Published in final edited form as:

J Neurosci Methods. 2015 January 30; 240: 128–140. doi:10.1016/j.jneumeth.2014.11.014.

Development and validation of consensus clustering-based framework for brain segmentation using resting fMRI

Srikanth Ryali¹, Tianwen Chen¹, Aarthi Padmanabhan¹, Weidong Cai¹, and Vinod Menon^{1,2,3}

¹Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305

²Program in Neuroscience, Stanford University School of Medicine, Stanford, CA 94305

³Department of Neurology & Neurological Sciences, Stanford University School of Medicine, Stanford

Abstract

Background—Clustering methods are increasingly employed to segment brain regions into functional subdivisions using resting-state functional magnetic resonance imaging (rs-fMRI). However, these methods are highly sensitive to the (i) precise algorithms employed, (ii) their initializations, and (iii) metrics used for uncovering the optimal number of clusters from the data.

New Method—To address these issues, we develop a novel consensus clustering evidence accumulation (CC-EAC) framework, which effectively combines multiple clustering methods for segmenting brain regions using rs-fMRI data. Using extensive computer simulations, we examine the performance of widely used clustering algorithms including K-means, Hierarchical, and Spectral clustering as well as their combinations. We also examine the accuracy and validity of five objective criteria for determining the optimal number of clusters: Mutual Information, Variation of Information, Modified Silhouette, Rand Index, and Probabilistic Rand Index.

Results—A CC-EAC framework with a combination of base K-means clustering (KC) and hierarchical clustering (HC) with Probabilistic Rand Index as the criterion for choosing the optimal number of clusters, accurately uncovered the correct number of clusters from simulated datasets. In experimental rs-fMRI data, these methods reliably detected functional subdivisions of the supplementary motor area, insula, intraparietal sulcus, angular gyrus, and striatum.

Comparison with Existing Methods—Unlike conventional approaches, CC-EAC can accurately determine the optimal number of stable clusters in rs-fMRI data, and is robust to initialization and choice of free parameters.

© 2014 Elsevier B.V. All rights reserved.

Correspondence Author: Srikanth Ryali, Ph.D, Department of Psychiatry & Behavioral Sciences, 401 Quarry Rd., Stanford University School of Medicine, Stanford, CA 94305-5719, sryali@stanford.edu, *Tel no:* 1-650-736-0128, *Fax no:* 1-650-736-7200.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—A novel CC-EAC framework is proposed for segmenting brain regions, by effectively combining multiple clustering methods and identifying optimal stable functional clusters in rs-fMRI data.

Keywords

Parcellation; Segmentation; Data clustering; Consensus clustering; resting-state fMRI

Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) is increasingly used to characterize the functional organization of the human brain (Barnes et al., 2010; Barnes et al., 2012; Beckmann and Smith, 2004; Calhoun et al., 2001; Cohen et al., 2008; Deen et al., 2011; Kelly et al., 2012; Kelly et al., 2010; Kim et al., 2010; Nelson et al., 2010; Ryali et al., 2012; Shen et al., 2010; Shen et al., 2013; Wig et al., 2013a; Wig et al., 2013b; Yeo et al., 2011; Zhang et al., 2010). Segmenting brain regions into functionally homogeneous subdivisions is critical for understanding the role of individual brain areas in perception and cognition (Ryali et al., 2012; Wig et al., 2013a; Yeo et al., 2011). Human brain parcellations have typically been constructed using anatomical features, such as cellular organization, gyral folding patterns, or neurotransmitter profiles (Brodmann, 1994; Zilles and Amunts, 2010). However, such segmentations are based on post-mortem brains in the elderly, which may differ from functional subdivisions in normal healthy individuals. Furthermore, they cannot be used to characterize individual differences in functional brain organization. Segmentation schemes using rs-fMRI have been shown to be informative for: (i) determining the functional boundaries of brain regions in normal healthy adults (Cohen et al., 2008; Kim et al., 2010; Ryali et al., 2012; Shen et al., 2010), (ii) characterizing the development of these boundaries from childhood to adulthood (Barnes et al., 2012; Uddin et al., 2010b), (iii) elucidating aberrations in the functional organization in clinical populations (Zarei et al., 2012), and (iv) selecting nodes for further seedbased whole brain functional connectivity analysis (Deen et al., 2011; Shen et al., 2013; Supekar and Menon, 2012; Uddin et al., 2011).

Segmentation of brain regions using rs-fMRI can be posed as a data clustering problem. Data clustering algorithms are unsupervised machine learning methods that group objects or voxels (in the case of brain imaging data), into clusters having high intra-cluster and low inter-cluster similarity (Kelly et al., 2012; Kelly et al., 2010; Kim et al., 2010; Shen et al., 2010). Several data clustering approaches including K-means (Bellec et al., 2010; Chang et al., 2012; Deen et al., 2011; Kim et al., 2010), spectral clustering (Kelly et al., 2012; Kelly et al., 2010; Shen et al., 2010; Shen et al., 2013), and hierarchical clustering (Cauda et al., 2011; Salvador et al., 2005) have been used to segment brain regions using rs-fMRI. Unlike supervised machine learning problems, data clustering problems are highly ill-posed (Duda, 2001; Hastie et al., 2001; Jain et al., 1999) and the clusters obtained from a given dataset are strongly dependent on the clustering method used (Fred and Jain, 2005). For example, the widely used K-means algorithm, with Euclidean as a distance measure, imposes equal-sized hyper-spherical cluster shapes and performs poorly when the true underlying clusters are arbitrarily shaped. More broadly, the results of a clustering critically depend on (i)

initialization, (ii) choice of parameters such as the number of clusters, (iii) choice of the distance metric (Euclidean, Manhattan or correlation), and (iv) choice of features used. Other widely used clustering methods including hierarchical and spectral clustering are also plagued with similar issues. Critically, determining the optimal number of arbitrary shaped stable clusters in rs-fMRI data remains a major challenge.

Recent advances in data clustering and machine learning have led to the suggestion that consensus clustering can overcome many of the aforementioned limitations of current methods (Carpineto and Romano, 2012; Fred and Jain, 2005; Monti et al., 2003; Strehl and Ghosh, 2002; Topchy et al., 2005, 2003). Consensus clustering is similar to ensemble methods in supervised learning where multiple weak classifiers are combined to build a strong classifier (Freund and Schapire, 1997; Hastie et al., 2001; Kittler et al., 1998; Quinlan, 1996). A key goal here is to discover clusters that are stable across multiple partitions obtained by applying different clustering algorithms, parameters, initializations and data features (Fred and Jain, 2005; Topchy et al., 2005).

Previous neuroimaging studies have suggested that consensus clustering is a potentially useful approach for functionally segmenting brain regions (Bellec et al., 2010; Kelly et al., 2012; Kelly et al., 2010). These approaches have used different clustering algorithms and objective measures for selecting the optimal number of clusters from the data. However, these methods have not been validated adequately, even using realistic simulated datasets where the underlying clusters are known. Therefore, it is unclear whether these methods are sensitive to the choice of the underlying clustering method, its initialization, the choice of its free parameters, or the metrics used for choosing the optimal number of clusters. A commonly used approach for combining multiple partitions in the computer vision literatures is based on evidence accumulation (Fred and Jain, 2005). Using such an approach, Fred and Jain showed that clusters obtained by “weak” methods, such as K-means, can be effectively combined to discover stable clusters of arbitrary shape and size.

Here we develop and validate a novel consensus clustering evidence accumulation (CC-EAC) framework for segmenting individual brain regions. We first validate our approach on several simulated datasets and then apply them on experimental rs-fMRI datasets. We apply CC-EAC to determine functional subdivisions in five brain areas in the right hemisphere: (i) supplementary motor area (SMA) and pre-SMA, (ii) insular cortex, (iii) intraparietal sulcus (IPS), (iv) angular gyrus (AG), and (v) striatum. While SMA-preSMA, insular and striatal regions have been previously segmented using different clustering approaches (Chang et al., 2012; Deen et al., 2011; Kelly et al., 2012; Kim et al., 2010; Ryali et al., 2012), to our best knowledge functional subdivisions of the IPS and AG have not been previously investigated.

Methods

CC-EAC

We develop a CC-EAC framework (Fred and Jain, 2005) to find the stable and robust clusters in a given region of interest (ROI) using rs-fMRI time series as features. One approach would be to use the functional connectivity profiles of each voxel in an ROI to the rest of the voxels in the brain as features for segmentation (Kelly et al., 2012; Kelly et al.,

2010; Wig et al., 2013a; Yeo et al., 2011). In this segmentation approach, the similarities of the whole-brain functional connectivity profiles of each voxel in the ROI are used to segment the ROI. CC-EAC is a general framework wherein similarity between functional connectivity features can also be used to segment a given ROI. Here, we use the time series of each voxel in an ROI as features for segmentation, thus allowing voxels with similar time series profiles to be clustered together.

