table 1: Demographics

The letter “x” in columns 2014-2017 indicate acquisition of rsfMRI and structural scans at the corresponding year.

The data was acquired as part of an ongoing effort to monitor the quality and comparability of MRI data collected across the CCNA imaging network. The schedule of visits did not follow a strict design, with an approximate goal of one visit a year, starting at site qualification. Informed consent was obtained from the subject for the overall study and before every scan session. Due to the multisite nature of the study, ethics approval was obtained from the institutional review board of each participating institution prior to scanning.

Anatomical scans included 3D isotropic T1-weighted (T1w) imaging for assessing fine anatomical detail with high resolution (voxel size = 1.0 x 1.0 x 1.0 mm^3) and acceleration factor of 2 (Siemens: MP-RAGE; GE: FSPGR; Philips: T1-TFE). Functional T2*-weighted images were obtained using a blood-oxygen-level-dependent (BOLD) sensitive single-shot echo-planar (EPI) sequence. Additional scan parameters are provided in Table 2. During the rsfMRI acquisitions, no specific cognitive tasks were performed, and the participant was instructed to keep his eyes open. No camera or physiological recordings were captured, as these equipment were not available at every site. It should be noted that we excluded the second MNI 2015 intra-session scan (Table 1) from our study, since it was the only intra-session scan acquired. Thereby, data from the remaining 24 scans was used in the study.

					Bold EPI				
Site	Field Strength T	Voxel Size mm^3	FA °	Matrix size	TE ms	TR ms	Volumes #	Scan Time	Slice Order
CHUM	3.0	3.5 iso	70	64x64	30	2110	300	10.55	sequential ascending
CHUS	3.0	3.5 iso	70	64x64	30	2110	300	10.55	sequential ascending
CINQ	3.0	3.5 iso	70	64x64	30	2110	300	10.55	sequential ascending
UBC	3.0	3.5 iso	70	64x64	30	2110	300	10.55	interleaved ascending
ISMD	3.0	3.5 iso	70	64x64	30	2110	300	10.55	sequential descending
IUGM	3.0	3.5 iso	70	64x64	30	2110	300	10.55	interleaved ascending
MNI	3.0	3.5 iso	70	64x64	30	2110	300	10.55	interleaved ascending
EDM	3.0	3.5 iso	70	64x64	30	2130	300	10.65	interleaved ascending
RRI	3.0	3.5 iso	70	64x64	30	2130	250	8.88	interleaved ascending
SASK	3.0	3.5 iso	70	64x64	30	2140	250	8.92	sequential descending
SUN	3.0	3.5 iso	70	64x64	30	2130	250	8.88	interleaved ascending
TWH	3.0	3.5 iso	70	64x64	30	2130	278	9.87	interleaved ascending
VIC	3.0	3.5 iso	70	64x64	30	2500	250	10.42	interleaved ascending

Table 2: Scan parameters.

rsfMRI BOLD EPI scan parameters: Scan time is provided in minutes and seconds. Abbreviations: FA, flip angle; ms, millisecond; TE, echo time; TR, repetition time; iso, isometric;

T, Tesla; Sites: CHUM, Centre Hospitalier Universitaire de Sherbrooke; CHUS, Centre Hospitalier Universitaire de Sherbrooke; CINQ, Le Consortium d'Imagerie en Neurosciences et Santé Mentale de Québec; ISDM, Institut Universitaire en Santé Mentale Douglas; IUGM, Institut Universitaire de Gériatrie de Montréal; MNI, Montreal Neurological Institute; EDM, Peter S. Allen MR Research Centre; SASK, Royal University Hospital; TWH, Toronto Western Hospital; RRI, Robarts Research Institute; SUN, Sunnybrook Health Sciences Center/Sunnybrook Research Institute; VIC, West Coast Medical Imaging.

2.2 Hangzhou Normal University dataset (HNU1)

The HNU1 dataset[7] includes 30 healthy adults 20-30 years of age (mean age 24.4 years), each receiving 10 scans across one month (one scan every three days) on a single 3T GE Discovery MR750 scanner (Zuo et al. 2014). Anatomical scans included 3D isotropic T1w imaging (voxel size = 1.0 x 1.0 x 1.0 mm^3 and acceleration factor of 2). Functional T2*-weighted images were obtained using a 10 min BOLD-sensitive single-shot EPI sequence: 3.4 mm isotropic voxels, 90 deg flip angle, 64 x 64 matrix size, 30 ms TE, 2000 ms TR, 300 time points, interleaved ascending slice acquisition order[8]. During rsfMRI scanning, subjects were presented with a fixation cross and were instructed to keep their eyes open, relax and move as little as possible while observing the fixation cross. Subjects were also instructed not to engage in breath counting or meditation.

2.3 Computational environment

The datasets were preprocessed and analyzed using the NeuroImaging Analysis Kit, version 1.1.3 (NIAK-COG[9], (Bellec et al. 2011)), executed within an Ubuntu 16.0.4

[7] http://fcon_1000.projects.nitrc.org/indi/CoRR/html/hnu_1.html
[8] http://fcon_1000.projects.nitrc.org/indi/CoRR/html/_static/scan_parameters/HNU_1_scantable.pdf
[9] https://hub.docker.com/r/simexp/niak-cog/

Singularity[10] container , running GNU Octave[11] version 4.2.1, and the MINC toolkit[12] version 1.9.15. We also used three Jupyter notebooks that can be executed online via the binder platform[13], and run in a docker container[14] built from a public configuration file. Python packages used in the Jupyter notebooks include Numpy (Oliphant 2006), Pandas (McKinney and Others 2010), Matplotlib (Hunter 2007), Scikit-learn (Pedregosa et al. 2011), SciPy (Jones, Oliphant, and Peterson 2016), Seaborn[15] and StatsModel (Seabold and Perktold 2010). Interactive plots were generated using Plotly[16].

2.4 Pre-processing of MRI data

With the exception of the T1w scan from UBC, which failed quality control, we averaged all T1w scans (*n*=23) from Csub, following iterative alignment using rigid-body registration. Initially, a brain mask was extracted from a single arbitrary T1w scan (CHUM 2014) using the CIVET pipeline (Ad-Dab'bagh et al. 2006). Remaining T1w scans were then aligned to this reference scan, followed by an averaging of the aligned T1w scans. The averaged image served as the reference scan for the second iteration of alignment and averaging. We performed three such iterations to obtain our final T1w average.

Data from each fMRI scan was corrected for slice timing by linear temporal interpolation. The first three volumes of each fMRI run were discarded to allow the magnetization to reach steady-state. Rigid-body motion was estimated for each time

---

[10] https://github.com/SIMEXP/niak/releases/download/v1.1.3/niak_singularity.tgz
[11] http://www.gnu.org/software/octave/
[12] http://www.bic.mni.mcgill.ca/ServicesSoftware/ServicesSoft-wareMincToolKit
[13] http://mybinder.org
[14] https://mybinder.org/v2/gh/SIMEXP/cdip_human_phantom/master
[15] https://zenodo.org/record/1313201#.XAGdbVZKgWo
[16] https://plot.ly/

frame, intra-run and inter-run, using CHUM 2014 as an arbitrary reference. Each session was comprised of one fMRI run (see Table 1).

The rigid-body, fMRI-to-T1w and T1w-to-stereotaxic transformations were all combined and used to transform the fMRI images into MNI space at a 3 mm isotropic sampling. The following nuisance covariates were regressed out from the fMRI time series: slow time drifts (basis of discrete cosines with a 0.01 Hz high-pass cut-off), average signals in conservative masks of the white matter and the lateral ventricles, as well as the first principal components (accounting for 95% variance) of the six rigid-body motion parameters and their squares (Giove et al. 2009). The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the preprocessing pipeline can be found on the NIAK website[17].

The HNU1 dataset was preprocessed using the NIAK pipeline (Bellec et al. 2011), using the first available structural scan as reference for alignment in stereotaxic space. Four individuals demonstrated subpar alignment of the brain around the meninges. These individuals were excluded from the dataset, and all additional analysis were carried out on the remaining 26 individuals.

