

# Voxelwise Multivariate Statistics and Brain-Wide Machine Learning Using the Full Diffusion Tensor

Anne-Laure Fouque<sup>1,2</sup>, Pierre Fillard<sup>1,3</sup>, Anne Bargiacchi<sup>2</sup>, Arnaud Cachia<sup>4</sup>,  
Monica Zilbovicius<sup>2</sup>, Benjamin Thyreau<sup>1</sup>, Edith Le Floch<sup>1,2</sup>, Philippe Ciuciu<sup>1</sup>,  
and Edouard Duchesnay<sup>1,2</sup>

<sup>1</sup> CEA, Neurospin, LNAO, Saclay, France

<sup>2</sup> INSERM-CEA U.1000 Imaging and Psychiatry, Orsay, France

<sup>3</sup> INRIA Saclay-Île-de-France, Parietal, Saclay, France

<sup>4</sup> UMR 894 INSERM - Paris Descartes University, Laboratory of Pathophysiology of  
Psychiatric Diseases, Sainte-Anne Hospital, Paris, France

**Abstract.** In this paper, we propose to use the full diffusion tensor to perform brain-wide score prediction on diffusion tensor imaging (DTI) using the log-Euclidean framework., rather than the commonly used fractional anisotropy (FA). Indeed, scalar values such as the FA do not capture all the information contained in the diffusion tensor. Additionally, full tensor information is included in every step of the pre-processing pipeline: registration, smoothing and feature selection using voxelwise multivariate regression analysis. This approach was tested on data obtained from 30 children and adolescents with autism spectrum disorder and showed some improvement over the FA-only analysis.

## 1 Introduction

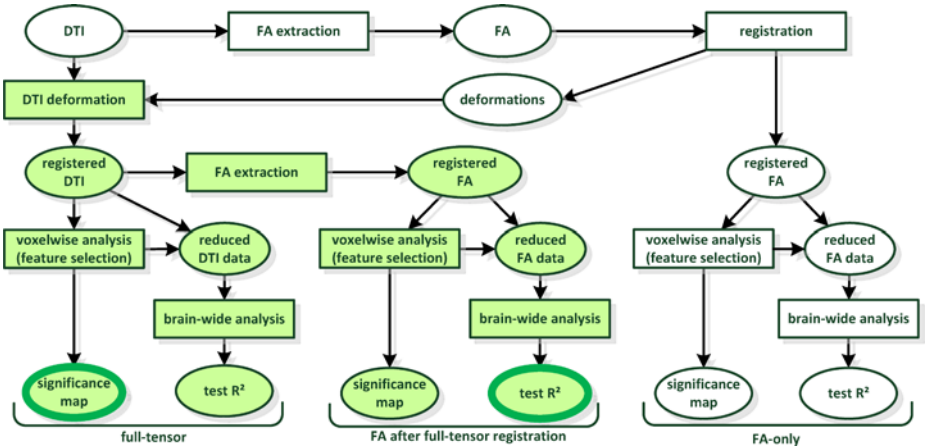
Our aim is to generalize to tensors some of the most commonly used diffusion-MRI processing tools in order to exploit the full information of the tensor in every step of the processing pipeline and ultimately perform brain-wide machine learning on the full tensor. We achieve this goal by making use of Log-Euclidean (LE) metrics [4], a simple and fast way to perform computations in the tensor space. Classical tensor-based analysis consists first in computing scalar images from the dataset, then in working only on the resulting scalar maps. Commonly used scalar features include the fractional anisotropy (FA) and the apparent diffusion coefficient (ADC) for diffusion tensor imaging (DTI) studies, and the Jacobian determinant in tensor-based morphometry (TBM) studies. However, these scalar values are computed only from the eigenvalues of the tensors, and therefore do not capture all the information available. Some information (*e.g.* tensor orientation) is thus lost in the process. Recently, several groups performed full-tensor voxelwise analysis of tensor fields with LE metrics, both in the DTI context [6,17] and in the TBM context [10,5]. However, to the best of our knowledge the LE framework has not yet been applied to voxelwise regression nor to brain-wide machine learning. The present paper aims at filling this gap.