CC-EAC uses two clustering methods: base and consensus to obtain stable clusters. The base clustering method is used to generate several partitions or groupings $P = \{p^1, p^2, \dots, p^N\}$ for a given set of observations Y^S using different initializations. The consensus clustering method is then used to find stable clusters across all these partitions that are not sensitive to the initializations of the base clustering method. These partitions could have been obtained by (i) applying different clustering methods, (ii) applying the same method with different parameter settings, (iii) applying clustering with different distance metrics, (iv) using different initializations of the same clustering method, (v) using different features of the data, (vi) or a combination of some or all of the above. In this work, we will generate different partitions of data by applying the same base clustering method several times with different initializations for each value of k (the number of clusters). We will then use several objective criteria to select the value of k for which the partitions are most stable, and thus determine the optimal number of clusters. The various steps that are involved in CC-EAC are outlined in Figure 1. Detailed descriptions of CCEAC are as follows:

In step 1, we extract the time series in each ROI voxel for every subject. Let $Y^S = \{y_i^s\}_{i=1}^M$ be the observed voxel time series, where M is the number of voxels in the ROI, y_i^s is the time series consisting of T fMRI observations at each voxel in ROI for a given subject s and S is the total number of subjects.

In step 2, we generate 100 different partitions of data from each subject for each k ranging from 2 to 10 by applying a base clustering method with different initializations. In this study, we use two clustering approaches (K-means and spectral clustering) as base clustering methods since these have been the two most widely used approaches in neuroimaging research. In both of these methods, we use the correlation between voxel time series as similarity (or distance) measures.

In step 3, we compute a co-association matrix $C_{s,k}$, for each k and s that finds similarities between these 100 different partitions. $C_{s,k}$ is a $M \times M$ matrix with the (i,j) -th entry defined as

$$C_{s,k}(i,j) = \frac{n_{ij}}{N}$$

Where, n_{ij} is the number of times the pair of voxels i and j are clustered together in $N = 100$ partitions of the data for a given k and S . An entry in $C_{s,k}(i,j)$ close to 1 means that the specific pair of voxels are clustered together consistently across the N partitions and the opposite is true when this is close to 0.

In step 4, we apply the consensus clustering method using $C_{s,k}$ as the similarity matrix for each k and s to generate consensus clusters across the 100 base partitions for each k and s .

In step 5, we quantify the stability of partitions for each k across all the subjects, using five different cluster similarity measures: normalized mutual information (NMI), variation of information (VI), rand index (RI), probabilistic rand index (PRI), and modified silhouette. The details of these measures are given in the next section.

In step 6, we choose the optimal number of clusters (K^*) where the similarity between the partitions is optimal. We compare the performance of each method in finding the optimal number of clusters from the data.

In step 7, we compute the group level average co-association matrix by averaging the coassociation matrices, corresponding to $k = k^*$, across all subjects (S).

Finally in step 8, we apply a consensus clustering method with the group co-association matrix as the similarity matrix to find k^* optimal number of clusters from the data. We use two different consensus clustering approaches: hierarchical with average linkage and spectral clustering, and also compare their relative performance.

In summary, we use the following combinations of base and consensus clustering methods within the CC-EAC framework: (1) K-means combined with hierarchical clustering (KC-HC) as in (Bellec et al., 2010), (2) K-means combined with spectral clustering (KC-SC), and (3) spectral clustering combined with spectral clustering (SC-SC) as in (Kelly et al., 2012). For each combination, we apply five objective criteria to find optimal number of clusters from the data. We finally compare the performance of these methods in finding stable clusters of several ROIs using both simulated and real rs-fMRI datasets. We run K-means and spectral clustering methods 100 times with different random initializations for each k (number of clusters) to generate base partitions. In the K-means clustering method, we use the correlation between two voxel time series as a “similarity” or “distance” metric. In this setting, correlation is equivalent to the Euclidean distance metric as the voxel time series have been normalized by subtracting their temporal means and dividing by their standard deviations. Further details of the rs-fMRI data preprocessing procedures are outlined below. We also note that the spectral clustering method used in this study is based on the work of Ng et al., (2002) and that the sample correlation matrix of the voxel time series is used as the similarity matrix in the spectral clustering method.

Objective criteria for determining the optimal number of clusters

We compare the following criteria that are used in determining the optimal number of clusters from the data (step 5 in Figure 1). Let $L^{i,k}$ and $L^{j,k}$ be the voxel labels for two subjects i and j for a given number of clusters k . The following criteria define the similarity between these two different partitions.

(a) *Normalized mutual information (NMI)* is an information theoretic criterion that can be used to measure the similarity between two partitions $L^{i,k}$ and $L^{j,k}$ (Fred and Jain, 2005; Strehl and Ghosh, 2002)

$$NMI(L^{i,k}, L^{j,k}) = \frac{2I(L^{i,k}, L^{j,k})}{H(L^{i,k}) + H(L^{j,k})} \quad (1)$$

Where $I(L^{i,k}, L^{j,k})$ is the mutual information between the two partitions, and $H(L^{i,k})$ and $H(L^{j,k})$ are their entropies. The mutual information between these two partitions is defined as

$$I(L^{i,k}, L^{j,k}) = \sum_{p=1}^k \sum_{q=1}^k \frac{m_{p,q}^{i,j}}{M} \log_2 \frac{\frac{m_{p,q}^{i,j}}{M}}{\frac{m_p^{i,k}}{M} \frac{m_q^{j,k}}{M}} \quad (2)$$

where, $m_{p,q}^{i,j}$ is the number of shared voxels between the clusters $L_p^{i,k}$ and $L_q^{j,k}$, m_p^i is the number of voxels in the cluster $L_p^{i,k}$ and m_q^j is the number of voxels in the cluster $L_q^{j,k}$. The entropy $H(L^{i,k})$ is defined as

$$H(L^{i,k}) = \sum_{p=1}^k \frac{m_p^i}{M} \log_2 \frac{m_p^i}{M} \quad (3)$$

NMI is the normalized measure of mutual information such that $0 \leq NMI \leq 1$, NMI is 0 if there is no similarity between the partitions and is 1 if they are identical. We compute the average NMI by averaging the NMI between all pairs of subjects for a given k . We then select the optimal number of clusters k^* for which the average NMI is maximized.

(b) *Variation of information (VI)* is an information theoretic metric introduced by (Meila, 2007) to quantify the dissimilarity between two partitions. VI for two partitions $L^{i,k}$ and $L^{j,k}$ is defined as

$$VI(L^{i,k}, L^{j,k}) = H(L^{i,k}) + H(L^{j,k}) - 2I(L^{i,k}, L^{j,k}) \quad (4)$$

As in the case of NMI , we compute the average VI between all pairs of subjects for a given k and choose the optimal number of clusters k^* for which the average VI is minimized.

(c) *Rand Index (RI)* is a measure for comparing similarity between two sets or between data partitions as in our case. RI for two partitions $L^{i,k}$ and $L^{j,k}$ can be defined as

$$RI(L^{i,k}, L^{j,k}) = \frac{n_{11} + n_{00}}{n_{00} + n_{01} + n_{10} + n_{11}} = \frac{n_{11} + n_{00}}{\binom{N}{2}} \quad (5)$$

Where

n_{11} : the number of pairs of voxels that are in the same cluster in both $L^{i,k}$ and $L^{j,k}$

n_{00} : the number of pairs of voxels that are in different clusters in both $L^{i,k}$ and $L^{j,k}$

n_{10} : the number of pairs of voxels that are in the same cluster $L^{i,k}$ but in different clusters in $L^{j,k}$

n_{01} : the number of pairs of voxels that are in different clusters in $L^{i,k}$ but in the same cluster in $L^{j,k}$

We compute the average RI between all pairs of the subjects for every k and choose the optimum k^* for which the average RI is maximized.

(d) *Probabilistic Rand Index (PRI)* proposed in (Carpineto and Romano, 2012) is a modified measure of RI for quantifying the similarity between two data clustering solutions. This measure gives different weights to n_{11} , n_{00} , n_{10} , and n_{01} depending on the number of clusters present in that particular partition. The rationale behind the different weightings is that if the number of clusters is small then the probability of grouping two voxels in a cluster is greater than when the number of clusters is higher. In other words, the information content of two voxels being clustered together is greater when the underlying number of clusters is larger. Therefore, PRI (Carpineto and Romano, 2012) gives more weight to n_{11} , n_{10} , and n_{01} for partitions having a higher number of clusters compared to the partitions having lower cluster numbers, and less weight to n_{00} . These weights are defined using the entropy criterion (Carpineto and Romano, 2012) as follows

$$w_h = -\log_2 p_h \quad (6)$$

Where, $h \in \{00,01,10,11\}$ is a binary indicator variable. We compute probabilities ($p_h = n_{11}, n_{00}, n_{10}$ and n_{01}) as defined in (Carpineto and Romano, 2012). Accordingly, PRI between two partitions $L^{i,k}$ and $L^{j,k}$ for subjects i and j for k number of clusters is defined as:

$$PRI(L^{i,k}, L^{j,k}) = \frac{w_{11}n_{11} + w_{00}n_{00}}{w_{00}n_{00} + w_{01}n_{01} + w_{10}n_{10} + w_{11}n_{11}} \quad (7)$$

We compute average PRI for selecting the optimal number of clusters from the data.

