Automatic group-wise whole-brain short association fiber bundle labeling based on clustering and cortical surface information

Andrea Vázquez
Universidad de Concepcion
Narciso López-López
Universidad de Concepcion
Josselin Houenou
Commissariat a l'Energie Atomique et aux Energies Alternatives Centre de Saclay
Cyril Poupon
Commissariat a l'Energie Atomique et aux Energies Alternatives Centre de Saclay
Jean-François Mangin
Commissariat a l'Energie Atomique et aux Energies Alternatives Centre de Saclay
Susana Ladra
Universidade da Coruna
Pamela Guevara (pamela.guevara@biomedica.udec.cl)
Universidad de Concepcion https://orcid.org/0000-0001-9988-400X

Research

Keywords: fiber labeling, clustering, fiber bundle, tractography, superficial white matter

Posted Date: January 9th, 2020

DOI: https://doi.org/10.21203/rs.2.20420/v1

Version of Record: A version of this preprint was published at BioMedical Engineering OnLine on June 3rd, 2020. See the published version at https://doi.org/10.1186/s12938-020-00786-z.

RESEARCH

Automatic group-wise whole-brain short association fiber bundle labeling based on clustering and cortical surface information

Andrea Vázquez[1†], Narciso López-López[1,2†], Josselin Houenou[3,4,5,6], Cyril Poupon[3], Jean-François Mangin[3], Susana Ladra[2] and Pamela Guevara[1*]

[*]Correspondence: pamela.guevara@biomedica.udec.cl
[1] Universidad de Concepción, Faculty of Engineering, Concepción, Chile
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Background: Diffusion MRI is the preferred non-invasive in vivo modality for the study of brain white matter connections. Tractography datasets contain 3D streamlines that can be analyzed to study the main brain white matter tracts. Fiber clustering methods have been used to automatically regroup similar fibers into clusters. However, due to inter-subject variability and artifacts, the resulting clusters are difficult to process for finding common connections across subjects, specially for superficial white matter.

Methods: We present an automatic method for labeling of short association bundles on a group of subjects. The method is based on an intra-subject fiber clustering that generates compact fiber clusters. Posteriorly, the clusters are labeled based on the cortical connectivity of the fibers, taking as reference the Desikan-Killiany atlas, and named according to their relative position along one axis. Finally, two different strategies were applied and compared for the labeling of inter-subject bundles: a matching with the Hungarian algorithm, and a well-known fiber clustering algorithm, called QuickBundles.

Results: Individual labeling was executed over four subjects, with an execution time of 3.6 minutes. An inspection of individual labeling based on a distance measure, showed good correspondence among the four tested subjects. Two inter-subject labeling were

successfully implemented and applied to 20 subjects, and compared using a set of distance thresholds, ranging from a conservative value of 10 mm to a moderate value of 21 mm. Hungarian algorithm led to high correspondence, but low reproducibility for all the thresholds, with 96 seconds of execution time. QuickBundles led to better correspondence, reproducibility and short execution time of 9 seconds. Hence, the whole processing for the inter-subject labeling over 20 subjects takes 1.17 hours.

Conclusion: We implemented a method for the automatic labeling of short bundles in individuals, based on an intra-subject clustering and the connectivity of the clusters with the cortex. The labels provide useful information for the visualization and analysis of individual connections, what is very difficult without any additional information. Furthermore, we provide two fast inter-subject bundle labeling methods. The obtained clusters could be used for performing manual or automatic connectivity analysis in individuals or across subjects.

Keywords: fiber labeling; clustering; fiber bundle; tractography; superficial white matter

Background

The preferred technique to non-invasively study structural brain connections is Diffusion-weighted Magnetic Resonance Imaging (dMRI), based on the measurement of water molecules movement [1, 2]. Diffusion tractography estimates the main white matter (WM) tracts, obtaining a set of 3D paths, called streamlines or fibers [3]. Tractography datasets can be analyzed to automatically extract or segment known WM bundles, with anatomical meaning. One strategy can be done through regions of interest (ROIs) that connect two zones of the cortex manually [4], or applying an atlas of regions of interest (ROI) and then use anatomical descriptions of the bundles to extract fibers connecting or passing through specific ROIs [5]. Automatic methods based on ROIs allow an easy modification or addition of bundle extraction rules, but do not include an analysis based on the trajectories of the fibers as a whole. Another strategy is based on clustering to regroup fibers with similar shape and position, and an atlas embedding anatomical bundle information

[6, 7, 8]. Also, other simpler algorithms have been implemented to extract bundles based on a multi-subject atlas [9, 10, 11]. Several atlases have been created to represent main deep white matter (DWM) bundles, which have been well described by anatomists and are very stable across subjects [6, 9, 12], i. e. present high similarity and can be found in all the subjects on medium to high quality databases. However, there exist several WM fiber bundles still unknown or not sufficiently described, because of their higher inter-subject variability and less reproducibility. This is the case of short association bundles, where only a few works have been focused on their description for the whole brain [13, 14]. Superficial white matter fibers can be studied using exploratory fiber clustering methods that aim to detect fiber tracts without having any reference to the start or end of WM fibers [15]. This type of algorithm, applied to a whole-brain tractography datasets, generates a set of fiber clusters representing the main WM connections in the analyzed brain. In the case of the works in [13, 14], SWM bundle atlases were obtained using different methods based on fiber clustering and the addition of anatomical information. Also, a recent work found a great amount of SWM bundles [16], but those were not labeled, requiring a posterior analysis for their study. Hence, existing methods have been focused on finding reproducible bundles across subjects, but not on the development of an automatic labeling of individual or inter-subject SWM clusters. Whole-brain fiber clustering methods, applied to individuals or to a population of subjects, do not return directly the identification of the obtained clusters, e. g. information about the anatomical areas connected by the fibers and their relative position in the cortex. Such identification or labeling could be very useful for the study of the human brain connectome in individuals and different populations. The labeled clusters could then be used to perform detailed analyses of known bundles, i. e., subdivisions of the main bundles, and also of unknown fascicles, such as short

association bundles. Furthermore, multi-subject analyses could be applied to create new bundle atlases.

We propose a method that automatically labels the clusters obtained from an intra-subject clustering, based on the regions connected by the clusters. This information is based on a cortical surface mesh, labeled with the Desikan-Killiany atlas. Direct correspondence between subjects is obtained for the connected anatomical regions. Furthermore, within each region, the clusters on individuals are labeled following and ordering criterion. Moreover, we apply two strategies for inter-subject cluster labeling. First, a matching method is implemented based on the Hungarian algorithm to find bundle correspondence across subjects. Also, a clustering algorithm is applied to perform inter-subject labeling, allowing the regrouping of similar bundles of a subject. Both implementations are fast, taking respectively about 100 and 10 seconds, over 20 subjects. The performance of both inter-subject implementations was compared in terms of reproducibility and inter-subject bundle distance. The methods are publicly available from [17].

Results

The experiments were executed on a computer with 4-core Intel Core i5-8250U CPU running at 1.60 GHz, 6MB of cache and 8GB of RAM, using Ubuntu 18.04.2 LTS with kernel 4.15.0-64 (64 bits). The programming language used to develop almost all stages is python 3.6. For the analysis, the tractographies, meshes and labels according to Desikan-Killiany atlas of 20 subjects were used. The intra-subject labeling was applied to all the subjects. First, the results for intra-subject labeling performed on four subjects are shown and analyzed. Next, the two inter-subject labeling methods, matching and clustering, were applied to the 20 subjects, using a set of distance thresholds, ranging from a conservative value of 10 mm to a moderate value of 21 mm. The reproducibility of the bundles was evaluated by counting the number of subjects that had each bundle. The quality of the labeling was evaluated

using a measure of distance between bundles across subjects. Also, heatmaps are shown to have an insight on the reproducibility and variability in terms on the number of fibers of the most reproducible bundles, for a restrictive threshold of 12 mm. Finally, some examples of bundles are displayed for a visual inspection of the results. Hungarian algorithm led to high correspondence, but low reproducibility for all the thresholds, with 96 seconds of execution time. QuickBundles led to better correspondence and reproducibility and short execution time, of only 9 seconds. Hence, the whole processing for the inter-subject labeling over 20 subjects takes on average 1.17 hours.

In the following we detail the results.

Intra-subject labeling

The intra-subject labeling was applied to the 20 subjects. For filtering (stage 2), a filter with a minimum cluster size of $min_{nf} = 10$ fibers, and minimum cluster length of $min_{len} = 30$ mm and a maximum cluster length of $max_{len} = 80$ mm was used to discard small and also short clusters, leading to an average of 1100 clusters per subject. The filtering values are similar to those previously used [13, 14]. The clusters of each subject were labeled according to the pair of anatomical regions connected by each bundle, and the position based on ascending order on y-axis (default configuration).

An example of the relative ordering for intra-subject bundle labeling is presented in Figure 1. We can appreciate that bundles connecting *postcentral (PoC)* and *precentral (PrC)* regions are ordered according to y-axis in ascending order. Bundles are ordered according to the PoC parcel, since it is indexed before in the Desikan-Killiany atlas.

Figure 1 Bundles connecting right PoC and PrC regions. Example of bundle labeling according to the relative position of the bundles connecting *PoC* and *PrC* regions for Subject001.

Even though the method performs only an intra-subject analysis, a degree of correspondence between the four first subjects can be found, according to their relative position index. Because of inter-subject variability, the correspondence is not perfect, nor do they all have the same number of bundles. Figure 2 displays the first five bundles of the four subjects, which connect the left *PrC* gyri with the *supramarginal (SM)* gyri.

Figure 2 Correspondence of intra-subject bundle labels across subjects. Comparison of the first five bundles from four subjects (001-004), connecting *PrC* and *SM* gyri.

A quantitative evaluation of the bundle correspondence among subjects is displayed in Figure 3, where the distance between the bundle centroids of each pair of subjects for each bundle is represented with a defined color. Bundles show a relative good correspondence among them, with distances between centroids ranging from seven to 36 mm, with an average of about 20 mm. Note that distances of 20-30 mm have been previously used for inter-subject analyses of superficial white matter [13, 14].

Figure 3 Bundle centroid distances between pairs of subjects for intra-subject labeling of four subjects. Distances in mm between bundle centroids for all the pairs of subjects, for five bundles connecting *PrC* and *SM* gyri. The colors represent the different pairs of subjects.

Finally, the average execution times for each stage of the intra-subject labeling are: 192s for *stage 1*, 10.23s for *stage 2*, 11s for *stage 3* and 2.49s for *stage 4*, taking on average a total time of 3.5 minutes.

Inter-subject labeling

The inter-subject labeling was applied to 20 subjects from the ARCHI database [18]. A comparison has been made between the two implemented methods, Matching with the *Hungarian* algorithm [19] and clustering with the *QuickBundles (QB)* algorithm [20]. Both algorithms work with an input parameter, which is the minimum average

direct-flip (MDF) distance [20] from one centroid to another. In addition, tests have been carried out with four different thresholds. The first threshold of 10 mm is very conservative, being the default threshold of *QB* for intra-subject clustering. We also used a 12 mm threshold, which is still conservative and aims to find similar bundles across subjects. Two other moderate thresholds were used: 18 mm and 21 mm, which are adequate considering that distances of 20-30 mm have been previously used for SWM inter-subject analyses [13, 14].

The reproducibility of the methods was evaluated by counting the number of subjects in which each bundle was found. Figure 4 shows the reproducibility of the bundles for both methods and the four thresholds. Table 1 lists three reproducibility indices for the two inter-subject labeling methods, separated by hemisphere: the maximum number of subjects for the bundles within the 20 most reproducible bundles, and the number of bundles with reproducibility greater than or equal to 50% and 75%. As expected, for both algorithms, the higher the distance threshold, the greater the reproducibility. As can be seen, the method that shows the highest reproducibility is *QB*, presenting 94 bundles with more than 50% of reproducibility for a distance threshold of 21 mm, what is a good number, based on previous studies [13, 14]. On the other hand, the *Hungarian* algorithm only found 34 bundles with more than 50% of reproducibility for the same threshold. Furthermore, the *Hungarian* algorithm found no bundles present in all subjects, while *QB* found 19 for 21 mm threshold. As the *Hungarian* algorithm tries to match 1-1 the bundles, leads to less reproducibility than *QB*.

Figure 4 Reproducibility of bundles with inter-subject labeling for the two methods. The number of subjects is shown on the x-axis while the y-axis shows the number of clusters in each range.

Figure 5 shows inter-subject labeling quality for both methods with the four tested distance thresholds. The quality is evaluated using the inter-cluster distance

Table 1 Reproducibility values between for the two inter-subject labeling methods. The left column identifies the method (Hungarian or QB), hemisphere (left or right), and the thresholds (12 mm, 18 mm or 21 mm). The second column lists the maximum number of subjects for the bundles within the 20 most reproducible bundles. Columns three and four show the number of bundles with reproducibility greater than or equal to 50% and 75%, respectively.

Method	max	# bundles $\geq$ 50%	# bundles $\geq$ 75%
Hungarian12_left	7	0	0
Hungarian12_right	6	0	0
QB12_left	19	12	3
QB12_right	14	14	0
Hungarian18_left	13	3	0
Hungarian18_right	11	3	0
QB18_left	20	41	9
QB18_right	19	42	6
Hungarian21_left	13	22	0
Hungarian21_right	13	12	0
QB21_left	20	49	10
QB21_right	20	45	9

(MDF), calculated per each bundle as the average distance between all the pairs of bundle centroids from the subjects where the bundle was labeled. Thus, the clusters classified with the same label are measured together, the closer the clusters are, the better the quality of the method. As expected, for both algorithms, the lower the distance threshold, the higher the quality. It can be seen that the most accurate algorithm is the *Hungarian* with a 10 mm threshold, at expenses of a low reproducibility, as shown above. The *QB* algorithm has a lower quality than *Hungarian*, because it groups clusters of the same subject and merges them, thus increasing the inter-cluster distance. However, the merging of close clusters leads to a final better reproducibility, while keeping a moderate intra-cluster distance across subjects, with values inferior to 30 mm, and an average of about 15 mm.

