Inferring Individual-level Variations in the Functional Parcellation of the Cerebral Cortex

Lei Nie, Paul M. Matthews*, and Yike Guo*

***Abstract*— *Objective:* Functional parcellation of the cerebral cortex is variable across different subjects or between cognitive states. Ignoring individual - or state - dependent variations in the functional parcellation may lead to inaccurate representations of individual functional connectivity, limiting the precision of interpretations of differences in individual connectivity profiles. However, it is difficult to infer the individual-level variations due to the relatively low robustness of methods for parcellation of individual subjects. *Methods:* We propose a method called "joint K-means" to robustly parcellate the cerebral cortex using functional magnetic resonance imaging (fMRI) data for contrasts between two states or subjects that intended to characterize variance in individual functional parcellations. The key idea of the proposed method is to jointly infer parcellations in contrasted datasets by iterative descent, while constraining the similarity of the two pathways in searches for local minima to reduce spurious variations. *Results:* Parcellations of resting-state fMRI datasets from the Human Connectome Project show that the similarity of parcellations for an individual subject studied on two sessions is greater than that between different subjects. Differences in parcellations between subjects are non-uniformly distributed across the cerebral cortex, with clusters of higher variance in the prefrontal, lateral temporal and occipito-parietal cortices. This pattern is reproducible across sessions, between groups and using different numbers of parcels. *Conclusion:* The individual-level variations inferred by the proposed method are plausible and consistent with the previously reported functional connectivity variability. *Significance:* The proposed method is a promising tool for investigating relationships between the cerebral functional organization and behavioral differences.**

***Index Terms*— Clustering, functional connectivity, functional parcellation, individual-level variations, K-means**

I. Introduction

THE human cerebral cortex can be parcellated into cytoarchitectonically distinct areas that have distinct functional specializations [1-2], which often are defined as nodes in brain network models of cognitive processing [3-6]. This cortical functional parcellation varies substantially across different individuals [7-9]. Nevertheless, approaches to defining these are still under development or in early validation stages; neuroimaging studies of brain networks typically still address parcellation at a group level [10], ignoring explicit representations of individual-level variations in parcellations. Individual functional parcellations are important for understanding human variation and pathology; inappropriate parcellations have a major impact on functional connectivity inference [11]. Use of group averaged parcellation templates may lead to inaccurate representations of functional connectivity for an individual [12], limiting the precision of interpretations of differences in individual connectivity profiles [13-17]. In such instances, it can be difficult to distinguish between differences in inferred connectivity arising from true differences in connectivity between brain regions or from differences in accuracy of representativeness of the functional parcellation template.

Current approaches to functional parcellations of the cerebral cortex using fMRI [18] can be grouped into two classes of methods. The first class includes hard parcellations, in which each vertex belongs to exactly one parcel. The methods for this class are mainly originated from boundary detection [19-22] and various clustering techniques, such as K-means [23-24], spectral clustering [25-26], hierarchical clustering [27] or model-based clustering [28-31]. The second class includes soft parcellations, in which a vertex can belong to more than one parcel. The methods for this class are mainly based on matrix factorization [32-34].

From the perspective of machine learning, all these methods mathematically formulate the parcellation task as an unsupervised learning problem which aims at inferring a function from unlabeled data. Compared with supervised learning, such as regression and classification, unsupervised learning tasks lack direct measures of success [35], which usually lowers their robustness. In addition, many unsupervised learning problems are NP-hard. Heuristic strategies have to be applied to greedily find a local minimum, which may further decrease the robustness.

Therefore, differences in inferred functional parcellations not only arise from inherent neurobiological differences, but also significantly caused from noise, parameter choices and initializations. If the variations are calculated by directly comparing individual parcellations, such false positive

Manuscript received January 18, 2016; revised April 25, 2016; accepted May 16, 2016. This work was supported a collaborative grant from Biogen and by the Imperial College EPSRC Institutional Sponsorship. *Asterisk indicates corresponding authors.*

L. Nie is with the Department of Computing, Imperial College London. (e-mail: l.nie13@imperial.ac.uk).

*P. M. Matthews is with the Division of Brain Sciences and Centre for Neurotechnology, Department of Medicine, Imperial College London. (e-mail: p.matthews@imperial.ac.uk).

*Y. Guo is with the Department of Computing, Imperial College London and also with School of Computer Engineering and Science, Shanghai University. (e-mail: y.guo@imperial.ac.uk).

variations may be found.

Recent studies addressed this problem in different ways. The most straightforward has been to acquire more fMRI data. To generate an individual functional parcellation, Laumann *et al.* accumulated 14 hours of resting-state fMRI data from one individual over more than a year [36]. A second approach is to introduce additional prior knowledge on the functional parcellation. Dhillon *et al*. proposed an individual sparse PCA with spatial anatomical priors [33]. Gordon et al. developed a procedure to match each cortical vertex in each subject to a group averaged parcellation template using the similarity of connectivity profiles [8]. Wang *et al*. designed another group template matching method based on the similarity of BOLD signals, in which the template is updated with each iteration [9]. Prior knowledge increases the robustness of the parcellation methods, but further research regarding methods for adapting the strength of the prior constraints is needed, especially in application intended to discover variations for which correspondences to a normative template are uncertain or themselves the focus of study.

Here we propose a method called "joint K-means". Our method only requires typical fMRI scan time and is intended to be used in the contrast of two states or subjects. No additional prior knowledge, such as anatomical atlases or group averaged parcellation template, is needed. The key idea of our method is to jointly infer two contrasted parcellations by iterative descent, while constraining the similarity of the two pathways in searches for local minima to reduce spurious variations. Our method is based on the K-means algorithm, which is an effective, simple and widely used clustering algorithm [37]. Given a set of points, the aim of the K-means algorithm is to find a clustering assignment that minimizes the total squared error between any point and the centroid of its associated cluster. For the joint K-means, its aim is to find two contrasted clustering assignments that minimize the weighted summation of the total squared error and the differences between the two clustering assignments. The penalty on the differences is intended to reduce spurious variations and is weighted by a regularization parameter λ. The regularization parameter λ can be tuned automatically. It is set to the minimal value that guarantees the generated parcellations of the original fMRI dataset and its circular block bootstrapped datasets that are similar. We report an evaluation of this method using resting-state fMRI datasets from the Human Connectome Project [38-39]. The results demonstrate test-retest reliability of the parcellations estimated from the two different fMRI sessions acquired from each subject and describe the reproducibility of inter-subject variance found across sessions, between groups and using different numbers of parcels.

II. Methods

Our method assumes that fMRI data are acquired with N vertices and T time points, which are represented by a data matrix $X_{N\times T}$. The n[th] row of the data matrix X is denoted as a vector x_n^{tr}, which is also called the n[th] data point. This represents the time series of blood-oxygen-level dependent (BOLD) signals of the n[th] vertex on the cerebral cortex. The symbol $(\cdot)^{tr}$ denotes the transpose of a vector or a matrix. For a given number of parcels K (defined in advance), a parcellation of the N vertices can be represented by a collection of sets $C = \{c_k | k \in [1\colon K]\}$, where $1\colon K$ indicates all integers from 1 to K. The k[th] set c_k includes all indexes of the points in the k[th] parcel. Equivalently, a parcellation can be represented by a binary matrix $P_{N\times K}$. If the element in the n[th] row and k[th] column of the parcellation matrix P is one, it means that the n[th] vertex belongs to the k[th] parcel in this parcellation. For two contrasted parcellations over a same vertex set, a vertex is called a variation if the indexes of the parcels that this vertex belongs to in the two parcellations are different. In this paper, a parcel and a parcellation are also called a cluster and a clustering assignment, respectively. We focus on hard parcellation, where each vertex belongs to exactly one parcel. Moreover, we do not impose the spatial contiguous constraint on parcellations, which means that there may not be a path between two vertexes in a parcel. Although we only consider parcellating cortical vertices in this paper, our method can be trivially extended to parcellate volumetric voxels.

A. K-means

K-means, one of the most popular and simple clustering algorithms in data analysis [37], has been used frequently for brain parcellation [23][40]. Assume all the N data points $\{x_n | n \in [1\colon N]\}$ are quantitative and the dissimilarity between any two points can be measured using squared Euclidean distance. Let c_k be an index set for the k[th] cluster, and μ_k be the mean value of all points associate with the k[th] cluster. The mean value μ_k is also called the centroid of the k[th] cluster. For the clustering assignment $C = \{c_k | k \in [1\colon K]\}$, the total squared error between any point and its associated mean value is defined as

$$J(C) = \sum_{k=1}^{K} \sum_{i \in c_k} \|x_i - \mu_k\|_2^2. \quad (1)$$

The goal of K-means is to find a clustering assignment C such that the total squared error $J(C)$ is minimized. It is NP-hard to solve this problem [41]. Thus, it is unlikely that polynomial time algorithms can be used to solve this problem [42]. The K-means algorithm greedily finds a local minimum by iterative descent. An iterative greedy algorithm always makes the choice that looks best at each step [43]. The sequence of locally optimal choices does not necessarily lead to a globally optimal result.

Let $|c_k|$ be the number of the elements of the set c_k. There are two major steps of each iteration in K-means (Algorithm 1). The first step (Line 3 to Line 6) is to calculate the centroid of each cluster based on previous clustering assignment. The second step (Line 7 to Line 10) is to assign all points to its closest centroid that is calculated in previous step. There are two user-specified parameters for the K-means algorithm: the number of cluster K and the initialization assignment C^0. The number of clusters in the initialization assignment C^0 should be equal to K. The centroids of the initialization C^0 usually are chosen to be K uniformly random points or K heuristically

Algorithm 1. K-means.

