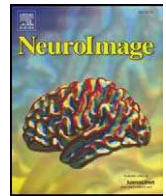




Contents lists available at ScienceDirect

NeuroImage

journal homepage: [www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg)

## Group analyses of connectivity-based cortical parcellation using repeated *k*-means clustering

Luca Nanetti<sup>a,b</sup>, Leonardo Cerliani<sup>a,b</sup>, Valeria Gazzola<sup>a,b</sup>, Remco Renken<sup>a,c</sup>, Christian Keysers<sup>a,b,\*</sup>

<sup>a</sup> BCN NeuroImaging Center, University of Groningen, The Netherlands

<sup>b</sup> Department of Neuroscience—Section Social Brain Lab, University Medical Center Groningen, The Netherlands

<sup>c</sup> Department of Neuroscience, University Medical Center Groningen, The Netherlands

### ARTICLE INFO

#### Article history:

Received 20 August 2008

Revised 7 May 2009

Accepted 3 June 2009

Available online xxx

#### Keywords:

Anatomical connectivity

*K*-means clustering

*K*-means

SMA–preSMA

Insula

### ABSTRACT

*K*-means clustering has become a popular tool for connectivity-based cortical segmentation using Diffusion Weighted Imaging (DWI) data. A sometimes ignored issue is, however, that the output of the algorithm depends on the initial placement of starting points, and that different sets of starting points therefore could lead to different solutions. In this study we explore this issue. We apply *k*-means clustering a thousand times to the same DWI dataset collected in 10 individuals to segment two brain regions: the SMA–preSMA on the medial wall, and the insula. At the level of single subjects, we found that in both brain regions, repeatedly applying *k*-means indeed often leads to a variety of rather different cortical based parcellations. By assessing the similarity and frequency of these different solutions, we show that ~256 *k*-means repetitions are needed to accurately estimate the distribution of possible solutions. Using nonparametric group statistics, we then propose a method to employ the variability of clustering solutions to assess the reliability with which certain voxels can be attributed to a particular cluster. In addition, we show that the proportion of voxels that can be attributed significantly to either cluster in the SMA and preSMA is relatively higher than in the insula and discuss how this difference may relate to differences in the anatomy of these regions.

© 2009 Elsevier Inc. All rights reserved.

### Introduction

DW-MRI (Diffusion Weighted-Magnetic Resonance Imaging) infers information about white matter structure in the brain from the differential attenuation of the spin echo signal, as modulated by the local spatial microstructure of the surrounding medium, and by the strength and direction of the applied magnetic diffusion gradient (Basser et al., 1994; Pierpaoli et al., 1996).

Using this method in conjunction with probabilistic tractography, one can estimate, for each individual voxel of the brain (seed) whether it is connected or not with all other voxels of the brain (target). This information is called the binarized *tractogram* (also known as binarized connectivity profile) of that voxel (Behrens et al., 2003; Hosey et al., 2005).

Johansen-Berg (2004) first illustrated how this information can be used to divide the medial motor wall in two subregions, which on the basis of their location and functional properties were likely to represent the supplementary and presupplementary motor area—SMA and preSMA. They used probabilistic tractography for all voxels within the SMA–preSMA complex to define the corresponding tractograms, took the correlation between the tractograms of each

pair of voxels in the SMA–preSMA as a measure of their similarity and calculated the full cross-correlation matrix (cc-matrix) of all tractograms. They then reordered the cc-matrix using a spectral reordering algorithm (Higham et al., 2007). Eye-balling then revealed a sudden discontinuity in the reordered matrix. Remapping the location of the voxels on either side of this discontinuity, they found that they fell within putative SMA and preSMA, respectively.

This method attracted much attention, because it is plausible that if two subregions have different connectivities, they may also have different functions. Requiring an experimenter to decide where to place the border between the regions however is undesirable. To circumvent this caveat, researchers turned towards *k*-means clustering (Hartigan, 1975; Hartigan and Wong, 1979) to decide where to place the border between clusters. *K*-means clustering is an iterative algorithm, which, for the case of connectivity profiles, ultimately divides the voxels of a seed region into *k* non overlapping clusters of voxels (the experimenter needs to decide the value of *k* based on functional and anatomical considerations). This is done in a hyper-space with as many dimensions as there are voxels in the seed region. Each seed voxel is represented as a point whose coordinates are the correlation of its tractogram with all the other voxels' tractograms. Different strategies exist to choose the initial putative centroids for the algorithm. Hartigan and Wong (1979) propose to randomly choose *k* points (i.e. voxels in our case) from the initial set, the coordinates of which become the centers of the *k* clusters. A frequently used

\* Corresponding author. Antonius Deusinglaan2, 9713AW Groningen, The Netherlands.  
Fax: +315036387500

E-mail address: [c.m.keysers@rug.nl](mailto:c.m.keysers@rug.nl) (C. Keysers).

alternative is to split the original set in  $k$ -subsets of (close to) equal size, and use their respective baricenters as putative centers (Lloyd, 1982). The following two steps are then repeated: (a) each voxel is assigned to the candidate center that happened to fall nearest to it, thus partitioning the voxels in  $k$  clusters; and (b) the  $k$  candidate centers are shifted to the real baricenter of each cluster. The iterative process stops when the candidate baricenters' positions change by less than a certain, preset and very small threshold. Every voxel is then finally assigned to one, and only one, cluster. Several studies have now used this method: Tomassini et al. (2007) segmented the premotor cortex in its dorsal and ventral parts; Anwender et al. (2007; Tomassini et al., 2007) segmented Broca's area and Klein et al. (2007) segmented BA44 from 45 and the SMA from the preSMA.

Because the end-position of the baricenters depends on the initial starting points,  $k$ -means does not necessarily provide the same solution each time it is run—but only a local minimum in the space of possible solutions (Hartigan and Wong, 1979; Klein et al., 2007; Selim and Ismail, 1994). A toy example illustrates this point in Fig. 1. Four points arranged in a square pattern are the set to be partitioned in two clusters. Depending from where the algorithm initially and randomly sets the centers (hatched crosses), six solutions can be found. Two solutions correspond to a partition of the original set in two subsets of 2 elements each (the ones of Figs. 1b and c). Four solutions (Fig. 1d and obvious variants) group the points in one cluster of 3 elements and one cluster of 1.

This means that repeating the algorithm can lead to different connectivity-based parcellation, and a single repetition of the algorithm can lead to a solution that can be very far from optimal (Kanungo et al., 2004).

Current implementations of  $k$ -means (e.g. in Matlab and R) only partially address this problem: they give the user the choice between (a, the default) performing a single  $k$ -means, and therefore outputting a single segmentation, or (b) performing several attempts, but again outputting a single segmentation, namely the one for which the sum of the (squared) distances between the baricenters and their voxels is minimal. While less arbitrary than solution (a), solution (b) is still

problematic because there is no obvious reason to believe that solution (b) is *biologically* the most valid of all parcellations.

These considerations generate a number of important questions that have received little attention in the  $k$ -means and DWI literature: how different are the various  $k$ -means solutions? How are they distributed in terms of relative frequency? Do different brain areas differ in how many different parcellations are found? How does this variability of parcellation affect the confidence with which voxels can be attributed to either cluster?

The aim of the present report is to examine these issues. We applied the  $k$ -means clustering not once but 1000 times to the DWI derived tractogram-cross-correlation-matrices of 10 participants in order to segment two different regions—the SMA–preSMA and the insula. We then examine and compare the distribution of  $k$ -means segmentations found for the 10 individuals and 2 brain areas.

We chose to confront segmentation of the SMA–preSMA with a segmentation of the insula because, based on anatomical considerations, these two brain regions should differ in terms of the challenges they pose to  $k$ -means clustering. The connectivity of SMA–preSMA has been shown by the previously cited studies—first and foremost Johansen-Berg (2004) to be well suited for clustering because the SMA and preSMA have remarkably distinct connectivity profiles (Behrens and Johansen-Berg, 2005), and the change of connectivity is relatively abrupt and occurs around  $y = 0$  in the standard space of the Montreal Neurological Institute (MNI).

Such a clear subdivision is unlikely to be found in the insula, as suggested by the blurred cytoarchitectonic boundary between the anteroventral agranular and postero-dorsal granular territories (Friedman et al., 1986; Jones and Burton, 1976; Mesulam and Mufson, 1985) and by the correspondent continuous variation in the anatomical connectivity patterns of the macaque (Amaral and Price, 1984; Chikama et al., 1997; Friedman et al., 1986; Fudge et al., 2005).

## Materials and methods

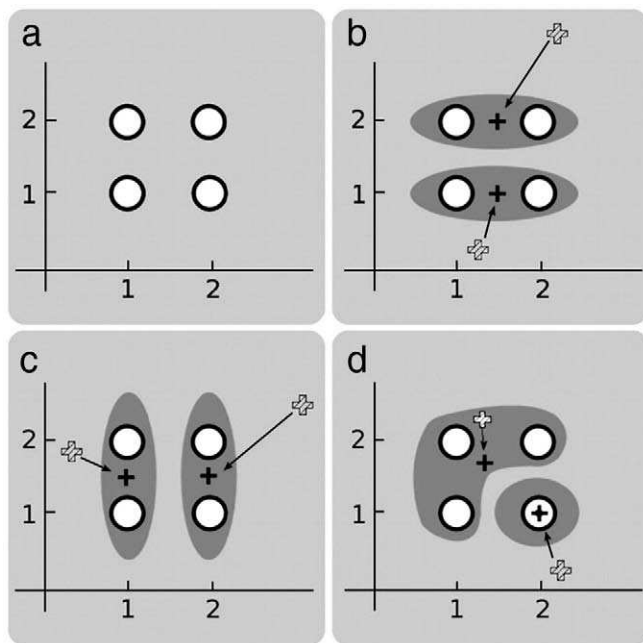
### Subjects

Ten healthy subjects, males, age  $31.8 \pm 6.9$ .