2.4 Quality control

To minimize artifacts due to excessive motion, all time frames showing a displacement > 0.5 mm were removed (Power et al. 2012). The number of censored volumes ranged from 0 to 118 time frames, with at least 179 volumes left to generate a connectivity map

[17] http://niak.simexp-lab.org/pipe_preprocessing.html

(see Supplementary Material Table S1). No scan was excluded due to excessive motion.

2.5 Connectivity maps

Using NIAK's connectome pipeline[18], for each rsfMRI scan (from both Csub and HNU1 datasets), we computed voxel-wise connectivity maps associated with each of the seven network templates extracted from a group-level functional brain atlas. The atlas, the Multiresolution Intrinsic Segmentation Template (MIST), was generated from 200 healthy subjects (Urchs et al. 2017). The MIST atlas consists of nine functional parcellations capturing successively finer levels of spatial detail, of which we used parcellations from resolution seven, consisting of seven commonly used large-scale network: cerebellar (CER), default-mode (DMN), frontoparietal (FPN), limbic (LIM), motor (MOT), salience (SAL), and visual (VIS). A network connectivity map was obtained per network by computing the Pearson's correlations between the average time course within the network template and the time course of every voxel in the brain.

2.6 Consistency of individual rsfMRI measures within/between sites

For each of the seven rsfMRI networks, a scan by scan similarity (Pearson's correlation) matrix was generated to summarize the consistency of connectivity maps across the 24 scans in the Csub dataset. A series of explanatory variables were assembled for a general linear model (GLM) analysis: (1) time between scans, expressed in years and corrected to a zero mean; (2) dummy variables encoding intra-vendor comparisons (three covariates: GE, Siemens and Philips); (3) dummy variables encoding intra-site

[18] http://niak.simexp-lab.org/pipe_connectome.html

comparisons (six covariates: CHUM, CHUS, CINQ, ISDM, IUGM, MNI; the other sites did not have multiple retest data available). An intercept was also added to the model, which, in combination with the other covariates of the model, captured the average consistency for comparisons across sites from different vendors. A linear mixture of the explanatory variables were adjusted on the inter-scan consistency measures (dependent variable) using ordinary least squares, for each network separately. For each network, we tested the significance of the effect of inter-vendor (t-test), intra-vendor (*F* test testing the combined effect of the three intra-vendor covariates), intra-site (*F* test testing the combined effect of the six intra-site covariates) and time (*t*-test). We adjusted the significance level of *p* values for multiple comparisons across networks using a Bonferroni correction (family-wise error 0.05, significance threshold $p < 0.0071$). We also examined the effects of each individual variable to assess which vendors and sites drove the significance of tests.

2.7 Consistency of rsfMRI measures within/between subjects

For each HNU1 subject and each network, we computed the average (and standard deviation) for the intra-subject consistency for all pairs with the 10 available scans. We also computed the average inter-subject consistency across all scans from different subjects, both within HNU1, and between HNU1 and Csub. To further statistically compare these consistency values, we implemented a single GLM analysis, in which the dependent variable was the measures of inter-scan consistency, and the explanatory variables included a series of dummy variables encoding separately the intra-subject comparisons (Csub and 26 HNU1 subjects), one dummy variable encoding

comparisons between Csub and HNU1 subjects, and one dummy variable encoding inter-subject comparisons in HNU1. A series of t-test were derived from the following contrasts: intra-subject in Csub vs intra-subject in HNU1, intra-subject in Csub vs inter-subject in HNU1, inter-subject Csub/HNU1 vs inter-subject HNU1, and intra-subject HNU1 vs inter-subject HNU1.

2.8 Fingerprinting of HNU1 participants and Csub

We assessed the ability of a simple data-driven cluster analysis to recover the identity of subjects based on connectivity maps of a single network, mixing the Csub single subject with the HNU1 subjects. A fingerprinting experiment consisted of the following steps: (i) for each subject, randomly select two scans out of all available scans (at least 10); (ii) assemble an inter-scan similarity matrix, using only the selected scans for all subjects; (iii) apply a hierarchical clustering on this similarity matrix (Ward's criterion); (iv) group the scans into as many clusters as there are subjects, based on the hierarchy; (v) for each subject, the fingerprinting experiment is considered successful if the two scans of this subject constitute a cluster.

The fingerprinting experiment was repeated B=1000 times using random scan selections, and independently for each network. The average accuracy of the fingerprinting for a given subject and network was derived as the proportion of successful fingerprinting experiments.

2.9 Data Records

Scripts used in this study are available on Github[19], as well as archived on zenodo[20]. The three Jupyter notebooks used in this study (graphs, stats_repro and stats-fingerprinting) can be executed online via the binder platform[21]. We have also made available on Github and zenodo two interactive dashboards containing (1) connectivity maps used to assess long-term consistency in Csub rsfMRI measures, and (2) connectivity maps from Csub and HNU1 datasets. Provided in each dashboard is the individual connectivity map per network, the average connectivity map per network, and the MIST parcellation at scale 7.

3. RESULTS

3.1. Connectivity maps

The key regions of all 7 networks were clearly identifiable at every session, as illustrated for DMN connectivity (Figure 1). Random fluctuations were also apparent, sometimes with strong shifts in global connectivity values (see for example IUGM vs EDM in Figure 1). We then quantified the consistency of rsfMRI maps generated at different sessions using the Pearson's correlation coefficient as a measure of spatial similarity. We selected this measure as it is invariant to the shifts in mean and variance we noted above.

[19] https://github.com/SIMEXP/cdip_human_phantom
[20] https://doi.org/10.5281/zenodo.1979984
[21] https://mybinder.org/v2/gh/SIMEXP/cdip_human_phantom/master