Input: the data $\{x_n | n \in [1:N]\}$, the number of clusters K, and an initialization assignment C^0

Output: the clustering assignment C

```
1   C = C^0
2   do
3     for k = 1:K
4       μ_k ← (Σ_{i∈c_k} x_i)/|c_k|
5       c_k ← ∅
6     end for
7     for n = 1:N
8       l ← argmin_k ||x_n − μ_k||_2^2
9       c_l ← c_l ∪ {n}
10    end for
11  until C stabilizes
```

random seeds [44]. Because K-means is a greedy algorithm and usually finds a local minimum, the initialization assignment significantly impacts the outputted clustering assignment [45]. Different initialization assignments may lead to different local minima.

B. Joint K-means

In order to introduce joint K-means, we reform the total squared error $J(C)$ minimization into a binary optimization problem [46]:

$$\begin{aligned} & min_P \|X - P\Omega\|_2^2 \\ s.t.\ & \Omega_{k,:} = \left(P_{:,k}^{tr} X\right) / \left\|P_{:,k}\right\|_2^2, \quad k \in [1:K] \\ & P_{n,:} \text{ is a binary unit vector, } n \in [1:N], \end{aligned} \tag{2}$$

where $X \in \mathbb{R}^{N \times T}$ is the data matrix; $P \in \{0,1\}^{N \times K}$ is the parcellation matrix; $P_{n,:}$ denotes the n[th] row of the matrix P; $P_{:,k}$ denotes the k[th] column of the matrix P; $\Omega \in \mathbb{R}^{K \times T}$ represents all centroids; $\Omega_{k,:}$ denotes the k[th] row of the matrix Ω, which is the k[th] centroid. Because each row of the matrix P is constrained to be a binary unit vector, exactly one element in each row is one and all other elements are zero. This constraint makes each point belong to exactly one cluster. $P_{:,k}$ indicates the membership of the k[th] cluster. The n[th] point belongs to the k[th] cluster if and only if the n[th] element of $P_{:,k}$ is one. $\left\|P_{:,k}\right\|_2^2$ counts the number of points in the k[th] cluster. $P_{:,k}^{tr} X$ adds together all points in the k[th] cluster. Therefore, $\left(P_{:,k}^{tr} X\right) / \left\|P_{:,k}\right\|_2^2$ is to calculate the centroid of the k[th] cluster. $P\Omega$ is a $N \times T$ dimensional real matrix. The n[th] row of the matrix $P\Omega$ is the centroid associated with the n[th] point. Thus, $\|X - P\Omega\|_2^2$ is the total squared error between any point and its associated centroid, which is equivalent to $J(C)$ in Equation 1.

The goal of K-means is to solve the binary optimization in Equation 2. The parcellation matrix P can be easily calculated from the clustering assignment C outputted by the K-means algorithm. Similarly, the clustering assignment C can be easily derived from the parcellation matrix P.

Given two contrasted data matrices X^1 and X^2, the joint K-means aims at clustering the two data matrices simultaneously. It tries to explore the inherent common structures of the two data matrices while allowing possible variations between them. The optimization problem of the joint K-means is defined as follows:

$$\begin{aligned} & min_{\{P^1, P^2\}} \textstyle\sum_{s=1}^{2} \|X^s - P^s \Omega^s\|_2^2 + \lambda \|P^1 - P^2\|_2^2 \\ s.t.\ & \Omega_{k,:}^s = \left(\left(P_{:,k}^s\right)^{tr} X^s\right) / \left\|P_{:,k}^s\right\|_2^2, \ k \in [1:K], s \in \{1,2\} \\ & P_{n,:}^s \text{ is a binary unit vector, } n \in [1:N], s \in \{1,2\}, \end{aligned} \tag{3}$$

where λ is a regularization parameter that controls the similarity of the two parcellation matrices and definitions of X^s, P^s, Ω^s are identical to those in Equation 2. Comparing Equation 3 with Equation 2, the only difference is that a penalty on variations between two parcellation matrices is added into the loss function. There are two effects of this penalty on variations. Data matrices are usually noisy, especially for fMRI data. The penalty in Equation 3 may reduce insignificant variations due to noises. Moreover, the clustering problem is inherently ill-posed and its binary optimization form is NP-hard. The algorithm finds a local minimal solution for each data matrix by following the corresponding greedy descent pathway. The penalty on variations $\lambda \|P^1 - P^2\|_2^2$ in the Equation 3 forces the two descent pathways to be close to each other; the penalty in Equation 3 is intended to reduce spurious variations. The regularization parameter λ is adjustable.

It is clear that Equation 2 can be reduced to Equation 3 by letting the regularization parameter λ be zero. Thus, both the problem in Equation 3 and that in Equation 2 are NP-hard. The joint K-means algorithm finds local minima for the problem in Equation 3 by greedy descent.

Similar to K-means, there are two major steps of each iteration in joint K-means (Algorithm 2). The first step (Line 3 to Line 9) is to calculate the centroid of each cluster for each data matrix based on previous clustering assignments. The second step (Line 10 to Line 21) is to assign each point to a cluster. Unlike Algorithm 1, Algorithm 2 does not always assign a point to its closest centroid. Instead, there are two options. The first option (Line 11 and Line 12) is to independently assign cluster labels to the two corresponding points in the two data matrix. The two cluster labels may be different. The second option (Line 13) is to give the two corresponding points same cluster labels. The square error of the first choice is always smaller than or equal to that of the second choice. According to Line 14 to Line 20 in Algorithm 2, the second option is chosen if the error difference between the two options is smaller than or equal to 2λ. Based on the clustering assignments C^1 and C^2, the parcellation matrices and variations between the two contrasted parcellations can be easily calculated.

The joint K-means algorithm can be regarded as a hard expectation–maximization algorithm. The first step is the expectation step. Based on the currently estimated parcellation matrices P^1 and P^2 (i.e., C^1 and C^2 in Algorithm 2), the first step is to estimate the centroid matrices Ω^1 and Ω^2 (i.e. $\left\{\mu_k^s \middle| k \in [1:K], s \in \{1,2\}\right\}$ in Algorithm 2) according to the

Algorithm 2. Joint K-means.

Input: the data $\{x_n^s | n \in [1:N], s \in \{1,2\}\}$, the number of clusters K, the regularization parameter λ, and an initialization assignment C^0

Output: two contrasted clustering assignments C^1 and C^2

$C^1 = C^0$
$C^2 = C^0$
do
  for $k = 1:K$
    for $s = 1:2$
      $\mu_k^s \leftarrow \left(\sum_{i \in c_k^s} x_i^s\right) / |c_k^s|$
      $c_k^s \leftarrow \emptyset$
    end for
  end for
  for $n = 1:N$
    $l^1 \leftarrow argmin_k \|x_n^1 - \mu_k^1\|_2^2$
    $l^2 \leftarrow argmin_k \|x_n^2 - \mu_k^2\|_2^2$
    $l^* \leftarrow argmin_k \sum_{s=1}^2 \|x_n^s - \mu_k^s\|_2^2$
    if $\sum_{s=1}^2 \|x_n^s - \mu_{l^*}^s\|_2^2 - \sum_{s=1}^2 \|x_n^s - \mu_{l^s}^s\|_2^2 \leq 2\lambda$
      $c_{l^*}^1 \leftarrow c_{l^*}^1 \cup \{n\}$
      $c_{l^*}^2 \leftarrow c_{l^*}^2 \cup \{n\}$
    else
      $c_{l^1}^1 \leftarrow c_{l^1}^1 \cup \{n\}$
      $c_{l^2}^2 \leftarrow c_{l^2}^2 \cup \{n\}$
    end if
  end for
until C^1 and C^2 stabilize

second line in Equation 3. The second step is the maximization step. Based on the currently estimated centroid matrices Ω^1 and Ω^2, the second step involves finding new parcellation matrices P^1 and P^2 that minimize the loss function in Equation 3.

The time complexity of the joint K-means algorithm is similar to that of the K-means algorithm. For each iteration, the required computation proportional to $O(KN)$. For a typical fMRI dataset, the joint K-means usually terminates within 100 iterations.

C. Tuning the Regularization Parameter

Comparing with the K-means algorithm, the joint K-means algorithm introduces the regularization parameter λ as an adjustable parameter. This subsection presents a method for automatically selecting an appropriate regularization parameter λ from fMRI data.

The regularization parameter λ controls the number of variations on parcellation matrices for the two contrasted data matrices. The two contrasted parcellation matrices from the joint K-means algorithm are identical if the regularization parameter λ is large enough. The two contrasted parcellation matrices from joint K-means algorithm are identical to the corresponding parcellation matrices independently estimated by the K-means algorithm if the regularization parameter λ is zero. An appropriate regularization parameter λ should eliminate spurious variations due to noises or low robustness, while maintaining real variations arising with changes inherent in the data.

Algorithm 3. Circular Block Bootstrap.

