Supplemental Results

Description	HCP field
Visual Episodic Memory	PicSeq_Unadj
Cognitive flexibility (DCCS)	CardSort_Unadj
Inhibition (Flanker task)	Flanker_Unadj
Fluid Intelligence (PMAT)	PMAT24_A_CR
Reading (pronounciation)	ReadEng_Unadj
Vocabulary (picture matching)	PicVocab_Unadj
Processing Speed	ProcSpeed_Unadj
Delay Discounting	DDisc_AUC_40K
Spatial orientation	VSPLOT_TC
Sustained Attention - Sens.	SCPT_SEN
Sustained Attention - Spec.	SCPT_SPEC
Verbal Episodic Memory	IWRD_TOT
Working Memory (list sorting)	ListSort_Unadj
Cognitive status (MMSE)	MMSE_Score
Sleep quality (PSQI)	PSQI_Score
Walking endurance	Endurance_Unadj
Walking Speed	GaitSpeed_Comp
Manual dexterity	Dexterity_Unadj
Grip strength	Strength_Unadj
Odor identificaiton	Odor_Unadj
Pain Interference Survey	PainInterf_Tscore
Taste intensity	Taste_Unadj
Contrast Sensitivity	Mars_Final
Emotional Face Matching	Emotion_Task_Face_Acc
Arithmetic	Language_Task_Math_Avg_Difficulty_Level
Story comprehension	Language_Task_Story_Avg_Difficulty_Level
Relational processing	Relational_Task_Acc
Social Cognition - random	Social_Task_Perc_Random
Social Cognition - interaction	Social_Task_Perc_TOM
Working Memory (n-back)	WM_Task_Acc
Agreeableness (NEO)	NEOFAC_A

Table S1. Lookup table showing the original HCP variable names with the corresponding descriptive labels used in the manuscript. More details of the behavioral measures can be found in the HCP data dictionary.

Description	HCP field
Openness (NEO)	NEOFAC_O
Conscientiousness (NEO)	NEOFAC_C
Neuroticism (NEO)	NEOFAC_N
Extraversion (NEO)	NEOFAC_E
Emot. Recog. - Total	ER40_CR
Emot. Recog. - Angry	ER40ANG
Emot. Recog. - Fear	ER40FEAR
Emot. Recog. - Happy	ER40HAP
Emot. Recog. - Neutral	ER40NOE
Emot. Recog. - Sad	ER40SAD
Anger - Affect	AngAffect_Unadj
Anger - Hostility	AngHostil_Unadj
Anger - Aggression	AngAggr_Unadj
Fear - Affect	FearAffect_Unadj
Fear - Somatic Arousal	FearSomat_Unadj
Sadness	Sadness_Unadj
Life Satisfication	LifeSatisf_Unadj
Meaning & Purpose	MeanPurp_Unadj
Positive Affect	PosAffect_Unadj
Friendship	Friendship_Unadj
Loneliness	Loneliness_Unadj
Perceived Hostility	PercHostil_Unadj
Perceived Rejection	PercReject_Unadj
Emotional Support	EmotSupp_Unadj
Instrument Support	InstruSupp_Unadj
Perceived Stress	PercStress_Unadj
Self-Efficacy	SelfEff_Unadj

Table S1 (cont.). Lookup table showing the original HCP variable names with the corresponding descriptive labels used in the manuscript. More details of the behavioral measures can be found in the HCP data dictionary.

HCP field	Initial search terms	Relevant (final) NeuroSynth term	Task-relevant networks
PicSeq_Unadj	Visual episodic memory	Episodic memory	DefaultA; ControlA
CardSort_Unadj	Cognitive flexibility, dimensional sort, Wisconsin card sorting, task switching	Cognitive control	DorsalAttentionA; VentralAttentionB; ControlA
Flanker_Unadj	Flanker task	Inhibitory control	ControlA; ControlC; VentralAttentionB
PMAT24_A_CR	Fluid intelligence	Intelligence	DorsalAttentionA; ControlA
ReadEng_Unadj	Reading decoding, pronounciation	Reading	DorsalAttentionA; ControlA; TemporalParietal
PicVocab_Unadj	Picture matching, vocabulary comprehension, vocabulary	NA	NA
ProcSpeed_Unadj	Processing speed	NA	NA
DDisc_AUC_40K	Delay discounting	Impulsivity	ControlA; ControlC; VentralAttentionB;
VSPLOT_TC	Spatial orientation, orientation, visual spatial	Visual perception	VisualA
SCPT_SEN	Sustained attention	Attention	DorsalAttentionA; ControlA; VisualA
SCPT_SPEC	Sustained attention	Attention	DorsalAttentionA; ControlA; VisualA
IWRD_TOT	Verbal episodic memory, word memory	Episodic memory	DefaultA; ControlA
ListSort_Unadj	Working memory	Working memory	DorsalAttentionA; ControlA;

Table S2. Task-relevant networks of 13 cognitive measures based on NeuroSynth database. The “initial search terms” were our initial queries in the NeuroSynth database for each cognitive measure. The “final search terms” were the terms that we finally utilized to retrieve the forward inference map. There was no appropriate search term for PicVocab_Unadj (picture vocabulary) and ProcSpeed_Unadj (processing speed). Each forward map was projected to fsLR surface space and compared with the group-level parcellation estimated from the HCP training set (Figure 2A) to select the task-relevant networks.

		Minimally correlated set of 5 behaviors	All 58 behaviors	13 Cognitive measures
Accuracy across 100 20-fold CVs (mean ± std)	**MS-HBM**	**0.1327±0.0065**	**0.0803±0.0032**	**0.1321±0.0053**
	YeoBackProject	0.1036±0.0080	0.0728±0.0032	0.1057±0.0060
	Gordon2017	0.0830±0.0080	0.0655±0.0036	0.0545±0.0062
	Wang2015	0.1039±0.0080	0.0755±0.0033	0.1202±0.0054
Percentage improvement (#CV splits MS-HBM outperforms other approaches)	**MS-HBM vs YeoBackProject**	28.55% (100/100)	10.35% (100/100)	25.17% (100/100)
	MS-HBM vs Gordon2017	61.07% (100/100)	22.88% (100/100)	145.12% (100/100)
	MS-HBM vs Wang2015	28.20% (100/100)	6.43% (95/100)	10.05% (97/100)

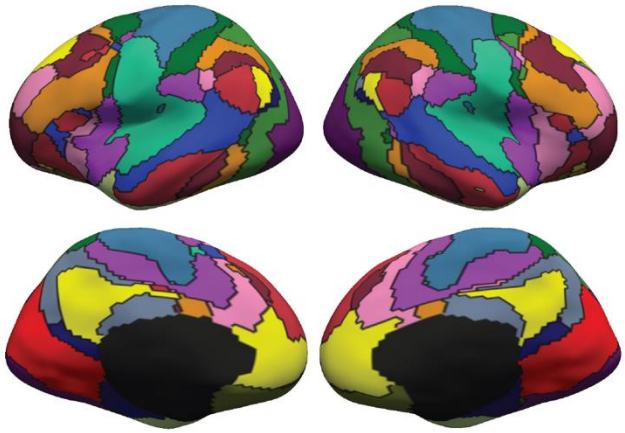
		NEO-5	Emotional recognition (ER)	Emotional measures excluding ER
Accuracy across 100 20-fold CVs (mean ± std)	**MS-HBM**	0.0955±0.0085	-0.0445±0.0101	**0.1038±0.0070**
	YeoBackProject	**0.1027±0.0084**	-0.0170±0.0092	0.0781±0.0070
	Gordon2017	0.0820±0.0080	**0.0297±0.0092**	0.0794±0.0072
	Wang2015	0.0772±0.0090	0.0068±0.0093	0.0802±0.0069
Percentage improvement (#CV splits MS-HBM outperforms other approaches)	**MS-HBM vs YeoBackProject**	-6.82% (16/100)	-264.70% (0/100)	33.46% (100/100)
	MS-HBM vs Gordon2017	17.20% (94/100)	-269.91% (0/100)	31.36% (100/100)
	MS-HBM vs Wang2015	24.61% (99/100)	-913.27% (0/100)	29.85% (100/100)

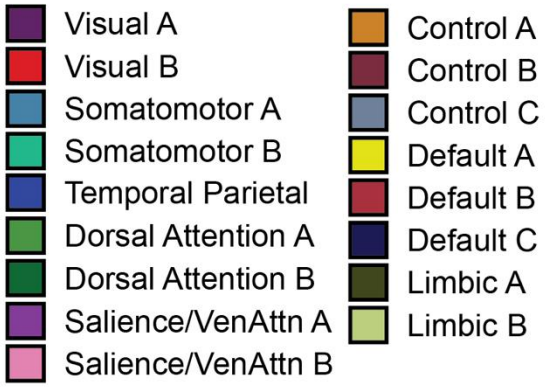
Table S3. Average prediction accuracies for different sets of behavioral measures (minimally correlated set of 5 behaviors, 58 behavioral measures, 13 cognitive measures, NEO-5 personality measures, emotion recognition measures and emotional measures) across different parcellation approaches. Prediction was based on individual-specific network topography. The mean accuracy and standard deviation was calculated across 100 20-fold cross-validations. The percentage improvement and number of cross-validations that MS-HBM outperforms other approaches across 100 20-fold cross-validations was reported.

	Accuracy (mean ± std) across 100 20-fold cross-validations			
	MS-HBM	**YeoBackProject**	**Gordon**	**Wang**
DVARS	0.3157±0.0182	0.4283±0.0169	0.2115±0.0180	0.3621±0.0175
FD	0.2120±0.0153	0.2733±0.0140	0.2565±0.0136	0.2066±0.0171

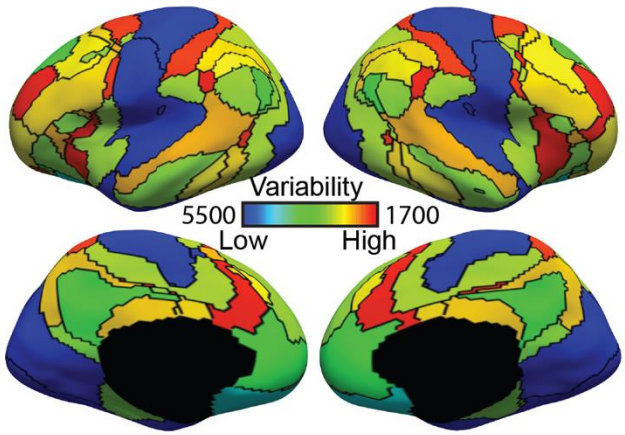
Table S4. Average prediction accuracies of DVARS and FD across different parcellation approaches. Prediction was performed based on individual-specific network topography without regressing any nuisance covariates. The mean accuracy and standard deviation was calculated across 100 20-fold cross-validations.

(A) Group parcellation





(B) Inter-subject RSFC variability



(C) Intra-subject RSFC variability

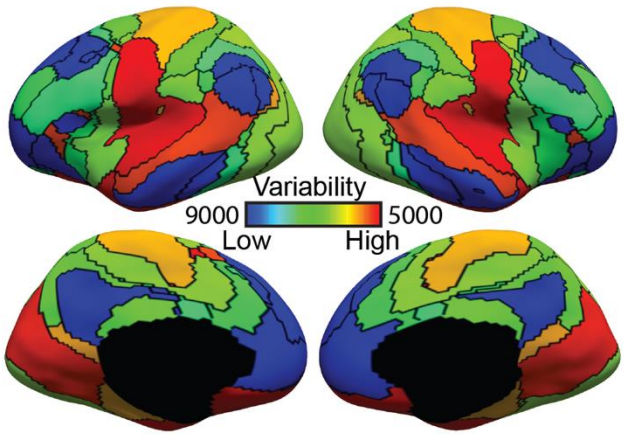


Figure S1. Sensory-motor networks exhibit lower inter-subject, but higher intra-subject, functional connectivity variability than association networks in the GSP training set. (A) 17-network group-level parcellation. (B) Inter-subject functional connectivity variability for different cortical networks. (C) Intra-subject functional connectivity variability for different cortical networks. Note that (B) and (C) correspond to the ϵ_i and σ_i parameters in Figure 1, where higher values indicate lower variability. .

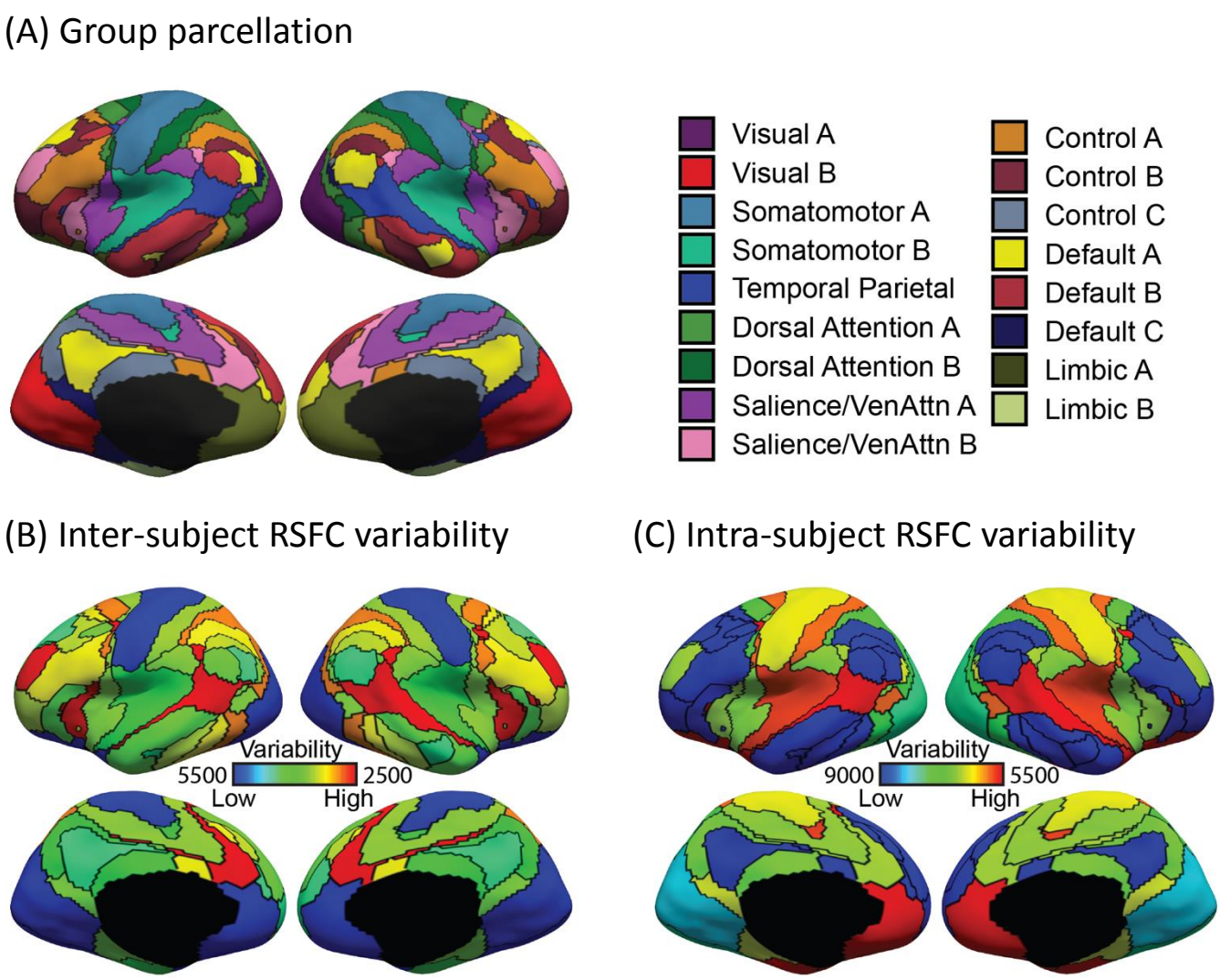


Figure S2. Sensory-motor networks exhibit lower inter-subject, but higher intra-subject, functional connectivity variability than association networks in the CoRR-HNU dataset. (A) 17-network group-level parcellation. (B) Inter-subject functional connectivity variability for different cortical networks. (C) Intra-subject functional connectivity variability for different cortical networks. Note that (B) and (C) correspond to the ϵ_i and σ_i parameters in Figure 1, where higher values indicate lower variability.

