Objectivity of the Threshold

We refer to the distribution of robust score. As shown in [fig ref], the distribution of the robust score has a symmetrical shape, and it is then possible to set the objective threshold for defining outlier fibers.

On the other hand, if the distribution of fiber length is a kind of asymmetrical form, it is impossible to set the threshold. Moreover, the threshold value of 4 is a very conservative value. The shape of the robust score and the conventional threshold contribute to the reliability of this study.

Quantitative Effect of Filtering for the Connectome

It was possible to selectively remove outlier fibers as shown in the [fig ref]. Also, it means that the tracked fibers with probabilistic tractography contain a certain number of outliers because the criteria for outlier fibers was considered conservative.

In previous studies, it was reported that even for simple data sets such as artificial simple data, spurious false-positive fibers are generated by the tracking. Therefore, it is reasonable that more erroneous fibers are generated in the tracking because the actual datasets contain many intersections.

Although it is possible to remove outlier fibers to some extent by the conventional length-based method, our proposed outlier removal method can remove them with more accuracy and objectivity. Therefore, we believe that our proposed method will contribute toward improving the accuracy and reliability of structural connectome analysis.

The reduction of the variance and its influence on the structural connectome analysis

The comparison of the difference between two groups is usually performed through an analysis of the structural connectome. In the case of the comparison, clinical studies on schizophrenia, Alzheimer's, epilepsy, and other psychiatric diseases have been reported. Statistical tests, such as the t-test, are conducted on each element of the structural connective matrix, and the correction for multiple comparisons is performed. In the t-test, the variance of the statistical value is a critical variable for the statistical power because the low variables reduce the scope of the null hypothesis.

If the variance due to outlier fibers can be suppressed, the statistical power of the t-static and Cohen's-d increase. For the t-static, the standard deviation equals n^2 (n: sample number). It is an advantage of this method that we can obtain the same effect by increasing the samples n^2. In the case of the 11 subjects, the variance decreased by an average of 7% (Table 2) as a result of the removal of outlier fibers. It is considered that there was an effect on the averages of five subjects (49%) from the sample. In clinical studies, it is often difficult to collect samples because of the many limitations. Then, this effect is one of the effective points of this method. This proposed method makes more accurate the comparison at the network level.

This proposed method also improves the studies on other metrics of fiber count such as FA, MD, or the other weighted connectome. The decreasing of outlier fibers improves the regions of the sampling, and it contributes to the reducing of the variance of these metrics.

Scalability

This proposed method is assumed to be used in a realistic environment in clinical or fundamental studies. It can be executed in a relatively short time even for a whole brain network in ~30 minutes with a conventional workstation. It is considered that time increases linearly as the number of fibers increases, but this proposed method can easily be expanded to multitasks. So, even if the number of fibers increases, we can deal with this scale problem by increasing the number of processors or using processors with more cores.

Limitations

As a limit of the proposed method, we suppose three limitations.

(1) Network with a too-small number of fibers

It is considered that it cannot be applied to a network containing too few fibers (~10 fibers in one network connectome). The reason is that it is difficult to statistically separate normal fibers and outlier fibers. However, it seems to be unlikely to occur because the network with too few fibers is not the subject of the structural connectome study.

(2) Network with a huge number of fibers

Neither our algorithm nor the usual tracking algorithm could be applied to networks with a huge number (1,000 million more) of fibers unless you use supercomputers.

(3) Networks other than the end-to-end networks

In this study, we applied the proposed method to the End 2 End network based on the AAL atlas. The AAL atlas is a commonly used atlas for structural connectome analysis. However, this study does not mention that this method can be applied to clustered fiber bundles. We would like to consider whether this proposed method contributes to improving the accuracy of clustered fibers in the near future. This study also does not refer to the application of the JHU Atlas, but it is applicable to the End 2 End network based on Desikan-Kiliany Atlas (Supplementary Material 1).