Input: the data $\{x_{n,t} | n \in [1:N], t \in [1:T]\}$, the success probability p

Output: the bootstrapped data $\{\hat{x}_{n,t} | n \in [1:N], t \in [1:T]\}$

1 $c \leftarrow 0$
2 **while** $c < T$
3 $\eta \leftarrow randi(T)$
4 $\kappa \leftarrow geornd(p)$
5 $\kappa \leftarrow \min(\kappa, T - c)$
6 **for** $n = 1:N$
7 $\hat{x}_{n,(c+1):(c+\kappa)} \leftarrow x_{n,(\eta \bmod T)+1:((\eta+\kappa-1) \bmod T)+1}$
8 **end for**
9 $c \leftarrow c + \kappa$
10 **end**

We assume that the number of variations on parcellation matrices should be very small if the two contrasted data matrices are generated from an identical model and the noise is low, e.g., the number of variations in parcellations should be small if the data matrices are generated from same subject acquired under same conditions.

Given a data matrix, another data matrix can be resampled from the original data matrix. We assume the original data matrix and the resampled data matrix approximately follow a similar distribution, and then find the minimal value of the regularization parameter λ that produces a sufficiently small number of variations in the two parcellations. Because brain parcellation using fMRI data is the targeted application of the joint K-means algorithm, each row of a data matrix is regarded as a time series. The temporal dependences should be preserved during resampling. Thus, we use the circular block bootstrap technique [47]. The efficiency of the circular block bootstrap (Algorithm 3) in replicating the distribution of spatial correlation has been demonstrated by previous studies [48][24].

Algorithm 3 generates a data matrix whose dimensions are identical to the original data matrix. The function $randi(T)$ generates a uniformly random integer between 1 and T. The function $geornd(p)$ generates random numbers from a geometric distribution with success probability p. Algorithm 3 samples several blocks with various starting points and lengths. Line 3 in Algorithm 3 samples a start point for a block uniformly random between 1 and T. Line 4 in Algorithm 3 generates the length of this block, which follows a geometric distribution with success probability p. Line 5 shortens the block length if it is too long. Line 6 to Line 8 stack the data in this block into the bootstrapped data matrix. Line 7 aligns the data circularly, so that a block falling at the end of the time series continues into the beginning. The only user-specified parameter for the circular block bootstrap is the success probability p, which controls the lengths of blocks. The mean length of blocks is $(1-p)/p$. It has been reported that the relationships between fMRI time series can be robustly identified roughly from 30 to 60 seconds in conventional acquisitions [49]. Thus, we recommend setting the success probability p be between 0.0164 and 0.0323.

The circular block bootstrap procedure is assumed to

Algorithm 4. Regularization parameter estimation.

Input: the data $\{x_n^s | n \in [1:N], s \in \{1,2\}\}$, the number of clusters K, an initialization assignment C^0, and the number of bootstrap variations Z

Output: the regularization parameter λ

1 $C^1 = C^0$
2 $C^2 = C^0$
3 **for** $r = 1:2$
4 **for** $k = 1:K$
5 **for** $s = 1:2$
6 $\mu_k^s \leftarrow \left(\sum_{i \in c_k^s} x_i^s\right) / |c_k^s|$
7 $c_k^s \leftarrow \emptyset$
8 **end for**
9 **end for**
10 $\Theta \leftarrow \emptyset$
11 **for** $n = 1:N$
12 $l^1 \leftarrow argmin_k \|x_n^1 - \mu_k^1\|_2^2$
13 $l^2 \leftarrow argmin_k \|x_n^2 - \mu_k^2\|_2^2$
14 $l^* \leftarrow argmin_k \sum_{s=1}^2 \|x_n^s - \mu_k^s\|_2^2$
15 $\theta \leftarrow \sum_{s=1}^2 \|x_n^s - \mu_{l^*}^s\|_2^2 - \sum_{s=1}^2 \|x_n^s - \mu_{l^s}^s\|_2^2$
16 $\Theta \leftarrow \Theta \cup \{\theta\}$
17 **end**
18 $\Theta \leftarrow sort(\Theta, \text{descend})$
19 $\lambda \leftarrow \max\{0.5 \times \Theta(Z+1), \lambda\}$
20 **for** $n = 1:N$
21 $l^1 \leftarrow argmin_k \|x_n^1 - \mu_k^1\|_2^2$
22 $l^2 \leftarrow argmin_k \|x_n^2 - \mu_k^2\|_2^2$
23 $l^* \leftarrow argmin_k \sum_{s=1}^2 \|x_n^s - \mu_k^s\|_2^2$
24 **if** $\sum_{s=1}^2 \|x_n^s - \mu_{l^*}^s\|_2^2 - \sum_{s=1}^2 \|x_n^s - \mu_{l^s}^s\|_2^2 \leq 2\lambda$
25 $c_{l^*}^1 \leftarrow c_{l^*}^1 \cup \{n\}$
26 $c_{l^*}^2 \leftarrow c_{l^*}^2 \cup \{n\}$
27 **else**
28 $c_{l^1}^1 \leftarrow c_{l^1}^1 \cup \{n\}$
29 $c_{l^2}^2 \leftarrow c_{l^2}^2 \cup \{n\}$
30 **end if**
31 **end for**
32 **end**

generate data with the same distribution as the original data. However, the bootstrap procedure can generate "outliers" that are very different from the original data. Thus, there should be some variations between the two parcellations derived from the original and bootstrapped data matrices. It is recommended that a threshold Z for bootstrap variations is set to a value of about 0.1% of the number of data points N, but not 0. This value is an empirical estimation on the number of the outliers caused by the bootstrap procedure.

Our aim is to find the minimal value of the regularization parameter λ such that the number of variations on parcellations between original and bootstrapped data matrices is smaller than Z. The exact relationship between the regularization parameter λ and the number of bootstrap variations calculated by the joint K-means algorithm is still unknown. Given the number of bootstrap variations Z, we only heuristically estimate a value of the regularization parameter λ (Algorithm 4).

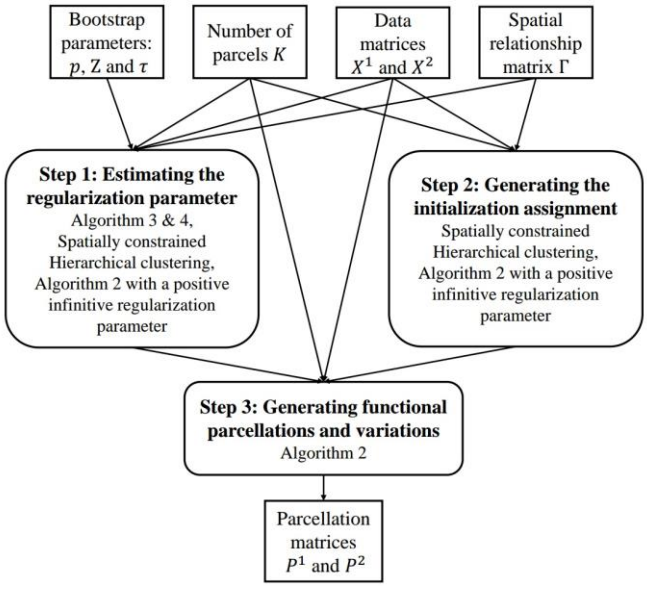

Fig. 1. Overview of the pipeline.

Algorithm 4 is similar to Algorithm 2. However, input for Algorithm 4 does not require the regularization parameter λ, but the number of bootstrap variations Z. The output of Algorithm 4 is an estimated value of the regularization parameter λ. There are three major steps of each iteration in Algorithm 4. The first step (Line 4 to Line 9) is to calculate the centroid of each cluster for each data matrix based on previous clustering assignments. The second step (Line 10 to Line 19) is to estimate a value of the regularization parameter λ based on the squared errors under two different conditions. One condition (Line 12 to Line 13) is that two corresponding points for the two data matrices can be assigned with different cluster labels. The other condition (Line 14) is that the cluster labels for two corresponding points have to be identical. The list Θ records the error differences between the two conditions for all points. The list Θ is then sorted in a descend order and the regularization parameter λ is set to 0.5 multiples the $(Z+1)^{\text{th}}$ element in the list Θ. This value of the regularization parameter λ ensures that there are Z bootstrap variations in this iteration. Using this estimated value of the regularization parameter λ, the third step (Line 20 to Line 31) is to assign each point to a cluster, which is identical to the second step in Algorithm 2. Unlike Algorithm 2, the three major steps are only iterated twice, which was found empirically to be sufficient for finding an appropriate value of the regularization parameter λ. It is noteworthy that Algorithm 4 is just a greedy algorithm. The regularization parameter λ estimated by Algorithm 4 does not necessarily yield Z bootstrap variations in the joint K-means algorithm. However, the number of bootstrap variations is close to Z.

D. The Pipeline

This subsection introduces a pipeline (Fig.1) for parcellation cerebral cortices into several functional areas using fMRI datasets and for discovering individual-level variations between these functional parcellations.

The input data for this pipeline include two contrasted data

matrices and a spatial relationship matrix. The two contrasted data matrices are denoted as X^1 and X^2 respectively, each row of which is the BOLD signals for a vertex. Typically, each data matrix is an fMRI session for one subject. There may be several runs in each session. The two contrasted data matrices can be from two states of one subject or two subjects. The spatial relationship matrix is denoted as Γ, which is symmetric and binary. If the element in the ith row and jth column of the spatial relationship matrix Γ is one, it means that the ith and the jth vertex are spatially connected to each other. The spatial relationship matrix Γ is optional. It is used to boost the speed of initialization assignment generation. There are four input parameters for this pipeline: the number of parcels K, the success probability p, the number of bootstrap variations Z, and the resampling number τ.

