Improving spatial normalization of brain diffusion MRI to measure longitudinal changes of tissue microstructure in human cortex and white matter.

Florencia Jacobacci[1], Jorge Jovicich[2], Gonzalo Lerner[1], Edson Amaro Jr[3], Jorge Armony[4], Julien Doyon[5], Valeria Della-Maggiore[1]

[1] Universidad de Buenos Aires. Facultad de Medicina. Departamento de fisiología y biofísica. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Instituto de Fisiología y Biofísica Houssay, Buenos Aires, Argentina,

[2] Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, Italy,

[3] PISA, LIM-44, Instituto de Radiologia, FMUSP, University of Sao Paulo, Sao Paulo, Brazil,

[4] Douglas Mental Health University Institute and McGill University, Montreal, Quebec, Canada,

[5] BIC, MNI, McGill University, Montreal, Quebec, Canada

Abstract

Scalar diffusion tensor imaging (DTI) measures, such as fractional anisotropy (*FA*) and mean diffusivity (*MD*), are increasingly being used to evaluate longitudinal changes in brain tissue microstructure. In this study, we aimed at optimizing the normalization approach of longitudinal DTI data in humans to improve registration in gray matter and reduce artifacts associated with multisession registrations. For this purpose, we examined the impact of different normalization features on the across-session test-retest reproducibility error of *FA* and *MD* maps from multiple scanning sessions. Diffusion data were pre-processed, fit to a tensor model to obtain *FA* and *MD* scalar maps and registered to standard stereotaxic space using different approaches that only differed in the features used in the normalization process, namely: 1) registration algorithm (FSL vs ANTs), 2) target image template (FMRIB58 *FA* vs MNI152 T1), 3) moving image (*FA*, *MD*, *b0*), and 4) normalization strategy (direct vs using an intermediate template). We found that a normalization approach using ANTs as the registration algorithm, MNI152 T1 template as the target image, *FA* as the moving image, and an intermediate *FA* template yielded the highest test-retest reproducibility in registering longitudinal DTI maps for both gray matter and white matter. Our optimized normalization pipeline opens a window to quantify longitudinal changes in microstructure at the cortical level.

Keywords

Diffusion weighted imaging; Diffusion tensor imaging; Registration; Normalization; Reproducibility; ANTs; FSL; longitudinal

1. Introduction

Scalar DTI measures, such as fractional anisotropy (*FA*) and mean diffusivity (*MD*), are increasingly being used in humans to evaluate longitudinal changes in tissue microstructure induced by learning (Landi et al., 2011), development (Krogsrud et al., 2016) or neurodegenerative disease (Keihaninejad et al., 2013). *FA* quantifies the directional diffusion preference of water molecules, primarily reflecting the alignment of fibers in white matter (Beaulieu, 2002). Hence, it has been a useful marker to assess the evolution of several neurological diseases such as multiple sclerosis, amyotrophic lateral sclerosis and Alzheimer's disease (Bodini and Ciccarelli, 2014). In healthy individuals, *FA* has been used to characterize plasticity induced by learning. For example, one week of visuomotor adaptation leads to an increase in *FA* in pyramidal tracts that correlates with the speed of learning (Landi et al., 2011), whereas 3 weeks of training on a juggling task increases *FA* in the posterior parietal cortex (Jan Scholz et al., 2009). Recent studies conducted in rats suggest that these macroscopic changes may reflect augmented myelination (Hughes et al., 2018; Sampaio-Baptista et al., 2013; Swire and ffrench-Constant, 2018).

In contrast, *MD* is a direction-independent measure of the average diffusivity, reflecting water motility, and thus, may be used to estimate microstructural changes both in gray matter (GM) and white matter (WM). *MD* too has long been a marker to probe tissue microstructure in neurological patients. For example, tissue cellularity in brain tumors is well correlated with *MD* (Gauvain et al., 2001), whereas cell-swelling in acute cerebral ischemia is characterized by a short-term decrease in *MD* in the affected region (Benveniste et al., 1992; Davis et al., 1994; Mintorovitch et al., 1991). In addition, this measure has been recently shown to have great potential to detect changes in gray matter induced by learning. Specifically, training on a spatial memory task reduces *MD* in the hippocampus 30 minutes post learning (Sagi et al., 2012). This decrease is associated at the microscopic level with glial hypertrophy, likely induced by LTP-like plasticity (Blumenfeld-Katzir et al., 2011; Sagi et al., 2012). Based on these findings it has been hypothesized that a learning-related reduction in cortical *MD* would be compatible with a drop in the interstitial volume associated with astrocyte hypertrophy.

In sum, both *FA* and *MD* may provide relevant information regarding learning-related alterations in brain microstructure that are physiologically sound. Yet, there are at least two caveats when it comes to using DTI to detect longitudinal changes in plasticity. One concerns the tissue of interest, namely whether it is white or gray matter. To date, there is no single volumetric normalization approach that serves to detect learning-related changes in both gray and white matter using DTI. In fact, most analytical tools including Tract-Based Spatial Statistics (TBSS, Smith et al., 2007, 2006) are optimized to detect differences in white matter microstructure but not in gray matter. The other caveat concerns the artifacts resulting from registering DTI images acquired from multiple sessions to the standard stereotaxic space.

Normal interindividual anatomical variability in fiber tracts can produce residual misalignment when registered to standard space (Smith et al., 2006). Longitudinal approaches are even more susceptible to normalization artifacts due to the potential misalignment between images acquired from the same subject in multiple magnetic resonance imaging (MRI) sessions. Voxel-based morphometry is, in fact, highly sensitive to errors in the normalization process which can yield to false positives (Bookstein, 2001; Schwarz et al., 2014). Yet, to date, there is no consensus regarding the optimal normalization method to process longitudinal DTI data (Papinutto et al., 2013). An optimized normalization pipeline that allows drawing valid conclusions from voxel-wise analysis conducted on multisession *FA* and *MD* maps is therefore indispensable.

In this study, we seek to optimize the normalization approach of longitudinal DTI data with the aim of improving registration in gray matter and reducing artifacts associated with multiple session registration. For this purpose, we compared the across-session test-retest reproducibility error of DTI images (*MD* and *FA*) using four –non-exclusive– normalization approaches. The first approach assessed the *Registration algorithm*. In the widely-used TBSS pipeline from FSL (Smith et al., 2007, 2006, 2004), scalar maps resulting from DTI analyses are normalized to the standard space using FSL's linear and non-linear registration algorithms (*FLIRT* and *FNIRT*, respectively; Smith et al., 2004). However, in recent years one registration tool has gained substantial attention: ANTs (Klein et al., 2009; Avants et al., 2011). Schwarz and collaborators (2014) have shown that the algorithm used by ANTs is more sensitive than TBSS for detecting white matter *FA* changes and leads to lower type I error rates (2014). Here, we aim to estimate the reproducibility of ANTs to detect longitudinal changes in gray matter, in both *MD* and *FA*.

The second normalization approach we evaluated in this study was centered on the *Target image*, i.e., the template used to register DTI images to standard stereotaxic space. The FMRIB58 template is widely used in DTI analyses. It is constructed out of the *FA* maps from 58 subjects normalized to MNI152 standard stereotaxic space. Yet, it is substantially smaller than the MNI152 T1 template due to erosion of the boundaries of the raw individual images to include values of *FA* larger than 0.2. This threshold, aimed at improving alignment where *FA* values are too low, gets rid of a significant amount of gray matter (Smith et al., 2006), rendering it inadequate to detect DTI changes in the cortex. Here, we examined how using the MNI152 T1 image as the standard normalization template impacts on the reproducibility of the DTI normalization process both for *MD* and *FA*.

The third normalization approach we evaluated in this study assessed the *Moving image*, i.e., the DTI image warped to the standard stereotaxic space. Traditional normalization pipelines for DTI (such as TBSS), perform an *FA*-*FA* registration so that the moving image (*FA* scalar map) and the final template (FMRIB58) are of the same modality (Smith et al., 2007,

2006). Yet, if the target image is the MNI152 T1, it is not obvious that *FA* would be the most adequate moving image when one is also interested in measuring *MD* in the cortex. Therefore, in this study, we compared the reproducibility of normalization using five different moving images: *FA*, *MD*, an image in which no diffusion gradient was applied and provides T2-weighted structural information (hereafter called as *b0*), the combination of *FA* and *b0* images, and the combination of *MD* and *b0* images.

Finally, the fourth normalization approach we examined in this study focused on the *Normalization strategy*. In cross-sectional studies, registration of DTI images is usually performed directly to a standard stereotaxic template. Although this approach seems appropriate for unrelated images it may not be optimal for MRI data obtained in repeated acquisitions. Longitudinal studies have the advantage that images acquired from the same subject share common structural information. Thus, the creation of a template for each individual using these images as inputs may reduce bias in registration by treating all of the time points for one given subject similarly. The normalization process will require aligning all of the acquired time points to the individual template and then registering each individual template to the one in standard stereotaxic space. Using an intermediate template for each individual has been recommended over pairwise registration (Klein et al., 2010; Reuter et al., 2012; Tustison et al., 2014) and it is more suitable for datasets with several acquisitions. On the other hand, using a group template as an intermediate step in the normalization process can also help avoiding biases in registration (Klein et al., 2010; Tustison et al., 2014). Here, we compared the direct normalization approach to that achieved using an intermediate individual or group template.

2. Methods

2.1. Participants

Twenty-one healthy subjects between 18 and 31 years old (11 female; ages: mean ± SD = 23.6 ± 3.1) participated in the study. All subjects were healthy volunteers with no self-reported history of psychiatric, neurological or cognitive impairment. Subjects provided written consent and were compensated for their participation. The experimental procedure was approved by the local Ethics Committee and performed according to the Declaration of Helsinki.

2.2. Image acquisition

The MRI dataset used in this study was acquired as part of an ongoing international collaborative project between the Quebec Brain Imaging Network (QBIN) and the Latin American Brain Mapping Network (LABMAN) aimed at studying plasticity induced by learning. Participants trained on three different tasks: visuomotor adaptation, motor sequence learning, and a control task involving no learning. Each task condition required subjects to be scanned before, 30 minutes and 24 hours after training. Hence, each individual was scanned nine times in total with different MRI modalities including diffusion-weighted images (DWI), T1-weighted, T2-weighted and EPI-BOLD images. Here, we report the acquisition parameters corresponding only to the DWI protocol.

