Processing the diffusion-weighted magnetic resonance imaging of the PING dataset

Noor B Al-Sharif[a,*], Etienne St-Onge[b,*], Guillaume Theaud[b], Alan C Evans[a], Maxime Descoteaux[b]

[a]*McGill Centre for Integrative Neuroscience (MCIN), McGill University, Canada*
[b]*Sherbrooke Connectivity Imaging Laboratory (SCIL), Université de Sherbrooke, Canada*

Abstract

Diffusion-weighted magnetic resonance imaging (dMRI) allows for the in-vivo assessment of anatomical white matter in the brain, thus allowing the depiction of structural connectivity. Using structural processing techniques and related methods, a growing body of literature has illustrated that connectomics is a crucial aspect to assessing the brain in health and disease. The Pediatric Imaging Neurocognition and Genetics (PING) dataset was collected and released openly to contribute to the assessment of typical brain development in a pediatric sample. This current work details the processing of diffusion-weighted images from the PING dataset, including rigorous quality assessment and fine-tuning of parameters at every step, to increase the accessibility of these data for connectomic analysis. This processing provides state-of-the-art diffusion measures, both classical diffusion tensor imaging (DTI) and more advanced HARDI-based metrics, enabling the evaluation not only of structural white matter but also of integrated multimodal analyses, i.e. combining structural information from dMRI with functional or gray matter analyses.

Keywords: diffusion MRI, connectome, tractography, white matter, pediatric development

1. Introduction

Brain structure & development

Post-natal neural development is characterized by widespread structural changes of the brain, in both gray and white matter regions [Sowell et al., 2004; Lenroot and Giedd, 2006; Giedd and Rapoport, 2010]. Microstructural changes in brain architecture, such as cortical thickness or myelination, lead to global morphometric changes detectable by magnetic resonance imaging (MRI) techniques, including anatomical images (such as T1 or T2* weighted) and diffusion-weighted imaging (dMRI). Just as these techniques can be used to

---

*These authors contributed equally: shared first and corresponding authors

10 assess differences between health and disease states, they also provide insight into underlying microstructural changes in the brain that characterize typical development *in vivo* [Lim and Helpern, 2002; Prayer and Prayer, 2003; Qiu et al., 2015]. Implementing multimodal imaging methods in the assessment of structural development throughout the lifespan enables a comprehensive per-

15 spective of anatomical growth and variance across the whole brain. Research has shown that a major component of post-natal brain development occurs in the strengthening or refining of local and whole-brain white matter connections [Lebel et al., 2008; Jernigan et al., 2011]. Many of these changes are experience-dependent and causally linked with myelination, synaptogenesis and synaptic

20 pruning in different areas of the brain at different developmental periods [Shatz, 1990; Benes et al., 1994; Paus et al., 1999; Stiles and Jernigan, 2010; Deoni et al., 2011]. Such changes in structural white matter provide a compelling case for the use of dMRI in assessing brain development.

Diffusion MRI & white matter processing

25 DMRI is a particular MRI sequence that quantifies water displacement in multiple directions. From diffusion-weighted images, local diffusion models are reconstructed to estimate the underlying characteristics of white matter anatomy [Jones, 2008; Descoteaux, 2015]. The dynamic changes of this microstructure in development can be characterized both in structural and func-

30 tional analyses [Klingberg et al., 1999; Mukherjee et al., 2002; Olesen et al., 2003; Sotiropoulos and Zalesky, 2019]. These changes, whether localized or not, underlie global changes in whole-brain networks, which form the basis of brain function and behaviour [Nagy et al., 2004]. As such, it is important to evaluate these changes in the context of brain connectivity, using techniques such

35 as dMRI to depict structural white matter changes [Jeurissen et al., 2019]. As of yet, we have a limited understanding of how white matter develops during childhood and adolescence and therefore require more insight into how these structures change over time.

Though the implementation of dMRI techniques has improved our under-

40 standing of anatomical white matter structure, processing of such data is known to be rife with complex challenges [Jones and Cercignani, 2010]. These challenges consist of low spatial resolution, noisy acquisition, MRI gradient-related artifacts and fiber organisation ambiguity (crossing, kissing, fanning) [Jbabdi et al., 2015; Dell'Acqua and Tournier, 2019; Jeurissen et al., 2019]. As such,

45 processing dMRI data needs to be reliable and reproducible to improve dMRI quality and better depict local and global structure of the human brain [Theaud et al., 2020].

PING

In order to depict stereotypical changes or patterns in structural develop-

50 ment, the use of large-scale population-representative datasets is essential. One such dataset is the Pediatric Imaging Neurocognition and Genetics (PING) study: a large-scale developmental cohort of subjects aged between 3 and 22 years, including imaging, cognitive and genetic data. This comprehensive study

provides a rich multimodal imaging perspective of development, from childhood to early adulthood, and enables the integration of behavioural and genetic information. PING publicly provides both the raw imaging and demographic data, as well as some processed derivatives of these data including cortical thickness, surface area, volumes and some diffusion-tensor measures. However, despite this notable amount of data made available by PING, structural white matter connectivity measures were not included, to our knowledge.

To best take advantage of these large-scale datasets, such data must be made accessible for use beyond their originating research groups. These types of open science practices promote scientific collaboration and discovery, such as the Human Connectome Project (HCP) [Glasser et al., 2013]. The creation and analysis of large-scale population datasets are arduous tasks best accomplished as collaborative efforts. Sharing data encourages better, more transparent research practices and also minimizes the necessity for repetition in laborious data collection and processing. Therefore, the derivatives of this current processing will be made publicly available and accessible through the National Database for Autism Research (NDAR) repository, a database of the National Institute of Mental Health Data Archive (NDA), which also hosts the original PING dataset.

2. Methods

In this section, we specify how the initial PING release and dMRI acquisition were harmonized and formatted [Jernigan et al., 2016]. The pre-processing and processing applied to dMRIs is detailed, along with any processing applied in parallel to the T1-weighted anatomical images. Finally, tractography and connectivity matrix generation are described. As detailed in Jernigan et al. [2016], written consent was obtained from parental guardians for participants below 18 years of age and from the participants themselves otherwise, in accordance with IRB protocols. For reproduction purposes, details are listed in the Usage Notes and Code availability sections. Figure 1 illustrates a comprehensive overview of the processing pipeline.

2.1. Input data collection and formatting

Imaging data collection

Original PING data were collected in Digital Imaging and Communications in Medicine (DICOM) format across nine sites on four 3T MR machine types: Siemens TrioTim, GE Discovery, GE Signa, and Philips Achieva. Scanning protocols were harmonized across all scanners and sites to minimize variance across scans. From PING, we gained access to a total of 954 subjects with dMRI scans, 634 of which also have at least 2 or more repeated scans within the same session. In total, 1669 unique acquisitions were used as input for the subsequent processing.