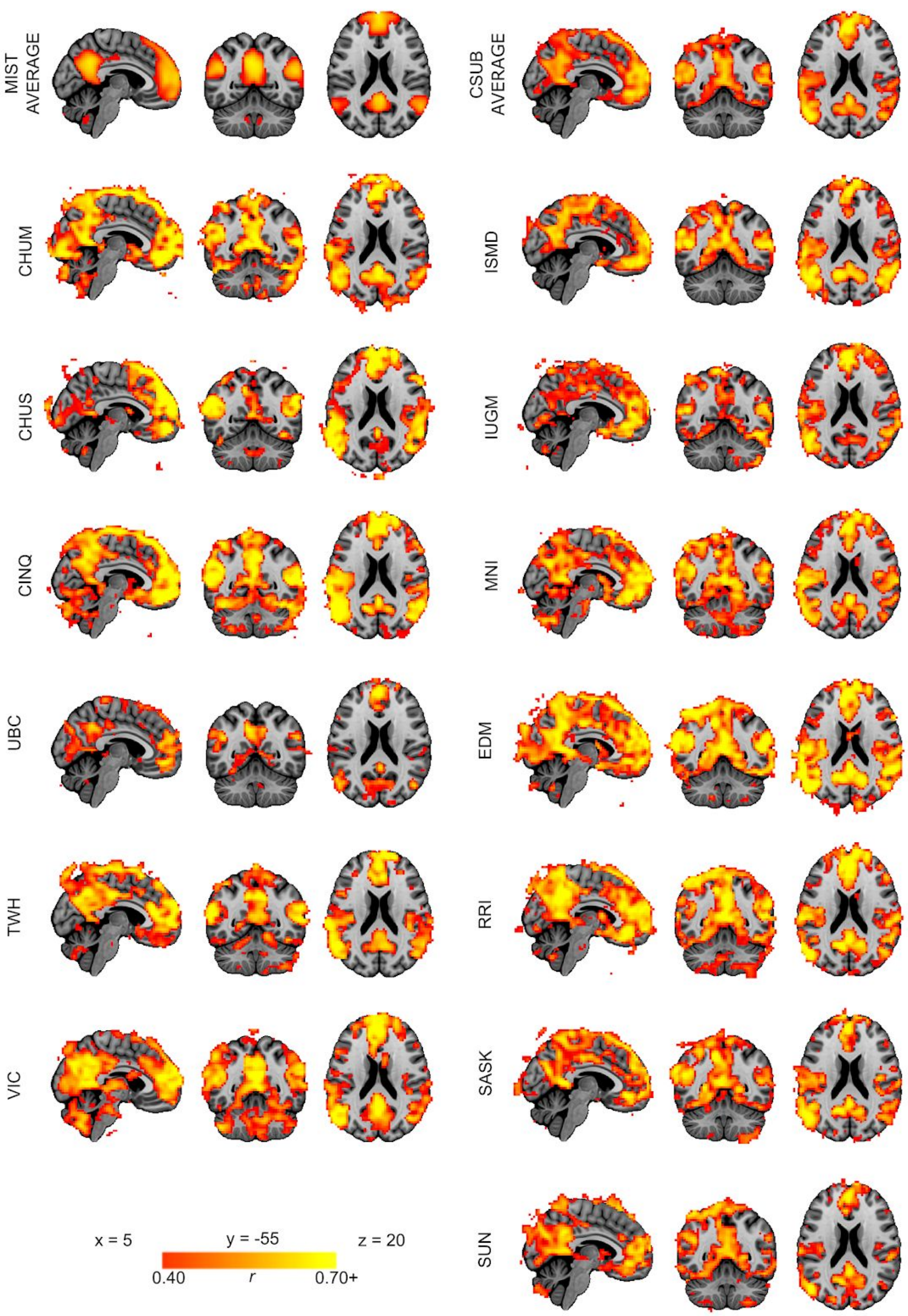


Figure 1: Default-mode network connectivity. The connectivity map for the DMN of each site is included. For sites with multiple sessions, only the most recent is shown. Brighter colours (orange-yellow) in the connectivity maps indicate stronger connectivity strength (higher Pearson

r correlation). Maps are superimposed onto the anatomic International Consortium for Brain Mapping (ICBM) 152 template. Top left map is a DMN average of the MIST parcellation dataset. Top right map is a DMN average of all 13 sites from the Csub dataset. Abbreviations: CHUM, Centre Hospitalier Universitaire de Sherbrooke; CHUS, Centre Hospitalier Universitaire de Sherbrooke; CINQ, Le Consortium d'Imagerie en Neurosciences et Santé Mentale de Québec; ISDM, Institut Universitaire en Santé Mentale Douglas; IUGM, Institut Universitaire de Gériatrie de Montréal; MNI, Montreal Neurological Institute; EDM, Peter S. Allen MR Research Centre; SASK, Royal University Hospital; TWH, Toronto Western Hospital; RRI, Robarts Research Institute; SUN, Sunnybrook Health Sciences Center/Sunnybrook Research Institute; VIC, West Coast Medical Imaging.

3.2 Consistency of individual fMRI measures within/between sites

The consistency of maps generated with rsfMRI data acquired on scanners from different vendors ranged from 0.57 ± 0.01 (limbic network) to 0.64 ± 0.01 (visual network), see Table 3. There was no substantial (or significant) effect of time between scanning sessions on consistency between maps. The estimated yearly rate of change in consistency (measured on a spatial correlation scale from -1 to 1) ranged from 4.22e-4, $p = 0.26$ (limbic network) to -3.98e-3, $p = 0.28$ (visual network), see Table 3, Figure 2.

There was a significant effect of vendors in two networks (cerebellar and frontoparietal), with trends ($p<0.05$ uncorrected) in three others (limbic, salience and visual), see Table 3 and Figure 3. This suggests that, for these networks, inter-site, intra-vendor consistency was significantly different from inter-site, inter-vendor consistency. The effect was driven by Siemens scanners, with markedly higher consistency in all seven networks (See Supplementary Material Table S2).

Intra-site consistency was also significantly higher than inter-site, inter-vendor consistency in two networks (cerebellar and visual), see Table 3, with trends (p<0.05 uncorrected) in three others (frontoparietal, motor and salience). The intra-site effects on consistency were highly heterogeneous, with some sites showing very small effects (e.g. cerebellar network: CINQ, difference in consistency 0.0037, p=0.942), while others were markedly different (e.g. cerebellar network: MNI, difference in consistency 0.1536, p=0.002), see Supplementary Material Table S2.

Network	Average Inter-site Inter-vendor consistency		Additional Intra-vendor effect			Additional Intra-site effect			Additional Effect of time	
	mean	p value	F value	dof	P value	F value	dof	p value	delta corr/yr	p value
CER	0.60	*<1e-15	7.41	265	*6.93 e-3	7.72	265	*5.85 e-3	-7.87 e-4	8.77 e-1
DMN	0.62	*<1e-15	2.75	265	9.82 e-2	1.81	265	1.79 e-1	-1.24 e-3	7.98 e-1
FPN	0.62	*<1e-15	9.92	265	*1.82 e-3	5.53	265	1.95 e-2	-2.81 e-3	5.53 e-1
LIM	0.57	*<1e-15	6.65	265	1.04 e-2	2.63	265	1.08 e-1	4.22 e-4	2.61 e-1
MOT	0.63	*<1e-15	1.60	265	2.07 e-1	4.55	265	3.38 e-2	-3.34 e-3	4.86 e-1
SAL	0.63	*<1e-15	5.73	265	1.74 e-2	5.82	265	1.66 e-2	-3.31 e-3	4.26 e-1
VIS	0.64	*<1e-15	5.26	265	2.26 e-2	7.74	265	*5.80 e-3	-3.98 e-3	2.87 e-1

Table 3: Consistency of rsfMRI connectivity measures

(*) indicates family-wise error < 0.05 (Bonferroni corrected for multiple comparisons across networks, adjusted threshold p<0.0071). Abbreviations: Networks: CER, cerebellar; DMN, default mode; FPN, frontoparietal; LIM, limbic; MOT, motor; SAL, salience; VIS, visual. Sites: CHUM, Centre Hospitalier Universitaire de Sherbrooke; CHUS, Centre Hospitalier Universitaire de Sherbrooke; CINQ, Le Consortium d'Imagerie en Neurosciences et Santé Mentale de Québec; ISDM, Institut Universitaire en Santé Mentale Douglas; IUGM, Institut Universitaire de Gériatrie de Montréal; MNI, Montreal Neurological Institute.