(e) *Modified Silhouette* (Bellec et al., 2010) is a cluster measure that is used to select the optimal number of clusters from the data and to quantify the stability of the clusters at both subject and group levels. Modified silhouette on a co-association matrix C^k is defined as

$$\text{Modified Silhouette}(k) = \frac{1}{M} \sum_{m=1}^M (w(C_k)_m - \max_{C \neq C_k} w(C)_m) \quad (8)$$

Where, $w(C_k)$ is the stability map generated for each cluster C as

$$w(C)_m = \sum_{n \in C, n \neq m} C_{m,n}^k \quad (9)$$

We select the optimal number of clusters k^* for which the modified silhouette is the largest.

Resting-state fMRI data

Resting-state fMRI data were acquired from 21 adult participants. The study protocol was approved by the Stanford University Institutional Review Board. The participants ranged in age from 19 to 22 y (mean age 20.40 y) with an IQ range of 97 to 137 (mean IQ: 112). The participants were recruited locally from Stanford University and neighboring community colleges.

Participants were instructed to keep their eyes closed and try not to move for the duration of the 8-min scan. Functional Images were acquired on a 3T GE Signa scanner (General Electric) using a custom-built head coil. Head movement was minimized during scanning by a comfortable custom-built restraint. A total of 29 axial slices (4.0 mm thickness, 0.5 mm skip) parallel to the AC-PC line and covering the whole brain were imaged with a temporal resolution of 2 s using a T2* weighted gradient echo spiral in-out pulse sequence (Glover and Law, 2001) with the following parameters: TR = 2,000 msec, TE = 30 msec, flip angle = 80 degrees. Slices are sequentially acquired from inferior to superior of the brain. The field of view was 20 cm, and the matrix size was 64×64, providing an in-plane spatial resolution of 3.125 mm. To reduce blurring and signal loss arising from field in homogeneities, an automated high-order shimming method based on spiral acquisitions was used before acquiring functional MRI scans.

Preprocessing of rs-fMRI data

Data were preprocessed using SPM8 (Statistical Parametric Mapping software, <http://www.fil.ion.ucl.ac.uk/spm>). The first eight image acquisitions of the rs-fMRI time series were discarded to allow for stabilization of the MR signal. Each of the remaining 232 volumes underwent the following preprocessing steps: realignment, slice-timing correction, normalization to the MNI template, and spatial smoothing using a 6-mm full-width half maximum Gaussian kernel. Excessive motion, defined as greater than 3.5 mm of translation or 3.5 degrees of rotation in any plane, was not present in any of the scans. The time series at each voxel were filtered using a band pass filter ($0.0083 \text{ Hz} < f < 0.15 \text{ Hz}$). For each ROI to be segmented, we regressed out the temporal mean and linear trend of each voxel time series within the ROI. Finally we divided each voxel time series by its standard deviation.

Synthetic fMRI datasets

Because the actual functional boundaries are unknown in experimental fMRI data, the performance of clustering methods and the suitability of cluster similarity measures were first validated on realistic simulated datasets in which the precise boundaries between the clusters are known. For this purpose, we constructed synthetic fMRI datasets using real fMRI signals extracted from 6 regions (3 in each hemisphere) in primary auditory cortex using the maximum probabilistic cytoarchitectonic maps available in the SPM Anatomy Toolbox (Eickhoff et al., 2005) (Figure 2). Voxel-wise time series were extracted from the experimental rs-fMRI dataset with 21 adult subjects that were preprocessed only up to the smoothing step to ensure that voxels were highly correlated both within and between regions. To create distinct brain regions or clusters, we destroyed between-region correlations by randomizing the phase of time courses in different regions (Prichard and Theiler, 1994). This manipulation has the advantage of maintaining the spectral information

of the original time series (Prichard and Theiler, 1994). Specifically, to destroy the correlation between regions, we added a randomly generated common phase to all time courses within a region, and similarly another random common phase to time series in other regions (Chen et al., 2013). As a result, correlations between voxels within each region still remained high and were spatially varied. In contrast, correlations between voxels in different regions were expected to be zero, thus creating distinct clusters.

Brain regions used for segmentation of experimental fMRI data

To investigate the performance of different consensus clustering methods and different similarity measures for obtaining the optimal number of clusters, we examined their ability to functionally segment different brain regions into spatially contiguous subdivisions.

(i) SMA, pSMA—We chose the right medial frontal cortex because it consists of two spatially contiguous cytoarchitectonically distinct subdivisions – the Supplementary Motor Area (SMA) (Brodmann area 4), and the preSMA (pSMA: Brodmann area 4, 6) (Ryali et al., 2012; Zilles et al., 1996). Importantly, the right SMA and pSMA are well demarcated anatomically by a vertical line, passing through the anterior commissure, perpendicular to the anterior-posterior commissure axis (Zilles et al., 1996). The right SMA and pSMA, shown in Figure 3A, were defined using a procedure similar to that used in previous studies (Aron et al., 2007; Johansen-Berg et al., 2004). The automated anatomical labeling (AAL) template (Tzourio-Mazoyer et al., 2002) was used to demarcate the SMA ($y < 0$) and pSMA ($y > 0$).

(ii) Insular cortex—The insula is a heterogeneous multimodal brain area situated deep within the lateral sulcus between the frontal and temporal lobes, encompassing Brodmann areas 13, 14, and 16. We used the right insular cortex as defined by the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) (Figure 3B).

(iii) Intraparietal Sulcus (IPS)—The right IPS is key area involved in visuospatial attention (Corbetta and Shulman, 2002). Cytoarchitectonic mapping studies in post-mortem brains have identified three subdivisions in the IPS (Choi et al., 2006; Scheperjans et al., 2008). In the current study, we merged the maximum probabilistic maps of three right IPS subdivisions using the Anatomy Toolbox in SPM (<http://www.fil.ion.ucl.ac.uk/spm/ext/#Anatomy>) to create a single IPS ROI (Figure 3C).

(iv) Angular Gyrus (AG)—The angular gyrus is a ventral posterior parietal cortex region involved in semantic memory, speech, and reasoning (Binder and Desai, 2011; Seghier, 2013). The subdivisions within this right ROI (PGa and PGp) were identified using cytoarchitectonic mapping (Caspers et al., 2006) (Figure 3D).

(v) Striatum—The striatum is a subcortical region that includes the caudate, putamen, and nucleus accumbens, which have distinct as well as overlapping functional properties and connectivity with the rest of the brain (Alexander et al., 1986; Di Martino et al., 2008; Polderman et al., 2006). We used the Harvard-Oxford subcortical atlases in FSL and merged

the maximum probabilistic maps of the right caudate, putamen, and nucleus accumbens to form a striatal ROI (Figure 3E).

Results

Synthetic fMRI data: Identification of number of clusters

Our first goal was to determine whether CC-EAC could accurately identify the number of clusters in simulated rs-fMRI data, where the cluster boundaries are known. Within the CC-EAC framework, we investigated the performance of three clustering approaches: KC-HC, KC-SC, and SC-SC for segmenting brain regions into functionally distinct subdivisions. We used five different metrics for determining the optimal number of clusters at the group level: Modified silhouette, PRI, RI, VI, and NMI.

Figure 4 shows the performance of the cluster comparison metrics for each combination of clustering methods on ROIs containing 2, 3, 4, 5, and 6 clusters. The first row of Figure 4 demonstrates that for ROIs with two clusters, all the five metrics for the three cluster combinations correctly identify the optimal number of clusters as 2. For ROIs with three clusters, except VI, all the other metrics correctly identified the optimal number of clusters to be 3, as shown in the second row of Figure 4. VI increased monotonically from 2 to 10 suggesting that 2 is the optimal number of clusters in this ROI. For ROIs with 4 clusters, similar performance was observed except for the cluster combination SC-SC as shown in the third row of Figure 4. None of the cluster comparison metrics was able to identify the correct number of clusters using SC-SC. In the case of cluster combinations KC-HC and KC-SC, with the exception of VI and RI, all other metrics were able to estimate the correct number of distinct clusters as 4 (third row of Figure 4). For ROIs with 5 clusters, the modified silhouette identified the correct number of clusters only for the cluster combination KC-HC while PRI was able to find the correct number of underlying clusters with both KC-HC and KC-SC (fourth row of Figure 4). In the case of ROIs with 6 clusters, all the metrics other than RI gave 6 clusters for KC-HC and KC-SC as shown in the fifth row of Figure 4. Although RI and NMI identified the correct number of clusters in some of the cases, they increased monotonically with increase in the number of clusters. These simulations also suggest that none of the cluster comparison metrics were able to find the optimal number of underlying clusters in ROIs with more than three subdivisions using spectral clustering (SC-SC).