Figure 5 Inter-cluster bundle distance for both inter-subject labeling method. X-axis represents the inter-cluster distance measured in mm. Y-axis shows the number of clusters in each range.

To have an insight of the reproducibility and variability of the most reproducible bundles in all subjects for the two labeling methods, we created heatmaps (Figures 6, 7, 8 and 9). The heatmaps were created separately for the 20 bundles of the left and

right hemispheres with the highest reproducibility in the 20 subjects, for a 12 mm threshold. Figures 6 and 7 display the heatmaps for the *Hungarian* algorithm, for left and right hemispheres, respectively, while Figures 8 and 9 show the heatmaps for the *QB* clustering algorithm. The bundles appear in descending order along the y-axis, according to the reproducibility between subjects, which appear along the x-axis. Empty (white) boxes indicate that a bundle does not exist in a certain subject. The colors represent the normalized number of fibers of each bundle (0-1), the darker, more fibers.

Figure 6 Reproducibility heatmap for Hungarian algorithm with threshold 12 mm, for the left hemisphere. On the x-axis are the subjects, on the y-axis are the 20 most reproducible bundles. The greater the number of fibers, the darker the color of the box on the heatmap that is normalized between 0 and 1.

Figure 7 Reproducibility heatmap for Hungarian algorithm with threshold 12 mm, for the right hemisphere. X-axis displays the subjects used, the 20 most reproducible bundles are shown on the y-axis. The darker boxes indicate a higher concentration of fibers in the bundle. These values are normalized between 0 and 1.

Figure 8 Reproducibility heatmap for QB with threshold 12 mm, for the left hemisphere. X-axis shows the subjects, while the y-axis shows the 20 most reproducible bundles among subjects in the left hemisphere. Darker boxes show bundles with more fibers in them. White boxes show absence of the bundle in the determined subject. The heat bar shows the values of normalized fibers between 0 and 1.

Figure 9 Reproducibility heatmap for QB with threshold 12 mm, for the right hemisphere. X-axis displays the subjects, and on the y-axis appears the 20 most reproducible bundles. The lighter the color of the box, the fewer fibers it contains. If the box is white, it indicates the absence of the bundle in the subject. The fiber values appear normalized between 0 and 1 in the heat bar.

It can be seen that, as in Figure 5, the method with the highest reproducibility is *QB*. The number of fibers seems to be more homogeneous for *QB*, with a tendency of a low normalized number of fibers. This does not mean that the bundles have few fibers, but that their number is of a given value for most of the subjects, with

very high values for a few subjects. The bundle with the highest reproducibility is *lh_PoC-SM_2* which was found in 19 subjects, followed by *lh_IP-SP_0* and *lh_Tr-RMF_0*, both found in 18 subjects. Reproducibility using matching is poorer, whose most reproducible bundle, *lh_IP-SP_69*, appears in only seven subjects.

Finally, some examples of bundles with high, medium and low reproducibility are displayed for a visual inspection of the results. Figure 10 shows the bundle *lh_PoC-SM_2*, belonging to the left hemisphere and classified by the *QB* clustering method with 12 mm threshold. This is the bundle with the highest reproducibility with this restrictive threshold, being present in 19 out of the 20 subjects, with the exception of *Subject008*, achieving a 95% of reproducibility. It can be seen how bundles connect approximately the same cortical regions in different subjects and have a similar main shape. Also, it can be seen that the number of fibers is very variable among subjects, what is usual in SWM bundles.

Figure 10 Bundle lh_PoC-SM_2, with the highest reproducibility in all subjects. The results show good reproducibility among subjects, appearing in 19 of the 20 subjects for the *QB* method with 12 mm threshold.

Figure 11 shows the bundle *lh_PoC-PrC_0*, of the left hemisphere and classified by the *QB* method with 12 mm of threshold. It appeared in 11 out of the 20 subjects, that is, a 55% of reproducibility. This is a small bundle connecting the *PoC* and *PrC* gyri.

Figure 11 Bundle PoC-PrC_0, with medium reproducibility. The *PoC-PrC_0* bundle appears in 11 out of 20 subjects, achieving 55% of reproducibility, for *QB* algorithm with a 12 mm threshold.

Lastly, Figure 12 shows for the left hemisphere the cluster *lh_RMF-SF_7*, classified by the *QB* method using a threshold of 12 mm. This is the least reproducible cluster of the heatmap of Figure 8, appearing in 9 out of the 20 subjects, reaching only 45% of reproducibility. However, it can be seen that the bundles connect the same area in all subjects, slightly varying the position and the number of fibers.

Figure 12 Bundle lh_RMF-SF_7, with low reproducibility. The bundle appears in nine out of 20 subjects using the *QB* clustering algorithm with a 12 mm threshold.

Discussion

In the last two decades, a great number of methods have been proposed for the analysis of tractography datasets. Most of the works have been focused on the study of deep white matter bundles, such as the arcuate fasciculus or the inferior fronto-occipital fasciculus. These bundles are in general larger and more stable across subjects, and has being described by neuroanatomists several decades ago. The methods have been focused on the study of these bundles, the creation of WM bundle atlases and the segmentation of WM bundles. Most of the studies have been developed with a combination of ROI-based and clustering-based methods, and the important guidance of neuroanatomy experts. In general, the applications analyze the segmented bundles across subjects and different populations of patients.

The methods have evolved with the increasing quality of the data. Tractography datasets have increased their size and complexity due to a higher resolution and better image quality, being able to provide a better representation of fiber crossing and small bundles. These advances are also associated to improved algorithms along all the processing pipeline, including artifacts and distortion corrections, diffusion local modeling and fiber tracking. Furthermore, the use of more accurate tractography propagation masks (e. g. based on T1 images [21]) has helped to achieve a better reconstruction of small and superficial white matter (SWM) fibers.

Hence, in the last decade, due to the better quality of dMRI images and processing algorithms, it has been possible to start studying the short association WM bundles. A first whole-brain study used an atlas of gray and white matter to extract short fibers connecting adjacent gyri [22]. Other works combined a hierarchical fiber clustering and cortical parcellation information to extract reproducible short association bundles [13]. A recent study, reported a great amount of short association

bundles, but without a labeling [16]. These works were mostly focused on the creation of SWM atlases. Hence, there is still a need of methods, open to the research community for the study of short association bundles in new databases.

The proposed methods provide an automatic labeling of SMW bundles. First, an efficient individual labeling was implemented. It generates compact clusters and labels them according to cortical a parcellation based on mesh information, for a high quality ROI-based labeling. Furthermore, the bundles connecting each pair of anatomical regions (gyri) are ordered following an axis orientation. The resulting clusters could be used for a fast an easy exploration of short association bundles in individual brains. Without a labelling, its exploration is very complex, since about one thousand clusters are produced for the whole brain.

Subsequently, an inter-subject method has been added, to obtain a correspondence between the clusters (or bundles) across subjects. We tested two methods, a matching, based on the *Hungarian* algorithm, and a clustering method, based on the *QB* algorithm. Even though we used available implementations of both methods, we have adapted them to the processing of labeled intra-subject clusters from different subjects to generate automatically labeled inter-subject clusters.

The results show a better reproducibility across subjects for the *QB* clustering method versus the *Hungarian* algorithm. *Hungarian* algorithm finds a good correspondence between subjects, with low inter-cluster distance, but at expenses of inferior reproducibility. Due to inter-subject variability, and the absence of bundles in some subjects, the one-to-one matching strategy seems not to be directly applicable for this kind of problem. On the other side, the clustering regroups similar bundles on subjects and do not impose the existence of clusters in all the subjects. Indeed previous inter-subject analyses based on clustering have included a reproducibility constraint, e. g. a minimum number of subjects present in the clusters. Hence, an advantage of the proposed labeling is that this reproducibility informa-

tion is easily extracted from the inter-subject labels, what is not the case for classic algorithms. Furthermore, even though the main goal of this work is not a study of reproducibility of SWM bundles, the results of our inter-subject clustering strategy are competitive with state-of-the-art methods, with 94 reproducible bundles for a moderate MDF distance of 21 mm, compared to about 100 hundred bundles obtained for atlases proposed in [13, 14], created with a maximum Euclidean distance of 30 mm.

Note that several factors impact the results, including the quality of the tractography datasets, and the registration strategy. It has been shown that using non-linear registration increases the number of SWM identified [14]. In our experiments we used affine registration to Talairach space, however, other registration algorithms can be applied without problem.

Finally, we highlight some advantages of the proposed methods. First, it is efficient, taking about 3.5 minutes for an intra-subject analysis and about 9 seconds to perform the inter-subject clustering. That is, for the whole inter-subject labeling processing, it takes about 1.17 hours on average. This time is reasonable for an inter-subject analysis. Furthermore, the algorithms are scalable, and can be applied to larger tractography datasets and databases.

The inter-subject labeling can be used to discover patterns of connections in different groups of healthy subjects and patients. The inter-subject clusters can be used to create WM bundle atlases, what require the inspection of experts in anatomy. Note that the clusters form tractography can contain artifacts and false positives [23]. Hence, the proposed algorithm can also contribute to the analysis of tractography datasets for the improvement of tractography methods, through the incorporation of anatomical information and filtering. Finally, other applications include the study of brain connectomes and methods for diffusion-based cortical parcellations.

Conclusions

We implemented a fast method for the automatic labeling of white matter fiber bundles, specifically for the *SWM*, based on an intra-subject clustering and the connectivity of the clusters with the cortical mesh, based on an anatomical ROI atlas. The algorithm also adds a label associated to the relative position of the bundles. Results for intra-subject labeling show a degree of correspondence between subjects, what is further improved with inter-subject labeling. A complete intra-subject labeling is executed on an average time of 3 minutes and 35 seconds for a tractography dataset of about one million fibers. This enables a fast and easy exploration, visualization and analysis of labeled short association bundles in individuals, what is very difficult without any additional information.

Besides, we developed an inter-subject labeling by using two methods. One approach is matching, in particular, the *Hungarian* algorithm, and the other is clustering, employing *QuickBundles* algorithm. The results show a better reproducibility across subjects for the clustering method versus the matching algorithm, while keeping a moderate inter-cluster distance, indicating a good quality of the clusters. Furthermore, the algorithm is scalable and the whole processing for the inter-subject labeling executes on a reasonable time, of about 1.17 hours for 20 subjects. The obtained clusters could be used to perform group-wise connectivity studies, such as the creation of WM bundle atlases, and the development of new methods for the analysis of brain connectome.

Future work will be focused on the application of the method in high quality databases, such as the *Human Connectome Project (HCP)* database, for the creation of a SWM atlas and diffusion-based cortical parcellations.

Methods

Database and tractography datasets

We used 20 subjects of the ARCHI database [18], which was acquired with a 3T MRI scanner (Siemens, Erlangen). The MRI protocol included the acquisition of a T1-weighted dataset using an MPRAGE sequence (160 slices; matrix=256x240; voxel size=1x1x1.1 mm) and a SS-EPI single-shell HARDI dataset along 60 optimized DW directions, b=1500 s/mm^2 (70 slices, TH=1.7 mm, TE=93 ms, TR=14,000 ms, FA=90, matrix=128x128, RBW=1502 Hz/pixel). BrainVISA/Connectomist 2.0 [24] was used for pre-processing the images. The HARDI dataset was corrected for artefacts, geometrical distortions induced by susceptibility effects, eddy currents and motion. The diffusion-weighted (DW) processing pipeline includes the calculation of the analytic q-ball diffusion model [25]. Whole-brain streamline deterministic tractography was calculated with one seed per voxel, from all the voxels of the T1-based propagation mask [21], in forward and backward directions, with a tracking step of 0.2 mm and a maximum curvature angle of 30°. Resulting tractography datasets contain about one million fibers per subject. Affine transformations between subjects' T1 and DW images were also calculated, as well as affine transformations from T1 to Talairach space. The cortical meshes and an automatic labeling of the anatomical regions according to the Desikan-Killiany atlas were obtained using FreeSurfer [26].

Automatic labeling of SWM bundles

To perform the automatic labeling of bundles of superficial white matter, a method consisting of four stages (see Figure 13) was developed, these are: *(1)* fiber clustering, *(2)* cluster filtering, *(3)* fiber intersection and *(4)* cluster labeling.

Stage 1: Fiber clustering

This first stage performs the clustering of a whole-brain tractography dataset, which returns a set of clusters of similar fibers (see Figure 13-(1)). The clustering

Figure 13 Schematics of the labeling method. Stage 1: Fiber clustering. Performs the clustering of the entire tractography. Stage 2: Cluster filtering. Filters out the small clusters and only keep the short bundles, obtained in the previous stage. Stage 3: Fiber intersection. Calculates the fiber bundle intersection with the cortical mesh. Stage 4: Cluster labeling. Renames the clusters based on the two connected regions of the cortex and their relative position.

method [27] is an improved version of an algorithm proposed in [28]. To apply the clustering, fibers must be first resampled with 21 equidistant points, as in [29, 9]. The method consists of 4 steps: *(1) Building clusters on a subset of fiber points*, where mini batch K-means is applied in parallel on a subset of fiber points, obtaining groups of points; *(2) Generating preliminary clusters*, which groups fibers sharing the point cluster labels from the previous step; *(3) Defining candidate clusters by reassigning small preliminary clusters*: reassigns small clusters to larger clusters based on a maximum distance threshold between clusters; *(4) Computing final clusters by merging close candidate clusters*: merges close clusters that share the central label obtained from step 1, according to a criterion of maximum Euclidean distance between clusters. Finally, a representative fiber of each cluster is selected, as its centroid, and resampled it with 21 equidistant points.