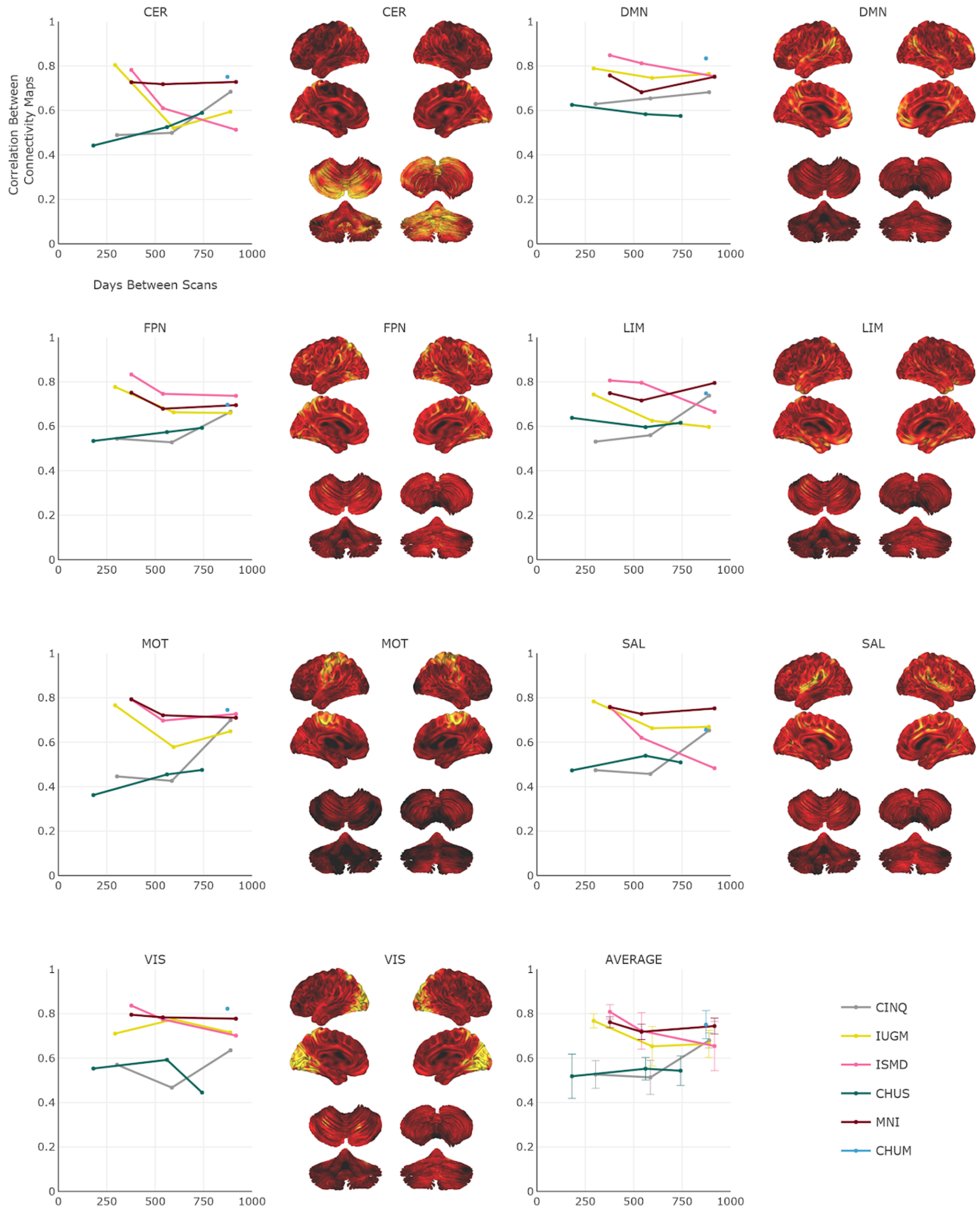


Figure 2: Intra-site consistency over time

Per network, the intra-site consistency (Pearson's correlation r) over time (range: 0 to 917 days) for the six sites. Longitudinal data are presented as line plots on the left, and the average

connectivity map is provided on the right. Brighter colours (orange-yellow) in the connectivity maps indicate stronger connectivity strength. Maps are superimposed onto the anatomic International Consortium for Brain Mapping (ICBM) 152 template and the SPM2_MNI aligned cerebellum surface (Van Essen et al. 2004). The average consistency across all networks are also shown. Abbreviations: Networks: CER, cerebellar; DMN, default mode; FPN, frontoparietal; LIM, limbic; MOT, motor; SAL, salience; VIS, visual. Sites: CHUM, Centre Hospitalier Universitaire de Sherbrooke; CHUS, Centre Hospitalier Universitaire de Sherbrooke; CINQ, Le Consortium d'Imagerie en Neurosciences et Santé Mentale de Québec; ISDM, Institut Universitaire en Santé Mentale Douglas; IUGM,Institut Universitaire de Gériatrie de Montréal; MNI, Montreal Neurological Institute. Note: Interactive graphs are provided in the "graphs" Jupyter notebook.

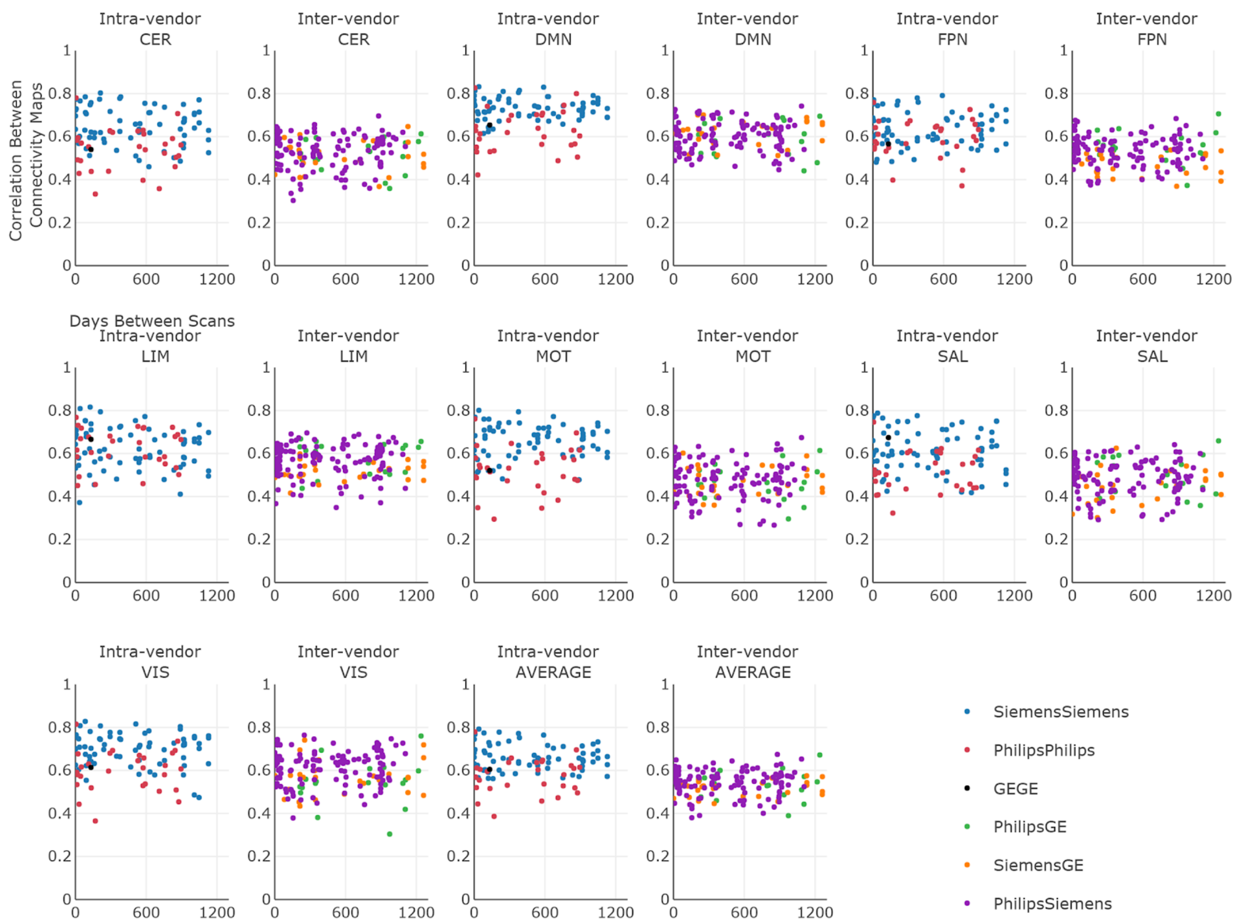


Figure 3: Inter-site, intra-vendor and inter-vendor consistency over time

Per network, the intra- and inter-vendor consistency over time (ranging from 0 to 1240 days) for GE, Philips and Siemens are presented as scatter plots. Intra- and inter-vendor average consistency across all networks are also shown. Abbreviations: Networks: CER, cerebellar; DMN, default mode; FPN, frontoparietal; LIM, limbic; MOT, motor; SAL, salience; VIS, visual. Note: Interactive graphs are provided in the "graphs" Jupyter notebook.

3.3 Consistency of fMRI measures within/between subjects

We evaluated both intra- and inter-subject consistency in HNU1 subjects for all seven networks (Table 4 and Supplementary Material Figure S1).