### DWI: acquisition, preprocessing

Diffusion weighted (DW) data from ten healthy subjects were acquired using a single-shot pulsed gradient spin echo EPI sequence (SENSE factor = 2, TR = 7463.57 ms, and TE = 67 ms) on a 3 T MR scanner (Intera, Philips Medical Systems, Best, The Netherlands). For each subject 15 DW images were collected using the 'medium' gradient scheme provided by the manufacturer, with a maximum gradient strength of 22 mT/m and a  $b$ -value of  $1000 \text{ s mm}^{-2}$ . In addition, one non-DW volume ( $b \approx 0 \text{ s mm}^{-2}$ , referred in the following as  $b_0$  image) was acquired. Each volume consisted of 76 transverse slices, slice thickness 2 mm, scanned with a  $96 \times 96$  matrix size (FOV =  $224 \times 224 \text{ mm}$ , voxel size =  $2.3 \times 2.3 \text{ mm}$ , slice thickness 2 mm, reconstructed matrix size  $128 \times 128 \text{ mm}$ , NEX = 2). In addition, one high resolution (1 mm isotropic) 3D anatomical scan (T1W, TFE CLEAR, TR = 7.6 s, TE = 3.6 s and flip angle = 8°) was acquired. Signal-to-noise ratio (SNR) in the  $b_0$  image was calculated in several white matter regions and resulted in a mean value of 24.79 dB (s.d. 0.76 dB). More details about SNR estimate in the [Supplementary methods and results](#), "Measurements of the signal-to-noise ratio (SNR): details".

The analysis of DW data was performed in FSL 3.3.11 ([www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)). After motion and eddy current correction, a Bayesian estimation of the diffusion parameters was calculated for each voxel using a simple partial volume model as described in Behrens et al. (Johansen-Berg (2003)). This made possible to perform the actual probabilistic tracking from each voxel in the seed regions (either the left insular cortex or the left SMA–preSMA) to every voxel in the



**Fig. 1.** An example of  $k$ -means' multiple solutions on clustering four points arranged in a square. Different starting points can lead to different clusters. Hatched crosses represent the centers' initial random positions. Solid black crosses are the final baricenters. Clusters are represented in dark gray. (a) is the set of points. (b), (c), and (d) are different solutions. Three more do exist, similar to (d).

ipsilateral hemisphere (target voxels). The result of the tracking is a series of maps, one for each (seed) voxel of the seed region under investigation; they contain for each (target) voxel a numerical information about the connectivity between the two voxels. Following the procedure described in [Johansen-Berg \(2004\)](#), those connectivity maps (also known as ‘connectivity profiles’) were binarized. The cross-correlation matrix (cc-matrix) of binarized connectivity profiles was thereafter computed.

#### Seed brain areas

The SMA–preSMA was drawn on the MNI template following the criteria provided in [Johansen-Berg \(2004\)](#): a sagittal slice at  $X_{\text{MNI}} = -2$ ,  $-22 < Y_{\text{MNI}} < 30$  in the antero-posterior direction, and along the  $Z_{\text{MNI}}$  axis from a small distance above the cingulate sulcus to the dorsal surface of the brain. This ROI is then converted onto each single subject's diffusion space using FSL's FLIRT ([Jenkinson et al., 2002; Jenkinson and Smith, 2001](#)). All single subject computations were performed in this space. The final  $k$ -means maps were then transformed back into the MNI space using FLIRT to enable group analysis.

For the insulae, we registered the T1 weighted image of each individual to the  $b_0$  diffusion weighted image of that individual. The insula was then drawn subject by subject on this T1 image following the indications of [Naidich et al. \(2004\)](#) and [Türe et al. \(1999\)](#). We thus considered the anterior, superior and inferior periinsular sulci as respective limits of the frontoorbital, frontoparietal, and temporal opercula, whilst the ideal rostro-ventral boundary between insular and orbitofrontal cortex was defined as the plane connecting the insular pole with the anterior limit of the inferior periinsular sulcus. Particular care has been used to ensure that no voxels from opercular areas were included in the tracing. At the end of the single subject computations in the participant's own diffusion space, the  $k$ -means summary maps were normalized onto the MNI space using FLIRT to enable group analyses.

#### Calculating the cross-correlation matrix

In accordance with the procedure described in [Johansen-Berg \(2004\)](#), tractograms corresponding to all voxels in both seed regions were calculated, binarized, and their cross-correlation matrix was calculated. The time necessary to calculate the cc-matrix is ~40 min. We used in-house software written in C which implements the correlation from Numerical Recipes ([Press et al., 1992](#)) compiled within the framework of the GNU C Compiler v 4.x and running on standard hardware/software (Pentium IV 2.4 GHz, 1 GB Ram, Linux kernel series 2.6).

#### The $k$ -means clustering algorithm

We employed the implementation of  $k$ -means in the stats package of R ([R Development Core Team, 2008](#)), which uses the algorithm of [Hartigan and Wong \(1979\)](#). The basic principles have been explained briefly in the [Introduction](#). We set  $k = 2$  to differentiate the SMA from the preSMA and the anterior insula from the posterior insula. Exploring the impact of using  $k = 3$  for these regions would be interesting, but goes beyond the scope of this particular report. The cross-correlation matrix of the SMA–preSMA was then fed 1000 times into the  $k$ -means routine, asking for 2 clusters. Each time two new random centers were automatically chosen, and a (potentially different) output was produced and saved. The same was done for the insula. Calculation time for 1000  $k$ -means was in the order of 30 min per region of interest and subject. Note that the meaning of “multiple repetitions of  $k$ -means” in this paper differs fundamentally from the concept of *iterations* of  $k$ -means used by previous authors (e.g. [Klein et al., 2007](#), 200 iterations; Tomassini, personal communication, 500 iterations). The latter referred to the number of

iterations within a single application of the intrinsically iterative  $k$ -means algorithm (personal communication, Valentina Tomassini), while we mean entirely independent applications of the algorithm with its multiple iterations. In these studies Lloyd's algorithm ([Lloyd, 1982](#)) was applied. In order to check whether the different strategies would affect either the number of different solutions found, or their relative frequency, or both (i.e., the overall stability of the process), we performed the same 1000 calculations for the insula employing Lloyd's algorithm: basically, dividing the initial dataset randomly in two, and deriving the initial centroids from these two subsets.

#### Identifying equivalent solutions

The output of the  $k$ -means clustering is a vector of labels, specifying for each voxel whether it belongs to one or the other cluster by assigning it the values  $-1$  or  $1$ . In order to examine how many of the 1000 repetitions of the  $k$ -means lead to the same solution it is important to consider that if two solutions simply have switched labels (i.e. all voxels with  $-1$  in the first solution have  $1$  in the second solution, and all the ones that have  $1$  in the first, have  $-1$  in the second), they are of course biologically equivalent. This means that if the correlation between the vector of labels of two solutions is  $1$  or  $-1$ , the solutions are equal, otherwise, they are different.

For each set of 1000  $k$ -means repetitions, we therefore identified how many different solutions exist. To ensure a meaningful comparison between solutions, we used the following automatic procedure to ensure a labeling consistent across solutions.

- i) the computed tractographies of all voxels belonging to a cluster are averaged, for all solutions and all subjects separately. These averages are arbitrarily labeled ‘a’ and ‘b’.
- ii) the averages are normalized into the standard Montreal Neurological Institute space, as for the template provided by FSL 4.1 (T1 MNI152, 2 mm)
- iii) the averages in MNI space are masked to the left hemisphere.
- iv) the cross-correlation matrix of (iii) is calculated. These matrices are reproduced in the [Supplementary methods and results, Tables S1 and S2](#).
- v) each cluster is checked against all clusters of all solutions, in order to find for each solution which cluster (a or b) better correlates with it. The solution with the highest combined (between clusters a and b) sum of correlations is chosen as the reference, i.e. the labeling of that solution becomes the labeling of choice. A pictorial representation of these sums is shown in the [Supplementary methods and results \(Fig. S1\)](#), separately for SMA–preSMA and insula.
- vi) the labeling of each solution is then kept or swapped following the result of the cross correlation between the tested clusters and the ones belonging to the reference solution.

All further analyses (e.g. correlations between solutions, number of distinct solutions) were done using these labelings.

#### Single subject summary map

For each individual subject, the results of the 1000  $k$ -means (after re-labeling see above) were summarized by averaging all 1000 solutions. A voxel always associated with the anterior cluster then received a value of  $1$ , a voxel always associated with the posterior cluster a value of  $-1$ . Values between  $1$  and  $-1$  represent voxels that were not always attributed to the same cluster. A value of  $0$  indicates that a voxel was attributed to the anterior and posterior cluster equally often.

#### Second level, group analysis

The single subject summary maps were used to examine how reliably a voxel can be attributed to the anterior or posterior cluster at

the level of the population. The rationale of this analysis is that under the null hypothesis ( $H_0$ : there is no spatial consistency in clustering), a voxel should on average have a summary value of 0, i.e. it is attributed to the posterior and anterior cluster with equal likelihood. If this null hypothesis can be rejected, over the population of participants, this voxel is more likely to be classified as belonging to one or the other of the two clusters. To test this hypothesis, we used SnPM (Nichols and Holmes, 2002) because the voxels' values were not normally distributed.

