Supra-Threshold Fiber Cluster Statistics for Data-Driven Whole Brain Tractography Analysis

Fan Zhang[1]([⊠]), Weining Wu[1,2], Lipeng Ning[1], Gloria McAnulty[3],
Deborah Waber[3], Borjan Gagoski[3], Kiera Sarill[3], Hesham M. Hamoda[3],
Yang Song[4], Weidong Cai[4], Yogesh Rathi[1], and Lauren J. O'Donnell[1]

[1] Brigham and Women's Hospital, Harvard Medical School, Boston, USA
fzhang@bwh.harvard.edu
[2] College of Computer Science and Technology,
Harbin Engineering University, Harbin, China
[3] Boston Children's Hospital, Harvard Medical School, Boston, USA
[4] School of Information Technologies, The University of Sydney, Sydney, Australia

Abstract. This work presents a supra-threshold fiber cluster (STFC) analysis that leverages the whole brain fiber geometry to enhance statistical group difference analysis. The proposed method consists of (1) a study-specific data-driven tractography parcellation to obtain white matter (WM) tract parcels according to the WM anatomy and (2) a nonparametric permutation-based STFC test to identify significant differences between study populations (e.g. disease and healthy). The basic idea of our method is that a WM parcel's neighborhood (parcels with similar WM anatomy) can support the parcel's statistical significance when correcting for multiple comparisons. The method is demonstrated by application to a multi-shell diffusion MRI dataset from 59 individuals, including 30 attention deficit hyperactivity disorder (ADHD) patients and 29 healthy controls (HCs). Evaluations are conducted using both synthetic and real data. The results indicate that our STFC method gives greater sensitivity in finding group differences in WM tract parcels compared to several traditional multiple comparison correction methods.

1 Introduction

Computational neuroimaging methods have aroused interest in identifying group differences in white matter (WM) via diffusion magnetic resonance imaging (dMRI). Traditional methods rely on techniques such as voxel-based morphometry (VBM) [2] and tract-based spatial statistics (TBSS) [24] for voxel-based analyses. Tractography-based analysis, on the other hand, has enabled measurement of macrostructural WM properties of specific subpopulations of fibers [5]. In this work, we focus on identifying WM group differences using whole brain tractography.

A standard tractography-based group difference analysis includes first selecting a tract of interest (e.g. the corpus callosum) and then comparing groups to find statistical differences in WM diffusion features (e.g. anisotropy), either using feature mean values [1] or along-tract measures [6,7,9]. These studies suggest that fibers with similar WM anatomy (fiber geometric trajectory) in general share similar diffusion properties. However, they are generally limited to a small number of selected tracts. Another strategy, which can efficiently identify multiple tracts, is to establish the tract correspondence directly based on fiber geometry [17]. One recent work has applied this strategy to identify 30 tracts (a subset of the WM, e.g. only corticospinal tract but not full corona radiata) and performed statistical analysis in the whole brain [28]. Our method is based on a study-specific whole brain WM parcellation into more regions (a total of 1416 WM parcels from all input tractography) and hence can allow identification of potential group differences more specific to local WM anatomy.

For group comparison, a hypothesis test (e.g. Student's t-test) is normally used to identify group differences, followed by multiple comparison correction (e.g. false discovery rate (FDR) [4] and Bonferroni [10] methods) for corrected statistical significance. Since these commonly-used correction methods can be less sensitive in finding significance, voxel-based multiple comparison correction has been conducted in a cluster-thresholding manner that utilizes spatial neighborhoods to boost belief in extended cluster areas [12]. Studies using voxel-cluster-thresholding methods have found WM group differences [19,23]. One work also applied the cluster-thresholding method to identify group differences in fiber segments [26]. However, we found no related work has been conducted to identify tractography-based group differences for whole brain analysis.

In light of the above, we propose a supra-threshold fiber cluster (STFC) method to identify WM group differences using whole brain tractography. The novelty is that, for the first time, the proposed method leverages the whole brain fiber geometry during the statistical analysis of tractography. Specifically, we define a WM parcel neighborhood according to the WM anatomy and we propose a novel method that uses these WM neighborhoods for determining statistical significance. The method uses a study-specific data-driven WM parcellation for parcel neighborhood construction (Sect. 2.2). The STFC test then leverages the neighborhoods to identify fiber clusters of multiple WM tract parcels for statistical significance with multiple comparison correction (Sect. 2.4). We demonstrate our method by application to a multi-shell dMRI dataset from attention deficit hyperactivity disorder (ADHD) patients and healthy controls (HCs).