Stage 2: Cluster filtering

The second stage filters out the small and long fiber clusters (see Fig. 13-(2)). Clusters are denoted as C_i, with $i = 1, ..., n$ the index of the cluster. The filter receives a minimum size of the cluster $min_{nf}(C_i)$ (number of fibers), to remove small fibers, and a minimum $min_{len}(C_i)$ and maximum cluster length $min_{len}(C_i)$ to only keep short fibers whithin a reasonable range. The length of each cluster is measured by means of the Euclidean distance between two adjacent points of the cluster's centroid.

Stage 3: Fiber intersection

This step calculates the intersection of the fibers with the cortical mesh, based on the algorithm proposed in [30] (see Fig. 13-(3)). The method first performs a

subdivision of $3D$ space into 1.5 mm size cells, which speeds up searches in the mesh. Next, each fiber endpoint is projected one point backward and two points forward to extend the search along the fiber trajectory on ending points and avoid missing intersections. All cells that include these points and their neighboring cells are selected. Finally, the intersection point of each fiber extremity with the cortical mesh triangles is determined using Möller-Trumbore equation [31], based on the analysis of the triangles contained in the selected cells.

The intersection algorithm is given by Equation 1:

$$O + tD = (1 - u - v)V_0 + uV_1 + vV_2 \tag{1}$$

where (u, v) are the exact coordinates of intersection with the mesh triangle, V_0, V_1 and V_2 are the vertices of a triangle, t is the direction, D is the normalized ray trajectory, and O is the ray from the point of origin.

Finally, for each hemisphere, the indexes of the start (Tri_{init}) and end triangles (Tri_{end}) where the fiber intersects the mesh are obtained, as well as the coordinates of the two exact points of initial ($Point_{init}$) and final ($Point_{end}$) intersection.

Stage 4: Cluster labeling

This stage labels all the clusters based on the cortical regions they connect, by using a cortical ROI atlas. For testing, we use the Desikan-Killiany atlas [32], consisting of 35 regions (gyri) per hemisphere. We use the cortical meshes, containing a list of vertices and triangles, and a labeling file, containing the cortical region label of each mesh vertex.

First, for each cluster, each fiber is labeled according to the triangle of the mesh that the fiber intersects, based on the region labels of the triangle vertices. The labeling of each triangle is defined as the most repeated label between its three vertices (see Fig. 13-(4.1.)). Next, the fibers require to be aligned, since in a tractography dataset, there is no a unique direction and fibers can be stored in direct

or inverse direction. Since after the clustering the fibers are grouped on compact clusters, these can be aligned so that the starting and ending points have the same orientation in a cluster (see Figure 13-(4.2.)). Hence, the fibers of a cluster are oriented based on the cluster centroid. To perform the alignment, we verify if the fibers are inverted with respect to the centroid. We denote f_i as the fiber i of the bundle, with $i = 1, ..., n$, and the centroid of the bundle as c_j, with $j = 1, ..., m$. Then, the Euclidean distance (d_E) is calculated between the first point of the fiber (f_{i1}) and both endpoints of the centroid (c_{i1} to c_{j21}). If $d_E(f_{i1}, c_{j1}) > d_E(f_{i1}, c_{j21})$, the fiber is inverted by flipping its fiber points.

Figure 14 Fiber bundle alignment by respect to its corresponding bundle centroid. The Euclidean distance (d_E) is calculated between the first point of the fiber (f_{i1}) and both end points of the centroid (c_{i1} to c_{j21}). If $d_E(f_{i1}, c_{j1}) > d_E(f_{i1}, c_{j21})$, the fiber is reversed.

Next, each cluster (or bundle) is labeled according to the most connected regions. For each bundle, the labels of both bundle extremities, i. e. the beginning and end of each bundle, are processed separately. The most common label (mcl) for bundle start (mcl_{init}) and end (mcl_{end}) is determined and used to name each bundle, with format mcl_{init}-mcl_{end} (see Figure 13-(4.3.)). For instance, a bundle connecting the post-central and pre-central anatomical regions will have the label *PoC-PrC*. Note that several bundles may connect the same pair of anatomical regions (gyri), as each cluster extremity only intersects a portion of a gyrus. Then, an order is defined for each pair of bundles defined by the index of the regions in the cortical region label file. For example, *PrC* has index 24 and *PoC* has index 22, then, the bundle is named as *PoC-PrC*. Subsequently, bundles with inverted names are flipped. For example, all the bundles labeled with *PrC-PoC* are inverted and named as *PoC-PrC*. Finally, as several bundles may have the same name, but connecting different specific sub-regions of the gyri, these are labeled with an extra index, indicating the relative position according to an axis in the brain. The intersection point of all

the bundle centroid in a gyrus are sorted based on a spatial coordinate (by default coordinate y), in ascending or descending order (by default ascending). Note that a particular axis and order can be given for each gyri.

Inter-subject labeling

In this section two methods are presented to obtain group-wise bundle labels of superficial white matter bundles, among the subjects of a population. Intra-subject labeling, presented in the previous section, labeled the bundles of a subject based on the connected brain regions individually, and an order based on the coordinates, producing a certain similarity between the subjects' bundles. However, this was not the main objective of the intra-subject labeling method and the correspondence between subjects can be improved applying inter-subject methods. The methods used to perform this processing are a matching algorithm and a clustering algorithm.

Matching algorithm for inter-subject labeling

The aim of this step is applying a matching [33] for finding a correspondence between similar bundles in the different subjects. Bipartite matching algorithms find correspondence between pairs of elements from two distinct sets. These algorithms are based on graph theory to find connections in two sets of vertices, where vertices in one set must match with vertices in the other set [34].

A well-known algorithm for a bipartite matching problem, is the *Hungarian* algorithm, that solves the minimum weight matching, i. e. the minimum distance between vertices from the two sets, A and B [19]. Being V the number of vertices from the two sets, the algorithm receives a matrix M, containing the distances between the vertices from the two sets. In our application, V is the total number of bundles from a pair of subjects and matrix M contains the distances between the bundle centroids from the two subjects, being one set represented in the rows, and the other in the columns. The original algorithm performs a perfect matching, i. e. each vertex (or bundle) in set A is matched with a vertex in set B, what requires

an equal number of vertices in both sets and produces a square matrix M. Our problem presents a different number of vertices in each set, as different number of bundles are found in each subject. Hence, we used an adapted algorithm that performs the analysis over non-squared matrices and leaves unmatched the most dissimilar elements.

More formally, each element $M[i, j]$ in matrix M represents the distance between bundle i of set A (subject A) and bundle j of set B (subject B), being the cost of matching between the two vertices. The result is an assignment of the elements of set A with set B by using the minimum assignment cost. The distance used is the minimum average direct-flip distance (MDF) between two pairs of fibers [20] (Equation 2), a distance commonly used for tractography fiber comparison.

$$\begin{aligned} d_{direct}(a,b) = d(a,b) &= \frac{1}{K}\sum_{i=1}^{K}|a_i - b_i| \\ d_{flipped}(a,b) &= d(a,b^F) = d(a^F,b) \\ MDF(a,b) &= min(d_{direct}(a,b), d_{flipped}(a,b)) \end{aligned} \tag{2}$$

The Hungarian algorithm has a complexity of $\mathcal{O}(V^3)$, however, as we perform the analysis separately for each pair of anatomical regions, the analyzed datasets are small with a low execution time.

The matching algorithm applied to inter-subject bundle labeling, first performs a bundle pre-processing. For each subject, previously labeled bundles with the intra-subject labeling, are separated into different groups depending on the pair of anatomical regions they connect. Then, for each region a map is created, whose key is the subject and the value is a list of the bundles that belong to the subject and region. For instance, for region *PoC-PrC* the bundles for *Subject001*, will be

stored in the key-value pair: *Subject001: [bundle0, bundle1, ..., bundleN]*. Next, the algorithm consists of four steps:

- *Step 1*. Once the maps of all the regions are obtained, the bundles of each region are processed sequentially. First, the subjects are ordered from highest to lowest, based on the number of bundles they contain. For each bundle, its centroid is calculated using the mean of the streamline point coordinates.
- *Step 2*. The analysis begins with the first subject in the list as a reference subject. This subject is compared with each of the following subjects using the *Hungarian* algorithm, receiving as input the distance matrix. This returns a matching based on the distance of one bundle centroid with another. The *Hungarian* algorithm receives as input the matrix of distances, which are calculated using the MDF distance (Equation 2) between all the bundle centroid pairs of all subjects. For each bundle, the algorithm returns the bundle that best matches it, according to the solution of the minimization problem. However, the distance between a pair of bundles could be higher, hence, the method evaluates all the distances between the matched bundle centroids and only keeps the pairs of bundles which distances do not exceed the established maximum distance threshold. This avoids the assignment of distant bundles, leaving them unassigned. Bundles that match each other are labeled with the same indexes, based on the label of the reference subject. For example, for two corresponding bundles, they would be called *PoC-PrC_0* even if they are from different subjects.
- *Step 3*. Two cases can happen with unassigned bundles: *(i) Bundles of the reference subject*. They are not similar to any other bundle in the dataset and they are labeled with a new index. *(ii) Bundles of the remaining subjects*. The bundles are stored. In the iteration in which the subject is taken as a reference, comparisons are made again with the rest of the subjects.

- *Step 4*. Repeat *Step 2* with the unassigned bundles of the following subjects, taking as reference the next subject in the list with unassigned bundles.

Figure 15 shows an example scheme of the algorithm for three subjects and the bundles connecting *PoC-PrC* regions. Each circle corresponds to a bundle. First, the subjects are ordered from highest to lowest number of bundles (see Figure 15-(1.)). Second, *Subject001* that is being compared with the rest is the reference. It is compared with the *Subject002* and does match only the first two bundles, leaving an unassigned bundle in *Subject002*, which will be saved for later comparison (see Figure 15-(2.)). Third, *Subject001* continues to be compared with the remaining subjects, in this case, with *Subject003*, what leads to the matching of bundles 1 and 2. In *Subject003* there remains an unassigned bundle (see Figure 15-(3.)). Finally, once *Subject001* is compared with all subjects, the reference subject becomes the next one, in this case, *Subject002*. Then, unassigned bundles are compared, for example, *Subject002* is compared with *Subject003* and the two unassigned bundles are matched (see Figure 15-(4.)). The bundle with the highest reproducibility in the example is the 1, since it is present in all subjects, and would be named as *PoC-PrC_1*, according to the label of the first reference subject of the bundle.

We used the implementation of the Hungarian algorithm available at scipy library [35].

Figure 15 Schematics of the Hungarian algorithm for inter-subject labeling of bundles connecting PoC-PrC regions. First, the bundles are ordered from highest to lowest number of bundles. Second, the reference subject, *Subject001*, is compared to *Subject002*, leaving unassigned bundles. Third, it continues comparing to the rest of the subjects. Finally, the reference passes to the next subject with unassigned bundles, *Subject002* and these are compared with the rest of the subjects. This process is repeated until all subjects are analyzed.

Clustering algorithm for inter-subject labeling

Clustering is an unsupervised classification method, which groups similar elements into subsets called clusters. Each cluster is made up of elements that have similar

characteristics, however, the elements of each cluster are different from each other [36].

The clustering method used to group the clusters is a well-known fiber clustering algorithm called *QuickBundles (QB)* [20]. *QB* is a clustering method specialized in grouping white matter fibers from tractography datasets quickly and with good quality. This unsupervised clustering algorithm groups the fibers into clusters, without recalculating the clusters, like classical methods such as K-means. The algorithm uses a distance threshold to define whether a new fiber will be assigned to the closest cluster or will start a new cluster. The algorithm has a single parameter, which is the minimum average direct-flip distance (MDF) between two pairs of fibers. It is one of the fastest methods that exist today, with runtime $\mathcal{O}(N^2)$, being N the size of the dataset.

Before applying *QB*, we apply the same bundle pre-processing as for the matching, to create a map for each pair of regions, with the bundles of each subject. Next, the *QB* algorithm is performed sequentially to each pair of regions. For each pair of regions, and all the subjects, the centroids of all clusters are calculated. The algorithm is applied to the complete set of clusters, i. e. from all subjects for the pair of regions. Once the inter-subject clusters are obtained, all intra-subject clusters belonging to the same inter-subject cluster are labeled with the same label. If several clusters of the same subject belong to the same inter-subject cluster, they are merged.

Figure 16 shows an example scheme for *QB* application to three subjects on the *PoC-PrC* regions. First, it starts with the computation of all the cluster centroids. Unlike matching, in this case it is not necessary that the clusters are ordered (see Figure 16-(1.)). Second, the *QB* method is applied to all clusters, generating inter-subject clusters. Bundles within inter-subject clusters are labeled with the same name (see Figure 16-(2.)). Finally, the clusters of a subject that are in the same

inter-subject clusters are merged (see Figure 16-(3.)). The clusters with the highest reproducibility are 0 and 3, since they appear in all subjects, whose tags would be: *PoC-PrC_0* and *PoC-PrC_3*. In addition, there may be some loose cluster, which will be individually labeled with another index.

Figure 16 Schematics of the QB algorithm for labeling inter-subject bundles for PoC-PrC regions. First, the cluster centroids are computed. Second, *QB* is applied to all the intra-subject clusters, to obtain inter-subject clusters. Bundles belonging to an inter-subject cluster are labeled using the same name. Finally, clusters of the same subject that belong to the same inter-subject cluster are merged.