Specifically, the 10 individual, spatially normalized summary maps were fed into SnPM ver. 5b separately for the two ROIs. To test whether the average summary value deviates from zero in each voxel we used a family wise error corrected threshold of  $p < 0.05$  for both regions separately. This identified regions of the SMA–preSMA and insula that can be attributed to one of the clusters more often than expected by chance. In addition, the number of voxels for which the null hypothesis was not rejected was compared across regions using a chi-square test (for independence) with Yates correction. Using family wise error correction is the best way to avoid false positives in the group analysis, however it means applying a more stringent voxel-wise threshold on larger regions, which would put the larger insula at a systematic disadvantage. To compare the number of classified voxels between regions, we therefore used a voxel-wise, uncorrected threshold of  $p < 0.002$ , which corresponded to  $p < 0.05$  F.W.E. corrected for the larger insula.

## Results and discussion

This study provides two main results. First, repeatedly applying  $k$ -means clustering to the same DWI dataset *can indeed* lead to multiple,

non-identical parcellations. Second, the distribution of these solutions can be used to infer statistically which voxels can (and which cannot) reliably be attributed to a cluster at the level of the population.

### Multiple $k$ -means solutions

The  $k$ -means algorithm was applied 1000 times on the cc-tractogram matrix of each subject, for both the ROIs (SMA–preSMA and insula). This led to 1000 outputs per ROI. Only for one of the ten participants, all 1000 solutions were identical in both ROIs (Subject 8). For 9 out of 10 participants, a number of non-identical solutions were found for at least one of the regions of interest. Tables 1a (for the insula) and 1b (for the SMA–preSMA) show the distribution of these solutions using two different implementations of  $k$ -means. On average, any solution accounted for less than 75% of the possible solutions (i.e. 25% of the solutions were different) and as many as 6 different solutions can be found for a given participant and ROI. Interestingly, the most frequent solution according to either of the  $k$ -means implementations is not always the solution with the minimum observed within-cluster Sum of Squared Distances (min-SSD). Given that the min-SSD solution is determined from the set of computed  $k$ -means solutions, this finding illustrates how computing multiple  $k$ -means reduces the risk of missing the global min-SSD.

Figs. 2 and 3 illustrate the variability of these solutions for two exemplary participants and Fig. S1 for all others.

For Subject 5 (Fig. 2), the SMA–preSMA was always divided in the same way, with a division around  $y = 0$  in MNI space. The insula however was divided in a number of different ways. The most frequent solution, accounting for 54.3% of the solutions divides the insula in an anterior and a posterior cluster, with the former

**Table 1a**  
Individual solutions for the insula.

| Subject | Solution | HW (# of solutions) | $r(\text{HW}, M_{1000})$ | Lloyd (# of solutions) | $r(\text{L}, M_{1000})$ | SSD      | $r(\text{sol}, \text{min-SSD})$ |
|---------|----------|---------------------|--------------------------|------------------------|-------------------------|----------|---------------------------------|
| 1       | 1        | 293                 | 0.969                    | 401                    | 0.856                   | 44524    | 1                               |
|         | 2        | 402                 | 0.976                    | 435                    | 0.855                   | 44530    | 0.966                           |
|         | 3        | 187                 | 0.460                    | 58                     | 0.738                   | 47259    | 0.301                           |
|         | 4        | 117                 | 0.354                    | 104                    | 0.161                   | 47587    | 0.351                           |
|         | 5        | 1                   | 0.365                    | 2                      | 0.173                   | 47588    | 0.363                           |
| 2       | 1        | 289                 | 0.999                    | 164                    | 0.991                   | 23893    | 1                               |
|         | 2        | 711                 | 0.994                    | 835                    | 0.999                   | 23894    | 0.988                           |
|         | 3        | 0                   | 0                        | 1                      | 0.343                   | 26882    | 0.351                           |
| 3       | 1        | 139                 | 0.970                    | 307                    | 0.971                   | 43736.4  | 1                               |
|         | 2        | 397                 | 0.970                    | 219                    | 0.971                   | 43736.5  | 1                               |
|         | 3        | 440                 | 0.957                    | 390                    | 0.955                   | 43769    | 0.857                           |
|         | 4        | 9                   | 0.961                    | 28                     | 0.964                   | 43783.21 | 0.897                           |
|         | 5        | 13                  | 0.962                    | 50                     | 0.965                   | 43783.23 | 0.900                           |
|         | 6        | 2                   | 0.962                    | 6                      | 0.965                   | 43783.3  | 0.900                           |
| 4       | 1        | 996                 | 0.999                    | 957                    | 0.999                   | 30147    | 1                               |
|         | 2        | 4                   | 0.202                    | 20                     | 0.227                   | 32601.7  | 0.198                           |
|         | 3        | 0                   | 0                        | 7                      | 0.1843                  | 32601.9  | 0.157245                        |
|         | 4        | 0                   | 0                        | 3                      | 0.1841                  | 32602.3  | 0.157241                        |
|         | 5        | 0                   | 0                        | 2                      | 0.179                   | 32602.6  | 0.152                           |
|         | 6        | 0                   | 0                        | 11                     | 0.142                   | 32615    | 0.116                           |
| 5       | 1        | 543                 | 0.971                    | 587                    | 0.941                   | 32493    | 1                               |
|         | 2        | 238                 | 0.932                    | 184                    | 0.880                   | 32761    | 0.935                           |
|         | 3        | 100                 | 0.573                    | 128                    | 0.669                   | 33825.9  | 0.353                           |
|         | 4        | 117                 | 0.559                    | 99                     | 0.656                   | 33826    | 0.368                           |
|         | 5        | 2                   | 0.774                    | 2                      | 0.807                   | 34020    | 0.693                           |
| 6       | 1        | 834                 | 0.994                    | 980                    | 0.999                   | 31068    | 1                               |
|         | 2        | 166                 | 0.843                    | 20                     | 0.788                   | 31754    | 0.781                           |
| 7       | 1        | 413                 | 0.999                    | 396                    | 0.993                   | 24978.09 | 1                               |
|         | 2        | 112                 | 0.996                    | 187                    | 0.995                   | 24978.13 | 0.996                           |
|         | 3        | 475                 | 0.973                    | 417                    | 0.991                   | 24980    | 0.972                           |
| 8       | 1        | 1000                | 1                        | 1000                   | 1                       | 35456    | 1                               |
| 9       | 1        | 1000                | 1                        | 1000                   | 1                       | 30400    | 1                               |
| 10      | 1        | 562                 | 0.999                    | 560                    | 0.999                   | 24724.2  | 1                               |
|         | 2        | 438                 | 0.998                    | 440                    | 0.996                   | 24724.9  | 0.995                           |

Column HW: frequency of a solution over 1000 repetitions of the  $k$ -means, obtained applying the Hartigan–Wong algorithm. Solutions are ordered following the Sum of Squared Distances value (SSD column). Column  $r(\text{HW}, M_{1000})$ : correlation between that solution and the weighted sum of all solutions found with the HW algorithm. The next two columns display the same information for the Lloyd algorithm. The last column contains the correlation of each solution with the one characterized by the smallest SSD.



**Table 1b**

Individual solutions for the SMA–preSMA.

| Subject | Solution | HW(# of solutions) | $r(\text{HW}, M_{1000})$ | Lloyd (#of solutions) | SSD      | $r(\text{sol}, \text{min-SSD})$ |
|---------|----------|--------------------|--------------------------|-----------------------|----------|---------------------------------|
| 1       | 1        | 706                | 0.999                    | 762                   | 15287    | 1                               |
|         | 2        | 294                | 0.998                    | 238                   | 15288    | 0.982                           |
| 2       | 1        | 528                | 0.967                    | 564                   | 11083.92 | 1                               |
|         | 2        | 184                | 0.964                    | 110                   | 11083.97 | 0.994                           |
|         | 3        | 5                  | 0.782                    | 8                     | 11663.76 | 0.601                           |
|         | 4        | 4                  | 0.791                    | 8                     | 11663.78 | 0.619                           |
|         | 5        | 278                | 0.791                    | 310                   | 11663.81 | 0.619                           |
| 3       | 1        | 455                | 0.924                    | 465                   | 13222    | 1                               |
|         | 2        | 302                | 0.338                    | 330                   | 13347    | 0.263                           |
|         | 3        | 240                | 0.468                    | 200                   | 13351    | 0.251                           |
|         | 4        | 3                  | 0.637                    | 4                     | 13807    | 0.644                           |
| 4       | 1        | 1000               | 1                        | 1000                  | 11164    | 1                               |
| 5       | 1        | 1000               | 1                        | 1000                  | 12135    | 1                               |
| 6       | 1        | 678                | 0.978                    | 647                   | 12978    | 1                               |
|         | 2        | 108                | 0.599                    | 141                   | 14041    | 0.448                           |
|         | 3        | 10                 | 0.670                    | 13                    | 14043    | 0.642                           |
|         | 4        | 3                  | 0.665                    | 2                     | 14044    | 0.637                           |
|         | 5        | 201                | 0.637                    | 197                   | 14330    | 0.612                           |
| 7       | 1        | 1000               | 1                        | 996                   | 6084     | 1                               |
|         | 2        | 0                  | 0                        | 4                     | 6839     |                                 |
| 8       | 1        | 1000               | 1                        | 1000                  | 10459    | 1                               |
| 9       | 1        | 406                | 0.713                    | 554                   | 10181    | 1                               |
|         | 2        | 105                | 0.896                    | 60                    | 10189    | 0.427                           |
|         | 3        | 307                | 0.950                    | 316                   | 11043    | 0.494                           |
|         | 4        | 182                | 0.126                    | 70                    | 11371    | 0.262                           |
| 10      | 1        | 1000               | 1                        | 1000                  | 11572    | 1                               |

Column HW: frequency of a solution over 1000 repetitions of the  $k$ -means, obtained applying the Hartigan-Wong algorithm. Solutions are ordered following the Sum of Squared Distances value (SSD column). Column  $r(\text{HW}, M_{1000})$ : correlation between that solution and the weighted sum of all solutions found with the HW algorithm. The next column contains the relative frequency of the solutions obtained using the Lloyd algorithm. The last column contains the correlation of each solution with the one characterized by the smallest SSD.