Magnetic resonance images were acquired with a 3T Siemens Tim TRIO scanner using a 12-channel head RF receive coil (Instituto Angel Roffo, University of Buenos Aires, Argentina). Each subject's head was positioned inside the head coil using the same anatomical landmarks as reference in all sessions. DWI were acquired using the multiband-accelerated sequence implemented by the Center for Magnetic Resonance Research (CMRR; Ugurbil et al., 2013; Xu et al., 2013). The following protocol was used for acquisition: voxel size=2x2x2 mm^3; field of view (FOV)=240x240 mm^2; 30 monopolar gradient directions uniformly distributed (Jones, 2004; Jones et al., 1999); 70 axial slices; repetition time (TR)=5208 ms; echo time (TE)=89 ms; acquisition time (TA)=3 minutes and 34 seconds; bandwidth (BW)=1488 Hz/Px; multiband (MB) acceleration factor=2, SENSE1 *coil-combine* mode, pure axial slice orientation with interleaved slice acquisition, anterior-posterior (A-P) phase encoding direction, with a b-value=1000 s/mm^2. Phase encoding in the anterior-posterior direction was chosen to preserve hemispheric symmetry (Smith et al., 2007). Eight *b0* volumes were acquired using an A-P phase encoding direction: two were acquired at the beginning of the sequence, one at the end and the rest interleaved every five b-1000 volumes. This configuration optimizes the signal-to-noise ratio of scalar images resulting from the fit of a diffusion tensor model (Jones et al., 1999). In addition, one *b0* volume was acquired with posterior-anterior (P-A) phase encoding direction to correct for susceptibility-induced geometric distortions (Andersson et al., 2003).

2.3. Image pre-processing

DWI *DICOM* images were converted to *NIFTI* format using the dcm2nii software (Li et al., 2016). Pre-processing steps for DWI were conducted for each of the nine scanning sessions separately, and included: i) correction of susceptibility-induced distortions using FSL's *topup* tool with *b0* volumes acquired with opposite phase encoding direction (Andersson et al., 2003), ii) correction of eddy currents-induced distortions, head motion correction and b-vector rotation using FSL's *eddy* tool (version 5.0.9; Andersson and Sotiropoulos, 2016). Next, DTIfit (FSL, Smith et al., 2004) was used to fit a diffusion tensor model to produce the scalar measures of interest: fractional anisotropy (*FA*) and mean diffusivity (*MD*). The "halo" of bright voxels that typically surrounds *FA* images due to eddy currents-induced distortions in cerebrospinal fluid (CSF; Bastin, 1999; Jones and Cercignani, 2010) was removed by eroding it with a spherical kernel of 6 mm radius (Smith et al., 2007, 2006). A brain mask generated out of the eroded *FA* image was subsequently applied to the associated *MD* and *b0* images. The resulting eroded *FA*, *MD* and *b0* images were then used to evaluate different normalization approaches.

2.4. Normalization approaches

After pre-processing, the DTI scalar maps were normalized using four non-exclusive approaches that varied only in the registration features chosen to bring them into stereotaxic space. Only DWI data acquired during the baseline and the 24h session of the control condition, from now on referred to as *test* and *retest images*, were used to compute the across-session test-retest reproducibility error. Given the short time interval between these sessions and the absence of a learning manipulation, we assumed that the across-session variability of tissue microstructure metrics would mostly reflect reproducibility errors related to the MRI acquisition protocol and the analysis pipelines.

The following features were assessed in each normalization approach: 1) the registration algorithm used to compute and apply the transformations for image normalization (FSL vs ANTs), 2) the target image, i.e. the image in standard stereotaxic space used as the reference in the normalization process (FMRIB58 vs MNI152 templates), 3) the moving image, i.e. the image warped to standard space (*MD*, *FA* or *b0*), and 4) the normalization strategy, i.e., whether they were directly warped to the stereotaxic space or through an intermediate template. These normalization approaches were assessed sequentially, taking the feature of the most reproducible pipeline as the default of the subsequent approach. These pipelines, with the respective manipulated variables and fixed parameters are outlined in Table 1 and described in detail as follows.

Table 1. Outline of the four brain spatial normalization approaches evaluated in this diffusion MRI study. For each approach, we show the parameters that were fixed and the tissues in which test-retest reproducibility errors were assessed. Acronyms: fractional anisotropy (FA), mean diffusivity (MD), mean unwweighted difusion volume (b0), white matter (WM), gray matter (GM).

Human brain spatial normalization approach	Pipeline	Fixed parameters	Tissue in which test-retest reproducibility error (%) was evaluated
1) Registration algorithm	FSL ANTs	Target template: FA FMRIB58 Moving image: FA Normalization strategy: Direct	Tissue: WM, from template mask Error metrics: FA, MD
2) Target image template	T1 MNI152 FA FMRIB58	Registration algorithm: ANTs Moving image: FA Normalization strategy: Direct	Tissue: WM, from template masks Error metrics: FA, MD
3) Moving image	FA MD b0 FA+b0 MD+b0	Registration algorithm: ANTs Target template: T1 MNI152 Normalization strategy: Direct	Tissue: WM and GM, from template mask Error metrics: FA, MD
4) Normalization strategy	Direct (no intermediate template) Intermediate individual FA template Intermediate group FA template	Registration algorithm: ANTs Target template: T1 MNI152 Moving image: FA	Tissue: WM and GM, from template mask Error metrics: FA, MD

2.4.1. Registration algorithm

Some of the most common pipelines for DTI analyses involve the use of normalization algorithms from FSL (Smith et al., 2007, 2006, 2004). However, recently ANTs' non-linear normalization algorithm (Avants et al., 2011) has been shown to outperform FSL's on various

metrics (Klein et al., 2009; Schwarz et al., 2014). To assess the reproducibility of the DTI normalization approach from FSL versus the one from ANTs, we contrasted the following:

i. A normalization pipeline based on FSL's *FLIRT* and *FNIRT* registration algorithms.
ii. A normalization pipeline based on ANTs' linear and non-linear registration algorithms.

Given that *FNIRT* from FSL is optimized to warp images of the same modality (Andersson et al., 2007; Smith et al., 2004), *FA* images from the *test* and *retest* sessions were registered to the FMRIB58 *FA* template for both pipelines. Because a large portion of the cortex is missing from the FMRIB58 template, reproducibility was evaluated only for the WM tissue (see section 2.5. for details on the mask).

For linear registration using FSL's *FLIRT* tool (FSL version 5.0.9), the default cost function (correlation ratio) was used, which normally allows the robust registration of all images including those with different contrasts. For non-linear registration using *FNIRT*, the only cost-function presently implemented is the "sum-of-squared differences" (Smith et al., 2004). We used the configuration file provided in FSL's toolbox for registration of *FA* images to FMRIB58 *FA* template (*FA_2_FMRIB58_1mm.cnf*). Linear and non-linear transformations were concatenated and applied to *FA* and *MD* maps using a single interpolation step with the tri-linear (default) method.

For linear registration using ANTs (version 2.2.0), translation, rigid and affine transformations were consecutively calculated using the following parameters: mutual information similarity metric, convergence threshold = $1x10^{-6}$, convergence window size = 20, gradient step = 0.1. For the non-linear transformation the symmetric normalization (*SyN)* algorithm (*antsRegistration* command) was used with the following parameters: mutual information similarity metric, 100x100x50 iterations in three resolution levels with shrink factors = 3x2x1 and smoothing sigmas = 4x2x1, convergence window size = 5, gradient step = 0.2, update field variance in voxel space = 3, total field variance in voxel space = 0. Linear and non-linear transformations were concatenated and applied to *FA* and *MD* maps using a single linear interpolation step.

Given that ANTs yielded better reproducibility than FSL we used ANTs' algorithm for the remaining normalization approaches.

2.4.2. Target image

If one is interested in analyzing diffusion parameters in the neocortex, the template traditionally used in DTI studies (FMRIB58) may be suboptimal (Fig. 1). This is due to the fact that the outermost edges of the cortex, containing voxels that are primarily GM or CSF were excluded

from the template. Thus, we propose the use of the MNI152 T1-based template as target of registration.

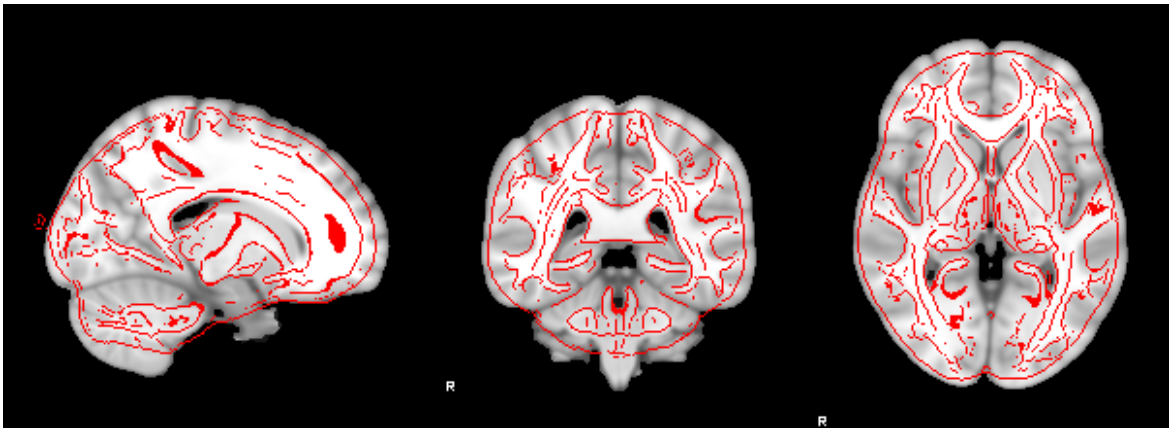

Figure 1. The choice of normalization template matters: comparison of the two adult human brain templates commonly used as targets for spatial normalization in group studies. The FMRIB58 FA template (in red) is shown overlaid on MNI152 T1 template. Note the substantial amount of the cortex excluded from the FA template with respect to the T1 template.