The output of this pipeline is two contrasted parcellations matrix, P^1 and P^2. The variations between these two contrasted parcellations can be calculated easily from the two contrasted parcellation matrices. As this pipeline is only based on individual data (not on group average data or group-level prior), it provides direct estimates of individual-level variations in functional parcellations.

There are three steps in this pipeline: estimating the regularization parameter, generating the initialization assignment, and generating functional parcellations and variations.

This pipeline firstly estimates a value of regularization parameter λ for each data matrix separately. For a data matrix X^s, Algorithm 3 is used to resample a data matrix $\hat{X}^s$ using the inputted success probability p and the number of bootstrap variations Z. The each row of data matrices, X^s and $\hat{X}^s$, are normalized to a unit vector with zero mean. If there are several runs in a session, the data in each run are normalized separately. For simplicity, the normalized matrices are still denoted as X^s and $\hat{X}^s$. Afterwards, the data matrices X^s and $\hat{X}^s$ are stacked horizontally into one matrix $[X^s, \hat{X}^s]$.

Spatially constrained hierarchical clustering [27][50] with Ward's linkage rules [51] is applied to the stacked data matrix $[X^s, \hat{X}^s]$ and the spatial relationship matrix Γ to generate K parcels. This parcellation is then used as an initialization assignment for Algorithm 2 (Joint K-means) to cluster the data matrices X^s and $\hat{X}^s$ with the regularization parameter λ set to positive infinity. This estimated parcellation by Algorithm 2 is used as the initialization assignment for Algorithm 4. At this point, Algorithm 4 is applied to the two data matrices X^s and $\hat{X}^s$, the initialization assignment estimated previously and the inputted number of bootstrap variations Z. Algorithm 4 outputs an estimated value of regularization parameter λ. This bootstrap process is executed τ times to generate τ values of the regularization parameter λ. We regard the 95th percentile of the τ values as an appropriate value of the regularization parameter λ for the data matrix X^s, which is denoted as $\hat{\lambda}^s$.

This pipeline then generates an initialization assignment for joint K-means. The original data matrices X^1 and X^2 are normalized and stacked horizontally into one matrix $[X^1, X^2]$. Similar to the previous step, spatially constrained hierarchical clustering [27][50] with Ward's linkage rules [51] is applied to the stacked data matrix $[X^1, X^2]$ and the spatial relationship matrix Γ to generate K parcels. This parcellation is then used as an initialization assignment for Algorithm 2 (joint K-means) to cluster the data matrix X^1 and X^2 with the regularization parameter λ set to positive infinity. The estimated parcellation is used as the initialization assignment for next step, which is denoted as $\hat{C}^0$.

This pipeline finally generates functional parcellations and their corresponding variations using Algorithm 2 (Joint K-means). The data matrices for Algorithm 2 are normalized X^1 and X^2. The number of clusters K is given in advance. The regularization parameter is set to $max\{\hat{\lambda}^1, \hat{\lambda}^2\}$, estimated by the first step of this pipeline. The initialization assignment is set to $\hat{C}^0$, which is generated by the second step of this pipeline. The output of the third step is the output of this pipeline.

E. Discussion

Compared with the standard K-means algorithm, the joint K-means algorithm can significantly increase the robustness to noises and initialization assignments. Given two data matrices X^1 and X^2, the standard approach would be to *independently* infer the clustering assignment for each data matrix using the K-means algorithm. The K-means algorithm is executed twice, one for the matrix X^1 and the other for the matrix X^2. Even if the initialization assignments for the two executions are identical and the the data matrices, X^1 and X^2, are similar to each other, the two inferred clustering assignments from the two executions can be very different because the two descent pathways of the two executions of the K-means algorithm can bifurcate due to minor differences between the data matrices, leading to two completely different local minima. However, for the *joint* K-means algorithm, minor differences between the data matrices are ignored because of the penalty term in Equation 3. Only major differences between the data matrices lead to variations in final clustering assignments. The regularization parameter λ controls the strength of the differences that are ignored.

In the Wang's method [9], spurious variations were reduced using a group averaged parcellation template. This template did not only include the initialization assignment, but also the pre-estimated inter-subject variability in functional connectivity and temporal signal-to-noise ratio (SNR) [9]. The role of the information on inter-subject variability and SNR is similar to penalty term in Equation 3, which reduces the effects of minor differences in data. In theory, results from this method are impacted by the template (i.e., the initialization assignment and variability). The applicability of the template may be confounded by large "true" inter-subject variability, which may be difficult to infer in advance or to generalize from one group to another group or one state to another state. In contrast, the inter-subject variability information is not required by the joint K-means algorithm.

The obvious drawback of the joint K-means algorithm is that it can only be used for comparing two data matrices. Another potential drawback is that it might be difficult to combine the two estimated regularization parameters $\hat{\lambda}^1$ and $\hat{\lambda}^2$ if the corresponding data matrices X^1 and X^2 are acquired using different protocols. In this paper, we implicitly assume that the two contrasted data matrices are acquired using similar protocols, so that the two estimated regularization parameters $\hat{\lambda}^1$ and $\hat{\lambda}^2$ are close to each other and can be combined using $max\{\hat{\lambda}^1, \hat{\lambda}^2\}$. Without this assumption, e.g., one data matrix has 2400 frames in 30 minutes and the other matrix has 280 frames in 5 minutes, the two estimated parameters may be very different from each other. It is unclear whether the maximum of the two estimated parameters are still a good choice.

F. Possible Extensions

Although this paper only focuses on discovering individual-level variations between two contrasted functional parcellations, the proposed method can be extended to define variations between multiple functional parcellations. Given a set of data matrices $\{X^s|s \in [1:S]\}$, Equation 3 can be extended as follows:

$$\begin{aligned} &min_{\{p^s\}} \sum_{s=1}^{S} \|X^s - P^s\Omega^s\|_2^2 + \sum_{s=1}^{S} \lambda^s \|P^s - \bar{P}\|_2^2 \\ s.t.\ &\Omega_{k,:}^s = \left(\left(P_{:,k}^s\right)^{tr} X^s\right) / \left\|P_{:,k}^s\right\|_2^2,\ k \in [1:K], s \in [1:S] \\ &P_{n,:}^s \text{ is a binary unit vector},\ n \in [1:N],, s \in [1:S], \end{aligned} \quad (4)$$

where λ^s is the regularization parameter for the data matrix X^s; $\bar{P}$ is the average parcellation. Because it is computationally expensive to compare the parcellations for every pair of data matrices, the average parcellation $\bar{P}$ is introduced.

Algorithm 5 is designed to greedily find a local minimum for the problem in Equation 4 using an iterative strategy, which is an extension of Algorithm 2. The symbol $label(C, n)$ represents the cluster label of the nth data point in the clustering assignment C. There are three major steps in Algorithm 5. The first step (Line 6 to Line 11) is to calculate the centroid of each cluster based on previous clustering assignment. The second step (Line 12 to Line 22) assigns all points to a centroid. A point is assigned to the closest centroid or the corresponding centroid defined in the current average parcellation. The third step (Line 23 to Line 29) is to update the average parcellation by weighted vote.

The joint K-means algorithm can be further extended to compute variations between different groups, e.g., a healthy and a disease state, in which case, there should be two average parcellations in Equation 4. However, the extended algorithm is memory expensive, because all data matrices are involved in each iteration. For example, the size of one session resting-state data from the Human Connectome Project is about 800MB. For 20 subjects, it costs about 16GB for storing the data. The total memory requirement for computation is approximately doubled. If the memory in a machine cannot store all data matrices, some data matrices have to be transferred between memory and hard disk in each iteration. The typical bandwidth of hard disks is only about 1% to 10% of memory.

Algorithm 5. Joint K-means for multiple data matrices.

Input: the data $\{x_n^s|n \in [1:N], s \in [1:S]\}$, the number of clusters K, the regularization parameters $\{\lambda^s|s \in [1:S]\}$, and an initialization assignment C^0

Output: the clustering assignments $\{C^s|s \in [1:S]\}$

```
1   for s = 1:S
2     C^s = C^0
3   end
4   C̄ = C^0
5   do
6     for k = 1:K
7       for s = 1:S
8         μ_k^s ← (Σ_{i∈c_k^s} x_i^s) / |c_k^s|
9         c_k^s ← ∅
10      end for
11    end for
12    for n = 1:N
13      l̄ ← label(C̄, n)
14      for s = 1:S
15        l ← argmin_k ||x_n^s − μ_k^s||_2^2
16        if ||x_n^s − μ_l̄^s||_2^2 − ||x_n^s − μ_l^s||_2^2 ≤ 2λ^s
17          c_l^s ← c_l^s ∪ {n}
18        else
19          c_l̄^s ← c_l̄^s ∪ {n}
20        end if
21      end for
22    end for
23    for k = 1:K
24      c̄_k ← ∅
25    end for
26    for n = 1:N
27      l ← argmin_k Σ_{s=1}^S λ^s |label(C^s, n) == k|
28      c̄_l = c̄_l ∪ {n}
29    end for
30  until {C^s|s ∈ [1:S]} stabilize
```

Spectral clustering is an alternative method that has been used for brain parcellation [25-26][50]. This method is based on a similarity matrix, each element of which measures the similarity of a pair of vertices. The spatial constraint can be easily integrated in this matrix by forcing the similarity between any disconnected vertices to be zero. The first K eigenvectors of the unnormalized or normalized Laplacian of the similarity matrix are computed and used as feature vectors for vertices. The K-means algorithm is then applied to these feature vectors. The joint K-means algorithm can be directly replace the K-means algorithm for comparing two similarity matrices.