2 Methods

2.1 Dataset

We used a multi-shell (b values of $1000/2000/3000$ s/mm^2) diffusion weighted imaging (DWI) dataset from 59 individuals (30 ADHD, 7 females and 23 males, age: 10.6 ± 1.7 years; 29 HC, 10 females and 19 males, age: 10.7 ± 1.7 years). The two groups were matched for age and socioeconomic status (SES). DWI

data were acquired using a multi-slice acquisition (x2) at spatial resolution of $2 \times 2 \times 2 \, \text{mm}^3$ with 70 gradient directions, and then processed using a semi-automated quality control to remove gradients with signal drop and to correct head motion.

We conducted whole brain tractography using the unscented Kalman filter method [21,22]. Tractography was seeded 5 times per voxel and return-to-the-origin probability (RTOP) was measured with a two tensor biexponential model [20] at each point while tracking. The RTOP estimates the net displacement of water molecules between the application of two diffusion gradients and it is known to be sensitive to the anisotropy of WM tissue [13]. We chose RTOP because it is known to be sensitive to the anisotropy of WM tissue and may increase pathophysiological specificity compared to traditional diffusion anisotropy measures, e.g. fractional anisotropy (FA) [3,18]. Therefore, we used this measurement to explore potential changes in WM anisotropy in ADHD. We performed diffusion MRI tractography visualization in 3D Slicer (www.slicer.org) via the SlicerDMRI project (dmri.slicer.org).

2.2   Data-Driven WM Parcellation and WM Parcel Neighborhood

Whole brain WM parcellation was conducted using a data-driven pipeline according to the common WM anatomy from the whole population. In brief, the parcellation started with a simultaneous joint alignment of tractography across all subjects (using affine then b-spline transforms) (Fig. 1a) [15]. Next, we learned a study-specific data-driven groupwise WM parcellation (atlas) using a spectral clustering of pairwise fiber trajectory distances across all subjects (Fig. 1b) [16]. Then, we applied the study-specific parcellation to each individual subject (Fig. 1c) [16]. Valid parcels were identified as those that passed a nonparametric one-tailed sign test in the population, based on fiber numbers in each parcel, as in [8]. We obtained a total of 1416 valid hemispheric and commissural parcels.

Parcel neighborhoods (Fig. 1d) were constructed according to the mean of the pairwise fiber distances between parcels in the atlas, i.e. $D_p = \frac{\sum_I \sum_J d_{ij}}{IJ}$, where d_{ij} is the fiber pair distance and I and J are the total numbers of fibers in the two parcels. A pair of parcels that had D_p smaller than a user-given distance threshold T_d were considered to be neighbors. We applied the mean closest point fiber distance [16] to measure d_{ij}, the same distance used in the spectral clustering. In this way, the neighborhood could capture the anatomical similarity between the parcels.
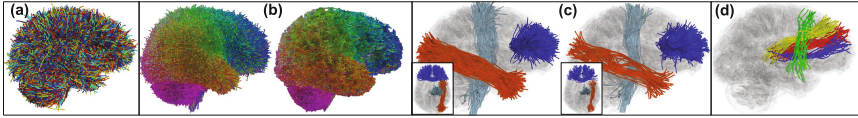
Fig. 1. Study-specific data-driven WM parcellation. (a) Tractography alignment using groupwise tractography registration, where color indicates subject. (b) Study-specific groupwise WM parcellation (atlas) (left) and example subject-specific tractography parcellation (right). Fibers from different parcels are colored differently, where similar colors represent white-matter-anatomy-similar fibers. (c) Example atlas WM tract parcels (left) and the corresponding subject-specific WM tract parcels (right). (d) Example neighbors of the red tract parcel include the yellow and the blue parcels but not the more distant green one.