The reproducibility associated with normalizing *FA* and *MD* images to two widely-used templates of different modalities was contrasted using the following pipelines:

i. A normalization pipeline using FMRIB58 *FA* template as the target image
ii. A normalization pipeline using MNI152 T1 template as the target image

Test and *retest FA* images were registered to either the *FA* template or the T1 template by concatenating linear and non-linear transformations using ANTs. The same parameters used for the *Registration algorithm* approach described in the previous section were chosen here. Transformations calculated for *FA* images were also applied to the corresponding *MD* images. We were only able to contrast the reproducibility for the pipelines at the level of white matter due to the lack of cortical gray matter information in the FMRIB58 template.

Given that using the MNI152 T1 template as target of normalization showed higher reproducibility for *MD* and *FA* than the FMRIB58 template, we chose it as the target image for the remaining normalization approaches.

2.4.3. Moving image

When the moving image and the target of normalization are of different modalities, as it is here, the choice of the DTI map to be warped to the standard stereotaxic space is not obvious. We compared the reproducibility for the normalization to the MNI152 template using five pipelines that differed only in the moving image:

i. A normalization pipeline using *FA* as moving image
ii. A normalization pipeline using *MD* as moving image
iii. A normalization pipeline using the *b0* as moving image
iv. A normalization pipeline combining information from *FA* and *b0* images (*FA*+*b0*) as moving image
v. A normalization pipeline combining information from *MD* and *b0* images (*MD*+*b0*) as moving image

In pipelines i) and ii), we assessed if the reproducibility of *FA* or *MD* is optimized when using the scalar image of interest as moving image for the normalization process. Using multiple modalities can improve pairwise registrations by providing complementary contrasts (Avants et al., 2011). Given that ANTs allows the use of one or more modalities to drive registration, in pipelines iii) through v) we examined if adding T2-weighted structural information from the *b0* image leads to a better registration of all tissues.

For each subject's *test* and *retest* sessions, the moving image in pipelines i) through v) were registered to the MNI152 T1 template by concatenating linear and non-linear transformations using ANTs (parameters for registration with ANTs used in the *Registration algorithm* approach were maintained). Then, these transformations were applied to the *FA* and *MD* maps for the same subject and session. Given that the MNI152 T1 template contains information from gray matter and white matter, we were able to compare the reproducibility error of *FA* and *MD* in both types of tissue. To this end, we created voxel-wise reproducibility error maps, and obtained the mean for gray and white matter within each corresponding mask (see section 2.5. for details).

Given that *FA* was the moving image yielding the best reproducibility both for *MD* and *FA*, we used this moving image for the last normalization approach.

2.4.4. Normalization strategy

The first three normalization approaches proposed in this work were based on pipelines in which registration of the moving image to the target was performed in a direct fashion. The direct approach does not differ from what is often used in cross-sectional studies, but it may not be the best normalization strategy to make the most out of longitudinal data. One of the benefits of having a longitudinal data set is that within-subject variability may be reduced by taking advantage of the images acquired at different time points to create an intermediate template for normalization. It has been shown that the use of an intermediate template improves alignment over direct pairwise registration (Klein et al., 2010; Reuter et al., 2012; Tustison et al., 2014). Therefore, we evaluated two pipelines using different intermediate templates created with ANTs:

i. A normalization pipeline using an individual *FA* template as intermediate template
ii. A normalization pipeline using a group *FA* template as an intermediate template

When having two time points for the same subject, a common strategy to create an individual template is to refer both images to their halfway point or mid-space (Thomas et al., 2009). In contrast, when the number of time points is larger, as in this case in which 9 images were obtained per subject, creating a mid-space for all time points is cumbersome. ANTs provides a tool that allows for the construction of an unbiased multivariate template using images from many sessions (and modalities) from the same subject (Avants et al., 2011, 2010). The intermediate templates proposed in pipelines i) and ii) were created using this tool: *antsMultivariateTemplateConstruction* (Avants et al., 2011, 2010).

DWI images from all nine sessions were used as input for the construction of each individual *FA* template. For the creation of the group *FA* template, individual *FA* templates from all subjects (twenty-one) were used as input. The following parameters were chosen for the template construction: rigid-body registration of inputs for the creation of an initial template used as seed, gradient step size=0.2, cross-correlation similarity metric, Greedy-SyN transformation model used for registration. Once the intermediate *FA* templates were produced (one individual template per subject, and one group template), they were registered to the MNI152 T1 using ANTs' linear and non-linear transformations.

The template construction tool automatically provides the transformations that map each one of the input images to the output template of choice. Hence, for the normalization of images in pipeline i), the transformations that map the *test* and *retest FA* images to the individual template were concatenated to those that map the individual template to MNI152 space. Similarly, for pipeline ii), the transformations that map the *test* and *retest FA* images to the individual template, the corresponding individual template to the group template and the group template to MNI152 space were concatenated. Then, transformations were applied to the *FA* and *MD* maps in subject space to align them to the MNI152 standard space in a single interpolation step.

To evaluate if adding an intermediate template to the normalization process improved reproducibility in comparison with a direct normalization approach, we contrasted the reproducibility from pipelines i) and ii) with that obtained for the pipeline in which normalization was performed directly to the MNI152 T1 template.

2.5. Data analysis

For all normalization approaches, the performance of the different pipelines was established based on the across-session test-retest reproducibility error using binary masks for each tissue type, namely gray and white matter. Furthermore, to assess the effect of applying the

optimal normalization pipeline on the integrity of the DTI images, we calculated the signal-to-noise ratio (SNR) of *MD* and *FA* on gray and white matter tissue.

2.5.1. Across-session test-retest reproducibility error

Across-session test-retest reproducibility errors (RE) of *FA* and *MD* were computed in a voxel-wise fashion for each subject and each normalization approach as the absolute difference between the *test* and the *retest* DTI measure, divided by the mean value of both sessions, and multiplied by 100 to express it as percent change (Papinutto et al., 2013).

$$RE = \frac{100 * |TEST - RETEST|}{0.5 * (TEST + RETEST)} \quad \text{(Eq.1)}$$

To assess differences in RE for each tissue type, we computed the average RE for each DTI measure over the voxels within a GM and a WM mask. Masks were created by segmenting the MNI152 T1 template into three tissues (GM, WM and CSF) using FSL's FAST tool (Zhang et al., 2001). The GM component obtained from the segmentation was used to create an atlas GM mask. An additional subcortical mask was created using the Harvard-Oxford subcortical structural atlas (Smith et al., 2004; thresholded=60% probability) to include a series of structures of interest that were missing from the automatically generated GM component. This included the right and left pallidum, putamen, caudate and thalamus. Cortical and subcortical GM masks were combined and binarized into a unique GM mask. On the other hand, the WM mask was created based on the intersection of the components automatically generated by segmenting WM tissues from the MNI152 template and the FMRIB58 template (the FMRIB58 is substantially smaller due to thresholding). This allowed comparing RE in WM for images normalized with either template.

2.5.2. Signal-to-noise ratio (SNR) assessment

To obtain the SNR, we computed the mean across each normalized *test* DTI map and divided it by its standard deviation (Farrell et al., 2007). Thus, for each subject we generated one SNR value per DTI measure (one for *MD* and one for *FA*) and tissue type (GM and WM masks).

2.5.3. Statistical analysis

Statistical analysis was performed using SPSS (IBM SPSS Statistics for Windows, version 25.0). RE and SNR values were statistically compared using a repeated measures analysis of variance (ANOVA), with *pipeline*, DTI measure and tissue type as within-subject factors. Normality of the data was checked using Shapiro-Wilk's test. For analyses involving comparisons between more than two pipelines, sphericity of the data was tested using

Mauchly's test. In the cases in which the sphericity assumption was not met, a Greenhouse-Geisser degrees of freedom correction was performed. One subject consistently appeared as an outlier in the *Moving image* approach. Consequently, it was removed from all the analyses. Post-hoc Tukey tests were used to examine specific differences between pipelines. Bonferroni correction was used to adjust the significance threshold for multiple comparisons.

2.6. Data and code availability statement

The source-code of the optimized pipeline proposed for longitudinal DTI normalization is publicly available in GitHub (see section 3.4 for the link). The dataset used for this work is available upon request.

3. Results

DWI data were pre-processed, fit to a tensor model to obtain *FA* and *MD* scalar maps and registered to standard stereotaxic space using four approaches that only differed in the features used in the normalization process, namely: 1) registration algorithm, 2) target image, 3) moving image and 4) normalization strategy.

3.1. Registration algorithm: ANTs yields lower reproducibility error than FSL

With the aim of finding the optimal registration algorithm we first inspected the quality of registration visually by overlaying registered FA maps (red edges) on the FMRIB58 FA template (background image). Visual inspection of normalized *FA* images yielded overall better anatomical alignment to the standard template for ANTs than for FSL. Figure 2 illustrates this observation for one representative image.