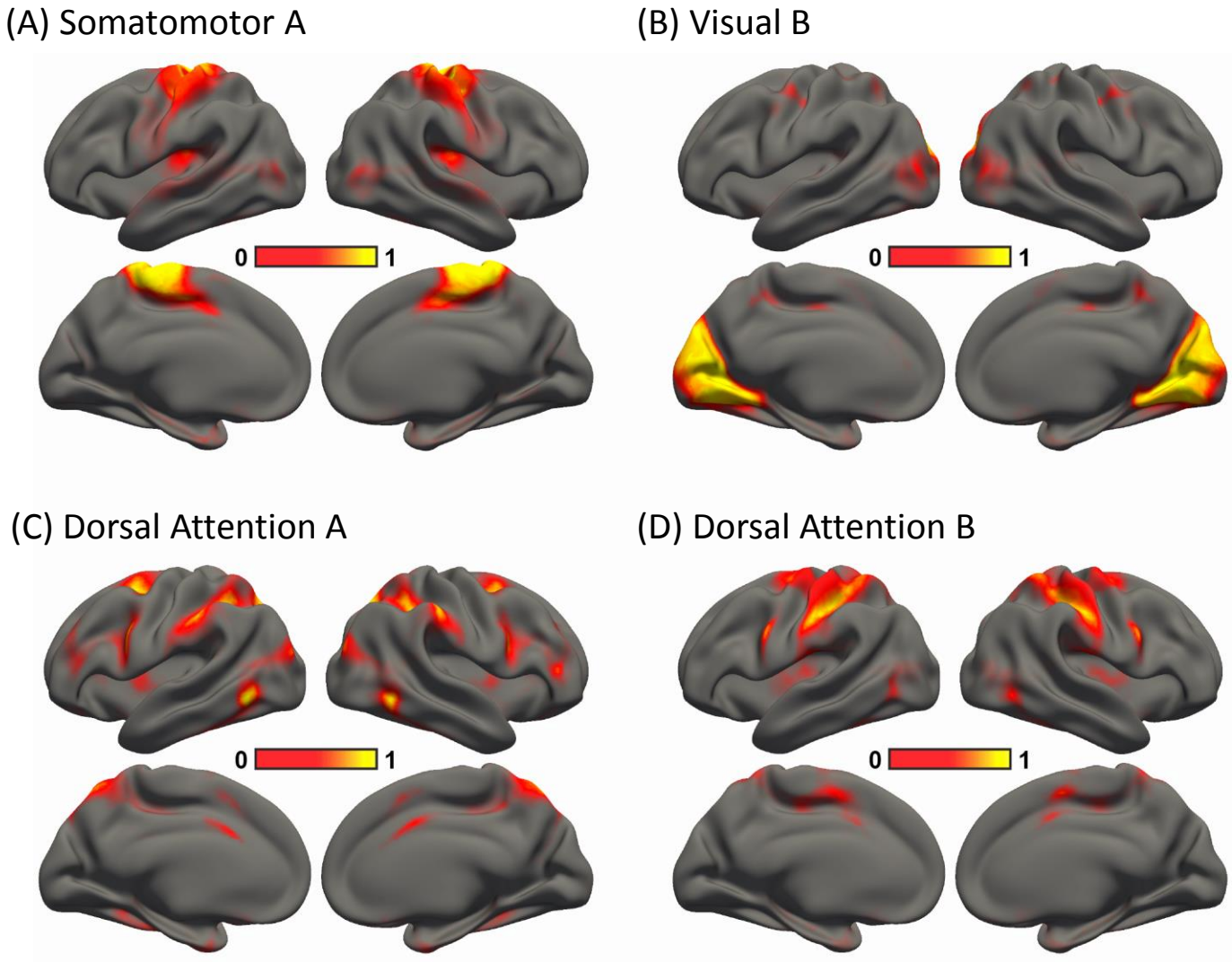


Figure S3. Sensory-motor networks are less spatially variable than association networks across subjects in the HCP training set. Spatial probability maps of (A) Somatomotor network A, (B) Visual network B, (C) Dorsal Attention network A, and (D) Dorsal Attention network B. A higher value (bright color) at a spatial location indicates high probability of a network appearing at that spatial location. Results were replicated in the GSP (Figure S4) and Corr-HNU (Figure S5) datasets. Note that this corresponds to the Θ_l parameter in Figure 1.

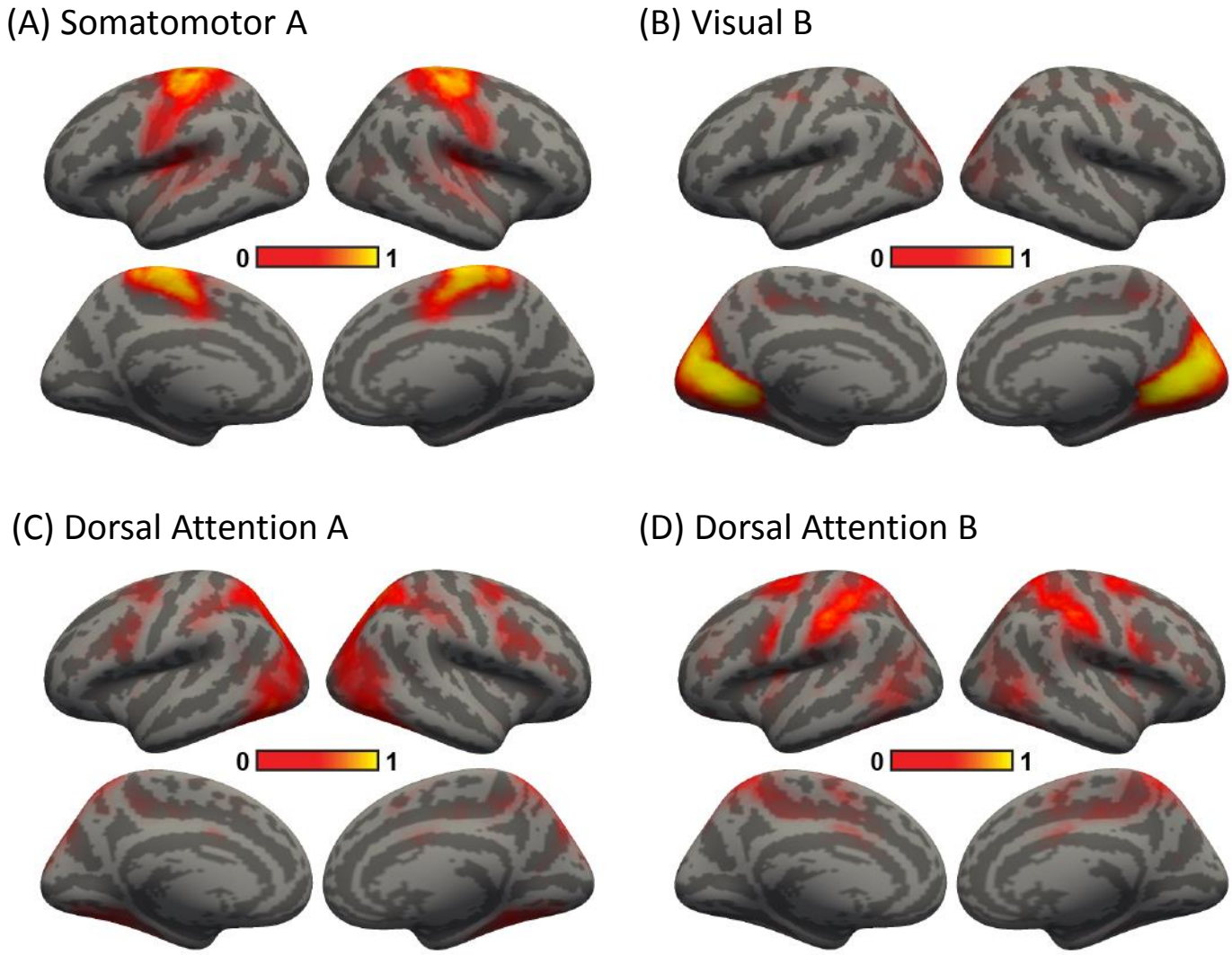


Figure S4. Sensory-motor networks are less spatially variable than association networks across subjects in the GSP dataset. Spatial probability maps of (A) Somatomotor network A, (B) Visual network B, (C) Dorsal Attention network A, and (D) Dorsal Attention network B. A higher value (bright color) at a spatial location indicates high probability of a network appearing at that spatial location. Note that this corresponds to the Θ_l parameter in Figure 1.

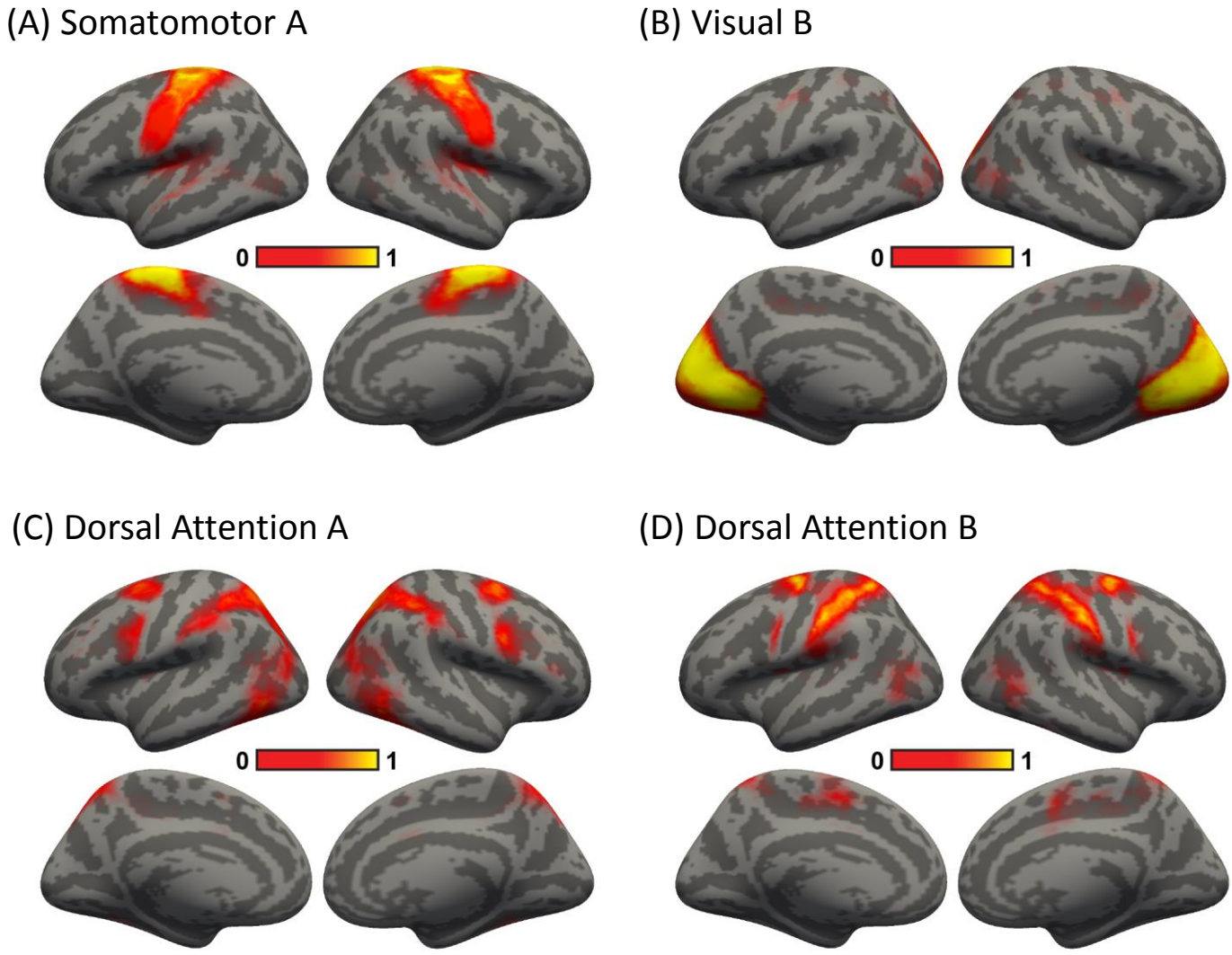


Figure S5. Sensory-motor networks are less spatially variable than association networks across subjects in the CoRR-HNU dataset. Spatial probability maps of (A) Somatomotor network A, (B) Visual network B, (C) Dorsal Attention network A, and (D) Dorsal Attention network B. A higher value (bright color) at a spatial location indicates high probability of a network appearing at that spatial location. Note that this corresponds to the Θ_l parameter in Figure 1.