In summary, the modified silhouette and PRI measures were able to correctly identify the number of clusters in all five ROIs with KC-HC as the cluster combination. The other three measures – RI, VI and NMI – were inconsistent in identifying the optimal number of clusters.

Synthetic fMRI data: Functional subdivisions

Our next goal was to examine how well the segmented functional subdivisions matched the simulated data. We segmented the simulated datasets using the KC-HC as the cluster combination and either the modified silhouette or PRI as the criterion for choosing the optimal number of clusters. We made this choice based on the relative performance of these

methods (Figure 4). Figure 5 shows the clusters obtained using KC-HC at the group level. Note that the clusters were obtained by applying hierarchical clustering with the co-association obtained for the optimal k (k^*) for each ROI as the similarity metric. The obtained stable clusters were similar to the actual clusters within each ROI (Figure 2). Note that the color coding used for different clusters in Figure 2 and Figure 5 is arbitrary. Similar stable clusters are also obtained using KC-SC (data not shown).

Experimental rs-fMRI data: Identification of number of clusters

Next, we applied CC-EAC to determine the optimal number of clusters in experimental rs-fMRI data. We applied CC-EAC to segment five different brain areas using rs-fMRI data acquired from 21 adult participants.

(i) SMA-pSMA—The first row of Figure 6 shows the performance of each metric in identifying the optimal number of clusters for the SMA-pSMA ROI. The optimal number of clusters according to the modified silhouette, PRI and VI is 2 for each cluster combination method (Figure 6), consistent with the cytoarchitectonic segmentation of this ROI (Zilles et al., 1996). Note that RI and NMI increased monotonically with the number of clusters combinations KC-HC and KC-SC.

(ii) Insular cortex—The second row of Figure 6 shows the five cluster comparison criteria as a function of the number of clusters for each combination: KC-HC, KC-SC and SC-SC. The modified silhouette identified two function divisions for all the three cluster combinations; PRI identified three clusters in the case of KC-HC and KC-SC; VI identified two while RI and NMI increased as a function of the number of clusters suggesting ten as the optimal number of clusters for each combination.

(iii) IPS—The third row of Figure 6 shows that the modified silhouette and PRI identified three subdivisions for the combinations KC-HC and KC-SC in this ROI, consistent with previous cytoarchitectonic mapping of this region (Caspers et al., 2008). VI, RI, and NMI increased as a function of the number of clusters for each combination and suggest two divisions by VI and ten by RI and NMI.

(iv) AG—In this ROI, the modified silhouette and VI suggest 2 whereas RI and NMI suggest 10 as the optimal number of clusters for each cluster combination as shown in the fourth row of Figure 6. PRI, on the other hand, suggests 3 as the optimal number of subdivisions for the cluster combinations KC-HC and KC-SC.

(v) Striatum—The fifth row of Figure 6 shows the profiles of the five cluster similarity metrics as the function of the number of clusters for three combination methods. Both the modified silhouette and PRI suggested three subdivisions for only the combination KC-HC. As before, VI, RI, and NMI increased as the function of the number of clusters for all combinations.

In summary, the objective measures VI, RI and NMI provided no useful information for choosing the optimal number of clusters (Figures 4 and 6). PRI and modified silhouette with KC-HC were the most accurate in identifying the correct number of functional clusters.

Functional segmentation in experimental rs-fMRI data

Our next goal was to examine the functional boundaries of the segmented functional subdivisions and to compare them with known anatomical boundaries. Based on findings from the previous sections, we focused on KC-HC as the clustering approach. The modified silhouette and PRI identified the same number of clusters in SMA-pSMA, IPS, and striatum, but different numbers in insula and AG. For insula and AG, we chose PRI as the criterion for selecting the number of clusters because it identified more clusters, thus providing a richer basis for examining regional heterogeneity.

(i) SMA-pSMA—CC-EAC framework together with the modified silhouette and PRI identified two subdivisions in this region. These two subdivisions overlapped with the SMA-pSMA, and their localization is consistent with the boundary at the anterior commissure (Picard and Strick, 2001; Zilles and Amunts, 2010) as shown in Figure 7A. The diagonal entries of Table 1A show that overlap of 92% and 86% with the corresponding SMA and pSMA cytoarchitectonic subdivisions. The nondiagonal entries respectively indicate the overlap of SMA with pSMA (8%) and pSMA with SMA (14%).

(ii) Insula—The right insular cortex was segmented into posterior, dorsal and anterior sub-regions (Figure 7B). The boundaries of these sub-regions are similar to those reported in previous studies using whole-brain rs-fMRI connectivity analysis (Chang et al., 2012; Deen et al., 2011).

(iii) IPS—Figure 7C depicts the functional boundaries of the IPS and Table 1B shows the percentage overlap of functional clusters with respect to its cytoarchitectonic segments. Using PRI we found three optimal functional clusters. The functional boundaries of the segmented clusters are similar, however with some differences, to the known boundaries of the cytoarchitectonic IPS subdivisions (Choi et al., 2006; Scheperjans et al., 2008). The overlap of functional clusters IPS-1, IPS-2 and IPS-3 with their corresponding cytoarchitectonic maps are 76%, 87%, and 65% respectively (Table 1B). The non-diagonal entries of the first row of Table-1B show the percentage overlap of functional subdivision of IPS-1 with the cytoarchitectonic subdivisions IPS-2 (20%) and IPS-3 (4%). The non-diagonal entries of the second row of Table-1B show the overlap of functional subdivision of IPS-2 with the cytoarchitectonic subdivisions of IPS-1 (13%) and IPS-3 (0%). The non-diagonal entries of the third row of Table-1B show the overlap of functional subdivision of IPS-3 with the cytoarchitectonic subdivisions of IPS-1 (8%) and IPS-2 (26%).

(iv) AG—Three stable clusters were identified in the AG. They encompass its posterior, dorsal and anterior aspects, as shown in Figure 7D and overlap with the cytoarchitectonic subdivisions PGp and PGa to different extents. The overlap of posterior and anterior functional segments of AG with PGp and PGa are 43% and 63% respectively. The additional dorsal functional segment consists of 43% of the PGp and 37% of the PGa (Table 1C). The posterior functional subdivision has an overlap of 14% with the anterior cytoarchitectonic subdivision PGa (Table 1C).

(v) Striatum—Figure 7E shows the segmented clusters of the right striatum using KC-HC as the cluster combination method. This method identified three clusters corresponding to the known functional and structural subdivisions of the striatum, including caudate, putamen, and nucleus accumbens (Alexander et al., 1986). Table 1D indicates that the agreement between functional and anatomical clusters is high with overlap of 73%, 97%, and 89% with the anatomically defined putamen, caudate, and nucleus accumbens, respectively. The non-diagonal entries of the first row of Table-1D show the overlap of functional subdivision of the putamen with the caudate (0%) and the nucleus accumbens (27%). The non-diagonal entries of the second row of Table-1D show the overlap of the caudate with the putamen (0%) and the nucleus accumbens (3%). The non-diagonal entries of the third row of Table-1D show the overlap of the nucleus accumbens with the putamen (0%) and caudate (11%).

Discussion

We developed a novel consensus clustering framework for segmenting brain regions using rs-fMRI voxel time series as features. We addressed three major challenges in brain segmentation: (i) choice of clustering methods, (ii) robustness of the clusters to initialization, and (iii) determination of the optimal number of clusters. Using extensive simulations, we compared the performance of three combinations of base and consensus clustering methods within CC-EAC framework: K-means with Hierarchical (KC-HC), K-means with Spectral (KC-SC), and Spectral with Spectral (SC-SC) clustering methods. These clustering pairs were used to compare the performance of five different objective criteria for determining the optimal number of clusters. Two of these criteria were based on information-theoretic measures: normalized mutual information (NMI) and variation of information (VI); another two were based on set comparison measures: rand index (RI) and probabilistic rand index (PRI), and the fifth was based on a cluster purity measure: modified silhouette. Below we characterize the performance of different clustering methods first using simulated datasets where the cluster boundaries are known and then on experimental fMRI datasets where the cluster boundaries are only partially known.

Validation on simulated datasets - Choice of cluster combinations

To test various clustering methods, we first created simulated rs-fMRI datasets with different numbers of clusters with arbitrary size and shape (Figure 2). These datasets, with known cluster boundaries, allowed us to examine the performance of CC-EAC. We used three different cluster combination methods KC-HC, KC-SC, and SC-SC together with PRI, modified silhouette, RI, VI, and NMI as metrics for obtaining the optimal number of clusters.

