A comparison of three fiber tract delineation methods and their impact on white matter analysis

Valerie Sydnor, Ana María Rivas-Grajales, Amanda Lyall, Fan Zhang, Sylvain Bouix, Sarina Karmacharya, Martha Shenton, Carl-Fredrik Westin, Nikos Makris, Demian Wassermann, et al.

HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2019 September 01.

Published in final edited form as:

Neuroimage. 2018 September ; 178: 318–331. doi:10.1016/j.neuroimage.2018.05.044.

A Comparison of Three Fiber Tract Delineation Methods and Their Impact on White Matter Analysis

Valerie J. Sydnor[#a], **Ana María Rivas-Grajales**[#a,b], **Amanda E. Lyall**[a,b], **Fan Zhang**[c,d], **Sylvain Bouix**[a], **Sarina Karmacharya**[a], **Martha E. Shenton**[a,b,d,e], **Carl-Fredrik Westin**[c,d], **Nikos Makris**[a,b], **Demian Wassermann**[a,f,g], **Lauren J. O'Donnell**[c,d], and **Marek Kubicki**[a,b,d]

[a]Psychiatry Neuroimaging Laboratory, Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[b]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, USA

[c]Laboratory for Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[d]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[e]VA Boston Healthcare System, Brockton Division, Brockton, MA, USA

[f]Athena, Université Cote d'Azur, Inria, France

[g]Parietal, CEA, Université Paris-Saclay, INRIA Saclay Île-de-France

[#] These authors contributed equally to this work.

Abstract

Diffusion magnetic resonance imaging (dMRI) is an important method for studying white matter connectivity in the brain *in vivo* in both healthy and clinical populations. Improvements in dMRI tractography algorithms, which reconstruct macroscopic three-dimensional white matter fiber pathways, have allowed for methodological advances in the study of white matter; however, insufficient attention has been paid to comparing post-tractography methods that extract white matter fiber tracts of interest from whole-brain tractography. Here we conduct a comparison of three representative and conceptually distinct approaches to fiber tract delineation: 1) a manual multiple region of interest-based approach, 2) an atlas-based approach, and 3) a groupwise fiber clustering approach, by employing methods that exemplify these approaches to delineate the arcuate fasciculus, the middle longitudinal fasciculus, and the uncinate fasciculus in 10 healthy male subjects. We enable qualitative comparisons across methods, conduct quantitative evaluations

Corresponding author: Marek Kubicki, M.D., Ph.D., Psychiatry Neuroimaging Lab, 1249 Boylston Street, 3rd Floor, Boston, MA 02215, Telephone: 617-525-6234, kubicki@bwh.harvard.edu.

Declarations of interest: None

of tract volume, tract length, mean fractional anisotropy, true positive and true negative rates, and report measures of intra-method and inter-method agreement. We discuss methodological similarities and differences between the three approaches and the major advantages and drawbacks of each, and review research and clinical contexts for which each method may be most apposite. Emphasis is given to the means by which different white matter fiber tract delineation approaches may systematically produce variable results, despite utilizing the same input tractography and reliance on similar anatomical knowledge.

Keywords

Diffusion MRI; Tractography; White Matter; Fiber Tract; Automatic classification of white matter tracts

1. Introduction

Diffusion magnetic resonance imaging (dMRI) is one of the most important methods for studying the connectivity and integrity of white matter pathways in the brain. Diffusion images are commonly analyzed using tractography, a method that reconstructs three-dimensional white matter fiber pathways *in vivo* based on water diffusion orientation information at each image voxel (Jones et al., 1999; Lazar et al., 2003; Malcolm et al., 2010; Mori et al., 1999). Tractography is widely used to investigate potential white matter pathology in psychiatric disorders (e.g., see review by Kubicki et al., 2007), cognitive disorders (e.g., Chua et al., 2008), nervous system disorders (e.g., Horsfield and Jones, 2002), mild traumatic brain injury (e.g., Shenton et al., 2012), and aging (e.g., Sullivan and Pfefferbaum, 2006), as well as in analyses that aim to uncover white matter trajectories and characterize the architecture of brain connectivity in healthy and diseased individuals (Catani et al., 2002; Essayed et al., 2017; Gong et al., 2009; Jolles et al., 2016; Makris et al., 2013a, 2013b, 2009, 2007, 2005, 2002; Rojkova et al., 2016; Sherbondy et al., 2005; Wakana et al., 2004).

