Materials and Method

The subject of this method is the end-to-end structural connectome networks by probabilistic tractography and the AAL atlas. Of course, there are other methods rather than probabilistic tractography, but we chose it because probabilistic tractography is commonly used for the structural connectome analysis.

As a dataset, eleven healthy subjects of the Human Connectome Project (HCP) were randomly extracted. The extracted sub-dataset contained 8 males and 3 females, and the average age was 25 years old. The Preprocessed Diffusion MRI data and the Preprocessed Structure Data of HCP were downloaded from the 1,200-subject release of the HCP dataset. The details of the HCP is referred to the paper (Van Essen et al. and Sotiropoulos et al.) for more details about the data acquisition and preprocessing.

・A Construction of Tractograms and Connectomes

The diffusion-weighted image data were processed following this anatomically-constrained tractography (ACT) procedure. In summary, a tissue-segmented image (5TT image) was generated from a T1 weighted image by MRtrix3 command (MRtrix3 command 5ttgen). The original diffusion-weighted image of HCP was multi-shell diffusion image, and we extracted the single-shell of b=2,000 images by MRtrix3 command (MRtrix3 command dwi2extracts) because the LiFE post-procedure required a huge memory consumption and computations. To avoid this memory consumption (approximately 500GB), we extracted single-shell images and reduced memory consumption. Then, the response function was estimated using the Tournier algorithm, and a single-shell Constrained Spherical Deconvolution (SCSD) was performed with the MRtrix3 commands dwi2fod and dwi2response. The parameter of each command was that the maximum spherical harmonic order (lmax) was 6.

Afterward, the initial tractogram was generated by the iFOD2 algorithm with the anatomy constraint. This processing generated one million streamlines in which maximum tract length = 250 millimeters, FA cutoff = 0.06, and seed\_image = GM\_WM interface. In the anatomy constraint, the streamlines that reached the GM boundaries were within both gray matter and white matter.

Then, Spherical Deconvolution Informed Filtering of Tractograms (SIFT) methodology and Linear Fascicle Evaluation (LiFE) were applied to the tractograms generated by the above procedure. The parameter for reducing the ratio for SIFT was 50%. The LiFE algorithm requires massive memory usage, so we down-sampled the diffusion-weighted image data into 3 mm x 3 mm x 3 mm-sized voxels and limited the number of fibers to one million in order to execute the LiFE algorithm in the ordinary workstations. As a conventional workstation, we used a Linux workstation with 98GB memory and 2 Xeon 3.6 GHz CPUs. The execution of the LiFE algorithm was performed by the python-implemented version in the Dipy [ref]. These post-processing procedures are known to reduce the tracking bias by optimizing the gaps between prediction and data.

For these reasons, the application of SIFT or LiFE in the successor procedure after probabilistic tracking is highly recommended to obtain a more informative structural connectome.

At this point, we obtained three whole brain tractograms—original, not post-processed whole brain tractograms; SIFT-applied whole brain tractograms; LiFE-applied whole brain tractograms.

Finally, these whole streamlines were parcellated into a set of 116 regions of the Automated Anatomical Labeling (AAL) atlas by MRtrix3 command tck2connectome, and we obtained the AAL based structural matrices with 116 nodes and 6,670 edges for each subject. The AAL atlas is one of the most common atlases used in the structural and functional connectome-based study.

・Proposal on a Novel Method for Outlier Removal

The procedure for this proposed method is summarized in [fig. ref]. The outlier fibers have some apparent features as they traverse looped pathways, too-long pathways, or isolated pathways. The simple and commonly used method for removing outlier fibers is length-based filtering. Too-long or too-short lengths are regarded as outlier fibers. However, in some cases, outlier fibers are not so long or short. Then, the simple criterion such as the deviation from the mean of lengths could not adequately remove the outlier fibers but also remove correct fibers in error. However, it is considered that outliers can be excluded with higher accuracy if combining a plurality of features and successfully defining the degree of dissociation.

Then, we incorporate other fiber features in addition to fiber length. Fibers are characterized by the three-dimensional feature vector (length, mean-density, mean-curvature),

[EQ features]

[EQ features]

where k means the index of the edge for the connectome matrix, and l means the lth fibers of the corresponding edge.

The mean-density and the mean-curvature were important as features because the outlier fibers are sometimes isolated, loop-structured or skewed trajectories. To obtain the mean-density and mean-curvature along fibers, the track-weighted density and track curvature images were calculated by the MRtrix3 command tckmap. Then, the mean-density and the mean-curvature along the fibers were calculated by averaging the density and the curvature of all voxels where the fibers crossed, respectively.

To convert these features into scalar statistical value, we adopted Maharabinos distance with a robust covariance estimated by the fast-MCD algorithm [ref]. This is because this V\_i,j contains many outliers (next section, we will estimate 5% ~ 15%), then the maximum likelihood estimator of the covariance tends to be overfitting to the outliers. The fast-MCD estimator is a robust and computational efficient algorithm for estimating the Minimum Covariance Determinant (MCD). The implementation of this fast-MCD was in the opensource python library of scikit-learn [ref].

To convert this Maharabinos distance into statistical value, we used the robust statical method, such that

[EQ stats],

 where MAD is a Median Absolute Deviation[ref].

This robust score is significantly more reliable than z-score, and we chose the cutoff for outlier as |4|. This threshold is thought to be very conservative because the z-score of 4 means 99.993 percent area. In the following discussion section, we address the shape of this robust score and the validity of this threshold.

・Evaluations for this proposed method

To quantitate the effect of this proposed method, we introduce the evaluation index,

[EQ stats]

E\_k,l = (A - B)/A time 100

, where k and l mean the row and column of the connectome matrix respectively. The variables A and B mean the number of fibers for the original edge and filtered edge respectively. Besides, to obtain the stability of this E\_k,l we neglect the edges with very few fibers below 30.