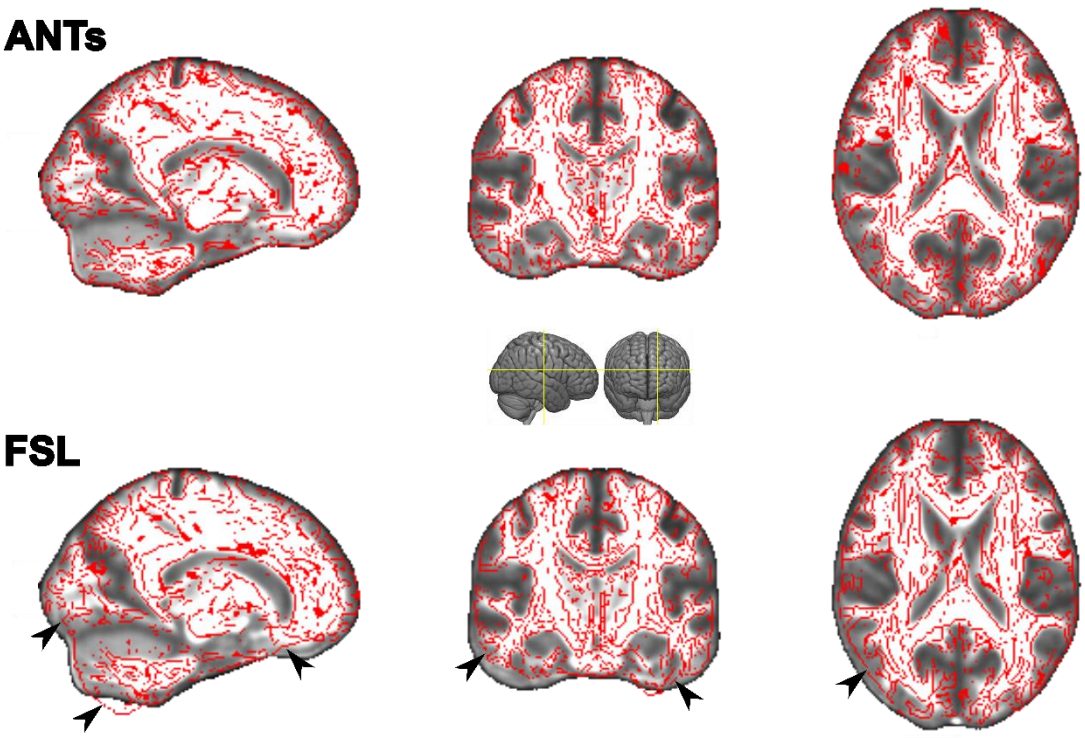

Figure 2. The choice of non-linear registration tool matters: comparison of registration to the FMRIB58 FA template (background image) from a sample subject using two different registration tools, ANTs (top row) and FSL (bottom row). Registered FA maps are displayed using red edges. Notice the suboptimal warping of FSL's tools in frontal, temporal and posterior regions (marked with arrows).

Next, we compared the RE obtained from a normalization pipeline using either FSL (Smith et al., 2007, 2006, 2004) or ANTs (Avants et al., 2011). Given that *FSL's* most used algorithm for non-linear registration (*FLIRT*) is optimized to warp images of the same modality,

FA images were registered to the FMRIB58 *FA* template for both pipelines. Thus, reproducibility was evaluated only for the WM tissue.

Quantitative assessment of the algorithm's reproducibility is shown in Figure 3. The upper row shows the distributions of reproducibility errors on the white matter mask of the FMRIB58 *FA* template, for *MD* (left) and *FA* (right). For both metrics we found that ANTs yielded lower percent reproducibility errors: *MD* (Mean±SE: RE=7.75±0.18% for FSL vs. RE=6.24±0.11% for ANTs; $F(1,19)=126.41$, $p<0.0001$) and *FA* (RE=17.11±0.28% for FSL vs RE=10.16±0.11% for ANTs; $F(1,19)=1103.83$, $p<0.0001$). The lower row of Figure 3 shows the color coded (% errors) voxel-wise spatial map of mean RE computed across the WM mask (FMRIB58 *FA* template). Note that the reproducibility errors were higher for *FA* than for *MD* in both normalization pipelines (main effect of DTI measure $F(1,19)=6095.44$, $p<0.0001$, with mean RE=13.64±0.19 for FA vs RE=6.99±0.13 for MD). More details can be found in Supplementary Table 1.

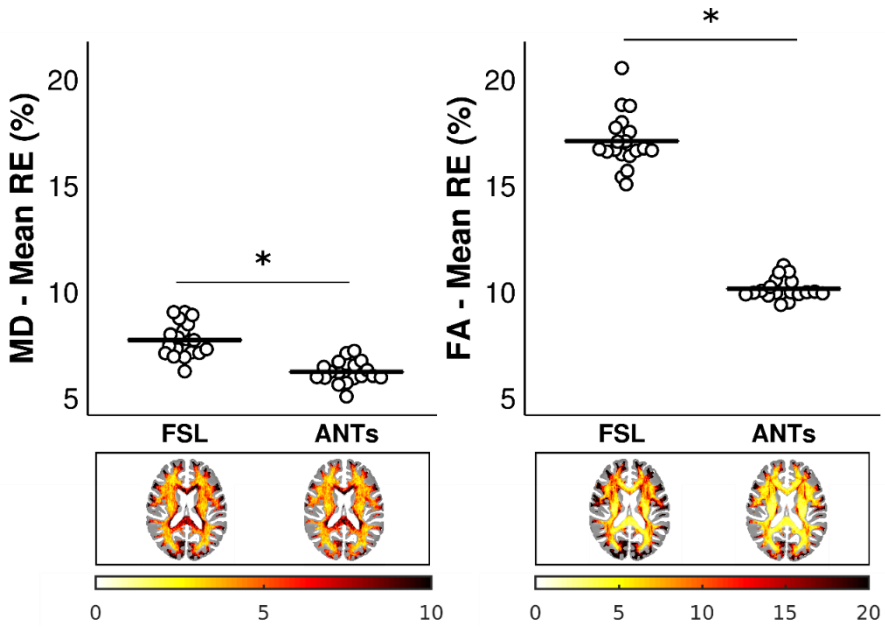

Figure 3. The choice of non-linear registration tool (FSL or ANTs) used with the FMRIB58 *FA* template affects test-retest reproducibility of diffusion scalars in white matter. ANTs yields significantly lower reproducibility errors than FSL (*p<0.0001).The upper plots show the mean percent test-retest reproducibility error for MD (left) and FA (right). The lower row shows the color-coded anatomical distribution of percent reproducibility errors in the white matter mask of the FA template; axial slice coordinate: z=19 mm. Acronyms: mean diffusivity (MD), fractional anisotropy (FA), test-retest reproducibility error (RE).

3.2. Target image: MNI152 T1 template yields lower reproducibility error in WM than FMRIB58 FA template

With the aim of optimizing the normalization process to include the cortex, we compared the RE associated with registering *MD* and *FA* maps to the FMRIB58 *FA* template with that

obtained from registering them to the MNI152 T1 template. RE was evaluated on the WM mask. We used ANTs for the registration given the improved reproducibility performance shown in Figure 3.

We found that reproducibility errors were significantly lower when using MNI152 as target image (Fig. 4) both for *MD* (RE=5.58±0.13% when normalized to MNI152 vs RE=6.24±0.11% when using FMRIB58 as target; $F(1,19)=108.15$; $p<0.0001$) and *FA* (RE=9.25±0.20% when normalized to MNI152 vs RE=10.16±0.11% when using FMRIB58 as target; $F(1,19)=58.76$; $p<0.0001$). In line with the previous normalization approach, RE was higher for *FA* than *MD* for both pipelines (main effect of DTI measure $F(1,19)=2993.47$, $p<0.0001$, with mean RE=9.70±0.15 for FA vs RE=5.91±0.12 for MD). More details can be found in Supplementary Table 2.

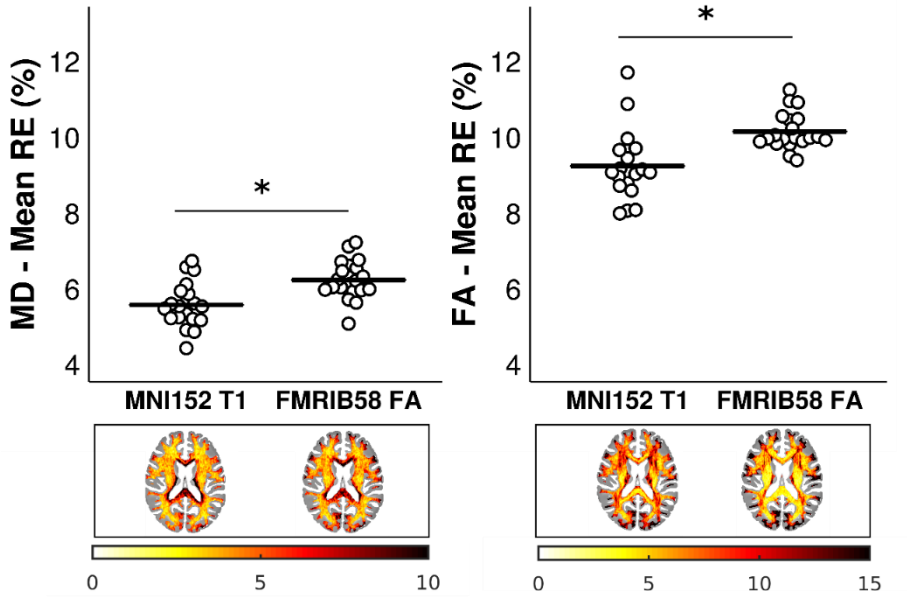

Figure 4. The choice of target template affects reproducibility errors in white matter *MD* and *FA*. MNI152 T1 template produced significantly lower reproducibility error than FMRIB58 FA template (*p<0.0001). Similarly to Fig. 3, the upper row shows the group % error distributions and the lower row the voxel-wise % error maps for each diffusion metric (*MD* and *FA*) and each target template (MNI152 T1 and FMRIB58 FA). Acronyms: mean diffusivity (MD), fractional anisotropy (FA), test-retest reproducibility error (RE).

3.3. Moving image: Using FA as moving image reduces reproducibility error both in GM and WM

Given that so far the optimal normalization approach is based on the MNI152 T1 template as target of normalization the choice of moving image is not obvious. Here, we contrasted the RE obtained from five normalization pipelines differing only in the moving image: i) *FA*, ii) *MD*, iii)

b0, iv) *FA*+*b0*, and v) *MD*+*b0*. In this case, we were able to assess the RE both in white matter and gray matter using corresponding tissue masks.

The assessment of RE in white-matter tissue yielded a significant effect of moving image both for *MD* ($F(1.909,36.272)=9.483$; $p<0.001$) and *FA* ($F(2.406,45.714)=199.158$; $p<0.0001$). A post-hoc test conducted on the *MD* measure, revealed that the "*b0*" pipeline yielded higher RE than the other four pipelines ($p<0.01$), while the rest of the pipelines did not differ from one another (Fig. 5, top left). On the other hand, a post-hoc test conducted on the *FA* measure, showed that "*FA*+*b0*" and "*FA*" pipelines yielded the lowest RE, and that these pipelines differed from the rest ($p<0.0001$). Combining *FA* and *b0* images did not significantly improve reproducibility over using *FA* alone ($p=0.107$). These results are depicted in Figure 5 (top, right).