2.3 Group Difference at Individual Parcel Level

We performed a null hypothesis test for each individual parcel to find the parcel-level WM difference. Specifically, we measured the median of the RTOP values of all points in each parcel, i.e. M_{RTOP}. An one-tailed Student's t-test was then performed under null hypothesis $H_0 : \mu_{HC}(M_{RTOP}) \leq \mu_{ADHD}(M_{RTOP})$, as studies in ADHD widely suggest decreased diffusion anisotropy [11]. We considered the parcels with p-value <0.05 as the ones with parcel-level differences.

2.4 Supra-Threshold Fiber Cluster Test

Then, the parcel-level differences were tested for significance in a multiple comparison correction using a permutation-based STFC test. This is similar to the process in a voxel-image-based supra-threshold cluster test [12]. However, our method leverages the fiber spatial relationships from the whole brain fiber geometry to build the parcel neighborhoods, while a voxel-image-based method relies on voxel spatial neighborhoods. In detail, an STFC was defined as a cluster of multiple parcels with parcel-level differences, where each parcel neighbored at least one other parcel(s) in the cluster (under the distance threshold T_d). The STFC test performed a nonparametric permutation test using a summary statistic of maximal STFC size ($maxSTFCS$). $N = 10000$ permutations were performed in all experiments, resulting in a distribution of $maxSTFCS$ (the null distribution) that enabled computation of corrected STFC significance. Algorithm 1 shows the pseudocode of the method. Specifically, we first computed a histogram of $maxSTFCS$ across multiple permutation runs (lines 1 to 5). Then, for each STFC from the correctly labeled groups, its corrected significance value was computed by comparing its STFC size ($STFCS$) to the $maxSTFCS$ histogram (lines 6 to 11).

**Algorithm 1.** Supra-threshold Fiber Cluster Test

---

1: **for** each permutation run in [1, N] **do**
2:    Randomly permute group labels of all subjects (0 for ADHD, 1 for HC).
3:    Identify parcels with group difference at parcel-level in the permuted data.
4:    Extract all STFCs and record the $maxSTFCS$.
5: Calculate the histogram of $maxSTFCS$ to produce the null distribution.
6: Identify parcels with group difference at parcel-level in the correctly labeled data.
7: Extract all STFCs and compute $STFCS$ for each parcel.
8: **for** each STFC **do**
9:    Locate its $STFCS$ in the $maxSTFCS$ histogram.
10:    Obtain the number of permutations with $maxSTFCS \geq STFCS(p)$ as $N(p)$.
11:    Compute the corrected significance value as $(N(p) + 1)/(N + 1)$.

---

3   Experimental Results

3.1   Synthetic Data

We first illustrate our method on synthetic data. To simplify the assessment and tractography visualization, we created a realistic synthetic dataset with true group difference in the corpus callosum (CC), generated as follows. We identified a total of 34 CC parcels from the whole brain parcellation, as shown in Fig. 2a. For each CC parcel, we added white Gaussian noise (signal-to-noise ratio at 1 [25]) to the actual measured features of all HC subjects. Repeating this process twice generated two synthetic groups of G1 and G2, each with 29 subjects. We then modified 15 CC parcels-of-interest to have true group difference by adding synthetic feature changes to the G2 subjects. For each of the 15 parcels, we decreased its group mean M_{RTOP} values in G2 (as a percent of its original feature mean) for a null hypothesis test: $H_0 : \mu_{G1}(M_{RTOP}) \leq \mu_{G2}(M_{RTOP})$. These 15 parcels were selected to form 3 different synthetic ground truth clusters (with sizes of 4, 5 and 6, as shown in Figs. 2b–d). Larger synthetic feature changes led to the parcels-of-interest showing more significant group differences. Then, the evaluation goal was to test if a method could correctly identify the parcels with true significance, even when the added change was small.