III. CASE STUDY

A. Datasets and Preprocessing

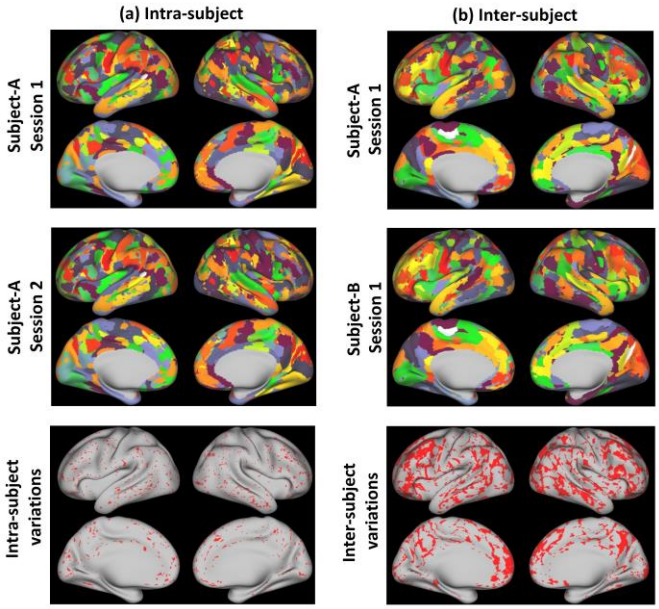

Fig. 2. Intra-subject and inter-subject demonstrations of functional parcellations and the corresponding variations. The number of parcels in each of the left and right cerebral hemispheres is 150.

Resting-state fMRI datasets from by the Human Connectome Project [38-39] were used for demonstration of the proposed method. These datasets were called "40 Unrelated Subjects" by the Human Connectome Project, which were publicly available at http://www.humanconnectome.org/. These datasets were generated from 38 unrelated subjects. All these subjects were healthy and their ages were between 22 and 35. There were 21 females and 17 males. To allow an independent test the reproducibility of the variations in the parcellation, the 38 subjects were divided into two groups. All subjects were sorted in an ascending order according to their identities. The first, group A, included the first 19 subjects; while the second, group B, included the others. The subjects' ages in both groups ranged from 22 to 35. There were 10 females in group A and 11 females in group B.

For each subject, the resting-sate fMRI data were acquired over four runs, each of 14 minutes and 33 seconds (including 1200 whole brain frames). The four acquisition runs were acquired in two independent imaging sessions conducted on sequential days. Within each session, the oblique axial acquisition of one run was with phase encoding in a right-to-left direction; phase encoding in a left-to-right direction was used for the other run. The HCP scanning protocol was approved by the local Institutional Review Board at Washington University in St. Louis.

The functional images were acquired using a gradient-echo EPI sequence with the following parameters: repetition time (TR) = 720ms, echo time (TE) = 33.1ms, flip angle = 52o, FOV = 208×180mm, slice thickness = 2.0mm, 72 slices. The data was preprocessed using FSL [52], FreeSurfer [53] and Connectome Workbench [54]. Individual images were registered to a group-average surface template using multimodal surface matching algorithm [55-56] applied with "cortical folding only" mode. The data were de-noised using the ICA-FIX method [57-58].

The CIFTI gray ordinates data [59] were used, but subcortical areas were discarded in our experiments. The exclusion of subcortical areas does not significantly impact the results, because the similarity of two vertices in this paper was calculated using the Euclidean distance between the normalized BOLD signals of these two vertices. It can be easily proved that the Euclidean distance of normalized vectors and their Pearson correlation coefficient are equivalent and related by a liner transformation. It has been demonstrated that the correlation-based measure and the connectivity-based measure perform similarly [27].

The left and right cerebral hemispheres were processed independently. Vertices without BOLD signals were discarded. 29,696 and 29,716 vertices were found for the left and right cerebral hemispheres, respectively. Considering that the total number of human neocortical areas was estimated to be about 150 to 200 per hemisphere [60] and the resting-state BOLD signals may not distinguish all cortical areas, three different numbers of parcels have been tried, which were 100, 150 and 200 for each cerebral hemisphere (i.e. 200, 300 and 400 parcels for the whole cerebral cortex). To estimate the regularization parameter λ, the success probability p, the number of bootstrap variations Z, and the number of resampling τ were set to 0.0164, 297 and 20. Thus, mean length of blocks was 60 (about 43.65 seconds); and there were approximate 0.1% of the total vertices were variations using original and bootstrapped fMRI datasets. The spatial relationship matrix was generated according to two Human Connectome Project defined group-average R440 surface files: Q1-Q6_R440.L.inflated.32k_fs_LR.surf.gii and Q1-Q6_R440.R.inflated.32k_fs_LR.surf.gii.

Pairs of contrasted parcellations were inferred using two types of pairwise (i.e. intra-subject and inter-subject) datasets; their variations were calculated by directly comparing the two contrasted parcellations. There were 19 pairs of intra-subject parcellations in each group, which were inferred in the contrast of session 1 and session 2 datasets acquired from same subjects. There were 171 pairs of inter-subject parcellations for each session in each group, which were inferred in the contrast of session 1 or session 2 datasets acquired from two different subjects. Each dataset was parcellated several times in different contrasted pairs; these parcellations for the dataset may not be same.

B. Demonstrations of parcellations and their variations

The intra-subject example was in the contrast of the session 1 and session 2 datasets acquired from a subject (Fig. 2a). For the left hemisphere, there were 1187 variations, which accounted for 4.0% of the vertices; for the right hemisphere, there were 1422 variations, which accounts for 4.8% of the vertices. The inter-subject example was in the contrast of the session 1 datasets acquired from two subjects (Fig. 2b). Approximately 24% of the vertices (7017 and 7244 on the left and right hemispheres respectively) showed variation between the two subjects. The number of inter-subject variations thus was approximate 5.5 times greater of that found within a

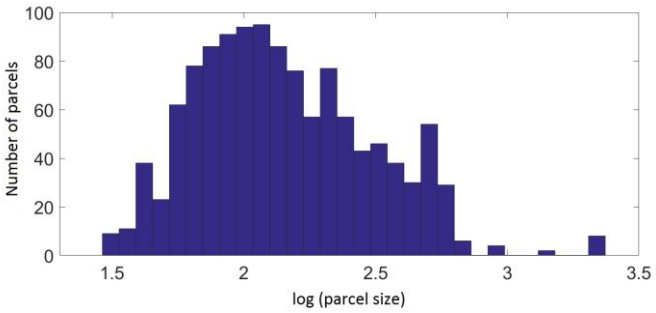

Fig. 3. Distribution of parcel sizes.

subject with these examples. The inter-subject variations were more heterogeneous than intra-subject variations, clustering in regions of prefrontal, lateral temporal and occipito-parietal cortices.

Sizes of parcels varied about 2 orders of magnitude (Fig. 3), which was consistent with previous estimation [60]. The largest parcel size in these parcellations was 2355; the smallest one was 29. Most of parcels included less than 1000 vertices. The sizes of about half of the parcels were between 50 and 150.

C. *Test-retest Reliability*

The Dice [61] and Jaccard [62] coefficients were used to measure the similarity between contrasted parcellations. Given two sets A and B, the Dice coefficient between two sets is defined as follows:

$$Dice = 2|A \cap B|/(|A| + |B|), \quad (5)$$

where $|\cdot|$ indicates the number of elements in a set, while the Jaccard coefficient is defined as:

$$Jaccard = |A \cap B|/|A \cup B|. \quad (6)$$

Higher Dice and Jaccard coefficients are found with more similar parcels. A summary Dice (or Jaccard) coefficient between two parcellations was defined as the arithmetic mean of all Dice (or Jaccard) coefficients between matched parcels.

The average Dice coefficients for contrasted intra-subject parcellations were significantly higher than those for contrasted inter-subject parcellations (Fig. 4). For group A, the intra-subject Dice coefficients were $94.1\% \pm 2.2\%$ ($mean \pm s.d.$) for the left hemisphere and $94.3\% \pm 2.0\%$ for the right hemisphere, while the inter-subject Dice coefficients were $73.5\% \pm 3.2\%$ for the left hemisphere and $74.4\% \pm 3.0\%$ for the right hemisphere. This observation was reproduced in group B. The intra-subject Dice coefficients were $94.3\% \pm 2.6\%$ for the left hemisphere and $94.3\% \pm 2.5\%$ for the right hemisphere; the inter-subject Dice coefficients were $75.3\% \pm 3.3\%$ for the left hemisphere and $74.9\% \pm 3.2\%$ for the right hemisphere. Moreover, the differences between intra-subject Jaccard coefficients and inter-subject Jaccard coefficients were even higher (approximately 27%). Together, these results suggest that the proposed method has a high test-retest reliability and sufficient sensitivity to detect inter-subject variance.