The assessment of RE in gray-matter tissue yielded a significant effect of moving image on RE both for *MD* ($F(2.413,45.838)=39.856$; $p<0.0001$) and *FA* ($F(2.395,45.500)=1078.86$; $p<0.0001$). A post-hoc test conducted on the *MD* measure (Fig. 5, bottom left), showed that both the "*FA*+*b0*" and the "*FA*" pipelines yielded the lowest RE ($p<0.05$), both being equally reliable ($p=0.177$). A post-hoc test conducted on the *FA* measure (Fig. 5, bottom right), showed that the "*FA*" pipeline yielded the lowest RE ($p<0.001$). Voxel-wise distribution of RE is depicted at the bottom of Figure 5. Refer to Supplementary Tables 3 and 4 for additional information regarding differences between these pipelines and post-hoc comparisons.

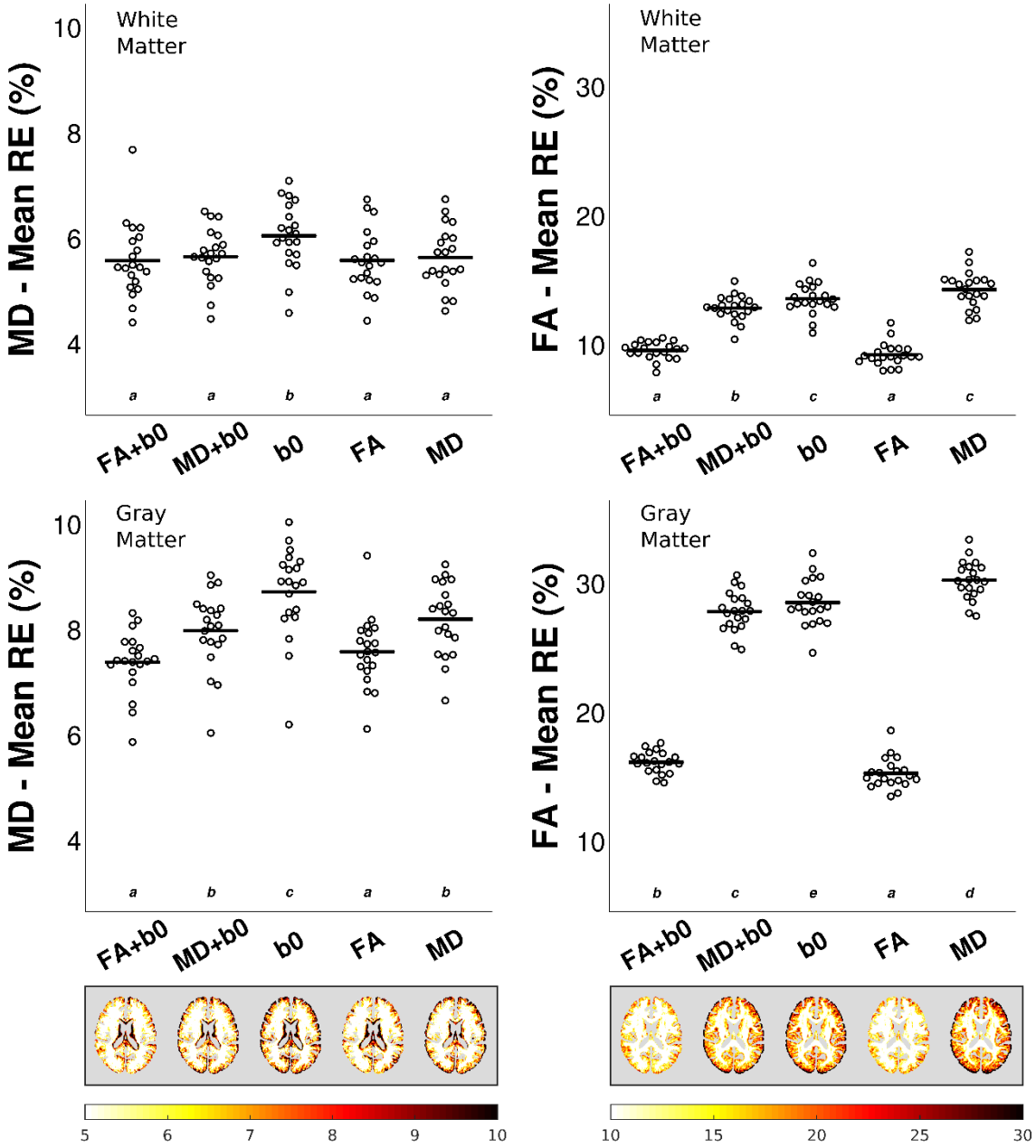

Figure 5. The choice of moving image warped to the MNI152 T1 template affects test-retest reproducibility error. Using FA as moving image reduces reproducibility error significantly in both brain tissues (GM and WM), for both FA and MD ($p<0.001$). Shown are the mean RE (top), and the color-coded anatomical distribution of RE for MD (left) and FA (right) in white and gray matter. Letters above the horizontal axis represent the compact display of all pair-wise comparisons using Tukey's test. Different letters express differences between pipelines with an adjusted p-value<0.05. Same letters indicate no statistical differences. As indicated in Eq. (1) RE is expressed as percent change. Acronyms: mean diffusivity (MD), fractional anisotropy (FA), test-retest reproducibility error (RE), gray matter (GM), white matter (WM), mean diffusion unweighted volume (b0).

In conclusion, regardless of the measure (*MD* or *FA*) and the tissue (WM or GM), *FA* is the most reliable moving image to use in the registration to the MNI152 T1 template.

Consistent with results from the previous sections, the reproducibility error associated with *FA* was higher than *MD* for all pipelines and across tissues (main effect of DTI measure $F(1,19)=6799.91$, $p<0.0001$, with mean RE=17.76±0.21% for *FA* vs RE=6.84±0.3% for *MD*). Note that reproducibility errors were higher in GM than in WM, regardless of diffusion metric (main effect of tissue type $F(1,19)=6151.91$, $p<0.0001$, with mean RE=15.79±0.18 for GM vs RE=8.81±0.14 for WM).

3.4. Normalization strategy: Using an intermediate individual FA template for normalization reduces reproducibility error in GM and WM.

So far, our results indicate that using ANTs for registration, the MNI152 T1 template as target image and the *FA* map as moving image yield the optimal normalization approach for longitudinal DTI images. However, normalization steps in previous sections were performed in a direct fashion, that is, the moving image was aligned to the target image with no intermediate registration steps. Here we examined the reproducibility error associated with pipelines involving two different intermediate templates: an individual *FA* template and a group *FA* template. We compared reproducibility results from these two pipelines with those from the direct pipeline using "*FA*" as moving image, which yielded the lowest RE for *FA* and *MD* in both tissues, as shown in the previous section.

Reproducibility error in white matter differed significantly both for *MD* ($F(2,38)=10.188$; $p<0.0001$) and *FA* ($F(1.510,28.684)=98.573$; $p<0.0001$). A post-hoc test showed that using an individual *FA* template as an intermediate step in the normalization process yielded the lowest RE (Fig. 6, top panel), and differed from the rest of the pipelines for both *MD* ($p<0.03$) and *FA* ($p<0.0001$).

Reproducibility error in gray matter also differed significantly both for *MD* ($F(1.447,27.498)=3.950$; $p=0.043$) and *FA* ($F(1.393,26.472)=230.70$; $p<0.0001$). A post-hoc test revealed that using an intermediate individual *FA* template yielded the lowest RE (Fig. 6, bottom panel), and differed from the rest of the pipelines for *FA* ($p<0.02$). For MD, using the individual FA template as an intermediate step produced lower RE than using the direct registration strategy ($p=0.001$) but it did not differ from using the group FA template ($p=0.2$). Voxel-wise maps for the mean RE across subjects are shown in Figure 6 to illustrate RE distribution. Refer to Supplementary Tables 5 and 6 for additional information regarding differences between these pipelines and post-hoc comparisons.