Anatomical images

Though the PING dataset contains both T1w and T2⋆w images, only T1w images are required for the processing described here. T1w images were collected using a real-time prospective motion correction (3D PROMO) technique to minimize the effects of motion, which is typically widespread in pediatric samples. Subjects were processed only if they had both dMRI and T1w images which passed quality assurance, i.e. minimal motion, no major artifacts, etc.
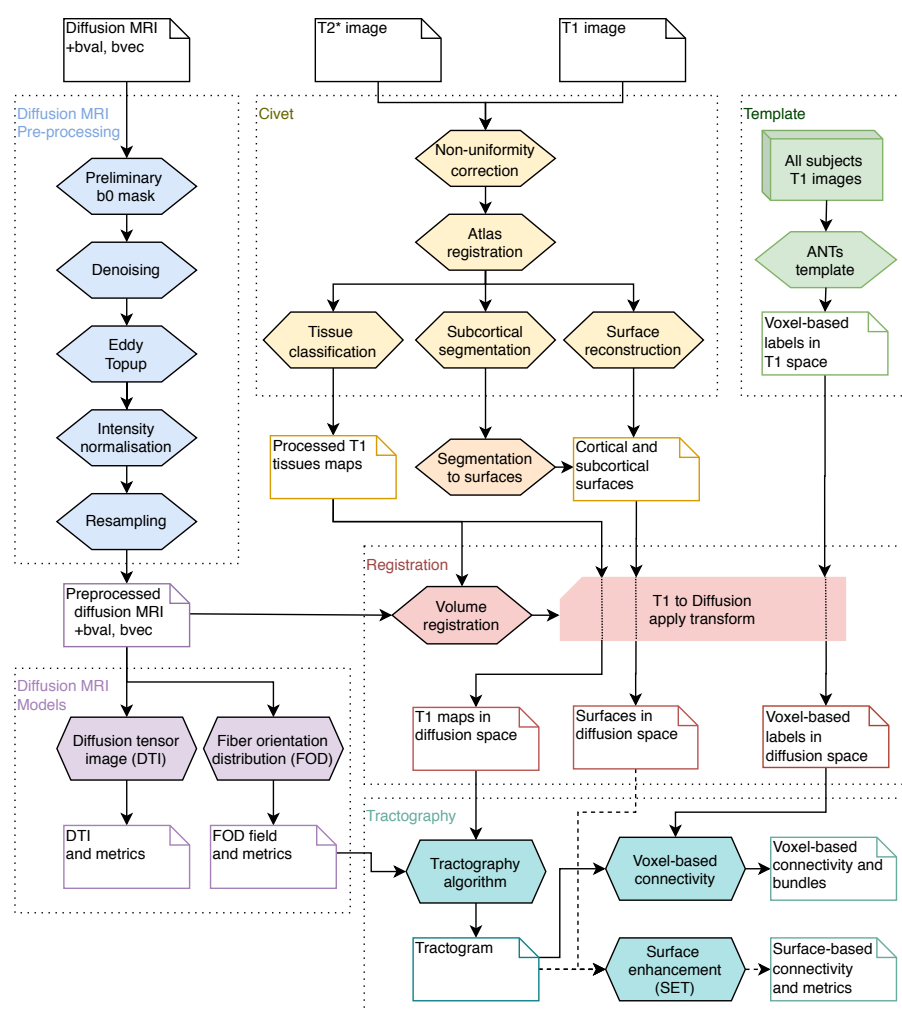

Figure 1: Processing pipeline architecture, for diffusion, T1 and T2⋆ weighted images: dMRI pre-processing (blue), CIVET for WM-GM-CSF tissue maps with surfaces (yellow), T1 template reconstruction (green), Registration to align T1 maps from anatomical to diffusion space (red), diffusion local models reconstruction (purple), and tractography with structural connectivity (cyan).

4

100 *Data conversion: DICOM to NIfTI*

Since most dMRI processing tools cannot directly input DICOM files, a conversion is necessary. First, DICOM ('.dcm') files were converted to Neuroimaging Informatics Technology Initiative (NIfTI) file format ('.nii') using *dcm2niix* [Li et al., 2016]. For dMRI volumes, this conversion also generates two accompa-

105 nying 'bval' and 'bvec' files which contain acquisition details of the b-values and b-vectors for each volume.

Transformation (space) convention

To facilitate the use of common dMRI software, all NIfTI files were modified to the RAS (left-to-right, posterior-to-anterior, inferior-to-superior) convention.

110 *Diffusion-weighted images (dMRI)*

Each dMRI acquisition is composed of a total of 32 diffusion-weighted three-dimensional volumes: 30 with diffusion directions and 2 baseline (b0 images). Of the two b0 images, one is acquired in the anterior-to-posterior (A-P) direction, which is in alignment with all volumes of the main dMRI (named 'b0.nii').

115 The second b0 is acquired in the reverse direction, posterior-to-anterior (P-A), (named 'b0_rev.nii'). This reverse b0 is used to correct for susceptibility-induced distortion in dMRI. Each direction measures the signal loss along the chosen orientation (b-vector) with a single b-value of 1000.

2.2. Diffusion MRI Pre-processing

120 A preliminary b0 mask ('bet_prelim_b0.nii') is estimated with *FSL-bet* from the 'b0.nii' dilated with *MRtrix-mask filter*. This is done to mask voxels determined to be outside of the brain tissue and to remove unnecessary computation of invalid voxels. Denoising was performed for all dMRI directions ('dwi.nii') inside the preliminary mask ('bet_prelim_b0.nii') with *MRtrix-dMRI denoise*[Tournier

125 et al., 2019].Topup correction map was computed with *FSL-topup*, to correct for common dMRI acquisition deformation, using both oppositely-acquired b0 images ('b0.nii', 'b0_rev.nii') and applied to all dMRI directions ('dwi.nii'). Eddy current correction was also applied using *FSL-eddy*. *FSL-eddy* applies both eddy current and distortion correction (generated from *FSL-topup*) to all diffu-

130 sion volumes; it also returns distortion-corrected 'bval' and 'bvec' files. A new brain mask ('bet_b0.nii') is computed after the deformation correction, minimally dilated with *MRtrix-mask filter*. Intensity normalization was employed to establish spatial uniformity of the signal intensity across the dMRI volume using *ANTs-N4BiasFieldCorrection*. Additionally, *MRtrix-dMRI normalise* was used

135 to harmonize the diffusion signal across subjects [Tournier et al., 2019]. Each diffusion volume was resampled with linear interpolation to an isotropic resolution of $1mm$ [Lehmann et al., 1999; Dyrby et al., 2014].

2.3. Diffusion MRI models

DTI. Dipy-TensorModel was used to estimate the Diffusion Tensor Image

140 (DTI) [Basser et al., 1994] at every spatial position (voxel), with a weighted

5

least squares method. Subsequently, multiple DTI measures were computed from reconstructed tensors (Figure 2), including the Fractional Anisotropy (FA), Mean Diffusivity (MD), Apparent Diffusion Coefficient (ADC), Radial Diffusivity (RD), and red-green-blue (RGB) map colored in the x-y-z orientation (Figure 3).