Across multiple simulated datasets, among the three cluster combination methods, KC-HC and KC-SC resulted in the most accurate estimation of the correct number of clusters (Figure 4). Although spectral clustering is generally thought to be a superior clustering method, its performance was not consistent across all the ROIs and metrics considered within the CC-EAC framework. Our results suggest that using a sophisticated method such as spectral clustering for generating base partitions does not necessarily result in superior

performance. Crucially, boosting and bagging methodologies using weaker algorithms can be used to build powerful classification methods (Fred and Jain, 2005; Freund and Schapire, 1997; Hastie et al., 2001; Kittler et al., 1998; Quinlan, 1996). Consensus clustering approaches inspired by boosting and bagging rely on the diversity of base clusters for discovering stable clusters. The superior performance of KC-HC and KC-SC over SC-SC can be attributed to the more diverse base partitions obtained by K-means as compared to spectral clustering methods (Figure 4). Another problem with using spectral clustering as the consensus method in CC-EAC is that it uses a K-means algorithm on the eigenvectors, obtained from the eigenvalue decomposition of the Laplacian of the correlation matrix of voxel time series (Meila, 2007; Ng et al., 2002). Therefore, the final clusters obtained using spectral clustering are highly sensitive to the initialization of K-means clustering. In contrast, our simulations demonstrate that KC-HC reliably and accurately recovers the underlying clusters (Figures 4 and 5).

Validation on simulated datasets - Choice of objective criteria for determining number of clusters

We also investigated five objective criteria for determining the optimal number of clusters. Specifically, within the CC-EAC framework, we examined which of the five measures - PRI, NMI, VI, RI, or modified silhouette would be ideal for determining the number of clusters in the simulated data.

NMI and VI are information theoretic measures that quantify the similarity or dissimilarity between two partitions based on the probability estimates of the voxels being in one cluster or the other. The relationship between these two quantities is defined in Equation (4). These measures are based on partitions rather than voxel pairs that are clustered together. NMI measures similarity between these two partitions whereas VI measures the dissimilarity between them. We then chose the optimal k for which the average NMI between the subjects is maximized or the average VI is minimized. RI and PRI are set-theoretic criteria that, like NMI, measure similarity between two partitions; however, the measures are based on individual pairs of voxels that are clustered together. RI estimates are based on the ratio of the number of pairs of voxels that are clustered together to the total number of pairs of voxels in two given partitions as shown in Equation (5). RI then selects the optimal k for which this average ratio among all pairs of partitions for a given k is the largest. RI gives equal weights to both agreements (n_{11}) and disagreements (n_{00}) (Carpineto and Romano, 2012). In this case, when k is large, n_{00} increases as the probability of voxels being placed in different clusters increases; thereby RI favors a large k . This phenomenon has been previously demonstrated with a toy example (see Figure 2 and Table 3 in (Carpineto and Romano, 2012)). In the present study, we observed a similar phenomenon using our CC-EAC approach (Figure 4). PRI (Carpineto and Romano, 2012) addresses this problem by giving different weights to the pairs of voxels being grouped together (Equation 7) across two partitions. The modified silhouette (Bellec et al., 2010) is based on the co-association matrix that quantifies intra- and inter-cluster similarity (Equation 8).

The performance of different objective criteria for selecting the optimal number of clusters from simulated datasets suggests that PRI and modified silhouette perform the best in

virtually every case (Figure 4). Although RI and NMI showed maxima at the current number of clusters on simulated datasets, they generally increased monotonically with increase in the number of clusters (Figure 4). This phenomenon is expected in the case of RI because it favors a large number of clusters and gives the same weightage to both agreements and disagreements of pair of voxels being clustered together, as explained above. VI, on the other hand, behaved in an opposite fashion and reached a minimum at 2 clusters suggesting 2 as the optimal choice in the majority of cases. A similar phenomenon was also observed by Kelly and colleagues (Kelly et al., 2012), requiring them to ignore the case of 2 clusters as a trivial choice. In summary, our simulations suggest that both PRI and modified silhouette provide robust objective criteria for choosing the optimal number of clusters.

Validation on experimental rs-fMRI

Functional subdivisions of SMA-pSMA—We applied CC-EAC on experimental rs-fMRI data by first segmenting the SMA and pSMA, which have been the focus of several studies (Behrens et al., 2006; Johansen-Berg et al., 2004; Kim et al., 2010; Nanetti et al., 2009; Ryali et al., 2012). In the case of the combined SMA and pSMA region, all the three cluster combinations: KC-HC, KC-SC and SC-SC identified two functional divisions (first row of Figure 6) using the PRI, modified silhouette, and VI as the objective criteria for selecting the optimal number of clusters. The functional boundary discovered by applying KC-HC on a group of 21 subjects (Figure 7A) is consistent with the known cytoarchitectonic boundary close to $y=0$ (Figure 3A) (Zilles et al., 1996). Crucially, there is a large overlap of functional and structural segments (Table 1A)..

Functional subdivisions of the insular cortex—Next, we applied our CC-EAC approach, together with the five criteria, to discover the optimal number of clusters in the insular cortex. Among the three cluster combination methods and five objective criteria, KC-HC together with PRI resulted in three subdivisions (Figures 6 and 7B). Consistent with the findings of SMA-pSMA, the other cluster combination methods (KC-SC and SC-SC) together with the other objective criteria resulted in either two clusters (Modified Silhouette and VI) or ten clusters (NMI and RI) for the insular cortex (Figure 6). Notably, KC-HC together with PRI resulted in three sub-divisions of insular cortex: posterior, anterior dorsal, and anterior ventral. The three sub-divisions identified by KC-HC and PRI help to clarify findings from previous studies. For example, Deen and colleagues (Deen et al., 2011) arbitrarily chose three clusters, while Chang and colleagues (Chang et al., 2012) used a metric based on the ratio of inter- and intra-cluster distances to select tripartite division of insula without examining the stability of their clustering approach. Kelly and colleagues used VI to select the optimal number of clusters that resulted in local minima ranging from 2 to 15 (Kelly et al., 2012). However, these studies lack rigorous validation of their methods, especially with respect to initialization and parameter settings. In the present study, we first validated our CC-EAC approach on simulated and experimental datasets where the functional boundaries are relatively known and then applied these methods to segment the insular cortex.

The three clusters identified in our study encompass the anterior-dorsal, anterior-ventral, and posterior insula regions crucial for cognition, emotion and introspection (Menon and Uddin,

2010). Our clustering results converge on and inform previous findings (Chang et al., 2012; Deen et al., 2011). Using similar sub-divisions as those obtained in the present study, meta-analysis with reverse inference revealed functional specificity for each of these three clusters (Chang et al., 2012). Specifically, the dorsal-anterior cluster has been linked to higher cognition and executive control, the anterior-ventral insula cluster to affective processing, and the posterior insula to sensorimotor processing and pain. These findings further converge on whole-brain functional connectivity analyses showing different whole-brain connectivity profiles of posterior, ventral and dorsal regions of the insula (Deen et al., 2011). A final noteworthy aspect of our findings is that the segmentations are consistent with histological and electrophysiological data, which have suggested three distinct insular subdivisions (Mesulam and Mufson, 1982). Taken together, these findings point to the validity of the CC-EAC approach for identifying stable and distinct clusters in brain regions such as the insula where the number of clusters is not known a priori.

Functional subdivisions of the IPS—The right IPS is a key node in the dorsal attention system (Corbetta and Shulman, 2002). It has been widely implicated in a variety of cognitive-motor processes, including visually guided motor control (Frey et al., 2005), visuospatial attention (Egner et al., 2008), spatial and object working memory (Belger et al., 1998), and numerical processing (Arsalidou and Taylor, 2011; Rosenberg-Lee et al., 2011). Previous investigations of post-mortem human brains have shown that the human IPS consists of at least 3 cytoarchitectonically distinct areas, including hIP1, hIP2, and hIP3 (Choi et al., 2006; Scheperjans et al., 2008). To our knowledge, no study has segmented the IPS using *in vivo* human functional brain imaging. KC-HC with PRI and the modified silhouette index revealed three optimal clusters. The functional divisions obtained using this approach showed both overlap as well as some distinctions with cytoarchitectonic subdivisions of the IPS (Figure 7C and Table 1B).

Our finding suggests that the three IPS clusters belong to distinct brain networks and likely serve different cognitive functions. Previous studies have shown that the anterior IPS, mostly hIP1 and hIP2, is involved in visually guided action control, whereas the posterior IPS, mostly hIP3, is involved in processing visual features (Shikata et al., 2003). Furthermore, the three cytoarchitectonic subdivisions of the IPS show different patterns of functional and structural connectivity at the whole-brain level (Uddin et al., 2010). Thus, converging results from postmortem data and functional connectivity analyses emphasize the robustness of our findings.