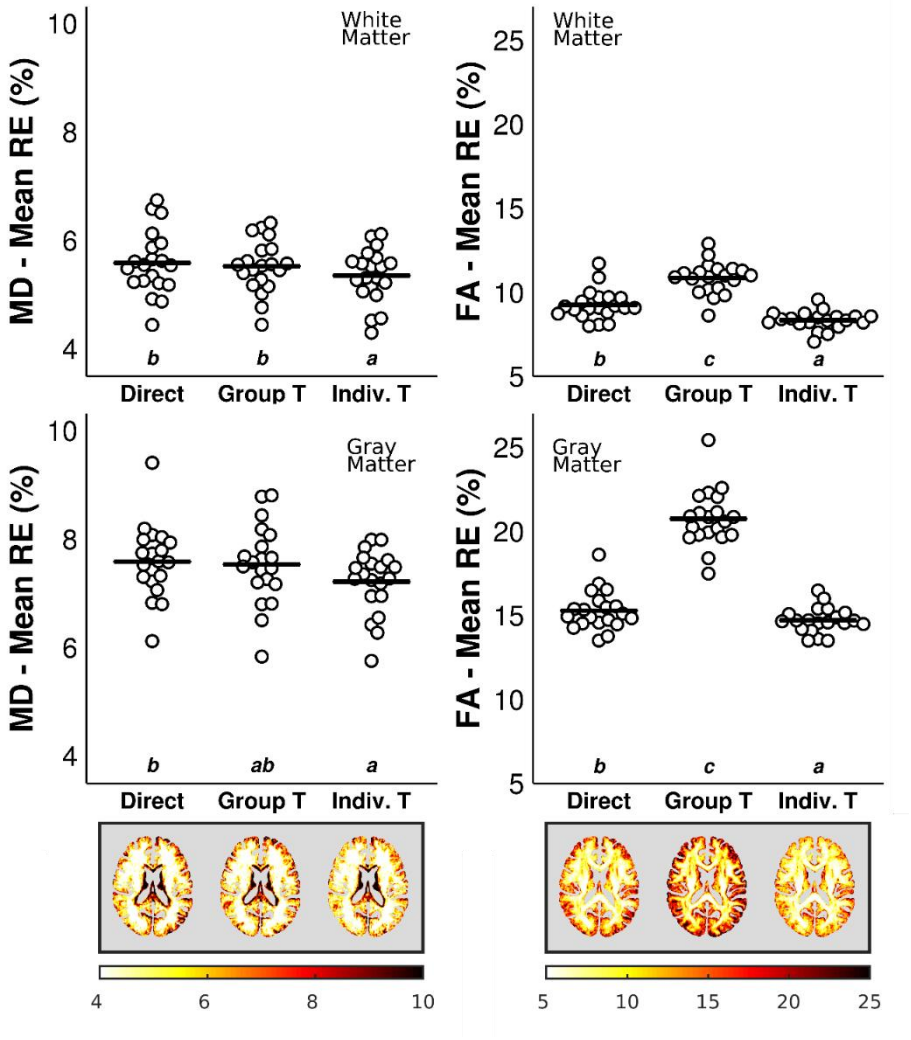

Figure 6. The intermediate individual FA template reduces reproducibility error in GM and WM. Shown are the mean RE (top), and the color-coded anatomical distribution of RE for MD (left) and FA (right) for white and gray matter. As indicated in Eq. (1) RE is expressed as percent change. **Direct**: direct normalization, **Group T**: normalization via group FA template (created out of 21 individual templates), and **Indiv. T**: normalization via individual FA template (created out of 9 scans per subject). Letters above the horizontal axis represent the compact display of all pair-wise comparisons using Tukey's test. Different letters express differences between pipelines with an adjusted p-value<0.05. Same letters indicate no statistical differences. Acronyms: mean diffusivity (MD), fractional anisotropy (FA), test-retest reproducibility error (RE), gray matter (GM), white matter (WM).

Consistent with results from the previous sections, the reproducibility error associated with *FA* was higher than *MD* for all pipelines and across tissues (main effect of DTI measure $F(1,19)=2866.98$, $p<0.0001$, with mean RE=13.20±0.18% for *FA* vs RE=6.46±0.11% for *MD*). Note that reproducibility errors were higher in GM than in WM, regardless of pipeline and diffusion metric (main effect of tissue type $F(1,19)=3590.21$, $p<0.0001$, with mean RE=12.18±0.16 for GM vs RE=7.48±0.12 for WM).

So far we have focused on the test-retest percent reproducibility error as index to evaluate normalization approaches. Yet, an optimal approach in terms of RE may also impact on the integrity of the image, hindering the signal-to-noise ratio (SNR). Figure 7 depicts the mean SNR computed for *FA* and *MD* in WM and GM for the *Normalization strategy* approach.

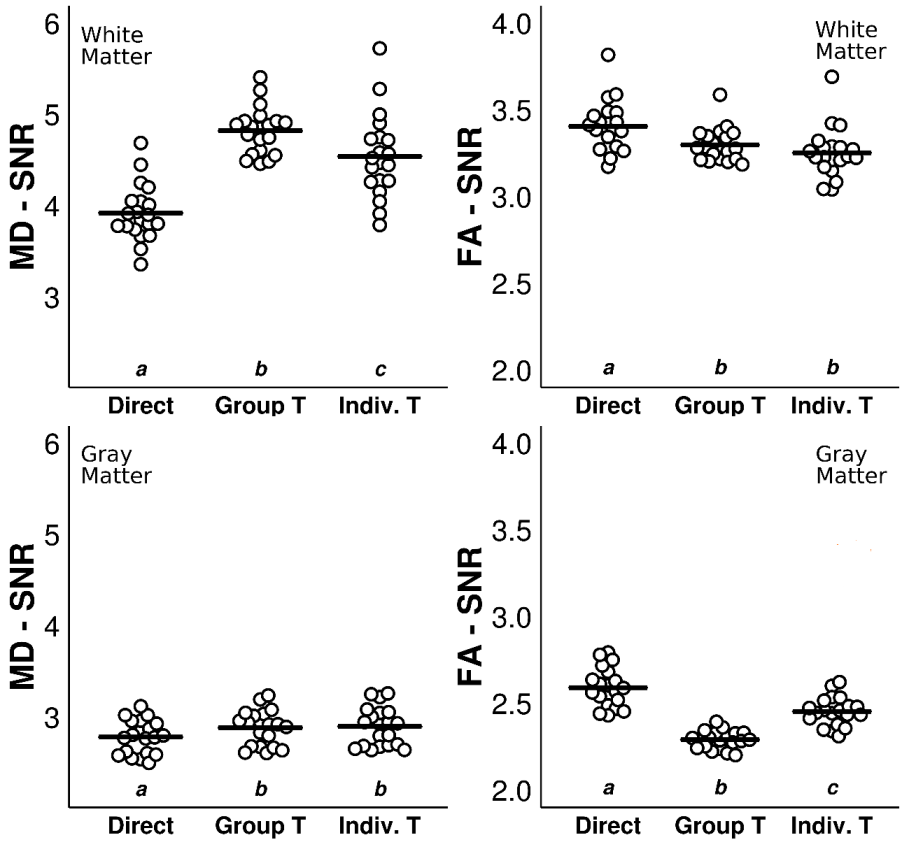

Figure 7. The choice of an intermediate template for brain spatial normalization strategy affects the global signal-to-noise ratio (SNR) in *MD* and *FA* maps. Note that different scales are used for panels showing MD and FA results. **Direct**: direct normalization, **Group T**: normalization via group FA template (created out of 21 individual templates), and **Indiv. T**: normalization via individual FA template (created out of 9 scans per subject). Letters above the horizontal axis represent the compact display of all pair-wise comparisons using Tukey's test. Different letters express differences between pipelines with an adjusted p-value<0.05. Same letters indicate no statistical differences. Acronyms: mean diffusivity (MD), fractional anisotropy (FA).

Statistical assessment of SNR in white matter identified a significant effect of normalization strategy for *MD* ($F(1.460,27.737)=167.543$; $p<0.0001$) and *FA* ($F(2,38)=19.056$; $p<0.0001$). A post-hoc test conducted on *MD* revealed that using a group *FA* template as an intermediate step for normalization pipeline yielded the highest SNR ($p<0.001$). On the other hand, a post-hoc test conducted on *FA* showed that the direct normalization using *FA* as moving image yielded the highest SNR ($p<0.005$).

Statistical assessment of SNR in gray matter also identified a significant effect of normalization strategy both for *MD* ($F(1.440,27.358)=22.518$; $p<0.0001$) and *FA* ($F(1.521,28.904)=168.093$; $p<0.0001$). A post-hoc assessment on *MD* showed that normalization pipelines using an intermediate template (individual *FA* template or group *FA* template) yielded the highest SNR ($p<0.0001$), with no difference between them. Post-hoc assessment on *FA* showed that the direct pipeline produced the highest SNR ($p<0.0001$). Refer to Supplementary Tables 7 and 8 for additional information regarding differences between these pipelines and post-hoc comparisons.

In sum, using an intermediate individual *FA* template yields the lowest RE for both *MD* and *FA* across tissues. This approach was associated with high SNR for *MD* both in GM and WM. In contrast, using a direct normalization strategy, i.e. registering *FA* directly to MNI152 template, yielded the highest SNR for *FA* in both tissues. In agreement with the previous approaches, RE was higher for *FA* than *MD*, and the SNR was higher for *MD* than for *FA*.

The codes for the optimal normalization pipeline are publicly available at: https://github.com/florjaco/DWIReproducibleNormalization.

4. Discussion

Brain diffusion MRI scalar maps like fractional anisotropy and mean diffusivity provide an increasingly valuable noninvasive tool to quantify structural plasticity (Keihaninejad et al., 2013; Krogsrud et al., 2016; Lam et al., 2014; Landi et al., 2011; Sagi et al., 2012; Scholz et al., 2009; Sexton et al., 2014). Yet, longitudinal changes in gray matter are usually overlooked in traditional pipelines of DTI analysis, which mostly focus on white matter tissue. Here we seek to optimize current spatial normalization approaches to detect longitudinal changes in diffusion scalar maps both at the level of the cortex and the white matter. To this end, we explored the impact of varying different features of the spatial normalization process on the across-session test-retest reproducibility error of *FA* and *MD* maps produced from multiple scanning sessions. We found that the most reliable approach, both for *MD* and *FA* in gray and white matter tissues, consisted on a spatial normalization pipeline using ANTs as the registration algorithm, MNI152 T1 template as the target image, *FA* as the moving image, and an individual *FA* template as an intermediate step.

The impact of the normalization parameters on the reproducibility of DTI registration has previously been explored by a few laboratories. For example, Liu and collaborators (2014) showed that traditional linear and non-linear algorithms such as *FLIRT* and *FNIRT* (FSL) or tensor-based algorithms were equally reliable. This was attributed to the high accuracy of DTI measures extracted from images acquired with more than 30 encoding directions. Around the same time, however, Schwarz and collaborators (2014) showed that the reproducibility of the spatial normalization process may in fact be improved further by using the non-linear algorithm from ANTs (*SyN*). Using a simulation approach, the authors showed that ANTs was more sensitive to detect true changes, yielding a lower rate of false positives than FSL. The ANTs registration algorithm was later also found to outperform FSL in terms of the alignment of white matter tracts, as assessed through quantification of fiber similarity (mean-square error) and *FA* profiles (Wang et al., 2017). Using an alternative approach to assess normalization reproducibility, in the present study we demonstrated that ANTs' registration algorithm yielded better reproducibility than FSL's. Thus, altogether, both Schwarz's and our study point to a better performance of ANTs for non-linear registration of DTI maps.

A unique contribution of our work is the application of ANTs to detect longitudinal changes of *FA* and *MD* in gray matter tissue. Estimating the reproducibility of these maps in both types of tissue was possible because we used a T1 stereotaxic template as the target image of normalization. Unlike DTI-based templates (e.g., Cabeen et al., 2017; Schwarz et al., 2014; Van Hecke et al., 2011; Zhang and Arfanakis, 2018), the MNI152 preserves the integrity of the cortex. To our knowledge, this is the first study to examine the direct registration of DTI maps to a T1 stereotaxic template. Interestingly, using the MNI152 T1 template not only improved reproducibility over the FMRIB58 in gray matter but also in white matter. Therefore,

using a T1-based standard as target image appears to be advantageous over DTI-based templates to improve detection of longitudinal changes in microstructure in both types of tissues.