The remainder of this paper is organized as follows. We first present the Log-Euclidean framework and how to use it to perform all the steps necessary to brain-wide machine learning (pre-processings, feature selection using voxelwise regression analysis and finally brain-wide score prediction). We then describe the dataset involved in our experiments and compare the results of the full-tensor approach with those of the classical FA-only approach.

## 2 Methods

In this section, we present the methods used in this study. Section 2 describes the Log-Euclidean framework. Next sections expose the different steps in the pipeline: registration, feature selection using voxelwise regression analysis and brain-wide score prediction. The whole pipeline is summarized in Fig. 1.

**Log-Euclidean Metrics.** The tensor space is the space of  $3 \times 3$  symmetric positive-definite matrices  $Sym_3^+$ . The usual Euclidean operations on the space of  $3 \times 3$  matrices suffer from many defects when applied to tensors [4]: positive-definiteness is not always preserved, and there is tensor swelling effect. To address these problems, affine-invariant metrics have been proposed [13], but they induce a huge numerical complexity. The Log-Euclidean framework [4] is a much simpler and faster way to make computations on tensors, while in practice yielding similar results as the affine-invariant framework. The LE approach relies on a vector space structure defined on tensor space  $Sym_3^+$  with a logarithmic multiplication  $S_1 \odot S_2 \stackrel{def}{=} \exp(\log S_1 + \log S_2)$  and a logarithmic scalar multiplication



**Fig. 1.** Flowchart of the whole pipeline – in white classical steps, in green new steps to account for the full-tensor information – thick green border marks the most efficient methods for voxelwise and brain-wide analysis respectively – some pre-processings have been omitted

$\lambda \otimes S \stackrel{def}{=} \exp(\lambda \cdot \log S)$ . Computations on the tensors can thus be converted into Euclidean computations on the tensor logarithms. Moreover, this yields a one-to-one mapping between tensor space and  $\mathbb{R}^6$  with the operator  $V(\mathbf{S}) = (\log(\mathbf{S})_{1,1}, \sqrt{2} \cdot \log(\mathbf{S})_{2,1}, \log(\mathbf{S})_{2,2}, \sqrt{2} \cdot \log(\mathbf{S})_{3,1}, \sqrt{2} \cdot \log(\mathbf{S})_{3,2}, \log(\mathbf{S})_{3,3})^T$  ([4], [13]). Computations are therefore further simplified by working on the resulting vectors.

**Registration and Smoothing.** The classical pipeline involves: (1) extracting scalar values (FA), (2) constructing a study-specific scalar template, (3) nonlinearly registering the individual scalar maps to the scalar template, (4) smoothing the resulting scalar maps to improve the power of any further statistical analysis.

To retain the full tensor information in our pipeline, we chose to use the deformations computed from FA registration in step 3, and apply them to the tensor fields (*cf* Fig. 1). During this warping, each tensor must be reoriented to remain consistent with the anatomy [2]. Moreover, the interpolation is performed in the Log-Euclidean domain. Smoothing is performed in the LE domain by applying an isotropic Gaussian convolution kernel to the vector representation of the tensor logarithms.

**Feature Selection using Voxelwise Statistical Analysis.** Brain-wide methods suffer from the *curse of dimensionality*: as neuroimaging data provides a very large number of features and a low number of subjects, there is a very high risk of overfitting the training data resulting in poor generalization. It is therefore necessary to select a smaller set of relevant features to train the machine learning system. To perform this feature selection step, we use voxel-based regression analysis which allows us to rank voxels according to their statistical significance. A regression is carried out in each voxel yielding an  $F$ -value. In the full-tensor case, the regression is multivariate and the  $F$ -value is computed via the Pillai-Bartlett's trace [8]. To visualize the results, the  $p$ -values maps are thresholded to show only the most statistically significant regions. To use as feature selection, we retain the best  $m$  voxels for the rest of the analysis.