FOD. Dipy-csdeconv package was used to compute fiber orientation distribution (FOD). Initially the "single fiber distribution" is estimated with the `auto_response` function and afterward the full distribution is computed at each position using `ConstrainedSphericalDeconvModel` with the `symmetric724` discretized sphere. The FOD represents the estimated orientation distribution of fibrous structure at each voxel [Tournier et al., 2007; Descoteaux et al., 2009]. Peaks, representing main diffusion directions, are extracted from local maxima of each FOD angular distribution, similar to a vector field (Figure 2). This FOD field is later used to compute the tractography and estimate the structural connectivity. In addition, some feature maps are estimated from the FOD reconstruction, illustrated in Figure 3, such as total apparent fiber density (AFD).

2.4. T1w image processing

Segmentation maps for white matter, gray matter, cerebrospinal fluid (CSF) and subcortical structures (SC) were generated by *CIVET 2.1*, from the native T1 anatomical image. A brain mask, also estimated from this segmentation, was applied to all T1w images before registration [Zijdenbos et al., 1998; Tohka et al., 2004]. Cortical surfaces were reconstructed with *CIVET 2.1* [Kim et al., 2005; Lyttelton et al., 2007], similarly to *FreeSurfer*. Subcortical regions were estimated with *ANIMAL* [Collins et al., 1999] and converted into meshes using a marching cube algorithm, and then combined with cortical surfaces in a single file. The anatomical T1 and diffusion volume were co-registered, transforming the T1 volume into native diffusion space. The resulting transformations (affine matrix and warp) from *antsRegistration*, are also used to align subsequent anatomical maps and cortical surfaces to diffusion space [Avants et al., 2011; Theaud et al., 2020].

Template construction

In parallel to the included T1w processing within *Tractoflow*, a PING specific T1w template was generated using *antsMultivariateTemplateConstruction2.sh* [Avants et al., 2011], which automates an iterative registration process to move any number of T1w volumes into a common space. Once the template has been generated, the resulting transformation files are used to register other template-aligned volumes into individual subject space. This is applied in subsequent processes which require aligning atlas segmentation labels into native T1w and diffusion volume space, as described below in creating voxel-based connectivity matrices. Averaged tissue maps in the template space can be oberved in Figure 4.
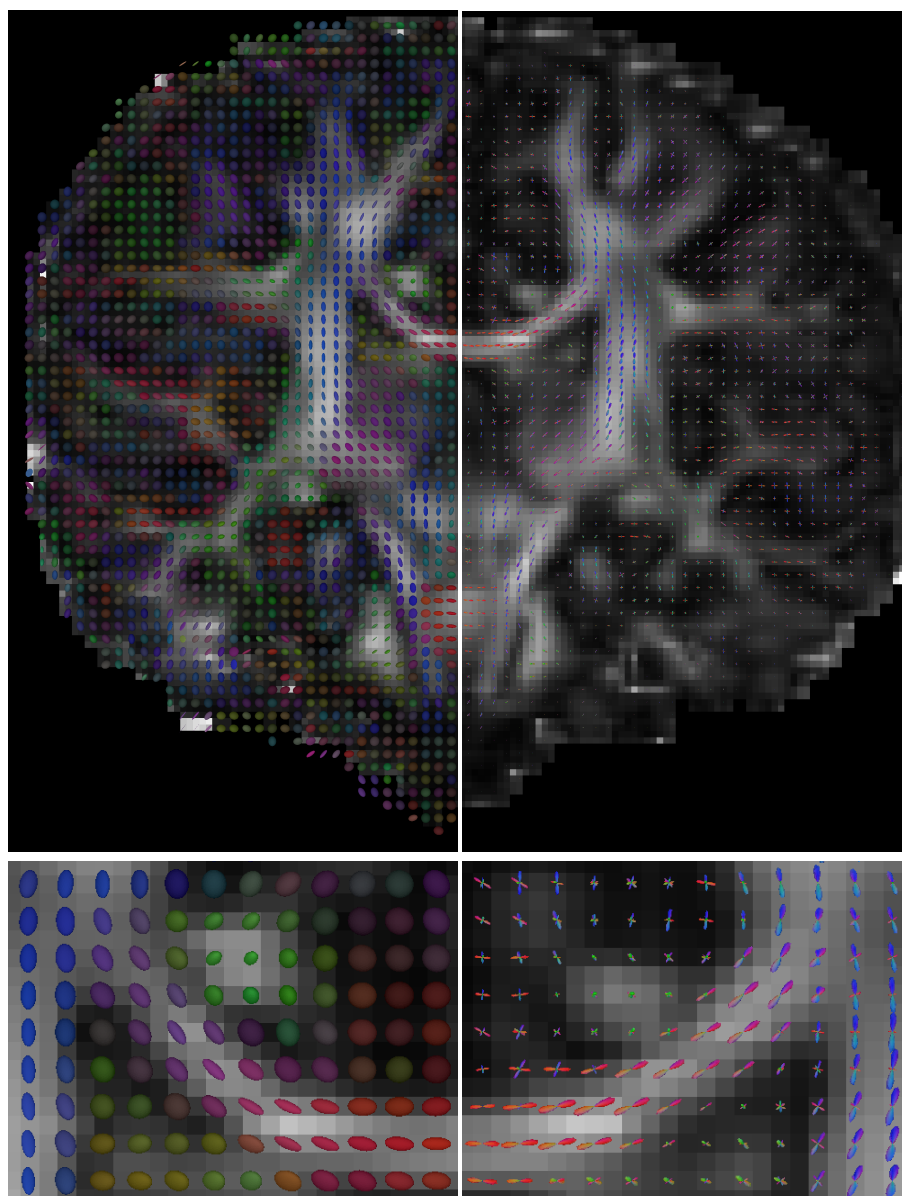
Figure 2: Coronal view of the reconstructed diffusion tensor image (DTI) on the left, and fiber orientation distribution (FOD) on the right. On the bottom row, a zoomed view of the *Corpus Callosum* and *Cingulum* regions.
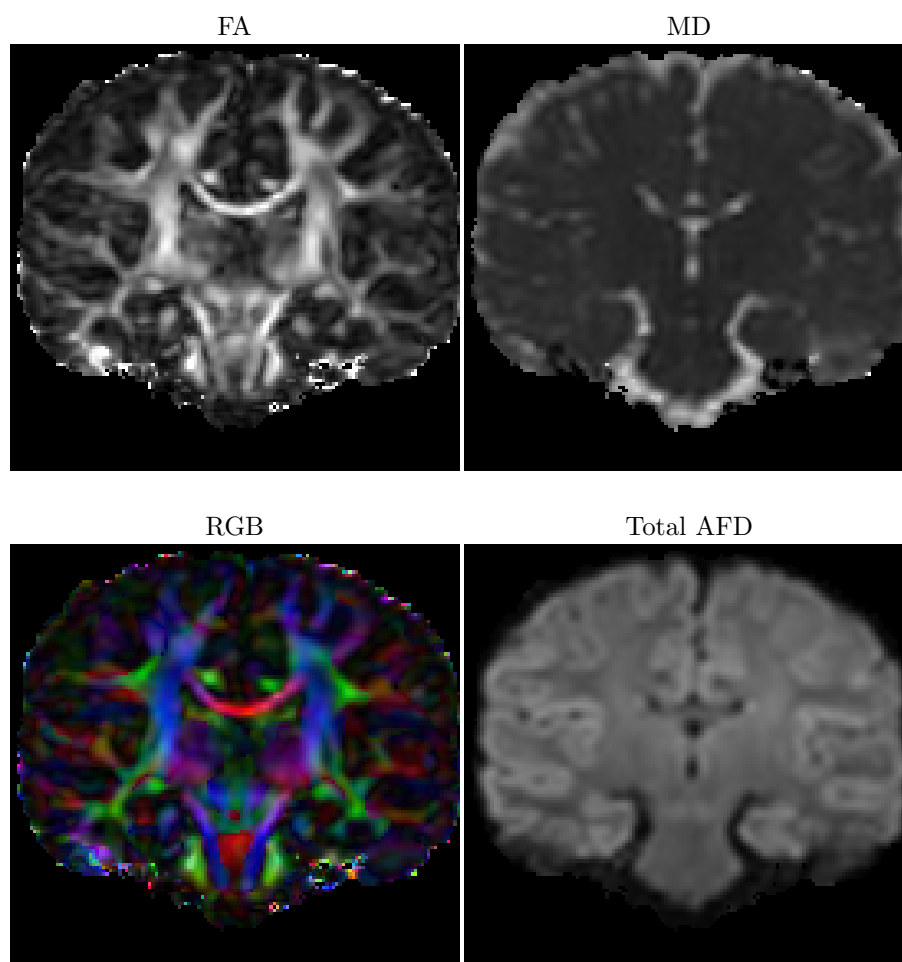
Figure 3: Diffusion tensor image (DTI) field measures: fractional anisotropy (FA), mean diffusivity (MD), and red-green-blue (RGB) map colored from the x-y-z orientation. Total apparent fiber density (AFD) estimated from the Fiber orientation distribution (FOD).