Functional subdivisions of the AG—The angular gyrus (AG) is a heteromodal region situated in the ventral division of the posterior parietal cortex. It is implicated in semantic processing, word reading and speech comprehension, number processing, memory retrieval, attention and spatial cognition, reasoning, and social cognition (Binder and Desai, 2011; Seghier, 2013; Uddin et al., 2010a). Lack of proper understanding of its functional subdivisions has contributed to imprecision in understanding its heterogeneous functions. Recent cytoarchitectonic mapping studies of the AG have pointed out that the AG consists of two distinct subdivisions: an anterior (PGa) and a posterior (PGp) (Caspers et al., 2006). Resting-state functional connectivity studies have revealed that these regions have distinct

connectivity patterns: the PGp is strongly coupled with other core nodes of the default mode network, including the ventral medial prefrontal cortex, the posterior cingulate, and medial temporal lobe (Uddin et al., 2010a). In contrast, the PGa is tightly coupled with the inferior frontal gyrus and the caudate (Uddin et al., 2010a). Fiber tracking of DTI data have provided converging evidence for distinct connectivity patterns with the PGa and PGp (Uddin et al., 2010a). These findings have pointed to the AG as a functionally heterogeneous region.

We found evidence for additional functional heterogeneity in the AG beyond the two cytoarchitectonic regions identified in post-mortem brain. We identified three functional divisions within the AG using PRI as the selection criterion, encompassing its anterior, dorsal, and posterior clusters. The functional posterior and anterior clusters overlapped 43% and 63% with the PGp and PGa respectively (Table 1C). The remaining voxels in the PGp and PGa are mainly localized to the dorsal functional cluster. Thus, our findings provide evidence for further heterogeneity within the AG beyond those identified by cytoarchitectonic features alone. Further research using meta-analysis and whole-brain functional connectivity is needed to clarify the functions of the three functional AG sub-regions. Our findings could provide an additional framework for understanding the complex patterns of activation and deactivation reported in the AG (Uddin et al., 2010a; Wu et al., 2009).

Functional subdivisions of the Striatum—Finally, we applied our CC-EAC approach together with the five criteria to discover the optimal number of clusters in the striatum. Among the three cluster combination methods and five objective criteria, KC-HC together with PRI and modified silhouette identified three subdivisions (Figure 6): the dorsal/anterior striatum (encompassing the caudate nucleus), the dorsal/posterior striatum (encompassing the posterior and superior aspects of the putamen), and the ventral/anterior striatum (encompassing the posterior/inferior aspects of the caudate, the nucleus accumbens, and the anterior/inferior aspects of the putamen) (Figure 7E). The diagonal entries of Table 1D suggest that these functional clusters show prominent overlap with the anatomical boundaries of the caudate, putamen and nucleus accumbens respectively. The three clusters identified in our study are consistent with prior functional and anatomical research implicating these regions with distinct functions in the primate brain (Afifi, 2003; Alexander et al., 1986; Postuma and Dagher, 2006; Yelnik, 2008). Briefly, the dorsal/anterior caudate has been consistently reported to be involved in cognitive control and functionally connected to dorsal and lateral aspects of the prefrontal cortex; the putamen is reported to be a primary motor structure, and has been shown to be significantly connected with primary motor areas of cortex; lastly, the ventral striatum (including the nucleus accumbens) is involved in reward and affective behaviors, with extensive connections to the anterior cingulate and orbitofrontal cortices as well as the amygdala (De Martino et al., 2008). Taken together, these findings emphasize the validity of the CC-EAC approach for identifying stable and distinct clusters in brain regions such as the striatum, where it has been particularly difficult to identify the boundaries of the nucleus accumbens.

Future Work

The methods developed here can also be used with other features such as functional connectivity of each voxel in the region of interest to other brain regions (Cohen et al., 2008; Wig et al., 2013a; Wig et al., 2013b; Yeo et al., 2011). Using these methods, one could compute the functional connectivity of each voxel within a ROI with all other voxels in the brain and segment the ROI based on the similarity of their connectivity profiles. However, segmenting the entire brain with either voxel time series or connectivity profiles still remains a challenge. We have also segmented primary sensory cortex, including visual and auditory cortex, using voxel time series as features (data not shown). However, the functional clusters obtained in these regions did not match well with areas V1, V2, V3 (in the case of vision) or areas TE 1.0, TE 1.1 and TE 1.2 (in the case of audition). We are presently investigating these discrepancies and exploring whether additional features such as functional and structural connectivity might help in finding their functional boundaries. To our knowledge, there have been no published reports of resting-state fMRI-based segmentation of these primary sensory areas.

CC-EAC also provides an elegant framework for combining features, including those from different imaging modalities such as diffusion tensor imaging (Fred and Jain, 2005). Combining multiple features from rs-fMRI (e.g. voxel time-series and functional connectivity) with structural imaging features has potential for further improving the reliability of segmentation as each of these features provides complimentary information on brain organization. This approach has the potential to further clarify the subdivisions of heteromodal brain regions such as the insular cortex (Cauda and Vercelli, 2012) and the lateral prefrontal cortex. Future studies will use multimodal imaging features and CC-EAC to address this question.

Conclusions

We have developed a novel clustering method for segmenting brain regions using voxel time-series in a given region of interest as features. Two important and novel aspects of our study include (1) combining complementary clustering methods as opposed to using just one clustering method, and (2) identifying the most accurate metric to determine the number of clusters. Using extensive computer simulations we demonstrate the power of the CC-EAC approach, while resolving multiple issues concerning the determination of the correct number of clusters and the accuracy of various objective criteria used for this purpose. Our results suggest that a combination of K-means and consensus Hierarchical clustering, together with the Probabilistic Rand Index and modified silhouette, is effective in uncovering the optimal number of stable clusters from simulated and experimental rs-fMRI datasets. Our consensus clustering framework is likely to be useful in future research on the human connectome.

Acknowledgements

This research was supported by grants from the National Institutes of Health (1K25HD074652, NS071221 and NS086085). We thank Drs. Claire Kelly and Pierre Bellec for helpful feedback on their prior work. We also thank Dr. Kaustubh Supekar for useful feedback.

References

- Afifi AK. The basal ganglia: a neural network with more than motor function. *Seminars in pediatric neurology*. 2003; 10:3–10. [PubMed: 12785742]
- Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*. 1986; 9:357–381.
- Aron AR, Behrens TE, Smith S, Frank MJ, Poldrack RA. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J Neurosci*. 2007; 27:3743–3752. [PubMed: 17409238]
- Arsalidou M, Taylor MJ. Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage*. 2011; 54:2382–2393. [PubMed: 20946958]
- Barnes KA, Cohen AL, Power JD, Nelson SM, Dosenbach YB, Miezin FM, Petersen SE, Schlaggar BL. Identifying Basal Ganglia divisions in individuals using resting-state functional connectivity MRI. *Front Syst Neurosci*. 2010; 4:18. [PubMed: 20589235]
- Barnes KA, Nelson SM, Cohen AL, Power JD, Coalson RS, Miezin FM, Vogel AC, Dubis JW, Church JA, Petersen SE, Schlaggar BL. Parcellation in left lateral parietal cortex is similar in adults and children. *Cereb Cortex*. 2012; 22:1148–1158. [PubMed: 21810781]
- Beckmann C, Smith S. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging*. 2004; 23:137–152. [PubMed: 14964560]
- Behrens TE, Jenkinson M, Robson MD, Smith SM, Johansen-Berg H. A consistent relationship between local white matter architecture and functional specialisation in medial frontal cortex. *Neuroimage*. 2006; 30:220–227. [PubMed: 16271482]
- Belger A, Puce A, Krystal JH, Gore JC, Goldman-Rakic P, McCarthy G. Dissociation of mnemonic and perceptual processes during spatial and nonspatial working memory using fMRI. *Hum Brain Mapp*. 1998; 6:14–32. [PubMed: 9673660]
- Bellec P, Rosa-Neto P, Lyttelton OC, Benali H, Evans AC. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *Neuroimage*. 2010; 51:1126–1139. [PubMed: 20226257]
- Binder JR, Desai RH. The neurobiology of semantic memory. *Trends Cogn Sci*. 2011; 15:527–536. [PubMed: 22001867]
- Brodman K. Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues Brodman's Localization in the Cerebral Cortex. *Nature Publishing Group*. 1994
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp*. 2001; 14:140–151. [PubMed: 11559959]
- Carpineto C, Romano G. Consensus Clustering Based on a New Probabilistic Rand Index with Application to Subtopic Retrieval. *Ieee T Pattern Anal*. 2012; 34:2315–2326.
- Caspers S, Eickhoff SB, Geyer S, Scheperjans F, Mohlberg H, Zilles K, Amunts K. The human inferior parietal lobule in stereotaxic space. *Brain Struct Funct*. 2008; 212:481–495. [PubMed: 18651173]
- Caspers S, Geyer S, Schleicher A, Mohlberg H, Amunts K, Zilles K. The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *Neuroimage*. 2006; 33:430–448. [PubMed: 16949304]
- Cauda F, D'Agata F, Sacco K, Duca S, Geminiani G, Vercelli A. Functional connectivity of the insula in the resting brain. *Neuroimage*. 2011; 55:8–23. [PubMed: 21111053]
- Cauda F, Vercelli A. How many clusters in the insular cortex? *Cereb Cortex*. 2012
- Chang LJ, Yarkoni T, Khaw MW, Sanfey AG. Decoding the Role of the Insula in Human Cognition: Functional Parcellation and Large-Scale Reverse Inference. *Cereb Cortex*. 2012
- Chen T, Ryali S, Qin S, Menon V. Estimation of resting-state functional connectivity using random subspace based partial correlation: A novel method for reducing global artifacts. *Neuroimage*. 2013; 82:87–100. [PubMed: 23747287]
- Choi HJ, Zilles K, Mohlberg H, Schleicher A, Fink GR, Armstrong E, Amunts K. Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of