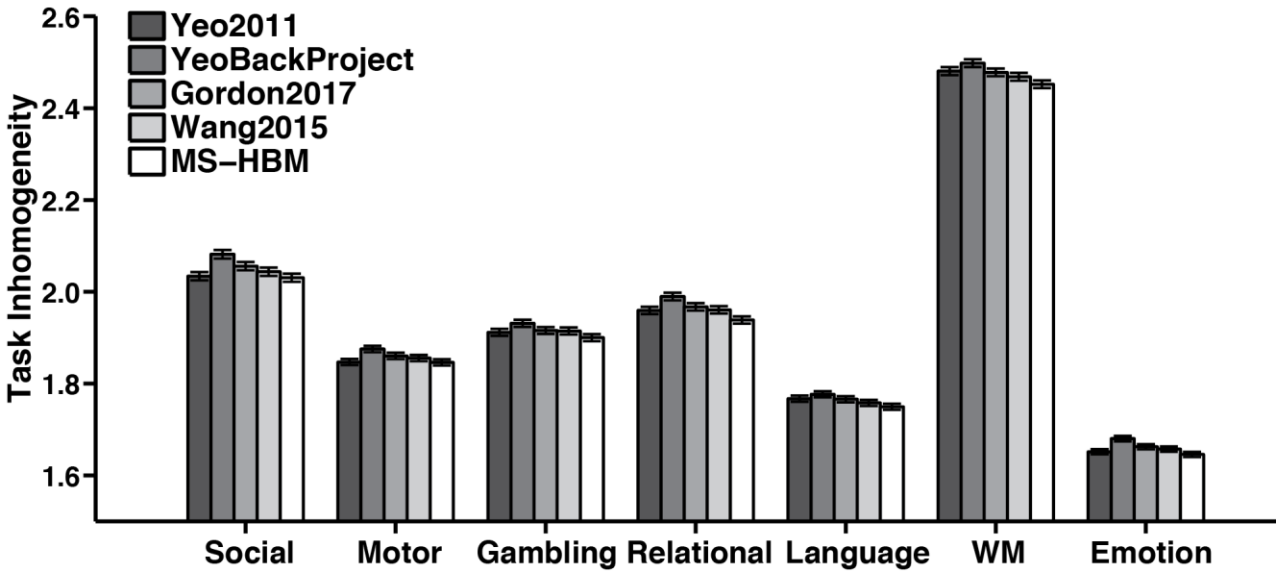


Figure S6. Task inhomogeneity of resting-state parcellations in the HCP dataset. 17-network individual-specific parcellations were estimated using one rs-fMRI session. Task inhomogeneity was then defined as the standard deviation of task activation within each network, and then averaged across all networks and contrasts within each behavioral domain. Lower value indicates better functional homogeneity. Error bars correspond to standard errors. Compared with Yeo2011, YeoBackProject, Gordon2017 and Wang2015, the MS-HBM individual-specific parcellations achieved a modest average improvement 0.63% (Cohen's d = 0.12, 0.09, 0.66, 1.0, 0.9, 1.1, 0.46 for social, motor, gambling, relational, language, working memory and emotion respectively), 2.0% (Cohen's d > 1.3 for all domains), 1.04% (Cohen's d > 0.99 for all domains) and 0.7% (Cohen's d > 0.79 for all domains) respectively.

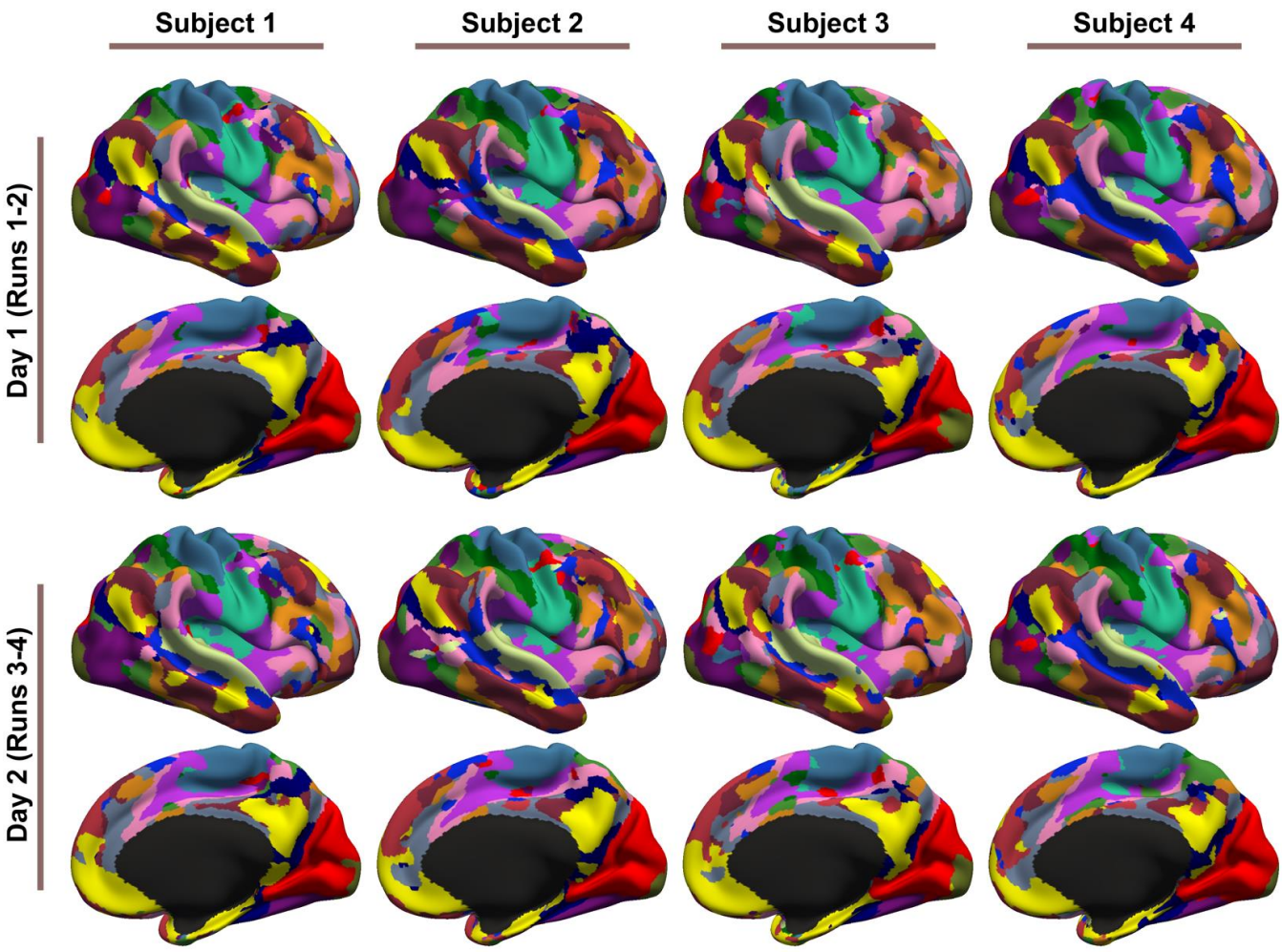


Figure S7. 17-network parcellations were estimated using runs 1-2 and runs 3-4 separately for each subject from the HCP test set. Parcellations of four representative subjects are shown here. Left hemisphere parcellations are shown in Figure 4.

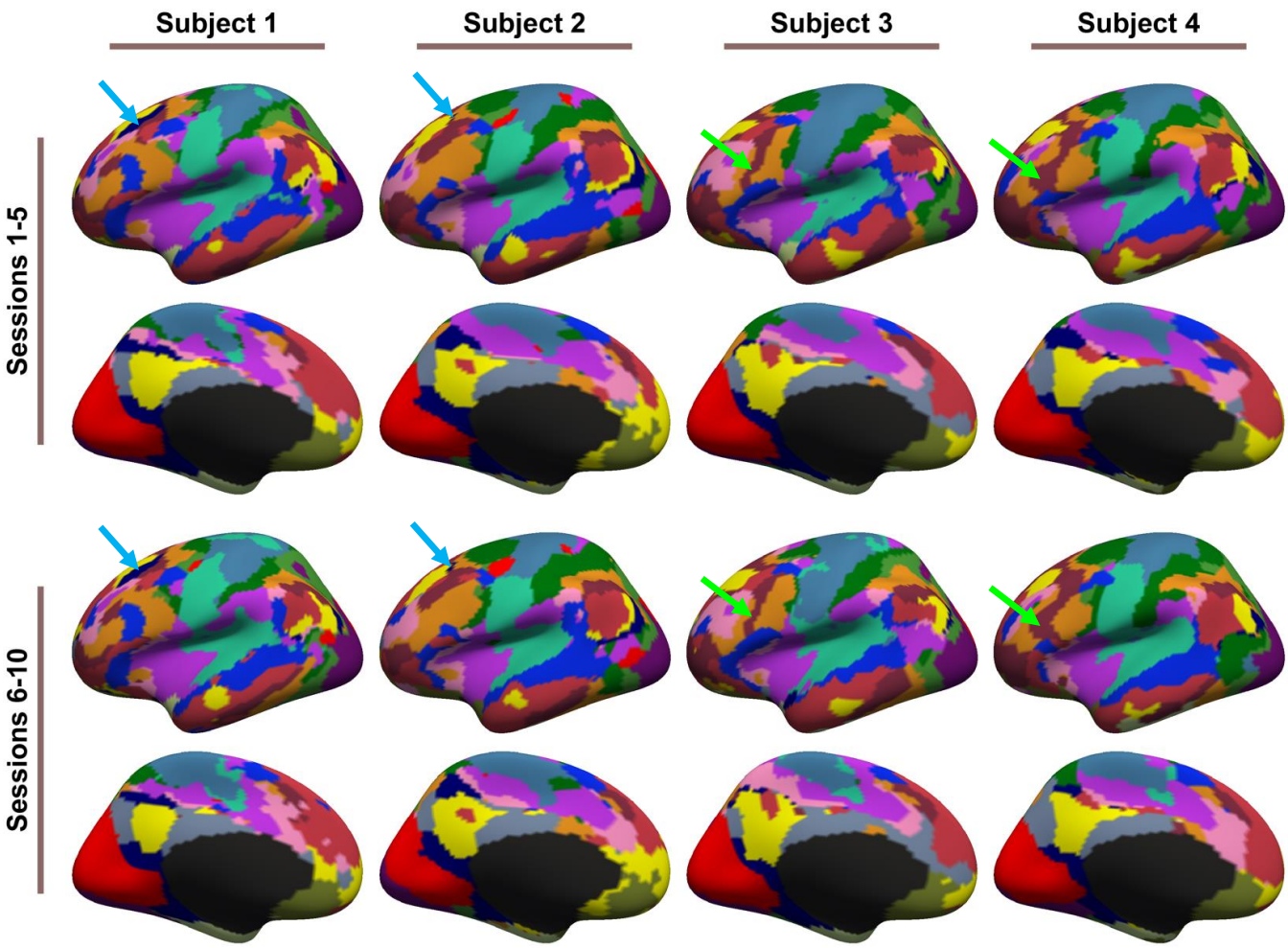


Figure S8. 17-network parcellations were estimated using sessions 1-5 and sessions 6-10 separately for each subject from the CoRR-HNU dataset. Parcellations of four representative subjects are shown here. Black and green arrows indicate individual-specific parcellation features. The Default C (dark blue) network exhibited a dorsal prefrontal component for certain subjects (blue arrows), but was missing in other subjects. As another example, the lateral prefrontal component of the Control A (orange) network was separated into two separate components by the Control B (brown) network (green arrows). These features were mostly replicated across sessions. Right hemisphere parcellations are shown in Figure S9.

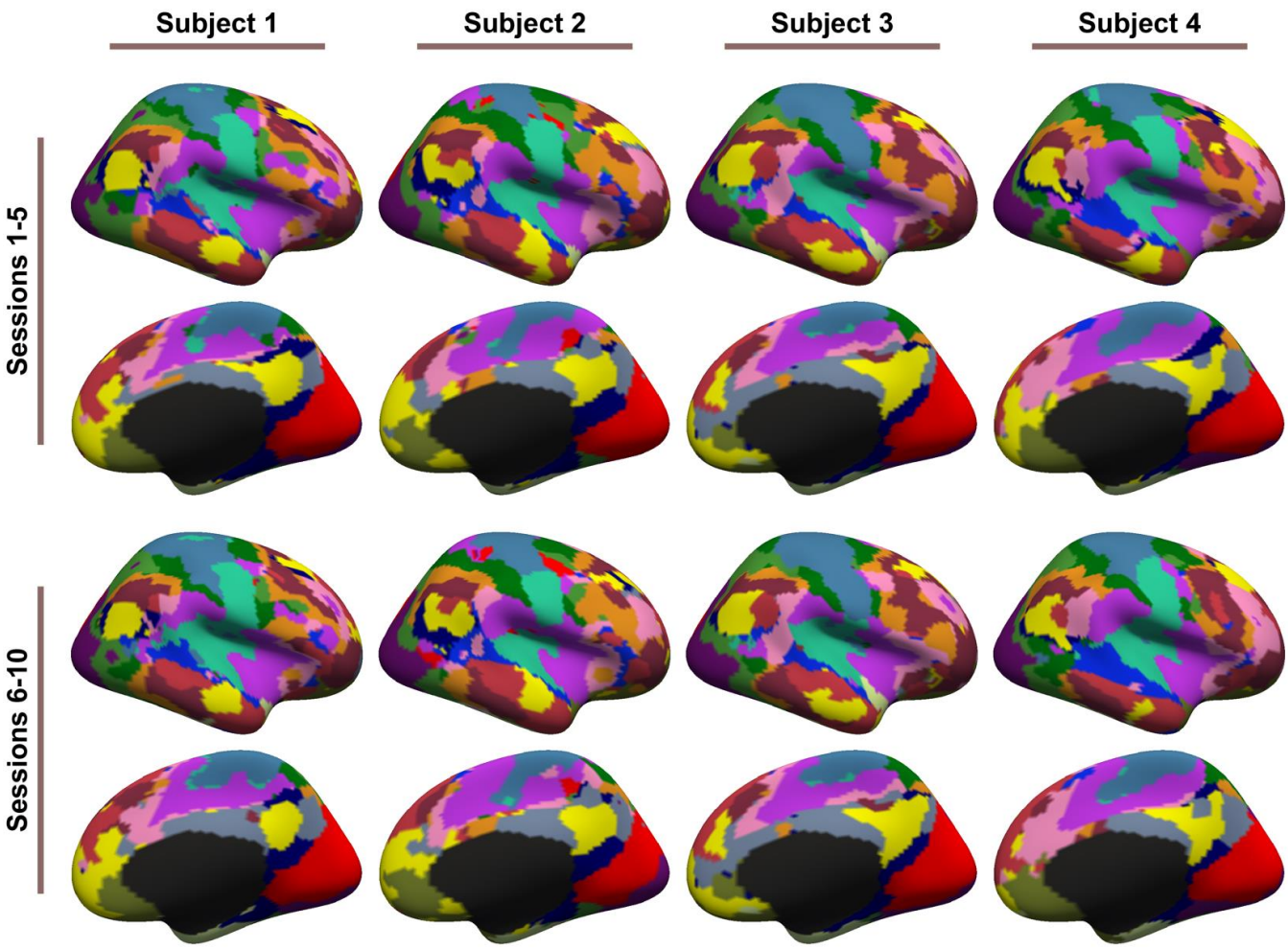


Figure S9. 17-network parcellations were estimated using sessions 1-5 and sessions 6-10 separately for each subject from the CoRR-HNU dataset. Parcellations of four representative subjects are shown here. Left hemisphere parcellations are shown in Figure S8.