Comparisons were conducted among the uncorrected t-test, the proposed STFC method, a standard permutation test (Perm-T, $N = 10000$) that used the minimal t-test-based p-value of all parcels for the summary statistic (as applied in [28]), and two traditional FDR and Bonferroni multiple comparison correction methods. The same significance level of $\alpha = 0.05$ was used for all compared methods. Figure 2b displays the number of significantly different parcels that were correctly identified in each method. For the STFC method, we displayed the result with $T_d = 44$ mm. As shown in Fig. 2c, the reasonable range for T_d in the synthetic data testing was 44 to 47 mm. This range corresponded to the settings that could identify the 3 ground truth clusters, as illustrated in Fig. 2d. We note that in the STFC experiment we considered a parcel as misidentified if it belonged to a STFC including parcels from more than one ground truth cluster.
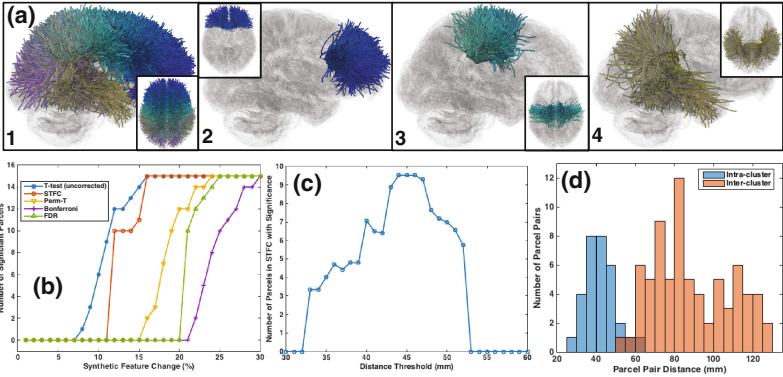
Fig. 2. Synthetic data experiment: (a) All CC parcels (a1, size 34) and 15 CC parcels-of-interest (a2, a3 and a4 cluster, with sizes of 4, 5 and 6 respectively) with synthetic group differences. (b) Number of correctly identified CC parcels, versus the level of synthetic group difference. (c) Mean number of correctly identified parcels across the different synthetic changes (as in b), versus the distance threshold T_d. T_d lower than 44 mm could not identify all ground truth parcels, while T_d larger than 47 mm grouped inter-cluster parcels together. (Note that T_d should be higher than the within-parcel mean fiber distance, which was 27.9 ± 5.3 mm.) (d) Distributions of the parcel pair distance D_p from intra-cluster (blue) and inter-cluster (orange) among the ground truth STFCs (a2–a4).
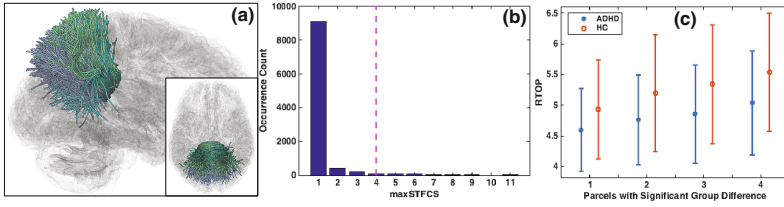
3.2   Real Data

Next, we show experimental results on the real data for the corpus callosum[1] and the whole brain. A one-tailed t-test was first performed with the null hypothesis $H_0 : \mu_{HC}(M_{RTOP}) \leq \mu_{ADHD}(M_{RTOP})$. Multiple comparison correction was then conducted using the STFC, the Perm-T, the FDR and the Bonferroni methods respectively, at a significance level of $\alpha = 0.05$. For the STFC method, the results were reported at the smallest T_d where we could identify significances.

Corpus Callosum: 8 of the 34 CC parcels passed the initial t-test. While none survived the Perm-T, the FDR or the Bonferroni methods, a STFC of 4 parcels (Fig. 3a) was identified using the proposed method when $T_d = 32$ mm. The corrected significance value of the identified STFC was 0.0271 as illustrated in Fig. 3b and the group mean feature values of the parcels are shown in Fig. 3c.

Whole Brain: For the whole brain parcellation of a total of 1416 parcels, 654 had a p-value smaller than 0.05 in the initial parcel-level t-test. The FDR and the Bonferroni methods did not find any significantly different parcels between the groups. The Perm-T method identified one individual significant parcel (corrected p-value 0.0272) connecting the middle precentral gyrus to the

---

[1] This tract's subdivisions are often studied, so we chose it as an illustrative example.