Recent emphasis on improving dMRI tractography algorithms has allowed for significant methodological advances in the study of white matter fiber tracts, namely, the capacity to handle branching or crossing fibers, and the ability to reconstruct tracts in areas of low signal (Behrens et al., 2003; Malcolm et al., 2010; Parker et al., 2003; Tournier J-Donald et al., 2012). To date, a fair amount of work has been dedicated to comparing different tractography algorithms (Behrens et al., 2007; Bürgel et al., 2009; Fillard et al., 2011; Jones, 2008; Khalsa et al., 2014; Tensaouti et al., 2011) and to assessing or validating tractography against post-mortem tract tracing (Dauguet et al., 2007; Donahue et al., 2016; Dyrby et al., 2007; Knösche et al., 2015; Seehaus et al., 2013), diffusion phantoms (Fillard et al., 2011; Fritzsche et al., 2010; Poupon et al., 2008; Pullens et al., 2010), and simulated ground truth bundles (Maier-Hein et al., 2017). After tractography algorithms generate potential macroscopic fiber pathways, or “streamlines”, however, anatomical and functional interpretation of the tractography results remains ambiguous until streamlines are grouped into distinctive white matter bundles based either on the morphology of known white matter tracts, or on an understanding of the brain structures the streamlines connect. Nevertheless,

insufficient attention has been paid to directly comparing tractography post-processing methods that facilitate white matter bundle delineation and extraction, despite the fact that different methodologies may produce inconsistent or discrepant output. A more comprehensive understanding of potential methodological differences is needed in order to enable (or prevent) appropriate (or inappropriate) comparisons of results across studies, and in order to better appreciate the relative advantages and limitations of diverse methods.

Tractography streamlines can be generated from whole brain diffusion data, i.e. “whole-brain tractography”, by simultaneously placing seed points throughout all voxels in the brain. Subsequently, white matter tracts of interest can be identified and extracted from the whole-brain tractography. Three methods are commonly employed to accomplish this: 1) manually placing regions of interest (ROIs) around the “stems” (the most compact portions) of fiber bundles (Catani et al., 2002; Catani and Thiebaut de Schotten, 2008a; Conturo et al., 1999; Mori et al., 2002; Sherbondy et al., 2005; Wakana et al., 2007); 2) selecting tractography streamlines based on the atlas-defined gray matter or white matter regions they penetrate or terminate in (Hua et al., 2008; Lawes et al., 2008; Suarez et al., 2012; Wassermann et al., 2016; Yendiki et al., 2011; Zhang et al., 2010); and 3) grouping streamlines into clusters based on fiber similarity properties (Brun et al., 2004; Ding et al., 2003; Garyfallidis et al., 2012; Guevara et al., 2012; O’Donnell and Westin, 2007).

