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

Ru Kong[1], Jingwei Li[1], Nanbo Sun[1], Mert R Sabuncu[2], Hesheng Liu[8], Alexander Schaefer[1], Csaba Orban[1], Xi-Nian Zuo[3,4], Avram J. Holmes[5], Simon Eickhoff[6,7], B.T. Thomas Yeo[1,8,9,10]

[1]Department of Electrical and Computer Engineering, ASTAR-NUS Clinical Imaging Research Centre, Singapore Institute for Neurotechnology and Memory Networks Program, National University of Singapore, Singapore [2]School of Electrical and Computer Engineering, Cornell University, USA [3]CAS Key Laboratory of Behavioral Sciences and Research Center for Lifespan Development of Brain and Mind (CLIMB), Institute of Psychology, Beijing, China [4]University of Chinese Academy of Sciences, Beijing, China [5]Yale University, New Haven, CT, USA [6]Institute for Systems Neuroscience, Medical Faculty, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany [7]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Center Jülich, Jülich, Germany [8]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA [9]Centre for Cognitive Neuroscience, Duke-NUS Medical School, Singapore [10]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore

Address correspondence to:

B.T. Thomas Yeo
ECE, ASTAR-NUS CIRC, SINAPSE & MNP
National University of Singapore
Email: thomas.yeo@nus.edu.sg

Abstract

Resting-state functional magnetic resonance imaging (rs-fMRI) offers the opportunity to delineate brain networks in individuals. A major question is whether individual-specific network topography (i.e., location and spatial arrangement) is behaviorally relevant. Here, we propose a multi-session hierarchical Bayesian model (MS-HBM) for estimating individual-specific cortical networks, and investigate whether individual-specific network topography can predict human behavior. The multiple layers of the MS-HBM explicitly differentiate intra-subject (within-subject) from inter-subject (between-subject) network variability. By ignoring intra-subject variability, previous individual-specific network mappings might confuse intra-subject variability for inter-subject differences. Compared with other approaches, MS-HBM parcellations generalized better to new rs-fMRI and task-fMRI data from the same subjects. Specifically, MS-HBM parcellations estimated from a single rs-fMRI session (10 minutes) showed comparable generalizability as parcellations computed by two state-of-the-art methods using data from five combined sessions (50 minutes). Using kernel regression, we showed that individual-specific network topography could predict 58 behavioral phenotypes across cognition, personality and emotion. Although anatomical size often relates to functional capacity across and within species, we found network topography to be more effective for behavioral prediction than network size. Thus, we introduce individual-specific network *topography* as a fingerprint of human behavior, highlighting the potential of looking beyond connectivity *strength* for behavioral prediction.

Introduction

One prominent tool for identifying large-scale human brain networks is resting-state functional connectivity (RSFC), which reflects the synchrony of fMRI signals between brain regions, while a subject is lying at rest without any goal-directed task (Biswal et al., 1995; Greicius et al. 2003; Fox and Raichle, 2007). RSFC brain networks correspond well to task-evoked activation patterns (Seeley et al. 2007; Smith et al., 2009; Cole et al., 2014; Yeo et al., 2015a). RSFC is also heritable (Glahn et al. 2010; Yang et al. 2016; Ge et al., 2017), correlates with gene expression across the cortical mantle (Hawrylycz et al. 2015; Richiardi et al. 2015; Krienen et al. 2016), and predicts individual differences in behavior (Hampson et al., 2006; van den Heuvel et al., 2009; Finn et al., 2015; Smith et al., 2015). Consequently, RSFC has been widely utilized to estimate population-average functional brain networks by averaging data across multiple subjects (Beckmann et al. 2005; Damoiseaux et al. 2006; Fox et al. 2006; Dosenbach et al. 2007; Margulies et al. 2007; Power et al. 2011; Yeo et al., 2011; Lee et al. 2012).

Population-average networks have provided important insights into the large-scale functional organization of the human brain (Buckner et al., 2013; Wig et al., 2017). However, since population-average networks might obscure individual-specific network organization, there has been a tremendous interest in estimating unique individual-specific brain networks (Beckmann et al., 2009; Bellec et al., 2010; Zuo et al., 2010; Varoquaux et al., 2011; Hacker et al., 2013; Wig et al., 2014; Chong et al., 2017). Indeed, many studies have documented that the size, location and spatial arrangement of individual-specific brain networks vary substantially across participants (Harrison et al., 2015; Laumann et al., 2015; Wang et al., 2015; Glasser et al., 2016; Gordon et al., 2017a, 2017b, 2017c; Braga and Buckner, 2017). Yet, the possible behavioral relevance of individual differences in network size and network topography (location and spatial arrangement) remains unknown.

We proposed a multi-session hierarchical Bayesian model (MS-HBM) for estimating individual-specific network parcellations of the cerebral cortex, and investigated whether individual-specific network topography and size are associated with human behavior. The multiple layers of the MS-HBM allowed explicit separation of inter-subject (between-subject) and intra-subject (within-session) functional connectivity variability. Previous individual-specific network mappings only accounted for inter-subject variability, but not intra-subject

variability[1]. However, inter-subject and intra-subject RSFC variability can be markedly different across regions (Mueller et al., 2013; Chen et al., 2015; Laumann et al., 2015). For example, the motor cortex exhibits high intra-subject functional connectivity variability, but low inter-subject functional connectivity variability (Laumann et al., 2015). Therefore, observed RSFC variability in the motor cortex might be incorrectly attributed to inter-subject spatial variability of brain networks, rather than just intra-subject sampling variability, resulting in sub-optimal network mapping. In this paper, we showed that compared with four other approaches, MS-HBM individual-specific parcellations generalized better to new resting and task fMRI data from the same individuals.

Having established that the MS-HBM generated high-quality individual-specific parcellations, we utilized kernel regression to demonstrate that individual differences in network topography (i.e., location and spatial arrangement) could predict 58 behavioral phenotypes across cognition, personality and emotion. Thus, individuals with more spatially overlapping cortical parcellations exhibited more similar behavioral phenotypes. The kernel regression framework allowed us to dissect the difference between network topography and network size in predicting behavior. Unlike network topography, there is significant literature suggesting that regional brain volume predicts functional capacity across and within species (Krebs et al., 1990; Barton et al., 2009; Erickson et al., 2011; Holmes et al., 2016; Sabuncu et al., 2016). For example, nocturnal primates have larger olfactory structures, but smaller striate visual cortices, than diurnal primates (Barton et al., 2009). In our analyses, we found that at least at the resolution of large-scale networks, network topography was more useful than network size in predicting behavior.

The contributions of this work are two-fold. First, by estimating inter-subject variability, intra-subject network variability and individual-specific networks within a unified statistical framework, the estimation of individual-specific cortical networks were greatly improved. For example, MS-HBM parcellations estimated from a *single* rs-fMRI session were comparable to those generated by two prominent algorithms using five times the amount of data (Wang et al., 2015; Gordon et al., 2017a; 2017b), as evaluated by generalizability to new rs-fMRI data from the same subjects. Second, our results suggest that

[1] We note that Wang and colleagues (2015) accounted for the signal-to-noise characteristics of fMRI data in their network estimation procedure. 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 signal-to-noise regions.

individual-specific cortical network topography might serve as a fingerprint of human behavior, in contrast to the vast majority of previous literature, which relies on functional connectivity strength for behavioral association or prediction (Hampson et al., 2006; Finn et al., 2015; Rosenberg et al., 2016; Smith et al., 2015; Yeo et al., 2015b).

Methods

Overview

We proposed a multi-session hierarchical Bayesian model (MS-HBM) to estimate functional network parcellations of the cerebral cortex in individual subjects. The model distinguished between inter-subject and intra-subject network variability. Subsequent analyses proceeded in four stages. First, to examine whether inter-subject and intra-subject variability could be reliably estimated across datasets, the MS-HBM was applied to three multi-session resting-state fMRI datasets. Second, the MS-HBM was compared with four other approaches using new rs-fMRI and task-fMRI data from the same subjects. Third, we examined the reproducibility of the MS-HBM parcellations and how well the parcellations captured inter-subject differences. Finally, we investigated whether individual differences in cortical parcellations reflected individual differences in behavior.

Multi-session fMRI datasets

The Genomic Superstruct Project (GSP) test-retest dataset (Holmes et al., 2015) consisted of structural MRI and resting-state fMRI from 69 healthy young adults (ages 18 to 35). All imaging data were collected on matched 3T Tim Trio scanners (Siemens Healthcare, Erlangen, Germany) at Harvard University and Massachusetts General Hospital using the vendor-supplied 12-channel phased-array head coil. Each participant has two sessions, acquired on two different days separated by less than 6 months. One or two rs-fMRI runs were acquired per session. Each BOLD run was acquired in 3mm isotropic resolution with a TR of 3.0 seconds and lasted for 6 minutes and 12 seconds. The structural data consisted of one 1.2mm isotropic scan for each session. Details of the data collection can be found elsewhere (Holmes et al., 2015).

The Hangzhou Normal University of the Consortium for Reliability and Reproducibility (CoRR-HNU) multi-session dataset (Zuo et al., 2014; Chen et al., 2015) consisted of structural MRI and resting-state fMRI from 30 young healthy adults (ages 20 to 30). All imaging data were collected on 3T GE Discovery MR750 using an 8-channel head coil. Each participant was scanned a total of 10 sessions across one month (one session every three days). One rs-fMRI run was collected in each session. Each fMRI run was acquired in 3.4mm isotropic resolution with a TR of 2.0 seconds and lasted for 10 minutes. The structural data consisted of one 1mm isotropic scan for each session. Details of the data collection can be found elsewhere (Zuo et al., 2014; Chen et al., 2015).

The Human Connectome Project (HCP) S900 release (Van Essen et al., 2012b; Smith et al., 2013) consisted of structural MRI, resting-state fMRI and task fMRI of 881 subjects. All imaging data were collected on a custom-made Siemens 3T Skyra scanner using a multiband sequence. Each participant has two fMRI sessions on two consecutive days. Two rs-fMRI runs were collected in each session. Each fMRI run was acquired in 2mm isotropic resolution with a TR of 0.72 seconds and lasted for 14 minutes and 33 seconds. The structural data consisted of one 0.7mm isotropic scan for each subject. Details of the data collection can be found elsewhere (Van Essen et al., 2012b; Smith et al., 2013).

It is worth noting the significant acquisition differences among the three datasets, including scanner type (e.g., GE versus Siemens) and acquisition sequence (e.g., multiband versus non-multiband). The interval between repeated visits were also very different, ranging from one day in the HCP dataset and up to six months in the GSP dataset. These differences allowed us to test the robustness of our individual-specific network estimation model.

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 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.

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 fsLR surface space (Van Essen et al. 2012a), smoothed by 2mm and denoised with ICA-FIX (Salimi-Khorshidi et al. 2014; Griffanti 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 fsLR 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 $\leq$ f $\leq$ 0.08 Hz) was then applied to the data. Finally, spatial smoothing was applied by iteratively averaging the data at each surface mesh vertex with its neighbors four times.

Population-level parcellation and functional connectivity profiles

We have previously developed an approach to derive a population-level parcellation of the cerebral cortex into large-scale resting-state networks (Yeo et al., 2011). The cortical networks were defined as sets of cortical regions with similar corticocortical functional connectivity profiles. Here we applied the same approach to the GSP, CoRR-HNU and HCP datasets. Our previous analyses (Yeo et al., 2011) identified 7 and 17 networks to be particularly stable. For simplicity, we will only consider 17 networks throughout this paper. Details of this approach have been previously described (Yeo et al., 2011). For completeness, we briefly described its application to the current datasets.

Recall that the preprocessed fMRI data from the CoRR-HNU and GSP subjects have been projected onto the fsaverage5 surface meshes. The fsaverage5 surface meshes consisted of 18715 cortical vertices. Following previous work (Yeo et al., 2011), the connectivity profile of a cortical region (vertex) was defined to be its functional coupling to 1175 regions of interest (ROIs). The 1175 ROIs consisted of single vertices uniformly distributed across the fsaverage5 surface meshes. For each rs-fMRI run of each subject, the Pearson's correlation between the fMRI time series at each spatial location (18715 vertices) and the 1175 ROIs were computed. The 18715 x 1175 correlation matrix were then binarized by keeping the top 10% of the correlations to obtain the final functional connectivity profiles. Outlier volumes (flagged during preprocessing) were ignored when computing the correlations.

In the case of the HCP dataset, the preprocessed fMRI data have been projected onto the fsLR surface space. The fsLR_32K surface meshes consisted of 59412 cortical vertices. We defined the connectivity profile of a cortical region (vertex) to be its functional coupling to 1483 ROIs. The 1483 ROIs consisted of single vertices uniformly distributed across the fsLR_32K surface meshes. For each rs-fMRI run of each subject, the Pearson's correlation between the fMRI time series at each spatial location (59412 vertices) and the 1483 ROIs were computed. The 59412 x 1483 correlation matrix were then binarized by keeping the top 10% of the correlations to obtain the final functional connectivity profile. Outlier volumes (flagged during preprocessing) were ignored when computing the correlations.

To obtain a population-level parcellation from a group of subjects, each vertex's connectivity profiles were averaged across all BOLD runs of all subjects. The averaged connectivity profiles were clustered using a mixture of von Mises–Fisher distributions (Lashkari et al., 2010; Yeo et al., 2011). The expectation-maximization (EM) algorithm operated by first randomly assigning the vertices (18715 in the GSP and CoRR-HNU datasets, or 59412 in the HCP dataset) to different networks. The algorithm then iterated between two steps (E-step and M-step) until convergence. In the M-step, the algorithm computed a network-level connectivity profile based on vertices assigned to the same network. In the E-step, the algorithm re-assigned the network membership of vertices based on the similarity between each vertex's connectivity profile and the network-level connectivity profile. The clustering algorithm was repeated 1000 times with different random initializations and the estimate with the best model likelihood was selected.

Multi-session hierarchical Bayesian model (MS-HBM)

The previous section described an approach to estimate a population-level parcellation from a group of subjects. Figure 1 illustrates the MS-HBM model for estimating individual-specific cerebral cortex parcellations using multi-session fMRI data. Let $X_n^{s,t}$ denote the (binarized) functional connectivity profile of cortical vertex n from session t of subject s. For example, Figure 1 (fourth row) illustrates the binarized functional connectivity profile for a posterior cingulate cortex vertex ($X_{PCC}^{1,1}$) and a precuneus vertex ($X_{pCun}^{1,1}$) from the 1st session of the 1st subject. The shaded circle indicates that $X_n^{s,t}$ is the only observation in the entire model. Based on the observed connectivity profiles of *all* vertices from *all* sessions of a single subject, the goal is to assign a network label l_n^s for each vertex n of subject s. Even though a vertex's connectivity profile is unlikely to be the same across different fMRI sessions, the vertex's network label was assumed to be the same across sessions. For example, Figure 1 (last row) illustrates the individual-specific parcellation of the 1st subject using data from *all* sessions.

Some of the model parameters (e.g., inter-subject variability) must be estimated from a training set of subjects. A new subject (possibly from another dataset) could then be parcellated without access to the original training data. Even though the model was defined on multi-session fMRI data, an effective workaround was provided for single-session fMRI data. The exact mathematical model is found in Supplemental Methods S1. Here we provide the intuition behind this model.

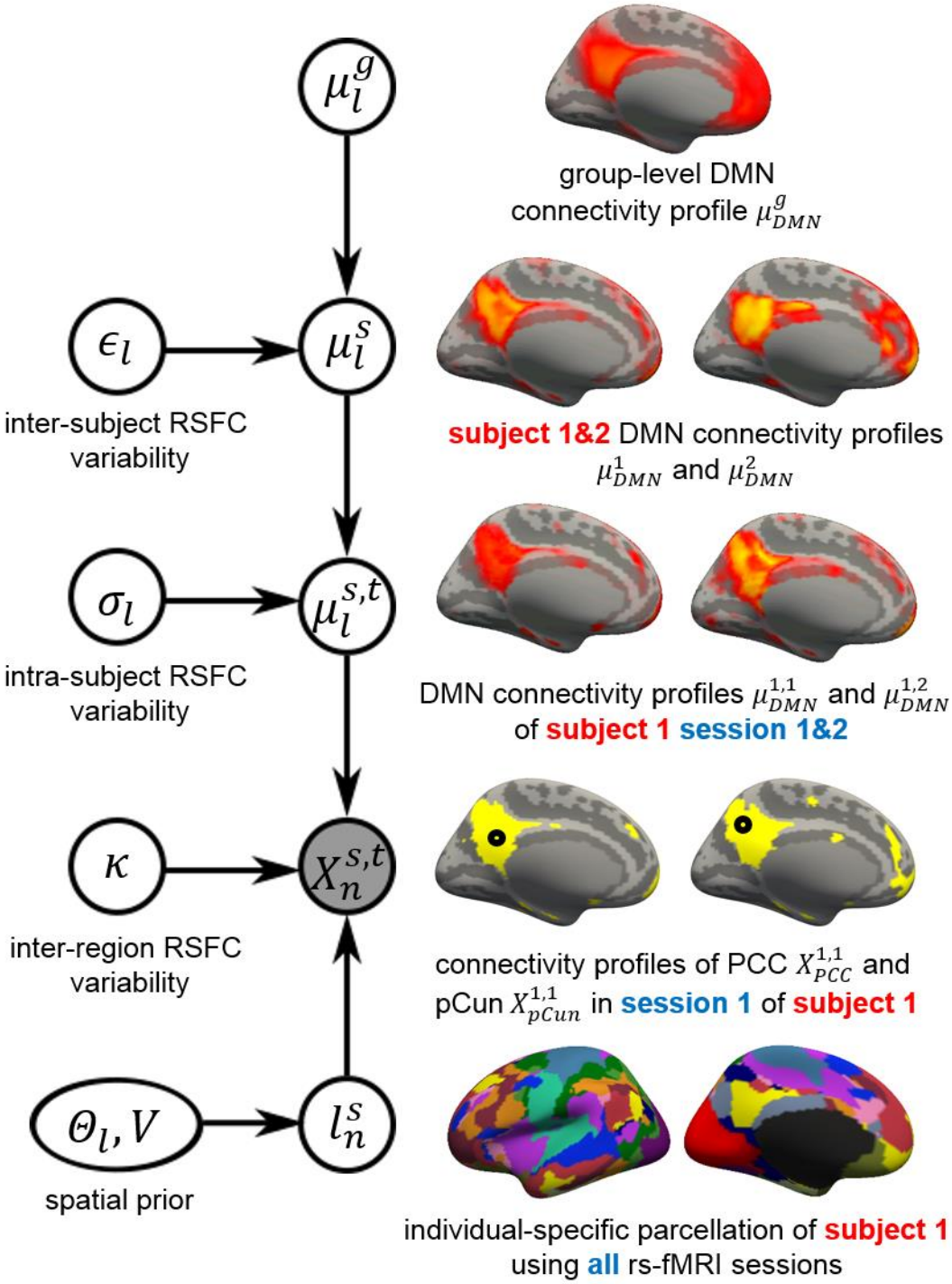

Figure 1. Multi-session hierarchical Bayesian model (MS-HBM) of individual-specific cortical parcellation. $X_n^{s,t}$ denote the RSFC profile at brain location n for subject s during rs-fMRI session t. The shaded circle indicates that $X_n^{s,t}$ are the only observed variables in the entire model. The goal is to estimate the network label l_n^s for subject s at each cortical location n given RSFC profiles from all sessions. μ_l^g is the group-level RSFC profile of network l. μ_l^s is the subject-specific RSFC profile of network l. A large ϵ_l indicates small inter-subject RSFC variability, i.e., the group-level and subject-specific RSFC profiles are very similar. $\mu_l^{s,t}$ is the subject-specific RSFC profile of network l during session t. A large σ_l indicates small intra-subject RSFC variability, i.e., the subject-level and session-level

RSFC profiles are very similar. κ captures inter-region RSFC variability. A large κ indicates small inter-region variability, i.e., two regions from the same network exhibit very similar RSFC profiles. Finally, Θ_l captures inter-subject variability in the spatial distribution of networks (e.g., probability of the DMN being located at the PCC), while smoothness prior V encourages network labels to be spatially smooth. See text for details.

To obtain the subject-specific parcellation, the MS-HBM assumed that each cortical network exhibited a distinctive RSFC profile. Let μ_l^g denote the group-level functional connectivity profile of network l. We can think of μ_l^g as the average connectivity profile of all vertices of all sessions of all subjects belonging to network l. For example, Figure 1 (top row) illustrates the group-level default mode network (DMN) connectivity profile (μ_{DMN}^g).

To model inter-subject RSFC variability, let μ_l^s denote the functional connectivity profile of network l and subject s. We can think of μ_l^s as the average connectivity profile of all vertices of all sessions of subject s belonging to network l. For example, Figure 1 (second row) illustrates the DMN connectivity profiles of two subjects (μ_{DMN}^1 and μ_{DMN}^2). The subject-specific connectivity profile μ_l^s was assumed to follow a von Mises-Fisher distribution with mean direction μ_l^g (group-level RSFC profile of network l) and concentration parameter ϵ_l. A large ϵ_l indicates low inter-subject functional connectivity variability, i.e., μ_l^s and μ_l^g are very similar. The subscript l indicates that ϵ_l is different for each network.