Fig. 3. Real data experiment for the corpus callosum: (a) The STFC consisting of 4 CC parcels identified with significant group difference. (b) 270 permutation tests had $maxSTFCS \geq 4$, leading to the corrected significance value of $271/10001 = 0.0271$. (c) Comparison of M_{RTOP} values per identified parcel, plotted in sorted order.
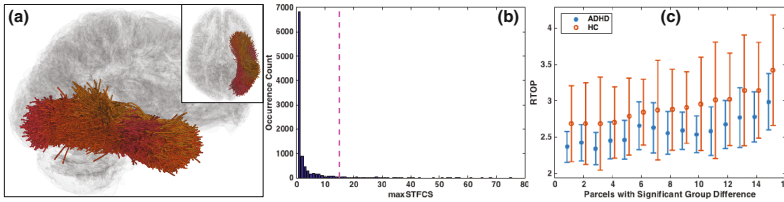
Fig. 4. Real data experiment for the whole brain WM: (a) The STFC consisting of 15 parcels identified with significant group difference. (b) 455 permutation tests had $maxSTFCS \geq 15$, leading to the corrected significance value of the identified STFC $456/10001=0.0456$. (c) Comparison of M_{RTOP} values per identified parcel, plotted in sorted order.

supra-marginal and the superior-parietal gyri in the right hemisphere. Our method identified one significant STFC of 15 WM parcels when $T_d = 24\,\mathrm{mm}$, which located in the temporal and occipital lobes (Fig. 4a). The corrected significance value of the identified STFC was 0.0456 as illustrated in Fig. 4b. A comparison of the group mean feature values is given in Fig. 4c.

4   Discussion and Conclusion

We tested our method on a synthetic dataset with known group differences. The results showed that the STFC method was more sensitive to detect the ground truth group differences, while the Perm-T, FDR and the Bonferroni methods could only find the significance when there were large differences (over 15% change of the group mean feature value). As for the experiments on the real data, in CC we found a significantly different STFC of 4 WM parcels connecting to the superior-parietal gyri and the precuneus, with decreased median RTOP values in the ADHD group. For the whole brain WM analysis, our method identified 15 parcels from a STFC with significant group difference in the right temporal and occipital lobes, which have been previously reported to be affected in ADHD [14,27]. These parcels had lower median RTOP values in the ADHD group when compared to the HC group, suggesting potentially reduced WM anisotropy in

ADHD. We did not find any significance using the FDR or Bonferroni methods in two real data tests. One parcel that potentially belonged to the anterior segment of right arcuate fasciculus was identified from the whole brain using the Perm-T method, but no parcel survived in the CC Perm-T analysis.

The STFC method had a parameter of distance threshold T_d that was used to form the WM parcel neighborhood. Our experiments using the synthetic data showed that too large or too small T_d values found fewer true significant parcels (Fig. 2c). Given the synthetic ground truth clusters, small T_d values (e.g. $T_d < 25$ mm, Fig. 2d) were not able to form any neighborhoods within a cluster, while larger values tended to group ground truth clusters together (e.g. $T_d > 65$ mm, Fig. 2d). For our real data tests, we applied the minimal setting of T_d that could find a significance. This allowed us to identify the parcels that were most similar to each other in terms of their WM anatomy. On the other hand, increasing T_d could help to find more WM structures. For example, in the whole brain analysis given $T_d = 25$ mm, we identified a larger significant STFC of 24 parcels that included the parcels from $T_d = 24$ mm and extended to the inferior parietal lobe.

In this paper, we have presented a novel STFC analysis to identify WM group differences using whole brain tractography. Experimental results suggest that our method in general is more sensitive for identifying WM group differences when compared to several traditional multiple comparison correction methods. Similar to voxel-cluster-thresholding analyses, the proposed method aims to find large clusters of WM parcels with significance; thus it could potentially miss some significantly different parcels located in small neighborhoods, e.g. the one individual significant parcel identified in the standard permutation test.