2.5. Structural connectivity analysis

Tractography

Structural pathways can be reconstructed from dMRI by following the local orientation of estimated local models with a process called tractography. In this work, we used the "particle filtering tractography" (PFT) [Girard et al., 2014], implemented in *Dipy*, taking advantage of the sharp orientation from the FOD field. Resulting streamlines represent estimated pathways of the brain macrostructure and connectivity. Similarly to anatomically-constrained tractography (ACT) [Smith et al., 2012], PFT takes advantage of previously computed WM-GM-CSF tissue maps, to define legal and illegal regions.

8

The "surface-enhanced tractography" (SET) [St-Onge et al., 2018], an adaptation of the PFT algorithm, was also used to initialize and terminate streamlines from cortical and subcortical surfaces at the WM-GM boundary. Since
195 CIVET 2.1 produces a standardized mesh, with a fixed number of vertices and triangles for every subject, SET allows for better comparison and individually variable correlation of connectivity features.

Connectivity matrices

Connectivity matrices were generated using two separate methods: voxel-
200 and surface-based. Voxel-based matrices were created using a 3D anatomical Desikan-Killiany-Tourville (DKT) [Desikan et al., 2006; Klein and Tourville, 2012] segmentation label in each acquisition's native diffusion space. To register the DKT segmentation label to native diffusion space, a number of image deformation steps were executed. The DKT segmentation label was moved
205 from MNI space to the PING template space and then inversely warped to
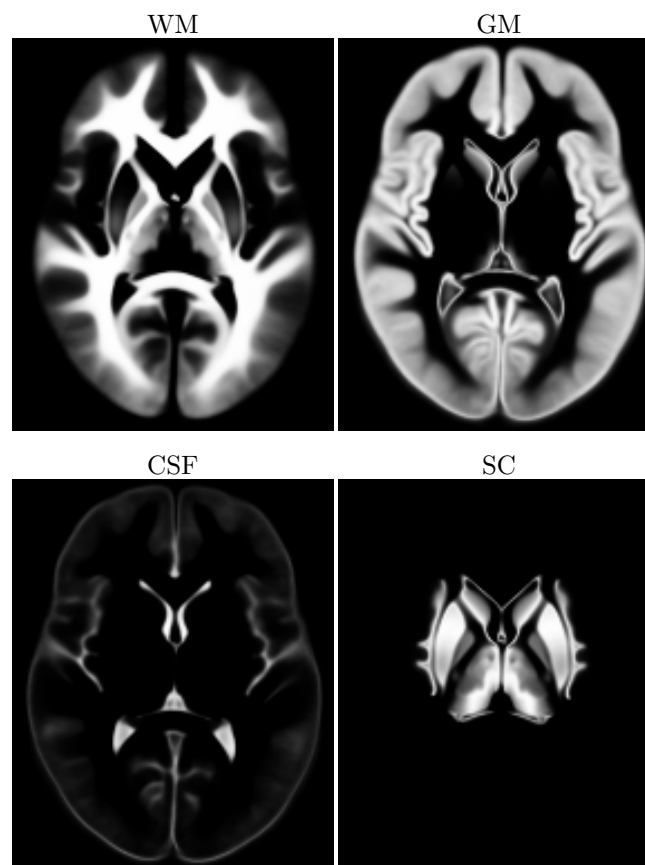

Figure 4: Averaged tissue maps in the template space, axial view.

subject-specific T1w native space. Using the transformation affine matrix and warp image from *Tractoflow* processing, the DKT label was then aligned from T1w to native diffusion space. This native-space label was generated for each diffusion acquisition and used as input with the corresponding tractogram in

210 *SCIL-scil_compute_robust_connectivity_matrix.py*, which outputs a connectivity matrix as a *npy* file. Depicted in Figure 5, connectivity is represented as the number of streamlines with endpoints within a region of interest (ROI) label as delineated by the DKT label.