Statistical analysis

Histograms have been used to evaluate the reproducibility of inter-subject bundles (clusters) in terms of the number of subjects in where they are found, for the two tested methods and different distance thresholds. The greater the number of subjects, the more reproducible the bundle is. Another histogram displays the number of inter-subject bundles obtained for different inter-cluster distance ranges. The greater the number of clusters with lower distance, the more precise the classification of the method is.

On the other hand, heatmaps were created to visually evaluate the reproducibility and variability of the bundles in the different subjects. First, the presence or absence of a bundle in a subject can be observed. Also, the normalized number of fibers is displayed using a heat colormap. The stronger the color indicates a higher number of fibers in the bundle.

Simple averages were used to compute the execution times over the different subjects.

Abbreviations

SWM: Superficial White Matter; DWM: Deep White Matter; WM: White Matter; QB: QuickBundles; PoC: Postcentral; PrC: Precentral; SM: Supra-marginal; RMF: Rostral Middle Frontal; SF: Superior Frontal; SP: Superior Parietal; IP: Inferior Parietal; Tr: Pars Triangularis; dMRI: Diffusion-weighted Magnetic Resonance Imaging; MRI:

Magnetic Resonance Imaging; HARDI: High Angular Resolution Diffusion Imaging; HCP: Human Connectome Project; ROI: Region of Interest; MDF: Minimum average Direct-Flip.

Acknowledgements

Not applicable.

Author's contributions

AV, NLL and PG designed the main research idea. AV, NLL and PG wrote the manuscript. CP provided the pre-processed ARCHI database. JFM, SL and PG provided guidance on the implementation and evaluation of the algorithms. JH provided guidance on the overall work based on the main aplications of the proposed method. AV and NLL wrote the main analysis code and performed all the experiments. All authors read and approved the final manuscript.

Funding

This work has received funding by the CONICYT PFCHA/ DOCTORADO NACIONAL/2016-21160342, CONICYT FONDECYT 1190701, CONICYT PIA/Anillo de Investigación en Ciencia y Tecnología ACT172121, Basal Center FB0008, and by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 690941. This work was also partially funded by the Human Brain Project, funded from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 785907 (SGA2) and No: 604102 (SGA1).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author upon request.

Ethics approval and consent to participate

The project was approved by the Local Ethical Committee, "Comité de Protection des Personnes Ile-de-France VII", with codes CPP100002/CPP10002, and all subjects signed an informed consent before inclusion.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests in this section.

Author details

[1] Universidad de Concepción, Faculty of Engineering, Concepción, Chile. [2]Universidade da Coruña, Centro de investigación CITIC, A Coruña, Spain. [3]NeuroSpin, CEA, Paris-Saclay University, Gif-sur-Yvette, France. [4]INSERM U955 Unit, Mondor Institute for Biomedical Research, Team 15 "Translational Psychiatry", Créteil, France. [5]Fondation Fondamental, Créteil, France. [6]AP-HP, Department of Psychiatry and Addictology, Mondor University Hospitals, School of Medicine, DHU PePsy, Créteil, France.

References

1. Le Bihan, D., Iima, M.: Diffusion Magnetic Resonance Imaging: what water tells us about biological tissues. PLOS Biology **13**(7), 1002203 (2015)
2. Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance, Series B **103**(3), 247–254 (1994)

3. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. Magnetic Resonance in Medicine **44**(4), 625–632 (2000)
4. Catani, M., Dell'Acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., Valabregue, R., De Schotten, M.T.: Short frontal lobe connections of the human brain. Cortex **48**(2), 273–291 (2012)
5. Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., Westin, C.-F.: The White Matter Query Language: a novel approach for describing human white matter anatomy. Brain Structure and Function **221**(9), 4705–4721 (2016)
6. O'Donnell, L.J., Westin, C.-F.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Transactions on Medical Imaging **26**(11), 1562–1575 (2007)
7. O'Donnell, L.J., Suter, Y., Rigolo, L., Kahali, P., Zhang, F., Norton, I., Albi, A., Olubiyi, O., Meola, A., Essayed, W.I., *et al.*: Automated white matter fiber tract identification in patients with brain tumors. NeuroImage: Clinical **13**, 138–153 (2017)
8. Garyfallidis, E., Côté, M.-A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., Descoteaux, M.: Recognition of white matter bundles using local and global streamline-based registration and clustering. NeuroImage **170**, 283–295 (2018)
9. Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Fillard, P., Le Bihan, D., Leboyer, M., Houenou, J., Mangin, J.-F.: Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. NeuroImage **61**(4), 1083–1099 (2012). doi:10.1016/j.neuroimage.2012.02.071
10. Labra, N., Guevara, P., Duclap, D., Houenou, J., Poupon, C., Mangin, J.-F., Figueroa, M.: Fast automatic segmentation of white matter streamlines based on a multi-subject bundle atlas. Neuroinformatics **15**(1), 71–86 (2017)
11. Vázquez, A., López-López, N., Labra, N., Figueroa, M., Poupon, C., Mangin, J.-F., Hernández, C., Guevara, P.: Parallel Optimization of Fiber Bundle Segmentation for Massive Tractography Datasets. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 178–181 (2019). IEEE
12. Ros, C., Güllmar, D., Stenzel, M., Mentzel, H.-J., Reichenbach, J.R.: Atlas-guided cluster analysis of large tractography datasets. PLoS ONE **8**(12), 83847 (2013)
13. Guevara, M., Román, C., Houenou, J., Duclap, D., Poupon, C., Mangin, J.-F., Guevara, P.: Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography. NeuroImage **147**, 703–725 (2017)
14. Román, C., Guevara, M., Valenzuela, R., Figueroa, M., Houenou, J., Duclap, D., Poupon, C., Mangin, J.-F., Guevara, P.: Clustering of whole-brain white matter short association bundles using HARDI data. Frontiers in Neuroinformatics **11**, 73 (2017)
15. O'Donnell, L.J., Golby, A.J., Westin, C.-F.: Fiber clustering versus the parcellation-based connectome. NeuroImage **80**, 283–289 (2013)
16. Zhang, F., Wu, Y., Norton, I., Rigolo, L., Rathi, Y., Makris, N., O'Donnell, L.J.: An anatomically curated fiber clustering white matter atlas for consistent white matter tract parcellation across the lifespan. NeuroImage **179**, 429–447 (2018)
17. Labeling. https://github.com/andvazva/Labeling.git/ Accessed 27/12/19
18. Schmitt, B., Lebois, A., Duclap, D., Guevara, P., Poupon, F., Rivière, D., Cointepas, Y., LeBihan, D., Mangin, J.-F., Poupon, C.: CONNECT/ARCHI: an open database to infer atlases of the human brain connectivity. In: ESMRMB (2012)
19. Frank, A.: On Kuhn's Hungarian method—a tribute from Hungary. Naval Research Logistics (NRL) **52**(1), 2–5 (2005)

20. Garyfallidis, E., Brett, M., Correia, M.M., Williams, G.B., Nimmo-Smith, I.: Quickbundles, a method for tractography simplification. Frontiers in Neuroscience **6**, 175 (2012)
21. Guevara, P., Duclap, D., Marrakchi-Kacem, L., Rivière, D., Cointepas, Y., Poupon, C., Mangin, J.: Accurate tractography propagation mask using T1-weighted data rather than FA. In: Proceedings of the International Society of Magnetic Resonance in Medicine, p. 2018 (2011)
22. Zhang, Y., Zhang, J., Oishi, K., Faria, A.V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G.B., Evans, A., Toga, A.W., Woods, R., Mazziotta, J.C., Miller, M.I., van Zijl, P.C.M., Mori, S.: Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. NeuroImage **52**(4), 1289–1301 (2010)
23. Maier-Hein, K.H., Neher, P.F., Houde, J.-C., et al.: The challenge of mapping the human connectome based on diffusion tractography. Nature Communications **8**(1349) (2017). doi:10.1038/s41467-017-01285-x
24. BrainVISA. http://brainvisa.info/web/index.html Accessed 27/12/19
25. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-ball imaging. Magnetic Resonance in Medicine **58**(3), 497–510 (2007)
26. FreeSurfer. https://surfer.nmr.mgh.harvard.edu/fswiki Accessed 27/12/19
27. Vázquez, A.: Método eficiente de clustering de fibras cerebrales basado en distribución de puntos. Master's thesis in computer science, Universidad de Concepción, Concepción (2019)
28. Sanchez, A., Hernández, C., Poupon, C., Mangin, J.-F., Guevara, P.: Clustering of tractography datasets based on streamline point distribution. In: International Society of Magnetic Resonance in Medicine Conference (2018). ISMRM 2018
29. Guevara, P., Poupon, C., Rivière, D., Cointepas, Y., Descoteaux, M., Thirion, B., Mangin, J.-F.: Robust clustering of massive tractography datasets. NeuroImage **54**(3), 1975–1993 (2011)
30. Silva, F., Guevara, M., Poupon, C., Mangin, J.-F., Hernández, C., Guevara, P.: Cortical Surface Parcellation Based on Graph Representation of Short Fiber Bundle Connections. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1479–1482 (2019). IEEE
31. Möller, T., Trumbore, B.: Fast, minimum storage ray/triangle intersection. In: ACM SIGGRAPH 2005 Courses, p. 7 (2005). ACM
32. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., *et al.*: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage **31**(3), 968–980 (2006)
33. Monge, A.E., Elkan, C., *et al.*: The Field Matching Problem: Algorithms and Applications. In: KDD, vol. 2, pp. 267–270 (1996)
34. Wilson, R.J.: Introduction to Graph Theory. John Wiley & Sons, Inc., New York, NY, USA (1986)
35. Solve the Linear Sum Assignment Problem. https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.optimize.linear_sum_assignment.html Accessed 27/12/19
36. Xu, R., Wunsch, D.C.: Survey of clustering algorithms (2005)

Figures

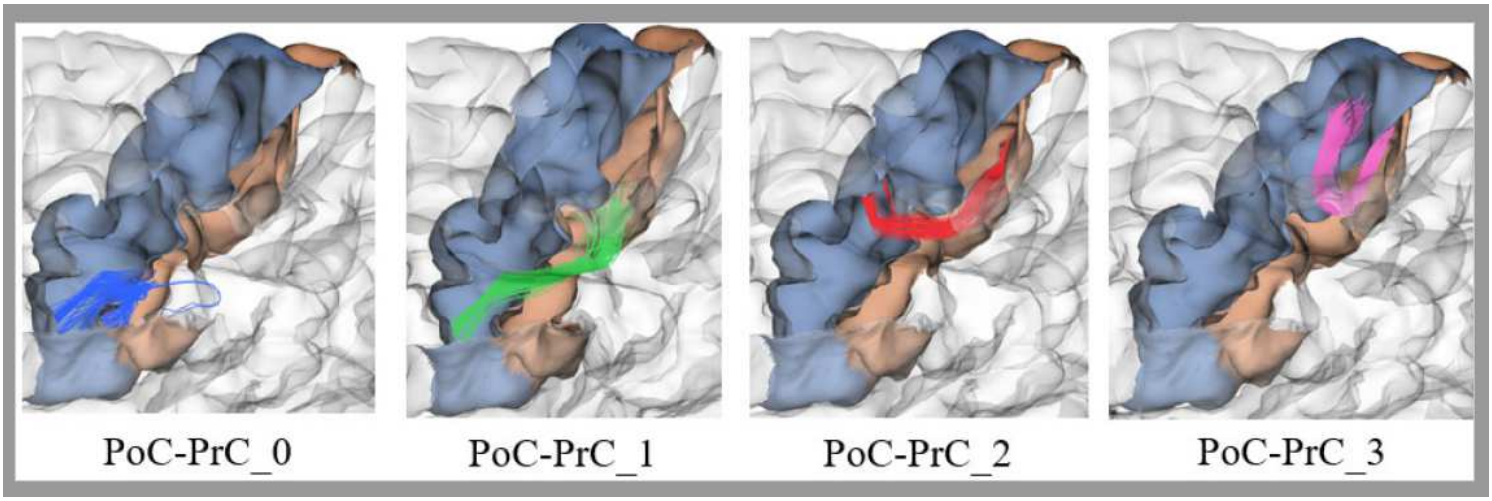

Figure 1

Bundles connecting right PoC and PrC regions. Example of bundle labeling according to the relative position of the bundles connecting PoC and PrC regions for Subject001.