References

1. Alexander, A.L., Lee, J.E., et al.: Diffusion tensor imaging of the corpus callosum in autism. Neuroimage **34**(1), 61–73 (2007)
2. Ashburner, J., Friston, K.J.: Voxel-based morphometry-the methods. Neuroimage **11**(6), 805–821 (2000)
3. Avram, A.V., et al.: Clinical feasibility of using mean apparent propagator (MAP) MRI to characterize brain tissue microstructure. NeuroImage **127**, 422–434 (2016)
4. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Ser. B (Methodological) **57**, 289–300 (1995)
5. Ciccarelli, O., Catani, M., et al.: Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. Lancet Neurol. **7**(8), 715–727 (2008)
6. Colby, J.B., Soderberg, L., et al.: Along-tract statistics allow for enhanced tractography analysis. Neuroimage **59**(4), 3227–3242 (2012)
7. Corouge, I., Fletcher, P.T., et al.: Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. Med. Img. Ana. **10**(5), 786–798 (2006)
8. Gong, G., He, Y., et al.: Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. Cereb. Cortex **19**(3), 524–536 (2009)

9. Goodlett, C.B., Fletcher, P.T., et al.: Group analysis of DTI fiber tract statistics with application to neurodevelopment. Neuroimage **45**(1), S133–S142 (2009)

10. Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian J. Stat. **6**, 65–70 (1979)

11. Nagel, B.J., et al.: Altered white matter microstructure in children with attention-deficit/hyperactivity disorder. J. Am. Acad. Child Adoles. Psychiatry **50**(3), 283–292 (2011)

12. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. **15**(1), 1–25 (2002)

13. Ning, L., et al.: Estimating diffusion propagator and its moments using directional radial basis functions. IEEE Trans. Med. Img. **34**(10), 2058–2078 (2015)

14. O'Conaill, C.R., Malisza, K.L., et al.: Visual search for feature conjunctions: an fMRI study comparing alcohol-related neurodevelopmental disorder (ARND) to ADHD. J. Neurodevelopmental Dis. **7**(1), 10 (2015)

15. O'Donnell, L.J., Wells, W.M., Golby, A.J., Westin, C.-F.: Unbiased groupwise registration of white matter tractography. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 123–130. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33454-2_16

16. O'Donnell, L., Westin, C.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Trans. Med. Img. **26**(11), 1562–1575 (2007)

17. O'Donnell, L., Westin, C.F., Golby, A.J.: Tract-based morphometry for white matter group analysis. Neuroimage **45**(3), 832–844 (2009)

18. Özarslan, E., et al.: Mean apparent ator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. NeuroImage **78**, 16–32 (2013)

19. Pérez-Iglesias, R., et al.: White matter defects in first episode psychosis patients: a voxelwise analysis of diffusion tensor imaging. Neuroimage **49**(1), 199–204 (2010)

20. Rathi, Y., Michailovich, O., Setsompop, K., Bouix, S., Shenton, M.E., Westin, C.-F.: Sparse multi-shell diffusion imaging. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6892, pp. 58–65. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23629-7_8

21. Rathi, Y., Gagoski, B., Setsompop, K., Michailovich, O., Grant, P.E., Westin, C.-F.: Diffusion propagator estimation from sparse measurements in a tractography framework. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8151, pp. 510–517. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40760-4_64

22. Reddy, C.P., Rathi, Y.: Joint multi-fiber noddi parameter estimation and tractography using the unscented information filter. Front. Neurosci. **10**, 166 (2016)

23. Rotarska-Jagiela, A., Oertel-Knoechel, V., et al.: Anatomical brain connectivity and positive symptoms of schizophrenia: a diffusion tensor imaging study. Psychiatry Res. Neuroimaging **174**(1), 9–16 (2009)

24. Smith, S.M., Jenkinson, M., et al.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage **31**(4), 1487–1505 (2006)

25. Smith, S.M., Nichols, T.E.: Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage **44**(1), 83–98 (2009)

26. Taquet, M., Scherrer, B., Commowick, O., Peters, J., Sahin, M., Macq, B., Warfield, S.K.: Registration and analysis of white matter group differences with a multi-fiber model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 313–320. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33454-2_39

27. Vance, A., Silk, T., et al.: Right parietal dysfunction in children with attention deficit hyperactivity disorder, combined type: a functional MRI study. Mol. Psychiatry **12**(9), 826–832 (2007)
28. Wang, D., et al.: Tractography atlas-based spatial statistics: statistical analysis of diffusion tensor image along fiber pathways. NeuroImage **125**, 301–310 (2016)