**Brain-Wide Machine Learning.** Some complex brain patterns where several regions interact cannot be detected by voxel-based analysis. Multivariate methods that analyze voxels jointly are thus required to detect these patterns. In this paper, we call these multivariate methods *brain-wide methods* to avoid confusion with multivariate voxel-based methods that work with multivariate objects on only one voxel. The regression method used in this paper is kernel ridge regression (see *e.g.* [9]).

Based on a dataset made of  $N$  subjects, we denote by  $\mathbf{y}$  the  $N \times 1$  vector of responses.  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$  are the  $N$  subject images (typically  $\mathbf{x}_i$  is of size  $m \times 1$  for FA-only,  $m \times 6$  for tensors, where  $m$  is the number of selected voxels).  $k(\mathbf{x}_i, \mathbf{x}_j)$  is the kernel function that transforms the data from the data space to the feature space and  $\mathbf{K}$  is the  $N \times N$  kernel matrix where  $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ . The most simple kernel function is the linear kernel, which corresponds to the inner product of the data space. The goal of the kernel ridge regression is to minimize the penalized residual sum of squares:  $\hat{f} = \arg \min_{f \in \mathcal{H}_k} \sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2 + \lambda \|f\|_{\mathcal{H}_k}^2$ .

As  $\hat{f}$  can be expanded into  $\hat{f}(\mathbf{x}) = \sum_{i=1}^N \alpha_i K(\mathbf{x}_i, \mathbf{x})$ , this is equivalent to  $\hat{\alpha} = \arg \min_{\alpha \in \mathbb{R}^N} (K\alpha - \mathbf{y})^T (K\alpha - \mathbf{y}) + \lambda \alpha^T K\alpha$ . The ridge regression solution reads:  $\hat{\alpha} = (K + \lambda NI)^{-1} \mathbf{y}$ . To validate the method, we perform a leave-one-out (LOO) cross-validation. Prediction accuracy is assessed by the explained variance  $R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$  where  $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$  is the mean of the true values and  $\hat{y}_i = \hat{f}(\mathbf{x}_i)$ .  $R^2 = 1$  corresponds to a perfect prediction, while we might get  $R^2 < 0$  if the prediction error is high.

Fig. 1 presents the two pipelines. In the FA-only pipeline, the data space is simply  $\mathbb{R}^m$  where  $m$  is the number of voxels selected to predict the score. We use the linear kernel and the regression is thus equivalent to a classical linear ridge regression. In the full-tensor pipeline, the data space is the space of vectors of tensors  $(Sym_3^+)^m$  where  $m$  is the number of voxels selected to predict the score: each subject is represented by a vector of  $m$  tensors. We construct an inner product on this space using the vector representation  $V$  of the tensors : with  $\mathbf{x}_i = (\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,m})^T$  and  $\mathbf{x}_j = (\mathbf{x}_{j,1}, \dots, \mathbf{x}_{j,m})^T$  members of  $(Sym_3^+)^m$ ,  $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \sum_{l=1}^m V(\mathbf{x}_{i,l})^T V(\mathbf{x}_{j,l})$  is an inner product on  $(Sym_3^+)^m$ . We can thus compute a linear kernel for the kernel ridge regression.

### 3 Application to Autism Spectrum Disorders (ASD)

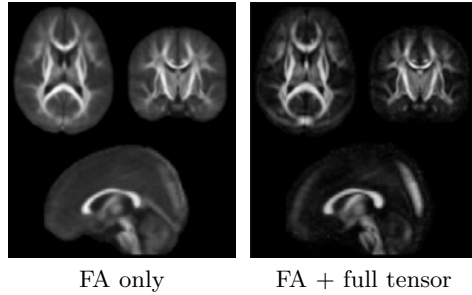
In this section we present the dataset used and compare the results obtained by both approaches for pre-processings, voxelwise regression analysis and brain-wide machine learning.

**Data Acquisition.** The dataset consisted of diffusion MRI scans and autism severity evaluation for thirty children and adolescents with ASD, and lacks control subjects. The autistic syndrome diagnosis was based on DSM IV-TR criteria (Diagnostic and Statistical Manual of Mental Disorders [3]) and its severity was evaluated with the Autism Diagnostic Interview-Revised (ADI-R) algorithm [11] (a high global ADI-R score means severe autism). For all regression analyses, we used a modified version of the global ADI-R score (mADI-R) previously used by Gendry Meresse et al. in [7].