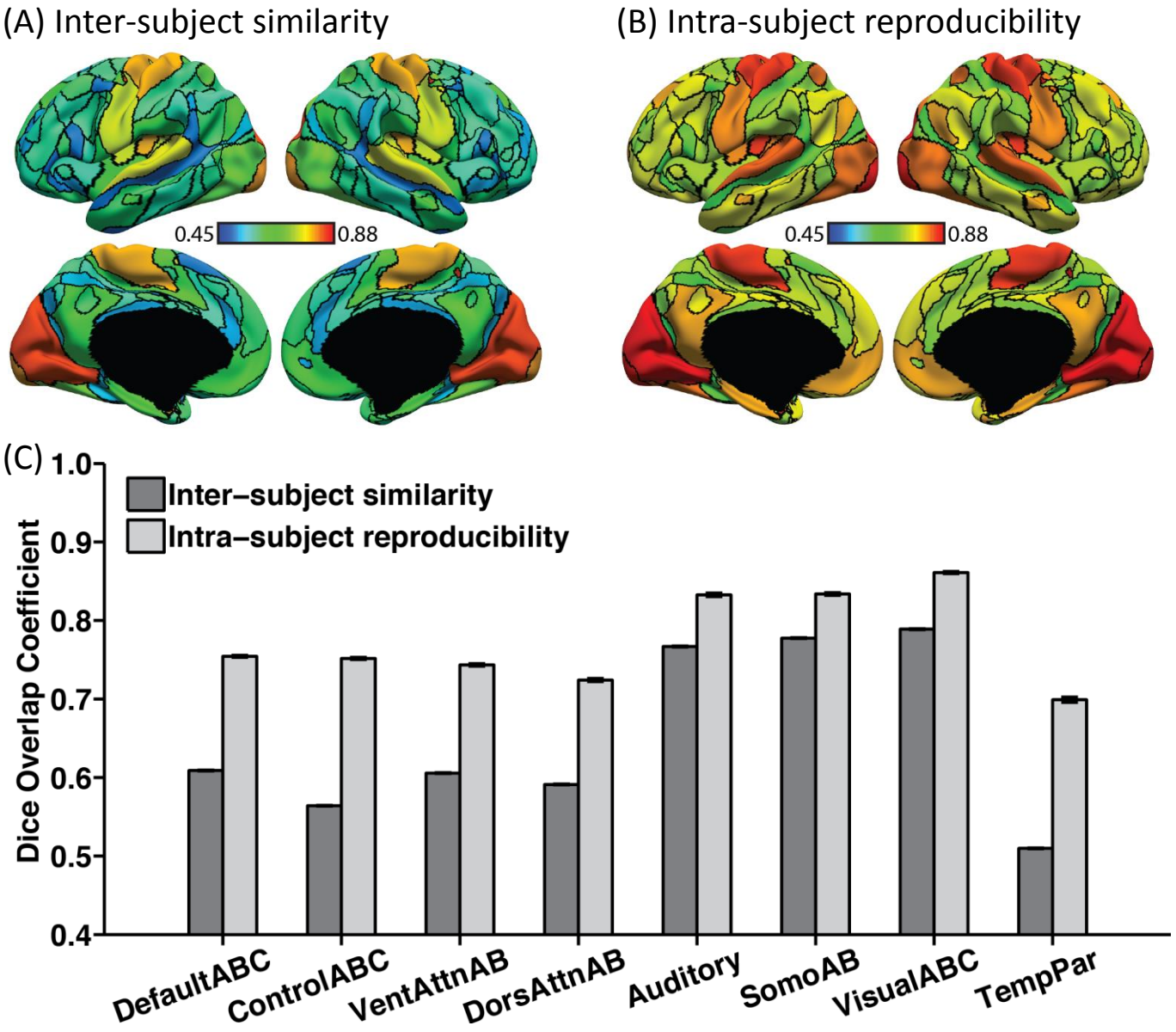


Figure S10. Individual-specific MS-HBM parcellations show high within-subject reproducibility and low across-subject similarity in the HCP test set. Individual-specific MS-HBM parcellations were generated by using the first two runs (day 1) and last two runs (day 2) separately for each subject. (A) Inter-subject spatial similarity for different networks. (B) Intra-subject reproducibility for different networks. Warm color indicates higher overlap. Cool color indicates lower overlap. (C) Quantification of inter-subject similarity and intra-subject reproducibility for different networks. “VentAttnAB” corresponds to Salience/Ventral Attention (A and B) networks. “SomoAB” corresponds to Somatomotor (A and B) networks. Error bars correspond to standard errors.

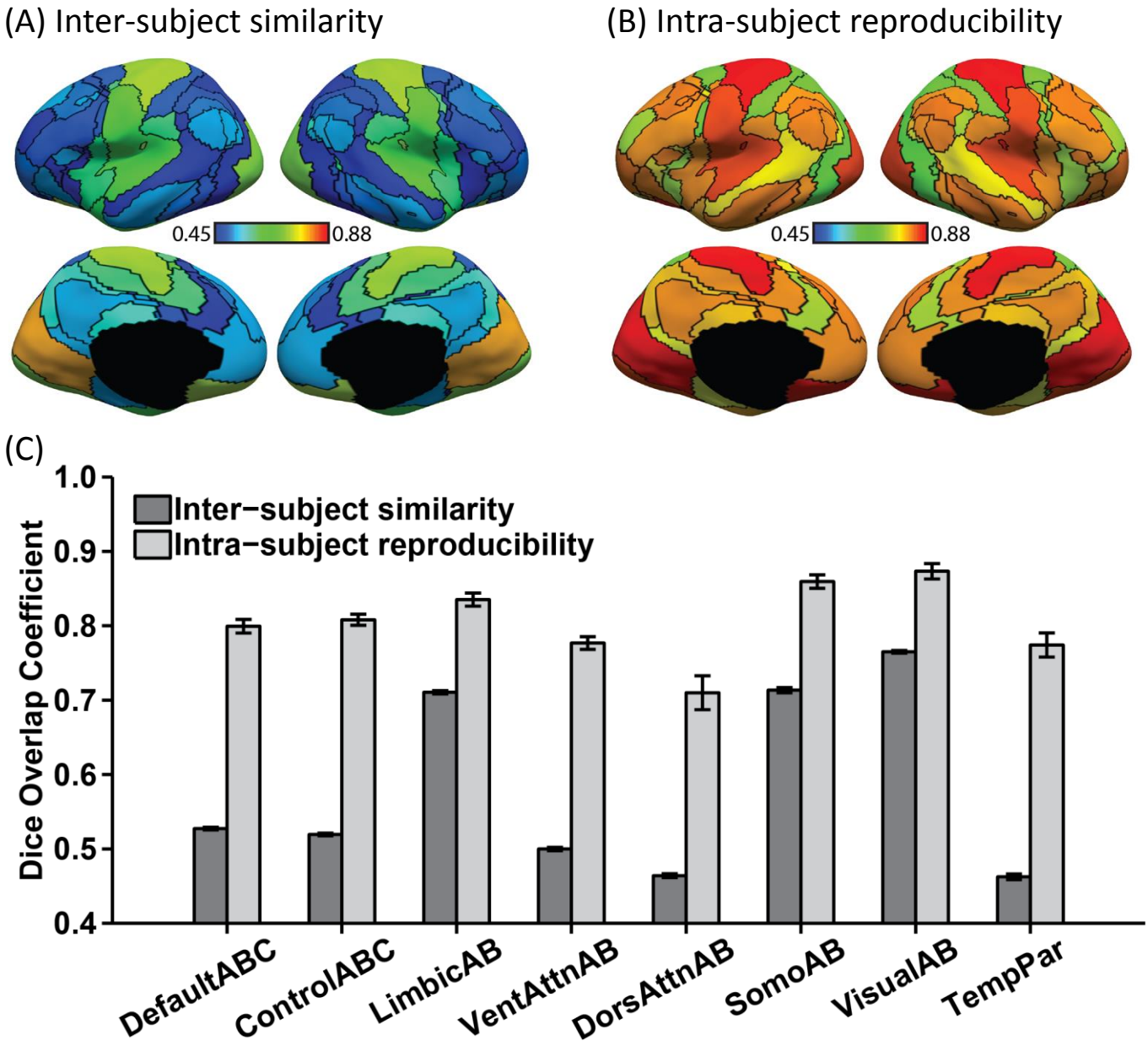


Figure S11. Individual-specific MS-HBM parcellations show high within-subject reproducibility (overlap = 81.6%) and low across-subject similarity (overlap = 59.4%) in the CoRR-HNU dataset. (A) Inter-subject spatial similarity for different networks. (B) Intra-subject reproducibility for different networks. Warm color indicates higher overlap. Cool color indicates lower overlap. (C) Quantification of inter-subject similarity and intra-subject reproducibility for different networks. “VentAttnAB” corresponds to Salience/Ventral Attention networks A and B. “SomoAB” corresponds to Somatomotor networks A and B. Error bars correspond to standard errors.

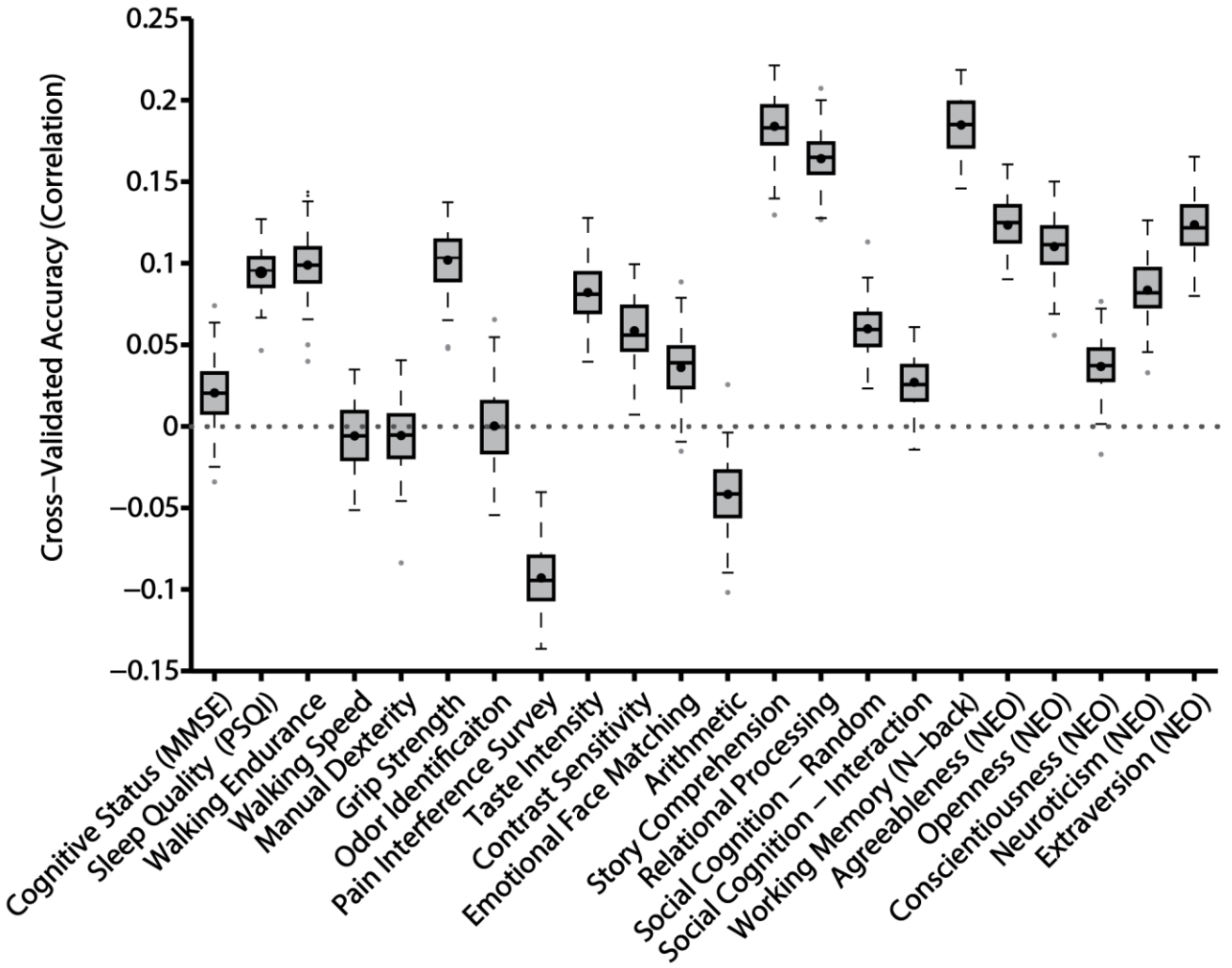


Figure S12. Prediction accuracy of 22 cognitive, emotion, personality and other non-imaging measures based on inter-subject differences in the spatial arrangement of cortical networks. Boxplots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean. In the case of the NEO-5 personality scores, average predication accuracy was r = 0.0955 ± 0.0085 (mean ± std). Other measures are found in Figures 5 and S13.