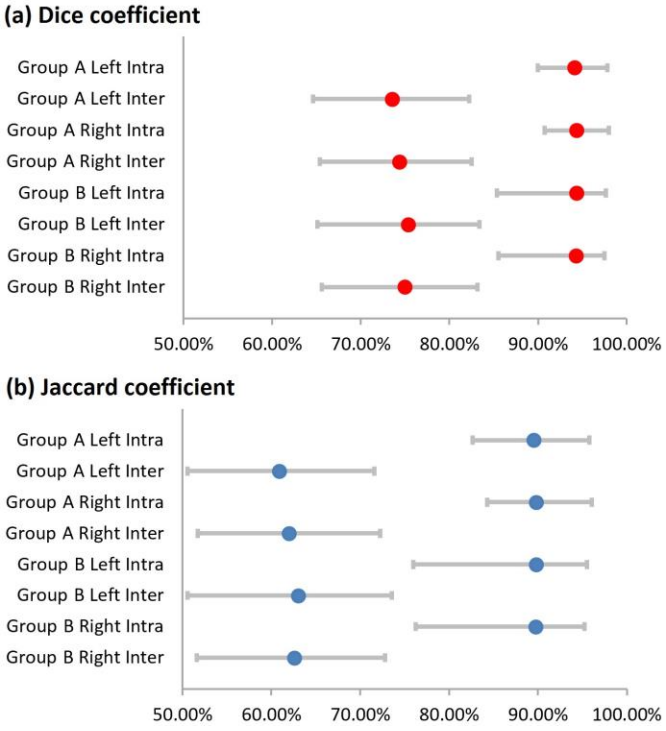

Fig. 4. Distributions of the Dice and Jaccard coefficients. The dots indicate the mean values, and the two ends of the lines represent the minimal and maximal values.

The *minimal* intra-subject Dice coefficient (85.4%) was higher than the *maximal* inter-subject Dice coefficient (83.4%); and the *minimal* intra-subject Jaccard coefficient (76.0%) was higher than the *maximal* inter-subject Jaccard coefficient (73.6%). The differences between intra- and inter-subject variances suggest the proposed method can be used to identify whether two resting-state fMRI datasets are from a same subject.

D. *Intra- and Inter-subject Variations*

The value of a vertex in a variation map indicated the probability of this vertex being a variation. For example, if the value of a vertex in an intra-subject map was 0.26 ($\approx 5/19$), this vertex belonged to different parcels for 5 intra-subject pairs of contrasted fMRI datasets; if the value of a vertex in an inter-subject map was 0.26 ($\approx 45/171$), this vertex belonged to different parcels for 45 inter-subject pairs of contrasted fMRI datasets.

The inter-subject variations were more non-uniformly distributed across the cerebral cortex than the intra-subject variations, clustering in the prefrontal, lateral temporal and occipito-parietal cortices for both groups (Fig. 5). These areas were correlated with cognitive differences between subjects [63-64]. Previous studies [65-66] showed that these areas had high inter-subject connectivity variability. These studies on connectivity were based on fixed parcellations across all subjects without considering possible inter-subject variability on parcellations.

For intra-subject variations, no remarkable and reproducible patterns were observed by us. Because there were only two fMRI sessions, it is difficult to define the causes of these

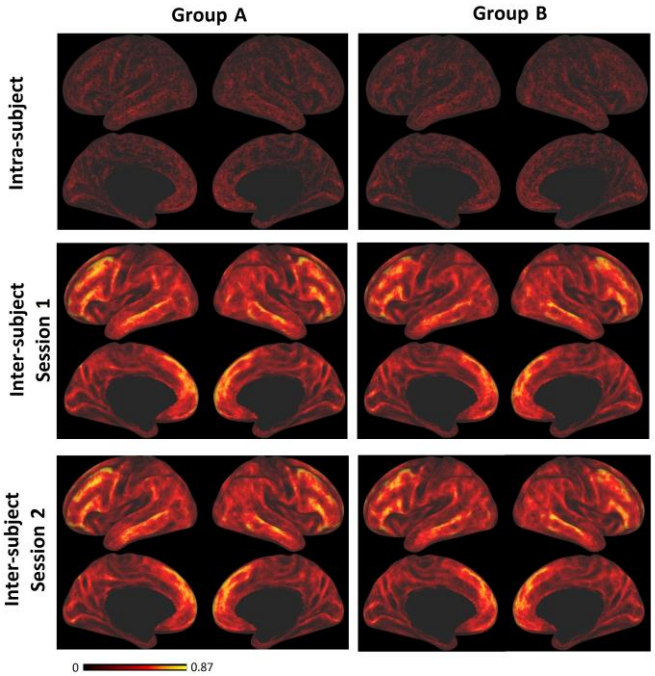

Fig. 5. Illustrative intra- and inter-subject variation maps. The maximal value in these maps is 0.87.

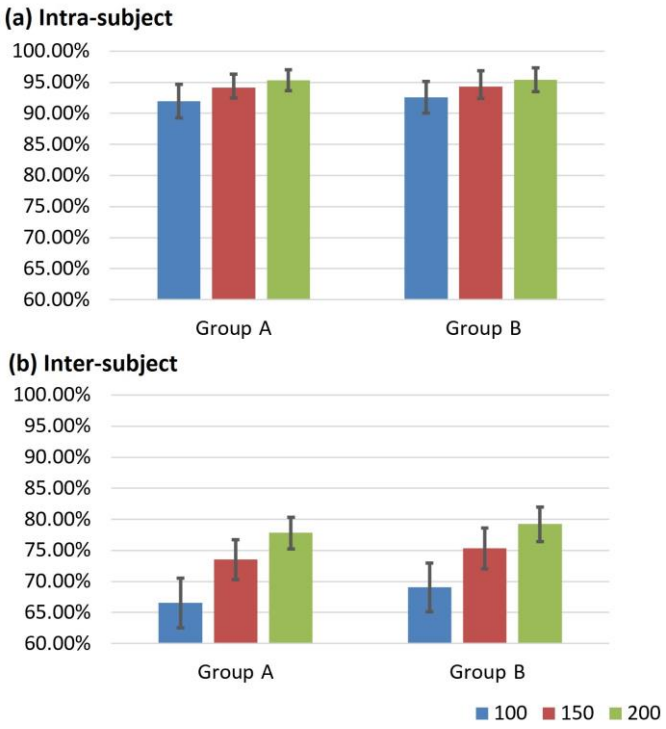

Fig. 6. Dice coefficients with different numbers of parcels.

variations. Intra-subject variations may be resulted from differences in conscious state between sessions. Other factors, e.g. radiofrequency power and distribution, sampling noise or differences in physiological states (e.g., respiratory rate), which also could contribute to intra-subject variations. Further research on this issue is needed.

A variation map can be reformed into a vector indexed by the positions of vertices. The spatial correlation coefficient, which was the Pearson correlation coefficient between the two reformed vectors, was used to measure the similarity of two variation maps. The coefficients for each cerebral hemisphere were calculated separately. The correlation coefficients between the intra-subject maps of group A and group B were 0.33 (left) and 0.34 (right). The correlation coefficients between the inter-subject maps of group A and group B were 0.84 (left, session 1), 0.83 (right, session 1), 0.84 (left, session 2) and 0.84 (right, session 2). The correlation coefficients between the inter-subject maps for the two session were as high as 0.92 for both cortices and groups. These results provided support for the high reproducibility of the inter-subject variation maps across sessions and groups.

The relationship between variations and parcel sizes was also investigated using spatial correlation coefficient. The value of each vertex in a parcel size map is the average size of the parcels that this vertex belonged to. The correlation coefficient was estimated between a variation map and its corresponding parcel size map. For intra-subject variation maps, the coefficients were 0.34 (left, group A) and 0.35 (right, group A), 0.27(left, group B) and 0.31 (right, group B). For inter-subject variation maps, the coefficients were -0.018 (left, group A, session 1), 0.22 (right, group A, session 1), -0.034 (left, group A, session 2), 0.0047 (right, group A, session 2), 0.0080 (left, group B, session 1), 0.054 (right, group B, session 1), -0.0039 (left, group B, session 2) and 0.010 (right, group B, session 2). These results showed that the parcel size might have some positive effect on intra-subject variations; but no major effect on inter-subject variations was detected.

E. Different Numbers of Parcels

Three different numbers of parcels, which were 100, 150 and 200 for the left hemisphere, were investigated. The intra-subject Dice coefficients were higher than the inter-subject Dice coefficients (Fig. 6). Both intra-subject and inter-subject Dice coefficients increased with the number of parcels, which was agreed with the observation in [27]. Comparing with intra-subject Dice coefficients, the inter-subject Dice coefficients had more obvious increases, from 66.5% to 77.8% in group A, and from 69.1% to 79.2% in group B. For any fixed number of parcels, the minimal intra-subject Dice coefficient was higher than the maximal inter-subject Dice coefficient. Similar phenomena were observed using the Jaccard coefficient.

