

Supporting Information

Biswal et al. 10.1073/pnas.0911855107

SI Methods

Image Preprocessing. Overview. All available resting-state scans were preprocessed using both AFNI (1) and FSL (www.fmrib.ox.ac.uk). Specific commands can be found in the preprocessing scripts that will be released at www.nitrc.org/projects/fcon_1000/ on publication of this paper. After the first five time points of every scan were discarded, to remove possible T1 stabilization effects, the data were corrected for motion by aligning each volume to the mean image volume using Fourier interpolation in AFNI. Then the data were spatially smoothed using a 6-mm FWHM Gaussian kernel. Mean-based intensity normalization was done by scaling all volumes by the same factor (10,000).

Seed-based correlation analyses. The data were temporally filtered using both a high-pass (Gaussian-weighted least squares straight-line fitting, with $\sigma = 100.0$ s) and low-pass (Gaussian low-pass temporal filtering, with a HWHM of 2.8 s) filter, followed by linear detrending to remove any residual drift.

Independent component analysis. Temporal concatenation group analysis. Consistent with common practice, temporal filtering for ICA analyses was limited to high-pass filtering (Gaussian-weighted least squares straight-line fitting, with $\sigma = 100.0$ s).

Dual regression. This step used the same preprocessed data as used in the seed-based correlation analyses.

ALFF/fALFF. No temporal filtering was carried out, because the data were examined in the frequency domain within select bands (2, 3). Temporal despiking with a hyperbolic tangent squashing function was performed, however, to limit extreme values. Linear trends were then removed from the data.

Registration and normalization. After the skull was removed using AFNI, registration of each individual's high-resolution anatomic image to a common stereotactic space [the Montreal Neurological Institute's 152-brain template (MNI152); 3 mm isotropic voxel size] was done using a 12-degrees of freedom linear affine transformation (FLIRT) (4, 5). The resulting transformation was then applied to each individual's functional dataset. We did not further optimize the normalization with a nonlinear algorithm, because of concerns about image quality and limited coverage in some datasets.

Functional Connectivity: Seed-Based Correlation Analysis. Nuisance signal regression. Consistent with common practice in the fMRI literature, nuisance signals were removed from the data via multiple regression before functional connectivity analyses were performed. This step is designed to control for the effects of physiological processes, such as fluctuations related to motion and cardiac and respiratory cycles. Specifically, each individual's 4D time series data were regressed on nine predictors: white matter (WM), cerebrospinal fluid (CSF), the global signal, and six motion parameters. The global signal regressor was generated by averaging across the time series of all voxels in the brain. The WM and CSF covariates were generated by segmenting each individual's high-resolution structural image (using FAST in FSL). The resulting segmented WM and CSF images were thresholded to ensure 80% tissue type probability. These thresholded masks were then applied to each individual's time series, and a mean time series was calculated by averaging across time series of all voxels within each mask. The six motion parameters were calculated in the motion-correction step during preprocessing. Movement in each of the three cardinal directions (X, Y, and Z) and rotational movement around three axes (pitch, yaw, and roll) were included for each individual.

Seed selection. Six 7.5-mm-radius seed regions of interest (ROIs) (containing 33 voxels) centered on the coordinates previously used

by Fox et al. (6) were created to examine functional connectivity for each of six regions, three regions within the "task-positive" network and three within the "default mode" network. The ROIs within the task-positive network were located in the IPS (-25, -57, 46), the middle temporal region (MT+; -45, -69, -2), and the right frontal eye field (FEF) region of the precentral sulcus (25, -13, 50). The default mode network seed ROIs were located in the left lateral parietal cortex (LP; -45, -67, 36), medial prefrontal cortex (MPF; -1, 47, -4), and PCC (-5, -49, 40).

Individual seed-based functional connectivity analysis. First, each individual's residual 4D time series data were spatially normalized by applying the previously computed transformation to the MNI152 standard space. Then the time series for each seed was extracted from these data. Time series were averaged across all voxels in each seed's ROI. For each individual dataset, the correlation between the time series of the seed ROI and that of each voxel in the brain was determined. This analysis was implemented using 3dfm+ (AFNI) to produce individual-level correlation maps of all voxels that were positively or negatively correlated with the seed's time series. Finally, these individual-level correlation maps were converted to Z-value maps using Fisher's r -to- z transformation.