Network	HNU1 intra-subject	HNU1 inter-subject	Csub intra-subject	Csub vs HNU1 inter-subject
CER	0.64 ± 0.10	0.42 ± 0.08	0.59 ± 0.09	0.37 ± 0.07
DMN	0.73 ± 0.13	0.48 ± 0.10	0.56 ± 0.10	0.33 ± 0.07
FPN	0.68 ± 0.12	0.39 ± 0.07	0.53 ± 0.11	0.26 ± 0.09
LIM	0.60 ± 0.12	0.37 ± 0.09	0.53 ± 0.12	0.30 ± 0.07
MOT	0.72 ± 0.10	0.50 ± 0.09	0.63 ± 0.10	0.43 ± 0.10
SAL	0.73 ± 0.11	0.47 ± 0.08	0.57 ± 0.09	0.35 ± 0.07
VIS	0.75 ± 0.10	0.56 ± 0.09	0.65 ± 0.09	0.48 ± 0.07

Table 4: Consistency of intra-subject and inter-subject connectivity measures
Abbreviations: Networks: CER, cerebellar; DMN, default mode; FPN, frontoparietal; LIM, limbic; MOT, motor; SAL, salience; VIS, visual.

Average intra-subject consistency ranged from 0.60 ± 0.12 (limbic network) to 0.75 ± 0.10 (visual network). Intra-subject consistency in HNU1 was higher than the average intra-Csub consistency across all networks (e.g. 0.73 ± 0.13 in HNU1 vs 0.56 ± 0.10 in Csub, for the DMN, all tests *p*<10^-15). Inter-subject consistency in HNU1 was lower than intra-subject consistency, both HNU1 and Csub (e.g. inter-subject of 0.48 ± 0.10 in HNU1, versus intra-subject of 0.73 ± 0.13 in HNU1 and intra-subject of 0.56 ± 0.10 in

Csub, for the DMN, all tests p<10^-15). The consistency between HNU1 participants and Csub was also lower than the consistency between HNU1 participants across all networks (e.g. inter-subject of 0.48 ± 0.10 in HNU1 vs 0.33 ± 0.0 for Csub vs HNU1, in the DMN, all tests p<10^-15). Overall, the site effects present in Csub scans seemed to decrease both the intra- and inter-subject consistency, compared to monosite HNU1 data. Yet, intra-subject Csub consistency remained higher than inter-subject HNU1 consistency.

3.4 Fingerprinting of HNU1 participants and Csub

We ran a fingerprinting experiment by mixing 2 random scans of Csub with 2 random scans for each of the HNU1 participants, and using an unsupervised cluster analysis to determine whether the two scans of a subject would be clustered together. The highest fingerprinting accuracy was reached for the salience and frontoparietal networks, with about 90+% successful identification on median across subjects (Figure 4).

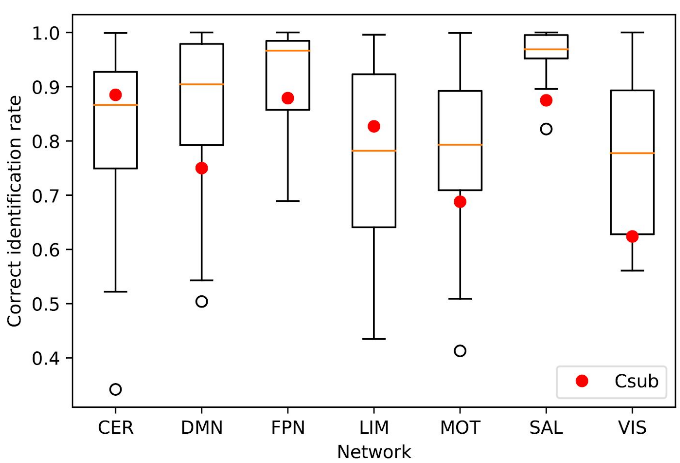


Figure 4: Fingerprinting accuracy
Distribution of the correct identification rate per network for the 26 HNU1 subjects. The red circle indicates the average identification accuracy of Csub per network.

Some networks reached lower median accuracy (~80%), but still much higher than chance level (1/27 subjects ~ 4%). For all networks, the accuracy observed for Csub was on the lower end of the normal range observed for HNU1 subjects. Overall, the observed decrease in intra-subject and inter-subject consistencies on Csub was small enough for fingerprinting Csub with resting-state maps at a high level of accuracy.

4. DISCUSSION

4.1 Consistency in rsfMRI connectivity measures within/between sites

In the present study, we assessed the consistency of rsfMRI connectivity measures in a single participant, scanned at 13 CDIP-compliant sites. We report consistencies of 0.53 ± 0.11 (frontoparietal network) to 0.65 ± 0.09 (visual network) for connectivity maps generated from data obtained at different sites, scanner vendors and time points separated by a wide range of durations, from 0 to 1262 days apart. We found significant effects of scanner vendors and sites, although only one vendor (Siemens) and two sites were associated with substantial effects of higher consistency. The finding that Siemens scanners have more consistent maps than Philips scanners is in agreement with the report of An and colleagues (An et al. 2017). We did not replicate the excellent consistency of GE scanners, but we had only one inter-GE scanner comparison available in the Csub sample. Such inter-vendor differences in consistency may be due to factors such as scanner drift (Friedman and Glover 2006) and smoothness of the raw images produced (Friedman et al. 2006). Our observation that site effects were often very small is in line with the observation of Noble and colleagues (Noble et al. 2017).

This study found that, especially with short time series (less than 10 min), the physiological variability of resting-state measures dominates the scanner variations. Our report extends upon this analysis with longer longitudinal follow up (years instead of weeks), more sites (13 instead of 8), and more vendors (three instead of two).

4.2 Consistency within/between subjects

We found in the HNU1 data that the similarity of maps generated between different individuals is much lower (e.g. 0.48 ± 0.10 for DMN) than the similarity observed intra-subject (e.g. 0.73 ± 0.13 for DMN) in the same sample, as well as intra-subject in our Csub participant (e.g. 0.56 ± 0.10 for DMN). This last consistency value is an average across many scan sites, vendors, and inter-scan intervals and, consequently, the intra-subject Csub consistency was lower than in HNU1 (on average across subjects). The observation suggests that, even with multisite, long-term longitudinal data and relatively short scan duration (about 10 minutes for both Csub and HNU1), it may be possible to implement reliable “brain fingerprinting”. Potential feasibility of fingerprinting was also reinforced by the observation that comparisons between Csub and HNU1 participants were lower on average (e.g. 0.33 ± 0.07 for DMN) than inter-subject comparisons in HNU1 (e.g. 0.48 ± 0.10 for DMN).

4.3 Fingerprinting

We found that it was possible to fingerprint Csub using the connectivity map of a single network with a fairly high level of accuracy (>85%), when Csub scans were pooled with short-term longitudinal scans from 26 HNU1 participants. This level of accuracy was on the lower end, but within the range of the accuracy observed with HNU1 participants.

Note that, on average in the frontoparietal and salience networks, the accuracy of the fingerprinting was over 95% in HNU1. These accuracy levels are only slightly lower than those reported using a connectome-based approach (Finn et al. 2015).

4.4 Study limitations

The major drawback of our study was its inclusion of only a single, male participant. This limitation is due to feasibility constraints, namely the time span of the study (scheduled to last at least five years) and the number of sites involved (set to increase to over 30 sites across Canada and other countries as CDIP is being rolled out to various recruiting sites in supported studies such as the CCNA). We report here on the first wave of data, collected over the initial 3.5 years. In order to assess that this single individual observation may be generalizable to other subjects, we confirmed that the intra-subject consistency we observed in Csub was close to what was observed on average in many HNU1 subjects (N=26) scanned 10 times over the course of one month at a single site. Our findings suggest that the consistency of connectivity maps remain of the same magnitude over several years, at least for a middle aged, healthy subject. However we also found that Csub was substantially less consistent with subjects from HNU1 than inter-subject consistency within HNU1. This observation may be due to the fact that HNU1 connectivity maps may be more similar because they were scanned at the same site. It also may reflect a difference in ethnicity between Csub (Caucasian) and HNU1 participants (a study based in China), and/or a systematic difference in age (Csub was older than HNU1 participants). These differences are a

limitation of the data sample available for this study, and may have contributed to inflate the results of the fingerprinting experiment.