Using the surface-based method, every streamline generated by SET starts
215 and terminates within a triangle that makes up the cortical surface mesh. Each triangular facet corresponds to a specific cortical label in the CIVET DKT atlas. Subsequently, the connectivity matrix is computed by counting the number of streamlines connecting two different cortical regions.
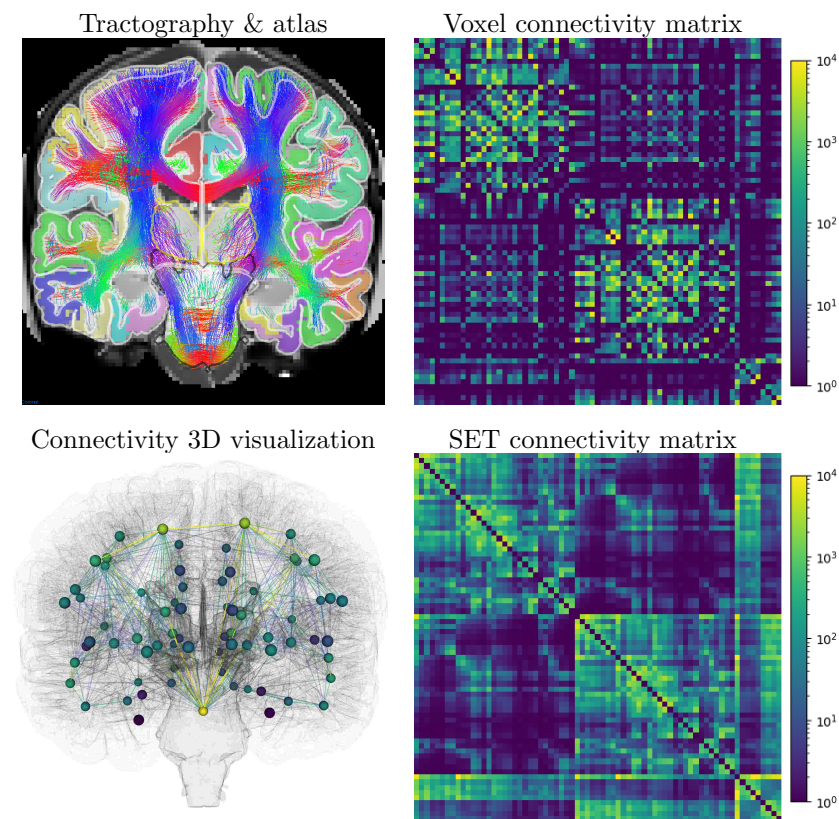

Figure 5: Tractography reconstruction visualized over the DKT atlas labels and resulting connectivity matrices, streamlines count per million in log scale. The connectivity matrix contains 35 cortical and subcortical labels per hemisphere (left, right) and the brainstem, for a total of 71×71 potential connections.

3. Technical Validation

3.1. Quality Control

Data formatting

Once converted to NIfTI and enforced to RAS orientation, each image header was verified automatically with *MRtrix-mrinfo*, to correct stride convention and data formatting.

Anatomical images

If a subject had multiple anatomical T1 acquisitions, a single best-quality image was chosen for each subject, validated visually. During initial assessment of the raw T1 acquisitions, subjects were excluded for severe motion, slicing artifacts, or missing data.

CIVET

Brain masks, tissue maps and surfaces from CIVET 2.1 were inspected with the given alignment QC image. Subjects were excluded from further analysis post-CIVET processing due to a) failing to complete processing or b) motion artifacts and incorrectly assigned data.

Eddy & Topup

Images were visually assessed before and after Eddy & Topup correction to ensure the correction was applied properly. This also validated that the b0 preliminary mask was not too restrictive.

Pre-processed dMRI and T1 registration

Registered pre-processed T1s (masked, denoised, normalized, resampled) were assessed with the fully pre-processed dMRI (prelim masked, denoised, eddy- plus topup-corrected, normalized, resampled) to ensure that the images were aligned and properly pre-processed. In addition, anatomical tissue segmentation maps (WM-GM-CSF) were visually inspected in the diffusion space prior to tractography.

Template alignment

Anatomical T1 images and tissue map alignment were verified both qualitatively and quantitatively with an alignment score. Quantitative scores are shown in the next sub-section.

DTI & FOD

Diffusion orientations acquired from the b-vector files and their respective eddy- and topup-corrected versions were validated by visualizing DTI main orientation in well-known brain structures: the corpus callosum and the cingulum. A flip in the x axis for b-vector values was required as the images were acquired in LAS orientation, but *Dipy* is based on RAS reference. Resulting local reconstructions were also qualitatively verified for shape, size and orientation discrepancies (Figure 2). This ensured that computed DTI and FOD values were correct and well-aligned with the white matter structure.

Diffusion measures

Diffusion measures were also qualitatively validated using maps for FA, MD, RD, RGB and AFD. The variability of these dMRI metrics for this dataset is presented in the next sub-section. Subjects were removed if their diffusion maps presented visible artifacts such as motion, poor registration, uncorrected deformations or signal loss.

Tractograms

The resulting tractograms, from 100 randomly chosen subjects, were inspected with 3D visualization, to ensure no reconstruction artifacts were present.

3.2. Dataset Variability

In addition to standard quality control and visual validation, quantitative measures and variability were computed after each of the following processing steps. This was done to ensure that no major outliers remained in the final dataset. These variability results could also serve as references for future brain development studies or comparison with the PING dataset.

Template alignment

After the template reconstruction (Figure 4), an alignment score (dice) to this averaged template was computed for each individual probabilistic tissue map (Figure 6a). This template alignment score (mean; standard deviation) for WM (0.897; 0.010), GM (0.877; 0.015) and SC (0.930; 0.019) is excellent, with small variance. The CSF score (0.574; 0.039) is lower, compared to others, mostly due to brain mask variation. Subjects misaligned to the template were easily detected as outliers from the WM dice score (z-score < -5) and confirmed through visual validation. In general, a data point is considered an outlier when the z-score is greater than 3 or less than -3. Removing these non-aligned subjects consistently reduced the dice standard deviation by a factor of 3 for WM (around 2 for other tissue maps). Almost identical results and corresponding outliers were found using other distance measures (L1 and L2) as well.
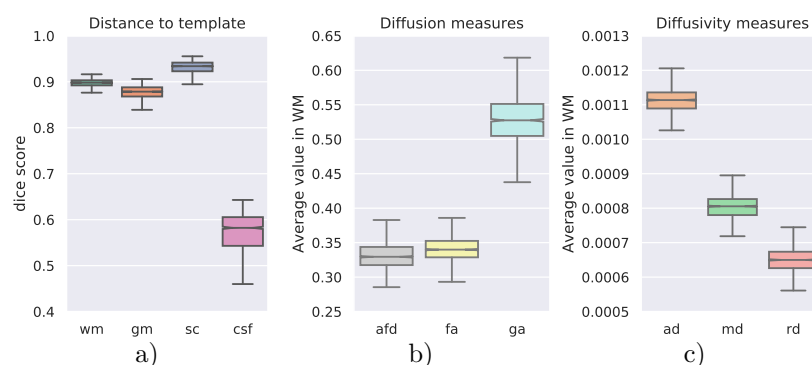

Figure 6: a) Dice alignment score to the template, for each subject. White matter measures of: b) diffusion (AFD, FA, GA) and c) diffusivity in mm^2/s (AD, MD, RD).

Diffusion measures

Measures estimated from dMRI models (DTI and FOD) were also evaluated for each subject. Intensities in each tissue structure were analysed with a histogram and average value per subject (see Figure 6, Table A.1). From this variability analysis, results suggest that some FOD metrics, such as NUFO, are too variable to measure any group differences in this dataset. Nonetheless, this limitation could be partially caused by the variable nature of this dataset (age variation, multiple acquisition site, etc).