Functional Connectivity: Independent Component Analysis. Overview. Temporal-concatenation group ICA (TC-GICA) was used to generate group-level components for the dataset (7) using MELODIC (FSL). Given computational resource limitations (e.g., 32 GB of physical memory), as well as a number of centers with a small number of time points due to repetition times >2.0 s, each TC-GICA run was applied to a dataset consisting of 18 participants/center from the 17 centers that collected a minimum of 165 functional volumes per scan. This approach also ensured that a single center's data would not drive the ICA components detected. Consistent with recent work on low-dimensional ICA (8), the number of components was fixed at 20. Given the potential for such factors as initial random values and subject sampling to affect ICA results, 25 TC-GICA analyses were performed, each using a unique resampling from each of the 17 centers. A meta-ICA analysis was then carried out across the 25 runs to extract the 20 spatially independent components consistently identified across the 25 runs. An alternative hierarchical clustering approach based on ICASSO (9) is described below. The two approaches yielded similar results. Dual regression (10, 11) was then carried out using the 20 resulting components as templates, to produce individual participant maps for each of the 20 components.

TC-GICA. Specifically, TC-GICA comprised five fundamental steps:

1. Each individual's preprocessed data were first truncated to the same number of time points (i.e., 165 EPI volumes).
2. A bootstrapping dataset was generated by randomly choosing 18 individual datasets per center, resulting in 306 individual functional datasets.
3. All 306 individual functional datasets were spatially averaged in MNI152 standard space and then used to estimate the mean covariance matrix.
4. The number of components was set at 20, and all individual functional data were projected into a subspace spanned by the first 20 eigenvectors of the mean covariance matrix, resulting in reduced individual fMRI data (in a common subspace).

- All 306 reduced individual datasets were temporally concatenated, reduced via principal component analysis to 20 dimensions, and fed into the probabilistic ICA algorithm with a random initial value (12).

This procedure produced 20 group-level components for each TC-GICA run. Finally, 500 (20 × 25) group-level components were generated from the 25 TC-GICA runs.

Generation of component templates for dual regression (meta-ICA). To provide more accurate and robust ICA component templates, we carried out another low-dimensional (20 components) TC-GICA. Here we concatenated the 500 components produced by the 25 TC-GICA runs as the input data of a single-session ICA in MELODIC. The resultant maps were used as final component templates for the dual regression procedure. Of note, this method was selected as the primary approach over the alternative approach described because it guarantees the spatial independence of the 20 components, whereas the alternative approach does not.

Generation of component templates for dual regression model (alternative approach). To emphasize the robustness of the findings of the meta-ICA, here we describe an alternative approach that yields nearly identical components to the meta-ICA. The findings of the two approaches differed notably for only one of the 20 components, for which the meta-ICA finding was more plausible. Given the high degree of similarity between the two methods, we present only the findings from the meta-ICA in the present work. In the alternative approach, we used the hierarchical clustering algorithm implemented in the ICASSO toolbox (9). ICASSO was designed for validating the robustness of ICA with respect to random initial values (of the ICA mixing matrix) and the ICA cost function optimization search strategy. However, due to limitations in computational resources (e.g., 32 GB of memory in the present work), TC-GICA cannot be carried out on the full datasets. Thus, we used the bootstrapping approach described above with 25 ICA analyses, in which initial values and the specific participants selected from each center varied from one ICA analysis to the next. Here the 500 group-level components (20 components per run × 25 runs) were sorted using hierarchical clustering. The number of clusters (20) was selected to match the number of components. The similarity between components was measured by the combination of both spatial R_s and temporal R_t correlations in Eq. (1) and the distance between components as defined in Eq. (2) (13):

$$S(i,j) = \lambda * R_s(i,j) + (1 - \lambda) * R_t(i,j) \quad (1)$$

$$D(i,j) = \sqrt{1 - S(i,j)}, 1 \leq i, j \leq 500 \quad (2)$$

Considering the spatial ICA, $\lambda = 0.8$ was chosen in our clustering procedure. Finally, the median value at each voxel for each of the 20 clusters was calculated to determine the final component templates for the dual regression procedure.