4.5 Alternate preprocessing

We did not evaluate the effects of field-map distortion correction on the consistency of rsfMRI measures. Recently, Togo and colleagues (Togo et al. 2017) reported improved detection of rsfMRI connectivity following field-map distortion correction on a 240 volume single-site dataset acquired on a 3T Siemens scanner. Connectivity was assessed with and without field-map distortion correction in several networks near the paranasal sinuses in the frontal lobe or the mastoid air cells and ear canals in the temporal lobe, brain regions most susceptible to distortion caused by magnetic field inhomogeneity (Jezzard and Balaban 1995). A significant increase in connectivity strength was shown in the default-mode network, a network demonstrating robust consistency in our study. However, only a modest improvement in detection of the cerebellar network was reported, a network with lower consistency in our study (Togo et al. 2017). Moreover, we did not evaluate the effects of physiological noise correction on consistency, since the effect of this correction on consistently measures lacks consensus (Marchitelli et al. 2016).

5. CONCLUSIONS

5.1 Precision medicine

The larger intra-subject consistency (relative to inter-subject consistency) and accurate fingerprinting results in our study suggests that it is possible to extract multivariate

biomarkers of brain diseases from multisite harmonized data. This is consistent with with recent papers (Orban et al. 2017; Abraham et al. 2016) demonstrating that usage of retrospectively pooled multisite rsfMRI data to train machine learning models generalized better to subjects from new unseen sites, than models trained on single site data. In conclusion, our study supports the feasibility of using standardized rsfMRI derivatives for discovery of biomarkers of neurodegeneration, as planned in the CCNA.

5. ACKNOWLEDGEMENTS

We would like to thank Dr. Bratislav Misic for suggesting the fingerprinting experiment, as well as the CCNA LORIS platform (Imaging, Database & Information Technology) for organizing the data, and Hanad Sharmarke for making available the Jupyter notebooks on binder. A.B. is currently supported by a CIHR Postdoctoral Fellowship, the CCNA, and the Courtois Foundation. At the start of the project A.B was supported by the Alzheimer Society of Canada Postdoctoral Fellowship. P.B. is supported by the CCNA and the Courtois Foundation. Financial support for I.C. and O.P. was obtained from the Alzheimer's Society of Canada [grant number 13-32], the Canadian Institute for Health Research [grant number 117121], and the Fonds de recherche du Québec – Santé / Pfizer Canada - Pfizer-FRQS Innovation Fund [grant number 25262]. J.V. is supported by a Vanier Canada Graduate Studies Doctoral scholarship. S.D. is a Research Scholar from the Fonds de recherche du Québec – Santé [grant number 30801]. The *Consortium d'identification précoce de la maladie d'Alzheimer* – *Québec* is financed through the Fonds de recherche du Québec – Santé / Pfizer Canada Innovation Fund [grant number 25262]. The Canadian Consortium on Neurodegeneration in Aging is supported by a grant from the

Canadian Institutes of Health Research with funding from several partners including the Alzheimer Society of Canada, Sanofi, and Women's Brain Health Initiative.

6. REFERENCES

Abraham, Alexandre, Michael Milham, Adriana Di Martino, R. Cameron Craddock, Dimitris Samaras, Bertrand Thirion, and Gael Varoquaux. 2016. "Deriving Reproducible Biomarkers from Multi-Site Resting-State Data: An Autism-Based Example." *NeuroImage*, November. https://doi.org/10.1016/j.neuroimage.2016.10.045.

Ad-Dab'bagh, Y., O. Lyttelton, J. S. Muehlboeck, C. Lepage, D. Einarson, K. Mok, O. Ivanov, et al. 2006. "The CIVET Image-Processing Environment: A Fully Automated Comprehensive Pipeline for Anatomical Neuroimaging Research." In *Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping*, 2266. Florence, Italy.

An, Hyeong Su, Won-Jin Moon, Jae-Kyun Ryu, Ju Yeon Park, Won Sung Yun, Jin Woo Choi, Geon-Ho Jahng, and Jang-Yeon Park. 2017. "Inter-Vender and Test-Retest Reliabilities of Resting-State Functional Magnetic Resonance Imaging: Implications for Multi-Center Imaging Studies." *Magnetic Resonance Imaging* 44 (December): 125–30.

Badhwar, Amanpreet, Angela Tam, Christian Dansereau, Pierre Orban, Felix Hoffstaedter, and Pierre Bellec. 2017. "Resting-State Network Dysfunction in Alzheimer's Disease: A Systematic Review and Meta-Analysis." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 8 (April): 73–85.

Bellec, P., F. M. Carbonell, V. Perlbarg, C. Lepage, O. Lyttelton, V. Fonov, A. Janke, J. Tohka, and A. C. Evans. 2011. "A Neuroimaging Analysis Kit for Matlab and Octave." In *Proceedings of the 17th International Conference on Functional Mapping of the Human Brain*, 2735–46.

Brown, Matthew R. G., Gagan S. Sidhu, Russell Greiner, Nasimeh Asgarian, Meysam Bastani, Peter H. Silverstone, Andrew J. Greenshaw, and Serdar M. Dursun. 2012. "ADHD-200 Global Competition: Diagnosing ADHD Using Personal Characteristic Data Can Outperform Resting State fMRI Measurements." *Frontiers in Systems Neuroscience* 6 (September): 69.

Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews. Neuroscience* 14 (April): 365.

Cheng, Wei, Lena Palaniyappan, Mingli Li, Keith M. Kendrick, Jie Zhang, Qiang Luo, Zening Liu, et al. 2015. "Voxel-Based, Brain-Wide Association Study of Aberrant Functional Connectivity in Schizophrenia Implicates Thalamocortical Circuitry." *NPJ Schizophrenia* 1 (May): 15016.

Cicchetti, D. V., and S. A. Sparrow. 1981. "Developing Criteria for Establishing Interrater Reliability of Specific Items: Applications to Assessment of Adaptive Behavior." *American Journal of Mental Deficiency* 86 (2): 127–37.

Dansereau, Christian, Yassine Benhajali, Celine Risterucci, Emilio Merlo Pich, Pierre Orban, Douglas Arnold, and Pierre Bellec. 2017. "Statistical Power and Prediction Accuracy in Multisite Resting-State fMRI Connectivity." *NeuroImage* 149 (April): 220–32.

Di Martino, Adriana, David O'Connor, Bosi Chen, Kaat Alaerts, Jeffrey S. Anderson, Michal

Assaf, Joshua H. Balsters, et al. 2017. “Enhancing Studies of the Connectome in Autism Using the Autism Brain Imaging Data Exchange II.” *Scientific Data* 4 (March): 170010.

Dong, Aoyan, Jon B. Toledo, Nicolas Honnorat, Jimit Doshi, Erdem Varol, Aristeidis Sotiras, David Wolk, John Q. Trojanowski, Christos Davatzikos, and Alzheimer’s Disease Neuroimaging Initiative. 2017. “Heterogeneity of Neuroanatomical Patterns in Prodromal Alzheimer’s Disease: Links to Cognition, Progression and Biomarkers.” *Brain: A Journal of Neurology* 140 (3): 735–47.

Drysdale, Andrew T., Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N. Fetcho, et al. 2017. “Resting-State Connectivity Biomarkers Define Neurophysiological Subtypes of Depression.” *Nature Medicine* 23 (1): 28–38.

Finn, Emily S., Xilin Shen, Dustin Scheinost, Monica D. Rosenberg, Jessica Huang, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable. 2015. “Functional Connectome Fingerprinting: Identifying Individuals Using Patterns of Brain Connectivity.” *Nature Neuroscience* 18 (11): 1664–71.