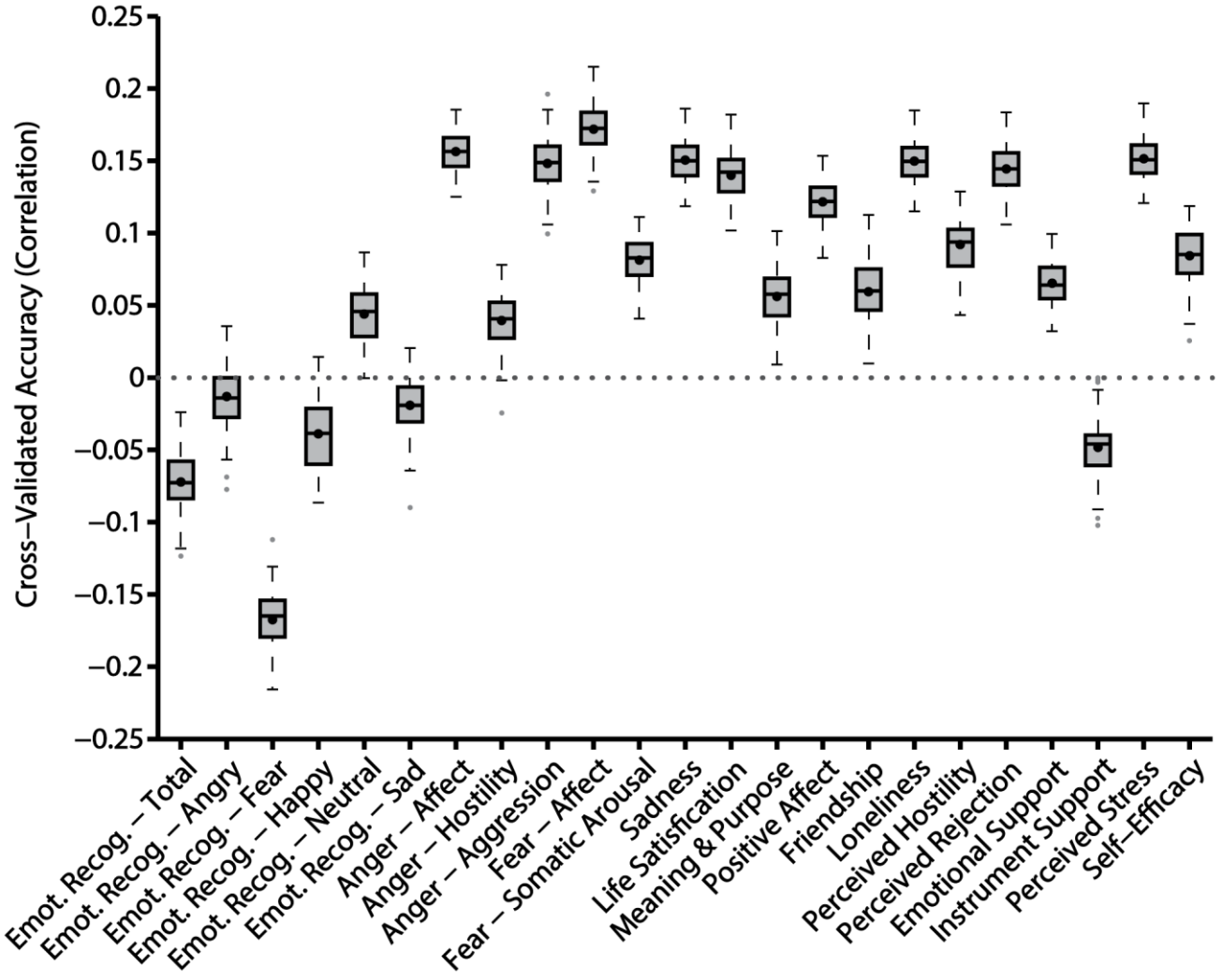
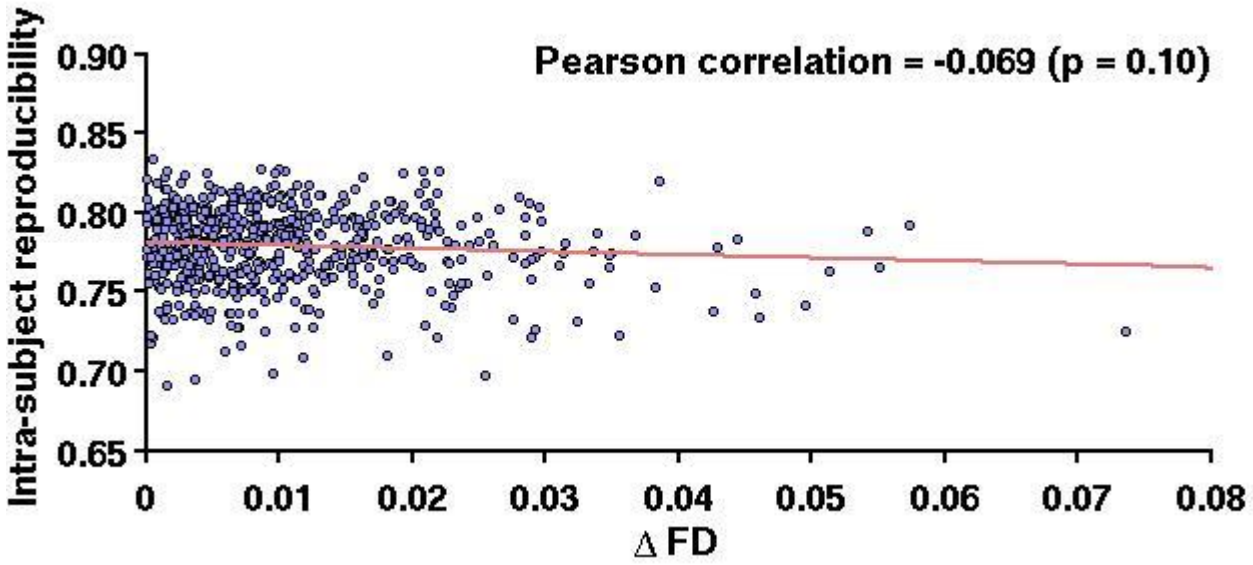


Figure S13. Prediction accuracy of 23 cognitive, emotion, personality and other non-imaging measures based on inter-subject differences in the spatial arrangement of cortical networks. Boxplots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean. In the case of the emotional measures (all items in Figure S15 except for emotional recognition), the average prediction accuracy was r = 0.1038 ± 0.0070 (mean ± std). Other measures are found in Figures 5 and S12. Interestingly, prediction accuracy for the emotion recognition task was poor.

(A)



(B)

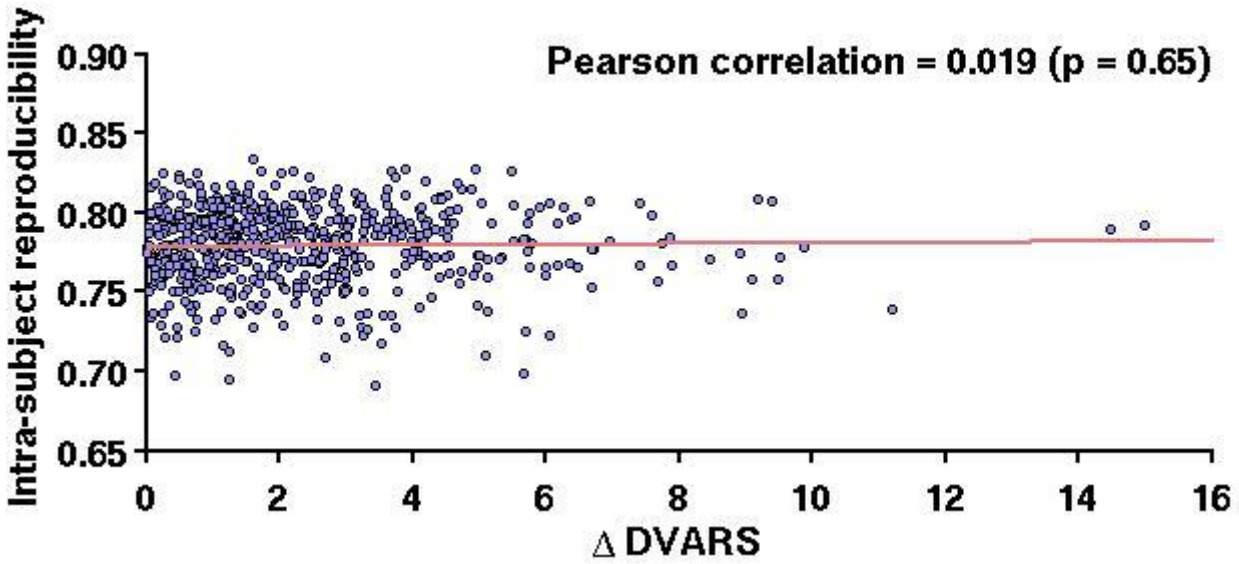


Figure S14. Intra-subject reproducibility is not significantly affected by FD or DVARS difference between the two scan days of the HCP subjects. (A) Scatterplot of intra-subject reproducibility versus absolute FD difference between the two scan days. (B) Scatterplot of intra-subject reproducibility versus absolute DVARS difference between the two scan days. Only HCP test set subjects with all 58 behavioral measures were considered. If network topography was corrupted by motion-related imaging artifacts, then one would expect subjects who moved by very different amounts on the two different scan days to have poorer intra-subject parcellation reproducibility than subjects who moved by similar amounts on both days. As shown above, this is not the case.

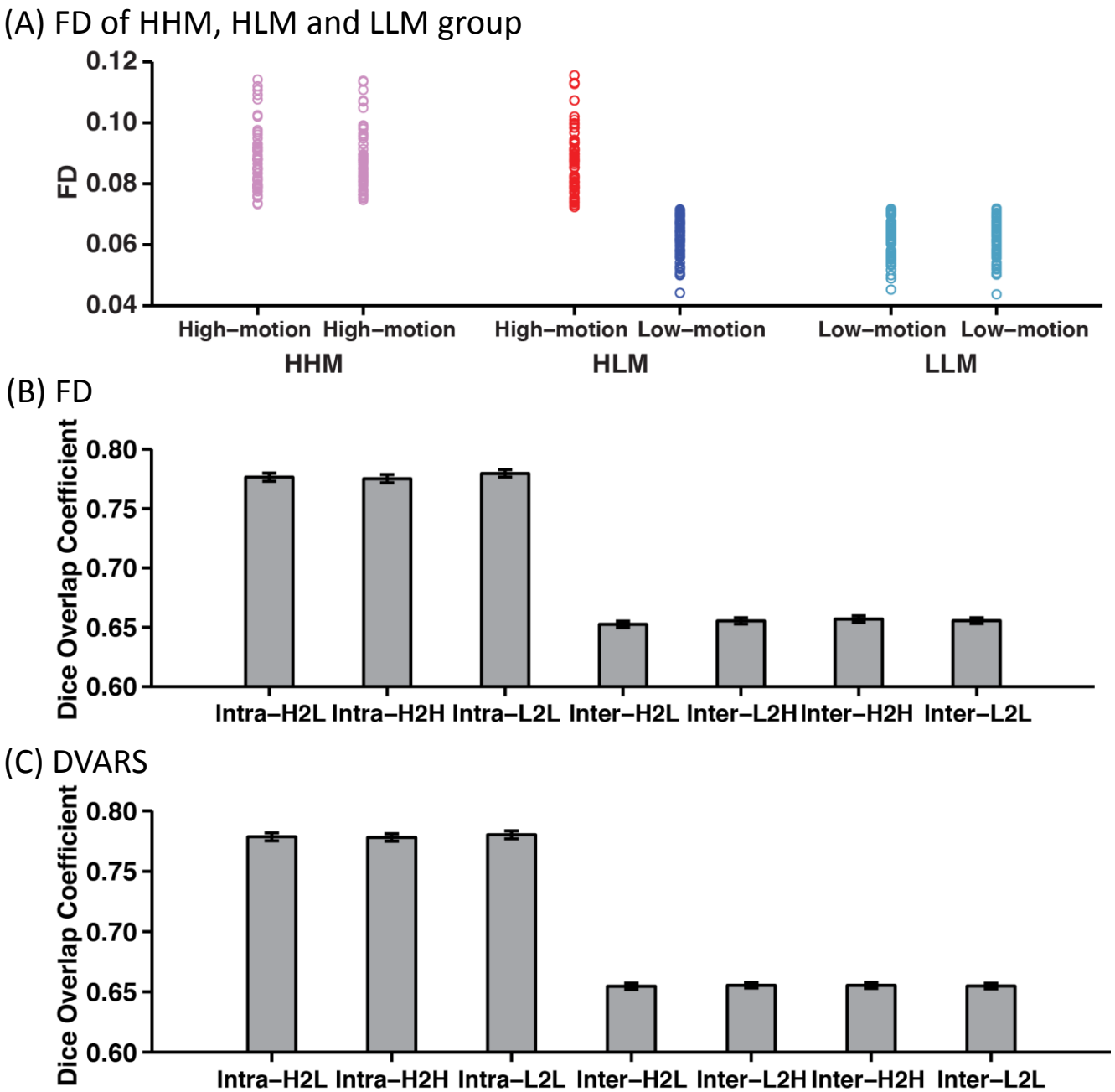


Figure S15. Motion-related imaging artifacts have little effect on intra-subject reproducibility and inter-subject similarity. (A) One issue with the previous analysis (Figure S14) is that a subject with high absolute FD difference between the two scan days might still exhibit low motion on both days (relative to other subjects). Therefore, we also considered three groups of 60 HCP subjects. The high-low-motion (HLM) subjects consisted of subjects, whose FD were above the median FD in one day, and below the median FD in another day. The high-high-motion (HHM) group consisted of subjects whose FD were high on both days and matched the FD of the HLM subjects on the high FD days. The low-low-motion (LLM) motion group consisted of subjects whose FD were low on both days and matched the FD of the HLM subjects on the low FD days. (B) Intra-subject parcellation reproducibility were basically identical among HLM subjects (Intra-H2L), HHM subjects (Intra-H2H) and LLM subjects (Intra-L2L). For comparisons, the inter-subject parcellation similarity between HHM subjects and HLM subjects during their high motion days (Inter-H2H), between LLM subjects and HLM subjects during their low motion days (Inter-L2L), between HHM subjects and HLM subjects during their low motion days (Inter-H2L) and between LLM subjects and HLM subjects during their high motion days (Inter-L2H) were basically identical and significantly lower than intra-subject parcellation reproducibility. (C) Similar results were obtained with DVARS.

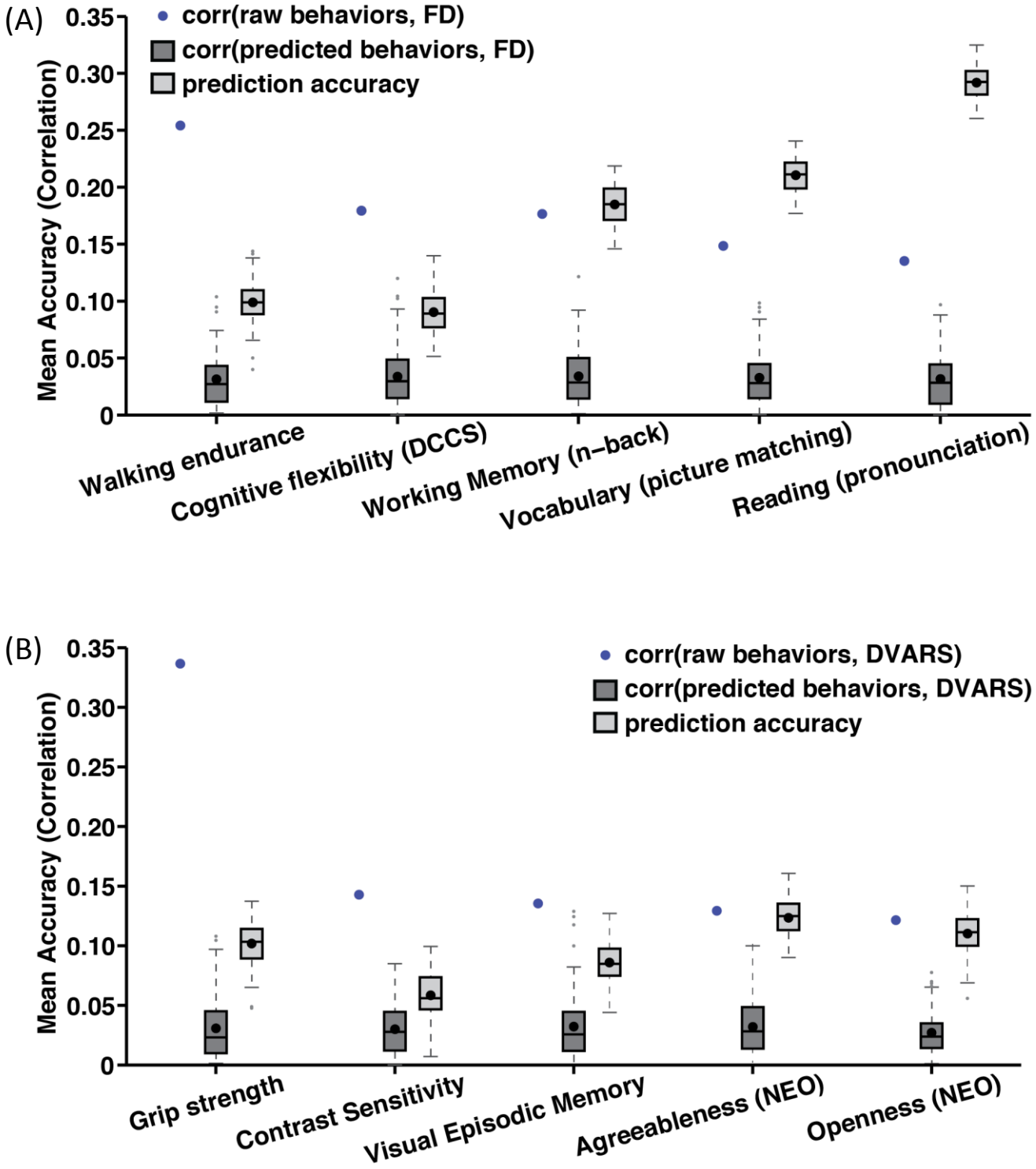


Figure S16. Prediction accuracies of 5 behavioral measures with the highest correlation with (A) FD and (B) DVARS. The correlation between the behavioral measures and FD/DVARS was depicted as the blue dot. Boxplots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean. The prediction accuracies of these 5 behaviors were higher than the correlation between the prediction of these 5 behavioral measures and FD/DVARS.