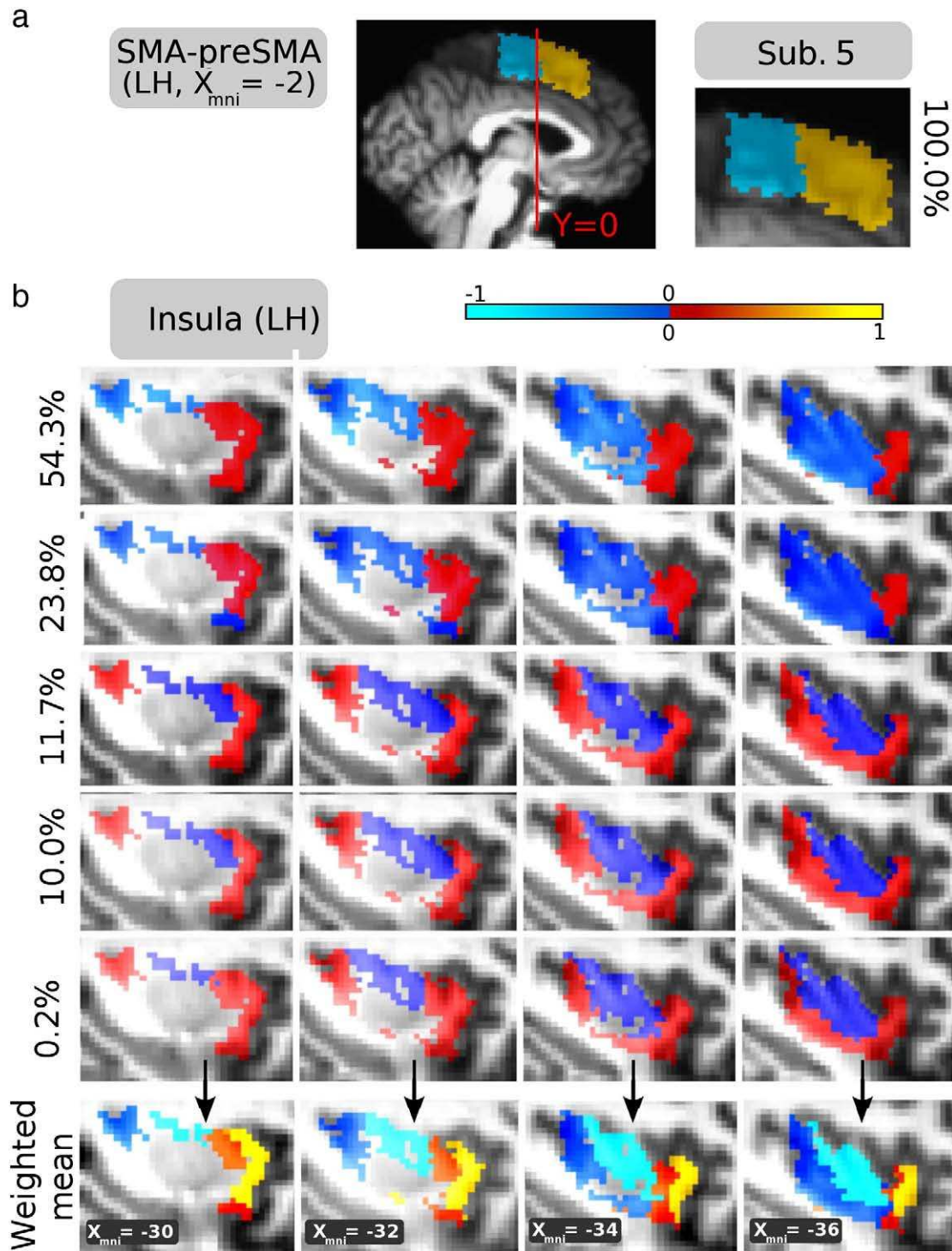
corresponding approximately to the anterior short gyrus. The second most frequent solution associates the anterior-ventral insula to the posterior cluster. The last three solutions cluster the anterior short and posterior long gyri of the insula together. This last type of parcellation may have occurred because one of the two centers got caught in a local minima within the wide-range connectivity of the dysgranular insula, coercing the anterior and posterior insulae into the single, remaining cluster (see the ‘Parabola’ example in the [Supplementary methods and results](#)). Averaging all 1000 solutions leads to a map dividing the insula in two compact clusters along the rostro-caudal axis.

For Subject 6 ([Fig. 3](#)), five distinct solutions were found for the SMA–preSMA. The second most frequent solution is peculiar, as it clusters the most posterior and most anterior parts of the SMA–preSMA together, in a way that resembles the three least frequent solutions for the insula of Subject 5. Such cases may suggest that a division in three ( $k=3$ ) instead of two clusters may lead to more stable solutions ([Anwander et al., 2007](#)). Since the aim of this report is to alert the reader to the existence of multiple solutions for a given, *a priori* chosen  $k$  value, we did not systematically explore the effect of varying  $k$ . However, to examine whether finding multiple  $k$ -means solutions is specific to a particular choice of  $k$ , we explored the effect of choosing values of  $k=3, 4, 5, 6, 7, 8, 9$ , and 10 for two example subjects we selected because at  $k=2$  they either showed no variability in solutions for either region of interest (Subject 8), or multiple solutions in both regions (Subject 3). We found that the data of Subject 8 (single solutions at  $k=2$ ) did generate a multitude of solutions at all  $k>2$  tested for both regions (number of repetitions=500). The data of Subject 3 (multiple solutions at  $k=2$ ) continued to generate multiple solutions for the insula at all tested  $k$  values (6 solutions at  $k=2$ , 23 at  $k=3$ , 51 at  $k=4$ , 25 at  $k=5$ , and 156 at  $k=10$ ). However, for the SMA–preSMA, the four solutions at  $k=2$  for Subject 3 turned into a single solution at  $k=3$ , but 18 at  $k=4$ , 33 at  $k=5$ , 42 at  $k=6$ . See [Table 2](#) for a summary of the solutions’ numerosities for these two ROIs and subjects. From these two example participant’s data, we can conclude (a) that the

variability in solutions *does* depend on the choice of  $k$  for any given participant and region, but (b) that variability depends on  $k$  in a not always monotonic fashion that varies from participant to participant and from brain region to brain region and (c) that for our pool of participants at least, there is no choice of  $k$  that abolishes variability for all participants. In other words, investigators should consider the risk of multiple  $k$ -means solutions independently of their choice of  $k$ .

For Subject 6, the average solution for the SMA–preSMA shows a relatively sharp transition between the posterior and anterior clusters, again around  $Y_{\text{MNI}}=0$ . For the insula, two solutions were found, that differ mainly in their attribution of the anterior-ventral insula, resulting in an average map that resembles that of Subject 5.

A more quantitative analysis of the variance in solutions can be performed by comparing individual solutions with the numerical average of all solutions ([Table 3](#) and [Fig. 4](#)). Let  $M_{1000}$ , the grand mean of 1000 outputs, be the estimate of the ‘true’ weighted mean of all possible solutions for a certain ROI and subject. Let also  $K_1$  be the solution found running the  $k$ -means only once. The ‘acceptability’ of an individual solution can then be defined using the  $r(M_{1000}, K_1)$ . Picking 100 solutions at random from the 1000 of each participant (with replacement) and correlating this solution with  $M_{1000}$  produced a distribution of 1000 values of correlation for each ROI (100 solutions per participant for 10 participants). Defining what an unacceptably bad solution is, is of course arbitrary, but if one considers a solution unacceptable if  $r<0.8$  (i.e. leaving more than 35% of the spatial variance in the  $M_{1000}$  unaccounted for), then according to the  $k$ -means implementation of [Hartigan and Wong \(1979\)](#) 6% of the insula and 16% of the SMA–preSMA clustering-just-once solutions are unacceptable ([Table 3](#) and [Figs. 4a](#) and [b](#)). These ratios goes up to 35% and 40%, respectively, if  $r<0.99$  is taken. But how ‘bad’ can a solution be? A way to assess this question is to look at the ‘worst’ 5% quantile in terms of correlation with the mean ( $M_{1000}$ ). This results in  $r=0.57$  for insula and  $r=0.47$  for SMA–preSMA ([Table 3](#) and [Figs. 4c](#) and [d](#)). In other words, accepting the result of running the  $k$ -means just once comes with a probability  $p=0.05$  to find a solution  $K_1$  which explains as little as 1/3 of the variance of  $M_{1000}$  for insula and 1/5 for