Most studies aimed at detecting longitudinal changes in *FA* have used the *FA* map as the moving image to register to stereotaxic space. This is likely due to the fact that these maps provide better white matter contrast (although see Bach et al., 2014 and Park et al., 2003 for tensor-based registration). Yet, one would expect the choice of moving image to be measure dependent, with *MD* yielding better reproducibility to examine changes in gray matter, and *FA* in white matter. Given that not all DTI measures have good tissue contrast, we also examined the impact of including the *b0* on the reproducibility error. We found that using *FA* as moving image yielded the best reproducibility for both DTI measures in both tissues, whereas using *MD* deteriorated reproducibility in general, even when the *b0* was included to improve tissue contrast. Tustison and colleagues (2014) have pointed out that conducting the statistics on the same DTI measure that was used as moving image could bias the analysis, raising the rate of false positives. This does not hold, however, when the similarity metric used in the registration is based on mutual information, the quantity of choice in our study. Why using *FA* as moving image yields better reproducibility than *MD*? One possible explanation may lie on the anatomical correspondence between moving image and template. *FA* quantifies the diffusion of water molecules along axons, thereby providing relatively good resolution of white matter tracks. Thus, an accurate spatial matching between white matter in *FA* images and in the MNI152 T1 would show the best results. On the contrary, *MD* maps offer little tissue contrast, probably hampering good correspondence to the T1 template.

Another relevant issue to consider at the time of normalizing longitudinal data is the use of an intermediate step (either a group or an individual template) in the registration process. The use of a group template has been shown to reduce normalization bias (Tustison et al., 2014), increase accuracy of the DTI analysis and preserve the underlying diffusion information in voxel-based analysis (Van Hecke et al., 2008). This approach may be optimal when dealing with different population samples such as patients and healthy subjects, whose brains may differ critically from the standard. Yet, when dealing with a longitudinal study, obtaining a good alignment of multiple scans from each subject is a critical step. The use of an individual template has been implemented successfully in *TBSS* (FSL) to improve registration to the skeleton (Madhyastha et al., 2014). Here, we found that preregistration to an individual *FA* template reduced reproducibility error both in gray matter and white matter for *FA* as well as *MD*. This approach outperformed the direct normalization to the standard space. Using an individual template was also better than using a group template, constructed based on all individual templates, as the intermediate step. Our results suggest that, for longitudinal studies of a uniform population of subjects (e.g., healthy volunteers), where intra-

individual correspondence is crucial, warping images to a group template may hinder reproducibility. This may be due to the high inter-subject variability in some white matter tracts (Smith et al., 2006).

To examine if the transformations imposed by the normalization pipelines compromised the integrity of the DTI images, we assessed their impact on the signal-to-noise ratio. Even though we found some differences between the three normalization strategies, the SNR was overall preserved. Given that in longitudinal studies intra-individual reproducibility is critical, we suggest that at acceptable SNR values priority is given to the normalization strategy that minimizes reproducibility error.

One systematic bias we observed across all normalization approaches was the magnitude of whole-brain reproducibility errors, which was higher, almost two-fold, for *FA* than for *MD* regardless of the feature examined and type of tissue. These findings are in agreement with previous 1.5T and 3T studies, which showed that *FA* has a higher spatial variability of reproducibility errors across brain regions than *MD* (Farrell et al., 2007; Liu et al., 2014; Marenco et al., 2006). Moreover, both *FA* and *MD* were more reproducible in white matter than in gray matter. This tissue dependent difference in reproducibility errors has been reported previously for *FA* in white and deep gray matter (Farrell et al., 2007) and for both *FA* and *MD* in gray and white matter (Marenco et al., 2006). Cortical voxels can be affected by partial-volume effects in the gray matter/CSF interface, which could explain why areas of gray matter in the cortex show higher reproducibility errors (for example, see voxel-wise maps of RE in Figure 5).

Despite its strengths, our work presents limitations. First, the reproducibility error was assessed at the level of tissue. Although this approach facilitates extracting conclusions on the global impact of the contrasted features, it overlooks any spatial variations. A voxel-wise approach would allow estimating the influence of other factors such as anatomical variability, signal-to-noise ratio and image inhomogeneities. Second, the reproducibility error provides only one way to measure the reproducibility of a spatial normalization approach. Other aspects such as the specificity and sensitivity of the method in question would allow establishing the false positive rate, false negative rate and the minimum detectable difference of the DTI measure of interest. The specificity and sensitivity of different normalization approaches have been addressed for *FA* maps in white matter (Schwarz et al., 2014; Zhang and Arfanakis, 2018), and may be extended to other DTI maps. Lastly, and rather beyond this study, a more fundamental challenge is related to the interpretation of *FA* or *MD* changes in relation to the specificity of tissue microstructure properties and limitations of the diffusion tensor model (Jones et al., 2013; Jones and Cercignani, 2010; Zatorre et al., 2012). Alternative diffusion scalar metrics have been proposed, for example using higher-order information derived by fiber orientation distribution estimations (Raffelt et al., 2012; Riffert et al., 2014). The sensitivity

of those metrics for plasticity effects in gray and white matter remains to be seen, and spatial normalization optimization approaches for them may follow the strategy proposed in this study.

In summary, our work explored different normalization approaches with the aim of optimizing the detection of subtle changes in microstructure both in gray and white matter tissue. We showed that using ANTs' non-linear algorithm to warp *FA* images to MNI152 T1 through an individual template yielded the best reproducibility. This pipeline outperformed traditional algorithms currently used to assess microstructure in white matter tracts. Furthermore, it allowed exploring changes in gray matter, opening a window to quantify plasticity at the cortical level.

Acknowledgments

We want to thank Chiara Maffei for her advice in early phases of this work, Lisa Novello for her help testing the distributed analysis scripts and related documentation, and Arnaud Boré for technical assistance. We also would like to thank the University of Buenos Aires and, especially, Vice-rector Juan Pablo Mas Velez for their support in getting the first MRI facility dedicated to research going in Buenos Aires. This work was supported by a collaborative grant from the Quebec Bioimaging Network (QBIN, Canada), a grant from the Argentinian Ministry of Defense (PIDDEF), and a grant from the Argentinian Agency for the promotion of Science and Technology (FONCyT, ANPCyT).

References

Andersson, J.L.R., Jenkinson, M., Smith, S., 2007. Non-linear registration, aka spatial normalisation. FMRIB Technical Report TR07JA2., Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK.

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. Neuroimage 20, 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Andersson, J.L.R., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063–1078. https://doi.org/10.1016/j.neuroimage.2015.10.019

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration. Neuroimage 54, 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025.A

Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C., 2010. The optimal template effect in hippocampus studies of diseased populations. Neuroimage 49, 2457–2466. https://doi.org/10.1016/j.neuroimage.2009.09.062

Bach, M., Laun, F.B., Leemans, A., Tax, C.M.W., Biessels, G.J., Stieltjes, B., Maier-Hein, K.H., 2014. Methodological considerations on tract-based spatial statistics (TBSS). Neuroimage 100, 358–369. https://doi.org/10.1016/j.neuroimage.2014.06.021

Bastin, M.E., 1999. Correction of eddy current-induced artefacts in diffusion tensor imaging using iterative cross-correlation. Magn. Reson. Imaging 17, 1011–1024. https://doi.org/10.1016/S0730-725X(99)00026-0

Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system - A technical review. NMR Biomed. 15, 435–455. https://doi.org/10.1002/nbm.782

Benveniste, H., Hedlund, L.W., Johnson, G.A., 1992. Mechanism of detection of acute cerebral ischemia in rats by diffusion-weighted magnetic resonance microscopy, Stroke. https://doi.org/10.1161/01.STR.23.5.746

Blumenfeld-Katzir, T., Pasternak, O., Dagan, M., Assaf, Y., 2011. Diffusion MRI of structural brain plasticity induced by a learning and memory task. PLoS One 6. https://doi.org/10.1371/journal.pone.0020678

Bodini, B., Ciccarelli, O., 2014. Diffusion MRI in Neurological Disorders, in: Diffusion MRI: From Quantitative Measurement to In Vivo Neuroanatomy: Second Edition. Elsevier, pp. 241–255. https://doi.org/10.1016/B978-0-12-396460-1.00011-1

Bookstein, F.L., 2001. "Voxel-Based Morphometry" Should Not Be Used With Imperfectly Registered Images. Neuroimage 14, 1454–1462.

https://doi.org/10.1006/nimg.2001.0770

Cabeen, R.P., Bastin, M.E., Laidlaw, D.H., 2017. A Comparative evaluation of voxel-based spatial mapping in diffusion tensor imaging. Neuroimage 146, 100–112. https://doi.org/10.1016/j.neuroimage.2016.11.020

Davis, D., Ulatowski, J., Eleff, S., Izuta, M., Mori, S., Shungu, D., Zijl, P.C.M. van, 1994. Rapid Monitoring of Changes in Water Diffusion Coefficients during Reversible Ischemia in Cat and Rat Brain. Magn. Reson. Med. 31, 454–460. https://doi.org/10.1007/BF01947359

Farrell, J.A.D., Landman, B.A., Jones, C.K., Smith, S.A., Prince, J.L., Van Zijl, P.C.M., Mori, S., 2007. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. J. Magn. Reson. Imaging 26, 756–767. https://doi.org/10.1002/jmri.21053

Gauvain, K.M., McKinstry, R.C., Mukherjee, P., Perry, A., Neil, J.J., Kaufman, B.A., Hayashi, R.J., 2001. Evaluating pediatric brain tumor cellularity with diffusion-tensor imaging. Am. J. Roentgenol. 177, 449–454. https://doi.org/10.2214/ajr.177.2.1770449

Hughes, E.G., Orthmann-Murphy, J.L., Langseth, A.J., Bergles, D.E., 2018. Myelin remodeling through experience-dependent oligodendrogenesis in the adult somatosensory cortex. Nat. Neurosci. 21, 696–706. https://doi.org/10.1038/s41593-018-0121-5