To model intra-subject RSFC variability, let $\mu_l^{s,t}$ denote the functional connectivity profile of network l and subject s during session t. We can think of $\mu_l^{s,t}$ as the average connectivity profile of all vertices from session t of subject s belong to network l. For example, Figure 1 (third row) illustrates the default network connectivity profiles of subject 1 during sessions 1 and 2 ($\mu_{DMN}^{1,1}$ and $\mu_{DMN}^{1,2}$). The session-specific connectivity profiles $\mu_l^{s,t}$ was assumed to follow a von Mises-Fisher distribution with mean direction μ_l^s (subject-specific RSFC profile) and concentration parameter σ_l. A large σ_l indicates low intra-subject functional connectivity variability, i.e., μ_l^s and $\mu_l^{s,t}$ profiles are very similar. The subscript l indicates that σ_l is different for each network.

The observed connectivity profiles of two regions belonging to the same network are unlikely to be identical. For example, the connectivity profiles of PCC and precuneus are similar, but not identical (Figure 1 fourth row) even though they might both belong to the DMN. To account for this intra-network (inter-region) variability, the observed connectivity profile $X_n^{s,t}$ of vertex n (that has been assigned to network l) was assumed to follow a von

Mises-Fisher distribution with mean direction $\mu_l^{s,t}$ (session-specific and subject-specific connectivity profile of network l) and concentration parameter κ. A large κ indicates low inter-region functional connectivity variability. For simplicity, κ was assumed to be the same across networks.

Finally, unlike group averaged connectivity profiles, the observed functional connectivity profiles of individual subjects are generally very noisy. If the observed profiles of PCC and pCun were too noisy, the model might not assign both of them to the same (DMN) network. Therefore, additional priors were imposed on the parcellation. First, the spatial smoothness prior V (Supplemental Methods S1) encouraged neighboring vertices (e.g., PCC and pCun) to be assigned to the same network. Second, the spatial prior $\Theta_{l,n}$ denote the probability of network l occurring at a particular spatial location n. For example, PCC might have high prior probability of being assigned to the default network.

Given a dataset of subjects with multi-session rs-fMRI data, the group-level network connectivity profiles μ_l^g, the inter-subject functional connectivity variability ϵ_l, the intra-subject functional connectivity variability σ_l, the spatial smoothness prior V and the inter-subject spatial variability prior Θ_l could be estimated. Given the estimated model parameters ($\mu_l^g, \epsilon_l, \sigma_l, V, \theta_l$), the parcellation of a new subject could then be inferred. Here we utilized a variational Bayes Expectation-Maximization (VBEM) algorithm to learn the model parameters from the training data and to estimate individual-specific parcellations. Details of the VBEM algorithm can be found in Supplementary Methods S2.

Although the MS-HBM was formulated for multi-session fMRI data, most studies only collect a single run of fMRI data. We considered the ad-hoc approach of splitting the single fMRI run into two and treating the resulting runs as two separate sessions. Our evaluations (see Results) suggest that this workaround worked surprisingly well.

Characterizing inter-subject and intra-subject network variability

We first evaluate whether the MS-HBM can yield robust estimates of inter-subject and intra-subject variability across datasets. For the purpose of subsequent experiments, the GSP dataset was divided into training (N = 37) and validation (N = 32) sets. The CoRR-HNU dataset (N = 30) was kept unchanged. The HCP dataset was divided into training (N = 40), validation (N = 40) and test (N = 596) sets. Furthermore, different fMRI runs within the same session were treated as data from different sessions. For example, each HCP subject underwent two fMRI sessions on two consecutive days. Within each session, there were two

rs-fMRI runs. For the purpose of our analyses, we treated each HCP subject as having four sessions of data. Future work might differentiate between intra-session and inter-session variability.

The group-level parcellation algorithm was applied to the GSP training dataset. The resulting group-level parcellation was then used to initialize the estimation of the group-level network connectivity profiles μ_l^g, the inter-subject functional connectivity variability ϵ_l, the intra-subject functional connectivity variability σ_l, and the inter-subject spatial variability prior Θ_l. For the purpose of computational efficiency, the spatial smoothness prior V was ignored in this analysis[2]. The procedure was repeated for the CoRR-HNU dataset and HCP training set.

Comparison with alternative approaches

Having confirmed previous literature (Mueller et al., 2013; Laumann et al., 2015) that inter-subject and intra-subject functional connectivity variability are indeed different across cortical networks, we compared MS-HBM with four alternative 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 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".

[2] Conceptually, ϵ_l and σ_l are estimated by averaging information across all vertices within each network across all subjects, while Θ_l is estimated for each vertex by averaging information across all subjects. On the other hand, the spatial smoothness V 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 will not significantly affect the estimates of ϵ_l, σ_l, Θ_l.

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 signal-to-noise (SNR) characteristics of the subject's rs-fMRI data. As previously noted, SNR is only one component of intra-subject variability, and consistent with Laumann and colleagues (Laumann et al., 2015), networks with the highest intra-subject variability are not necessarily low SNR regions. We refer to this fourth approach as "Wang2015".

All algorithms were applied to the CoRR-HNU dataset and the HCP test set. In the case of the CoRR-HNU dataset, the model parameters of all algorithms were estimated from the GSP dataset and then utilized to infer the parcellations of CoRR-HNU subjects. This is especially important for the MS-HBM because inter-subject and intra-subject variability might differ across datasets, so it was important to evaluate whether model parameters estimated from one dataset could be generalized to another dataset. More specifically, the training procedure for the MS-HBM was the same as the previous section, except that the GSP validation set was also used to tune the spatial smoothness prior V. Similarly, "free" parameters in Wang2015 and Gordon2017 were tuned using the GSP validation set.

In the case of the HCP dataset, recall that the HCP data were in a different surface space from the GSP data, so the GSP model parameters could not be applied to the HCP subjects. Instead, the model parameters of all algorithms were re-estimated from the HCP training and validation sets, and then utilized to infer the parcellation of each subject in the HCP test set.

Quantitative evaluation measures

Evaluating the quality of individual-specific resting-state parcellations is difficult because of a lack of groundtruth. Here, we considered two common evaluation metrics utilized in previous studies (Gordon et al., 2016; Chong et al., 2017; Gordon et al., 2017c; Schaefer et al., in press): resting-state connectional homogeneity and task functional inhomogeneity measures. These metrics encode the principle that if an individual-specific

parcellation captured the system-level organization of the individual's cerebral cortex, then each network should have homogeneous connectivity and function:

1. Resting-state connectional homogeneity. Resting-state connectional homogeneity was computed by averaging the Pearson's correlations between the resting-state fMRI time courses of all pairs of vertices within each network (Schaefer et al., in press). The average correlations are then averaged across all networks while accounting for network size:

$$\frac{\sum_{l=1}^{L} \rho_l |l|}{\sum_{l=1}^{L} |l|}, \qquad (1)$$

   where ρ_l is the resting-state homogeneity of network l and $|l|$ is the number of vertices within network l (Schaefer et al., in press). For each subject from CoRR-HNU (N = 30) and HCP test set (N = 596), we used one session to infer the individual-specific parcellation and computed the resting-state homogeneity of the individual-specific parcellation with the remaining sessions. Because the HNU dataset has the most amount of data (100 min), we also parcellated each CoRR-HNU subject using one or more fMRI sessions, and evaluated the resting-state homogeneity with the remaining sessions. This allowed us to estimate how much the various algorithms would improve with more data. When comparing between parcellations, a two-sided paired-sample t-test (dof = 29 for CoRR-HNU, dof = 595 for HCP) was performed. It is worth emphasizing that the evaluation utilized new rs-fMRI data not used for estimating the individual-specific parcellations.

2. Task functional inhomogeneity. The HCP task-fMRI data consisted of seven functional domains: social cognition, motor, gambling, working memory, language processing, emotional processing and relational processing, each with multiple task contrasts (Barch et al., 2013). For a given task contrast, task inhomogeneity was defined as the standard deviation of (activation) z-values within each network (Gordon et al., 2017c; Schaefer et al., in press). A lower standard deviation indicates higher functional homogeneity within the network. The standard deviations are averaged across all networks while accounting for network size:

$$\frac{\sum_{l=1}^{L} std_l |l|}{\sum_{l=1}^{L} |l|}, \qquad (2)$$

   where std_l is the standard deviation of task activation z-values for network l and $|l|$ is the number of vertices in parcel l (Gordon et al., 2017c; Schaefer et al., in press).

For each subject in the HCP test set (N = 596), the first rs-fMRI run from the first session was used to infer the individual-specific parcellation. The individual-specific parcellation was then utilized to evaluate task inhomogeneity for each task contrast (Eq. (2)) and then averaged across all contrasts within a functional domain, resulting in a single functional inhomogeneity measure per functional domain. The number of task contrasts per functional domain ranged from three for the emotion domain to eight for the working memory domain. When comparing between parcellations, the inhomogeneity metric (Eq. (2)) was averaged across all contrasts within a functional domain before a two-sided paired-sample t-test (dof = 595) was performed for each functional domain.

We note that a cortical parcellation with more networks would on average perform better on the proposed evaluation metrics. The reason is that a cortical parcellation with more networks will have smaller networks (on average), resulting in higher connectional homogeneity and lower functional inhomogeneity. For example, if a network consisted of only two vertices, then it would be highly homogeneous. However, this was not an issue in our experiments because all approaches were constrained to estimate the same number of networks (i.e., 17 networks). Furthermore, the evaluation metrics (Equations (1) and (2)) accounted for network size, so a network with only two vertices would only contribute minimally to the final homogeneity metric.

Characterizing the MS-HBM parcellations

Having established that the MS-HBM was better than other approaches in generating individual-specific parcellations, we further characterized the reproducibility of individual-specific MS-HBM networks using the CoRR-HNU data and HCP test set. Given that intra-subject and inter-subject network variability were different across networks, we were interested in evaluating whether intra-subject network reproducibility and inter-subject network similarity were also different across networks.

Individual-specific MS-HBM parcellations were independently inferred using the first two runs and the last two runs of the HCP test set. Therefore, there were two individual-specific parcellations for each subject based on data from two independent sets of rs-fMRI data. MS-HBM parcellations were also independently inferred using sessions 1-5 and sessions 6-10 of the CoRR-HNU dataset. Therefore, there were two individual-specific parcellations for each subject based on data from two independent sets of five sessions.

To evaluate the reproducibility of individual-specific parcellations, the Dice coefficient was computed for each network from the two parcellations of each subject. The Dice coefficients were then averaged across all networks and all subjects to provide an overall measure of intra-subject parcellation reproducibility. To evaluate inter-subject parcellation similarity, for each pair of subjects, the Dice coefficient was computed for each network. Since there were two parcellations for each subject, there were a total of four Dice coefficients for each network, which were then averaged. The Dice coefficients were then averaged across all networks and all pairs of subjects to provide an overall measure of inter-subject parcellation similarity.

Does individual-specific network spatial topography predict behavior?

Given that individual-specific functional networks exhibited unique topographical features not observed in group-level networks, we further investigated whether the spatial configuration of individual-specific cortical parcellations was behaviorally meaningful. Since the HCP dataset has a rich repertoire of behavior data, we selected 58 behavioral phenotypes measuring cognition, personality and emotion (Table S1). Individual-specific MS-HBM parcellations were estimated for each HCP test subject (N = 596) using all four rs-fMRI runs. 17 subjects were excluded from further analyses because they did not have all behavioral phenotypes, resulting in a final set of 578 subjects.

Kernel regression (Murphy et al., 2012) was utilized to predict each behavioral phenotype in individual subjects. Suppose y is the behavioral measure (e.g., fluid intelligence) and l is the individual-specific parcellation of a test subject. In addition, suppose y_i is the behavioral measure (e.g., fluid intelligence) and l_i is the individual-specific parcellation of the i-th training subject. Then kernel regression would predict the behavior of the test subject as the weighted average of the behaviors of the training subjects: $y \approx \sum_{i \in \text{training set}} \text{Similarity}(l_i, l) y_i$, where $\text{Similarity}(l_i, l)$ was the Dice overlap coefficient between corresponding networks of the test subject and i-th training subject, averaged across 17 networks. Therefore, successful prediction would indicate that that subjects with more spatially overlapping networks (i.e., network topography) have similar behavioral measures.

In practice, we included an l_2-regularization term (i.e., kernel ridge regression) to reduce overfitting (Supplementary Methods S3; Murphy et al., 2012). The l_2-regularization parameter was determined via an inner-loop cross-validation procedure. More specifically, we performed 20-fold cross-validation for each behavioral phenotype. Care was taken so that

family members were not split between folds. For each test fold, 20-fold cross-validation was repeatedly applied to the remaining 19 folds with different regularization parameters (i.e., inner-loop cross-validation). The optimal regularization parameter from the inner-loop cross-validation was then used to predict the behavioral phenotype in the test fold. Accuracy was measured by correlating the predicted and actual behavioral measure across all subjects within the test fold (Finn et al., 2015). By repeating the procedure for each test fold, each behavior yielded 20 correlation accuracies.

To test whether the predictions were statistically better than chance, prediction accuracies were averaged across all behaviors for *each* of the 20 folds and then a two-sided corrected resampled t-test (dof = 19) was performed (Nadeau and Bengio, 2000; Bouckart and Frank, 2004). The corrected resampled t-test was specifically developed for statistical testing within the cross-validation framework. The test corrected for the fact that prediction accuracies are not independent across folds (because of overlapping training data across folds).

We also considered a permutation testing framework, where prediction accuracies were averaged across all behaviors and 20 folds, and then compared against a null histogram of accuracies obtained by permuting the behavioral data across subjects and repeating the above cross-validation procedure. The permutation procedure respected the family structure in the HCP data (Smith et al., 2015). We found that the corrected resampled t-test and permutation testing led to identical conclusions in practice. Furthermore, the corrected resampled t-test appeared to provide slightly more conservative p values than the permutation procedure, so we will only report the results of the corrected resampled t-test.

It is worth noting that although the previous statistical tests do not rely on assumptions of independence among the behavioral measures, the results might be skewed if a behavioral construct is represented many times (e.g., many different measurements of fluid intelligence). However, the set of 58 behavioral measures was selected to cover a wide range of behavioral attributes across cognition, personality and emotion. Even within the cognitive domain, the measures ranged from working memory to episodic memory and vocabulary.

Finally, we should mention that certain behavioral measures are known to correlate with motion (Siegel et al., 2016). Therefore, age, sex and motion were regressed from the behavioral data before kernel ridge regression. To prevent information from the training data to leak to the test data, for each test fold, the nuisance regression was performed on the training folds and the regression coefficients were applied to the test fold.

Network size versus topography

The kernel regression framework is highly flexible and allowed us to test whether different network properties are responsible for predicting human behavior. In the previous section, a test subject's behavior was predicted based on the similarity between the individual-specific parcellations of the test subject and the training subjects. Similarity was measured using the Dice coefficient, i.e., two parcellations where corresponding networks spatially overlapped significantly would yield a higher Dice coefficient.

Given that sizes of anatomical structures often predict functional capabilities both across and within species (e.g., Krebs et al., 1990; Barton et al., 2009; Erickson et al., 2011; Holmes et al., 2016; Sabuncu et al., 2016), we also investigated whether individual differences in network size could predict behavior. This was achieved by defining the similarity between two parcellations to be the correlation between network sizes. More specifically, let s and s_i be 17 x 1 vectors with the j-th entries corresponding to the surface areas belonging to the j-th network of the test subject and i-th training subject respectively. Here, surface areas were measured in the subjects' native space. Then the similarity between the individual-specific parcellations of the test subject and i-th training subject was set to be the correlation between s and s_j. Therefore, test subject and i-th training subject might be similar due to corresponding networks having similar sizes, even though the networks might not significantly overlap. The two-sided corrected resampled t-test was employed to investigate whether there was statistical difference in the prediction accuracies when using network size, rather than average network overlap (previous section) for predicting behavior[3]. If the prediction accuracies were significantly higher for network overlap (compared with network size), then this would suggest that network topography (location), and not network size, was driving the behavioral prediction.

Topography of task-relevant networks

We also investigated whether the topography of task-relevant networks might contribute to the prediction of various behavioral measures. For example, the frontoparietal

[3] Note that the previous permutation procedure is not appropriate here because the behavioral data are not exchangeable across subjects under the null hypothesis that the prediction accuracies using network size and network overlap are the same. Since the corrected resampled t-test appeared to be providing more conservative p values than the permutation procedure in the previous section, we were confident that the corrected resampled t-test probably provided a reasonable control of type-I error.

control network is typically activated during working memory tasks. Would the topography of the individual-specific frontoparietal control network be more predictive of working memory performance than the topography of all networks?

To explore this question in a systematic fashion, we considered 13 cognitive measures highlighted in the HCP data dictionary (first 13 items in Table S1). For each cognitive measure, we searched for the "forward inference" map of the most relevant term in the NeuroSynth database (Yarkoni et al., 2011). The forward inference map quantified the likelihood that a particular brain voxel was activated in studies using that search term. The reverse inference maps were not considered because they turned out to be extremely sparse. Table S2 shows the search term utilized to obtain the forward inference map for each cognitive measure. There was no appropriate search term for two cognitive measures (processing speed and picture vocabulary), so they were excluded from further analyses. Each forward inference map was projected to fsLR surface space (Buckner et al., 2011; Van Essen et al., 2012a) and compared with the group-level parcellation estimated from the HCP training set, in order to select the task-relevant networks (Table S2). When predicting a particular behavior, the similarity between the test subject and i-th training subject was set to be the dice coefficient averaged across task- relevant networks. For example, in the case of fluid intelligence, Control network A and Dorsal Attention network A overlapped the most with the forward inference map associated with "intelligence" (Table S2). When applying kernel regression to predict fluid intelligence, similarity between the test subject and i-th training subject was set to be the average dice coefficient for Control network A and Dorsal Attention network A. The two-sided corrected resampled t-test was again employed to investigate whether there was statistical difference in the prediction accuracies when using task-relevant networks, rather than all networks (previous section) for predicting behavior.

Code availability

Code for this work is freely available at the github repository maintained by the Computational Brain Imaging Group (https://github.com/ThomasYeoLab/CBIG). More specifically, the GSP and CoRR-HNU datasets were preprocessed using an in-house pipeline (https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/preprocessing/CBIG_fMRI_Preproc2016). The group-level parcellation code (Yeo et al., 2011) are available here (https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Yeo2011_fcMRI_clustering). Finally, the individual-specific parcellation code is also available (GITHUB_LINK_TO_BE_ADDED).

Results

Overview

The MS-HBM (Figure 1) was applied to three multi-session rs-fMRI datasets to ensure that the model can reliably estimate inter-subject and intra-subject variability despite significant acquisition differences across datasets. After confirming previous literature (Mueller et al., 2013; Laumann et al., 2015) that inter-subject and intra-subject RSFC variability were different across networks, we then established that the MS-HBM produced better parcellations than other approaches. Finally, we investigated whether the topography (location and spatial arrangement) and size of individual-specific cortical networks were behaviorally relevant.

Sensory-motor networks exhibit lower inter-subject, but higher intra-subject, functional connectivity variability than association networks.

Figure 2A shows the 17-network population-level parcellation estimated from the HCP training set. The 17 networks were divided into eight groups (Visual, Somatomotor, Auditory, Dorsal Attention, Salience/Ventral Attention, Control, Default and TempPar), which broadly corresponded to major networks discussed in the literature. The 17 networks are referred to as "Default A", "Default B" and so on (Figure 2A).

The HCP population-level parcellation was replicated in the GSP (Figure S1A) and CoRR-HNU (Figure S2A) datasets, although there were some interesting distinctions, likely due to acquisition differences. For example, the Limbic (A and B) networks from the GSP population-level parcellation (Figure S1A) were absorbed into the Default (A and B) networks in the HCP population-level parcellation (Figure 2A). Instead, there were two additional networks in the HCP population-level parcellation: Visual C and Auditory networks. The Visual C network (Figure 2A) might correspond to the foveal representation within the primary visual cortex, while the Auditory network (Figure 2A) appeared to have split off from the Somatomotor B network in the GSP population-level parcellation (Figure S1A). The higher resolution HCP data might allow the separation of the Auditory and Somatomotor B networks, which were in close spatial proximity.