Fleiss, Joseph L., and Jacob Cohen. 1973. “The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability.” *Educational and Psychological Measurement* 33 (3): 613–19.

Friedman, Lee, and Gary H. Glover. 2006. “Report on a Multicenter fMRI Quality Assurance Protocol.” *Journal of Magnetic Resonance Imaging: JMRI* 23 (6): 827–39.

Friedman, Lee, Gary H. Glover, Diana Krenz, Vince Magnotta, and FIRST BIRN. 2006. “Reducing Inter-Scanner Variability of Activation in a Multicenter fMRI Study: Role of Smoothness Equalization.” *NeuroImage* 32 (4): 1656–68.

Giove, Federico, Tommaso Gili, Vittorio Iacovella, Emiliano Macaluso, and Bruno Maraviglia. 2009. “Images-Based Suppression of Unwanted Global Signals in Resting-State Functional Connectivity Studies.” *Magnetic Resonance Imaging* 27 (8): 1058–64.

Gratton, Caterina, Timothy O. Laumann, Ashley N. Nielsen, Deanna J. Greene, Evan M. Gordon, Adrian W. Gilmore, Steven M. Nelson, et al. 2018. “Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation.” *Neuron* 98 (2): 439–52.e5.

Hunter, John D. 2007. “Matplotlib: A 2D Graphics Environment.” *Computing in Science & Engineering* 9 (3): 90–95.

Jezzard, P., and R. S. Balaban. 1995. “Correction for Geometric Distortion in Echo Planar Images from B0 Field Variations.” *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 34 (1): 65–73.

Jones, Eric, Travis Oliphant, and Pearu Peterson. 2016. “Others. SciPy: Open Source Scientific Tools for Python. 2001.” *URL Http://www. Scipy. Org*.

Jovicich, Jorge, Ludovico Minati, Moira Marizzoni, Rocco Marchitelli, Roser Sala-Llonch, David Bartrés-Faz, Jennifer Arnold, et al. 2016. “Longitudinal Reproducibility of Default-Mode Network Connectivity in Healthy Elderly Participants: A Multicentric Resting-State fMRI Study.” *NeuroImage* 124 (Pt A): 442–54.

Koo, Terry K., and Mae Y. Li. 2016. “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.” *Journal of Chiropractic Medicine* 15 (2): 155–63.

Li, Hongming, Theodore D. Satterthwaite, and Yong Fan. 2018. “BRAIN AGE PREDICTION BASED ON RESTING-STATE FUNCTIONAL CONNECTIVITY PATTERNS USING CONVOLUTIONAL NEURAL NETWORKS.” *Proceedings / IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE International Symposium on Biomedical*

Imaging 2018 (April): 101–4.

Li, Hui-Jie, Xiao-Hui Hou, Han-Hui Liu, Chun-Lin Yue, Yong He, and Xi-Nian Zuo. 2015. "Toward Systems Neuroscience in Mild Cognitive Impairment and Alzheimer's Disease: A Meta-Analysis of 75 fMRI Studies." *Human Brain Mapping* 36 (3): 1217–32.

Marchitelli, Rocco, Ludovico Minati, Moira Marizzoni, Beatriz Bosch, David Bartrés-Faz, Bernhard W. Müller, Jens Wiltfang, et al. 2016. "Test-Retest Reliability of the Default Mode Network in a Multi-Centric fMRI Study of Healthy Elderly: Effects of Data-Driven Physiological Noise Correction Techniques." *Human Brain Mapping* 37 (6): 2114–32.

Matthews, Paul M., and Adam Hampshire. 2016. "Clinical Concepts Emerging from fMRI Functional Connectomics." *Neuron* 91 (3): 511–28.

McKinney, Wes, and Others. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, 445:51–56. Austin, TX.

Nielsen, Jared A., Brandon A. Zielinski, P. Thomas Fletcher, Andrew L. Alexander, Nicholas Lange, Erin D. Bigler, Janet E. Lainhart, and Jeffrey S. Anderson. 2013. "Multisite Functional Connectivity MRI Classification of Autism: ABIDE Results." *Frontiers in Human Neuroscience* 7 (September): 599.

Noble, Stephanie, Dustin Scheinost, Emily S. Finn, Xilin Shen, Xenophon Papademetris, Sarah C. McEwen, Carrie E. Bearden, et al. 2017. "Multisite Reliability of MR-Based Functional Connectivity." *NeuroImage* 146 (February): 959–70.

Oliphant, Travis E. 2006. *A Guide to NumPy*. Vol. 1. Trelgol Publishing USA.

Orban, Pierre, Christian Dansereau, Laurence Desbois, Violaine Mongeau-Pérusse, Charles-Édouard Giguère, Hien Nguyen, Adrianna Mendrek, Emmanuel Stip, and Pierre Bellec. 2017. "Multisite Generalizability of Schizophrenia Diagnosis Classification Based on Functional Brain Connectivity." *Schizophrenia Research*, June. https://doi.org/10.1016/j.schres.2017.05.027.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research: JMLR* 12 (Oct): 2825–30.

Power, Jonathan D., Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2012. "Spurious but Systematic Correlations in Functional Connectivity MRI Networks Arise from Subject Motion." *NeuroImage* 59 (3): 2142–54.

Saggar, Manish, Eva Tsalikian, Nelly Mauras, Paul Mazaika, Neil H. White, Stuart Weinzimer, Bruce Buckingham, Tamara Hershey, Allan L. Reiss, and Diabetes Research in Children Network (DirecNet). 2017. "Compensatory Hyperconnectivity in Developing Brains of Young Children With Type 1 Diabetes." *Diabetes* 66 (3): 754–62.

Seabold, Skipper, and Josef Perktold. 2010. "Statsmodels: Econometric and Statistical Modeling with Python." In *Proceedings of the 9th Python in Science Conference*, 57:61. SciPy society Austin.

Skåtun, Kristina C., Tobias Kaufmann, Nhat Trung Doan, Dag Alnæs, Aldo Córdova-Palomera, Erik G. Jönsson, Helena Fatouros-Bergman, et al. 2017. "Consistent Functional Connectivity Alterations in Schizophrenia Spectrum Disorder: A Multisite Study." *Schizophrenia Bulletin* 43 (4): 914–24.

Sperling, Reisa A., Paul S. Aisen, Laurel A. Beckett, David A. Bennett, Suzanne Craft, Anne M. Fagan, Takeshi Iwatsubo, et al. 2011. "Toward Defining the Preclinical Stages of Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 7 (3): 280–92.

Togo, Hiroki, Jaroslav Rokicki, Kenji Yoshinaga, Tatsuhiro Hisatsune, Hiroshi Matsuda,

Nobuhiko Haga, and Takashi Hanakawa. 2017. “Effects of Field-Map Distortion Correction on Resting State Functional Connectivity MRI.” *Frontiers in Neuroscience* 11 (December): 656.

Urchs, Sebastian, Jonathan Armoza, Yassine Benhajali, Jolène St-Aubin, Pierre Orban, and Pierre Bellec. 2017. “MIST: A Multi-Resolution Parcellation of Functional Brain Networks.” *MNI Open Research* 1 (December): 3.

Vemuri, Prashanthi, David T. Jones, and Clifford R. Jack Jr. 2012. “Resting State Functional MRI in Alzheimer’s Disease.” *Alzheimer’s Research & Therapy* 4 (1): 2.

Yan, Chao-Gan, R. Cameron Craddock, Xi-Nian Zuo, Yu-Feng Zang, and Michael P. Milham. 2013. “Standardizing the Intrinsic Brain: Towards Robust Measurement of Inter-Individual Variation in 1000 Functional Connectomes.” *NeuroImage* 80 (October): 246–62.

Zuo, Xi-Nian, Jeffrey S. Anderson, Pierre Bellec, Rasmus M. Birn, Bharat B. Biswal, Janusch Blautzik, John C. S. Breitner, et al. 2014. “An Open Science Resource for Establishing Reliability and Reproducibility in Functional Connectomics.” *Scientific Data* 1 (December): 140049.