- the human intraparietal sulcus. *The Journal of comparative neurology*. 2006; 495:53–69. [PubMed: 16432904]
- Cohen AL, Fair DA, Dosenbach NU, Miezin FM, Dierker D, Van Essen DC, Schlaggar BL, Petersen SE. Defining functional areas in individual human brains using resting functional connectivity MRI. *Neuroimage*. 2008; 41:45–57. [PubMed: 18367410]
- Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 2002; 3:201–215. [PubMed: 11994752]
- De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*. 2008; 43:44–58. [PubMed: 18672070]
- Deen B, Pitskel NB, Pelphrey KA. Three systems of insular functional connectivity identified with cluster analysis. *Cereb Cortex*. 2011; 21:1498–1506. [PubMed: 21097516]
- Di Martino A, Scheres A, Margulies DS, Kelly AM, Uddin LQ, Shehzad Z, Biswal B, Walters JR, Castellanos FX, Milham MP. Functional connectivity of human striatum: a resting state FMRI study. *Cereb Cortex*. 2008; 18:2735–2747. [PubMed: 18400794]
- Duda, RO.; Hart, PE.; Stork, DG. *Pattern Classification*. John Wiley and Sons; 2001.
- Egner T, Monti JM, Trittschuh EH, Wieneke CA, Hirsch J, Mesulam MM. Neural integration of top-down spatial and feature-based information in visual search. *J Neurosci*. 2008; 28:6141–6151. [PubMed: 18550756]
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*. 2005; 25:1325–1335. [PubMed: 15850749]
- Fred ALN, Jain AK. Combining multiple clusterings using evidence accumulation. *Ieee T Pattern Anal*. 2005; 27:835–850.
- Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci*. 1997; 55:119–139.
- Frey SH, Vinton D, Norlund R, Grafton ST. Cortical topography of human anterior intraparietal cortex active during visually guided grasping. *Brain research. Cognitive brain research*. 2005; 23:397–405. [PubMed: 15820646]
- Glover GH, Law CS. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn Reson Med*. 2001; 46:515–522. [PubMed: 11550244]
- Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Springer; 2001.
- Jain AK, Murty MN, Flynn PJ. Data clustering: A review. *Acm Comput Surv*. 1999; 31:264–323.
- Johansen-Berg H, Behrens TE, Robson MD, Drobjnak I, Rushworth MF, Brady JM, Smith SM, Higham DJ, Matthews PM. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc Natl Acad Sci U S A*. 2004; 101:13335–13340. [PubMed: 15340158]
- Kelly C, Toro R, Di Martino A, Cox CL, Bellec P, Castellanos FX, Milham MP. A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage*. 2012
- Kelly C, Uddin LQ, Shehzad Z, Margulies DS, Castellanos FX, Milham MP, Petrides M. Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. *Eur J Neurosci*. 2010; 32:383–398. [PubMed: 20662902]
- Kim JH, Lee JM, Jo HJ, Kim SH, Lee JH, Kim ST, Seo SW, Cox RW, Na DL, Kim SI, Saad ZS. Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: functional connectivity-based parcellation method. *Neuroimage*. 2010; 49:2375–2386. [PubMed: 19837176]
- Kittler J, Hatef M, Duin RPW, Matas J. On combining classifiers. *Ieee T Pattern Anal*. 1998; 20:226–239.
- Meila M. Comparing clusterings - an information based distance. *J Multivariate Anal*. 2007; 98:873–895.
- Menon V, Uddin LQ. Saliency, switching, attention and control: a network model of insula function. *Brain Struct Funct*. 2010; 214:655–667. [PubMed: 20512370]

- Mesulam MM, Mufson EJ. Insula of the old world monkey. I. Architectonics in the insulo-orbito-temporal component of the paralimbic brain. *The Journal of comparative neurology*. 1982; 212:1–22. [PubMed: 7174905]
- Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*. 2003; 52:91–118.
- Nanetti L, Cerliani L, Gazzola V, Renken R, Keysers C. Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. *Neuroimage*. 2009; 47:1666–1677. [PubMed: 19524682]
- Nelson SM, Cohen AL, Power JD, Wig GS, Miezin FM, Wheeler ME, Velanova K, Donaldson DI, Phillips JS, Schlaggar BL, Petersen SE. A parcellation scheme for human left lateral parietal cortex. *Neuron*. 2010; 67:156–170. [PubMed: 20624599]
- Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. *Adv Neur In*. 2002; 14:849–856.
- Picard N, Strick PL. Imaging the premotor areas. *Curr Opin Neurobiol*. 2001; 11:663–672. [PubMed: 11741015]
- Polderman TJ, Gosso MF, Posthuma D, Van Beijsterveldt TC, Heutink P, Verhulst FC, Boomsma DI. A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. *Acta neurologica Belgica*. 2006; 106:191–207. [PubMed: 17323837]
- Postuma RB, Dagher A. Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. *Cereb Cortex*. 2006; 16:1508–1521. [PubMed: 16373457]
- Prichard D, Theiler J. Generating surrogate data for time series with several simultaneously measured variables. *Phys Rev Lett*. 1994; 73:951–954. [PubMed: 10057582]
- Quinlan, JR. Bagging, boosting, and C4.5. *Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference*, Vols 1 and 2; 1996. p. 725-730.
- Rosenberg-Lee M, Chang TT, Young CB, Wu S, Menon V. Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: a cytoarchitectonic mapping study. *Neuropsychologia*. 2011; 49:2592–2608. [PubMed: 21616086]
- Ryali S, Chen T, Supekar K, Menon V. A parcellation scheme based on von Mises-Fisher distributions and Markov random fields for segmenting brain regions using resting-state fMRI. *NeuroImage*. 2012
- Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb Cortex*. 2005; 15:1332–1342. [PubMed: 15635061]
- Scheperjans F, Eickhoff SB, Homke L, Mohlberg H, Hermann K, Amunts K, Zilles K. Probabilistic maps, morphometry, and variability of cytoarchitectonic areas in the human superior parietal cortex. *Cereb Cortex*. 2008; 18:2141–2157. [PubMed: 18245042]
- Seghier ML. The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*. 2013; 19:43–61.
- Shen X, Papademetris X, Constable RT. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. *Neuroimage*. 2010; 50:1027–1035. [PubMed: 20060479]
- Shen X, Tokoglu F, Papademetris X, Constable RT. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*. 2013; 82:403–415. [PubMed: 23747961]
- Shikata E, Hamzei F, Glauche V, Koch M, Weiller C, Binkofski F, Buchel C. Functional properties and interaction of the anterior and posterior intraparietal areas in humans. *Eur J Neurosci*. 2003; 17:1105–1110. [PubMed: 12653987]
- Strehl, A.; Ghosh, J. Cluster ensembles - A knowledge reuse framework for combining partitionings; *Eighteenth National Conference on Artificial Intelligence (Aaai-02)/Fourteenth Innovative Applications of Artificial Intelligence Conference (Iaai-02)*, Proceedings; 2002. p. 93-98.