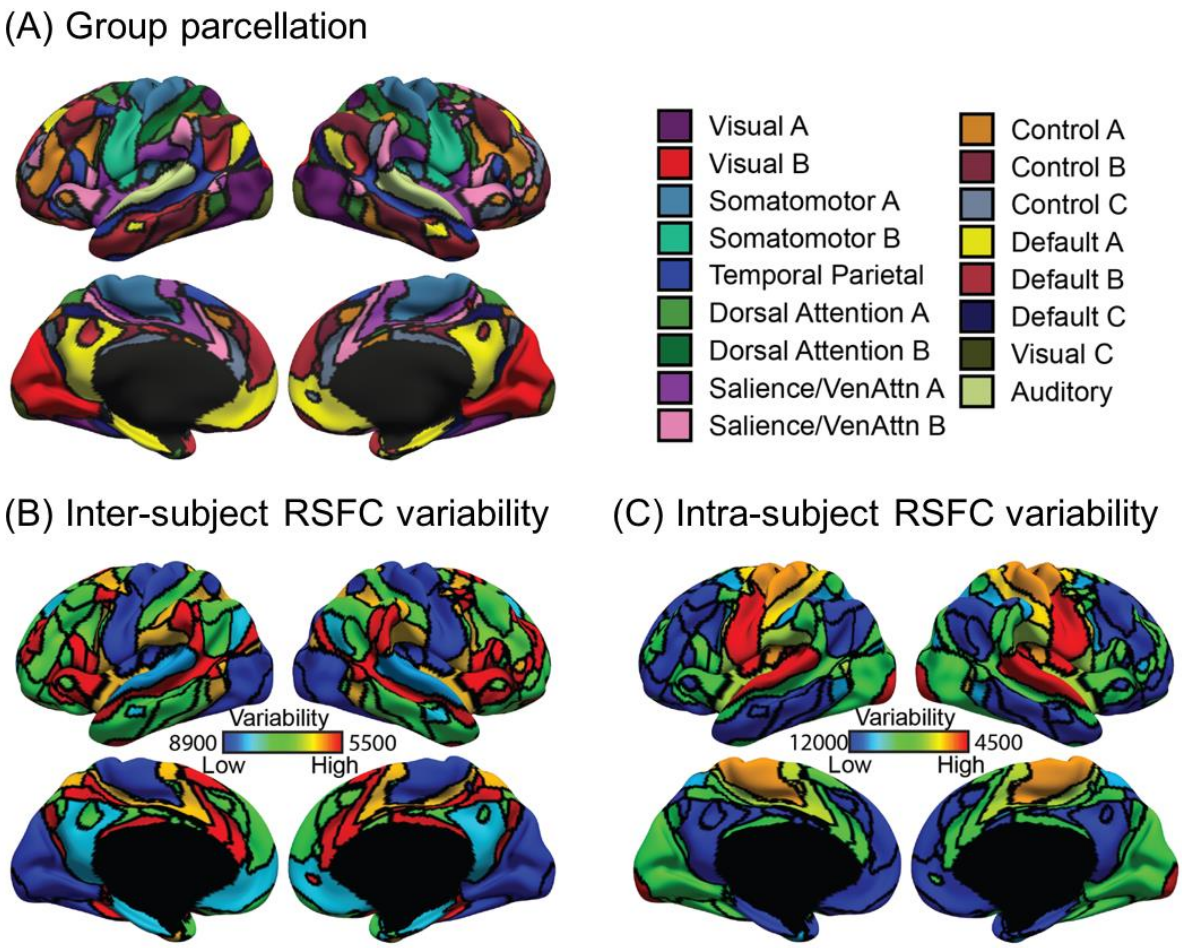

Figure 2. Sensory-motor networks exhibit lower inter-subject, but higher intra-subject, functional connectivity variability than association networks in the HCP training set. (A) 17-network group-level parcellation. (B) Inter-subject functional connectivity variability for different cortical networks. (C) Intra-subject functional connectivity variability for different cortical networks. Results were replicated in the GSP (Figure S1) and CoRR-HNU (Figure S2) datasets. Note that (B) and (C) correspond to the ϵ_l and σ_l parameters in Figure 1, where higher values indicate lower variability.

Recall that the inter-subject functional connectivity variability ϵ_l was estimated for each network. Hence, ϵ_l could be visualized by coloring each corresponding population-level network from Figure 2A. Figure 2B shows the inter-subject functional connectivity variability estimated from the HCP training set. Consistent with Laumann et al. (2015), sensory-motor networks exhibited lower inter-subject functional connectivity variability than association networks. More specifically, Somatomotor (A and B) and Visual (A and B) networks were the least variable, while Salience/Ventral Attention B network was the most variable. The results were largely consistent in the GSP (Figure S1B) and CoRR-HNU (Figure S2B) datasets, although there were some notable differences. For example, the

Somatomotor B network exhibited low variability in both the GSP and HCP datasets, but intermediate variability in the CoRR-HNU dataset.

Similar to ϵ_l, recall that the intra-subject functional connectivity variability σ_l was estimated for each network. Hence, σ_l could be visualized by coloring each corresponding population-level network from Figure 2A. Figure 2C shows the intra-subject functional connectivity variability estimated from the HCP training set. Consistent with Laumann et al. (2015), association networks exhibited lower intra-subject functional connectivity variability than sensory-motor networks. More specifically, Default (A and B) networks were the least variable, while Somatomotor (A and B), Auditory and Visual C networks were the most variable. The results were largely consistent in the GSP (Figure S1C) and CoRR-HNU (Figure S2C) datasets, although there were some interesting differences. Of particular note is that Visual B Network exhibited high intra-subject functional connectivity variability in the GSP dataset, but low or intermediate functional connectivity variability in the CoRR-HNU and HCP datasets. This difference might be due to the fact that subjects were told to fixate on a cross in the CoRR-HNU and HCP datasets, while subjects were told to keep their eyes open (with no fixation cross) in the GSP dataset.

It is worth noting that in the MS-HBM (Figure 1), higher values of ϵ_l and σ_l indicate lower variability. The values in Figure 2C are much larger than Figure 2B, suggesting that intra-subject functional connectivity variability is much lower than inter-subject functional connectivity variability. These results were replicated in the GSP (Figure S1) and CoRR-HNU (Figure S2) datasets.

Sensory-motor networks are less spatially variable than association networks across subjects

The MS-HBM model differentiated between inter-subject RSFC and network spatial variability. Like inter-subject functional connectivity variability, the sensory-motor networks were found to be less spatially variable than association networks across subjects. For example, Figure S3 shows the inter-subject spatial variability maps of four representative networks from the HCP training set. Yellow color at a spatial location indicates that across subjects, there is a high probability of the network appearing at that spatial location, suggesting low inter-subject spatial variability. The Somatomotor A network and Visual B network showed higher probabilities (more yellow color) than the Dorsal Attention networks, suggesting that Somatomotor A network and Visual B network exhibited lower inter-subject

spatial variability than Dorsal Attention networks. These results were consistent in the GSP (Figure S4) and CoRR-HNU (Figure S5) datasets.

Individual-specific networks generated by MS-HBM exhibit higher resting-state homogeneity than other approaches

Individual-specific parcellations were estimated using one rs-fMRI session from the CoRR-HNU dataset and HCP test set. The resting-state homogeneity of the parcellations were evaluated in the leave-out sessions (Figure 3A). Across both CoRR-HNU and HCP datasets, the group-level parcellation (Yeo2011) achieved the worst resting-state homogeneity, while MS-HBM performed the best. In the CoRR-HNU dataset, compared with Yeo2011, YeoBackProject, Gordon2017 and Wang2015, the MS-HBM achieved an improvement of 16.6% ($p = 3.2e-21$), 5.3% ($p = 4.5e-18$), 6.9% ($p = 1.2e-17$) and 4.2% ($p = 1.4e-17$) respectively. In the HCP dataset, compared with Yeo2011, YeoBackProject, Gordon2017 and Wang2015, the MS-HBM achieved an improvement of 9.8% ($p < 5e-324$), 9.5% ($p < 5e-324$), 5.7% ($p < 5e-324$) and 4.4% ($p < 5e-324$) respectively. All significant p-values survived false discovery rate ($q < 0.05$) correction.

Individual-specific parcellations were estimated with increasing number of rs-fMRI sessions using the CoRR-HNU dataset. The resting-state homogeneity of the parcellations were evaluated in the leave-out sessions (Figure 3B). Not surprisingly, performance of the Yeo2011 group-level parcellation remained constant regardless of the amount of data. The remaining approaches (YeoBackProject, Gordon2017, Wang2015 and MS-HBM) exhibited higher homogeneity with increased number of sessions. Critically, the improvement of our model over the other approaches grew with the inclusion of additional fMRI sessions. For example, as the number of sessions was increased from two to three to four to five, the MS-HBM achieved improvement of 5.4%, 5.9%, 6.1% and 6.4% respectively over Gordon2017. Interestingly, the improvement of our approach over Gordon2017 was largest when only one rs-fMRI session was utilized (6.9%). On the other hand, the MS-HBM achieved improvement of 3.6%, 4.4%, 4.8% and 5.0% respectively over Wang2015. Furthermore, using just one fMRI sessions (10 min), the MS-HBM was able to match the homogeneity achieved with the Wang2015 and Gordon2017 approaches using five fMRI sessions (50 min).

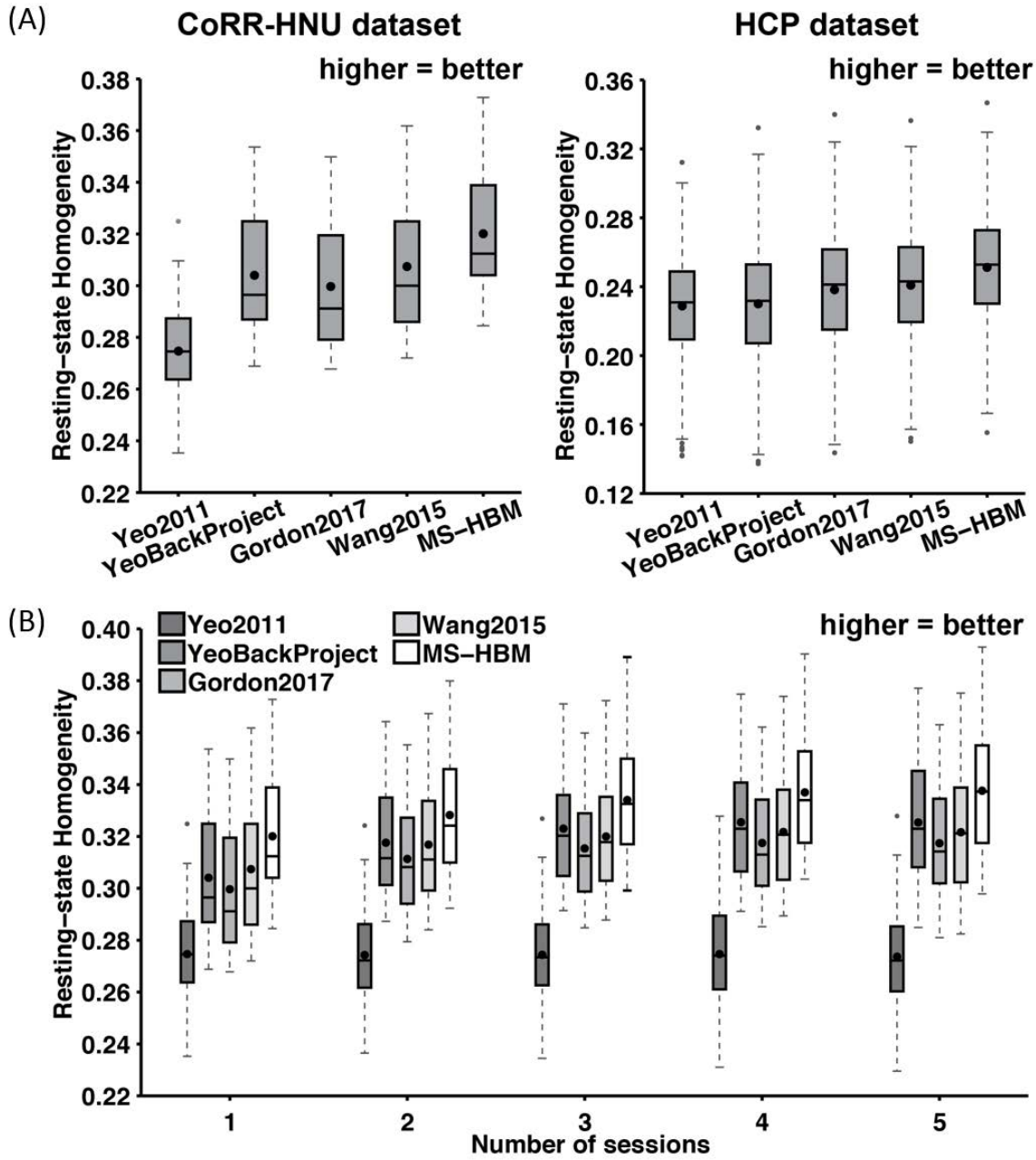

Figure 3. Resting-state homogeneity in the CoRR-HNU and GSP dataset. (A) 17-network individual-specific parcellations were estimated using one rs-fMRI session and resting-state homogeneity were computed on the remaining sessions for each subject from the CoRR-HNU and HCP dataset. (B) 17-network individual-specific parcellations were estimated using different number of rs-fMRI sessions and resting-state homogeneity were computed on the remaining sessions for each subject from the CoRR-HNU dataset. Using just one single fMRI sessions (10 min), the MS-HBM algorithm was able to match the homogeneity achieved with Wang2015 and Gordon2017 using five fMRI sessions (50 min). Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

Individual-specific networks generated by the MS-HBM exhibit lower task functional inhomogeneity than other approaches

Individual-specific parcellations were estimated using one rs-fMRI session (15 min) from the HCP test set. Figure S6 shows the task inhomogeneity of the different approaches. Compared with Yeo2011, YeoBackProject, Gordon2017 and Wang2015, the MS-HBM achieved a small, but statistically significant improvement of 0.63% ($p = 0.003$ for social, $p = 0.02$ for motor, $p < 5e-324$ for other 5 domains), 2.0% ($p < 1.3e-137$ for all domains), 1.04% ($p < 1.5e-89$ for all domains) and 0.7% ($p < 6.3e-64$ for all domains) respectively. All p-values that were less than 0.05 survived false discovery rate ($q < 0.05$) correction. Interestingly, the Yeo2011 group-level parcellation performed as well as (or even better than) YeoBackProject and Gordon2017.

Individual-specific MS-HBM parcellations exhibit high intra-subject reproducibility and low inter-subject similarity

To assess intra-subject reproducibility and inter-subject similarity, our model (Figure 1) was tuned on the HCP training and validation sets, and then applied to the HCP test set. Individual-specific parcellations were generated by using the first two runs and last two runs separately for each subject. Figures 4 and S7 show the parcellations of four representative subjects. The 17 networks were present in all individual-specific parcellations. However, network shapes, sizes and topologies were varied across subjects, consistent with previous studies of individual-specific brain networks (Harrison et al., 2015; Laumann et al., 2015; Wang et al., 2015; Gordon et al., 2017c; Braga and Buckner, 2017).

For example, the Default A (yellow) network exhibited a posterior temporal component for certain subjects (black arrows in Figure 4), but was missing in other subjects. As another example, the two lateral prefrontal components of the Control A (orange) network (Figure 2A) were fused into a single component in certain subjects (green arrows in Figure 4). These features were mostly replicated across sessions. Examples from the CoRR-HNU dataset are shown in Figures S8 and S9.

Figure 5A shows the across-subject spatial similarity (Dice coefficient) of individual-specific parcellations. A higher value (hot color) indicates greater inter-subject agreement. Figure 5B shows the within-subject reproducibility (Dice coefficient) of individual-specific parcellations. A higher value (hot color) indicates greater inter-session agreement within subjects. Further quantification is shown in Figure 5C, where the Dice coefficients were averaged across sub-networks.

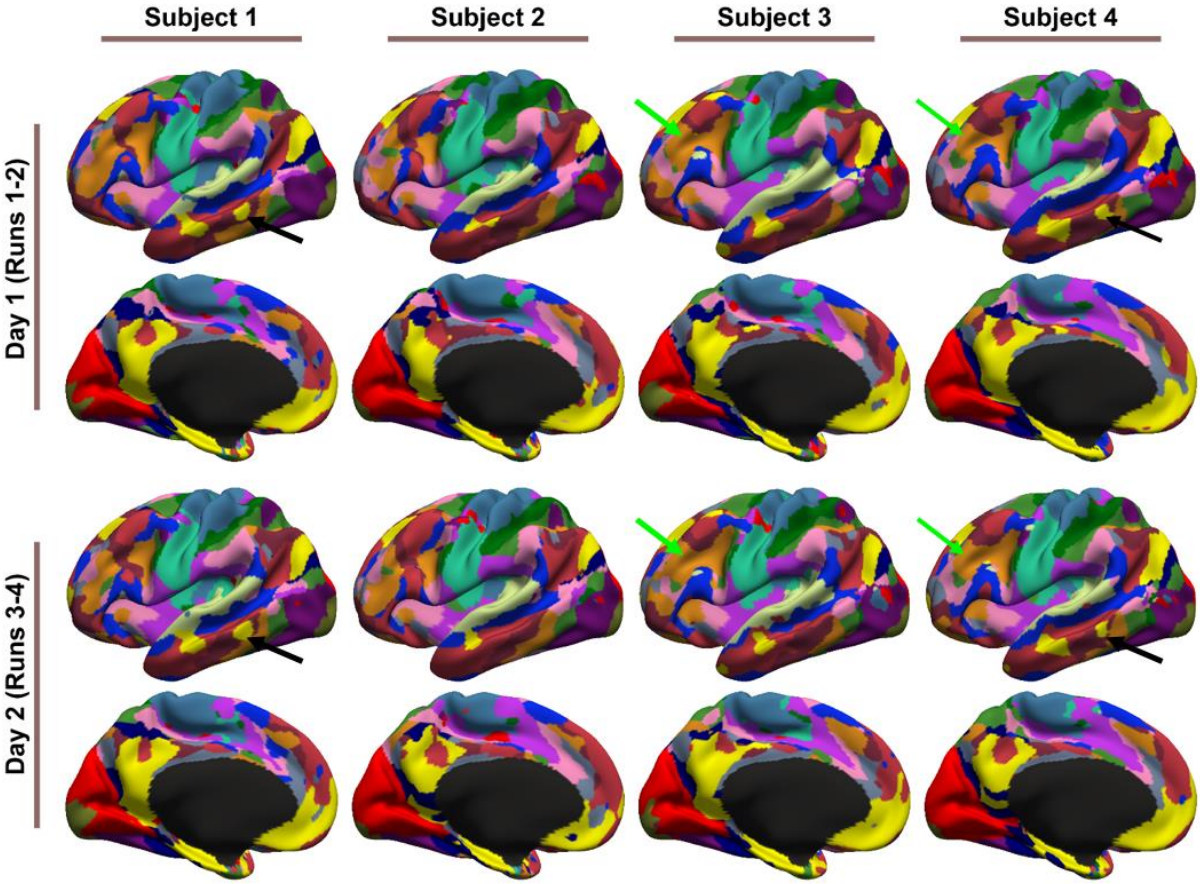

Figure 4. 17-network parcellations were estimated using runs 1-2 and runs 3-4 separately for each subject from the HCP test set. Parcellations of four representative subjects are shown here. Black and green arrows indicate individual-specific parcellation features. Right hemisphere parcellations are shown in Figure S7.

Across all networks, intra-subject reproducibility was greater than inter-subject similarity. Compared with association networks, the Somatomotor (A and B) networks and Visual (A and B) networks were more spatially similar across subjects, but also exhibited greater within subject inter-session reproducibility. Overall, the MS-HBM parcellation model achieved 77.9% intra-subject reproducibility and 65.4% inter-subject similarity.

The results are similar in the CoRR-HNU dataset (Figure S10), although intra-subject reproducibility was higher (81.6%) and inter-subject similarity was lower (59.4%). The improvement might be the result of longer scan duration in the CoRR-HNU dataset (50 min versus 30 min).