The variation maps with different numbers of parcels showed similar patterns (Fig. 7). For all numbers of parcels, the inter-subject variations were non-uniformly distributed and cluster in the prefrontal, lateral temporal and occipito-parietal cortices. Areas of higher variance became smaller increasing numbers of parcels. The spatial correlation coefficient was used to measure the reproducibility of variation maps. The correlation coefficients between intra-subject variation maps were 0.58 (100 vs. 150), 0.66 (150 vs. 200) and 0.50 (100 vs. 200). The correlation coefficients between inter-subject variation maps were much higher at 0.96 (100 vs. 150), 0.97 (150 vs. 200), and 0.91 (100 vs. 200). These results demonstrated that the inter-subject variation maps are robust to

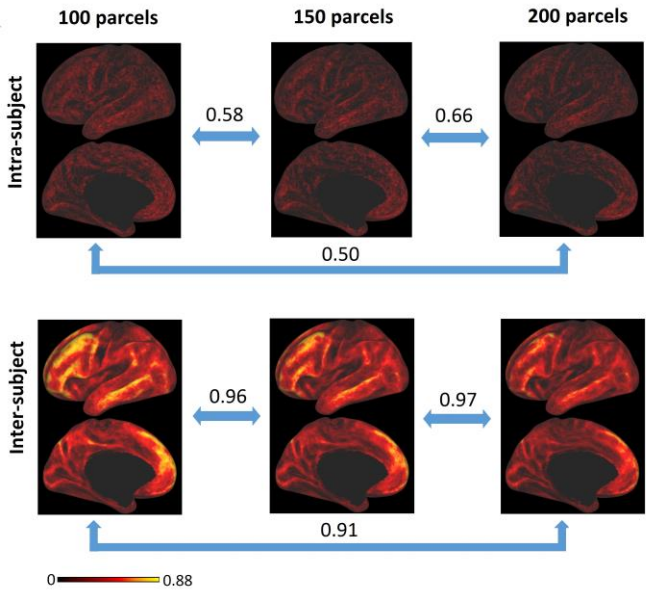

Fig. 7. Variation maps with different numbers of parcels. The maximal value in these maps is 0.88.

the number of parcels, which is an important user-specified parameter in the joint K-means algorithm.

F. Bootstrapped λ

The regularization parameter λ is another important parameter in the joint K-means algorithm. Unlike the number of parcels that has to be specified by users, a bootstrap procedure was designed to automatically find an appropriate value of the regularization parameter λ.

The number of bootstrap variations Z was set to 0.1% of the total number of the vertices (297 for each hemisphere). To investigate sensitivity of the bootstrapped regularization parameter λ to the number of bootstrap variations Z, two additional values for Z were used to generate bootstrapped λ for the left hemisphere in group A. The two values were 0.05% and 0.15% of the total number of the vertices (149 and 446, respectively, for the left hemisphere). The mean values of the bootstrapped λ were 0.00915, 0.0110 and 0.0139 for 149, 297 and 446, respectively.

There were 20 bootstrapped values for each session of a subject; and the values for each hemisphere were bootstrapped independently. Thus, there were 3040 ($= 20 \times 38 \times 2 \times 2$) values in total. The distribution of these values was an approximate Gaussian distribution with 0.0082 mean value and 0.0016 standard deviation (Fig. 8). As described in the methods section, the 95th percentile of the bootstrapped values for each cortex of each subject was used. The mean and standard deviation of these 95th percentiles were 0.011 and 0.0012 respectively. The low standard deviation suggests that it may be appropriate to fix a value of λ for all subjects in some cases. Thus, the total number of resampling can be reduced.

IV. Conclusion

Here we have proposed the joint K-means algorithm to robustly parcellate the cerebral cortex based on resting-state fMRI data from an individual subject studied across two sessions or between two subjects. A Matlab implementation of the proposed method is publicly available at https://github.com/LNie/JKMeans.

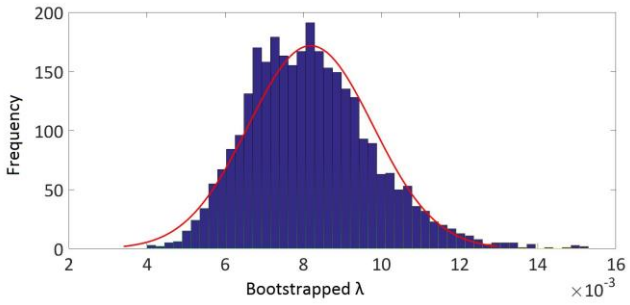

Fig. 8. Distribution of the bootstrapped values of the regularization parameter λ

Parcellations of resting-state fMRI datasets from the Human Connectome Project demonstrated that the inferred individual-level variations are plausible. The Dice and Jaccard coefficients of contrasted intra-subject parcellations were higher than those of contrasted inter-subject parcellations, suggesting that the proposed method has both a good test-retest reliability and sufficient sensitivity to detect between subjects differences. Inter-subject variations were distributed non-uniformly across the cerebral cortex and similar to inter-subject connectivity variability, with apparent clusters of higher variance in cortical regions associated with differences in cognitive performance between individuals. This preliminary observation deserves follow up study.

Acknowledgments

P. M. Matthews also thanks the Edmond Safra Foundation and Lily Safra for personal support and the Imperial College Healthcare Trust Biomedical Research Center for infrastructure support. Imperial College has received honoraria or consultancy fees for work by PMM conducted for GSK, IXICO, Transparency Life Sciences, GSK, Novartis and Biogen. Data were 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

[1] K. Brodmann. (2006). *Brodmann's: Localisation in the cerebral cortex*. Available: http://www.springer.com/us/book/9780387269177

[2] A. W. Toga *et al.* (2006, Dec.). Towards multimodal atlases of the human brain. *Nature Reviews Neuroscience*. [Online]. 7(12), pp. 952-966. Available: http://www.nature.com/nrn/journal/v7/n12/abs/nrn2012.html

[3] M. Rubinov and O. Sporns. (2010, Sep.). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*. [Online]. 52(3), pp. 1059-1069. Available: http://www.sciencedirect.com/science/article/pii/S105381190901074X

[4] K. J. Friston. (2011, Apr.). Functional and effective connectivity: a review. *Brain Connectivity*. [Online]. 1(1), pp. 13-36. Available: http://online.liebertpub.com/doi/abs/10.1089/brain.2011.0008

[5] R. C. Craddock, S. Jbabdi and *et al*. (2013, Apr.). Imaging human connectomes at the macroscale. *Nature Methods*. [Online]. *10(6)*, pp.