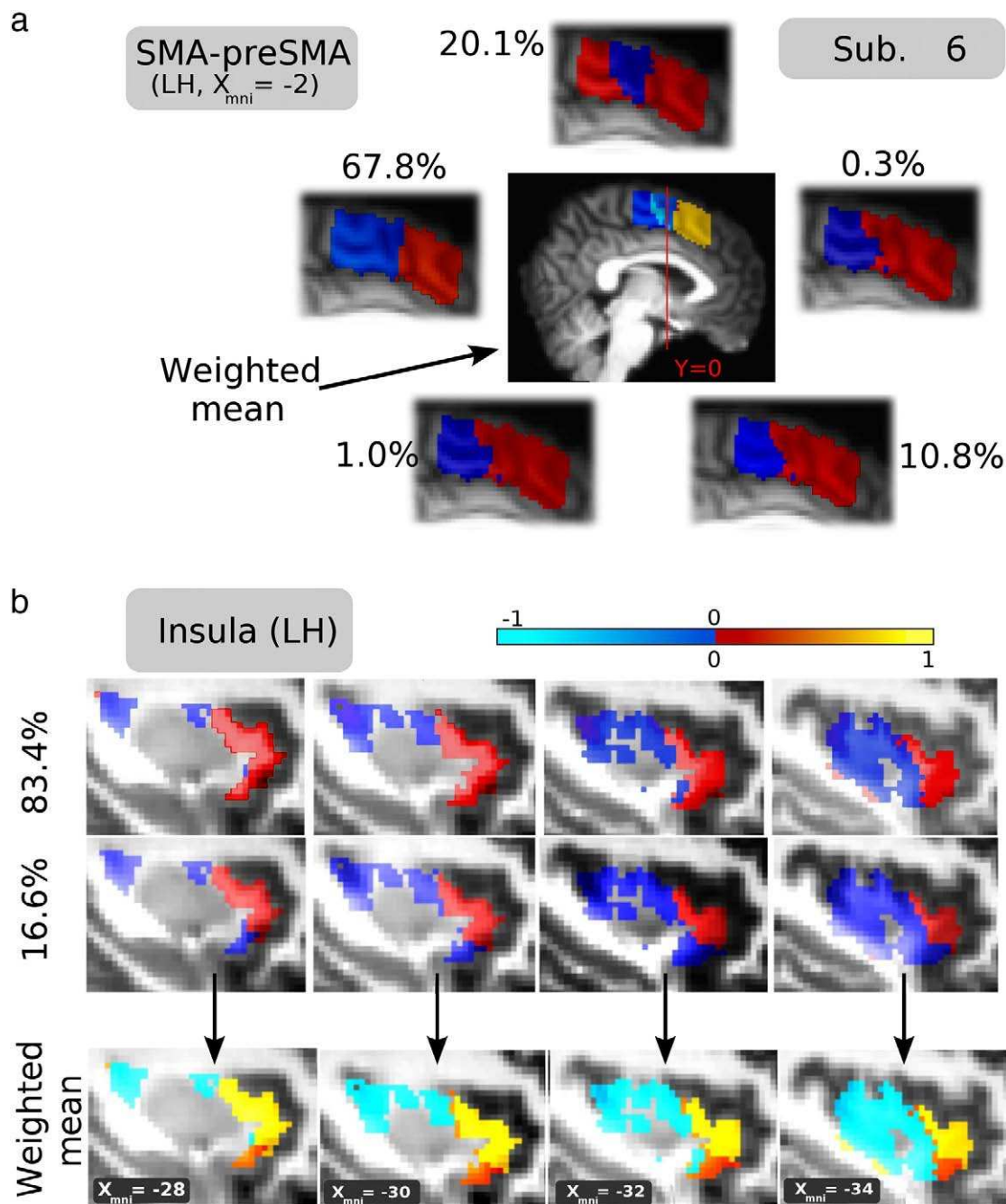
**Fig. 2.** (a) The single solution found for the SMA–preSMA is shown on a sagittal slice on the participant's anatomy ( $X_{MNI} = -2$ ). (b) Five solutions ordered by relative frequency are shown on 4 sagittal sections through the insula. The average solution is shown on the bottom row. Only the two most frequent ones show an anterior–posterior division of the insula. The color scale bar refers to the weighted means in (a) and (b). Clusters are shown in red and blue for the individual solutions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

SMA–preSMA. The risk clearly increases with a stricter threshold: at  $p = 0.01$  (1% quantile) the correlation values are 0.46 for insula and 0.13 for SMA–preSMA.

Taken together these results show that applying  $k$ -means just once on DWI data can lead to peculiar solutions that will not necessarily correspond to those another investigator would obtain performing another, single  $k$ -means clustering on the same data. Given that reproducibility is a key aim of modern scientific method, this is undesirable if not worrisome situation. Surprisingly, to our knowl-

edge, this issue has so far not been raised in the DWI literature. Using a different initialization algorithm for  $k$ -means (Lloyd, 1982) or comparing solutions against the min-SSD instead of the  $M_{1000}$  leads to very similar conclusions (Table 3).

An obvious solution is to run the  $k$ -means multiple times, but how often does it need to be applied to lead to a stable estimate? Let  $K_i$  be the mean of  $i$  randomly sampled  $k$ -means solutions. Figs. 4a–d show how similar  $K_i$  and  $M_{1000}$  are as a function of  $i$  (using values of  $i = 1, 2, 4, 8, 16, 32, 64, 128, 256, 512$ , and 900). Overall, the similarity increases



**Fig. 3.** Conventions as in Fig. 2 but five solutions have been found for the SMA–preSMA (a) and two for the insula (b). The color scale bar refers to the weighted means in (a) and (b). Clusters are shown in red and blue for the individual solutions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

monotonically, but for  $i \geq 256$ , little further improvement is observed. This means that for our ROIs, performing 256  $k$ -means or more, and averaging their solutions (after re-labeling as described in Materials and methods) leads to a stable estimate. Given how relatively similar the behavior of these parameters are in the insula and SMA–preSMA, it is likely that other ROIs would behave somewhat similarly. However the homogeneity of connectivity profiles within a cluster, the difference in the connectivity profiles between clusters and the signal-to-noise ratio of the DWI data are likely to influence how quickly these parameters will reach an asymptote. We suggest, therefore, running more than 256  $k$ -means, e.g. 1000, to ‘stay on the safe side’.

In summary, we suggest repeating  $k$ -means clustering 1000 times for each subject, re-labeling the solutions based on an automatic procedure,

and numerically averaging the solutions to achieve a stable and reproducible estimate of the connectivity-based parcellation of a brain region.

An alternative approach might be to keep only the solution corresponding to the putative global minimum of the cost function. However, the differences in SSD between any two solutions (including the ‘best’ and ‘second best’) are often very small ( $\sim 0.1\%$ ). In the face of such small differences, noise might easily swap the rank of a solution. Examining Fig. S2 for the SMA–preSMA for instance, it becomes apparent that the 1st solution (as ranked by SSD, i.e. the min-SSD solution) of Subject 9, resembles the 3rd, 4th and 5th best solutions of Subject 2, the 3rd of Subject 3 and the 2nd of Subject 6, but not the 1st solution of these participants. If one exclusively focuses on the min-SSD solution, one would therefore need to conclude that the SMA–preSMA of Subject 9 fundamentally differs from that of the other



**Table 2**  
Number of solutions found using  $k$ -means as a function of  $k$ .

| $k$ | SMA–preSMA |           | Insula    |           |
|-----|------------|-----------|-----------|-----------|
|     | Subject 3  | Subject 8 | Subject 3 | Subject 8 |
| 2   | 4          | 1         | 6         | 1         |
| 3   | 1          | 10        | 23        | 7         |
| 4   | 18         | 7         | 51        | 12        |
| 5   | 33         | 7         | 25        | 38        |
| 6   | 42         | 28        | 40        | 82        |
| 7   | 44         | 64        | 54        | 125       |
| 8   | 82         | 122       | 79        | 164       |
| 9   | 123        | 201       | 106       | 219       |
| 10  | 219        | 312       | 156       | 321       |

Number of different solutions found by 500 applications of  $k$ -means as a function of  $k$  (the number of clusters requested) separately for two ROIs subjects. Subject 3 and Subject 8 represent the extremes of interindividual variability at  $k=2$ , with Subject 3 showing multiple solutions in both ROIs, and Subject 8 a single solution in both ROIs. Note that there is no  $k$  value for which all participants evidence a single solution.

participants. Examining the distribution of solutions instead leads to what we believe to be a biologically more plausible conclusion: the participants are actually quite similar, but certain factors have swapped the rank order of their solutions. We therefore favor the average of the solutions ( $M_{1000}$ ) over the min-SSD as a summary map.

In the **Supplementary methods and results**: “issues about angular resolution” we present data from 3 additional participants acquired using two different DWI gradient directions schemes (15 and 60 direction schemes). Applying the same clustering procedure to both data sets shows that the issue of finding multiple solutions is not restricted to datasets acquired at relatively low angular resolution. In addition it confirms the fact that the variability of solutions is larger for the insula compared to the SMA–preSMA independently of angular resolution. Finally, the comparison shows that the summary map we propose, the  $M_{1000}$ , correlates more highly between two different acquisitions of the same participants than the solution with the min-SSD, see the table in **Supplementary methods and results**, “issues about angular resolution”. The same section contains the pictorial representation of all individual solutions found with both gradient direction schemes as well as the corresponding weighted averages.