FOD model data response (frf)

Fiber response function (frf) was first estimated for each dMRI acquisition in the center of the WM structure in high FA values near the corpus callosum. Results were then compared for each scanner and acquisition site (Figure 7a, Table A.2). Since all frf were similar across PING, values used to compute FOD were fixed to $18, 4, 4$ (ratio of $18/4$), the dataset average.

Connectivity matrices

Connectivity matrices were employed to measure inter-subject distance. For subjects with multiple dMRI acquisitions, intra-subject distance was also measured. In addition, variability analysis was performed over connectivity matrices to estimate outliers and the differences between acquisition sites and age groups. Figure 7b displays L1-distance, from the average connectivity to each subject's connectivity matrix, grouped per acquisition site (Table A.3).

Metrics

Dice score and distance equations, in between images A and B at each voxel position v in the whole 3D volume $\mathcal{V}$ is as follows:

$$\text{Dice score}(A, B) := \frac{\sum_v^{\mathcal{V}} A_v B_v}{\sum_v^{\mathcal{V}} A_v + \sum_v^{\mathcal{V}} B_v}$$
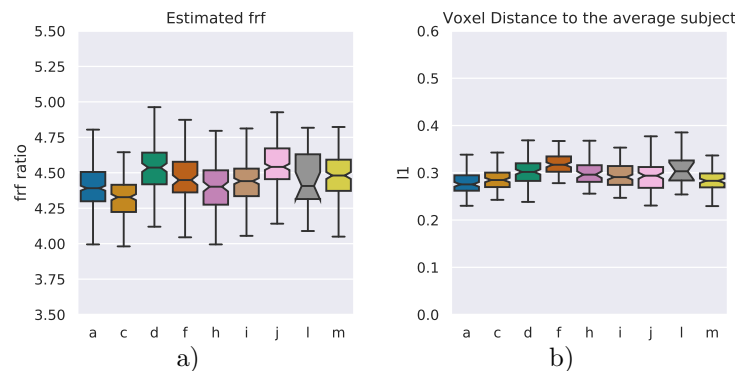

Figure 7: a) Estimated fiber response function (frf) ratio, b) connectivity matrices distance to the average subject, across acquisition sites. MRI scanner for each site: Philips Achieva (a,h), GE Signa HDx (c), GE Discovery MR750 (i) and Siemens TrioTim (d,f,j,l,m).

4. Conclusion

In this work, we created a processed and quality-assessed set of connectivity derivatives of the PING dataset, using a robust dMRI processing pipeline, *Tractoflow* [Theaud et al., 2020]. Resulting derivatives were assessed and validated at multiple steps to ensure a high-quality, normative dataset. Using a state-of-the-art pipeline and openly publishing the dataset will encourage the use of dMRI metrics within the scope of PING-related connectomic analyses. The resulting dataset includes:

- processed T1 and diffusion weighted images, from established software,
- brain tissue segmentation maps along with cortical surfaces from CIVET,
- advanced dMRI processing using fiber orientation distribution (FOD) in addition to diffusion tensor imaging (DTI),
- WM pathways reconstruction from cutting edge tractography algorithms,
- structural connectivity matrices, estimated using voxel and surface approach.

Code availability

An adapted version of *Tractoflow* for PING processing is available in the complementary files on NDA, named *Tractoflow-pve*. Within this project folder, both 'USAGE' and 'README.md' give information about usage, 'nextflow.config' contains configuration details, and 'main.nf' includes all processing steps and details. The singularity image 'tractoflow_2.0.0.img' contains all requirements to execute the 'tractoflow-pve/' pipeline. A secondary sub-pipeline 'set-nf/' is provided to compute the "surface-enhanced tractography" and surface-based measures, using the same file organisation as in 'tractoflow-pve' with the addition of 'set_ping.img' to execute it. The 'set_ping.img' is also used to execute the 'civet-nf' pipeline which organizes PVE output files from CIVET 2.1 for 'tractoflow-pve'. All code is packaged within the 'Processing_PING' archive file available on NDA.

All dMRI and T1w processing software employed in this work are available online and free of use for researchers: *ANTs*, *CIVET*, *dcm2niix*, *FreeSurfer*, *FSL*, *MRtrix*, *Nibabel*, and *Dipy*. Processing tools utilized for scripts and pipelining: *Nextflow*, *Singularity*, and *Python*. Visualisation tools: *Fury*, *MI-Brain*, *MisterI*, and *MRtrix mrview*.

Data Records

Data and research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. Dataset identifier(s): 10.15154/1519178. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDA. Information on the

original PING dataset, including acquisition protocols, is detailed in Jernigan
et al. [2016].

The processed derivatives described here and the code used to produce
them are all available on the NDA website, study ID #932 titled *Diffusion-weighted derivatives of PING*, doi: 10.15154/1519178. All code and file descriptions are found at the study's page under 'Data Analysis,' and resulting derivative data are packaged in the 'Results' section. Descriptions for all derivative files and in which data package they are organized can be found in *ping_dwiderivatives_ndarinfo.xlxs* under 'Data Analysis.'

Processed anatomical maps

All anatomical 3D volumes and derived maps are saved in *NIfTI* format,
supported in *dcm2niix*, *ANTs*, *Nibabel*, and *MRtrix*.

Processed dMRI

Images at each processing step are saved in 3D or 4D (representing a series of 3D images) *NIfTI*: b0 volumes (A-P, P-A), dMRI directions, as well as upsampled denoised-, eddy- & topup-corrected images.

Diffusion bval & bvec

Eddy-corrected b-values and b-vectors are saved in the same format as initially converted by *dcm2niix*: b-values are listed in a text file (space-separated), b-vector directions are listed in x-y-z format in distinct rows.

Cortical surfaces

The Visualization Toolkit (VTK) format was used for all meshes, with vertices saved in world coordinates (LPS). In this format, cortical surfaces can be visualized and overlaid onto the T1 image with *MI-Brain* (see Figure 5).

DTI

Tensor images were saved with *Dipy* and *Nibabel* in NIfTI format. The resulting image is a 4D volume: spatially represented in 3D with a fourth dimension containing all 6 symmetric tensor coefficients. Tensor coefficients were saved in this given order along the fourth dimension axis, where: $[D_{00}, D_{01}, D_{02}, D_{11}, D_{12}, D_{22}]$, where

$$D = \begin{bmatrix} D_{00} & D_{01} & D_{02} \\ D_{01} & D_{11} & D_{12} \\ D_{02} & D_{12} & D_{22} \end{bmatrix}.$$

FOD

FOD fields were saved with *Dipy* using the default spherical harmonics (SH) representation [Descoteaux et al., 2009]. Only even-ordered SH coefficients (c_l^m) were saved, as computed FOD are symmetric. The SH file is ordered by SH order (l) first and phase (m) second: $[c_0^0; c_2^{-2}, c_2^{-1}, c_2^0, c_2^1, c_2^2; c_4^{-4}, c_4^{-3}, c_4^{-2}, c_4^{-1}, c_4^0, c_4^1, c_4^2, c_4^3, c_4^4; c_6^{-6}, c_6^{-5}, c_6^{-4}, c_6^{-3}, c_6^{-2}, c_6^{-1}, c_6^0, c_6^1, c_6^2, c_6^3, c_6^4, c_6^5, c_6^6]$.