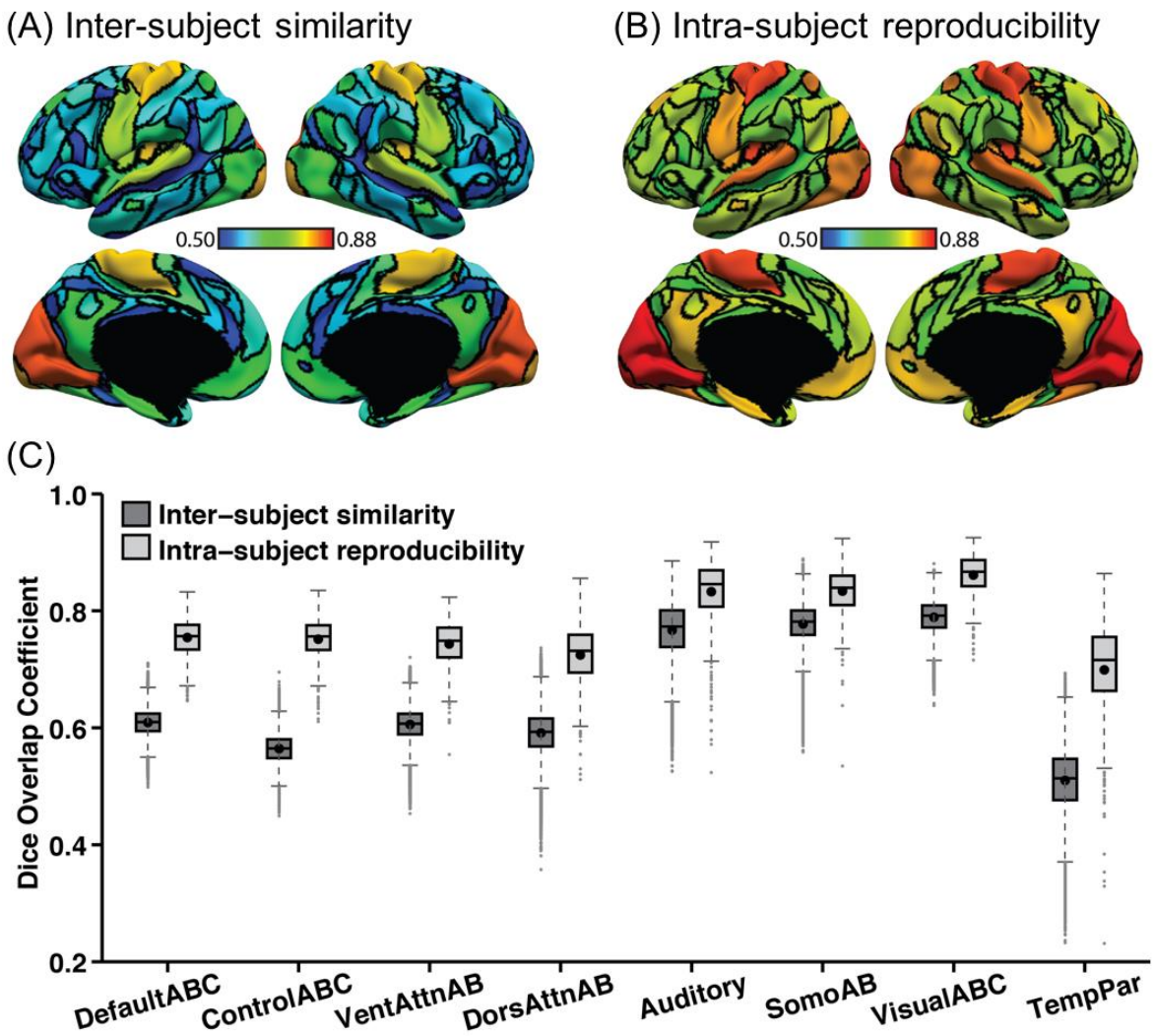

Figure 5. Individual-specific MS-HBM parcellations show high within-subject reproducibility and low across-subject similarity in the HCP test set. (A) Inter-subject spatial similarity for different networks. (B) Intra-subject reproducibility for different networks. Warm color indicates higher overlap. 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. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

Individual differences in network topography can predict cognition, personality and emotion

While it is well-known that individual-specific networks exhibit unique topographic features that are replicable across sessions (Laumann et al., 2015; Gordon et al., 2017c; Braga and Buckner, 2017), a major unanswered question is whether they are behaviorally relevant. Here, we found that individual-specific network topography was able to predict the 58

behavioral measures with an average prediction accuracy of r = 0.084 (p = 1.6e-10). While the accuracy might seem modest, they were comparable to (if not better than) other studies using functional connectivity *strength* for behavioral prediction (HCP MegaTrawl; https://db.humanconnectome.org/megatrawl/; Noble et al., 2017; Dubois et al., biorxiv). For example, of the 58 behavioral measures, 49 of them were also utilized in the HCP MegaTrawl. For the 300-dimensional group-ICA results, HCP MegaTrawl achieved an average accuracy of r = 0.059 (original data space), while kernel regression yielded an average accuracy of r = 0.091.

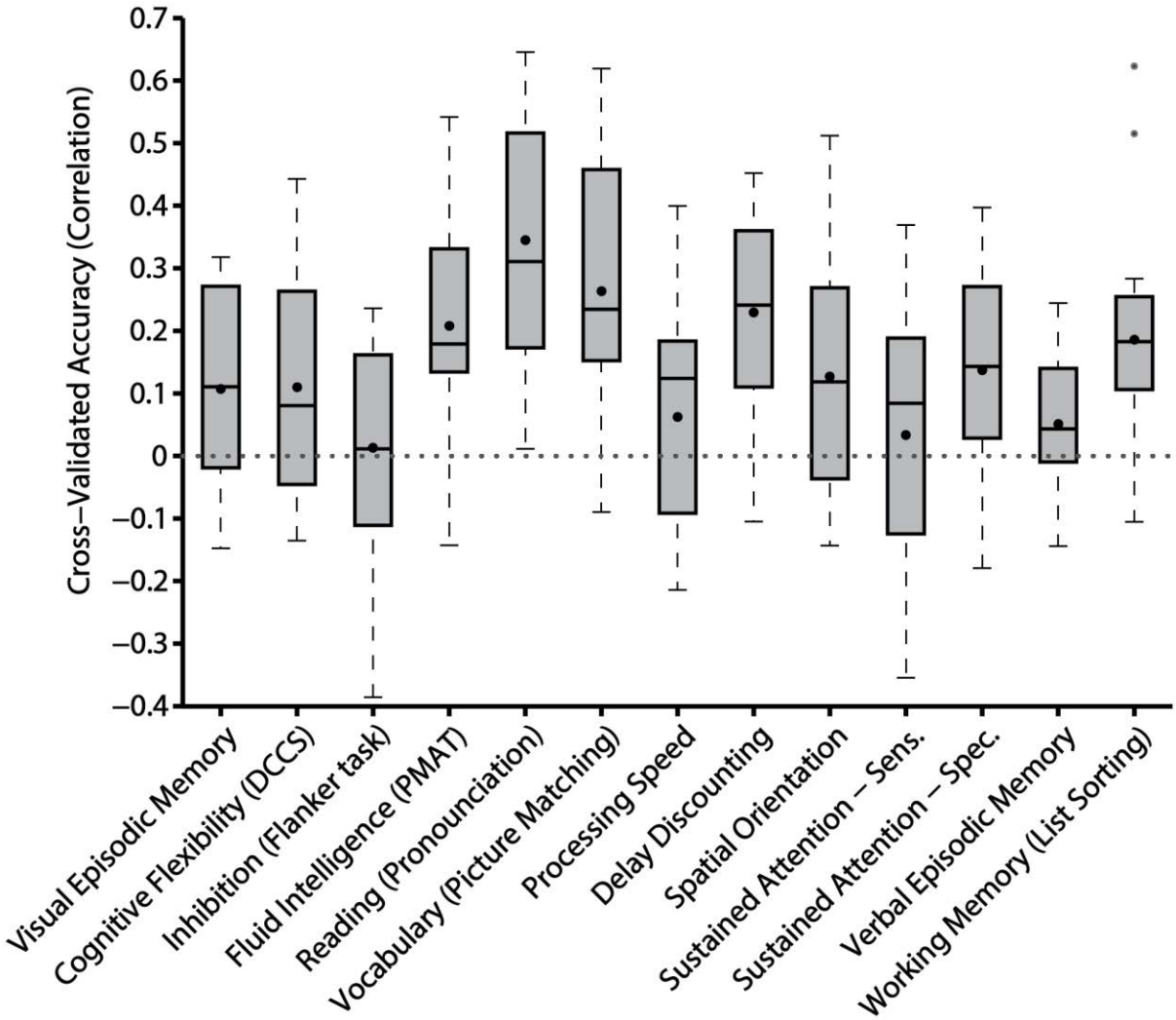

Figure 6. Prediction accuracy of 13 cognitive measures based on inter-subject differences in the spatial arrangement of cortical networks. Average prediction accuracy was r = 0.14 (p = 2e-8) for the 13 measures. Other behavioral measures are found in Figures S11 and S12. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

Figure 6 shows the prediction accuracy for 13 cognitive measures highlighted in the HCP data dictionary. Average prediction accuracy was r = 0.14 (p = 2e-8). The prediction accuracies for the remaining cognitive, emotion and personality measures are found in Figures S11 and S12. In the case of the NEO-5 personality scores (Figure S11), average predication accuracy was r = 0.10 (p = 0.001). Interestingly, the prediction of emotional recognition (Figure S12) was poor with an average prediction accuracy of r = -0.034 (p = 0.18). In the case of the emotional measures (all items in Figure S12 except for emotional recognition), the average prediction accuracy was r = 0.10 (p = 7.6e-4). All significant p-values (i.e., p < 0.05) survived false discovery rate (q < 0.05) correction. As previously mentioned, employing a permutation statistical testing framework yielded identical conclusions with less conservative p values.

Network size versus network topography

When utilizing individual differences in network size to predict behavior, the average accuracy (across 58 behaviors) was r = 0.047 (p = 6.6e-4; Figure 7). This was significantly worse (p = 0.0016) than using network topography (Figure 7). Average prediction accuracy for the 13 cognitive measures was r = 0.067 (p = 0.0021). In the case of the NEO-5 personality scores, average prediction accuracy was r = 0.080 (p = 0.021), while the average prediction accuracy of the emotional measures was r = 0.086 (p = 0.0064). Thus, individual difference in network size (at least at the resolution of large-scale networks) could not account for the ability of network topography to predict behavior.

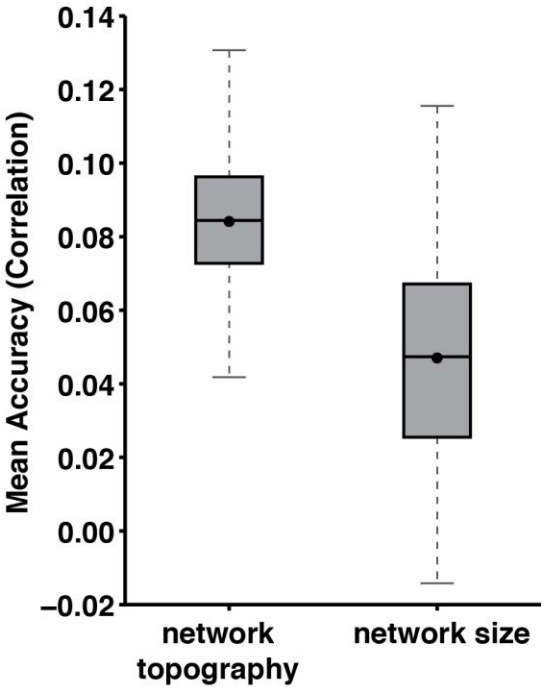

Figure 7. Average prediction accuracy across 58 behaviors based on inter-subject differences in network topography or network size. Average prediction accuracy based on network topography was r = 0.084 (p = 1.6e-10). Average prediction accuracy based on network size was r = 0.047 (p = 6.6e-4). Difference was significant (p = 0.0016). Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

Topography of task-relevant networks

When utilizing the topography of only task-relevant networks (rather than all networks) to predict behavior, the average accuracy across 11 cognitive measures (see Methods) was r = 0.12 (p = 1.7e-7; Figure 8). This was slightly worse than utilizing the topography of all networks (r = 0.14; p = 1.2e-8; Figure 8), but the difference was not statistically significant (p = 0.33).

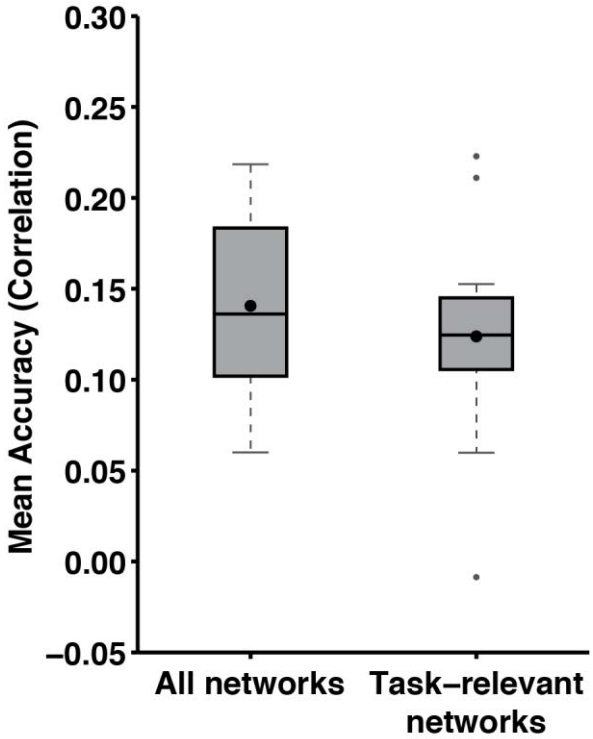

Figure 8. Average prediction accuracy across 11 cognitive measures based on topography of all networks or only task-relevant networks. Average prediction accuracy based on all networks was r = 0.14 (p = 1.2e-8). Average prediction accuracy based only on task-relevant networks was r = 0.12 (p = 1.7e-7). Difference between the two approaches was not significant (p = 0.33). Prediction accuracy of each cognitive measure is found in Figure S13. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

Discussion

Using a novel multi-session hierarchical Bayesian model (MS-HBM), we demonstrate that individually-specific cortical network topography could predict 58 behavioral phenotypes across cognition, personality and emotion. The prediction accuracies could not be accounted for by individual differences in network size. The MS-HBM allowed the joint estimation of inter-subject variability, intra-subject variability and individual-specific cortical networks within the same statistical framework. The resulting MS-HBM individual-specific parcellations were more homogeneous than parcellations derived with four alternative approaches during both resting and task states. These analyses suggest that the *spatial topography* of individuals-specific brain networks might serve as a generalizable fingerprint of human behavior, in contrast to the preponderance of studies utilizing functional connectivity *strength*.

Neurobiological interpretation of MS-HBM networks

Here, we focused on parcellating the cerebral cortex into a small number of (less than twenty) networks. Each spatial (e.g., parietal) component of a network likely spans multiple cytoarchitectonically, functionally and connectionally distinct cortical areas (Kaas 1987; Felleman and Van Essen 1991; Amunts and Zilles 2015; Eickhoff et al., in press) or neurobiological "atoms" (Eickhoff et al., in press; Schaefer et al., in press). One could extend the MS-HBM to estimate a finer division of the cerebral cortex that approximate these finer regions, i.e., individual-specific areal-level parcellation. The main challenge is that because of strong long-range functionally connectivity (Sepulcre et al., 2010), the current model will always result in spatially distributed networks even when estimating large number (e.g., hundreds) of networks. We are working on an additional spatial prior to ensure parcels are spatially localized, but not necessarily spatially connected (Glasser et al., 2016).

By assuming individual-specific parcellations to be the same across sessions, the MS-HBM essentially treats inter-session differences as noise. The implication is that individual-specific MS-HBM parcellations seek to capture stable, trait-like network organization in individuals. However, it is well-known that certain factors (e.g., caffeine intake, sleepiness, attention) result in different brain states and thus functional network organization (Tagliazucchi and Laufs, 2014; Laumann et al., 2015; Poldrack et al., 2015; Yeo et al., 2015b; Wang et al., 2016; Shine et al., 2016). Moreover, in longitudinal studies of certain populations, e.g., Alzheimer's Disease dementia, the goal is to detect neurological changes between consecutive sessions (Misra et al., 2009; Raj et al., 2015; Risacher et al., 2010;

Zhang et al., 2016; Lindemer et al., 2017). To capture transient session-specific or longitudinal changes in brain network organization, the model could be modified to allow for spatial differences in individual-specific parcellations across sessions.

Spatial configuration of individual-specific cortical networks is behaviorally meaningful

Recent work has suggested that individual-specific functional networks exhibit unique topological features not observed in group-level networks (Harrison et al., 2015; Laumann et al., 2015; Wang et al., 2015; Glasser et al., 2016; Langs et al., 2016; Braga & Buckner, 2017; Gordon et al., 2017a; 2017b; 2017c). This was also clearly the case with individual-specific MS-HBM parcellations (Figures 4, S8). While we have pointed out two examples (Default A and Control A networks), it was also obvious that many other individual-specific parcellation features are replicable across sessions.

A major unanswered question in the literature is whether individual differences in cortical parcellations are behaviorally meaningful. Here, kernel regression was utilized to demonstrate that the spatial arrangement of individual-specific cortical networks can be used to predict behavior in individual subjects (Figures 6, S11, S12). More specifically, kernel regression models the possibility that subjects with more similar parcellations exhibited similar behavior. Successful prediction suggests that inter-subject variation in the spatial configuration of cortical networks is strongly related to inter-subject variation in behavior.

Interestingly, the topography of task-relevant networks was not more predictive of cognitive performance than the topography of all networks (Figure 8). It might seem somewhat counterintuitive that the inclusion of task-irrelevant networks (as measured by meta-analysis of task activation) did not dilute prediction accuracy. However, previous studies have suggested that regions not activated by a task might nevertheless exhibit connectivity changes during task performance (Cole et al., 2014; Krienen et al., 2014). Therefore, some of the task-irrelevant networks might potentially be involved in executing a task, despite not being activated through traditional task contrast or subtraction analyses. Alternatively, non-task regions might provide predictive information about traits associated with task performance without direct involvement in the task mechanisms.

It is also worth noting that the task-relevant networks were selected based on forward inference maps extracted from NeuroSynth. However, the forward inference maps do not overlap with the 17 large-scale networks exactly. We hope to revisit this issue in the future by extending MS-HBM to estimate individual-specific areal-level parcellations (see previous section), which would likely match the forward inference maps better.

Finally, a vast body of literature has utilized inter-region functional connectivity *strength* to predict behavior (Hampson et al., 2006; Finn et al., 2015; Rosenberg et al., 2016; Smith et al., 2015; Yeo et al., 2015b), in some cases implicitly, e.g. by binarizing a functional connectivity matrix and then utilizing the resulting graph metrics for behavioral association (van den Heuvel et al., 2009). It would be worthwhile to investigate whether individual differences in network topography and individual differences in functional connectivity strength can be combined to further improve behavioral prediction.

Network size versus network topography

It is well-known that the amount of brain "real estate" devoted to a cognitive function often predicts functional importance or capability both across and within species. For example, nocturnal primates have larger olfactory structures than diurnal primates (Barton et al., 2009). Similarly, the human association cortex, which supports higher cognitive functions, is vastly enlarged relative to other non-human primates (Buckner and Krienen, 2013; Holmes and Patrick, in press). Within humans, larger hippocampal size has also been associated with better memory (Erickson et al., 2011).

Consistent with the vast literature, our results suggest that individual differences in network size can predict human behavior (Figure 7). However, the prediction accuracy is significantly weaker than network topography. This is surprising given that there is little evidence in the literature that the *location* of functional areas or networks might be functionally important.

One important caveat is that most previous studies associating anatomical size with functional importance (e.g., Krebs et al., 1990; Barton et al., 2009) are focused on smaller brain structures, rather than large-scale networks. Therefore, it is possible that prediction accuracies could significantly improve with individual-specific areal-level parcellations. We hope to revisit this issue in the future by extending the MS-HBM to estimate individual-specific areal-level parcellations.

Individual-specific MS-HBM parcellations are more homogeneous than other approaches during resting and task states

If an individual-specific parcellation is capturing the unique network organization of a subject's cerebral cortex, then regions within the same network should have similar resting-state time series, as well as similar activation amplitude for any given task contrast (Gordon et al., 2017c; Schaefer et al., in press). Across the CoRR-HNU and HCP datasets, individual-

specific MS-HBM parcellations exhibited greater resting-state functional connectivity homogeneity than parcellations from four other approaches (Figure 3), suggesting that MS-HBM parcellations better capture the "intrinsic" organization of individuals' cerebral cortex. Importantly, model parameters (e.g., inter-subject and intra-subject variability) estimated from the GSP dataset could improve the estimation of individual-specific parcellations in the CoRR-HNU dataset (Figure 3A). This is important because estimates of inter-subject and intra-subject functional connectivity variability were similar, but not the same across datasets (Figures 2, S1, S2). Therefore, our results suggest that the MS-HBM approach can be used to parcellate individuals from new datasets (using the same preprocessing pipeline), without having to re-estimate the model parameters (e.g., inter-subject and intra-subject functional connectivity variability).

In the HCP dataset, individual-specific MS-HBM parcellations also exhibited greater task functional homogeneity than parcellations from four other approaches (Figure S6), suggesting that MS-HBM parcellations better capture the "extrinsic" organization of individuals' cerebral cortex. Given the strong link between task fMRI and resting-state fMRI (Smith et al., 2009; Mennes et al., 2010; Cole et al., 2014; Krienen et al., 2014; Bertolero et al., 2015; Yeo et al., 2015a; Tavor et al., 2016), this might not seem surprising. However, it is worth pointing out that the group-level parcellation performed as well as, if not better than two of the individual-specific parcellation approaches (Figure S6). Furthermore, the MS-HBM approach only demonstrated small improvements over the group-level parcellation in five of seven functional domains, while there was no statistical difference in the two remaining two functional domains. One possibility is that the large-scale networks were too coarse to capture the finer details of task activation, consistent with our previous observation that the NeuroSynth forward inference maps do not cleanly match subtle features of the group-level large-scale networks. For example, the right-hand motor task preferentially activates the hand region of the left somatomotor cortex. However, Somatomotor A network is bilateral and covers the hand, foot and body regions of bilateral somatomotor cortex. As such, even if individual-specific Somatomotor A network was highly accurate, the resulting task inhomogeneity might still be relatively high.