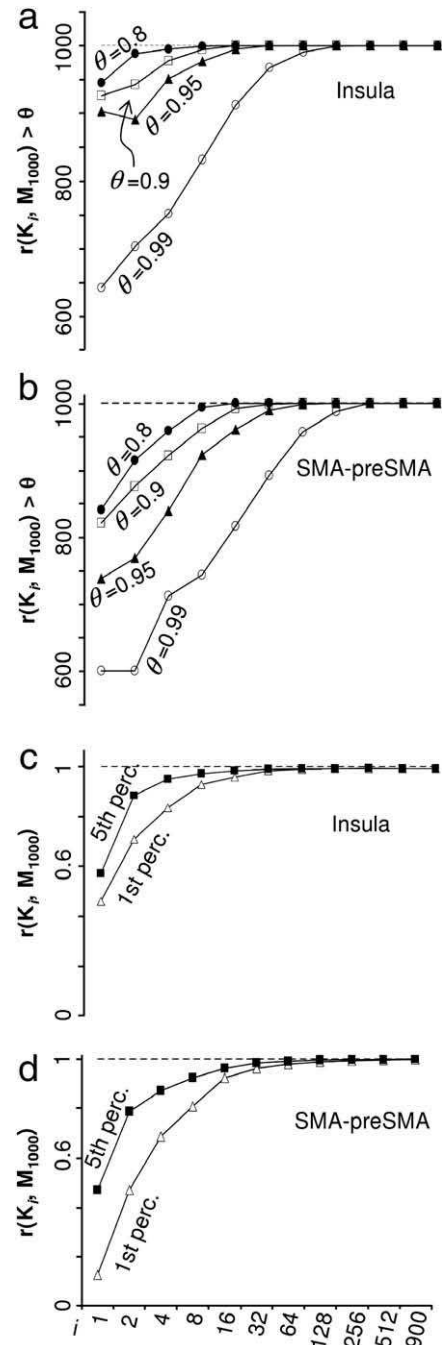
#### Group analysis

Comparing Figs. 2 and 3 shows that  $M_{1000}$  differs between Subjects 5 and 6, and Fig. S1 illustrates how variable  $M_{1000}$  is across our sample of 10 participants. The aim of most neuroimaging studies is to draw inferences about the human brain in general, and not about the specific brains of a small sample.

**Table 3**  
Impact of summary statistics and  $k$ -means initialization.

|                       | 1st percentile | 5th percentile | $r < 0.8$ | $r < 0.9$ | $r < 0.95$ | $r < 0.99$ |
|-----------------------|----------------|----------------|-----------|-----------|------------|------------|
| <b>Insula</b>         |                |                |           |           |            |            |
| $M_{1000}$            | 0.46           | 0.57           | 6%        | 7%        | 10%        | 35%        |
| min-SSD               | 0.30           | 0.37           | 7%        | 11%       | 14%        | 29%        |
| <b>SMA–preSMA</b>     |                |                |           |           |            |            |
| $M_{1000}$            | 0.13           | 0.47           | 16%       | 19%       | 23%        | 40%        |
| min-SSD               | 0.25           | 0.26           | 18%       | 18%       | 18%        | 22%        |
| <b>Insula (Lloyd)</b> |                |                |           |           |            |            |
| $M_{1000}$            | 0.23           | 0.85           | 5%        | 15%       | 21%        | 31%        |
| min-SSD               | 0.35           | 0.85           | 5%        | 9%        | 12%        | 30%        |

A sample of 1000 solutions for the  $k$ -means clustering performed only once ( $K_1$ ) was drawn from the estimated distribution of solutions at population level, and each  $K_1$  was correlated with both the summary statistics ( $M_{1000}$  and min-SSD). For the insula, the initialization was also performed according to Lloyd's algorithm. The first percentile of  $K_1$  has an average correlation with the summary statistics of  $0.29 \pm 0.11$ , the fifth of  $0.56 \pm 0.24$ . Neither of the initialization methods or summary statistics leads to systematically higher correlation values, suggesting that the choice of summary statistic or initialization strategy does not have a consistent impact on the central problem illustrated in this paper: a single application of  $k$ -means leads to potentially irreproducible results.



**Fig. 4.** Distribution of the correlation between the average of 1000 solutions and the average of  $i$   $k$ -means solutions as a function of  $i$ . (a) Frequency of finding solutions that correlate with  $M_{1000}$  more than a threshold value  $\theta$  as a function of  $i$  (x-axis) and  $\theta$  for the insula. (b) The same for the SMA–preSMA. (c) the 1st and the 5th percentile correlations with  $M_{1000}$  as a function of  $i$ . (d) same as (c) for the SMA–preSMA.

In functional neuroimaging, this issue has received much attention, and has led to the development of a two level statistical procedure that has been widely adopted (Holmes and Friston, 1998). At the first level, the activations observed in all the trials of the conditions to be compared are summarized in a single contrast volume of parameter estimates. At the second level, the contrast volume of each participant is entered into a voxel-wise statistic that tests whether the contrast significantly differs from zero. A significant difference in a voxel, at the second level, justifies the inference that (with a limited probability of type I error) the contrast would differ from zero in a new participant drawn from the general population. Non-significant differences can be due to two factors (or their combination). On the



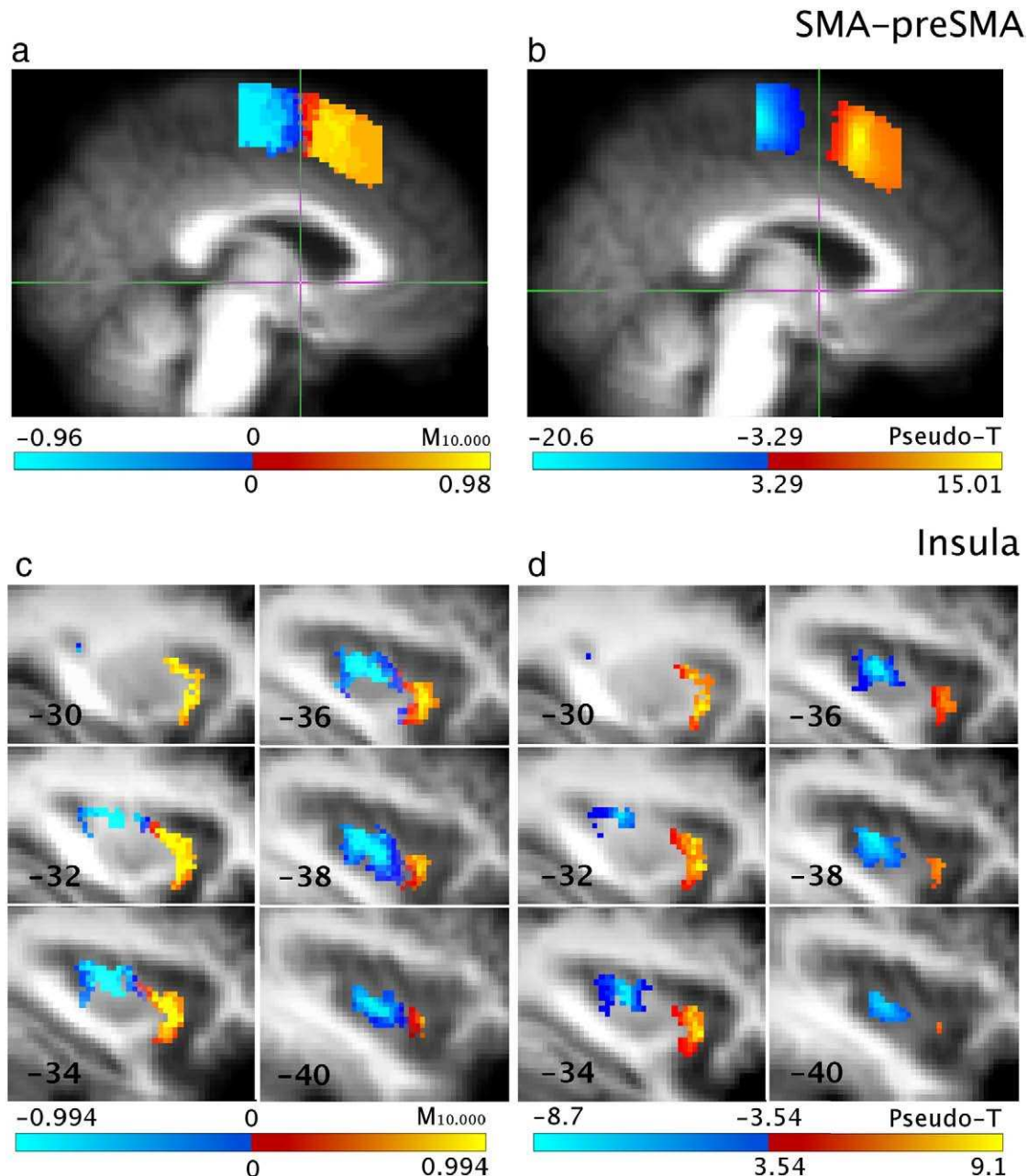
one hand, individual participants can have significantly non-zero contrasts at the first level of analysis but the variability across subjects is such that one cannot infer whether a new participant would show a difference in a particular direction. On the other hand, individual participants can fail to show evidence for a difference between conditions even at the first level.

The same approach could be applied to the investigation of DWI data. We propose to test, for each voxel in our group of participants, the null hypothesis that the voxel is equally likely to be associated with one or the other cluster (i.e. does not reliably belong to a specific cluster). In the single subject summary ( $M_{1000}$ ) maps, each voxel has a value of 1 if it is always associated with one of the clusters, of  $-1$  if it is always associated with the other cluster, and of 0 if it is associated equally often with both clusters. Numerically, this null hypothesis therefore translates into testing voxel-by-voxel, whether the mean of  $M_{1000} = 0$ .

To conduct this second level (group) analysis we normalized the  $M_{1000}$  of each individual into the MNI space. We then used SnPM to test the null hypothesis separately for the insula and SMA–preSMA. Fig. 5a shows the average  $M_{1000}$  value for each voxel of the SMA–preSMA, while Fig. 5b shows those voxels for which the pseudo  $t$ -value exceeds the significance threshold ( $p < 0.05$  Family Wise Error corrected over the extent of the ROI). Figs. 5c and d show the same for the insula. For each ROI, this analysis evidences that two compact clusters of voxels were reliably classified.

For the SMA–preSMA, the divide between the two regions fell around  $y=0$  in the MNI space, where the anatomical divide is thought to be located (Behrens and Johansen-Berg, 2005; Zilles et al., 1996), supporting the validity of our group analysis approach.