All diffusion MRI scans were acquired on a GE Signa 1.5 MRI system (General Electric) using a birdcage head coil. The sequence was a dual spin echo echo-planar imaging sequence (echo times  $TE = 70$  ms; repetition times  $TR = 8800$  ms; 60 axial slices, 2 mm slice thickness, in plane resolution  $1.875 \times 1.875$  mm, matrix  $128 \times 128$ ). For each slice, five images without diffusion weighting ( $b = 0$ ), and 41 images with diffusion gradients ( $b = 1500$  s/mm<sup>2</sup>) applied along 41 non-collinear directions were acquired. We used the diffusion model pipeline of BrainVISA 3.1<sup>1</sup> to perform a correction of the diffusion images for echo-planar distortions and the evaluation of the diffusion tensor from these corrected images.

**Registration.** We extracted FA values from the tensor images with the diffusion model pipeline of BrainVISA 3.1<sup>1</sup>. All FA images were nonlinearly registered to

<sup>1</sup> BrainVISA, <http://www.brainvisa.info>



**Fig. 2.** Left: mean of the 30 FA images after FA-only registration – Right: FA of the Log-Euclidean mean of the 30 tensor images after FA + full tensor registration

a study-specific FA template with SPM5<sup>2</sup> running on Matlab 7.6. Before further analysis, all FA maps were smoothed with an 8-mm isotropic Gaussian filter. To obtain registered full-tensor images, we used TTK<sup>3</sup> to apply the deformations computed during the FA registration to the tensor fields (*cf* Fig. 1). This warping was done with a log-Euclidean interpolation scheme and a finite-strain reorientation strategy [2]. All maps were smoothed in the log-Euclidean domain with an 8-mm isotropic Gaussian filter. Using the same level of smoothing in the two approaches enables comparison of the results.

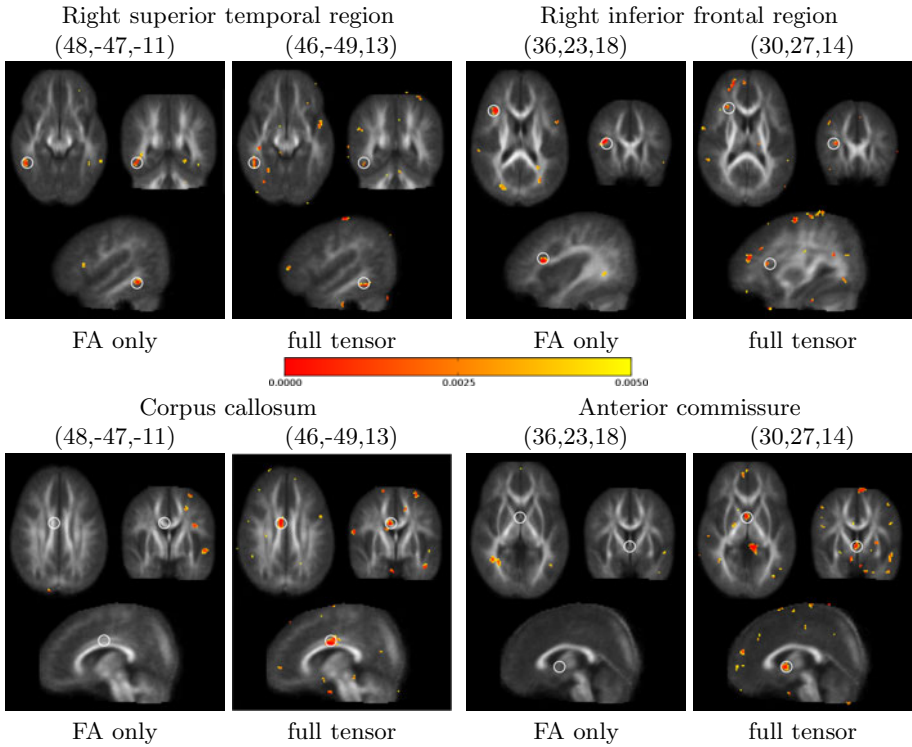
To assess the accuracy of the two methods, we computed the FA mean of the registered images before smoothing that we compared to the FA extracted from the full-tensor mean image (Fig. 2). The FA map obtained from the full-tensor registration is more contrasted than the one obtained from the FA-only registration. This means that the individual images are better aligned to one another, even though the same deformation field is used. This result confirms the importance of the interpolation scheme and the reorientation of tensors.