MS-HBM approach works well with single-session rs-fMRI data

Increasing the scan duration of resting-state fMRI can improve the reliability of functional connectivity measures (Van Dijk et al., 2010; Xu et al., 2016). While earlier studies have suggested that 5 to 12 minutes of resting-state scan might be sufficient to

provide reliable measurements (Van Dijk et al., 2010; Birn et al., 2013), more recent studies have suggested the need for 25 to 30 minutes of data (Anderson et al., 2011; Laumann et al., 2015; Gordon et al., 2017c). However, it is important to note that the amount of data necessary for reliable measurements depends on the functional connectivity measures being computed (Gordon et al., 2017c), as well as the methods employed.

Consistent with previous studies, our experiments showed that the quality of the individual-specific parcellations improved with more rs-fMRI data, although the improvements plateaued after around 30 to 40 minutes of data (Figure 3B). Importantly, even though the MS-HBM was developed for multi-session rs-fMRI, the algorithm performed well even with single-session data. For example, the individual-specific MS-HBM parcellations estimated with one rs-fMRI session (10 minutes) exhibited comparable resting-state connectional homogeneity with parcellations estimated by two prominent approaches using five times the amount of data (Wang et al., 2015; Gordon et al., 2017a, 2017b).

Methodological considerations

Consistent with recent studies (Mueller et al., 2013; Chen et al., 2015; Laumann et al., 2015), we found that association networks exhibited higher inter-subject RSFC variability than sensory-motor networks (Figures 2, S1, S2). One important methodological consideration is that previous studies assumed functional correspondence across subjects after macro-anatomical alignment (Mueller et al., 2013; Chen et al., 2015; Laumann et al., 2015). However, it is well-known that macro-anatomical alignment (or even functional alignment) is not sufficient to achieve perfect functional correspondence across subjects (Fischl et al 1999b; Yeo et al., 2010a; 2010b; Robinson et al., 2014; Harrison et al., 2015; Langs et al., 2016; Glasser et al., 2016). Indeed, a portion of the inter-subject functional connectivity variability observed in previous studies might be the result of residual functional misalignment across subjects (Harrison et al., 2015; Bijsterbosch et al., biorxiv).

Consequently, despite the use of functionally aligned HCP data, we avoided assuming vertex-level functional correspondence provided by the MSMAll functional alignment. Instead, the MS-HBM estimated inter-subject and intra-subject variability at the network level, and also explicitly differentiated between inter-subject spatial variability and inter-subject RSFC variability. This allowed the possibility that for certain networks, inter-subject variability might be attributed to spatial variability, rather than RSFC variability (also see Bijsterbosch et al., biorxiv). Nevertheless, we found that networks with higher inter-subject

functional connectivity variability (Figures 2, S1, S2) also exhibited greater inter-subject spatial variability (Figures S3 to S5).

Although the MS-HBM approach did not account for inter-site variability, we demonstrated that model parameters estimated from one site can generalize to another site with a different acquisition protocol (Figures 3, S8 to S10). Given the increasing availability of multi-session rs-fMRI (Zuo et al., 2014; Holmes et al., 2015; Poldrack et al., 2015; Filevich et al., 2017; Gordon et al., 2017c), it might be possible to add another layer to the hierarchical model to account for inter-site variability, in addition to intra-subject and inter-subject variability. Furthermore, our experiments did not differentiate between rs-fMRI runs collected within the same session versus rs-fMRI runs collected from different sessions. Another layer could again be inserted into the model to differentiate between within-subject intra-session and within-subject inter-session variability. However, we suspect diminishing returns.

Conclusions

We developed a multi-session hierarchical Bayesian model (MS-HBM) that differentiated between inter-subject and intra-subject variability when estimating individual-specific cortical network parcellations. Using a single rs-fMRI session (10 min), our approach yielded parcellations comparable to those estimated by two state-of-the-art algorithms using five rs-fMRI sessions (50 min), as evaluated by generalizability to new rs-fMRI data from the same subjects. Finally, inter-subject variation in the spatial topography (location and arrangement) of cortical networks were strongly related to inter-subject variation in behavior, suggesting their potential utility as fingerprints of human behavior.

Acknowledgement

This work was supported by Singapore MOE Tier 2 (MOE2014-T2-2-016), NUS Strategic Research (DPRT/944/09/14), NUS SOM Aspiration Fund (R185000271720), Singapore NMRC (CBRG/0088/2015), NUS YIA, the Singapore National Research Foundation (NRF) Fellowship (Class of 2017). HL was supported by the National Institute of Mental Health (1R01NS091604, P50MH106435) and the Beijing Municipal Science & Technology Commission (Z161100002616009). AJH was supported by the National Institute of Mental Health (573 K01MH099232). XNZ was supported by the National Basic Research (973) Program (2015CB351702), the Natural Science Foundation of China (81471740, 81220108014), and the Beijing Municipal Science and Tech Commission

(Z161100002616023, Z171100000117012). Our research also utilized resources provided by the Center for Functional Neuroimaging Technologies, P41EB015896 and instruments supported by 1S10RR023401, 1S10RR019307, and 1S10RR023043 from the Athinoula A. Martinos Center for Biomedical Imaging at the Massachusetts General Hospital. Data were also provided by the Brain Genomics Superstruct Project of Harvard University and the Massachusetts General Hospital (Principal Investigators: Randy Buckner, Joshua Roffman, and Jordan Smoller), with support from the Center for Brain Science Neuroinformatics Research Group, the Athinoula A. Martinos Center for Biomedical Imaging, and the Center for Human Genetic Research. Twenty individual investigators at Harvard and MGH generously contributed data to the overall project. Data were also provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

Amunts K, Zilles K. 2015. Architectonic mapping of the human brain beyond Brodmann. Neuron. 88:1086–1107.

Anderson JS, Ferguson MA, Lopez-Larson M, Yurgelun-Todd D. 2011. Reproducibility of Single-Subject Functional Connectivity Measurements. American Journal of Neuroradiology. 32:548–555.

Barton RA, Purvis A, Harvey PH. 1995. Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. Philosophical Transactions of the Royal Society B: Biological Sciences. 348:381–392.

Beckmann C, Mackay C, Filippini N, Smith S. 2009. Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. 15th Annual Meeting of Organization for Human Brain Mapping. poster 441 SU-AM.

Beckmann CF, DeLuca M, Devlin JT, Smith SM. 2005. Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc B Biol Sci. 360:1001–1013.

Bellec P, Rosa-Neto P, Lyttelton OC, Benali H, Evans AC. 2010. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. Neuroimage. 51:1126–1139.

Bertolero MA, Yeo BTT, D'Esposito M. 2015. The modular and integrative functional architecture of the human brain. Proc Natl Acad Sci. 112:E6798–E6807.

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

Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS. 1995. Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. Magn Reson Med. 34:537–541.

Bijsterbosch JD, Woolrich MW, Glasser MF, Robinson EC, Beckmann CF, Van Essen DC, Harrison SJ and Smith SM. 2017. The relationship between spatial configuration and functional connectivity of brain regions. bioRxiv. doi: https://doi.org/10.1101/210195

Bouckaert RR, Frank E. 2004. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. In: Information Processing in Medical Imaging. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 3–12.

Braga RM, Buckner RL. 2017. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. Neuron. 95:457-471.

Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BTT. 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. J Neurophysiol. 106:2322–2345.

Buckner RL, Krienen FM, Yeo BTT. 2013. Opportunities and limitations of intrinsic functional connectivity MRI. Nat Neurosci. 16:832–837.

Buckner RL, Krienen FM. 2013. The evolution of distributed association networks in the human brain. Trends Cogn Sci. 17:648–665.

Burgess GC, Kandala S, Nolan D, Laumann TO, Power JD, Adeyemo B, Harms MP, Petersen SE, Barch DM. 2016. Evaluation of Denoising Strategies to Address Motion-Correlated Artifacts in Resting-State Functional Magnetic Resonance Imaging Data from the Human Connectome Project. Brain Connectivity. 6:669–680.

Calhoun VD, Adali T. 2012. Multisubject Independent Component Analysis of fMRI: A Decade of Intrinsic Networks, Default Mode, and Neurodiagnostic Discovery. IEEE Rev Biomed Eng. 5:60–73.

Calhoun VD, Liu J, Adalı T. 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage. 45:S163-S172.

Chen B, Xu T, Zhou C, Wang L, Yang N, Wang Z, Dong H-M, Yang Z, Zang Y-F, Zuo X-N, Weng X-C. 2015. Individual Variability and Test-Retest Reliability Revealed by Ten Repeated Resting-State Brain Scans over One Month. PLoS ONE. 10:e0144963–21.
Chong M, Bhushan C, Joshi AA, Choi S, Haldar JP, Shattuck DW, Spreng RN, Leahy RM. 2017. Individual parcellation of resting fMRI with a group functional connectivity prior. NeuroImage. 156:87–100.
Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. 2014. Intrinsic and task-evoked network architectures of the human brain. Neuron. 83:238–251.
Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis. I. Segmen- tation and surface reconstruction. Neuroimage. 9:179–194.
Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. 2006. Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci U S A. 103:13848–13853.
Dosenbach NUF, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RAT, Fox MD, Snyder AZ, Vincent JL, Raichle ME, others. 2007. Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci. 104:11073–11078.
Dubois J, Galdi P, Han Yanting, Paul LK, Adolphs R. 2017. Predicting personality traits from resting-state fMRI. doi: https://doi.org/10.1101/215129
Eickhoff SB, Constable RT, Yeo BTT. 2017. Topographic organization of the cerebral cortex and brain cartography. NeuroImage. In press
Erickson KI, Voss MW, Prakash RS, Basak C, Szabo A, Chaddock L, Kim JS, Heo S, Alves H, White SM, Wojcicki TR, Mailey E, Vieira VJ, Martin SA, Pence BD, Woods JA, McAuley E, Kramer AF. 2011. Exercise training increases size of hippocampus and improves memory. Proceedings of the National Academy of Sciences. 108:3017–3022.
Felleman DJ, Van Essen DC. 1991. Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb Cortex. 1:1–47.
Filevich E, Lisofsky N, Becker M, Butler O, Lochstet M, Martensson J, Wenger E, Lindenberger U, Kühn S. 2017. Day2day: investigating daily variability of magnetic resonance imaging measures over half a year. BMC Neuroscience. 18:161.
Filippini N, MacIntosh BJ, Hough MG, Goodwin GM, Frisoni GB, Smith SM, Matthews PM, Beckmann CF, Mackay CE. 2009. Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. Proceedings of the National Academy of Sciences. 106:7209-7214.
Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci. 18:1664-1671.
Fischl B, Liu A, Dale AM. 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging. 20:70–80.
Fischl B, Sereno MI, Dale AM. 1999a. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage. 9:195–207.
Fischl B, Sereno MI, Tootell RBH, Dale AM. 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp. 8:272–284.
Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. 2006. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proc Natl Acad Sci. 103:10046–10051.
Fox MD, Raichle ME. 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci. 8:700–711.
Ge T, Holmes AJ, Buckner RL, Smoller JW, Sabuncu MR. 2017. Heritability analysis with repeat measurements and its application to resting-state functional connectivity. Proceedings of the National Academy of Sciences. 114:5521–5526.

Glahn DC, Winkler AM, Kochunov P, Almasy L, Duggirala R, Carless MA, Curran JC, Olvera RL, Laird AR, Smith SM, Beckmann CF, Fox PT, Blangero J. 2010. Genetic control over the resting brain. Proc Natl Acad Sci U S A. 107:1223–1228.
Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Uğurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC. 2016. A multi-modal parcellation of human cerebral cortex. Nature Publishing Group. 536:171–178.
Glasser MF, Sotiropoulos SN, Wilson JA, Coalson T, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M. 2013. The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage. 80:105-124.
Gordon EM, Laumann TO, Adeyemo B, Gilmore AW, Nelson SM, Dosenbach NUF, Petersen SE. 2017b. Individual-specific features of brain systems identified with resting state functional correlations. NeuroImage. 146:918–939.
Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. 2016. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cereb Cortex. 26:288–303.
Gordon EM, Laumann TO, Adeyemo B, Petersen SE. 2017a. Individual Variability of the System-Level Organization of the Human Brain. Cereb Cortex. 27:386-399.
Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF. 2017c. Precision Functional Mapping of Individual Human Brains. Neuron. 1–25.
Greicius MD, Krasnow B, Reiss AL, Menon V. 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci. 100:253–258.
Greve DN, Fischl B. 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage. 48:63–72.
Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM. 2014. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage. 95:232–247.
Hacker CD, Laumann TO, Szrama NP, Baldassarre A, Snyder AZ, Leuthardt EC, Corbetta M. 2013. Resting state network estimation in individual subjects. Neuroimage. 82:616–633.
Hampson M, Driesen NR, Skudlarski P, Gore JC, Constable RT. 2006. Brain connectivity related to working memory performance. J Neurosci. 26:13338-13343.
Harrison SJ, Woolrich MW, Robinson EC, Glasser MF, Beckmann CF, Jenkinson M, Smith SM. 2015. Large-scale Probabilistic Functional Modes from resting state fMRI. Neuroimage. 109:217–231.
Hawrylycz M, Miller JA, Menon V, Feng D, Dolbeare T, Guillozet-Bongaarts AL, Jegga AG, Aronow BJ, Lee C-K, Bernard A, others. 2015. Canonical genetic signatures of the adult human brain. Nat Neurosci. 18:1832–1844.
Holmes AJ, Hollinshead MO, Roffman JL, Smoller JW, Buckner RL. 2016. Individual Differences in Cognitive Control Circuit Anatomy Link Sensation Seeking, Impulsivity, and Substance Use. Journal of Neuroscience. 36:4038–4049.
Holmes AJ, Hollinshead MO, O'Keefe TM, Petrov VI, Fariello GR, Wald LL, Fischl B, Rosen BR, Mair RW, Roffman JL, Smoller JW, Buckner RL. 2015. Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. Sci data. 2:150031.
Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 17:825–841.

Kaas JH. 1987. The Organization of Neocortex in Mammals: Implications for Theories of Brain Function. Annual Review of Psychology. 38:129–151.
Krebs JR. 1990. Food-Storing Birds: Adaptive Specialization in Brain and Behaviour? Philosophical Transactions of the Royal Society B: Biological Sciences. 329:153–160.
Krienen FM, Yeo BTT, Buckner RL. 2014. Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Phil Trans R Soc B. 369:20130526.
Krienen FM, Yeo BTT, Ge T, Buckner RL, Sherwood CC. 2016. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. Proc Natl Acad Sci. 113:E469–E478.
Langs G, Wang D, Golland P, Mueller S, Pan R, Sabuncu MR, Sun W, Li K, Liu H. 2016. Identifying Shared Brain Networks in Individuals by Decoupling Functional and Anatomical Variability. Cereb Cortex. 26:4004–4014.
Lashkari D, Vul E, Kanwisher N, Golland P. 2010. Discovering structure in the space of fMRI selectivity profiles. NeuroImage. 50:1085–1098.
Laumann TO, Gordon EM, Adeyemo B, Snyder AZ, Joo SJ, Chen M-Y, Gilmore AW, McDermott KB, Nelson SM, Dosenbach NUF, others. 2015. Functional system and areal organization of a highly sampled individual human brain. Neuron. 87:657–670.
Lee MH, Hacker CD, Snyder AZ, Corbetta M, Zhang D, Leuthardt EC, Shimony JS. 2012. Clustering of resting state networks. PLoS One. 7:e40370.
Lindemer ER, Greve DN, Fischl BR, Augustinack JC, Salat DH. 2017. Regional staging of white matter signal abnormalities in aging and Alzheimer's disease. NeuroImage: Clinical. 14:156–165.
Margulies DS, Kelly AMC, Uddin LQ, Biswal BB, Castellanos FX, Milham MP. 2007. Mapping the functional connectivity of anterior cingulate cortex. Neuroimage. 37:579–588.
Mennes M, Kelly C, Zuo X-N, Di Martino A, Biswal BB, Castellanos FX, Milham MP. 2010. Inter-individual differ- ences in resting-state functional connectivity predict task-induced BOLD activity. Neuroimage. 50:1690–1701.
Misra C, Fan Y, Davatzikos C. 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. NeuroImage. 44:1415–1422.
Mueller S, Wang D, Fox MD, Yeo BTT, Sepulcre J, Sabuncu MR, Shafee R, Lu J, Liu H. 2013. Individual variability in functional connectivity architecture of the human brain. Neuron. 77:586–595.
Murphy KP. 2012. Machine Learning. MIT Press.
Nadeau C and Bengio Y. 2000. Inference for the generalization error. In Advances in neural information processing systems, pp. 307-313.
Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. 2017. Influences on the Test–Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. Cerebral Cortex. 27:5415-29.
Poldrack RA, Laumann TO, Koyejo O, Gregory B, Hover A, Chen M-Y, Gorgolewski KJ, Luci J, Joo SJ, Boyd RL, Hunicke-Smith S, Simpson ZB, Caven T, Sochat V, Shine JM, Gordon E, Snyder AZ, Adeyemo B, Petersen SE, Glahn DC, Reese Mckay D, Curran JE, Göring HHH, Carless MA, Blangero J, Dougherty R, Leemans A, Handwerker DA, Frick L, Marcotte EM, Mumford JA. 2015. Long-term neural and physiological phenotyping of a single human. Nature Communications. 6:8885.
Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE. 2011. Functional network organization of the human brain. Neuron. 72:665–678.

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage. 84:320–341.
Raj A, LoCastro E, Kuceyeski A, Tosun D, Relkin N, Weiner M. 2015. Network Diffusion Model of Progression Predicts Longitudinal Patterns of Atrophy and Metabolism in Alzheimer's Disease. Cell Reports. 10:359–369.
Richiardi J, Altmann A, Milazzo A-C, Chang C, Chakravarty MM, Banaschewski T, Barker GJ, Bokde ALW, Bromberg U, Büchel C, others. 2015. Correlated gene expression supports synchronous activity in brain networks. Science (80-). 348:1241–1244.
Risacher SL, Shen L, West JD, Kim S, McDonald BC, Beckett LA, Harvey DJ, Jack CR Jr., Weiner MW, Saykin AJ. 2010. Longitudinal MRI atrophy biomarkers: Relationship to conversion in the ADNI cohort. Neurobiology of Aging. 31:1401–1418.
Robinson EC, Jbabdi S, Glasser MF, Andersson J, Burgess GC, Harms MP, Smith SM, Van Essen DC, Jenkinson M. 2014. MSM: A new flexible framework for Multimodal Surface Matching. NeuroImage. 100:414–426.
Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, Chun MM. 2016. A neuromarker of sustained attention from whole-brain functional connectivity. Nat Neurosci. 19:165–171.
Sabuncu MR, Ge T, Holmes AJ, Smoller JW, Buckner RL, Fischl B, the Alzheimer's Disease Neuroimaging Initiative. 2016. Morphometricity as a measure of the neuroanatomical signature of a trait. Proceedings of the National Academy of Sciences. 113:E5749–E5756.
Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. 2014. Automatic denoising of functional MRI Data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage. 90:449-468.
Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BTT. 2017. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb Cortex. 1–20.
Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci. 27:2349–2356.
Ségonne F, Pacheco J, Fischl B. 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans Med Imaging. 26:518–529.
Sepulcre J, Liu H, Talukdar T, Martincorena I, Yeo BTT, Buckner RL. 2010. The Organization of Local and Distant Functional Connectivity in the Human Brain. PLoS Comput Biol. 6:e1000808–e1000815.
Shine JM, Koyejo O, Poldrack RA. 2016. Temporal metastates are associated with differential patterns of time-resolved connectivity, network topology, and attention. Proceedings of the National Academy of Sciences. 113:9888–9891.
Siegel JS, Mitra A, Laumann TO, Seitzman BA, Raichle M, Corbetta M, Snyder AZ. 2017. Data Quality Influences Observed Links Between Functional Connectivity and Behavior. Cereb Cortex. 27:4492–4502
Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen S, Power J, Salimi-Khorshidi G, Snyder AZ, Vu AT, Woolrich MW, Xu J, Yacoub E, Uğurbil K, Van Essen DC, Glasser MF. 2013. Resting-state fMRI in the Human Connectome Project. Neuroimage. 80:144–168.
Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, others. 2009. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci. 106:13040–13045.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney D, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 23:S208–S219.
Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TE, Glasser MF, Ugurbil K, Barch DM, Van Essen DC, Miller KL. 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat Neurosci. 18:1565-1567.
Tagliazucchi E, Laufs H. 2014. Decoding Wakefulness Levels from Typical fMRI Resting-State Data Reveals Reliable Drifts between Wakefulness and Sleep. Neuron. 82:695–708.
Tavor I, Jones OP, Mars RB, Smith SM, Behrens TE, Jbabdi S. 2016. Task-free MRI predicts individual differences in brain activity during task performance. Science (80-). 352:216–220.
van den Heuvel MP, Stam CJ, Kahn RS, Hulshoff Pol HE. 2009. Efficiency of functional brain networks and intellectual performance. J Neurosci. 29:7619–7624
Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. 2010. Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. Journal of Neurophysiology. 103:297–321.
Van Essen DC, Glasser MF, Dierker DL, Harwell J, Coalson T. 2012a. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. Cereb Cortex. 22:2241-62.
Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, Della Penna S, Feinberg D, Glasser MF, Harel N, Heath AC, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen SE, Prior F, Schlaggar BL, Smith SM, Snyder AZ, Xu J, Yacoub E. 2012b. The Human Connectome Project: A data acquisition perspective. Neuroimage. 62:2222–2231.
Varoquaux G, Gramfort A, Pedregosa F, Michel V, Thirion B. 2011. Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity. In: Information Processing in Medical Imaging. p. 562–573.
Wang C, Ong JL, Patanaik A, Zhou J, Chee MWL. 2016. Spontaneous eyelid closures link vigilance fluctuation with fMRI dynamic connectivity states. Proceedings of the National Academy of Sciences. 113:9653–9658.
Wang D, Buckner RL, Fox MD, Holt DJ, Holmes AJ, Stoecklein S, Langs G, Pan R, Qian T, Li K, Baker JT. 2015. Parcellating cortical functional networks in individuals. Nat Neurosci, 18:1853–1860.
Wig GS, Laumann TO, Cohen AL, Power JD, Nelson SM, Glasser MF, Miezin FM, Snyder AZ, Schlaggar BL, Petersen SE. 2014. Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. Cereb Cortex. 24:2036–2054.
Wig GS. 2017. Segregated Systems of Human Brain Networks. Trends in Cognitive Sciences. 21:981–996.
Xu T, Opitz A, Craddock RC, Wright MJ, Zuo X-N, Milham MP. 2016. Assessing Variations in Areal Organization for the Intrinsic Brain: From Fingerprints to Reliability. Cereb Cortex. 26:4192–4211.
Yang Z, Zuo X-N, McMahon KL, Craddock RC, Kelly C, de Zubicaray GI, Hickie I, Bandettini PA, Castellanos FX, Milham MP, others. 2016. Genetic and Environmental Contributions to Functional Connectivity Architecture of the Human Brain. Cereb Cortex. 26:2341–2352.
Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. 2011. Large-scale