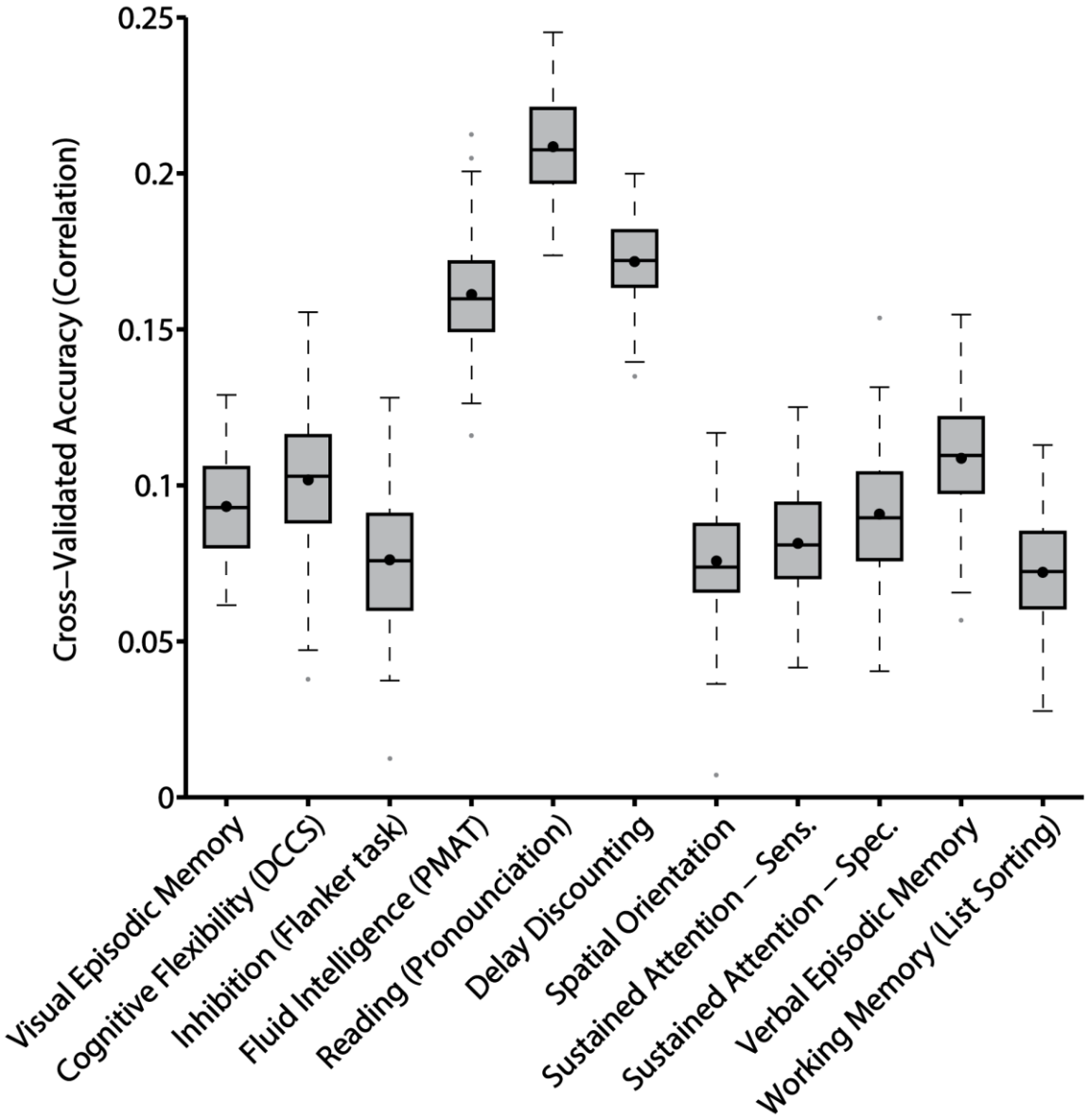


Figure S17. Prediction accuracy of 11 cognitive measures based on inter-subject differences in the spatial arrangement of task-relevant cortical networks. Boxplots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean. Average prediction accuracy based only on task-relevant networks was r = 0.1129 ± 0.0062 (mean ± std).

Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality and Emotion

Supplemental Material

This supplemental material is divided into *Supplemental Methods* and *Supplemental Results* to complement the Methods and Results sections in the main text, respectively.

Supplementary Methods

This section provides additional information about preprocessing, multi-session hierarchical Bayesian model (MS-HBM) and alternative parcellation algorithms that we compared with. Section S1 provides more details about the preprocessing of the GSP and CoRR-HNU data. Section S2 provides more details about the preprocessing of the HCP data. Section S3 provides mathematical details about the MS-HBM. Section S4 describes the algorithms for estimating group-level priors and deriving the individual-specific parcellations and how "free" parameters of the model are set. Section S5 provides more details about the alternative parcellation algorithms that were compared against MS-HBM. Section S6 describes the kernel ridge regression model for behavioral prediction.

S1. Processing of GSP and CoRR-HNU data

Structural data were processed using FreeSurfer. FreeSurfer constitutes a suite of automated algorithms for reconstructing accurate surface mesh representations of the cortex from individual subjects' T1 images (Dale et al., 1999; Fischl et al., 2001; Ségonne et al., 2007). The cortical surface meshes were then registered to a common spherical coordinate system (Fischl et al. 1999a; 1999b). The GSP subjects were processed using FreeSurfer 4.5.0 (Holmes et al., 2015), while the CoRR-HNU subjects were processed using FreeSurfer 5.3.0.

Resting-state fMRI data of GSP and CoRR-HNU were initially pre-processed with the following steps: (i) removal of first 4 frames, (ii) slice time correction with the FSL package (Jenkinson et al., 2002; Smith et al., 2004), (iii) motion correction using rigid body translation and rotation with the FSL package. The structural and functional images were aligned using boundary-based registration (Greve and Fischl 2009) using the FsFast software package (http://surfer.nmr.mgh.harvard.edu/fswiki/FsFast).

Framewise displacement (FD) and voxel-wise differentiated signal variance (DVARS) were computed using fsl_motion_outliers (Smith et al., 2004). Volumes with FD > 0.2mm or DVARS > 50 were marked as outliers. Uncensored segments of data lasting fewer than 5 contiguous volumes were

also flagged as outliers (Gordon et al., 2016). BOLD runs with more than half of the volumes flagged as outliers were removed completely. For the CoRR-HNU dataset, no session (and therefore no subject) was removed. For the GSP subjects, only one run was removed (out of a total of 222 runs). No individuals in the GSP dataset lost an entire session, and therefore, all subjects were retained.

Linear regression using multiple nuisance regressors was applied. Nuisance regressors consisted of global signal, six motion correction parameters, averaged ventricular signal, averaged white matter signal, as well as their temporal derivatives (18 regressors in total). The flagged outlier volumes were ignored during the regression procedure. The data were interpolated across censored frames using least squares spectral estimation of the values at censored frames (Power et al., 2014). Finally, a band-pass filter ($0.009\ Hz \leq f \leq 0.08\ Hz$) was applied.

The preprocessed fMRI data was projected onto the FreeSurfer fsaverage6 surface space (2mm vertex spacing). The projected fMRI data was smoothed using a 6mm full-width half-maximum (fwhm) kernel and then downsampled onto fsaverage5 surface space (4mm vertex spacing). Smoothing on the fsaverage6 surface, rather than in the volume minimized the blurring of fMRI signal across sulci.

S2. Processing of HCP data

Details of the HCP preprocessing can be found elsewhere (HCP S900 manual; Van Essen et al. 2012b; Glasser et al. 2013; Smith et al. 2013). Of particular importance is that the rs-fMRI data has been projected to the fs_LR surface space (Van Essen et al. 2012a), smoothed by 2mm fwhm and denoised with ICA-FIX (Salimi-Khorshidi et al. 2014; Griffanti et al., 2014) and aligned with MSMAll (Robinson et al., 2014).

However, recent studies have shown that ICA-FIX does not fully eliminate global and head-motion related artifacts (Burgess et al., 2016; Siegel et al., 2016). Therefore, further processing steps were performed on the rs-fMRI data in fs_LR surface after ICA-FIX denoising, which included nuisance regression, motion censoring and interpolation, and band-pass filtering. Volumes with FD > 0.2mm or DVARS > 75, as well as uncensored segments of data lasting fewer than 5 contiguous volumes were flagged as outliers. BOLD runs with more than half the volumes flagged as outliers were completely removed. Consequently, 56 subjects were removed. Furthermore, for this work, only subjects with all four runs remaining (N = 676) were considered.

Nuisance regression utilized regressors consisting of global signal, six motion parameters, averaged ventricular signal, averaged white matter signal, and their temporal derivatives (18 regressors in total). The outlier volumes were ignored during the regression procedure. The data were interpolated across censored frames using least squares spectral estimation (Power et al., 2014). A band-pass filter

(0.009 Hz $\leqslant$ f $\leqslant$ 0.08 Hz) was then applied to the data. Finally, the data was smoothed by 4mm fwhm. Given that the HCP team has already smoothed the data by 2mm, this results in an effective smoothing of 6mm fwhm.

S3.Mathematical model (MS-HBM)

In this section, we describe our model for individual-level parcellation of the cerebral cortex. We assume a common surface coordinate system, where the cerebral cortex is represented by left and right hemisphere spherical meshes such as FreeSurfer fsaverage surface meshes. Each mesh consists of a collection of vertices and edges connecting neighboring vertices into triangles (https://en.wikipedia.org/wiki/Triangle_mesh).

Let N denote the total number of vertices, T denote the number of resting-state fMRI (rs-fMRI) sessions, S denote the number of subjects, L denote the number of networks, and $\mathcal{N}_n$ denote the neighboring vertices of vertex n (as defined by the cortical mesh). For each subject s and session t, there is a preprocessed rs-fMRI time course associated with each vertex n. For each subject s, there is an unknown parcellation label l_n^s at vertex n. Note that the parcellation label is assumed to be the same across sessions (hence there is no index on the session). In this work, we use $1{:}S$ to denote a set of subjects $\{1, 2, \dots, S\}$, $1{:}T$ to denote a set of sessions $\{1, 2, \dots, T\}$, $1{:}N$ to denote a set of vertices $\{1, 2, \dots, N\}$, $1{:}L$ to denote a set of parcellation labels $\{1, 2, \dots, L\}$.

For each subject s at a particular session t, we computed the functional connectivity profile of each vertex (of the cortical mesh) by correlating the vertex's fMRI time course with the time courses of uniformly distributed cortical regions of interests[1] (ROIs). For the GSP and HNU datasets, the preprocessed data were in fsaverage5 surface space. In this case, the ROIs consisted of 1175 vertices approximately uniformly distributed across the two hemispheres (Yeo et al., 2011). For the HCP dataset, the preprocessed data is in fs_LR32k surface space. In this case, the ROIs consisted of 1483 vertices spaced approximately uniformly distributed across the two hemispheres. Each vertex's connectivity profile was binarized (see Methods in main manuscript) and normalized to unit length. Let $X_n^{s,t}$ denote the binarized, normalized functional connectivity profile of subject s at vertex n during

[1] We note that with uniformly distributed ROIs, spatially extensive networks might be over-represented because they contribute disproportionately more ROIs to the computation of the connectivity profiles. However, regions within these spatially extensive networks do not necessarily have homogeneous connectivity patterns, so the uniformly distributed ROIs can capture these heterogeneous connectivity patterns. Furthermore, given that we do not know the networks a-priori, it would be challenging to define the ROIs a-priori, such that large patches of homogeneous regions would not be over-represented.

session t. Let D denote the total number of ROIs and hence the length of $X_n^{s,t}$. We denote the connectivity profiles from all sessions of all subjects at all cortical vertices as $X_{1:N}^{1:S,1:T}$.

Figure 1 (main text) illustrates the schematic of the multi-session hierarchical Bayesian model (MS-HBM). Following previous work (Yeo et al., 2011), the functional connectivity profile $X_n^{s,t}$ of subject s from a session t at vertex n is assumed to be generated from a von Mises-Fisher distribution,

$$p\left(X_n^{s,t}\middle|l_n^s = l, \mu_{1:L}^{s,t}, \kappa\right) = p\left(X_n^{s,t}\middle|\mu_l^{s,t}, \kappa\right) = z_D(\kappa)\exp\left(\kappa\langle X_n^{s,t}, \mu_l^{s,t}\rangle\right), \qquad (1)$$

where l_n^s is the parcellation label at vertex n of subject s, and $\langle\,,\rangle$ denote inner product. $\mu_l^{s,t}$ and κ are the mean direction and concentration parameter of the von Mises-Fisher distribution for network label l of subject s during session t. $\mu_{1:L}^{s,t}$ are the mean directions for networks 1 to L. We can think of $\mu_l^{s,t}$ as the mean connectivity profile of network label l normalized to unit length. If functional connectivity profile $X_n^{s,t}$ is similar to mean connectivity profile $\mu_l^{s,t}$ (i.e., $\langle X_n^{s,t}, \mu_l^{s,t}\rangle$ is big), then vertex n is more likely to be assigned to network l.

The concentration parameter κ controls the variability of the functional connectivity profiles within each network. A higher κ results in a lower dispersion (i.e., lower variance), which means that vertices belonging to the same network are more likely to possess functional connectivity profiles that are close to the mean connectivity profile of the network. In theory, it might make sense for κ to be different across networks because certain networks might exhibit more inter-region functional connectivity variability (Gordon et al., 2017a). However, in practice, certain networks (e.g., limbic networks) with low signal-to-noise ratio (SNR) would end up with very low κ, resulting in their encroachment into the peripheries of other networks. Therefore, κ was set to be the same across all networks, subjects and sessions. Finally, $z_D(\kappa)$ is a normalization constant to ensure a valid probability distribution (Banerjee et al., 2005):

$$z_D(\kappa) = \frac{\kappa^{\frac{D-1}{2}-1}}{(2\pi)^{\frac{D-1}{2}} I_{\frac{D-1}{2}-1}(\kappa)}\,, \qquad (2)$$

where $I_{\frac{D-1}{2}-1}(\cdot)$ is the modified Bessel function of the first kind with order $\frac{D-1}{2}-1$.