**Voxelwise Regression Analysis.** Analyses were performed on the whole brain, with an uncorrected voxelwise threshold of  $p < 0.005$ . The full-tensor approach finds all of the regions found by the classical FA-only approach and known to be involved in ASD such as superior temporal regions and inferior frontal regions (Fig. 3, top). Moreover, some significant regions highlighted by the full-tensor approach were not found by the FA-only method (Fig. 3, bottom). In particular, the corpus callosum and the anterior commissure have been mentioned in other studies of autism: [1,12,14,15]. The fact that these regions were detected with the full-tensor approach but not with the FA-only approach confirms that the relevant information lies in the orientation of the tensor as we hypothesized, and not only in its eigenvalues.

**Brain-Wide Score Prediction.** We tested the prediction accuracy with a LOO cross validation. The accuracy was computed for several values of  $m$

<sup>2</sup> Statistical Parametric Mapping, <http://www.fil.ion.ucl.ac.uk/spm/>

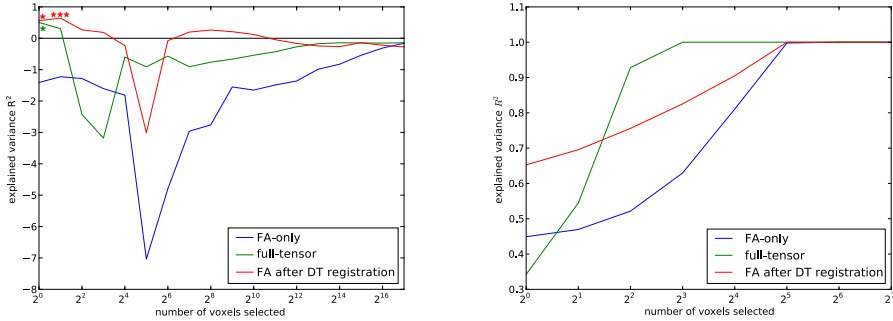
<sup>3</sup> The Tensor Toolkit, <https://gforge.inria.fr/projects/ttk/>



**Fig. 3.** Thresholded  $p$ -values maps for some regions of the brain ( $p < 0.005$ ) – coordinates of the most significant voxel are in Talairach space – Top row: consistent findings of the two approaches in the right superior temporal and inferior frontal regions – Bottom row: the full-DTI approach identified differences in the corpus callosum and the anterior commissure, that were missed by the FA-only analysis.

(number of selected voxels) and the  $p$ -values of the resulting explained variances were corrected across the multiple experiments with a maxT procedure [16].

Fig. 4 (left) shows that the FA-only approach does not yield a satisfying result: the cross-validated  $R^2$  is always negative. Taking the full tensor into account does reach an explained variance of 0.50 with  $p < 0.05$ . However, Fig. 4 (right) shows that the full-tensor approach is more prone to overfitting than the FA-only one. When using the full tensors the dimension of the data space is 6 times the number of voxels selected, so the method overfits as soon as  $6m > 29$  (number of training subjects). Therefore we tried an intermediary alternative between the two approaches: we applied deformations to the tensors but extracted the FA for further analysis. This approach achieves a fair trade-off: the brain-wide kernel regression yields a cross-validated  $R^2$  of 0.64 (with a significance of  $p < 0.005$ ). Extracting a scalar value seems to counteract the overfitting problem. Moreover, extracting the FA to perform the brain-wide prediction can be seen as a way to construct a kernel from the tensor values.