380 *Tractograms*

Tractography reconstructions were saved in *TrackVis* format ('.trk'). This format is well supported in *Nibabel*, *Dipy* and *MI-Brain*.

Connectivity matrices

All resulting connectivity matrices were stored as nummpy array matrices
385 ('.npy') format. The ordered list of labels is given in the 'labels_co_matrix.txt' file.

Usage Notes

Command lines

Nextflow and *Singularity* need to be installed; a complete example is given in 'PING_example.sh'.

```
390 nextflow  run civet-nf/main.nf  --civet  path/to/civet˙output/ "
         -with-singularity set˙ping.img

     sh civet-nf/tree˙for˙tractoflow.sh  -f  path/to/data/  -c  civet-nf/results/ "
         -o  tractoflow˙input/
395
     nextflow  run  tractoflow-pve/main.nf   --root tractoflow˙input/ "
         --dti˙shells "0 1000"   --fodf˙shells "0 1000" "
         -profile  civet˙pve    -with-singularity  tractoflow˙2.0.0.img

400 nextflow  run  set-nf/main.nf   --tractoflow  tractoflow˙nf/results/ "
         --civet  path/to/civet˙output/   -profile  civet˙DKT "
         -with-singularity  set˙ping.img
```

Pediatric Imaging Neurocognition and Genetics (PING)

Acknowledgements

16

Author contributions statement

The concept of this work, processing the structural connectivity of the PING dataset, was proposed and realized by Noor B Al-Sharif, supervised by Alan C Evans. Diffusion MRI processing planning and work outline was done by Etienne St-Onge, supervised by Maxime Descoteaux. *Tractoflow* pipeline initial implementation, and technical assistance was done by Guillaume Theaud. Processing pipeline adaptation for CIVET tissue maps, surfaces and SET was implemented by Etienne St-Onge, and rigorously tested by Noor B Al-Sharif, resulting in *Tractoflow-pve*.

Conflict of Interest

Maxime Descoteaux is co-founder of Imeka Solutions Inc. Other authors declare they have no actual or potential competing financial interests.

Data Citations

Data and research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. Dataset identifier(s): 10.15154/1519178. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDA.

References

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.

Basser, P.J., Mattiello, J., LeBihan, D., 1994. Mr diffusion tensor spectroscopy and imaging. Biophysical journal 66, 259–267.

Benes, F.M., Turtle, M., Khan, Y., Farol, P., 1994. Myelination of a key relay zone in the hippocampal formation occurs in the human brain during childhood, adolescence, and adulthood. Archives of general psychiatry 51, 477–484.

Collins, D.L., Zijdenbos, A.P., Baaré, W.F., Evans, A.C., 1999. Animal+ insect: improved cortical structure segmentation, in: Biennial International Conference on Information Processing in Medical Imaging, Springer. pp. 210–223.

Dell'Acqua, F., Tournier, J.D., 2019. Modelling white matter with spherical deconvolution: How and why? NMR in Biomedicine 32, e3945.

Deoni, S.C., Mercure, E., Blasi, A., Gasston, D., Thomson, A., Johnson, M., Williams, S.C., Murphy, D.G., 2011. Mapping infant brain myelination with magnetic resonance imaging. Journal of Neuroscience 31, 784–791.

Descoteaux, M., 2015. High angular resolution diffusion imaging (hardi). Wiley Encyclopedia of Electrical and Electronics Engineering , 1–25.

Descoteaux, M., Deriche, R., Knösche, T.R., Anwander, A., 2009. Deterministic and probabilistic tractography based on complex fibre orientation distributions. Medical Imaging, IEEE Transactions on 28, 269–286.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage 31, 968–980.

Dyrby, T.B., Lundell, H., Burke, M.W., Reislev, N.L., Paulson, O.B., Ptito, M., Siebner, H.R., 2014. Interpolation of diffusion weighted imaging datasets. NeuroImage 103, 202–213.

Giedd, J.N., Rapoport, J.L., 2010. Structural mri of pediatric brain development: what have we learned and where are we going? Neuron 67, 728–734.

Girard, G., Whittingstall, K., Deriche, R., Descoteaux, M., 2014. Towards quantitative connectivity analysis: reducing tractography biases. Neuroimage 98, 266–278.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the human connectome project. Neuroimage 80, 105–124.

Jbabdi, S., Sotiropoulos, S.N., Haber, S.N., Van Essen, D.C., Behrens, T.E., 2015. Measuring macroscopic brain connections in vivo. Nature neuroscience 18, 1546–1555.

Jernigan, T.L., Baaré, W.F., Stiles, J., Madsen, K.S., 2011. Postnatal brain development: structural imaging of dynamic neurodevelopmental processes, in: Progress in brain research. Elsevier. volume 189, pp. 77–92.

Jernigan, T.L., Brown, T.T., Hagler Jr, D.J., Akshoomoff, N., Bartsch, H., Newman, E., Thompson, W.K., Bloss, C.S., Murray, S.S., Schork, N., et al., 2016. The pediatric imaging, neurocognition, and genetics (ping) data repository. Neuroimage 124, 1149–1154.

Jeurissen, B., Descoteaux, M., Mori, S., Leemans, A., 2019. Diffusion mri fiber tractography of the brain. NMR in Biomedicine 32, e3785.

Jones, D.K., 2008. Studying connections in the living human brain with diffusion mri. cortex 44, 936–952.

Jones, D.K., Cercignani, M., 2010. Twenty-five pitfalls in the analysis of diffusion mri data. NMR in Biomedicine 23, 803–820.

Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I., Evans, A.C., 2005. Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. Neuroimage 27, 210–221.

Klein, A., Tourville, J., 2012. 101 labeled brain images and a consistent human cortical labeling protocol. Frontiers in neuroscience 6, 171.

Klingberg, T., Vaidya, C.J., Gabrieli, J.D., Moseley, M.E., Hedehus, M., 1999. Myelination and organization of the frontal white matter in children: a diffusion tensor mri study. Neuroreport 10, 2817–2821.

Lebel, C., Walker, L., Leemans, A., Phillips, L., Beaulieu, C., 2008. Microstructural maturation of the human brain from childhood to adulthood. Neuroimage 40, 1044–1055.

Lehmann, T.M., Gonner, C., Spitzer, K., 1999. Survey: Interpolation methods in medical image processing. IEEE transactions on medical imaging 18, 1049–1075.

Lenroot, R.K., Giedd, J.N., 2006. Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. Neuroscience & biobehavioral reviews 30, 718–729.

Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: Dicom to nifti conversion. Journal of neuroscience methods 264, 47–56.

Lim, K.O., Helpern, J., 2002. Neuropsychiatric applications of dti–a review. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo 15, 587–593.