Individual component reconstruction via the dual regression model. To reconstruct component maps for each participant, the recently developed dual regression procedure (11, 14) was applied to each of the 1,093 individual participants' datasets. Specifically, in the present work, dual regression consisted of two linear regressions carried out independently for each of the 20 component maps identified in temporal concatenation ICA. For each component template, the first regression model used the template as a spatial predictor for the participant's 4D data, producing a set of individual regression weights in the time domain (i.e., a time series for each spatial map). Using this time series as a temporal predictor for the 4D BOLD data, the second regression equation estimated the individual regression weights in the spatial domain (i.e., the participant-level individual spatial map). Both regressions used the same data set used for the seed-based con-

nectivity approaches, that is, each participant's 4D dataset after removal of the nine nuisance covariates. Component time series were demeaned in both regressions, but no variance normalization was used. The dual regression procedure was carried out for all 1,093 participants included across 24 centers, not just those used for the generation of ICA-based templates. For each component, these individual spatial maps were then used to evaluate group-level statistics.

Amplitude of spontaneous low-frequency fluctuations. To examine the potentially meaningful information contained within the ALFF, two fast-Fourier transformation (FFT)-based indices, ALFF and fALFF, were used to compute the amplitude of low-frequency fluctuations in the frequency domain (2, 3, 15). For each individual, ALFF and fALFF were computed to identify those voxels with significantly detectable low-frequency fluctuation amplitudes. Specifically, at each voxel, ALFF is calculated as the sum of amplitudes within a specific low frequency range (0.01–0.1 Hz). fALFF is the normalized ALFF, calculated by dividing the ALFF value by the total sum of amplitudes across the entire frequency range measured in a given time series. Voxelwise ALFF and fALFF maps were calculated for each participant in native space, and then transformed into the MNI152 standard brain space with 3-mm isotropic voxel size. Before statistical analyses, each individual ALFF or fALFF map was Z-transformed (i.e., by subtracting the mean voxelwise ALFF or fALFF obtained for the individual's entire brain, and then dividing by the corresponding SD) to improve its suitability for group-level parametric analyses. The individual Z-transformed ALFF or fALFF maps were used in subsequent group- and center-level analyses.

Unified group-level statistical model. For all three types of R-fMRI measures (seed-based correlations, ICA, and ALFF/fALFF), a unified general linear model frame was developed for center-level statistical analyses. The unified statistical model is a one-way ANOVA, treating centers as the between factor. F-contrasts were used to measure the effect of centers. Overall group mean contrasts across all centers were modeled as well. Specifically, a one-factor 24-level ANOVA (factor: center; 1,093 participants), with age and sex as covariates, was used to examine the effects of age, sex, and center on the three R-fMRI measures. Multiple comparisons were corrected at the cluster level using Gaussian random field theory (min $Z > 2.3$; cluster significance: $P < 0.05$, corrected).

SI Results

Center-Related Variability. The results presented in Fig. 1 show that the effects ascribable to center can be either interpreted as negligible, as indicated by the high between-center Kendall's W (row 2), or substantial, accounting for much of the variance (row 3). These opposite interpretations are not mutually exclusive. The high between-center Kendall's W indicates that the resting-state measures (i.e., functional connectivity, fluctuation amplitude) obtained from different centers have a high degree of similarity. Nevertheless, systematic differences exist between centers, and these are easily quantified by ANOVA. In Fig. S5 for each center, the mean functional connectivity across 40 peak voxels derived from the center effect map for the PCC seed is depicted. As the figure shows, there are between-center differences in the height of the functional connectivity values. Some centers have overall higher functional connectivity values than others; these differences in the height of functional connectivity values drive the significant between-center effects. The variability in functional connectivity values could be related to a number of factors (e.g., the specific scanner used, scanner sequence, sample characteristics, specific instructions to participants, degree of variability in participant wakefulness). Because there was no previous coordination among centers regarding scanning parameters, each of these parameters could contribute to across-center differences. Specific examination of these factors is beyond the scope of the present work, but we anticipate that it will be