automated synthesis of human functional neuroimaging data. Nature Methods. 8:665–670.
Yeo BTT, Krienen FM, Eickhoff SB, Yaakub SN, Fox PT, Buckner RL, Asplund CL, Chee MWL. 2015a. Functional Specialization and Flexibility in Human Association Cortex. Cereb Cortex. 25:3654–3672.
Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 106:1125–1165.
Yeo BTT, Sabuncu MR, Vercauteren T, Ayache N, Fischl B, Golland P. 2010a. Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration. IEEE Trans Med Imaging. 29:650–668.
Yeo BTT, Sabuncu MR, Vercauteren T, Holt DJ, Amunts K, Zilles K, Golland P, Fischl B. 2010b. Learning Task-Optimal Registration Cost Functions for Localizing Cytoarchitecture and Function in the Cerebral Cortex. IEEE Trans Med Imaging. 29:1424–1441.
Yeo BTT, Tandi J, Chee MWL. 2015b. Functional connectivity during rested wakefulness predicts vulnerability to sleep deprivation. Neuroimage. 111:147–158.
Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BTT, the Alzheimer's Disease Neuroimaging Initiative. 2016. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proceedings of the National Academy of Sciences. 113:E6535–E6544.
Zuo XN, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, Breitner JC, Buckner RL, Calhoun VD, Castellanos FX, Chen A. 2014. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci data. 1:140049.
Zuo XN, Kelly C, Adelstein JS, Klein DF, Castellanos FX, Milham MP. 2010. Reliable intrinsic connectivity networks: Test–retest evaluation using ICA and dual regression approach. NeuroImage. 49:2163–2177.
Zuo XN, Xing XX. 2014. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. Neuroscience and Biobehavioral Reviews. 45:100–118.

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 mathematical and implementation details of the multi-session hierarchical Bayesian model (MS-HBM). Section S1 provides mathematical details about the generative model. Section S2 describes the algorithms for estimating group-level priors and deriving the individual-specific parcellations and how "free" parameters of the model are set. Section S3 describes the kernel ridge regression model for behavioral prediction.

S1. Mathematical model

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 (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 fsLR32k 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 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. κ is assumed to be the same for 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.

S2. Model estimation

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.

S2.1 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. Under this 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.

S2.1.1 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:

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

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

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

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 $q(l_{1:N}^{s}) = \prod_{n=1}^{N} q(l_{n}^{s} | \lambda_{n,1:L}^{s})$:

$$\begin{aligned}
&\log p\left(X_{1:N}^{1:S,1:T} \middle| \mu_{1:L}^{1:S,1:T}, \kappa, \Theta_{1:N,1:L}\right) p\left(\mu_{1:L}^{1:S,1:T} \middle| \sigma_{1:L}, \mu_{1:L}^{1:S}\right) p(\mu_{1:L}^{1:S} | \epsilon_{1:L}, \mu_{1:L}^{g}) \\
&= \sum_{s=1}^{S} \log p\left(X_{1:N}^{s,1:T} \middle| \mu_{1:L}^{s,1:T}, \kappa, \Theta_{1:N,1:L}\right) + \sum_{s=1}^{S} \sum_{l=1}^{L} \log p\left(\mu_{l}^{s,1:T} \middle| \sigma_{l}, \mu_{l}^{s}\right) p\left(\mu_{l}^{s} \middle| \epsilon_{l}, \mu_{l}^{g}\right) && (9) \\
&= \sum_{s=1}^{S} \log \sum_{l_{1:N}^{s}} p\left(X_{1:N}^{s,1:T}, l_{1:N}^{s} \middle| \mu_{1:L}^{s,1:T}, \kappa, \Theta_{1:N,1:L}\right) + \sum_{s=1}^{S} \sum_{l=1}^{L} \log p\left(\mu_{l}^{s,1:T} \middle| \sigma_{l}, \mu_{l}^{s}\right) p\left(\mu_{l}^{s} \middle| \epsilon_{l}, \mu_{l}^{g}\right) && (10) \\
&\geq \sum_{s=1}^{S} \sum_{l_{1:N}^{s}} q(l_{1:N}^{s}) \log \frac{p\left(X_{1:N}^{s,1:T}, l_{1:N}^{s} \middle| \mu_{1:L}^{s,1:T}, \kappa, \Theta_{1:N,1:L}\right)}{q(l_{1:N}^{s})} + \sum_{s=1}^{S} \sum_{l=1}^{L} \log p\left(\mu_{l}^{s,1:T} \middle| \sigma_{l}, \mu_{l}^{s}\right) p\left(\mu_{l}^{s} \middle| \epsilon_{l}, \mu_{l}^{g}\right) && (11) \\
&= \sum_{s=1}^{S} \sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{l_{n}^{s}=1}^{L} \lambda_{n,l_{n}^{s}}^{s} \log p\left(X_{n}^{s,t} \middle| \mu_{l_{n}^{s}}^{s,t}, \kappa\right) + \sum_{s=1}^{S} \sum_{n=1}^{N} \sum_{l_{n}^{s}=1}^{L} \lambda_{n,l_{n}^{s}}^{s} \log \Theta_{n,L_{n}^{s}} \\
&\qquad - \sum_{s=1}^{S} \sum_{n=1}^{N} \sum_{l_{n}^{s}=1}^{L} \lambda_{n,l_{n}^{s}}^{s} \log \lambda_{n,l_{n}^{s}}^{s} + \sum_{s=1}^{S} \sum_{l=1}^{L} \left(\sum_{t=1}^{T} \log p(\mu^{s,t} | \sigma_{l}, \mu_{l}^{s}) + \log p(\mu_{l}^{s} | \epsilon_{l}, \mu_{l}^{g})\right) && (12) \\
&= \mathcal{L}(\lambda, \Omega), && (13)
\end{aligned}$$

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:

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

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:

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

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

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

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:

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

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

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

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.

S2.1.2 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.

S2.2 Individual-level parcellation estimation

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

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

$$\underset{\Psi}{\operatorname{argmax}} \log p\left(\kappa, \mu_{1:L}^{s,1:T}, \mu_{1:L}^{s} \middle| X_{1:N}^{s,1:T}, \epsilon_{1:L}, \sigma_{1:L}, \mu_{1:L}^{g}, \Theta_{1:N,1:L}\right). \tag{28}$$

Assuming a uniform (improper) prior on κ, and by introducing the parcellation labels $l_{1:N}^{s}$ 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:

$$\begin{aligned}\mathcal{L}(\lambda^{s}, \Psi) = &\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^{s} \log p\left(X_n^{s,t} \middle| \mu_{l_n^s}^{s,t}, \kappa\right) + \alpha \sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^{s} \log \Theta_{n,L_n^s} \\ &- c \sum_{n=1}^{N}\sum_{m\in\mathcal{N}_n}\sum_{l_n^s=1}^{L}\sum_{l_m^s=1}^{L} \lambda_{n,l_n^s}^{s} \lambda_{m,l_m^s}^{s} V(l_n^s, l_m^s) \\ &- \sum_{n=1}^{N}\sum_{l_n^s=1}^{L} \lambda_{n,l_n^s}^{s} \log \lambda_{n,l_n^s}^{s} + \sum_{l=1}^{L}\left(\sum_{t=1}^{T} \log p(\mu^{s,t} | \sigma_l, \mu_l^s) + \log p\left(\mu_l^s \middle| \epsilon_l, \mu_l^g\right)\right),\end{aligned} \tag{29}$$

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:

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

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:

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

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

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

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

S3. 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 $\boldsymbol{y} = [y_1, y_2, \dots, y_M]^T$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_3]^T$ 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 $\boldsymbol{K}_s = [K(l_1, l_s), K(l_2, l_s), \ldots, K(l_M, l_s)]$, 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

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

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

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

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.

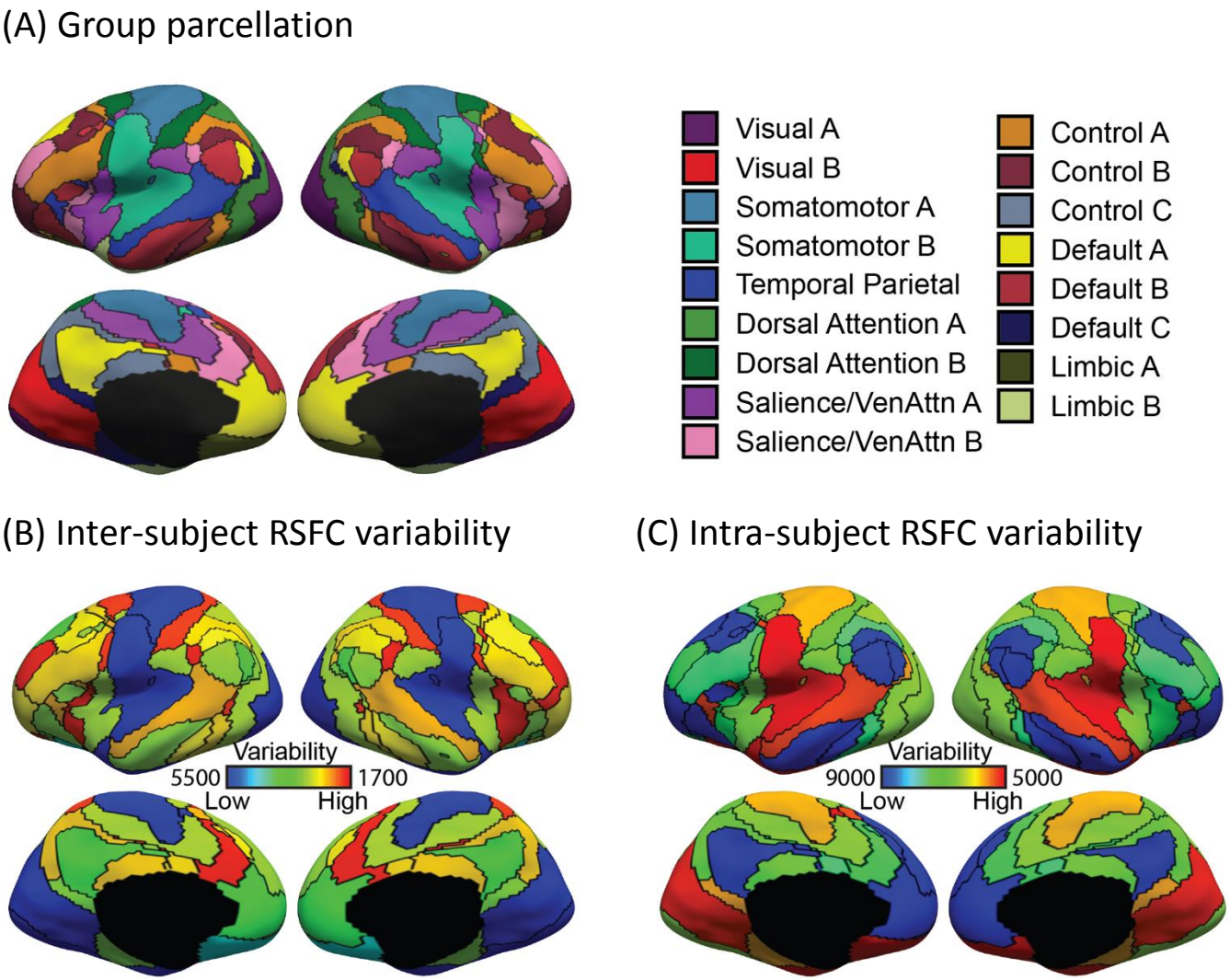

Figure S1. Sensory-motor networks exhibit lower inter-subject, but higher intra-subject, functional connectivity variability than association networks in the GSP training set. (A) 17-network group-level parcellation. (B) Inter-subject functional connectivity variability for different cortical networks.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

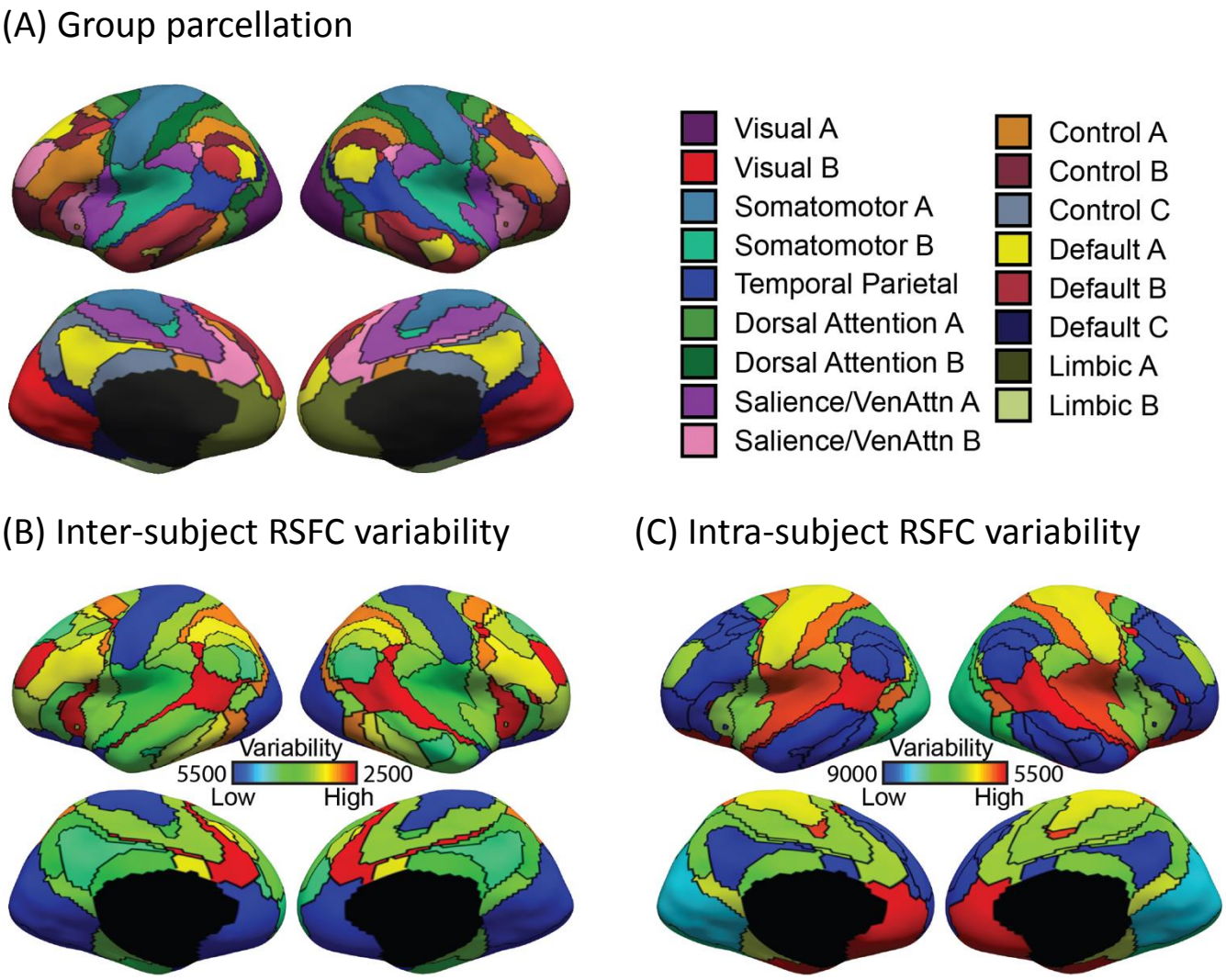

Figure S2. Sensory-motor networks exhibit lower inter-subject, but higher intra-subject, functional connectivity variability than association networks in the CoRR-HNU dataset. (A) 17-network group-level parcellation. (B) Inter-subject functional connectivity variability for different cortical networks.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

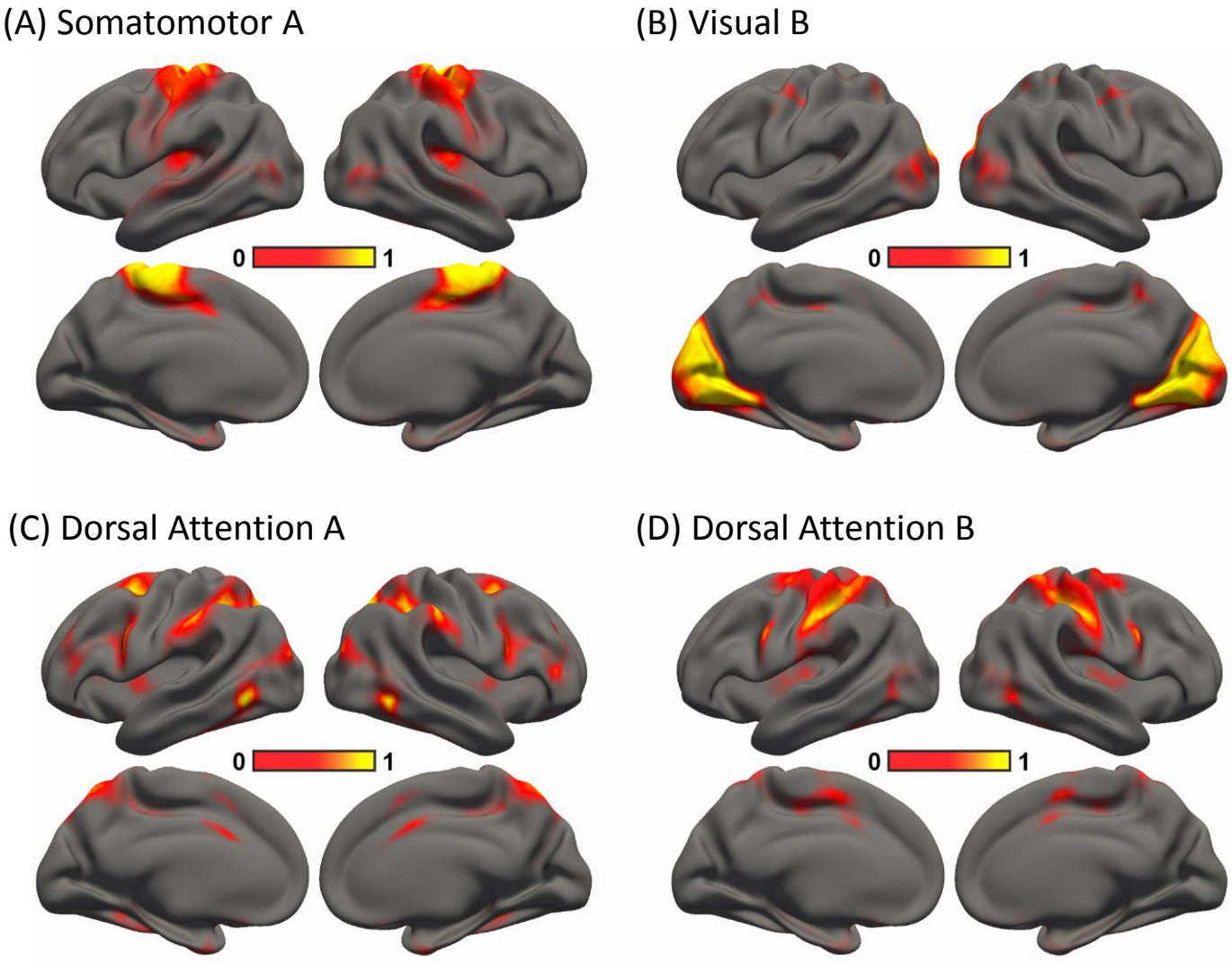

Figure S3. Sensory-motor networks are less spatially variable than association networks across subjects in the HCP training set. Spatial probability maps of (A) Somatomotor network A, (B) Visual network B, (C) Dorsal Attention network A, and (D) Dorsal Attention network B. A higher value (bright color) 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.