**Fig. 4.** Test  $R^2$  (left) and train  $R^2$  (right) for different numbers of voxels selected – Significance codes: 0 \*\*\* 0.005 \*\* 0.01 \* 0.05 1 – Full-tensor approach overfits rapidly (100% of train explained variance is reached with only  $2^3 = 8$  voxels). This contrasts with FA-only that needed 32 voxels to obtain a perfect fit. Finally, the FA after applying deformations to DTI showed an intermediate behavior.

## 4 Conclusion

The main contribution of this paper was to adapt the DTI analysis pipeline to use the full tensor information in a log-Euclidean framework, with a view to performing brain-wide machine learning on tensor images. We confirmed already observed results: the information contained in the full tensor enhances the correspondence between images and full-tensor voxelwise analysis is more sensitive than the FA-only analysis. We showed that the full-tensor brain-wide score prediction was more prone to overfitting than scalar-based prediction, but that computing the FA after applying deformations to the full tensors yielded better results than the traditional FA-only pipeline.

## References

1. Alexander, A.L., Lee, J.E., Bigler, E.D., DuBray, M.B., Froehlich, A., Lange, N., Fletcher, T.P., Chung, M.K., Lainhart, J.E.: White Matter is Diffusely Affected in Autism. In: Proc. 17th ISMRM, vol. 23, p. 639 (2009)
2. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. *IEEE TMI* 20(11), 1131–1319 (2001)
3. APA: Diagnostic and Statistical Manual of Mental Disorders, 4th edn. (2000)
4. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine* 56(2), 411–421 (2006)
5. Brun, C., Nicolson, R., Leporé, N., Chou, Y.-Y., Vidal, C.N., DeVito, T.J., Drost, D.J., Williamson, P.C., Rajakumar, N., Toga, A.W., Thompson, P.M.: Mapping brain abnormalities in boys with autism. *Human Brain Mapping* 30(12), 3887–3900 (2009)

6. Commowick, O., Fillard, P., Clatz, O., Warfield, S.K.: Detection of DTI white matter abnormalities in multiple sclerosis patients. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 975–982. Springer, Heidelberg (2008)
7. Gendry Meresse, I., Zilbovicius, M., Boddaert, N., Robel, L., Philippe, A., Sfaello, I., Laurier, L., Brunelle, F., Samson, Y., Mouren, M.C., Chabane, N.: Autism severity and temporal lobe functional abnormalities. *Annals of Neurology* 58(3), 466–469 (2005)
8. Hand, D.J., Taylor, C.C.: *Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioural Scientists*. Chapman and Hall, Boca Raton (1987)
9. Hastie, T., Tibshirani, R., Friedman, J.H.: *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York (2001)
10. Leporé, N., Brun, C., Chou, Y.Y., Chiang, M.C., Dutton, R.A., Hayashi, K.M., Luders, E., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE TMI* 27(1), 129–141 (2008)
11. Lord, C., Rutter, M., Le Couteur, A.: Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism. Dev. Disord.* 24(5), 659–685 (1994)
12. Noriuchi, M., Kikuchi, Y., Yoshiura, T., Kira, R., Shigeto, H., Hara, T., Tobimatsu, S., Kamio, Y.: Altered white matter fractional anisotropy and social impairment in children with autism spectrum disorder. *Brain Research*, 1–9 (2010)
13. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. *International Journal of Computer Vision* 66, 41–66 (2006)
14. Stanfield, A.C., McIntosh, A.M., Spencer, M.D., Philip, R., Gaur, S., Lawrie, S.M.: Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. *European Psychiatry* 23(4), 289–299 (2008)
15. Thomas, C., Humphreys, K., Jung, K.-J., Minshew, N., Behrmann, M.: The anatomy of the callosal and visual-association pathways in high-functioning autism: A DTI tractography study. *Cortex*, 1–11 (2010)
16. Westfall, P.H., Young, S.S.: *Resampling-based multiple testing: Examples and methods for  $p$ -value adjustment*. Wiley, Chichester (1993)
17. Whitcher, B., Wisco, J.J., Hadjikhani, N., Tuch, D.S.: Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magnetic Resonance in Medicine* 57(6), 1065–1074 (2007)