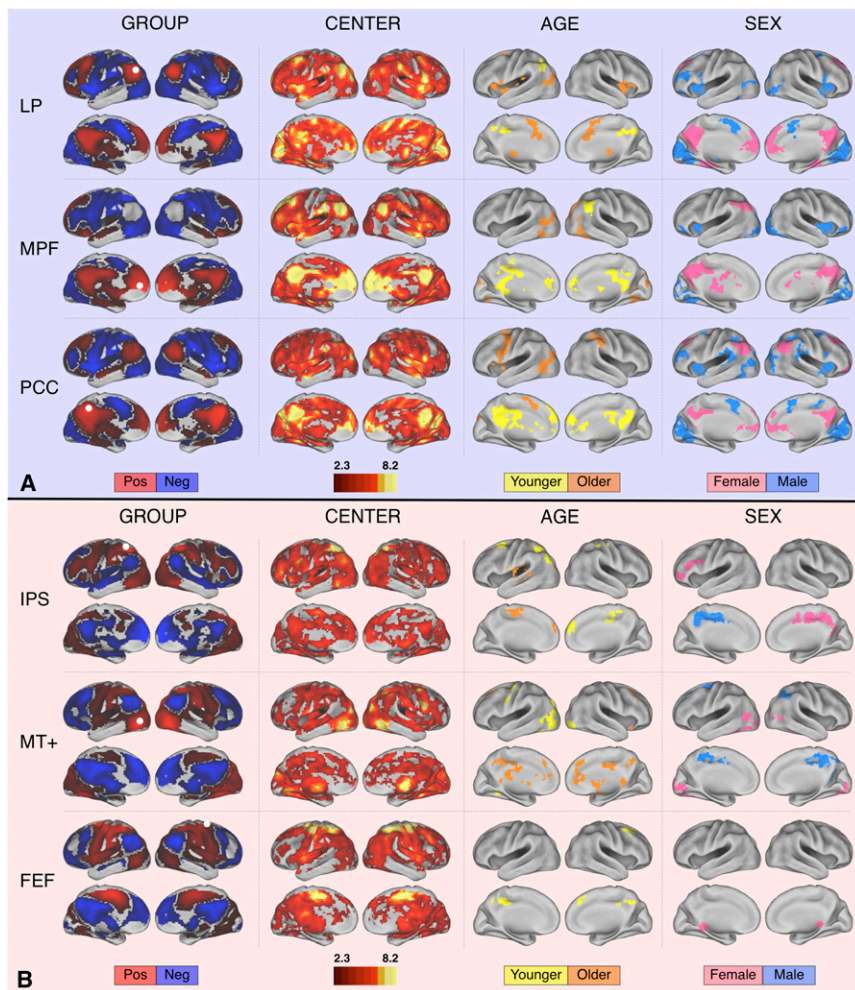


Fig. S1. Center-, age-, and sex-related variations detected in R-fMRI measures of functional connectivity using seed-based correlation analyses. The first column depicts group-level functional connectivity maps for three representative "default mode" seeds (A) and three "task-positive" network seeds (B). The seed ROIs are shown as white circles. The second column depicts voxels exhibiting significant effects of center, as detected by one-way ANOVA (across 24 centers, including 1,093 participants). Columns 3 and 4 depict voxels exhibiting age- and sex-related variations (modeled as covariates). Center, sex, and age findings were corrected for multiple comparisons ($Z > 2.3$; $P < 0.05$, corrected). All supplementary cortical surface maps are arrayed as shown in Fig. 1, with lateral views in upper rows, medial views in lower rows, left hemisphere on the left, and right hemisphere on the right. "Male" refers to significantly greater connectivity in males; similarly, "female" refers to significantly greater connectivity in females. "Older" refers to significantly increasing connectivity with increasing age, whereas "younger" refers to significantly increasing connectivity with decreasing age. "Pos", positive functional connectivity; "neg", negative functional connectivity.