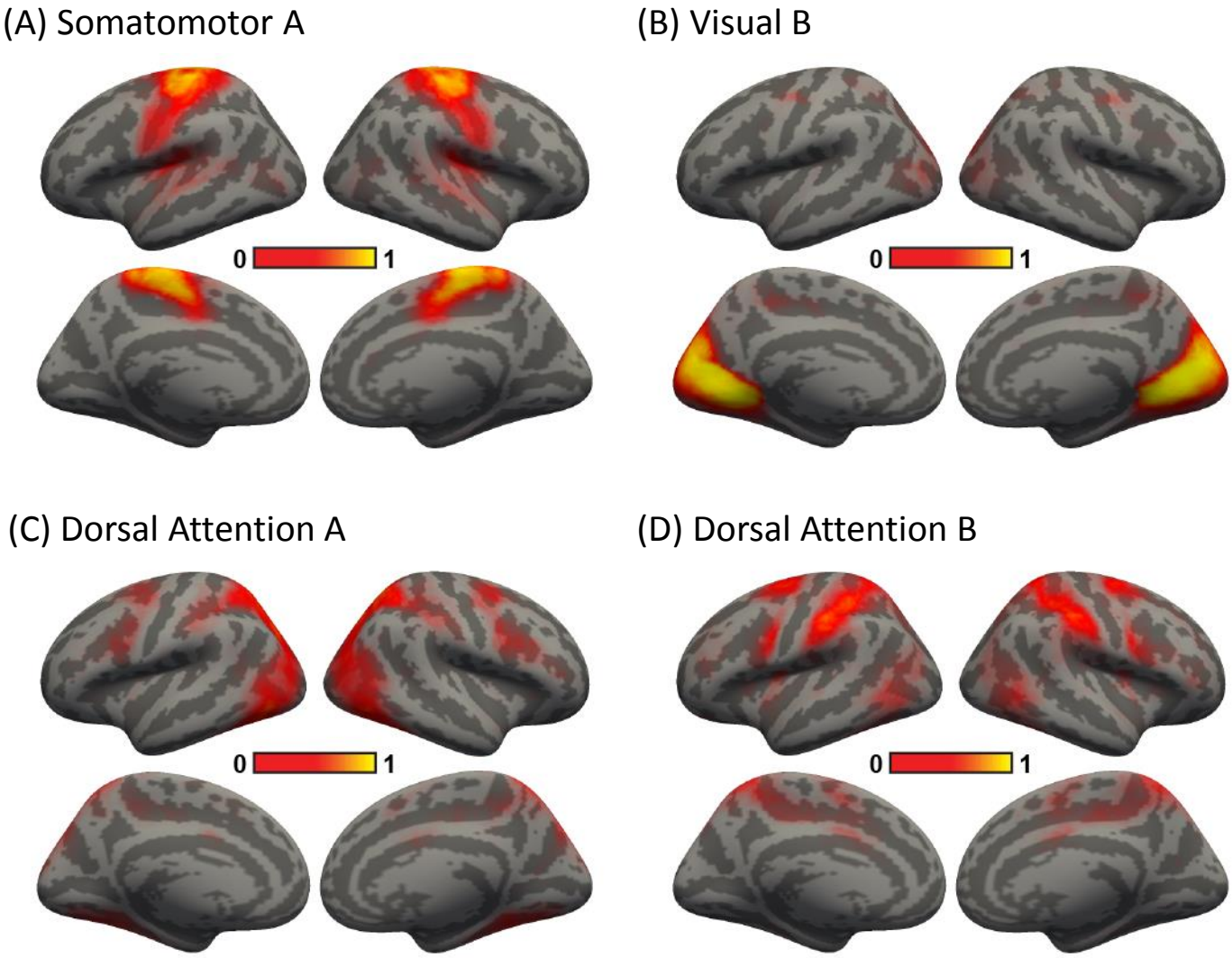

Figure S4. Sensory-motor networks are less spatially variable than association networks across subjects in the GSP dataset. Spatial probability maps of (A) Somatomotor network A, (B) Visual network B, (C) Dorsal Attention network A, and (D) Dorsal Attention network B. A higher value (bright color) at 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.

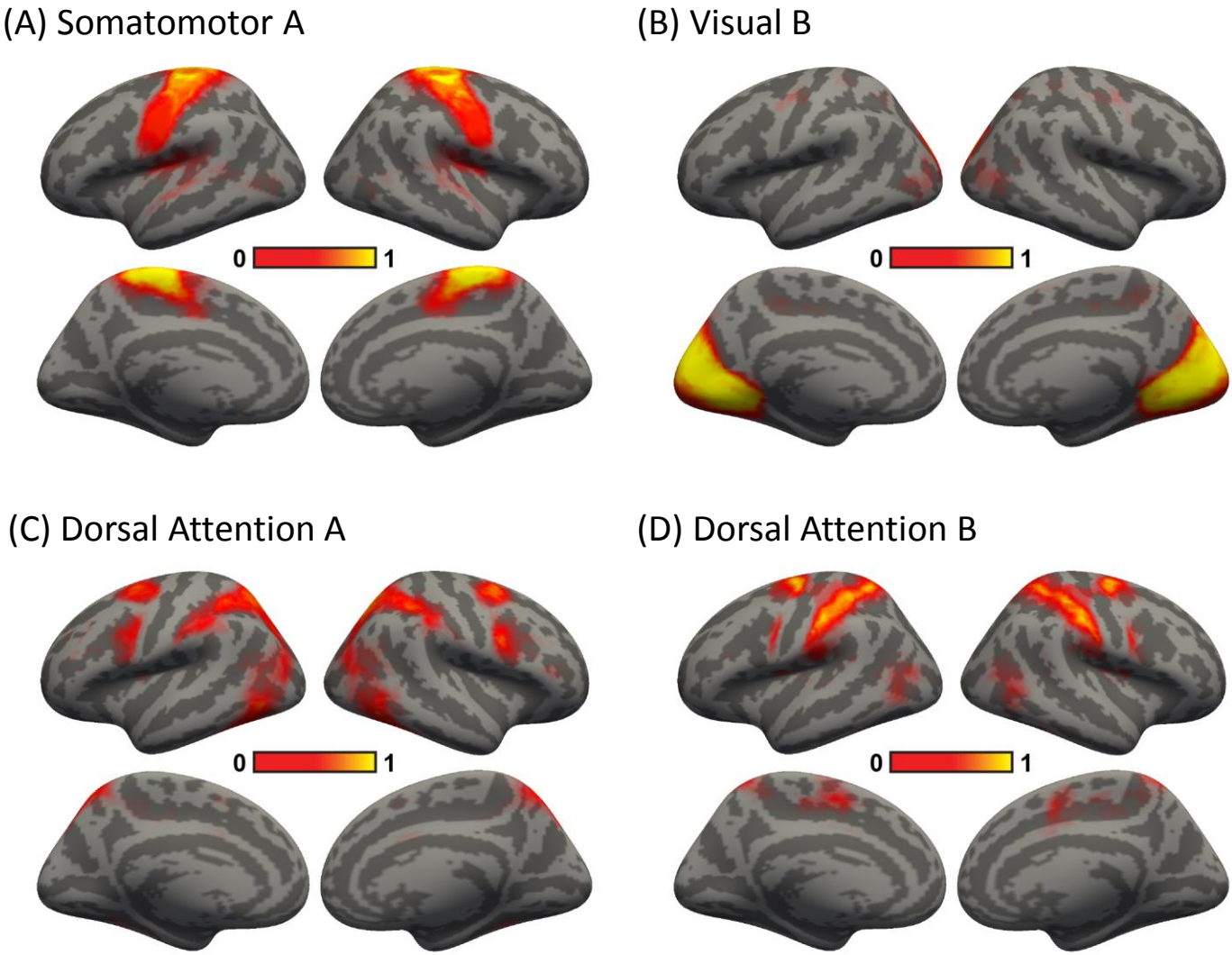

Figure S5. Sensory-motor networks are less spatially variable than association networks across subjects in the CoRR-HNU dataset. Spatial probability maps of (A) Somatomotor network A, (B) Visual network B, (C) Dorsal Attention network A, and (D) Dorsal Attention network B. A higher value (bright color) 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.

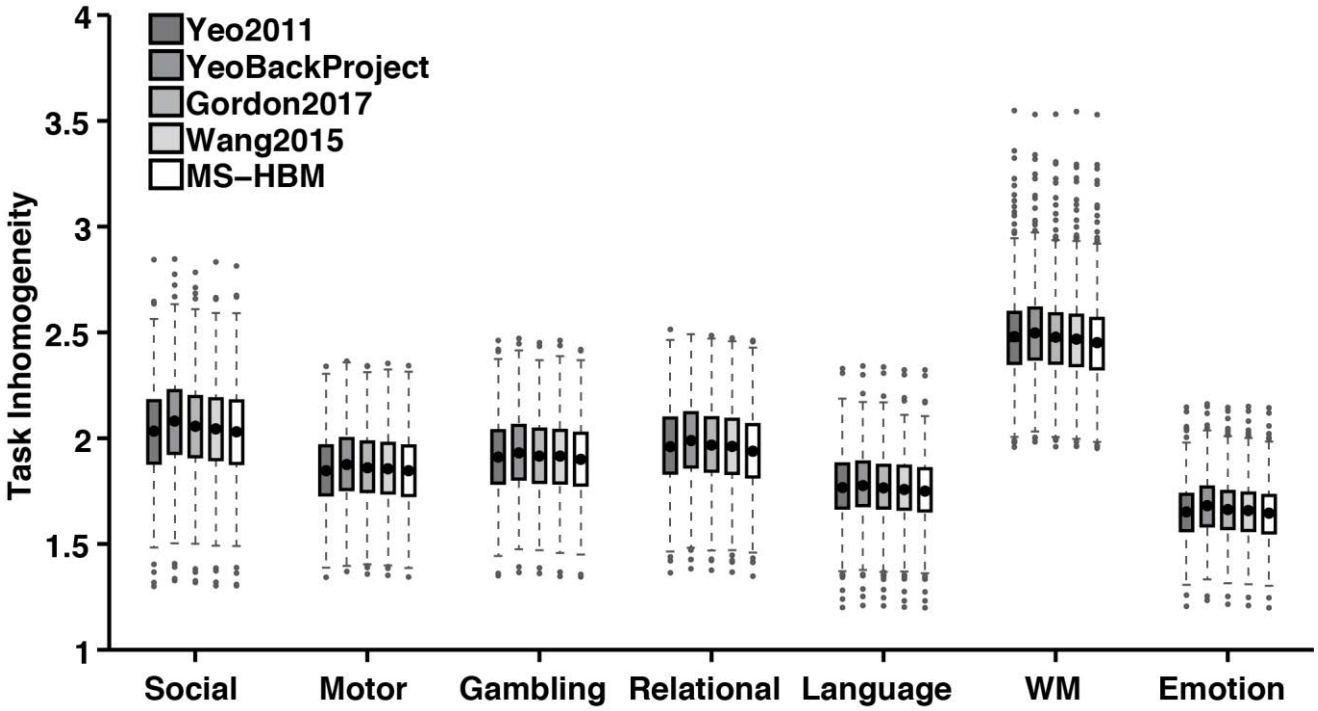

Figure S6. Task inhomogeneity of resting-state parcellations in the HCP dataset. 17-network individual-specific parcellations were estimated using one rs-fMRI session. Task inhomogeneity was then defined as the standard deviation of task activation within each network, and then averaged across all networks and contrasts within each behavioral domain. Lower value indicates better functional homogeneity. Compared with Yeo2011, YeoBackProject, Gordon2017 and Wang2015, the MS-HBM individual-specific parcellations achieved a modest average improvement of 0.63% ($p = 0.003$ for social, $p = 0.02$ for motor, $p < 5e\text{-}324$ for other 5 domains), 2.0% ($p < 1.3e\text{-}137$ for all domains), 1.04% ($p < 1.5e\text{-}89$ for all domains) and 0.7% ($p < 6.3e\text{-}64$ for all domains) respectively. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

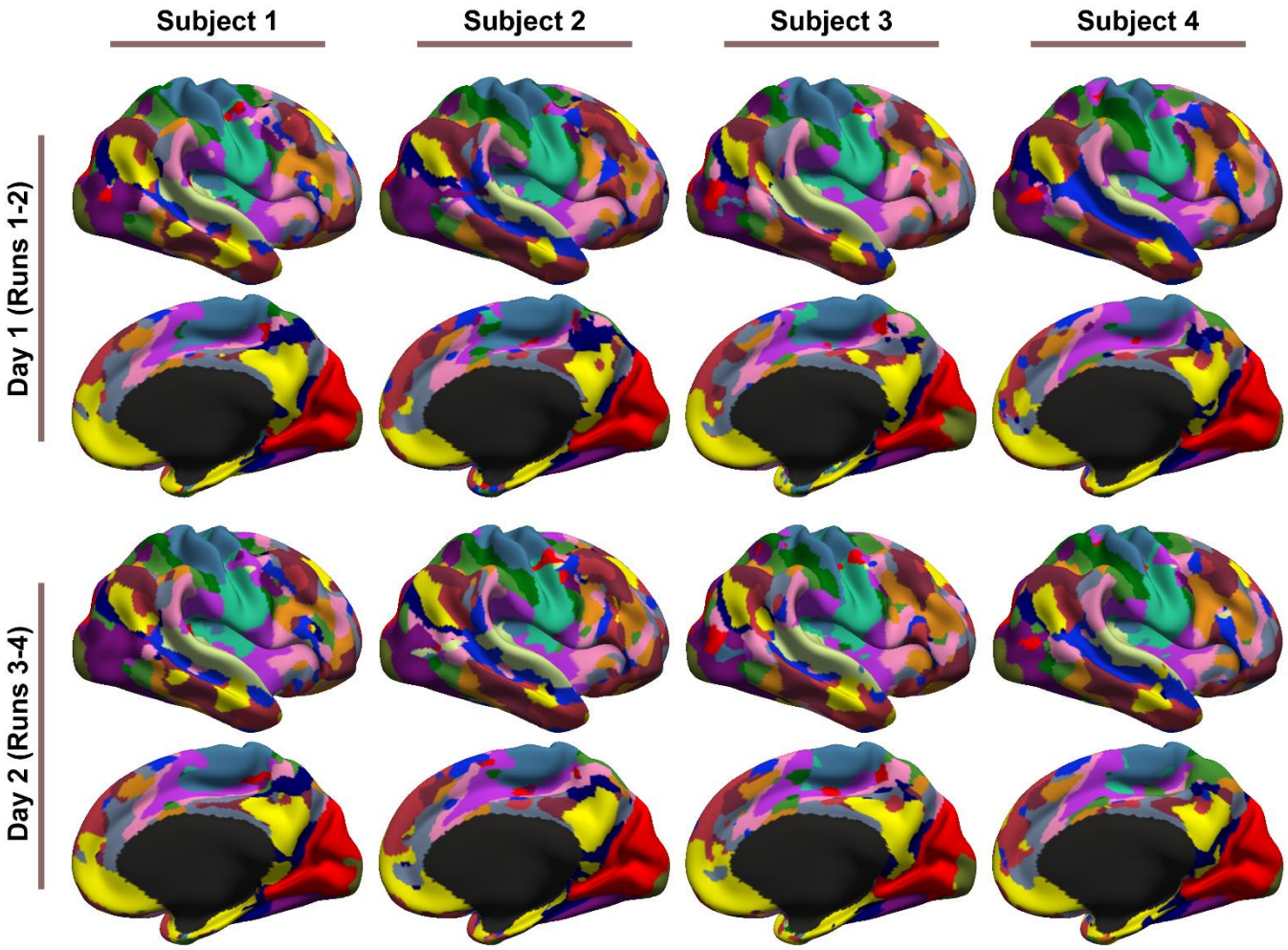

Figure S7. 17-network parcellations were estimated using runs 1-2 and runs 3-4 separately for each subject from the HCP test set. Parcellations of four representative subjects are shown here. Left hemisphere parcellations are shown in Figure 4.

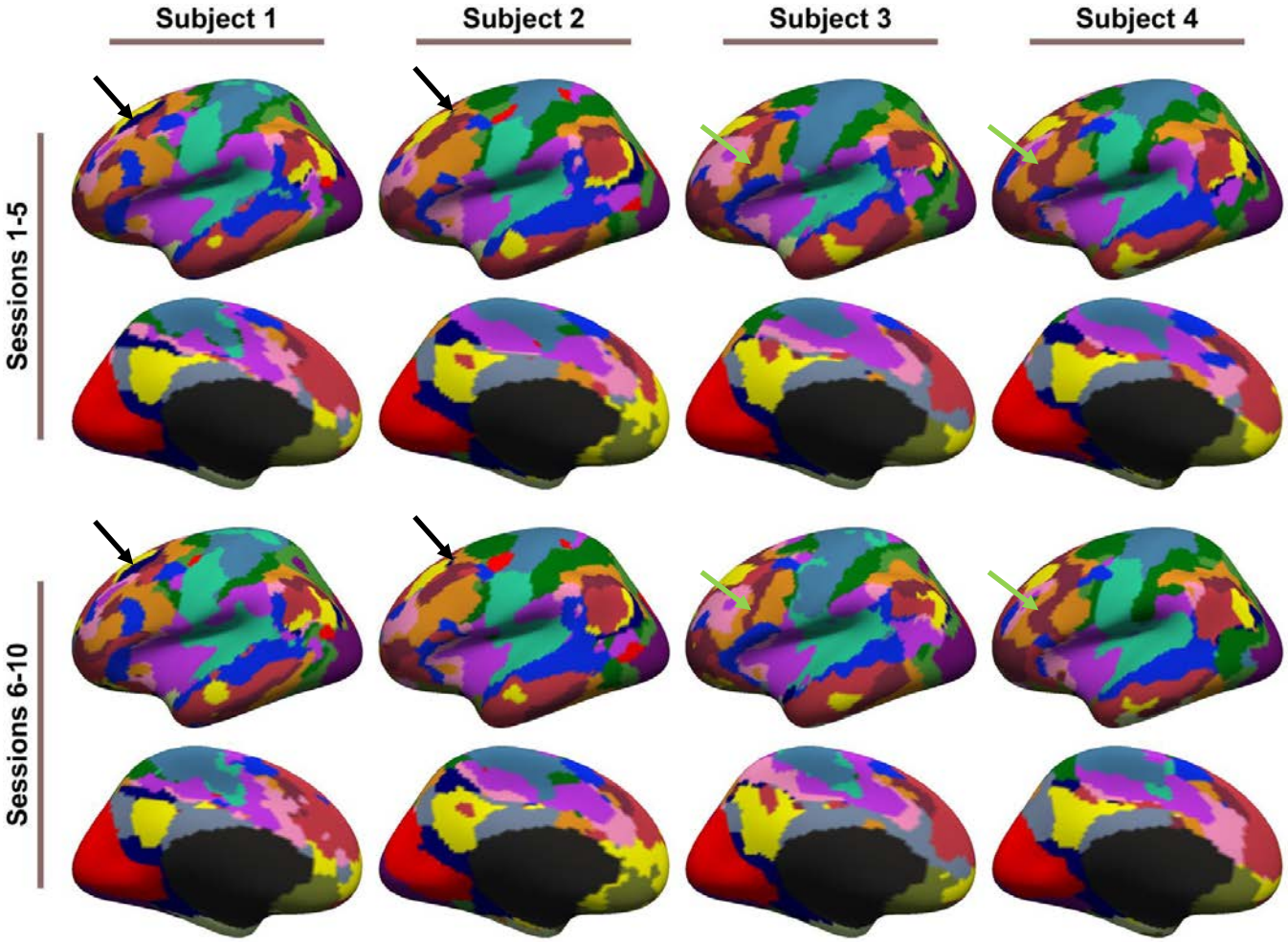

Figure S8. 17-network parcellations were estimated using sessions 1-5 and sessions 6-10 separately for each subject from the CoRR-HNU dataset. Parcellations of four representative subjects are shown here. Black and green arrows indicate individual-specific parcellation features. The Default C (dark blue) network exhibited a dorsal prefrontal component for certain subjects (black 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.