Manual ROI methods (see 1 above) involve identifying and drawing inclusionary and exclusionary ROIs on Fractional Anisotropy (FA) or color-by-orientation Diffusion Tensor Image (DTI) maps, and selecting only the streamlines that pass through inclusionary ROIs. White matter fiber pathways are typically organized into three portions, a compact, homogenous bundle, called the “stem”, an area of fiber divergence, the “spray”, and a widespread peripheral area with fiber terminations, referred to as the “extreme periphery” (Makris et al., 1997). Manual methods traditionally involve the placement of ROIs around tract stems, with ROI placement being guided by methodologies established in the literature, or by a neuroanatomist who is an expert in white matter anatomy. Manual ROI methods are considered the gold standard for white matter delineation given their subject-specific nature, their reliance on expert neuroanatomical knowledge, and their potential for high inter-rater and intra-rater reliability with regards to ROI placement, ROI volume, tract streamline number, and tract measures such as FA (Catani and Thiebaut de Schotten, 2008a; Malykhin et al., 2008; Wakana et al., 2007). Nonetheless, proper placement of ROIs, essential for accuracy, reliability and reproducibility, necessitates that the experimenter has neuroanatomical knowledge and is precise with ROI placements, and that the ROI methodology being utilized is robust. Manual methods additionally have inherent shortcomings in that they are very time-intensive, and they can only be utilized to delineate previously identified white matter pathways with known, uncomplicated trajectories (Zhang et al., 2010).

To address issues of time, labor, and ease of reproducibility, automated approaches to white matter tract delineation have been developed. Atlas-based methods (see 2 above) are automated approaches that require anatomical atlases and diffusion images or DTIs to be co-registered prior to tract extraction. White matter tract extraction then follows a set of instructions, which dictates which atlas-defined gray and/or white matter ROIs each fiber

tract should (and should not) pass through. Atlas-based methods allow for the extraction of a large number of white matter tracts in an efficient and reproducible manner, using protocols that are transparent to users and that minimize experimenter bias and the need for anatomical knowledge. Furthermore, in addition to being utilized to study well known white matter tracts, these methods are perhaps best suited for exploratory analyses. However, high quality output using these methods is contingent upon accurate registration of the atlas in use to each subject's data, and tract definitions are constrained to ROIs that are included in the atlas (e.g., atlases from Desikan et al., 2006; Fischl et al., 2002; Mori et al., 2011; Oishi et al., 2009; Varentsova et al., 2014).

Fiber clustering approaches (see 3 above) are alternate automated methods for white matter analysis, wherein whole-brain tractography is grouped into a number of "fiber clusters" in a data-driven manner. Specifically, fiber clustering methods group streamlines that follow similar anatomical pathways into distinctive clusters, and groupwise clustering methods additionally identify and extract homologous clusters from all subjects (O'Donnell and Westin, 2007). This data-driven method could be considered initially anatomically unbiased, as no *a priori* anatomical information is input into the clustering algorithm. Subsequent to clustering, however, individual clusters can be visualized and identified as part of larger white matter tracts, though this often requires advanced neuroanatomical knowledge. Groupwise clustering methods may prove most beneficial for reliably identifying anatomically similar portions of white matter across subjects (O'Donnell et al., 2013); however, they may also reduce the ability to examine between-subject variability in tract shape and architecture. Furthermore, groupwise methods require co-registration of subjects' whole brain tractography to the same three-dimensional space, which introduces the potential for registration errors that can affect cluster classification.

Validation of automated white matter delineation methods typically entails demonstrating that a given method produces similar results to those obtained from a manual ROI method, as these are presumed to be the most specific and anatomically accurate methods available. Typically, a kappa value (k) is calculated to measure inter-method agreement between an automated and a manual ROI method, with a $k > .60$ indicating "substantial" agreement between methods, and a $k > .80$ indicating "almost perfect" agreement (Landis and Koch, 1977). Previous comparisons of manual and atlas-based white matter delineation methods have generally reported substantial agreement, with kappa values averaging around .75, and ranging from .65 to .95 (Wassermann et al., 2016; Zhang et al., 2010, 2008). Similar kappa values have also been found for comparisons of fiber clustering results to manual methods (Voineskos et al., 2009). It is not entirely evident, however, how fiber clustering and atlas-based methods may compare (although see Guevara et al., 2012; Yoo et al., 2015), or what analyses that extend beyond kappa values will reveal. The ways in which manual ROI-based, atlas-based, and groupwise fiber clustering methods differentially impact results obtained from the same dataset has yet to be extensively investigated.