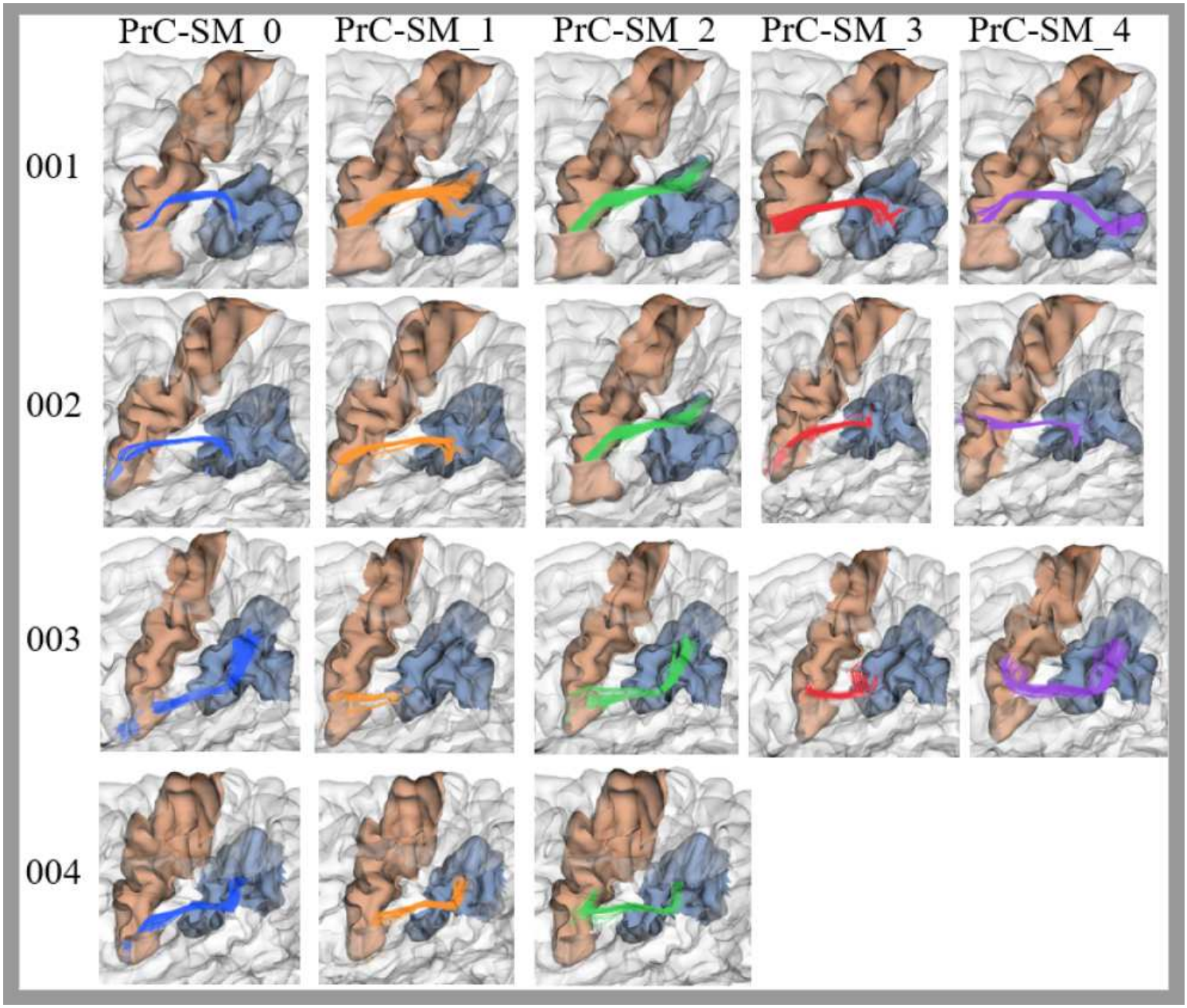

Figure 2

Correspondence of intra-subject bundle labels across subjects. Comparison of the first five bundles from four subjects (001-004), connecting PrC and SM gyri.

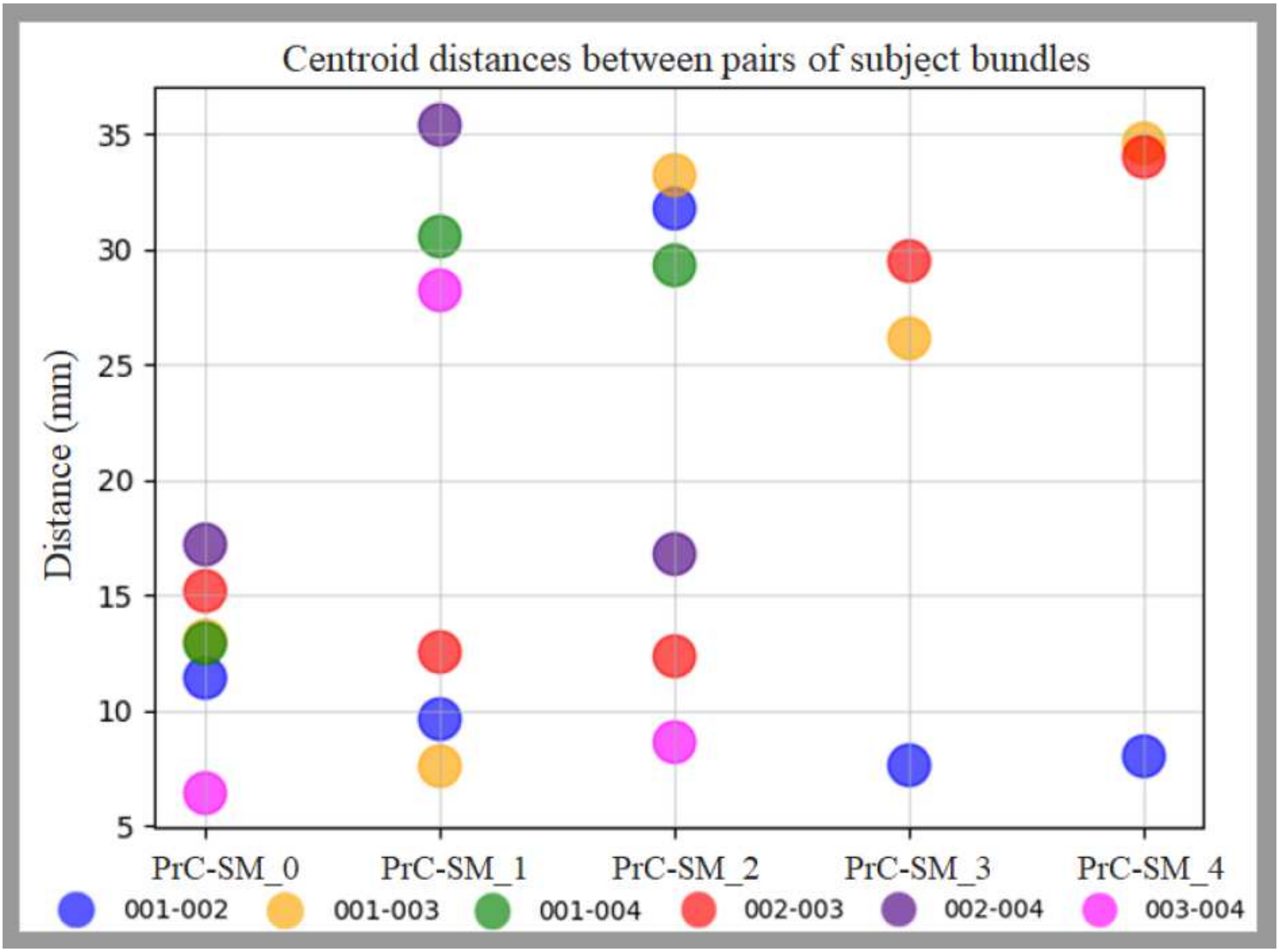

Figure 3

Bundle centroid distances between pairs of subjects for intra-subject labeling of four subjects. Distances in mm between bundle centroids for all the pairs of subjects, for five bundles connecting PrC and SM gyri. The colors represent the different pairs of subjects.

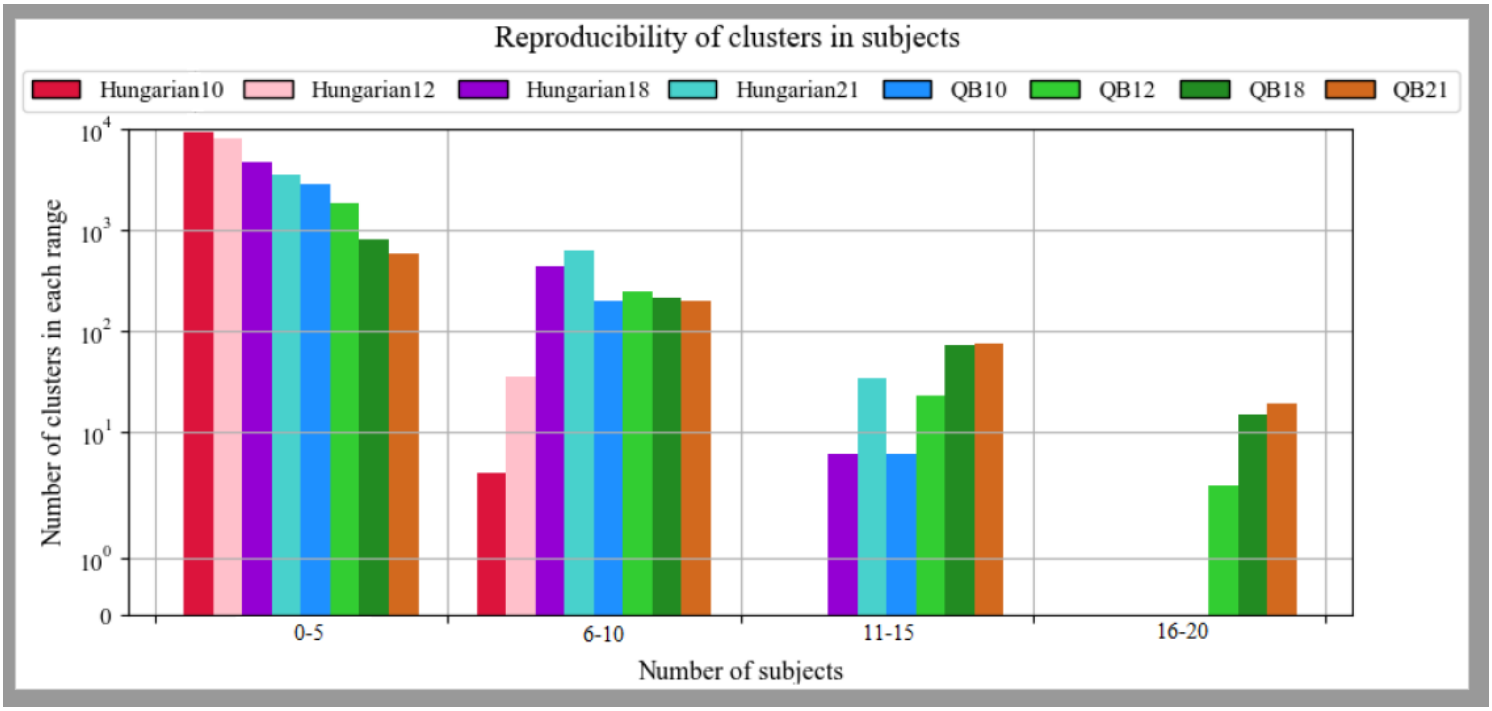

Figure 4

Reproducibility of bundles with inter-subject labeling for the two methods. The number of subjects is shown on the x-axis while the y-axis shows the number of clusters in each range.

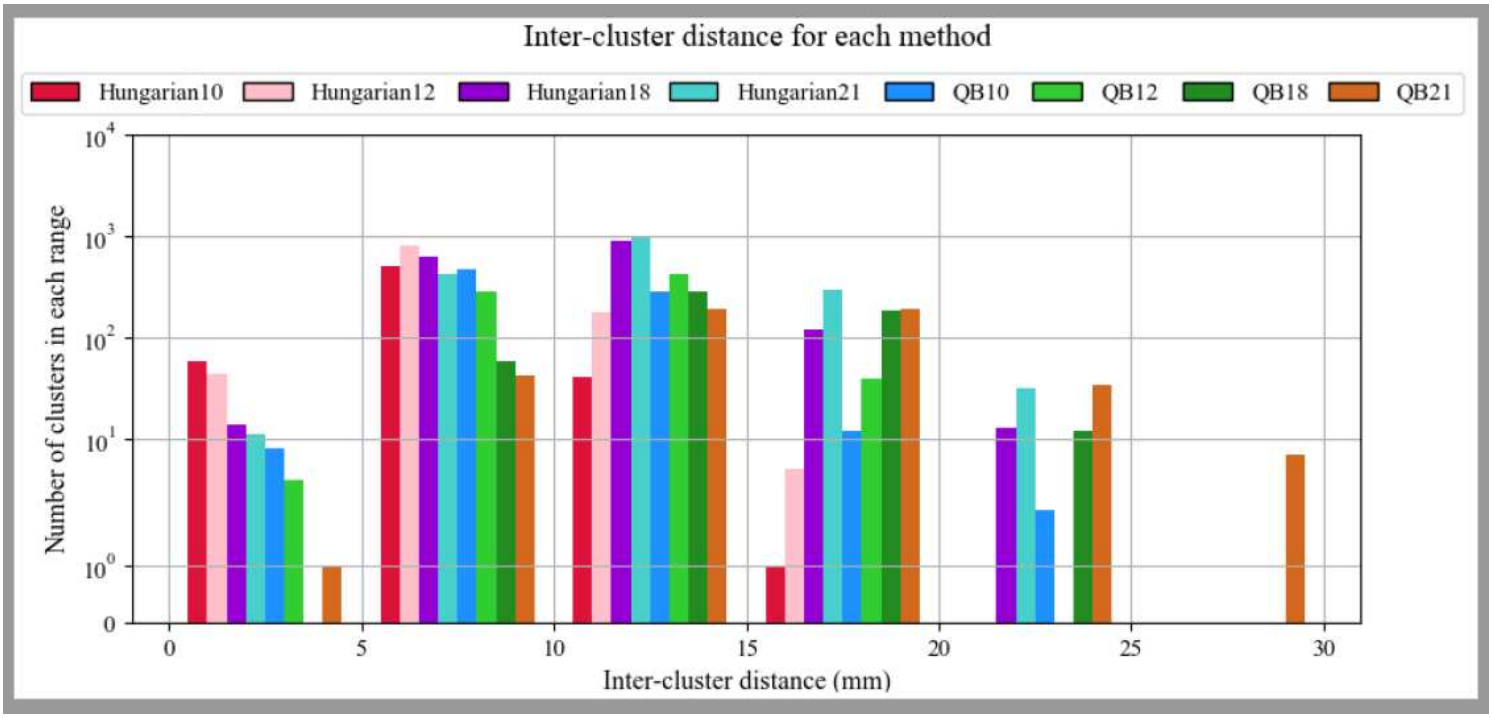

Figure 5

Inter-cluster bundle distance for both inter-subject labeling method. X-axis represents the inter-cluster distance measured in mm. Y-axis shows the number of clusters in each range.