For the insula, voxels above threshold clustered in two regions: a more rostral cluster covering the anterior short gyrus, extending ventrally to the subapical pole, and an adjacent part of the middle



**Fig. 5.** Group analysis. (a) The normalized grand average cluster value obtained by averaging the  $M_{1000}$  of SMA–preSMA for the 10 participants shown on the average anatomy at  $X_{MNI} = -2$ . (b) Voxels where  $M_{1000} \neq 0$  at  $p < 0.05$ , F.W.E. corrected. (c) and (d) same for insula.

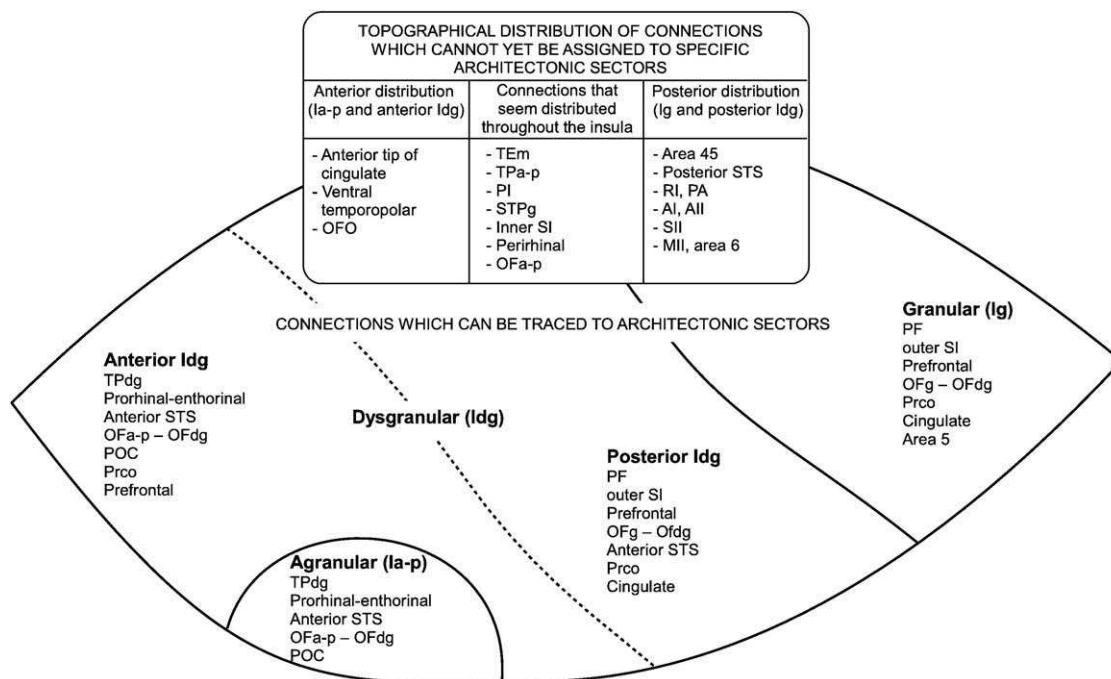
short gyrus, with higher values in more anterior cortical fields; and a caudal cluster, located in the posterior long gyri, with higher pseudo *t*-values in the most dorsal posterior insular territory. Voxels below threshold were located in the middle insula, in a region approximately corresponding with the middle and posterior short gyri and the anterior part of the long gyri. The insular cortex is known to feature a heterogeneous distribution of cortical types ranging from agranular (Ia: insula agranular) to granular (Ig: insula granular) cortex. This organization was already described by Brodmann (1909) in both humans (p. 122 and p.176) and monkeys (p. 132). In this study the variation of cortical types was shown along the rostro-caudal axis, but also noting that “on the edge of the anterior agranular half of the insula a cortical type of quite rudimentary structure can be distinguished” (p. 122), and stressing the need for further investigation on this territory. A similar rostro-caudal axis of variation in cortical types was described in the human brain by Rose (1928) and von Economo (Triarhou, 2007; von Economo, 1927). In humans, a dorso-ventral trajectory of variation in the cytoarchitecture (Brocahaus, 1940) and in the myeloarchitecture (Vogt, 1910) was also reported, nonetheless compatible with the former studies (Zilles, 2004). Subsequent studies in monkeys raised evidence for a radial organization of the cortical types around the piriform cortex (Amaral and Price, 1984; Jones and Burton, 1976; Mesulam and Mufson, 1982a; Roberts and Akert, 1963), with Ia occupying the most ventral anterior territories, Ig the postero-dorsal extent of the insula, and the dysgranular (Id: insula dysgranular) transition the zone lying in between and comprising also the most dorsal rostral extent of the insula. A similar topographic organization was found also for humans (Bonthuis et al., 2005; Mesulam and Mufson, 1982a; Öngür et al., 2003). The insular connectivity has been shown to have a strong relationship with the distribution of the cortical types: Ia and Ig have very different connectivity, often with brain regions featuring the same cytoarchitecture, while Id displays a wide spectrum of connectivity, ranging from allocortex to homotypical isocortex (Amaral and Price, 1984; Augustine, 1985; 1996 for a review; Chikama et al., 1997; Friedman et al., 1986; Fudge et al., 2005; Mesulam and Mufson, 1982b, 1985). Also different territories of Id display different connectivity patterns

according to the proximity with either Ia or Ig (see Fig. 6, adapted from Mesulam and Mufson (1982b)). For this reason, Mesulam and Mufson (1982b) concluded that, for purposes of topographical demarcation, the insula can be divided in an anteroventral sector (comprising the Ia and the anterior portion of Id) and a postero-dorsal sector (comprising the Ig and the posterior Id). Since our study aims at recovering regions with different connectivity in the insular cortex, we expected the trajectory of variation in connectivity patterns to be consistent with this anteroventral/dorsocaudal subdivision, rather than with the radial organization which appears to be more descriptive of the cytoarchitectonic distribution.

Interestingly, the gap between the classifiable voxels (i.e. those violating the null hypothesis) seems narrower for the SMA–preSMA and larger for the insula. To ensure that this is not due to the family wise error correction being more stringent in the larger insula, we applied the same uncorrected  $p < 0.002$  (corresponding to 0.05 F.W.E. correction for the insula) threshold to both regions and compared the proportion of significant voxels. In the SMA–preSMA 242 out of the 331 (73%) of the voxels were classifiable (i.e. violated the null hypothesis), whereas in the insula only 333/559 (i.e. 60%) were. A chi-squared test (with Yate's correction) shows that this difference is significant ( $p \sim 0.002$ ). A wider gap between classifiable zones can, as discussed above, reflect the presence of a larger number of voxels with large variability in the clusterisations within (i.e. individual  $M_{1000} \sim 0$ ) and/or between participants (i.e. individual  $M_{1000}$  could differ from zero but permutation test indicates that median ( $M_{1000}$ ) is not significantly different from 0).

The same analysis performed on the distribution of min-SSD solutions does not provide significantly different results in terms of spatial localization of clusters (see Supplementary methods and results, Fig. S5). However, the percentage of voxels above threshold decreased, due to the increased overall variance caused by the dissimilarity among the ‘best’ individual solutions. The classified/total voxels ratio was 211/331 for SMA–preSMA and 210/559 for the insula.

To examine if a systematic difference in signal-to-noise ratio of our DWI acquisitions between the insula and the SMA–preSMA could account for this difference, we compared the signal-to-noise ratio in



**Fig. 6.** Diagram of insular topographical connections (afferents and efferents). The dysgranular territory is arbitrarily bisected into anterior and posterior by the dotted line. The anteroventral subdivision of the insula is constituted by Ia-p and the anterior Idg, while the postero-dorsal subdivision encompasses Ig and posterior Idg. Adapted from Mesulam and Mufson (1982b).

the insula and SMA–preSMA, but found no significant difference ( $p > 0.58$ , see [Supplementary methods and results](#)).

Why can we assign a significantly higher proportion of SMA–preSMA voxels to either cluster compared to the insula? Anatomically, the insula is characterized by a wide region of transitional cytoarchitecture (Friedman et al., 1986; Jones and Burton, 1976; Mesulam and Mufson, 1985). In contrast the SMA–preSMA has two regions of relatively homogeneous cytoarchitectonics separated by a comparatively narrow boundary. Given the acknowledged relationship between cytoarchitectonics and connectivity in these regions (Chikama et al., 1997; Mesulam and Mufson, 1982a; Picard and Strick, 1996) that has formed the basis of most connectivity-based parcellation studies (Johansen-Berg, 2004; Klein et al., 2007; Tomassini et al., 2007), one would expect a difference in the degree of similarity between connectivity profiles as a function of distance within our two ROIs. In particular, if one were to examine the connectivity profile of two neighboring seed voxels in the SMA–preSMA region, in the vast majority of cases, they would correlate very highly, because they would be drawn from the relative homogeneous SMA or preSMA, and very seldom from the narrow transition zone between these regions. If one were to do the same in the insula, the correlation should be lower, because they would more likely derive from the much broader transition zone. If one were to compare two voxels taken at an increasing distance from each other, the relationship should invert, because they would increasingly incorporate voxels from the two very different regions in the SMA–preSMA, whilst continuing to sample voxels along the transition zone in the insula. For the largest distances finally, correlations should be close to zero in both regions. To verify this prediction, we plotted the correlation between connectivity profiles in our data as a function of distance for both regions (Figs. S3a, b). We found exactly the expected interaction: for distances between voxels ranging from 4 mm to 14 mm, connectivity profiles were more highly correlated in the SMA–preSMA, while for longer distances ranging from 28 mm to 54 mm, they were more highly correlated in the insula ( $t$ -tests,  $p < 0.05$  Bonferroni corrected for 31 distances). This pattern was common to all participants.