To model intra-subject functional connectivity variability, we assume a conjugate prior on the subject-specific and session-specific mean connectivity profiles $\mu_l^{s,t}$, which turns out to also be a von Mises-Fisher distribution:

$$p\left(\mu_l^{s,t}\middle|\mu_l^s,\sigma_l\right) = z_D(\sigma_l)\exp(\sigma_l\langle\mu_l^{s,t},\mu_l^s\rangle), \qquad (3)$$

where μ_l^s and σ_l are the mean direction and concentration parameter of the von Mises-Fisher distribution for network label l of subject s. We can think of μ_l^s as the individual-specific functional connectivity profile of network l of subject s. The concentration parameter σ_l controls how much the session-specific mean direction $\mu_l^{s,t}$ of subject s during session t can deviate from the subject-specific mean direction μ_l^s. A higher σ_l would imply lower intra-subject functional connectivity variability across sessions. σ_l is network-specific but is assumed to be the same for all subjects.

To model inter-subject functional connectivity variability, we assume a conjugate prior on the subject-specific mean connectivity profiles μ_l^s, which is again a von Mises-Fisher distribution whose mean direction corresponded to the group-level mean direction μ^g:

$$p\left(\mu_l^s\middle|\mu_l^g,\epsilon_l\right) = z_D(\epsilon_l)\exp(\epsilon_l\langle\mu_l^s,\mu_l^g\rangle), \qquad (4)$$

where μ^g and ϵ_l are the mean direction and concentration parameter of the von Mises-Fisher distribution for network label l. We can think of μ_l^g as the group-level functional connectivity profile of network l. The concentration parameter ϵ_l controls how much the individual-specific connectivity profile μ_l^s can deviate from the group-level connectivity profile μ_l^g. A higher ϵ_l would imply lower inter-subject functional connectivity variability across subjects.

Because the functional connectivity profiles of individual subjects are generally very noisy, we impose a MRF prior on the hidden parcellation labels $l_{1:N}^s$

$$p(l_{1:N}^s) = \frac{1}{Z(\alpha,c)}\exp(\alpha\textstyle\sum_{n=1}^N \log U(l_n^s|\Theta) - c\sum_{n=1}^N\sum_{m\in\mathcal{N}_n} V(l_n^s,l_m^s)), \qquad (5)$$

where $Z(\alpha,c)$ is a normalization term (partition function) to ensure $p(l_{1:N}^s)$ is a valid probability distribution. $\log U(l_n^s = l|\Theta) = \log\theta_{l,n}$ is a singleton potential encouraging certain vertices to be associated with certain labels. $V(l_n^s,l_m^s)$ is a pairwise potential (Potts model) encouraging neighboring vertices to have the same parcellation labels:

$$V(l_n^s, l_m^s) = \begin{cases} 0, & \text{if } l_n^s = l_m^s \\ 1, & \text{if } l_n^s \neq l_m^s \end{cases}, \qquad (6)$$

The parameters α and c are tunable parameters greater than zero, and control the tradeoffs between the various terms in the generative model. Assuming that $\alpha = 1$ and $c = 0$, then $\theta_{l,n}$ can be interpreted as the probability of label l occurring at vertex n of subject s.

S4. Model estimation for MS-HBM

In this section, we describe how model parameters are estimated from a training set and a validation set (Section S2.1), and how the parameters can be used to parcellate a new subject (Section S2.2). Throughout the entire section, we assume that the number of networks $L = 17$ without loss of generality.

S4.1 Learning model parameters

Our goal is to estimate the model parameters $\{\epsilon_{1:L}, \sigma_{1:L}, \Theta_{1:N,1:L}, \mu_{1:L}^g, c, \alpha\}$ from a training set and a validation set of binarized and normalized functional connectivity profiles, which can then be utilized for estimating individual-specific parcellations in unseen data of new subjects (Section S2.2). As a reminder, $\epsilon_{1:L}$ is a group prior representing inter-subject functional connectivity variability, $\sigma_{1:L}$ is a group prior corresponding to intra-subject functional connectivity variability, $\Theta_{1:N,1:L}$ is a group prior representing inter-subject spatial variability and reflects the probability of a network occurring at particular spatial location, and $\mu_{1:L}^g$ is the group-level connectivity profile for each network. The parameters α and c tradeoff between various terms in the generative model. Because the partition function $Z(\alpha, c)$ (Eq. (5)) is NP-hard to compute, for computational efficiency, we first assume $\alpha = 1$, $c = 0$ in order to estimate $\{\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^g\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L}\}$ from the training dataset[2]. Under this

[2] Conceptually, κ is estimated by averaging information across all vertices of all subjects. $\epsilon_{1:L}$, $\sigma_{1:L}$and $\mu_{1:L}^g$ are estimated by averaging information across all vertices within each network across all subjects, $\mu_{1:L}^{1:S}$ and $\mu_{1:L}^{1:S,1:T}$ are estimated by averaging information across all vertices within each network for each subject, while $\Theta_{1:N,1:L}$ is estimated for each vertex by averaging information across all subjects. On the other hand, the spatial smoothness V (parameterized by c) serves to "clean up" individual-specific parcellations by removing isolated islands of vertices assigned to particular networks. Because these isolated islands constitute only a small fraction of the networks, excluding the spatial smoothness V (i.e., set $c = 0$) will not significantly affect the estimates of $\epsilon_{1:L}$, $\sigma_{1:L}, \Theta_{1:N,1:L}$.

scenario, $Z(\alpha, c) = 1$, and $\theta_{l,n}$ can be interpreted as the probability of label l occurring at vertex n of subject s. The tunable parameters α and c are then estimated in the validation set using a grid search.

S4.1.1 Estimating $\{\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^{g}\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L}\}$ *from training set*

Given observed binarized, normalized functional connectivity profiles $X_{1:N}^{1:S,1:T}$ from the training set, we seek to estimate $\{\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^{g}\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L}\}$ using Expectation-Maximization (EM). As previously explained, we assume $\alpha = 1, c = 0$.

Let $\Omega = \{\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^{g}\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L}\}$. We consider the following maximum-a-posterior (MAP) estimation problem:

$$\underset{\Omega}{\operatorname{argmax}} \log p\left(\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^{g}\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L} \middle| X_{1:N}^{1:S,1:T}\right). \quad (7)$$

Assuming a uniform (improper) prior on $\{\Theta_{1:N,1:L}, \kappa, \sigma_{1:L}, \epsilon_{1:L}\}$, the MAP problem can be written as

$$\underset{\Omega}{\operatorname{argmax}} \log p\left(X_{1:N}^{1:S,1:T} \middle| \mu_{1:L}^{1:S,1:T}, \kappa, \Theta_{1:N,1:L}\right) p\left(\mu_{1:L}^{1:S,1:T} \middle| \sigma_{1:L}, \mu_{1:L}^{1:S}\right) p\left(\mu_{1:L}^{1:S} \middle| \epsilon_{1:L}, \mu_{1:L}^{g}\right). \quad (8)$$

We then introduce the parcellation labels $l_{1:N}^{s}$ for each subject s as latent variables, and use Jensen's inequality to define a lower bound $\mathcal{L}(\lambda, \Omega)$, where $\lambda = \lambda_{1:N,1:L}^{1:S}$ are the parameters of the q functions $q(l_{1:N}^{s}) = \prod_{n=1}^{N} q(l_{n}^{s} | \lambda_{n,1:L}^{s})$:

$$\begin{aligned}
&\log p\left(X_{1:N}^{1:S,1:T} \middle| \mu_{1:L}^{1:S,1:T}, \kappa, \Theta_{1:N,1:L}\right) p\left(\mu_{1:L}^{1:S,1:T} \middle| \sigma_{1:L}, \mu_{1:L}^{1:S}\right) p\left(\mu_{1:L}^{1:S} \middle| \epsilon_{1:L}, \mu_{1:L}^{g}\right) \\
&= \sum_{s=1}^{S} \log p\left(X_{1:N}^{s,1:T} \middle| \mu_{1:L}^{s,1:T}, \kappa, \Theta_{1:N,1:L}\right) + \sum_{s=1}^{S} \sum_{l=1}^{L} \log p\left(\mu_{l}^{s,1:T} \middle| \sigma_{l}, \mu_{l}^{s}\right) p\left(\mu_{l}^{s} \middle| \epsilon_{l}, \mu_{l}^{g}\right) \quad (9) \\
&= \sum_{s=1}^{S} \log \sum_{l_{1:N}^{s}} p\left(X_{1:N}^{s,1:T}, l_{1:N}^{s} \middle| \mu_{1:L}^{s,1:T}, \kappa, \Theta_{1:N,1:L}\right) + \sum_{s=1}^{S} \sum_{l=1}^{L} \log p\left(\mu_{l}^{s,1:T} \middle| \sigma_{l}, \mu_{l}^{s}\right) p\left(\mu_{l}^{s} \middle| \epsilon_{l}, \mu_{l}^{g}\right) \quad (10) \\
&\geq \sum_{s=1}^{S} \sum_{l_{1:N}^{s}} q(l_{1:N}^{s}) \log \frac{p\left(X_{1:N}^{s,1:T}, l_{1:N}^{s} \middle| \mu_{1:L}^{s,1:T}, \kappa, \Theta_{1:N,1:L}\right)}{q(l_{1:N}^{s})} + \sum_{s=1}^{S} \sum_{l=1}^{L} \log p\left(\mu_{l}^{s,1:T} \middle| \sigma_{l}, \mu_{l}^{s}\right) p\left(\mu_{l}^{s} \middle| \epsilon_{l}, \mu_{l}^{g}\right) \quad (11)
\end{aligned}$$

$$= \sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s \log p\left(X_n^{s,t}\middle|\mu_{l_n^s}^{s,t},\kappa\right) + \sum_{s=1}^{S}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s \log \Theta_{n,L_n^s}$$

$$- \sum_{s=1}^{S}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s \log \lambda_{n,l_n^s}^s + \sum_{s=1}^{S}\sum_{l=1}^{L}\left(\sum_{t=1}^{T} \log p(\mu^{s,t}|\sigma_l,\mu_l^s) + \log p(\mu_l^s|\epsilon_l,\mu_l^g)\right) \quad (12)$$

$$= \mathcal{L}(\lambda,\Omega), \quad (13)$$

where equality is achieved when $q(l_{1:N}^s) = \lambda_{1:N,1:L}^s$ are the posterior probability of the individual-specific parcellation of subject s given the parameters Ω. Therefore, instead of maximizing the original MAP problem (Eq. (7)), we instead maximize the lower bound:

$$\{\lambda^*,\Omega^*\} = \underset{\lambda,\Omega}{\operatorname{argmax}}\, \mathcal{L}(\lambda,\Omega). \quad (14)$$

In the E-step, we fix $\Omega = \{\epsilon_{1:L},\sigma_{1:L},\kappa,\mu_{1:L}^g\,\mu_{1:L}^{1:S},\mu_{1:L}^{1:S,1:T},\Theta_{1:N,1:L}\}$, and estimate λ:

$$\lambda = \underset{\lambda}{\operatorname{argmax}}\, \mathcal{L}(\lambda,\Omega) \quad (15)$$

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s \log p(X_n^{s,t}|\mu_{l_n^s}^{s,t},\kappa) + \sum_{s=1}^{S}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s \log \Theta_{n,l_n^s}$$

$$- \sum_{s=1}^{S}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s \log \lambda_{n,l_n^s}^s + \sum_{s=1}^{S}\sum_{n=1}^{N} \eta_n^s \left(\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^s - 1\right), \quad (16)$$

where η_n^s are the Lagrange multipliers enforcing the constraint $\sum_{l=1}^{L} \lambda_{n,l}^s = 1$. Optimizing Eq. (16), we get:

$$\log \lambda_{k,l}^s \propto \sum_{t=1}^{T} \log p\left(X_k^{s,t}\middle|\mu_l^{s,t},\kappa\right) + \log \Theta_{k,l} \quad (17)$$

$$= \sum_{t=1}^{T} \log z_D(\kappa) \exp\left(\kappa\langle X_n^{s,t},\mu_l^{s,t}\rangle\right) + \log \Theta_{k,l} \quad (18)$$

$$= T \log z_D(\kappa) + \sum_{t=1}^{T} \kappa\langle X_n^{s,t},\mu_l^{s,t}\rangle + \log \Theta_{k,l} \quad (19)$$

In the M-step, we fix λ and estimate Ω:

$$\Omega = \underset{\Omega}{\operatorname{argmax}}\, \mathcal{L}(\lambda, \Omega). \qquad (20)$$

By using the constraints that $\langle \mu_l^{s,t}, \mu_l^{s,t} \rangle = 1$, $\langle \mu_l^s, \mu_l^s \rangle = 1$, $\langle \mu_l^g, \mu_l^g \rangle = 1$, $\kappa > 0$, $\sigma_l > 0$, $\epsilon_l > 0$, and differentiating $\mathcal{L}(\lambda, \Omega)$ with respect to $\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^g\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L}$, and setting the derivatives to zero, we get the following update equations:

$$\mu_l^{s,t} = \frac{\kappa \sum_{n=1}^{N} \lambda_{n,l}^s X_n^{s,t} + \sigma_l \mu_l^s}{\left\| \kappa \sum_{n=1}^{N} \lambda_{n,l}^s X_n^{s,t} + \sigma_l \mu_l^s \right\|} \qquad (21)$$

$$\kappa = \frac{(D-2)\Gamma^\kappa}{1-\Gamma^{\kappa 2}} + \frac{(D-1)\Gamma^\kappa}{2(D-2)}, \Gamma^\kappa = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l=1}^{L} \lambda_{n,l}^s \langle \mu_l^{s,t}, X_n^{s,t} \rangle}{T \sum_{s=1}^{S}\sum_{n=1}^{N}\sum_{l=1}^{L} \lambda_{n,l}^s} \qquad (22)$$

$$\mu_l^s = \frac{\sigma_l \sum_{t=1}^{T} \mu_l^{s,t} + \epsilon_l \mu_l^g}{\left\| \sigma_l \sum_{t=1}^{T} \mu_l^{s,t} + \epsilon_l \mu_l^g \right\|} \qquad (23)$$

$$\sigma_l = \frac{(D-2)\Gamma_l^\sigma}{1-\Gamma_l^{\sigma 2}} + \frac{(D-1)\Gamma_l^\sigma}{2(D-2)}, \Gamma_l^\sigma = \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} \langle \mu_l^s, \mu_l^{s,t} \rangle \qquad (24)$$

$$\mu_l^g = \frac{\sum_{s=1}^{S} \epsilon_l \mu_l^s}{\left\| \sum_{s=1}^{S} \epsilon_l \mu_l^s \right\|} = \frac{\sum_{s=1}^{S} \mu_l^s}{\left\| \sum_{s=1}^{S} \mu_l^s \right\|} \qquad (25)$$

$$\epsilon_l = \frac{(D-2)\Gamma_l^\epsilon}{1-\Gamma_l^{\epsilon 2}} + \frac{(D-1)\Gamma_l^\epsilon}{2(D-2)}, \Gamma_l^\epsilon = \frac{1}{S}\sum_{s=1}^{S} \langle \mu_k^g, \mu_l^s \rangle \qquad (26)$$

$$\Theta_{n,l} = \frac{1}{S}\sum_{s=1}^{S} \lambda_{n,l}^s \ , \qquad (27)$$

where D is the length of $X_n^{s,t}$ (i.e., number of ROIs in each functional connectivity profile), S is the number of subjects, T is the number of sessions, and $\|\cdot\|$ corresponds to the l_2-norm. Therefore, the estimate of the functional connectivity profile $\mu_l^{s,t}$ (Eq. (21)) of network l of subject s during session t is the weighted sum of the average time course of vertices constituting network l of subject s during session t ($\sum_{n=1}^{N} \lambda_{n,l}^s X_n^{s,t}$) and the subject-specific mean direction μ_l^s, with weights κ and σ_l for each term, normalized to be unit norm. If σ_l is much greater than κ, then $\mu_l^{s,t}$ is more likely to be dominated by subject-specific mean direction μ_l^s, which means that the functional connectivity profile of network l is highly stable across sessions. Similarly, the estimate of the functional connectivity profile μ_l^s (Eq. (23)) of network l of subject s is the weighted sum of the average session-specific mean directions across all sessions for network l of subject s ($\sum_{t=1}^{T} \mu_l^{s,t}$) and the group-level mean direction μ_l^g, with

weights σ_l and ϵ_l for each term, normalized to be unit norm. If ϵ_l is much greater than σ_l, then μ_l^s is more likely to be dominated by group-level mean direction μ_l^g, which means that the functional connectivity profile of network l is highly stable between subjects. Finally, the estimate of the group-level functional connectivity profile μ_l^g (Eq. (25)) of network l is the sum of the subject-specific mean directions across all subjects for network l ($\sum_{s=1}^{S} \mu_l^s$), normalized to be unit norm. The estimate of $\Theta_{n,l}$ (Eq. (27)) is the posterior probability of network l being assigned to vertex n, averaged across all the subjects.