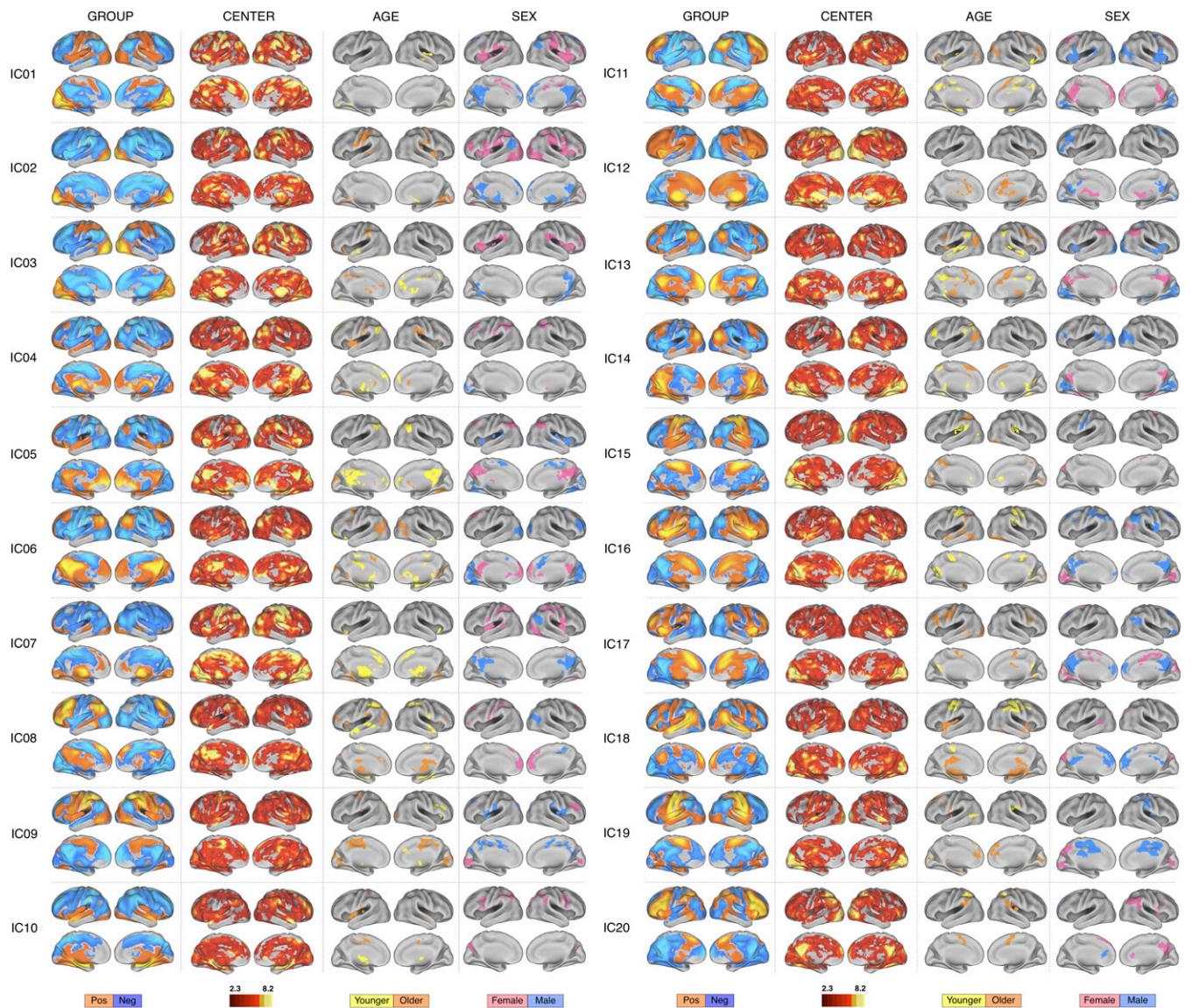


Fig. S2. Center-, age-, and sex-related differences detected in R-fMRI measures of functional connectivity combining independent component and dual regression analyses. The first column depicts group-level maps for 20 functional connectivity ICs. For each component, the second column depicts voxels exhibiting significant effects of center, as detected by one-way ANOVA (across 24 centers, including 1,093 participants). Columns 3 and 4 depict voxels exhibiting age- and sex-related variations. Center, age and sex findings were corrected for multiple comparisons ($Z > 2.3$; $P < 0.05$, corrected). “Male” refers to significantly greater connectivity in males; similarly, “female” refers to significantly greater connectivity in females. “Older” refers to significantly increasing connectivity with increasing age, whereas “younger” refers to significantly increasing connectivity with decreasing age. “Pos,” positive group effect; “neg,” negative group effect.