What impact would the observed difference in correlation between connectivity profiles, observed in our data between the insula and SMA–preSMA, have on the reliability with which voxels can be clustered using  $k$ -means? Mathematically, there is unfortunately no simple relationship between the smoothness of the transition between datasets to be clustered and the stability of  $k$ -means clustering: for instance 1000  $k$ -means on the sequence [1,2,3,4,5,6,7,8,8,9,10] always creates the same two clusters [1,2,3,4,5] and [6,7,8,9,10] despite the perfect smoothness of the transition. However, a fluctuation in the position of the border between two clusters of  $\sim 30$  mm (which is the width of the non-classifiable corridor of our data for the insula) corresponds to a larger fluctuation in the corresponding distances in our  $k$ -means space for the SMA–preSMA compared to the insula (because distance  $\sim 1 - r$ , and  $r$  is larger on average in the insula between voxels with a distance  $> 18$  mm). Given that  $k$ -means clustering operates on minimizing within-cluster distances, this may mean, that for a commensurable fluctuation between solutions in  $k$ -means space, one would expect a larger fluctuation in anatomical space for the insula. Although tentative, this interpretation would suggest a relationship between (a) the smoothness of the anatomical transition within a region and (b) the leeway for variability in solutions within (where the differences in distance between local minima and global minimum are smaller in smoother datasets, see Fig. S4) and/or between participants left to the  $k$ -mean clustering algorithm.

## Conclusions

This study investigated for the first time whether a single application of the  $k$ -means algorithm is a reliable way to cluster data

acquired using DWI. We show that it is not: single applications of the algorithm can lead to irreproducible results.

In contrast we show that applying  $k$ -means 1000 times and averaging the results leads to a stable and reproducible estimate of how a region can be clustered in two clusters using connectivity profiles. Averaging multiple  $k$ -means has the advantage of generating graded summary maps for each participant, in which the value of each voxel reflects how reliably a voxel is attributed to a cluster. We further propose that average  $k$ -means maps should be submitted to hypothesis testing at the second, group level to provide maps that estimate whether in the population, voxels belong to one or the other of two clusters. This approach represents an important evolution of connectivity-based cortical parcellation by providing a reproducible and statistically meaningful map of cortical segmentation.

## Acknowledgments

The research was supported by a VIDI grant and a Cognition Pilot Grant to CK from the Dutch National Science Foundation (N.W.O.) and a Marie Curie Excellence Grant to CK from the European Commission. The authors want also to thank Harma Meffert, for her invaluable help during the collection of the data.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2009.06.014](https://doi.org/10.1016/j.neuroimage.2009.06.014).

## References

- Amaral, D.G., Price, J.L., 1984. Amygdalo-cortical projections in the monkey (*Macaca fascicularis*). *J. Comp. Neurol.* 230, 465–496.
- Anwander, A., Tittgemeyer, M., von Cramon, D.Y., Friederici, A.D., Knösche, T.R., 2007. Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.
- Augustine, J.R., 1985. The insular lobe in primates including humans. *Neurol. Res.* 7, 2–10.
- Augustine, J.R., 1996. Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res. Brain Res. Rev.* 22, 229–244.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267.
- Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088.
- Behrens, T.E., Johansen-Berg, H., 2005. Relating connective architecture to grey matter function using diffusion imaging. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360, 903–911.
- Bonhthuis, D.J., Solodkin, A., Van Hoesen, G.W., 2005. Pathology of the insular cortex in Alzheimer disease depends on cortical architecture. *J. Neuropathol. Exp. Neurol.* 64, 910–922.
- Brockhaus, H., 1940. Die cyto- und myeloarchitektonik des Cortex claustralis und des Claustrum beim Menschen. *J. Psychol. Neurol.* 49, 249–348.
- Brodman, K., 1909. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues (Page numbering according to the 2006 english version: Brodmann's Localization in the Cerebral Cortex. Berlin, Springer). JA Barth, Leipzig.
- Chikama, M., McFarland, N.R., Amaral, D.G., Haber, S.N., 1997. Insular cortical projections to functional regions of the striatum correlate with cortical cytoarchitectonic organization in the primate. *J. Neurosci.* 17, 9686–9705.
- Friedman, D.P., Murray, E.A., O'Neill, J.B., Mishkin, M., 1986. Cortical connections of the somatosensory fields of the lateral sulcus of macaques: evidence for a corticolimbic pathway for touch. *J. Comp. Neurol.* 252, 323–347.
- Fudge, J., Breitbart, M., Danish, M., Pannoni, V., 2005. Insular and gustatory inputs to the caudal ventral striatum in primates. *J. Comp. Neurol.* 490, 101–118.
- Hartigan, J.A., 1975. *Clustering Algorithms*. Wiley.
- Hartigan, J.A., Wong, M.A., 1979. A  $k$ -means clustering algorithm. *Appl. Stat.* 28, 100–108.
- Higham, D.J., Kalna, G., Kibble, M., 2007. Spectral clustering and its use in bioinformatic. *J. Comput. Appl. Math.* 204, 25–37.
- Holmes, A.P., Friston, K.J., 1998. Generalisability, random effects and population inference. *NeuroImage* 7.
- Hosey, T., Williams, G., Ansorge, R., 2005. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn. Reson. Med.* 54, 1480–1489.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–841.



- Johansen-Berg, H., 2004. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc. Natl. Acad. Sci.* 101, 13335–13340.
- Jones, E.G., Burton, H., 1976. Areal differences in the laminar distribution of thalamic afferents in cortical fields of the insular, parietal and temporal regions of primates. *J. Comp. Neurol.* 168, 197–247.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2004. A local search approximation algorithm for *k*-means clustering. *Comput. Geom.* 28, 89–112.
- Klein, J., Behrens, T., Robson, M., Mackay, C., Higham, D., Johansenberg, H., 2007. Connectivity-based parcellation of human cortex using diffusion MRI: establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *NeuroImage* 34, 204–211.
- Lloyd, S.P., 1982. Quantization in PCM. *IEEE Trans. Inf. Theory* 28, 129–137.
- Mesulam, M.M., Mufson, E.J., 1982a. Insula of the old world monkey. I. Architectonics in the insulo-orbito-temporal component of the paralimbic brain. *J. Comp. Neurol.* 212, 1–22.
- Mesulam, M.M., Mufson, E.J., 1982b. Insula of the old world monkey. III: Efferent cortical output and comments on function. *J. Comp. Neurol.* 212, 38–52.
- Mesulam, M., Mufson, E., 1985. The insula of Reil in man and monkey. Architectonics, connectivity, and function. In: Jones, P. (Ed.), *Cerebral Cortex*. Plenum Press, New York, pp. 179–226.
- Naidich, T.P., Kanga, E., Fatterpekar, G.M., Delman, B.N., Gultekin, S.H., Wolfe, D., Ortiz, O., Yousry, I., Weismann, M., Yousry, T.A., 2004. The Insula: Anatomic Study and MR Imaging Display at 1.5 T. *Am. J. neuroradiol.* 25, 222–232.
- Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.
- Öngür, D., Ferry, A., Price, J., 2003. Architectonic subdivision of the human orbital and medial prefrontal cortex. *J. Comp. Neurol.* 460, 425–449.
- Picard, N., Strick, P.L., 1996. Motor areas of the medial wall: a review of their location and functional activation. *Cereb. Cortex* 6, 342–353.
- Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A., Di Chiro, G., 1996. Diffusion tensor MR imaging of the human brain. *Radiology* 201, 637–648.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. *Numerical Recipes in C—The Art of Scientific Computing*, Second Edition. Cambridge University Press.
- R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Roberts, T.S., Akert, K., 1963. Insular and opercular cortex and its thalamic projection in *Macaca mulatta*. *Schweizer Archiv für Neurologie, Neurochirurgie und Psychiatrie=Archives suisses de neurologie, neurochirurgie et de psychiatrie* 92, 1–43.
- Rose, M., 1928. Die Inselrinde des Menschen und der Tieren. *J. Psychol. Neurol.* 37, 464:624.
- Selim, S.Z., Ismail, M.A., 1994. *K*-means-type algorithms: a generalized convergence theorem and characterization of local optimality. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 81–87.
- Tomassini, V., Jbabdi, S., Klein, J., Behrens, T., Pozzilli, C., Matthews, P., Rushworth, M., Johansen-Berg, H., 2007. Diffusion-weighted imaging tractography-based parcellation of the human lateral premotor cortex identifies dorsal and ventral subregions with anatomical and functional specializations. *J. Neurosci.* 27, 10259–10269.
- Triarhou, L., 2007. The Economo–Koskinas atlas revisited: cytoarchitectonics and functional context. *Stereotact. Funct. Neurosurg.* 85, 195–203.
- Türe, U., Yasargil, D.C., Al-Mefty, O., Yasargil, M.G., 1999. Topographic anatomy of the insular region. *J. Neurosurg.* 90, 720–733.
- Vogt, O., 1910. Nouvelle contribution à l'étude de la myéloarchitecture de l'écorce cérébrale. XXe Congrès des Médecins Aliénistes et Neurologistes de France et des pays de langue française, Bruxelles.
- von Economo, K., 1927. *Zellaufbau der Grosshirnrinde des Menschen*. Springer, Berlin.
- Zilles, K., 2004. Architecture of the human cerebral cortex—regional and laminar organization. In: Paxinos, G. (Ed.), *The Human Nervous System*. Academic Press, San Diego, pp. 997–1055.
- Zilles, K., Schlaug, G., Geyer, S., Luppino, G., Matelli, M., Qü, M., Schleicher, A., Schormann, T., 1996. Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. *Adv. Neurol.* 70, 29–43.