524-539. Available: http://www.nature.com/nmeth/journal/v10/n6/abs/nmeth.2482.html
[6] N. B. Turk-Browne. (2013, Nov.). Functional interactions as big data in the human brain. *Science*. [Online]. *342(6158)*, pp. 580-584. Available: http://www.sciencemag.org/content/342/6158/580.short
[7] M. A. Frost, and Rainer Goebel. (2012, Jan.). Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage*. [Online]. 59(2), pp. 1369-1381. Available: http://www.sciencedirect.com/science/article/pii/S1053811911009281
[8] E. Gordon *et al.* (2015, Oct.). Individual variability of the system-level organization of the human brain. *Cerebral Cortex*. [Online]. bhv239. Available: http://cercor.oxfordjournals.org/content/early/2015/10/12/cercor.bhv239.short
[9] D. Wang *et al*. (2015, Nov.). Parcellating cortical functional networks in individuals. *Nature Neuroscience*. [Online]. 18(12), pp. 1853–1860. Available: http://www.nature.com/neuro/journal/v18/n12/full/nn.4164.html
[10] M. Stanley *et al.* (2013, Nov.). Defining nodes in complex brain networks. *Frontiers in Computational Neuroscience*. [Online]. 7, pp.169. Available: http://journal.frontiersin.org/article/10.3389/fncom.2013.00169/full
[11] S. M. Smith *et al*. (2011, Jan.). Network modelling methods for fMRI. *Neuroimage*. [Online]. *54(2)*, pp. 875-891. Available: http://www.sciencedirect.com/science/article/pii/S1053811910011602
[12] S. W. Sohn *et al*. (2015, Aug.). Influence of ROI selection on resting state functional connectivity: an individualized approach for resting state fMRI analysis. *Frontiers in Neuroscience*. [Online]. 9, pp. 280. Available: http://journal.frontiersin.org/article/10.3389/fnins.2015.00280/abstract
[13] S. M. Smith *et al.* (2015, Nov.). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*. [Online]. 18(11), pp. 1565-1567. Available: http://www.nature.com/neuro/journal/v18/n11/pdf/nn.4125.pdf
[14] E. Finn *et al.* (2015, Nov.). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*. [Online]. 18(11), pp. 1664-1671. Available: http://www.nature.com/neuro/journal/v18/n11/full/nn.4135.html
[15] R. P. Monti, C. Anagnostopoulos, and G. Montana. (2015, Dec.). Learning population and subject-specific brain connectivity networks via Mixed Neighborhood Selection. *arXiv preprint*. [Online]. arXiv:1512.01947. Available: http://arxiv.org/abs/1512.01947
[16] S. M. Smith. (2016, Jan.). Linking cognition to brain connectivity. *Nature Neuroscience*. [Online]. 19(1), pp. 7-9. Available: http://www.nature.com/neuro/journal/v19/n1/full/nn.4206.html
[17] M. Rosenberg *et al*. (2016, Jan.). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*. [Online]. 19(1), pp. 165-171. Available: http://www.nature.com/neuro/journal/v19/n1/full/nn.4179.html
[18] R. A. Poldrack and M. J. Farah. (2015, Oct.). Progress and challenges in probing the human brain. *Nature*. [Online]. 526, pp. 371-379. Available: http://www.nature.com/nature/journal/v526/n7573/abs/nature15692.html
[19] A. L. Cohen *et al*. (2008, May). Defining functional areas in individual human brains using resting functional connectivity MRI. *Neuroimage*. [Online]. 41(1), pp. 45-57. Available: http://www.sciencedirect.com/science/article/pii/S1053811908001171
[20] G. S. Wig, T. O. Laumann, and S. E. Petersen. (2014, Jun.). An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage*. [Online]. 93, pp. 276-291. Available: http://www.sciencedirect.com/science/article/pii/S105381191300791X
[21] G. S. Wig *et al.* (2014, Aug.). Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cerebral Cortex*. [Online]. 24 (8), pp. 2036-2054. Available: http://cercor.oxfordjournals.org/content/24/8/2036
[22] E. M. Gordon *et al.* (2016, Jan.). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex*. [Online]. 26(1), pp. 288–303. Available: http://cercor.oxfordjournals.org/content/26/1/288
[23] L. Nanetti *et al*. (2009, Oct.). Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. *Neuroimage*. [Online]. 47(4), pp. 1666-1677. Available: http://www.sciencedirect.com/science/article/pii/S1053811909006260
[24] P. Bellec *et al*. (2010, Jul.). Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *Neuroimage*. [Online]. 51(3), pp. 1126-1139. Available: http://www.sciencedirect.com/science/article/pii/S1053811910002697
[25] R. C. Craddock *et al*. (2012, Aug.). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*. [Online]. 33(8), pp. 1914-1928. Available: http://onlinelibrary.wiley.com/doi/10.1002/hbm.21333/full
[26] X. Shen *et al*. (2013, Nov.). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*. [Online]. 82, pp. 403-415. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005818
[27] T. Blumensath *et al*. (2013, Aug.). Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *Neuroimage*. [Online]. 76, pp. 313-324. Available: http://www.sciencedirect.com/science/article/pii/S1053811913002668
[28] B. T. T. Yeo *et al*. (2011, Sep.) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*. [Online]. 106(3), pp. 1125-1165. Available: http://jn.physiology.org/content/106/3/1125.short
[29] S. Ryali *et al*. (2013. Jan.). A parcellation scheme based on von Mises-Fisher distributions and Markov random fields for segmenting brain regions using resting-state fMRI. *NeuroImage*. [Online]. 65, pp. 83-96. Available: http://www.sciencedirect.com/science/article/pii/S1053811912009858
[30] C. Baldassano, D. M. Beck, and F. Li. (2015, Feb.). Parcellating connectivity in spatial maps. *PeerJ*. [Online]. pp. e784. Available: https://peerj.com/articles/784/
[31] N. Honnorat *et al*. (2015, Feb.). GraSP: geodesic graph-based segmentation with shape priors for the functional parcellation of the cortex. *Neuroimage*. [Online]. 106, pp. 207-221. Available: http://www.sciencedirect.com/science/article/pii/S1053811914009100
[32] G. Varoquaux *et al*. (2010, May). A group model for stable multi-subject ICA on fMRI datasets. *Neuroimage*. [Online] 51(1), pp. 288-299. Available: http://www.sciencedirect.com/science/article/pii/S1053811910001618
[33] P. Dhillon *et al*. (2014, Oct.). Subject-specific functional parcellation via prior based eigenanatomy. *NeuroImage*. [Online]. 99, pp. 14-27. Available: http://www.sciencedirect.com/science/article/pii/S1053811914003917
[34] S. J. Harrison *et al*. (2015 Apr.). Large-scale probabilistic functional modes from resting state fMRI. *NeuroImage*. [Online]. 109, pp. 217-231. Available: http://www.sciencedirect.com/science/article/pii/S1053811915000208
[35] T. Hastie, R. Tibshirani and J. Friedman. (2009). The elements of statistical learning. (Second Edition) [Online]. Available: http://statweb.stanford.edu/~tibs/ElemStatLearn/
[36] T. O. Laumann *et al.* (2015, Aug.). Functional system and areal organization of a highly sampled individual human brain. *Neuron*. [Online]. 87(3), pp. 657-670. Available: http://www.sciencedirect.com/science/article/pii/S0896627315006005
[37] A. K. Jain. (2010, Jun.). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*. [Online]. *31(8)*, pp. 651-666. Available: http://www.sciencedirect.com/science/article/pii/S0167865509002323
[38] D. C. Van Essen *et al*. (2013, Oct.). The wu-minn human connectome project: an overview. *Neuroimage*. [Online]. *80*, pp. 62–79. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005351
[39] S. M. Smith *et al*. (2013, Oct.). Resting-state fMRI in the human connectome project. *NeuroImage*. [Online], 80, pp. 144-168. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005338
[40] S. Eickhoff *et al.* (2015, Dec.). Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*. [Online]. 36(12), pp. 4771-4792. Available: http://onlinelibrary.wiley.com/doi/10.1002/hbm.22933/full
[41] P. Drineas *et al.* (2004, Jul.). Clustering large graphs via the singular value decomposition. *Machine Learning*. [Online]. 56(1-3), pp. 9-33. Available: http://link.springer.com/article/10.1023/b:mach.0000033113.59016.96
[42] M. Sipser. (2006). *Introduction to the Theory of Computation*.
[43] H. Thomas *et al*. (2009). *Introduction to algorithms*. (3rd ed.).
[44] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc.* SODA, 2007, pp. 1027-1035. Available: http://dl.acm.org/citation.cfm?id=1283494
[45] S. Bubeck, M. Meilă, and U. von Luxburg. (2012, Jan.). How the initialization affects the stability of the k-means algorithm. *ESAIM: Probability and Statistics*. [Online]. 16, pp. 436-452. Available: http://journals.cambridge.org/abstract_S1292810012000134
[46] Z. Akata, C. Thurau, and C. Bauckhage. "Non-negative matrix factorization in multimodality data for segmentation and label prediction,"

in Proc. 16th Computer Vision Winter Workshop, 2011. Available: https://hal.inria.fr/hal-00652879/

[47] W. Härdle, J. Horowitz, and J. Kreiss. (2003, Aug.). Bootstrap methods for time series. *International Statistical Review*. [Online]. 71(2), pp. 435-459. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2003.tb00485.x/abstract

[48] P. Bellec, G. Marrelec, and H. Benali. (2008). A bootstrap test to investigate changes in brain connectivity for functional MRI. *Statistica Sinica*. [Online]. 18(4), pp. 1253. Available: http://www3.stat.sinica.edu.tw/sstest/oldpdf/A18n42.pdf

[49] R. Hutchison *et al.* (2013, Oct.). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*. [Online]. 80, pp. 360-378. Available: http://www.sciencedirect.com/science/article/pii/S105381191300579X

[50] B. Thirion *et al.* (2014, Jul.). Which fMRI clustering gives good brain parcellations?. *Frontiers in Neuroscience*. [Online]. 8, pp.167. Available: http://journal.frontiersin.org/article/10.3389/fnins.2014.00167/full

[51] J. H. J. Ward (1963, Apr.). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. [Online]. 58(301), pp. 236-244. Available: http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845

[52] M. Jenkinson *et al*. (2012, Aug.). FSL. *Neuroimage*. [Online]. *62(2)*, pp. 782–790. Available: http://www.sciencedirect.com/science/article/pii/S1053811911010603

[53] A. M. Dale, B. Fischl and M. I. Sereno (1999, Feb.). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*. [Online]. *9(2)*, pp. 179–194. Available: http://www.sciencedirect.com/science/article/pii/S1053811998903950

[54] D. S. Marcus *et al*. (2013, Oct.). Human Connectome Project informatics: Quality control, database services, and data visualization. *Neuroimage*. [Online]. *80*, pp. 202–219. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005776

[55] E. C. Robinson *et al.* (2014, Oct.). MSM: A new flexible framework for Multimodal Surface Matching. *Neuroimage*. [Online]. 100, pp. 414-426.Available: http://www.sciencedirect.com/science/article/pii/S1053811914004546

[56] S. M. Smith *et al.* (2013, Dec.). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*. [Online]. 17(12), pp. 666-682. Available: http://www.sciencedirect.com/science/article/pii/S1364661313002209

[57] L. Griffanti *et al*. (2014, Jul.). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage*. [Online]. *95*, pp. 232–247. Available: http://www.sciencedirect.com/science/article/pii/S1053811914001815

[58] G. Salimi-Khorshidi *et al*. (2014, Apr.). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*. [Online]. *90*, pp. 449–468. Available: http://www.sciencedirect.com/science/article/pii/S1053811913011956

[59] M. F. Glasser *et al*. (2013, Oct.). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*. [Online]. 80, pp. 105–124. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005053

[60] V. Essen *et al*. (2012, Oct.). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral Cortex*. [Online]. 22(10), pp. 2241-2262. Available: http://cercor.oxfordjournals.org/content/22/10/2241.short

[61] L. R. Dice. (1945, Jul.). Measures of the amount of ecologic association between species. *Ecology*. [Online]. 26(3), pp. 297-302. Available: http://www.jstor.org/stable/1932409

[62] P. Jaccard. (1912, Feb.). The distribution of the flora in the alpine zone. *New Phytologist*. [Online]. 11(2), pp. 37-50. Available: http://www.jstor.org/stable/2427226

[63] R. E. Jung, and R. J. Haier. (2007, Apr.). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences*. [Online]. 30(02), pp. 135-154. Available: http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=1305780&fileId=S0140525X07001185

[64] I. J. Deary, L. Penke, and W. Johnson. (2010, Mar.). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*. [Online]. 11(3), pp. 201-211. Available: http://www.nature.com/nrn/journal/v11/n3/abs/nrn2793.html

[65] S. Mueller *et al*. (2013, Feb.). Individual variability in functional connectivity architecture of the human brain. *Neuron*. [Online]. 77(3), pp. 586-595. Available: http://www.sciencedirect.com/science/article/pii/S0896627313000044

[66] B. Chen *et al*. (2015, Dec.). Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month. *PloS One*. [Online]. 10(12), pp. e0144963. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144963