Given the training set, the algorithm first estimates a group-level parcellation (Yeo et al., 2011), which is then used to initialize the EM algorithm. The EM algorithm iterates E-step (Eq. (19)) and M-step (Eqs. (21-27)) till convergence. We note that the update equations (Eqs. (21-27)) in the M-step are dependent on each other. Therefore, within the M-step, the update equations (Eqs. (21-27) are iterated till convergence.

S4.1.2 Estimating tunable parameters c and α

In the previous subsection (Section S2.1.1), the training set was used to estimate $\Omega = \{\epsilon_{1:L}, \sigma_{1:L}, \kappa, \mu_{1:L}^g\ \mu_{1:L}^{1:S}, \mu_{1:L}^{1:S,1:T}, \Theta_{1:N,1:L}\}$, assuming $\alpha = 1, c = 0$. To tune the parameters c and α, we assume access to a validation set.

Recall that each subject in the validation set has multiple rs-fMRI sessions. We consider $c \in \{10,20,30,40,50,60\}$ and $\alpha \in \{100,150,200,250\}$. For a given pair of (c, α), and given $\{\epsilon_{1:L}, \sigma_{1:L}, \Theta_{1:N,1:L}, \mu_{1:L}^g\}$ estimated from the training set, we estimate for each subject in the validation set, the individual-specific parcellation based on a subset of rs-fMRI sessions (see Section S2.2 for algorithm). Resting-state homogeneity (Eq. (1) in main text) is then computed in the remaining rs-fMRI sessions of the validation subjects. The pair of (c, α) with the highest homogeneity in the unseen rs-fMRI sessions of the validation subjects is then utilized for parcellating new subjects.

In the case of the GSP data, the optimal pair of parameters is $c = 30$ and $\alpha = 200$. In the case of the HCP data, the optimal pair of parameters is $c = 40$ and $\alpha = 200$. Note that we do expect the parameters to be different between the GSP and HCP datasets because of resolution differences between the fsaverage5 and fs_LR32k surface meshes.

Throughout the paper (main text), the reported quality (Figures 5, 6 and 7) of the individual-specific parcellations was evaluated using subjects not used to tune the parameters. For example, in the case of the CoRR-HNU subjects (Figures 5 and 6), the model parameters were estimated from the GSP training and validation sets. In the case of the HCP data (Figures 6 and 7), model parameters were

estimated the HCP training and validation sets, while the reported quality of the individual-specific parcellations was evaluated using the HCP test set.

S4.2 Individual-level parcellation estimation

Using parameters $\{\epsilon_{1:L}, \sigma_{1:L}, \Theta_{1:N,1:L}, \mu^g_{1:L}\}$ estimated from the training set (Section S2.1.1), and for a particular pair of (c, α), we can estimate the individual-specific parcellation $l^s_{1:N}$ of a new subject s with T sessions by employing the variational Bayes expectation maximization (VBEM) algorithm.

Let $\Psi = \{\kappa, \mu^{s,1:T}_{1:L}, \mu^s_{1:L}\}$. We consider the following maximum-a-posterior (MAP) estimation problem:

$$\underset{\Psi}{\operatorname{argmax}} \log p\left(\kappa, \mu^{s,1:T}_{1:L}, \mu^s_{1:L} \middle| X^{s,1:T}_{1:N}, \epsilon_{1:L}, \sigma_{1:L}, \mu^g_{1:L}, \Theta_{1:N,1:L}\right). \tag{28}$$

Assuming a uniform (improper) prior on κ, and by introducing the parcellation labels $l^s_{1:N}$ of the new subject s as latent variables, the lower bound $\mathcal{L}(\lambda, \Psi)$ of the MAP problem (Eq. (28)) can be written as:

$$\begin{aligned}\mathcal{L}(\lambda^s, \Psi) = &\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l^s_n=1}^{L} \lambda^s_{n,l^s_n} \log p\left(X^{s,t}_n \middle| \mu^{s,t}_{l^s_n}, \kappa\right) + \alpha \sum_{n=1}^{N}\sum_{l^s_n=1}^{L} \lambda^s_{n,l^s_n} \log \Theta_{n,L^s_n} \\ &- c\sum_{n=1}^{N}\sum_{m\in\mathcal{N}_n}\sum_{l^s_n=1}^{L}\sum_{l^s_m=1}^{L} \lambda^s_{n,l^s_n}\lambda^s_{m,l^s_m} V(l^s_n, l^s_m) \\ &- \sum_{n=1}^{N}\sum_{l^s_n=1}^{L} \lambda^s_{n,l^s_n} \log \lambda^s_{n,l^s_n} + \sum_{l=1}^{L}\left(\sum_{t=1}^{T} \log p(\mu^{s,t}|\sigma_l, \mu^s_l) + \log p\left(\mu^s_l \middle| \epsilon_l, \mu^g_l\right)\right),\end{aligned} \tag{29}$$

where equality is achieved when λ^s is the posterior probability of the individual-specific parcellation of subject s given the parameters Ψ. Similar to Section S2.1.1, we can maximize the lower bound (Eq. (29)) by iteratively updating λ^s and Ψ. Unlike Section S2.1.1, we cannot compute the exact posterior probability λ^s because of the pairwise potentials in the Markov random field (Wainwright and Jordan, 2008). Using the mean-field approximation (Wainwright and Jordan, 2008), an approximate posterior probability λ^s is estimated in the variational E-step, while Ψ is updated in the variational M-step.

More specifically, in the variational E-step, Ψ is fixed and λ^s is estimated as follows:

$$\log \lambda_{n,l}^{s} \propto T \log z_D(\kappa) + \sum_{t=1}^{T} \kappa \langle X_n^{s,t}, \mu_l^{s,t} \rangle - 2c \sum_{m \in \mathcal{N}_n} \sum_{l_m^s=1}^{L} \lambda_{m,l_m^s}^{s} V(l_n^s, l_m^s) + \alpha \log \Theta_{n,l}. \quad (30)$$

In the variational M-step, λ^s is fixed and $\Psi = \{\kappa, \mu_{1:L}^{s,1:T}, \mu_{1:L}^{s}\}$ is estimated as follows:

$$\mu_l^{s,t} = \frac{\kappa \sum_{n=1}^{N} \lambda_{n,l}^{s} X_n^{s,t} + \sigma_l \mu_l^s}{\left\| \kappa \sum_{n=1}^{N} \lambda_{n,l}^{s} X_n^{s,t} + \sigma_l \mu_l^s \right\|} \quad (31)$$

$$\kappa = \frac{(D-2)\Gamma^\kappa}{1-\Gamma^{\kappa 2}} + \frac{(D-1)\Gamma^\kappa}{2(D-2)}, \Gamma^\kappa = \frac{\sum_{s=1}^{S} \sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{s} \langle \mu_l^{s,t}, X_n^{s,t} \rangle}{T \sum_{s=1}^{S} \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{s}} \quad (32)$$

$$\mu_l^s = \frac{\sigma_l \sum_{t=1}^{T} \mu_l^{s,t} + \epsilon_l \mu_l^g}{\left\| \sigma_l \sum_{t=1}^{T} \mu_l^{s,t} + \epsilon_l \mu_l^g \right\|}. \quad (33)$$

Once the VBEM algorithm converges, vertex n of subject s will be assigned to label l with the highest (approximate) posterior probability.

S5. Alternative parcellation approaches

We compared the MS-HBM with four alternative parcellation approaches. The first approach was to apply the population-level parcellation (Yeo et al., 2011) to individual subjects. We will refer to this approach as "Yeo2011". For the second approach, recall that the population-level parcellation algorithm was an expectation-maximization (EM) algorithm, which iteratively computed a network connectivity profile based on vertices assigned to the same network (M-step) and then re-assigned the network membership of vertices based on the similarity between each vertex's connectivity profile and the network connectivity profile (E-step). Using the network connectivity profiles from the Yeo2011 population-level parcellation, we could estimate networks in an individual subject by assigning a network label to each vertex based on the similarity between the vertex's connectivity profile (for that subject) and the population-level network connectivity profile (i.e., E-step). Since this approach is analogous to the ICA back-projection algorithm (Calhoun et al., 2009; Beckmann et al., 2009; Filippini et al., 2009; Zuo et al., 2010; Calhoun and Adali 2012), we will refer to this second alternative approach as "YeoBackProject".

We also implemented the influential individual-specific parcellation algorithm of Gordon and colleagues (Gordon et al., 2017a; Gordon et al., 2017b), where the binarized functional connectivity map of each cortical vertex was matched to binarized network templates derived from the Yeo2011 population-level parcellation. Care was taken to verify that our implementation was consistent with

Gordon's algorithm (whose code is publicly available). We refer to this third approach as "Gordon2017".

Finally, we considered the prominent individual-specific parcellation algorithm of Wang and colleagues (Wang et al., 2015), which adapted the Yeo2011 population-level parcellation to an individual subject, while accounting for inter-subject RSFC variability and SNR characteristics of the subject's rs-fMRI data. However, scanner noise is just one component contributing to intra-subject variability. As shown by others (Mueller et al., 2013; Laumann et al., 2015) and also in our results, brain networks with the highest intra-subject variability do not correspond to low SNR regions. We refer to this fourth approach as "Wang2015".

S6. Behavioral prediction model

In this section, we describe our model for behavioral prediction based on individual differences in the spatial arrangement of cortical networks. Kernel regression (Murphy et al., 2012) was utilized to predict each behavioral phenotype in individual subjects. Suppose we have M training subjects, y_i is the behavioral measure (e.g., fluid intelligence) and l_i is the individual-specific parcellation of the i-th training subject. Given $\{y_1, y_2, \dots, y_M\}$ and $\{l_1, l_2, \dots, l_M\}$, the kernel regression model will be:

$$y_i = \sum_{j=1}^{M} \alpha_j K(l_j, l_i), \qquad (34)$$

where $K(l_j, l_i)$ is the Dice overlap coefficient between corresponding networks of the i-th and j-th training subjects, averaged across 17 networks. The classical way to estimate α in Eq. (34) is to minimize the quadratic cost:

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{M} \left(y_i - \sum_{j=1}^{M} \alpha_j K(l_j, l_i) \right)^2. \qquad (35)$$

Defining $\boldsymbol{y} = [y_1, y_2, \dots, y_M]^T$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_M]^T$ and $\mathbb{K}$ to be an $M \times M$ matrix, whose (j, i)-th element is $K(l_j, l_i)$, Eq. (35) can be written as:

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2} (\boldsymbol{y} - \mathbb{K}\boldsymbol{\alpha})^T (\boldsymbol{y} - \mathbb{K}\boldsymbol{\alpha}). \qquad (36)$$

Differentiating Eq. (36) with respect to $\boldsymbol{\alpha}$, we can get

$$\boldsymbol{\alpha} = \mathbb{K}^{-1}\boldsymbol{y}. \qquad (37)$$

To predict the behavior measure y_s (e.g., fluid intelligence) of a test subject s with its individual-specific parcellation l_s, we can compute $K(l_i, l_s)$, which is the Dice overlap coefficient

between corresponding networks of subject s and i-th training subject, averaged across 17 networks. The predicted behavior measure y_s can be calculated as

$$y_s = \sum_{i=1}^{M} \alpha_i K(l_i, l_s), \qquad (38)$$

where α_i is estimated by Eq. (37). If we denote $\boldsymbol{K}_s = [K(l_1, l_s), K(l_2, l_s), \dots, K(l_M, l_s)]$, then Eq. (38) can be written as:

$$y_s = \boldsymbol{K}_s \boldsymbol{\alpha} = \boldsymbol{K}_s \mathbb{K}^{-1} \boldsymbol{y}\,. \qquad (39)$$

In practice, $\mathbb{K}^{-1}$ is a symmetric matrix whose diagonal elements are roughly the same and ~100 times larger than the off-diagonal elements. Therefore, the predicted behavior measure y_s can be seen as the weighted average of the behaviors of the training subjects: $y_s \approx \boldsymbol{K}_s \boldsymbol{y} = \sum_{i=1}^{M} K(l_i, l_s) y_i$. If the individual-specific parcellation l_s of test subject s is more similar to the parcellation of training subject i than training subject j, then weight $K(l_i, l_s)$ will be larger than $K(l_j, l_s)$, and so y_s will be more similar to y_i than y_j.

To reduce overfitting, an l_2-regularization term (i.e., kernel ridge regression) is typically added to cost function (Eq. (36)), resulting in a new regularized cost function:

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2} (\boldsymbol{y} - \mathbb{K}\boldsymbol{\alpha})^T (\boldsymbol{y} - \mathbb{K}\boldsymbol{\alpha}) + \frac{\lambda}{2} \boldsymbol{\alpha}^T \mathbb{K} \boldsymbol{\alpha}\,, \qquad (40)$$

where λ is a tuning parameter, which controls the importance of the regularization term. Differentiating Eq. (40) with respect to $\boldsymbol{\alpha}$, we get

$$\boldsymbol{\alpha} = (\mathbb{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{y}. \qquad (41)$$

To predict the behavior measure y_s (e.g., fluid intelligence) of a test subject s, Eq. (41) is substituted into Eq. (38), resulting in

$$y_s = \boldsymbol{K}_s \boldsymbol{\alpha} = \boldsymbol{K}_s (\mathbb{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{y}. \qquad (42)$$