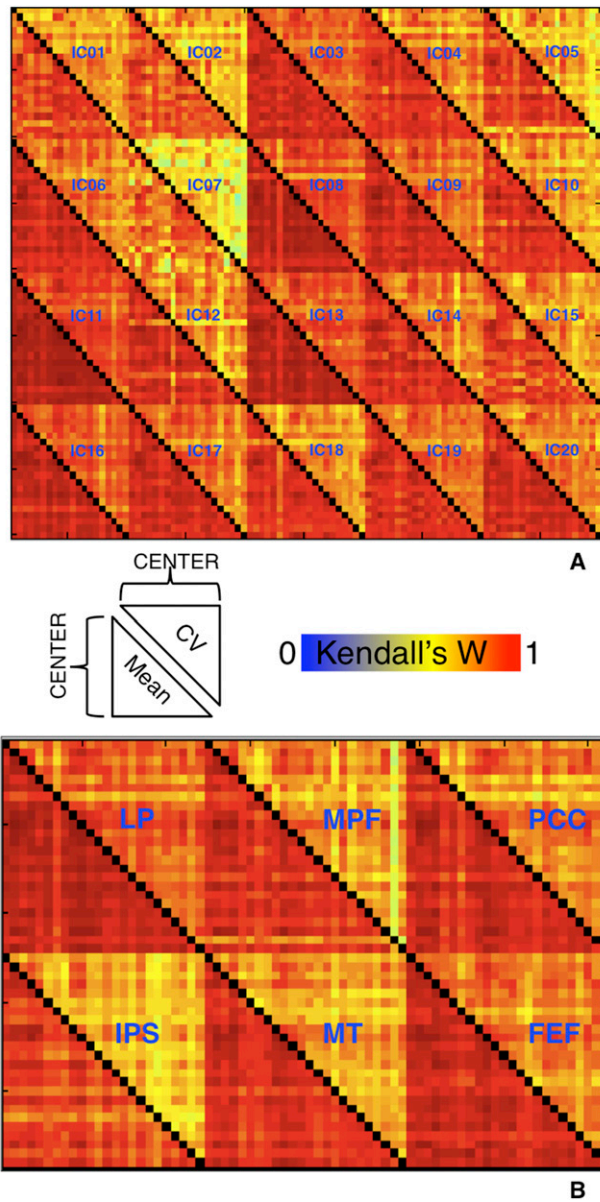


Fig. 54. Consistency of R-fMRI measures across centers: ICA combined with dual regression (A), and seed-based correlation (B). For each center, the voxelwise mean and coefficient of variation was calculated for each R-fMRI measure. The Kendall's W concordance of the mean or coefficient of variation maps between any two centers was calculated. The coefficient of variation is depicted above the diagonal, the mean below.