- Supekar K, Menon V. Developmental maturation of dynamic causal control signals in higher-order cognition: a neurocognitive network model. *PLoS Comput Biol*. 2012; 8:e1002374. [PubMed: 22319436]
- Topchy A, Jain AK, Punch W. Clustering ensembles: Models of consensus and weak partitions. *Ieee T Pattern Anal*. 2005; 27:1866–1881.
- Topchy, A.; Jain, AK.; Punch, W. Combining multiple weak clusterings; Third Ieee International Conference on Data Mining, Proceedings; 2003. p. 331-338.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002; 15:273–289. [PubMed: 11771995]
- Uddin LQ, Supekar K, Amin H, Rykhlevskaia E, Nguyen DA, Greicius MD, Menon V. Dissociable connectivity within human angular gyrus and intraparietal sulcus: evidence from functional and structural connectivity. *Cereb Cortex*. 2010a; 20:2636–2646. [PubMed: 20154013]
- Uddin LQ, Supekar K, Menon V. Typical and atypical development of functional human brain networks: insights from resting-state FMRI. *Front Syst Neurosci*. 2010b; 4:21. [PubMed: 20577585]
- Uddin LQ, Supekar KS, Ryali S, Menon V. Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development. *J Neurosci*. 2011; 31:18578–18589. [PubMed: 22171056]
- Wig GS, Laumann TO, Cohen AL, Power JD, Nelson SM, Glasser MF, Miezin FM, Snyder AZ, Schlaggar BL, Petersen SE. Parcellating an Individual Subject's Cortical and Subcortical Brain Structures Using Snowball Sampling of Resting-State Correlations. *Cereb Cortex*. 2013a
- Wig GS, Laumann TO, Petersen SE. An approach for parcellating human cortical areas using restingstate correlations. *Neuroimage*. 2013b
- Wu SS, Chang TT, Majid A, Caspers S, Eickhoff SB, Menon V. Functional heterogeneity of inferior parietal cortex during mathematical cognition assessed with cytoarchitectonic probability maps. *Cereb Cortex*. 2009; 19:2930–2945. [PubMed: 19406903]
- Yelnik J. Modeling the organization of the basal ganglia. *Revue neurologique*. 2008; 164:969–976. [PubMed: 18808769]
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zoller L, Polimeni JR, Fischl B, Liu HS, Buckner RL. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*. 2011; 106:1125–1165. [PubMed: 21653723]
- Zarei M, Beckmann CF, Binnewijzend MA, Schoonheim MM, Oghabian MA, Sanz-Arigita EJ, Scheltens P, Matthews PM, Barkhof F. Functional segmentation of the hippocampus in the healthy human brain and in Alzheimer's disease. *Neuroimage*. 2012; 66C:28–35. [PubMed: 23128076]
- Zhang D, Snyder AZ, Shimony JS, Fox MD, Raichle ME. Noninvasive functional and structural connectivity mapping of the human thalamocortical system. *Cereb Cortex*. 2010; 20:1187–1194. [PubMed: 19729393]
- Zilles K, Amunts K. Centenary of Brodmann's map--conception and fate. *Nat Rev Neurosci*. 2010; 11:139–145. [PubMed: 20046193]
- Zilles K, Schlaug G, Geyer S, Luppino G, Matelli M, Qu M, Schleicher A, Schormann T. Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. *Adv Neurol*. 1996; 70:29–43. [PubMed: 8615210]

- Develop and validate consensus clustering framework for segmenting brain regions
- Develop and validate criteria to determine optimal number of clusters
- Performance evaluated using extensive computer simulations and resting-state fMRI
- Methods optimal for group-level segmentation clarified

Figure 1. Flow chart for consensus clustering with evidence accumulation (CC-EAC)

CC-EAC can segment a given brain region (ROI) using rs-fMRI data. Step 1: extract time series from a given ROI for each subject; Step 2: generate several base clusters for each subject and for each k ranging from 2 to 10 using a base clustering algorithm starting from different initializations using K-means or Spectral Clustering methods; Step 3: compute the co-association matrices $C_{s,K}$ for each k and s , which indicates how many times a pair of voxels are clustered together across the base clusters; Step 4: apply consensus clustering method with $C_{s,K}$ as the similarity matrix to find stable clusters across all the initialization for each k and s ; Step 5: compute cluster comparison criteria: Normalized Mutual Information (NMI), Variation of Information (VI), Rand Index (RI), Probabilistic Rand Index (PRI) and Modified Silhouette across all the subjects for each k ; Step 6: Identify the optimal number of clusters k^* based on the cluster comparison criteria; Step 7: Compute co-association matrix at the group level for $k = k^*$ by averaging subject level co-association matrices C_{s,k^*} ; Step 8: Find the consensus stable and k^* optimal number of clusters by applying consensus clustering method using the group level coassociation matrix as the similarity matrix.

Synthetic Data Clusters

Figure 2. Simulated clusters used to validate CC-EAC

ROIs with (A) 2, (B) 3, (C) 4, (D) 5, and (E) 6 clusters with arbitrary sizes and shapes constructed from resting state fMRI. The clusters are formed by adding a constant phase (randomly generated) to each voxel time series within a given cluster. This procedure introduces a greater correlation between voxel time series within a cluster while destroying the correlations between voxel time series belonging to different clusters. Additionally, this procedure preserves the spectral profiles of each voxel.

ROIs

Figure 3. Regions of Interest (ROIs) for segmentation

(A) Right SMA and pSMA (SMApSMA), (B) right Insular cortex, (C) right Intraparietal sulcus (IPS), (D) right Angular Gyrus (AG) and (E) right Striatum.

Figure 4. Performance of CC-EAC with various cluster consensus methods on simulated data Modified silhouette, Probabilistic Rand Index (PRI), Rand Index (RI), Variation of Information (VI), and Normalized Mutual Information (NMI) are distinct criteria used for identifying the optimal number of clusters in ROIs. K-means-Hierarchical (KC-HC) and K-means-Spectral Clustering (KC-SC) and Spectral-Spectral (SC-SC) are three different base and consensus clustering pairs examined in the CC-EAC framework. The performance was examined on 2, 3, 4, 5, and 6 distinct ROIs shown in Figure 3. KC-HC with PRI and modified silhouette consistently identifies the correct number of clusters from each ROI. VI suggests 2 as the optimal number of clusters in majority of the cases while RI and NMI monotonically increases with the number of clusters.

KC-HC

Figure 5. Segmentation of simulated regions of interest using CC-EAC

Consensus clustering with KC and HC as base and consensus clustering methods respectively (KC-HC) is able to reliably recover the actual functional boundaries on simulated datasets consisting of 2, 3, 4, 5 and 6 clusters in a group of 21 subjects. The clusters obtained with KC-HC is very similar to the actual clusters shown in Figure 2 (color coding in these two figures is arbitrary).

Figure 6. Performance of various measures for cluster consensus on experimental rs-fMRI datasets

Performance of modified silhouette, Probabilistic Rand Index (PRI), Rand Index (RI), Variation of Information (VI), and Normalized Mutual Information (NMI) in identifying the optimal number of clusters in right SMA-pSMA, Insular cortex, intraparietal sulcus (IPS), angular gyrus (AG), and striatum. Modified silhouette, PRI and VI identify the optimal number of correct number of clusters as two in SMA-pSMA using all the three cluster combination methods. The modified silhouette suggests two clusters for Insula and AG while three clusters for IPS and Striatum with KC-KC as the cluster combination. VI is biased towards two clusters while RI and NMI increase monotonically for all the five ROIs and cluster combinations. PRI resulted in three clusters for Insula, IPS, AG and Striatum with KC-HC as the cluster combination.

KC-HC

Figure 7. Segmentation of brain regions using CC-EAC

Consensus clustering with K-means and Hierarchical clustering (KC-HC) as base and consensus clustering methods respectively together with Probabilistic Rand Index (PRI) is used to find stable and optimal number of clusters at group level from (A) SMA-pSMA is segmented into two clusters using KC-HC at group level whose functional boundaries are similar to the cytoarchitectonic boundaries (Figure, 3A) (B) Insular cortex is segmented into three clusters, (C) IPS into three clusters: hip1, hip2 and hip3 with close correspondence to cytoarchitectonic maps, (D) AG into posterior, anterior and dorsal segments, and (E) striatum into putamen, caudate and nucleus accumbens.

Table 1
Percentage overlaps of functional segments with cytoarchitectonic/anatomical segments

(A) SMA-pSMA, (B) IPS, (C) AG and (D) Striatum. The diagonal elements (shown in red) in each table represent the percent overlap with respect to their cytoarchitectonic or anatomical clusters. The non-diagonal entries show the percentage overlap of a functional cluster with the other adjacent anatomical or cytoarchitectonic segments.

(A) SMA-pSMA			
	SMA	pSMA	
SMA	92	8	
pSMA	14	86	

(B) IPS			
	IPS-1	IPS-2	IPS-3
IPS-1	76	20	4
IPS-2	13	87	0
IPS-3	8	26	65

(C) AG			
	Posterior	Anterior	Dorsal
Posterior	43	14	43
Anterior	0	63	37

(D) Striatum				
	Putamen	Caudate	Nucleus-accumbens	
Putamen	73	0	27	
Caudate	0	97	3	
Nucleus-accumbens	0	11	89	