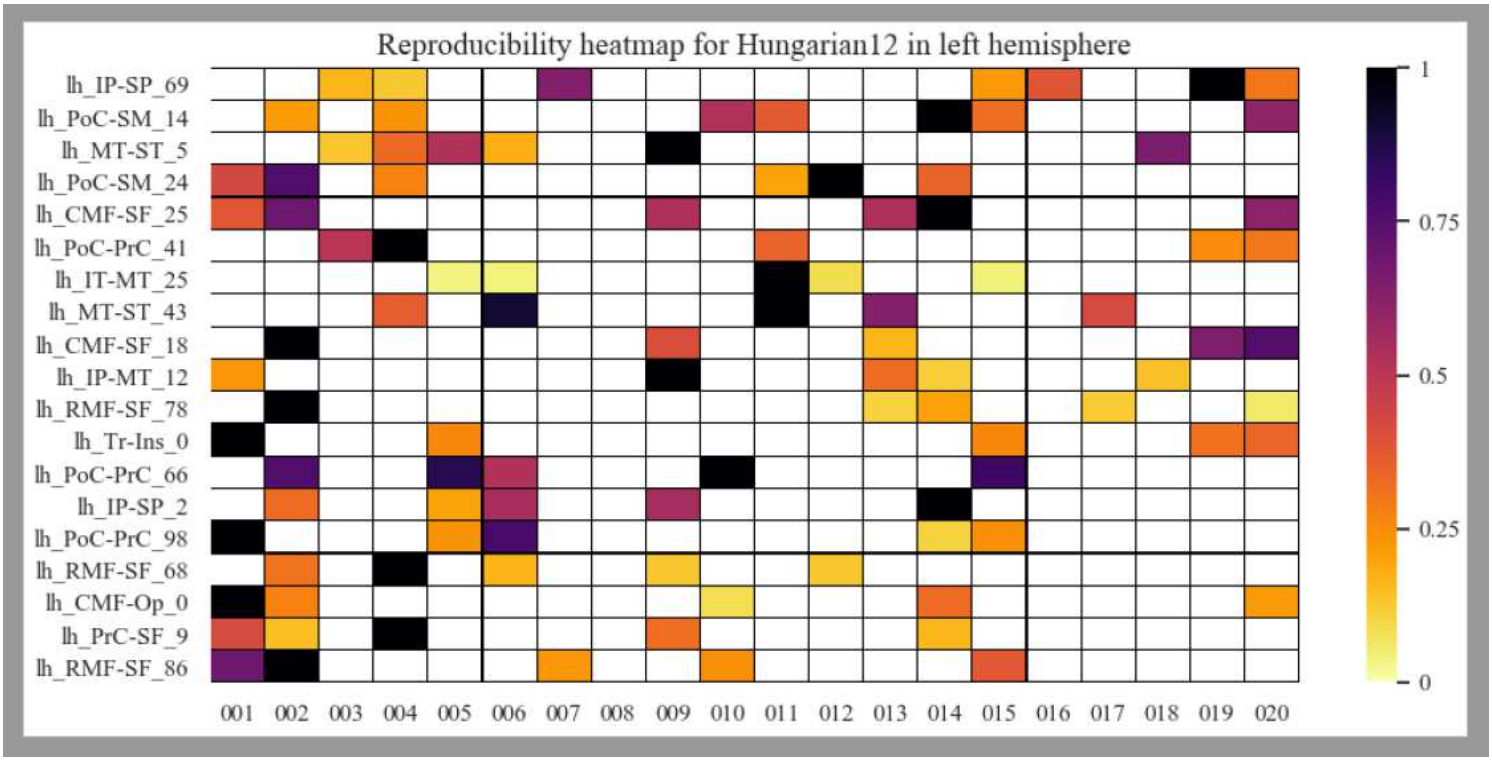

Figure 6

Reproducibility heatmap for Hungarian algorithm with threshold 12 mm, for the left hemisphere. On the x-axis are the subjects, on the y-axis are the 20 most reproducible bundles. The greater the number of fibers, the darker the color of the box on the heatmap that is normalized between 0 and 1.

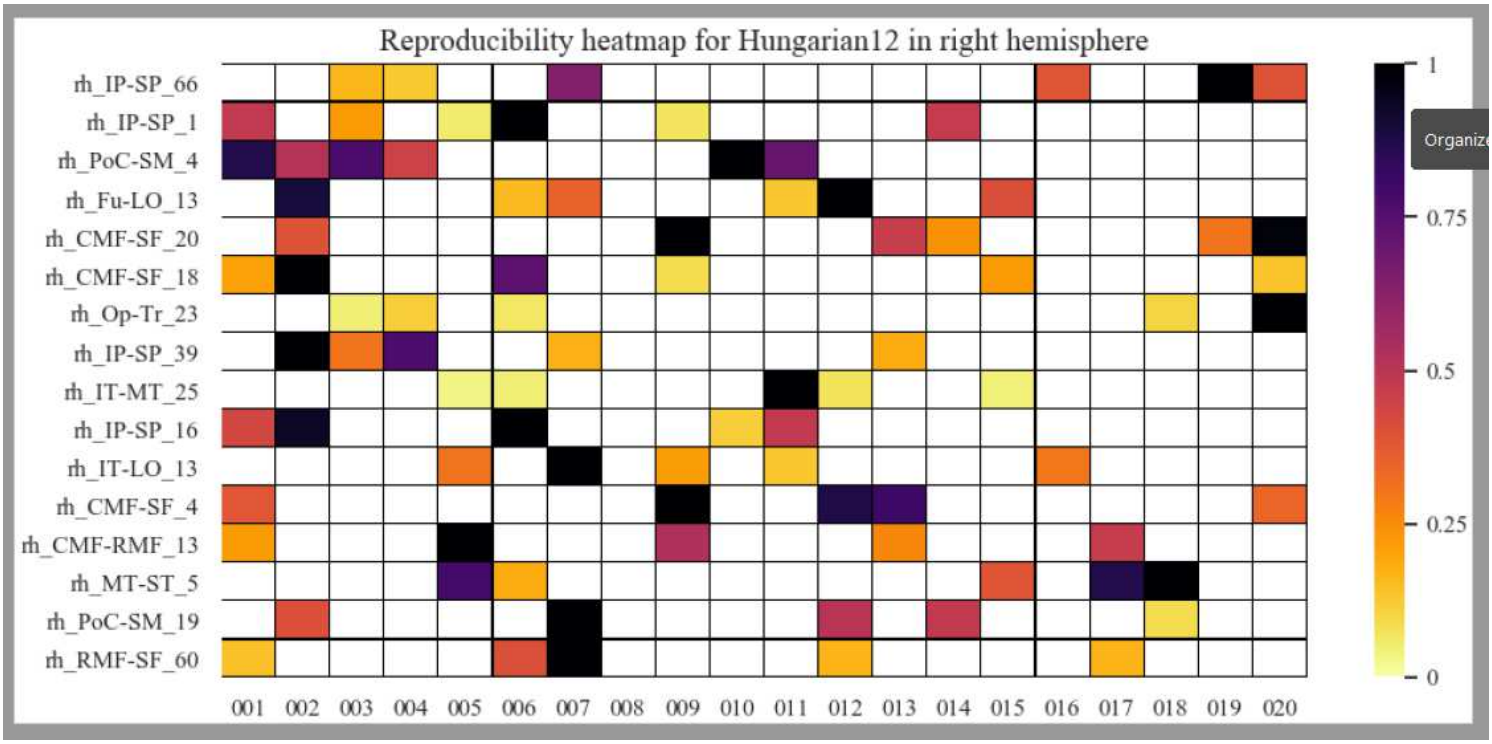

Figure 7

Reproducibility heatmap for Hungarian algorithm with threshold 12 mm, for the right hemisphere. X-axis displays the subjects used, the 20 most reproducible bundles are shown on the y-axis. The darker boxes indicate a higher concentration of fibers in the bundle. These values are normalized between 0 and 1.

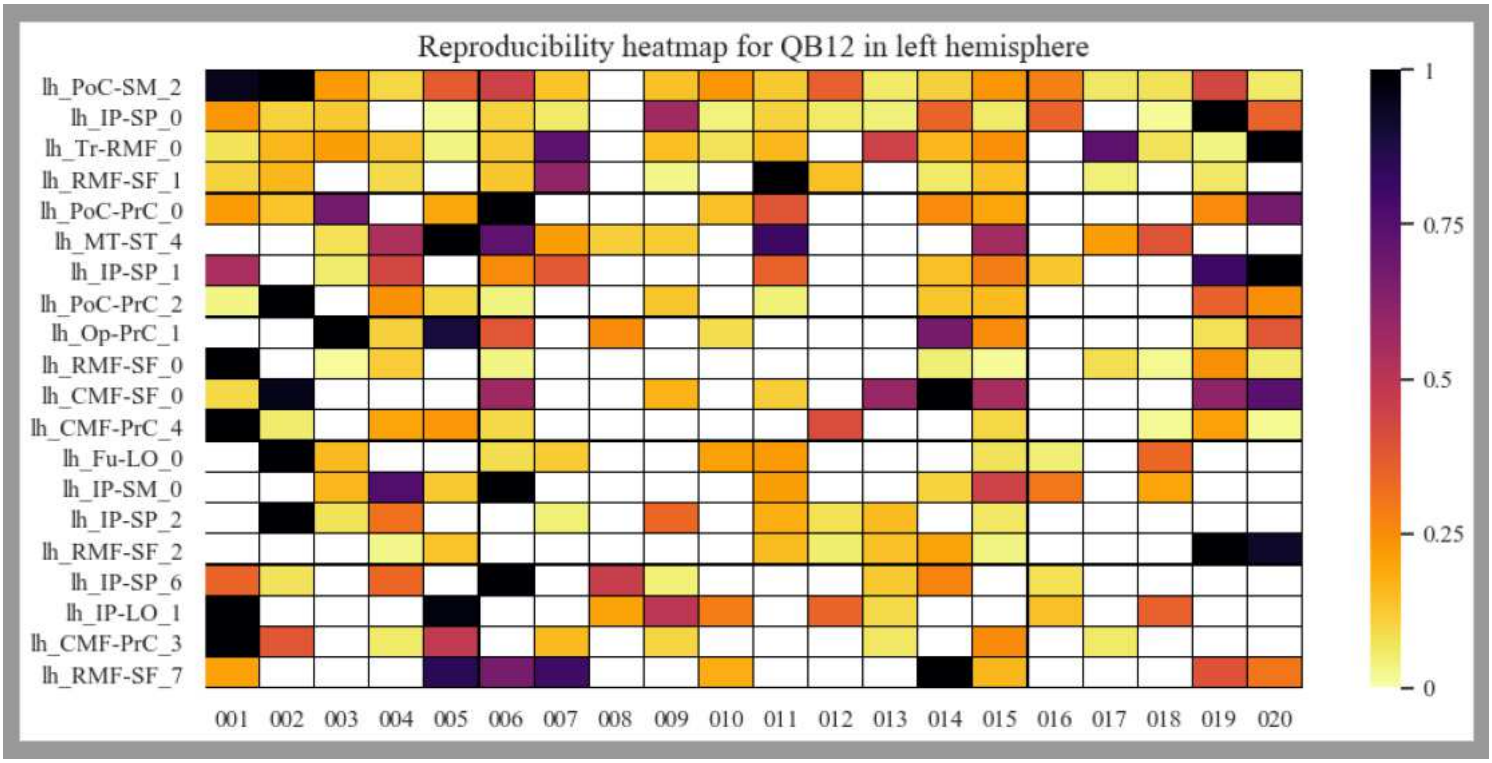

Figure 8

Reproducibility heatmap for QB with threshold 12 mm, for the left hemisphere. X-axis shows the subjects, while the y-axis shows the 20 most reproducible bundles among subjects in the left hemisphere. Darker boxes show bundles with more fibers in them. White boxes show absence of the bundle in the determined subject. The heat bar shows the values of normalized fibers between 0 and 1.