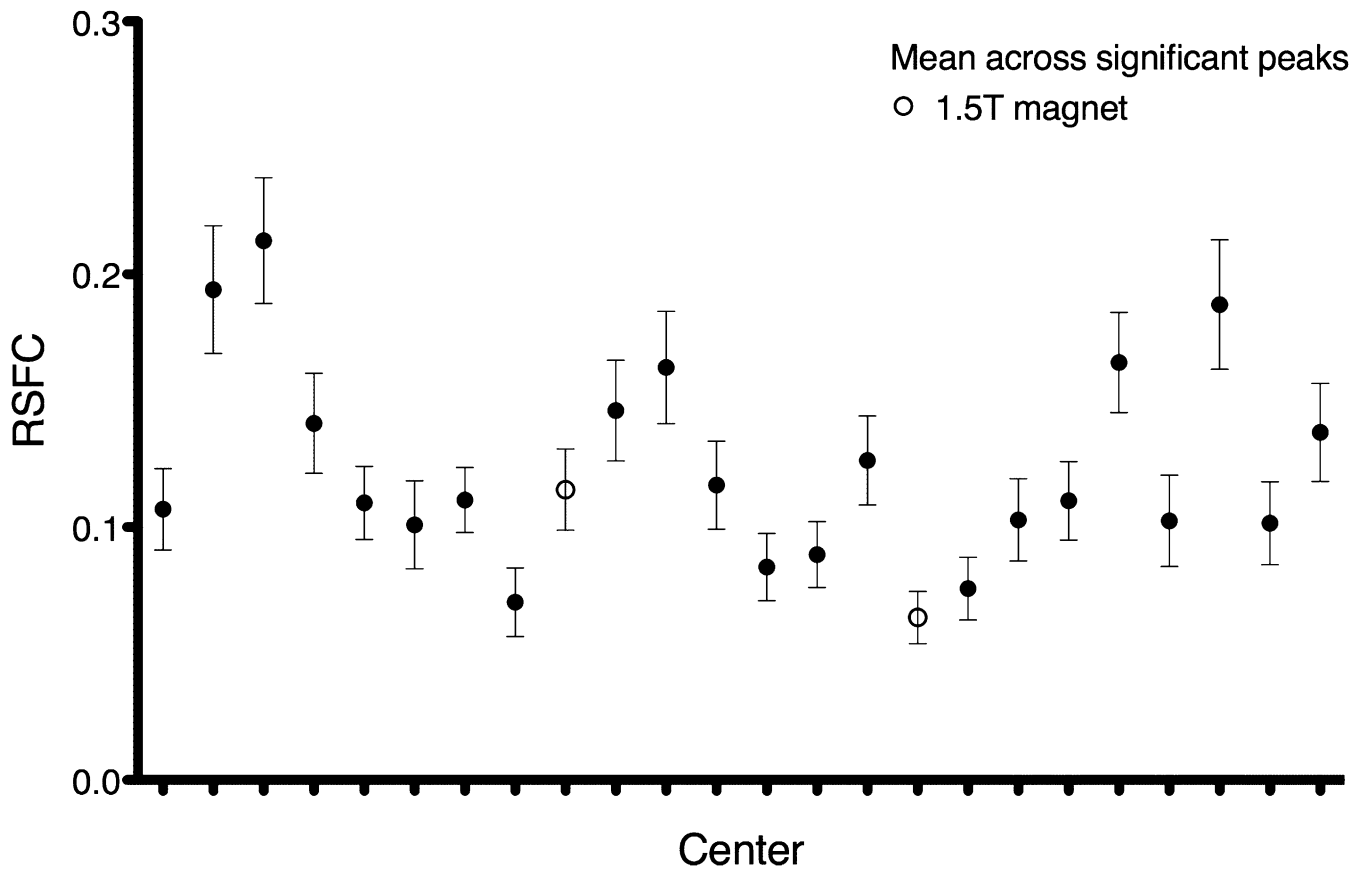


Fig. S5. Functional connectivity values observed at peak locations of between-center differences. For each center we calculated the mean across a 3 mm radius sphere centered at each of the 40 most significant voxels indexing the effect of center for the PCC seed ROI (Fig. 1, column 1, row 3). Connectivity values indexed the functional connectivity between the 3 mm radius sphere and the PCC seed ROI. All centers included in the analyses are shown ($n = 24$). Although the strength of functional connectivity values observed across centers clearly varies, the within-center variability is relatively low. This indicates that the differences in functional connectivity strength among centers are relatively stable across the brain. A center that shows higher functional connectivity in one area of the brain compared with another center most likely also shows higher functional connectivity in other areas of the brain.