Jones, D.K., 2004. The Effect of Gradient Sampling Schemes on Measures Derived from Diffusion Tensor MRI: A Monte Carlo Study. Magn. Reson. Med. 51, 807–815. https://doi.org/10.1002/mrm.20033

Jones, D.K., Cercignani, M., 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed. 23, 803–820. https://doi.org/10.1002/nbm.1543

Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn. Reson. Med. 42, 515–525. https://doi.org/10.1002/(SICI)1522-2594(199909)42:3<515::AID-MRM14>3.0.CO;2-Q

Jones, D.K., Knösche, T.R., Turner, R., 2013. White matter integrity , fiber count , and other fallacies: The do's and don'ts of diffusion MRI. Neuroimage 73, 239–254. https://doi.org/10.1016/j.neuroimage.2012.06.081

Keihaninejad, S., Zhang, H., Ryan, N.S., Malone, I.B., Modat, M., Cardoso, M.J., Cash, D.M., Fox, N.C., Ourselin, S., 2013. An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's disease. Neuroimage 72, 153–163. https://doi.org/10.1016/j.neuroimage.2013.01.044

Klein, A., Andersson, J.L.R., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M., Christensen, G., Collins, L., Hellier, P., Song, P., Hyun, J., Lepage, C., Pennec, X., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R. V, 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46, 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037.Evaluation

Klein, A., Ghosh, S.S., Avants, B., Yeo, B.T.T., Fischl, B., Ardekani, B.A., Gee, J.C., Mann, J.J., Parsey, R. V, 2010. Evaluation of volume-based and surface-based brain image registration methods. Neuroimage 51, 214–220. https://doi.org/10.1016/j.neuroimage.2010.01.091.Evaluation

Krogsrud, S.K., Fjell, A.M., Tamnes, C.K., Grydeland, H., Mork, L., Due-Tønnessen, P., Bjørnerud, A., Sampaio-Baptista, C., Andersson, J., Johansen-Berg, H., Walhovd, K.B., 2016. Changes in white matter microstructure in the developing brain-A longitudinal diffusion tensor imaging study of children from 4 to 11years of age. Neuroimage 124, 473–486. https://doi.org/10.1016/j.neuroimage.2015.09.017

Lam, B.Y.K., Halliday, G.M., Irish, M., Hodges, J.R., Piguet, O., 2014. Longitudinal white matter changes in frontotemporal dementia subtypes. Hum. Brain Mapp. 35, 3547–3557. https://doi.org/10.1002/hbm.22420

Landi, S.M., Baguear, F., Della-Maggiore, V., 2011. One week of motor adaptation induces structural changes in primary motor cortex that predict long-term memory one year later. J. Neurosci. 31, 11808–11813. https://doi.org/10.1523/JNEUROSCI.2253-11.2011

Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Methods 264, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001

Liu, X., Yang, Y., Sun, J., Yu, G., Xu, J., Niu, C., Tian, H., Lin, P., 2014. Reproducibility of diffusion tensor imaging in normal subjects: An evaluation of different gradient sampling schemes and registration algorithm. Neuroradiology 56, 497–510. https://doi.org/10.1007/s00234-014-1342-2

Madhyastha, T., Mérillat, S., Hirsiger, S., Bezzola, L., Liem, F., Grabowski, T., Jäncke, L., 2014. Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. Hum. Brain Mapp. 4555, 4544–4555. https://doi.org/10.1002/hbm.22493

Marenco, S., Rawlings, R., Rohde, G.K., Barnett, A.S., Honea, R.A., Pierpaoli, C., Weinberger, D.R., 2006. Regional distribution of measurement error in diffusion tensor imaging. Psychiatry Res. - Neuroimaging 147, 69–78. https://doi.org/10.1016/j.pscychresns.2006.01.008

Mintorovitch, J., Moseley, M.E., Chileuitt, L., Shimizu, H., Cohen, Y., Weinstein, P.R., 1991.

Comparison of diffusion- and T2-weighted MRI for the early detection of cerebral ischemia and reperfusion in rats. Magn. Reson. Med. 18, 39–50. https://doi.org/10.1002/mrm.1910180106

Papinutto, N.D., Maule, F., Jovicich, J., 2013. Reproducibility and biases in high field brain diffusion MRI: An evaluation of acquisition and analysis variables. Magn. Reson. Imaging 31, 827–839. https://doi.org/10.1016/j.mri.2013.03.004

Park, H.-J., Kubicki, M., Shenton, M.E., Guimond, A., McCarley, R.W., Maier, S.E., Kikinis, R., Jolesz, F.A., Westin, C.-F., 2003. Spatial normalization of diffusion tensor MRI using multiple channels. Neuroimage 20, 1995–2009. https://doi.org/10.1016/j.neuroimage.2003.08.008

Raffelt, D., Tournier, J., Rose, S., Ridgway, G.R., Henderson, R., Crozier, S., Salvado, O., Connelly, A., 2012. Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. Neuroimage 59, 3976–3994. https://doi.org/10.1016/j.neuroimage.2011.10.045

Reuter, M., Schmansky, N.J., Rossa, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 64, 1402–1418. https://doi.org/doi:10.1016/j.neuroimage.2012.02.084.

Riffert, T.W., Schreiber, J., Anwander, A., Knösche, T.R., 2014. Beyond fractional anisotropy : Extraction of bundle-specific structural metrics from crossing fiber models. Neuroimage 100, 176–191. https://doi.org/10.1016/j.neuroimage.2014.06.015

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., Assaf, Y., 2012. Learning in the Fast Lane: New Insights into Neuroplasticity. Neuron 73, 1195–1203. https://doi.org/10.1016/j.neuron.2012.01.025

Sampaio-Baptista, C., Khrapitchev, a a, Foxley, S., Schlagheck, T., Scholz, J., Jbabdi, S., DeLuca, G.C., Miller, K.L., Taylor, a, Thomas, N., Kleim, J., Sibson, N.R., Bannerman, D., Johansen-Berg, H., 2013. Motor Skill Learning Induces Changes in White Matter Microstructure and Myelination. J. Neurosci. 33, 19499–19503. https://doi.org/10.1523/JNEUROSCI.3048-13.2013

Scholz, J., Klein, M.C., Behrens, T.E., Johansen-Berg, H., 2009. Training induces changes in white-matter architecture. Nat Neurosci 12, 1370–1371. https://doi.org/10.1038/nn.2412

Schwarz, C.G., Reid, R.I., Gunter, J.L., Senjem, M.L., Przybelski, S.A., Zuk, S.M., Whitwell, J.L., Vemuri, P., Josephs, K.A., Kantarci, K., Thompson, P.M., Petersen, R.C., Jack, C.R., 2014. Improved DTI registration allows voxel-based analysis that outperforms Tract-Based Spatial Statistics. Neuroimage 94, 65–78. https://doi.org/10.1016/j.neuroimage.2014.03.026

Sexton, C.E., Walhovd, K.B., Storsve, A.B., Tamnes, C.K., Westlye, L.T., Johansen-Berg,

H., Fjell, A.M., 2014. Accelerated Changes in White Matter Microstructure during Aging: A Longitudinal Diffusion Tensor Imaging Study. J. Neurosci. 34, 15425–15436. https://doi.org/10.1523/JNEUROSCI.0203-14.2014

Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J., 2006. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487–1505. https://doi.org/10.1038/nprot.2007.45

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, 208–219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Smith, S.M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T.E., Klein, J.C., Robson, M.D., Jones, D.K., Behrens, T.E.J., 2007. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. Nat. Protoc. 2, 499–503. https://doi.org/10.1038/nprot.2007.45

Swire, M., ffrench-Constant, C., 2018. Seeing Is Believing: Myelin Dynamics in the Adult CNS. Neuron 98, 684–686. https://doi.org/10.1016/j.neuron.2018.05.005

Thomas, A.G., Marrett, S., Saad, Z.S., Ruff, D.A., Martin, A., Bandettini, P.A., 2009. Functional but not structural changes associated with learning: An exploration of longitudinal Voxel-Based Morphometry (VBM). Neuroimage 48, 117–125. https://doi.org/10.1016/j.neuroimage.2009.05.097

Tustison, N.J., Avants, B.B., Cook, P.A., Kim, J., Whyte, J., Gee, J.C., Stone, J.R., 2014. Logical circularity in voxel-based analysis: Normalization strategy may induce statistical bias. Hum. Brain Mapp. 35, 745–759. https://doi.org/10.1002/hbm.22211

Ugurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. Neuroimage 80, 80–104. https://doi.org/10.1016/j.neuroimage.2013.05.012

Van Hecke, W., Leemans, A., Sage, C.A., Emsell, L., Veraart, J., Sijbers, J., Sunaert, S., Parizel, P.M., 2011. The effect of template selection on diffusion tensor voxel-based analysis results. Neuroimage 55, 566–573. https://doi.org/10.1016/j.neuroimage.2010.12.005

Van Hecke, W., Sijbers, J., D'Agostino, E., Maes, F., De Backer, S., Vandervliet, E., Parizel,

P.M., Leemans, A., 2008. On the construction of an inter-subject diffusion tensor magnetic resonance atlas of the healthy human brain. Neuroimage 43, 69–80. https://doi.org/10.1016/j.neuroimage.2008.07.006

Wang, Y., Shen, Y., Liu, D., Li, G., Guo, Z., Fan, Y., Niu, Y., 2017. Evaluations of diffusion tensor image registration based on fiber tractography. Biomed. Eng. Online 16, 1–20. https://doi.org/10.1186/s12938-016-0299-2

Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D.A., Yacoub, E., Uğurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3T. Neuroimage 83, 991–1001. https://doi.org/10.1016/j.neuroimage.2013.07.055

Zatorre, R.J., Fields, R.D., Johansen-Berg, H., 2012. Plasticity in gray and white: neuroimaging changes in brain structure during learning. Nat. Neurosci. 15, 528–36. https://doi.org/10.1038/nn.3045

Zhang, S., Arfanakis, K., 2018. Evaluation of standardized and study-specific diffusion tensor imaging templates of the adult human brain: Template characteristics, spatial normalization accuracy, and detection of small inter-group FA differences. Neuroimage 172, 40–50. https://doi.org/10.1016/j.neuroimage.2018.01.046

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57. https://doi.org/10.1109/42.906424