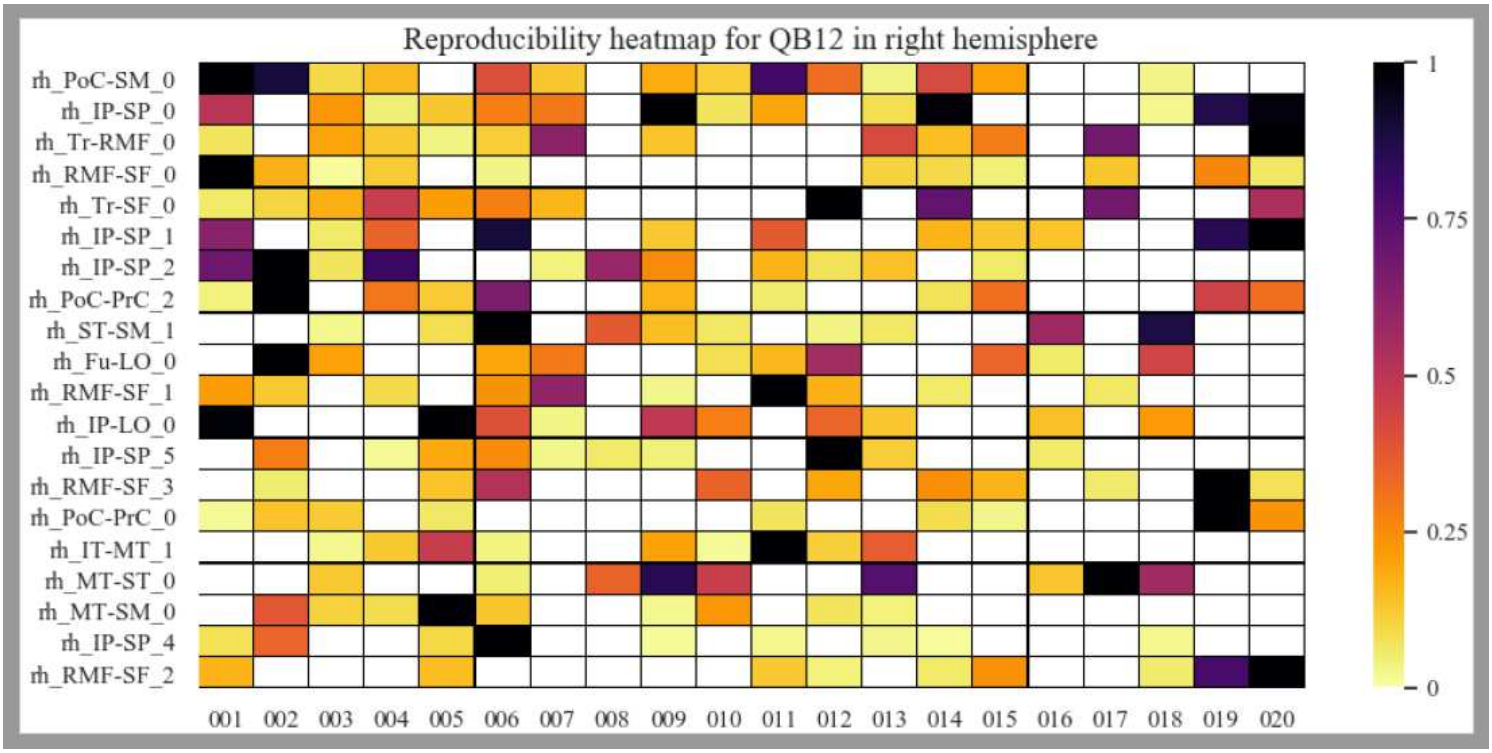

Figure 9

Reproducibility heatmap for QB with threshold 12 mm, for the right hemisphere. X-axis displays the subjects, and on the y-axis appears the 20 most reproducible bundles. The lighter the color of the box, the fewer fibers it contains. If the box is white, it indicates the absence of the bundle in the subject. The fiber values appear normalized between 0 and 1 in the heat bar.

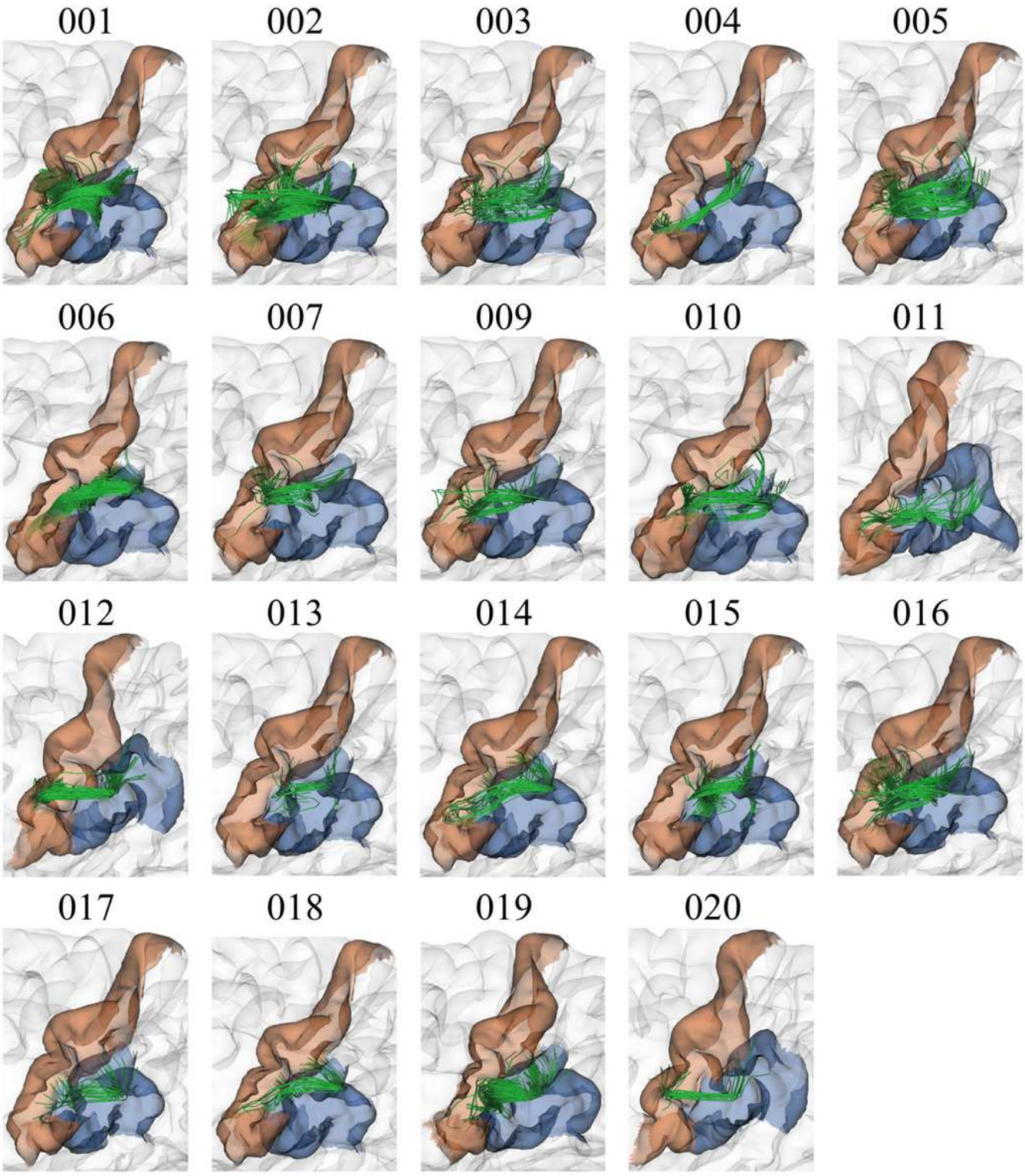

Figure 10

Bundle lh PoC-SM 2, with the highest reproducibility in all subjects. The results show good reproducibility among subjects, appearing in 19 of the 20 subjects for the QB method with 12mm threshold.

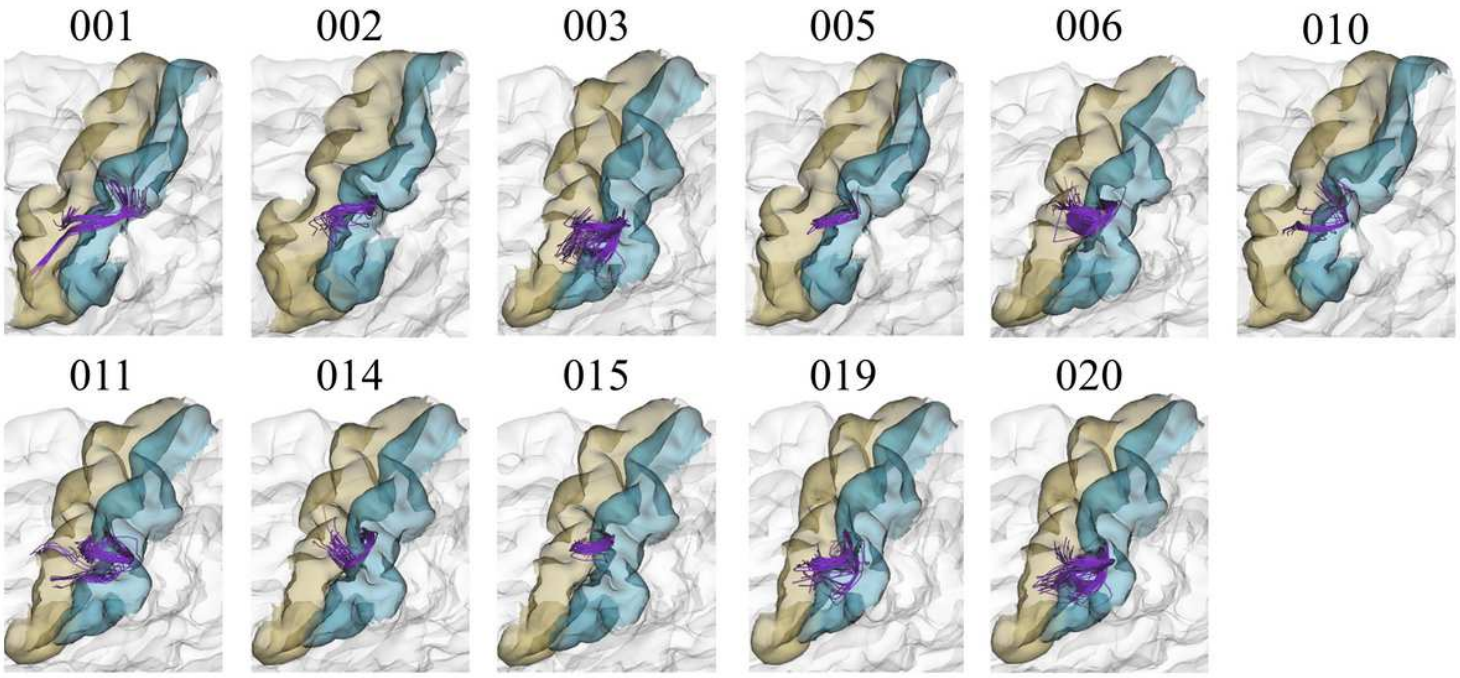

Figure 11

Bundle PoC-PrC 0, with medium reproducibility. The PoC-PrC 0 bundle appears in 11 out of 20 subjects, achieving 55% of reproducibility, for QB algorithm with a 12mm threshold.

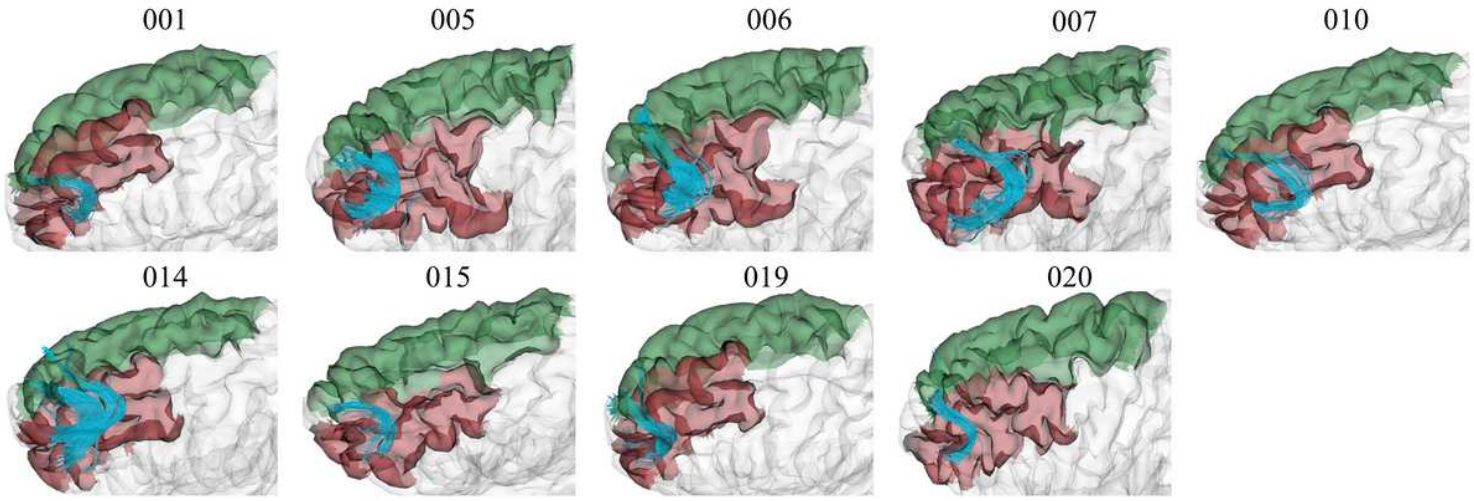

Figure 12

Bundle lh RMF-SF 7, with low reproducibility. The bundle appears in nine out of 20 subjects using the QB clustering algorithm with a 12mm threshold.