7. SUPPLEMENTARY MATERIAL

Provided are Table S1 referred to in Section 2.5 of the manuscript, and Table S2 referred to in Section 3.2 of the manuscript.

Network	Total volume	Volumes scrubbed	Volumes remaining	% scrubbed	% remaining
CHUM1	297	4	293	1.35	98.65
CHUM2	297	42	255	14.14	85.86
CHUS1	297	8	289	2.69	97.31
CHUS2	297	91	206	30.64	69.36
CHUS3	297	34	263	11.45	88.55
CINQ1	297	86	211	28.96	71.04
CINQ2	297	118	179	39.73	60.27
CINQ3	297	41	256	13.80	86.20
UBC1	297	42	255	14.14	85.86
ISMD1	297	0	297	0	100
ISMD2	297	4	293	1.35	98.65
ISMD3	297	0	297	0	100
IUGM1	297	0	297	0	100
IUGM2	297	0	297	0	100
IUGM3	297	0	297	0	100
MNI1	296	0	296	0	100
MNI2	297	0	297	0	100
MNI3	297	0	297	0	100
EDM	297	0	297	0	100
SASK	247	18	229	7.29	92.71
TWH	247	0	247	0	100
RRI	247	0	247	0	100
SUN	275	0	275	0	100
VIC	247	12	235	4.86	95.14

Table S1: Time frames removed due to excessive motion.

Time frames showing a displacement >0.5 mm were removed. Abbreviations: CHUM, Centre Hospitalier Universitaire de Sherbrooke; CHUS, Centre Hospitalier Universitaire de Sherbrooke; CINQ, Le Consortium d'Imagerie en Neurosciences et Santé Mentale de Québec; ISDM, Institut Universitaire en Santé Mentale Douglas; IUGM, Institut Universitaire de Gériatrie de Montréal;

MNI, Montreal Neurological Institute; EDM, Peter S. Allen MR Research Centre; SASK, Royal University Hospital; TWH, Toronto Western Hospital; RRI, Robarts Research Institute; SUN, Sunnybrook Health Sciences Center/Sunnybrook Research Institute; VIC, West Coast Medical Imaging.

	CER	DMN	FPN	LIM	MOT	SAL	VIS
CHUM	0.1210	*0.1822	0.1431	*0.1964	*0.1945	0.0955	*0.1949
CHUS	-0.0298	-0.0257	-0.0131	-0.0665	-0.0557	-0.0246	-0.0259
CINQ	0.0037	-0.0087	0.0110	-0.0366	-0.0414	-0.0171	0.0218
ISMD	*0.1036	-0.0193	-0.0148	0.0617	0.0587	*0.0912	0.0536
IUGM	0.0569	-0.0027	0.0817	0.0218	0.0519	0.0677	0.0400
MNI	*0.1536	0.0641	*0.1085	0.0717	0.0887	0.0778	0.0169
GE	0.1412	0.0349	0.1470	0.0563	0.0011	0.0445	0.0099
Philips	*0.0418	-0.0048	0.0096	0.0291	0.0134	*0.0458	0.0131
Siemens	*0.0553	*0.1091	*0.0955	*0.1507	*0.0901	*0.0811	*0.1249

Table S2:

Effect of each individual vendor and site for each of the seven rsfMRI networks. * indicated $p<0.05$, uncorrected for multiple comparisons.

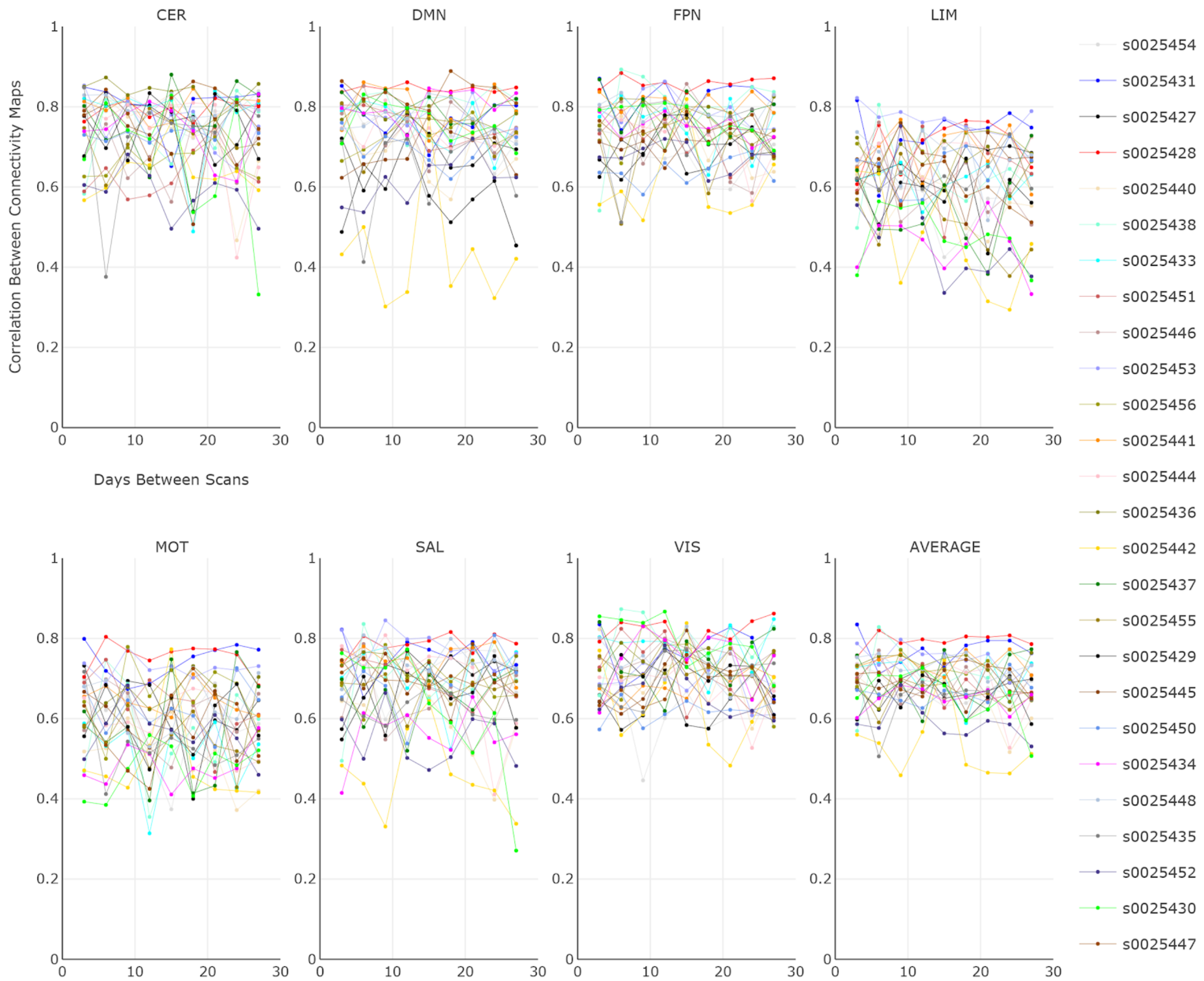


Figure S1: Intra- and inter-subject consistencies

Intra- and inter-subject consistency for the 26 individuals in the HNU1 dataset are plotted per network. The average consistency across all networks are also provided. To visualize the effect of time on intra-subject consistency, we plotted the consistency between connectivity maps at scan session 0 with that at scan sessions 1 to 9 (e.g. subject s0025438: scan session 0 vs scan session 1, subject s0025438: scan session 0 vs scan session 2). Per subject, the time interval between scan sessions was 3 days, with a total of 10 scans across one month. Abbreviations: CER, cerebellar network; DMN, default mode network; FPN, frontoparietal network; LIM, limbic network; MOT, motor network; SAL, salience network; VIS, visual network. Note: Interactive graphs are provided in the "graphs" Jupyter notebook.