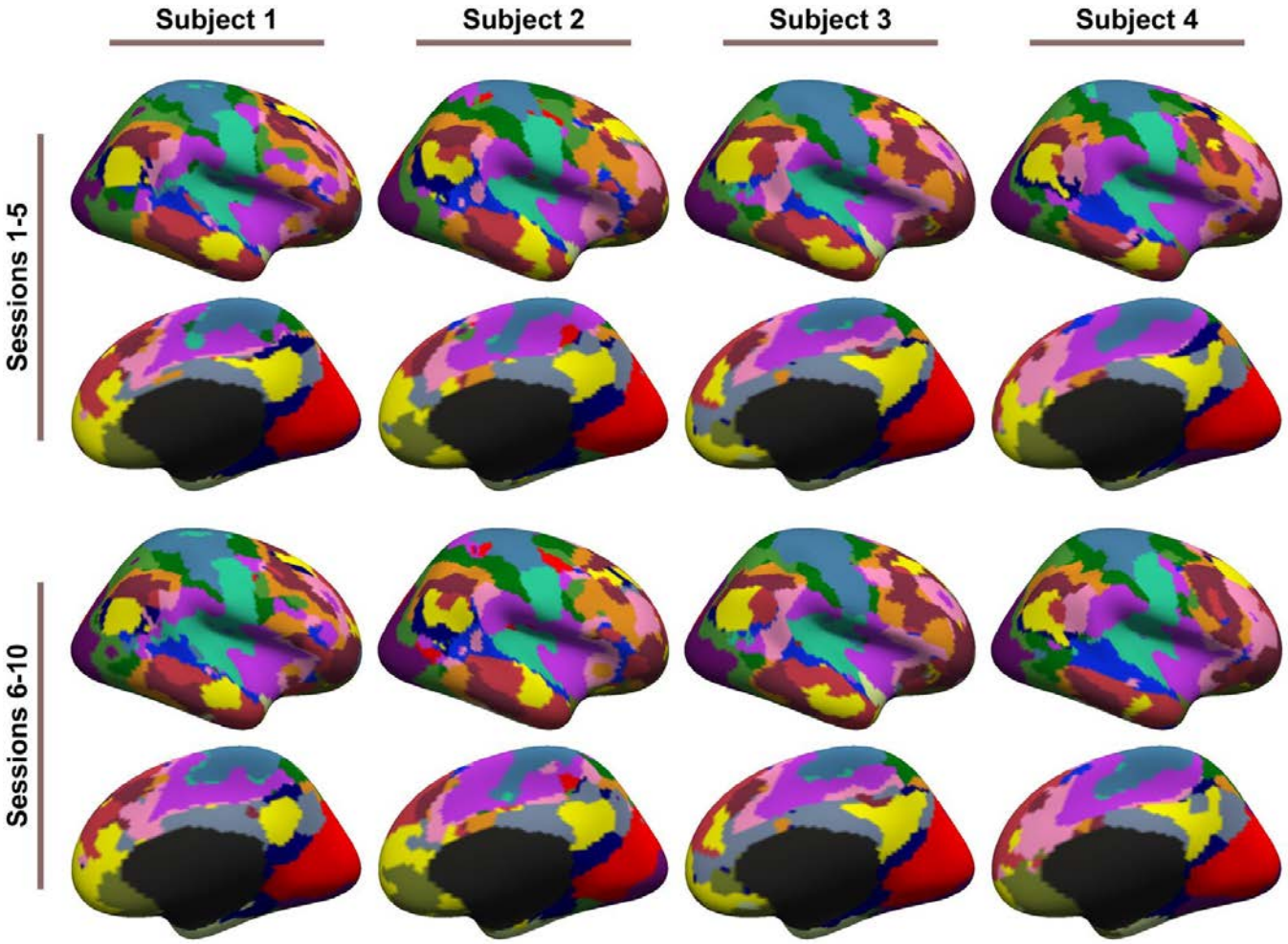

Figure S9. 17-network parcellations were estimated using sessions 1-5 and sessions 6-10 separately for each subject from the CoRR-HNU dataset. Parcellations of four representative subjects are shown here. Left hemisphere parcellations are shown in Figure S8.

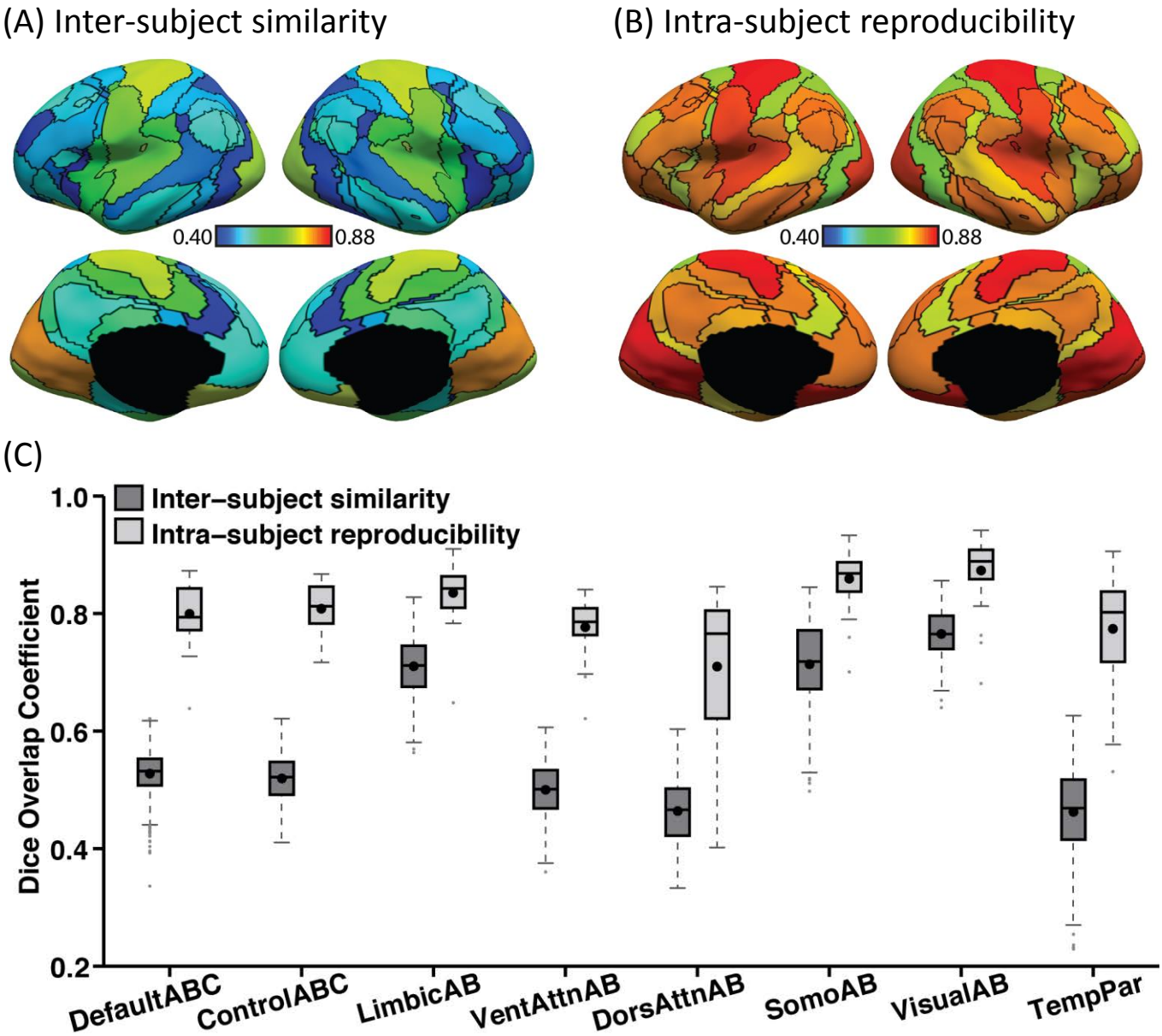

Figure S10. Individual-specific MS-HBM parcellations show high within-subject reproducibility (overlap = 81.6%) and low across-subject similarity (overlap = 59.4%) in the CoRR-HNU dataset. (A) Inter-subject spatial similarity for different networks. (B) Intra-subject reproducibility for different networks. 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. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

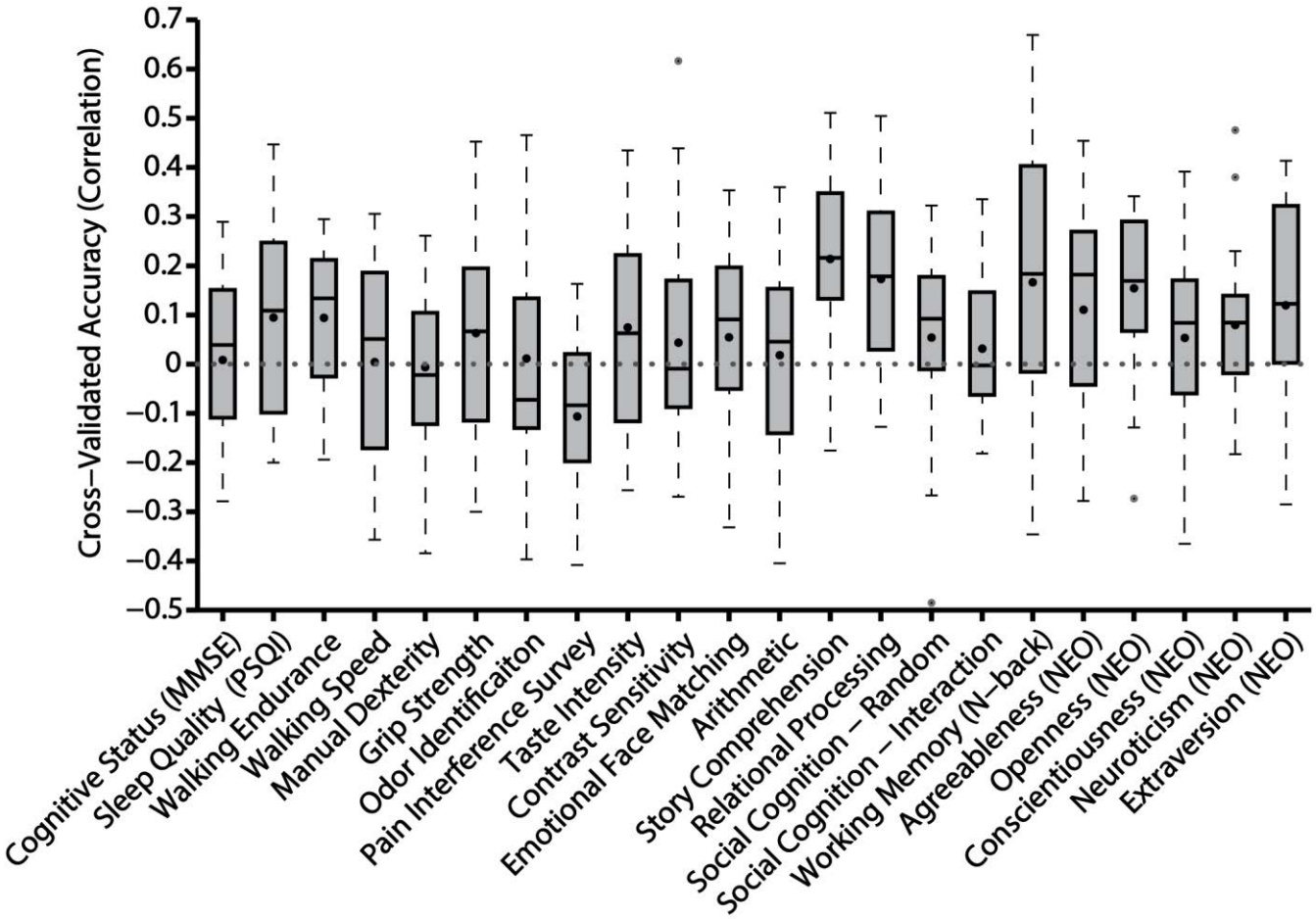

Figure S11. Prediction accuracy of 22 cognitive, emotion, personality and other non-imaging measures based on inter-subject differences in the spatial arrangement of cortical networks. In the case of the NEO-5 personality scores, average predication accuracy was r = 0.10 (p = 0.001). Other measures are found in Figures 6 and S12. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

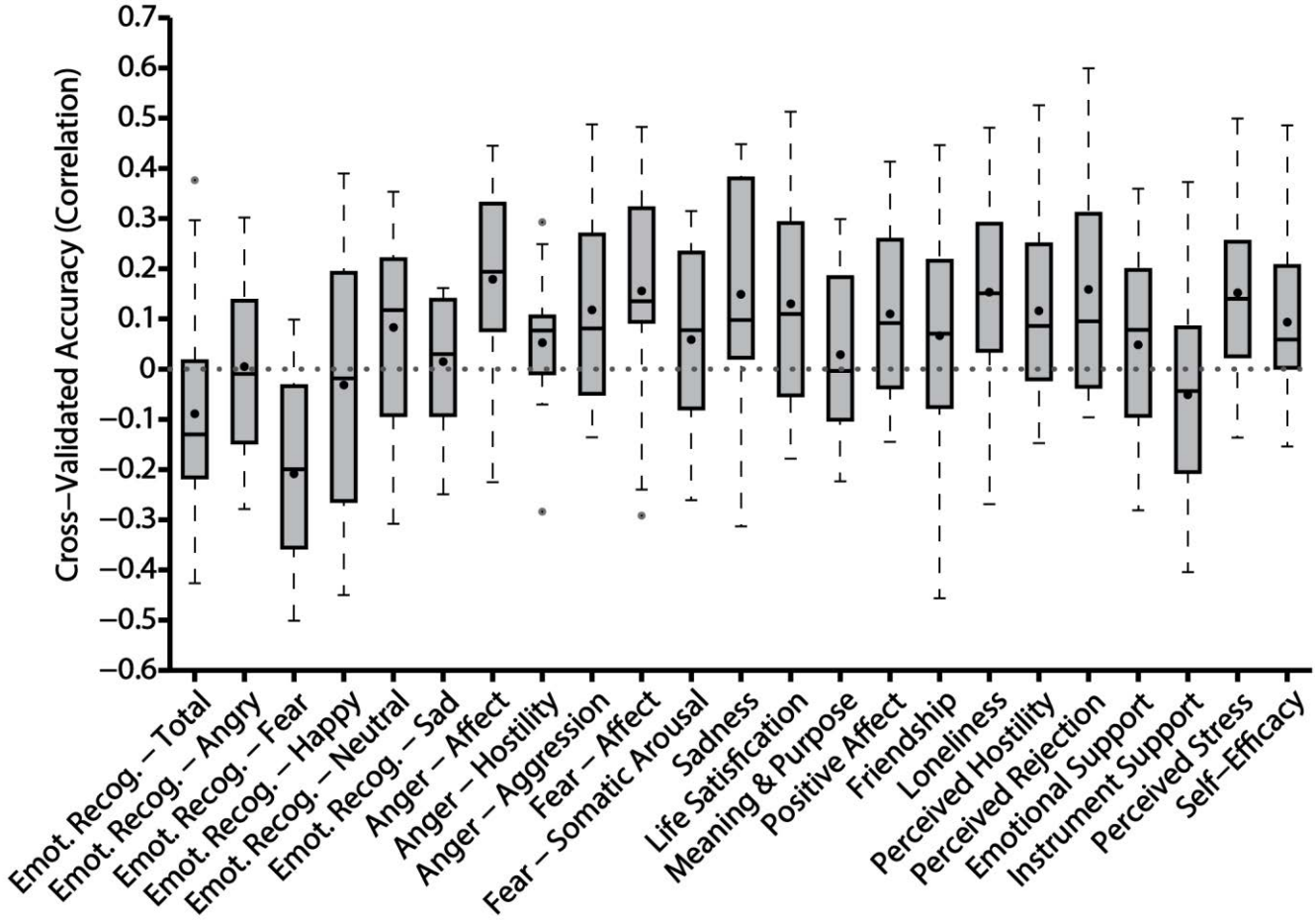

Figure S12. Prediction accuracy of 23 cognitive, emotion, personality and other non-imaging measures based on inter-subject differences in the spatial arrangement of cortical networks. In the case of the emotional measures (all items in Figure S10 except for emotional recognition), the average prediction accuracy was r = 0.10 (p = 7.6e-4). Other measures are found in Figures 6 and S11. Interestingly, prediction accuracy for the emotion recognition task was poor. Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.

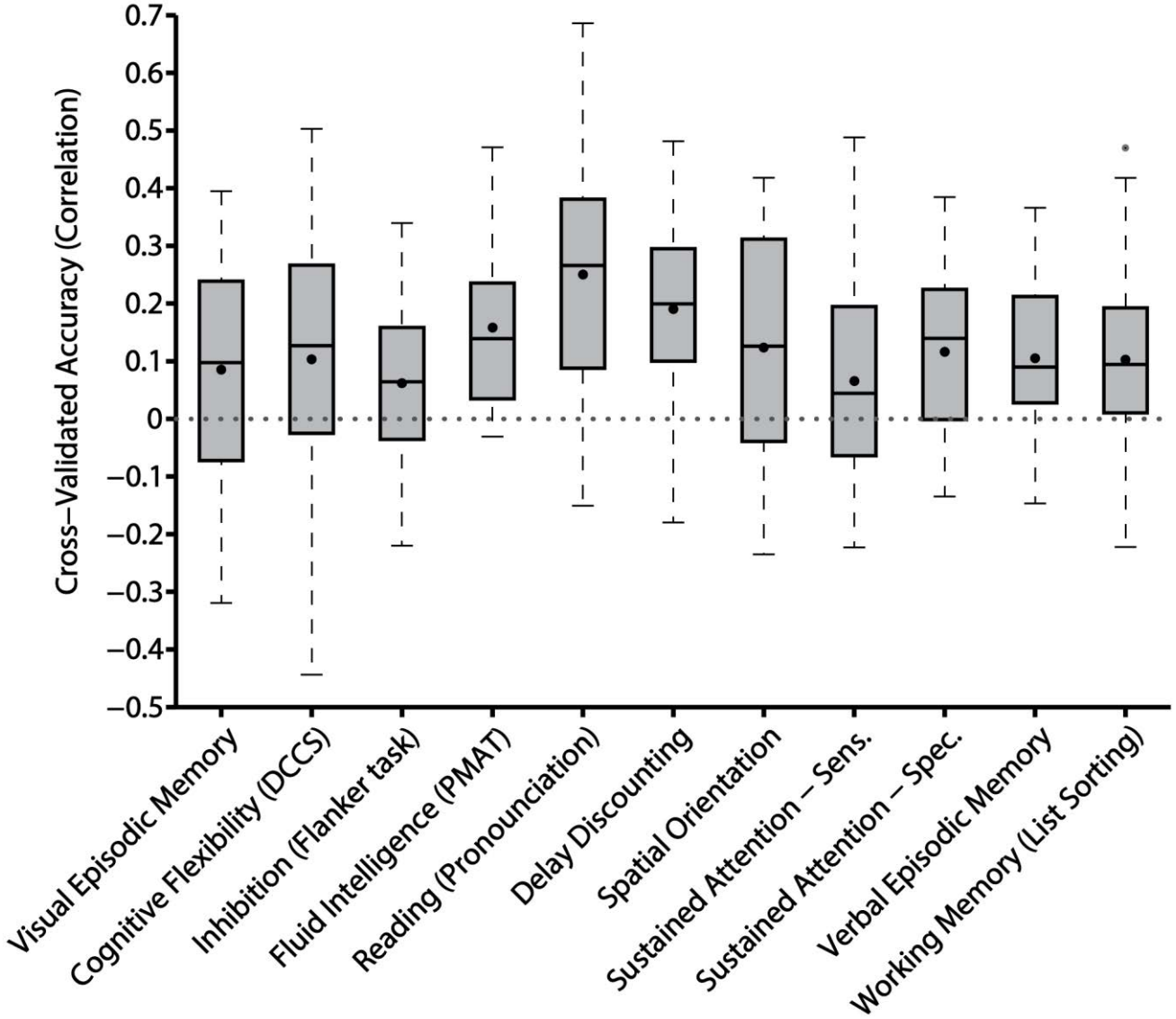

Figure S13. Prediction accuracy of 11 cognitive measures based on inter-subject differences in the spatial arrangement of task-relevant cortical networksAverage prediction accuracy based only on task-relevant networks was r = 0.12 (p = 1.7e-7). Box plots utilized default Matlab parameters, i.e., box shows median and inter-quartile range (IQR). Whiskers indicate 1.5 IQR. Dot indicates mean.