Table S1. Data currently included in the 1,000 Functional Connectomes Project

	Center	PI	<i>N</i>	<i>n</i> *	Age years, mean (SD)	Age range years	Male sex %
1.	Baltimore, MD	J. J. Pekar/S. H. Mostofsky	23		29.26 (5.46)	20–40	35%
2.	Bangor, UK	S. Colcombe	20		23.4 (5.32)	19–38	100%
3.	Beijing, China	YF. Zang	198	193	21.16 (1.83)	18–26	39%
4.	Beijing, China	XC. Weng	28	27	20.41 (1.39)	18–24	27%
5.	Berlin, Germany	D. Margulies	26		29.77 (5.21)	23–44	50%
6.	Bethesda, MD	M. Ernst	18		33.00 (13.31)	18–53	22%
7.	Cambridge, MA	R. L. Buckner	198		21.03 (2.31)	18–30	38%
8.	Cambridge, MA	S. Whitfield-Gabrieli	39	35	25.09 (3.53)	20–32	49%
9.	Cleveland, OH	M. J. Lowe	31		43.55 (11.14)	24–60	35%
10.	Dallas, TX	B. Rypma	24		42.63 (20.07)	20–71	50%
11.	Hvidovre, Denmark	A.-M. Dogonowski/K. Madsen	28		41.75 (10.7)	21–68	50%
12.	Leiden, The Netherlands	S. A. R. B. Rombouts	31		22.19 (2.57)	18–28	74%
13.	Leipzig, Germany	A. Villringer	37		26.22 (5)	20–42	43%
14.	Magdeburg, Germany	M. Walter	29	28	30.43 (5.71)	22–43	93%
15.	Milwaukee, WI	SJ. Li	64		53.59 (5.79)	44–65	64%
16.	New Haven, CT	M. Hampson	19	18	31.61 (10.27)	18–48	56%
17.	New York, NY [†]	M. Milham/F. X. Castellanos	59		32.78 (8.83)	20–49	68%
18.	New York, NY [†]	M. Milham/F. X. Castellanos	20		29.75 (9.94)	18–46	40%
19.	Newark, NJ	B. B. Biswal	19		24.11 (3.91)	21–39	47%
20.	Orangeburg, NY [‡]	M. J. Hoptman	21	20	40.65 (11.03)	20–55	75%
21.	Oulu, Finland [‡]	V. J. Kiviniemi/J. Veijola	103		21.52 (0.57)	20–23	36%
22.	Oxford, UK	S. M. Smith/C. Mackay	22		29 (3.79)	20–35	55%
23.	Queensland, Australia	K. McMahon	19	18	26.28 (3.71)	20–34	61%
24.	St. Louis, MO	B. L. Schlaggar/S. E. Petersen	31		25.1 (2.31)	21–29	45%

Data from the following centers will be included in the 1000 Functional Connectomes data release but are not included in the current analyses: Ann Arbor, MI: C. S. Monk/R. D. Seidler/S. J. Peltier; Atlanta, GA: H. S. Mayberg; Berlin, Germany: S. Schmidt; Durham, NC: D. J. Madden; Durham, NC: L. Wang; London, Ontario, Canada: P. Williamson; Munich, Germany, C. Sorg/V. Riedl; Nanjing, China: G.J. Teng/HY. Zhang; Pittsburgh, PA: G.J. Siegle; Portland, OR: D. Fair/B. J. Nagel; Taipei, Taiwan: C.P. Lin; Vienna, Austria: C. Windischberger.

*Actual number of participants included in the analysis, if different from *N*.

[†]Data from the same magnet, different sequence.

[‡]1.5-T magnet.

Funding sources for each contributor (numbered by site): 1: R01 MH085328, R01 MH078160, HD-24061 (Intellectual Disabilities Research Center), M01 RR00052 (Johns Hopkins General Clinical Research Center) and P41 RR15241 (National Center for Research Resources); 3: NSFC (No.30621130074); 4: Chinese Ministry of Science and Technology (No. 2007CB512300); 5: Berlin School of Mind and Brain (DFG); 7: Howard Hughes Medical Institute; 9: National Multiple Sclerosis Society; 13: Competence Net Stroke (BMBF) and Berlin School of Mind and Brain (DFG); 17-18: NIDA (RO1DA016979), NIMH (RO1MH083246), Stavros Niarchos Foundation; 19: NINDS (RO1NS049176); 20: R01 MH064783, R21 MH084031, R01 MH0663674; 21: Academy of Finland (Grant codes 124257, 212181, 214273); 23: Australian Research Council (ARC) Discovery grant (DP0452264); 24: NIH NS053425; Atlanta, G.A.: H.S. Mayberg: URC Grant, Emory University; Durham, NC: D.J. Madden: NIH/NIA R01 AG011622; Durham, NC: L. Wang: Paul B. Beeson Career Developmental Awards (K23-AG028982); Portland, OR: D. Fair/B. Nagel: Oregon Clinical and Translational Research Institute, Medical Research Foundation, UNCF-Merck, Ford Foundation, Dana Foundation; Taipei, Taiwan: C-P. Lin: National Health Research Institute grant (NHRI-EX98-9813EC), Taiwan; Vienna, Austria: C. Windischberger: OeNB-P11468 and OeNB-P12982.