Lyttelton, O., Boucher, M., Robbins, S., Evans, A., 2007. An unbiased iterative group registration template for cortical surface analysis. Neuroimage 34, 1535–1544.

Mukherjee, P., Miller, J.H., Shimony, J.S., Philip, J.V., Nehra, D., Snyder, A.Z., Conturo, T.E., Neil, J.J., McKinstry, R.C., 2002. Diffusion-tensor mr imaging of gray and white matter development during normal human brain maturation. American Journal of Neuroradiology 23, 1445–1456.

Nagy, Z., Westerberg, H., Klingberg, T., 2004. Maturation of white matter is associated with the development of cognitive functions during childhood. Journal of cognitive neuroscience 16, 1227–1233.

Olesen, P.J., Nagy, Z., Westerberg, H., Klingberg, T., 2003. Combined analysis of dti and fmri data reveals a joint maturation of white and grey matter in a fronto-parietal network. Cognitive Brain Research 18, 48–57.

19

Paus, T., Zijdenbos, A., Worsley, K., Collins, D.L., Blumenthal, J., Giedd, J.N., Rapoport, J.L., Evans, A.C., 1999. Structural maturation of neural pathways in children and adolescents: in vivo study. Science 283, 1908–1911.

Prayer, D., Prayer, L., 2003. Diffusion-weighted magnetic resonance imaging of cerebral white matter development. European journal of radiology 45, 235–243.

Qiu, A., Mori, S., Miller, M.I., 2015. Diffusion tensor imaging for understanding brain development in early life. Annual review of psychology 66, 853–876.

Shatz, C.J., 1990. Impulse activity and the patterning of connections during cns development. Neuron 5, 745–756.

Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2012. Anatomically-constrained tractography: improved diffusion mri streamlines tractography through effective use of anatomical information. Neuroimage 62, 1924–1938.

Sotiropoulos, S.N., Zalesky, A., 2019. Building connectomes using diffusion mri: why, how and but. NMR in Biomedicine 32, e3752.

Sowell, E.R., Thompson, P.M., Toga, A.W., 2004. Mapping changes in the human cortex throughout the span of life. The Neuroscientist 10, 372–392.

St-Onge, E., Daducci, A., Girard, G., Descoteaux, M., 2018. Surface-enhanced tractography (set). Neuroimage 169, 524–539.

Stiles, J., Jernigan, T.L., 2010. The basics of brain development. Neuropsychology review 20, 327–348.

Theaud, G., Houde, J.C., Boré, A., Rheault, F., Morency, F., Descoteaux, M., 2020. Tractoflow: A robust, efficient and reproducible diffusion mri pipeline leveraging nextflow & singularity. NeuroImage , 116889.

Tohka, J., Zijdenbos, A., Evans, A., 2004. Fast and robust parameter estimation for statistical partial volume models in brain mri. Neuroimage 23, 84–97.

Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. NeuroImage 35, 1459–1472.

Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.H., Connelly, A., 2019. Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation. NeuroImage , 116137.

Zijdenbos, A., Forghani, R., Evans, A., 1998. Automatic quantification of ms lesions in 3d mri brain data sets: validation of insect, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 439–448.

20

Appendix A. Full Comparison

	WM	GM	SC	CSF
AD	$0.00111, 0.000033$	$0.00116, 0.000072$	$0.00104, 0.000033$	$0.00160, 0.000146$
MD	$0.00080, 0.000033$	$0.00103, 0.000059$	$0.00082, 0.000032$	$0.00144, 0.000125$
RD	$0.00065, 0.000035$	$0.00096, 0.000054$	$0.00070, 0.000033$	$0.00135, 0.000117$
AFD	$0.332, 0.022$	$0.332, 0.026$	$0.360, 0.027$	$0.244, 0.030$
FA	$0.340, 0.019$	$0.129, 0.011$	$0.254, 0.015$	$0.114, 0.020$
GA	$0.529, 0.038$	$0.191, 0.026$	$0.375, 0.024$	$0.190, 0.050$
dice	$0.897, 0.010$	$0.877, 0.015$	$0.930, 0.019$	$0.574, 0.039$

Table A.1: Diffusivity measures (AD, MD, RD) in mm^2/s, diffusion measures (AFD, FA, GA) and tissue maps dice alignment score, average & variance over all subjects, for each segmentation maps.

	first eigenvalue	second eigenvalue	ratio
a	$0.001709, 0.000124$	$0.000388, 0.000026$	$4.40, 0.16$
c	$0.001686, 0.000076$	$0.000391, 0.000019$	$4.32, 0.14$
d	$0.001887, 0.000196$	$0.000416, 0.000041$	$4.54, 0.18$
f	$0.001850, 0.000207$	$0.000413, 0.000041$	$4.47, 0.17$
h	$0.001836, 0.000201$	$0.000418, 0.000039$	$4.39, 0.18$
i	$0.001716, 0.000087$	$0.000387, 0.000019$	$4.44, 0.16$
j	$0.001932, 0.000211$	$0.000425, 0.000042$	$4.54, 0.18$
l	$0.001735, 0.000128$	$0.000390, 0.000022$	$4.45, 0.18$
m	$0.001782, 0.000183$	$0.000397, 0.000036$	$4.49, 0.19$

Table A.2: Estimated fiber response function (frf) in mm^2/s across acquisition sites: first eigenvalue, second eigenvalue, ratio. Philips Achieva (a,h), GE Signa HDx (c), GE Discovery MR750 (i) and Siemens TrioTim (d,f,j,l,m).

	voxel L1	voxel L2	SET L1	SET L2
a	$0.278, 0.027$	$0.0234, 0.0029$	$0.385, 0.040$	$0.0331, 0.0055$
c	$0.287, 0.026$	$0.0242, 0.0032$	$0.382, 0.036$	$0.0303, 0.0068$
d	$0.301, 0.028$	$0.0254, 0.0030$	$0.375, 0.041$	$0.0278, 0.0050$
f	$0.317, 0.024$	$0.0266, 0.0031$	$0.417, 0.041$	$0.0331, 0.0058$
h	$0.299, 0.026$	$0.0251, 0.0029$	$0.435, 0.037$	$0.0370, 0.0048$
i	$0.291, 0.024$	$0.0245, 0.0028$	$0.358, 0.038$	$0.0269, 0.0051$
j	$0.291, 0.033$	$0.0243, 0.0033$	$0.378, 0.044$	$0.0292, 0.0057$
l	$0.310, 0.031$	$0.0252, 0.0027$	$0.429, 0.038$	$0.0347, 0.0063$
m	$0.283, 0.023$	$0.0236, 0.0026$	$0.363, 0.039$	$0.0274, 0.0052$

Table A.3: Connectivity matrices L1, L2 and χ^2 distances, per acquisition site, to the average matrix. Philips Achieva (a,h), GE Signa HDx (c), GE Discovery MR750 (i) and Siemens TrioTim (d,f,j,l,m).