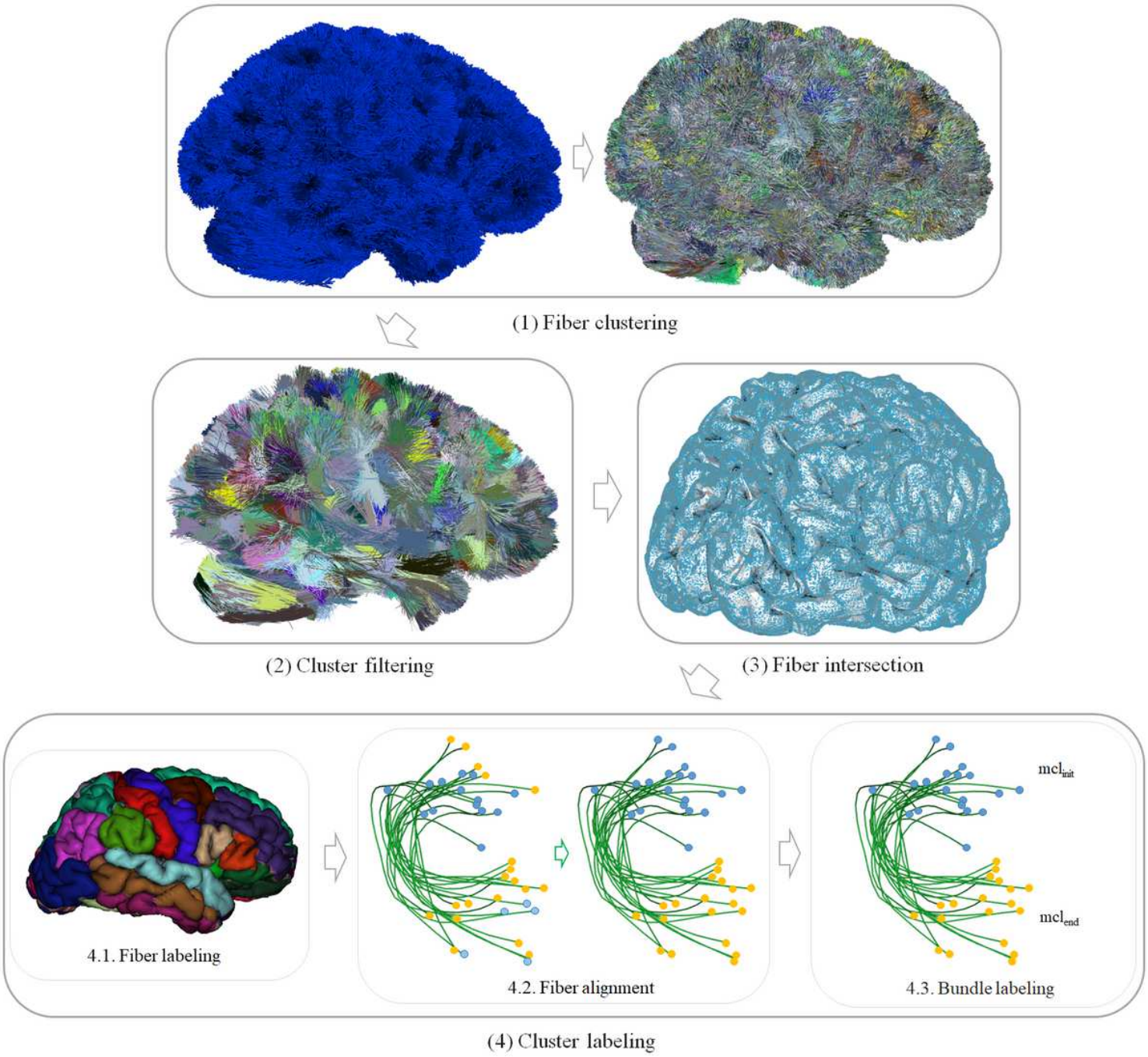

Figure 13

Schematics of the labeling method. Stage 1: Fiber clustering. Performs the clustering of the entire tractography. Stage 2: Cluster filtering. Filters out the small clusters and only keep the short bundles, obtained in the previous stage. Stage 3: Fiber intersection. Calculates the fiber bundle intersection with the cortical mesh. Stage 4: Cluster labeling. Renames the clusters based on the two connected regions of the cortex and their relative position.

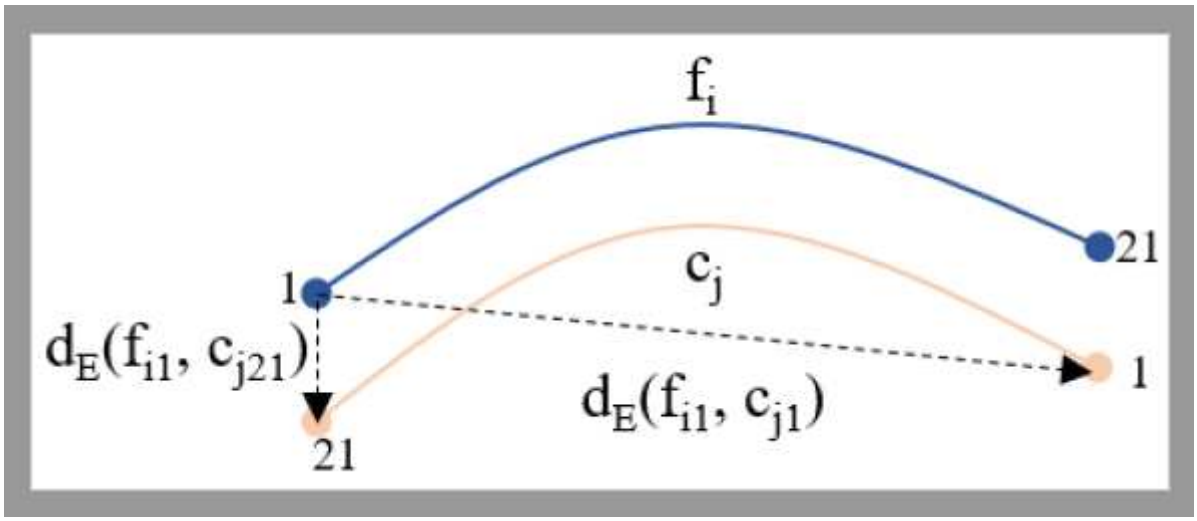

Figure 14

Fiber bundle alignment by respect to its corresponding bundle centroid. The Euclidean distance (dE) is calculated between the firpoint of the fiber (fi1) and both end points of the centroid (ci1 to cj21). If dE(fi1; cj1) > dE(fi1; cj21), the fiber is reversed.

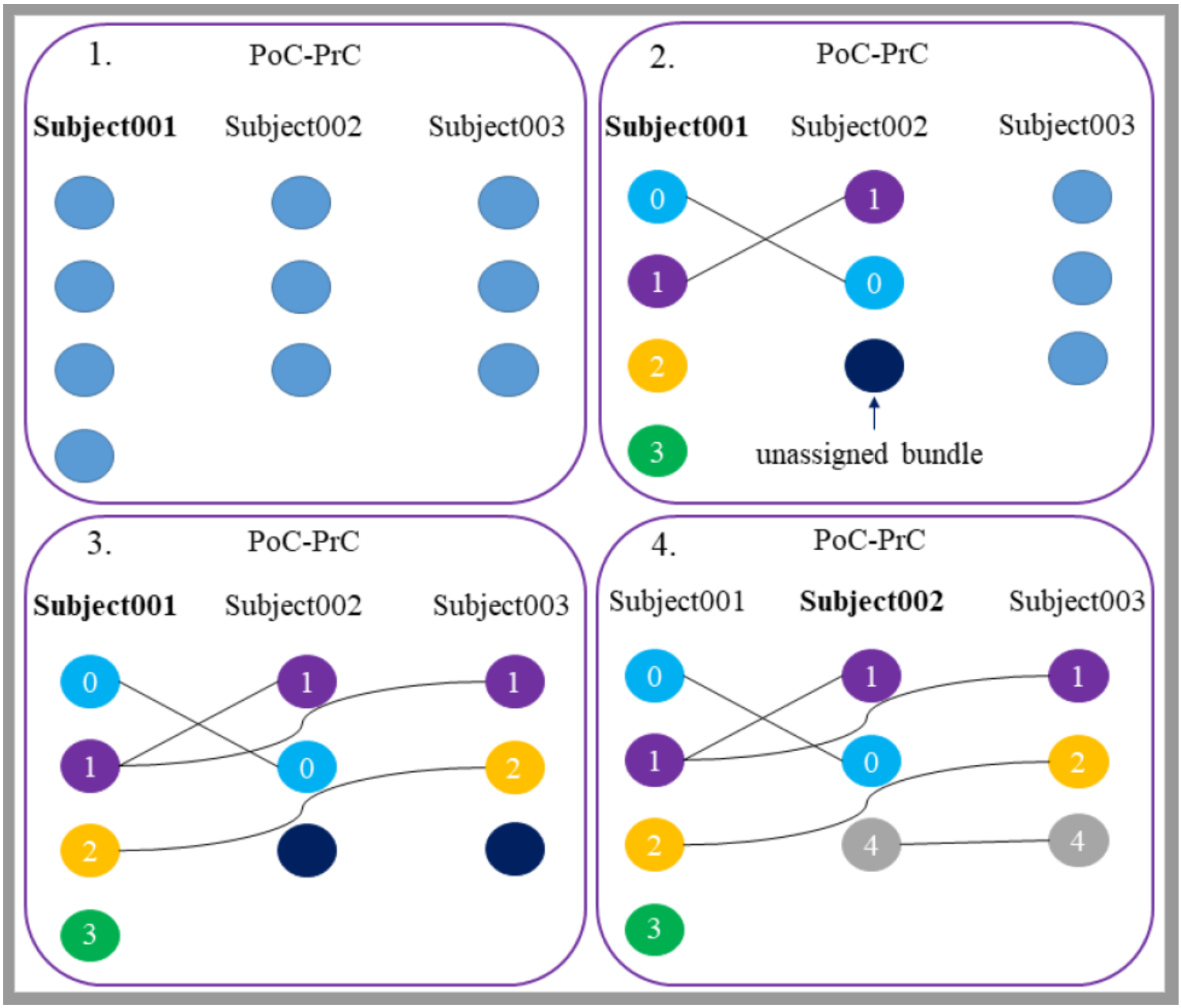

Figure 15

Schematics of the Hungarian algorithm for inter-subject labeling of bundles connecting PoC-PrC regions. First, the bundles are ordered from highest to lowest number of bundles. Second, the reference subject, Subject001, is compared to Subject002, leaving unassigned bundles. Third, it continues comparing to the rest of the subjects. Finally, the reference passes to the next subject with unassigned bundles, Subject002 and these are compared with the rest of the subjects. This process is repeated until all subjects are analyzed.

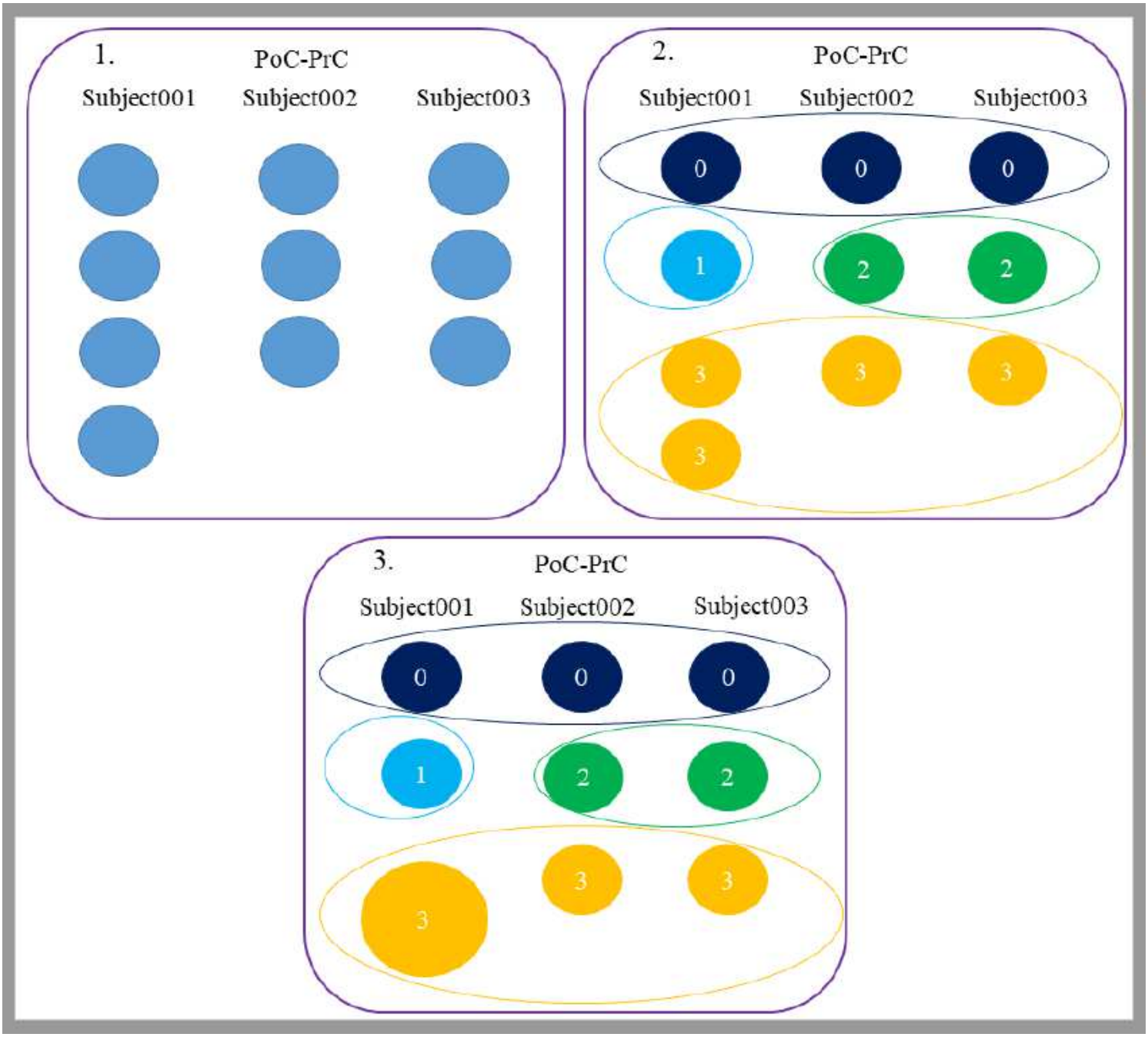

Figure 16

Schematics of the QB algorithm for labeling inter-subject bundles for PoC-PrC regions. First, the cluster centroids are computed. Second, QB is applied to all the intra-subject clusters, to obtain inter-subject clusters. Bundles belonging to an inter-subject cluster are labeled using the same name. Finally, clusters of the same subject that belong to the same inter-subject cluster are merged.