The goal of the present study is to conduct a head to head comparison of three distinct post-processing methods used for white matter tract extraction that exemplify three major conceptual approaches to tract delineation. These methods are separately applied to the same data, thus ensuring that preprocessing and tractography generation steps are equivalent

across methods. This is the first study, to the best of our knowledge, to comprehensively compare examples from a manual ROI-based, a cortical atlas-based, and a data-driven approach. We utilize three methods that represent these approaches—Manual Multiple-ROI Labeling (guided by Catani and Thiebaut de Schotten (2008a)), The White Matter Query Language (Wassermann et al., 2016), and Data-Driven Groupwise Fiber Clustering (O'Donnell and Westin, 2007)—to extract three widely studied, functionally significant, and anatomically distinctive white matter tracts, namely: the arcuate fasciculus (a large, arcing fiber bundle), the middle longitudinal fasciculus (a long, straight, horizontal fiber bundle), and the uncinate fasciculus (a hook-shaped fiber bundle with a compact stem). We examine each method's influence on tract shape, tract volume, tract length, fractional anisotropy and anatomical connectivity, enable visualization of areas within each tract where the three methods are predominantly convergent and divergent, and compute Jaccard Coefficients and Intraclass Correlation Coefficients, standard methods employed to evaluate inter-method spatial overlap and inter-method reliability, respectively. We additionally report intra-method sensitivity, a measure of the extent to which each method correctly identifies fibers that do belong to a given tract (i.e. true positives), and intra-method specificity, a measure of how often each method correctly identifies white matter regions that are not part of a given tract (i.e. true negatives). Methodological similarities and differences across the three approaches are discussed, and the clinical and research contexts for which each approach may be most apposite are reviewed.

2. Methods

2.1. Image Acquisition and Preprocessing

Magnetic Resonance (MR) Images from 10 healthy, right-handed male subjects (mean age: 22.29 years, standard deviation: 1.68) were included in the analysis. Images were acquired as part of the larger Boston Center for Intervention Development and Applied Research (CIDAR) study (www.bostoncidar.org). Inclusion and exclusion criteria are described in Lee et al. (2013). Images were obtained on a 3-Tesla whole body MRI Echo speed system General Electric scanner (GE Medical Systems, Milwaukee) at the Brigham and Women's Hospital, Boston, MA. Diffusion images were acquired using an echo planar image sequence with the following parameters: TR/TE=17000 ms/78 ms; 51 gradient directions with b=900 s/mm^2 and 8 additional b=0 images. Each volume consisted of 85 axial slices of 1.7 mm thickness, and had an acquisition matrix of 144×144 in a field of view of 240×240 mm^2, and a voxel dimension of 1.7×1.67×1.67 mm. T1-weighted and T2-weighted images were acquired with the following parameters: 176 slices, TR/TE= 7.4 ms/3 ms, TI = 600, 10° flip angle, acquisition matrix of 256 × 256, field of view of 256×256 mm^2, and a voxel dimension of 1×1×1 mm.

Images were visually examined to ensure high quality prior to preprocessing. Preprocessing steps for the diffusion data included motion and eddy current correction using an affine registration algorithm in FSL (http://www.fmrib.ox.ac.uk/fsl), followed by EPI geometric distortion correction using an in-house script and advanced normalization tools' (ANTS) symmetric diffeomorphic mapping (Avants et al., 2008) to warp the diffusion image along the phase encoding direction to an undistorted T2-weighted image (https://github.com/

pnlbwh/pnlutil). Diffusion tensor images were estimated from the diffusion data in 3D Slicer Version 4.5 (https://www.slicer.org) via the SlicerDMRI project (http://dmri.slicer.org) (Norton et al., 2017) using a weighted-least-squares estimation. Whole-brain two-tensor tractography was computed using a multi-fiber tracking algorithm that utilizes an Unscented Kalman Filter (UKF) (described in (Malcolm et al., 2010)), with a free water correction applied (Baumgartner et al., 2012). This multi-tensor UKF tractography algorithm recursively estimates tensor model parameters (e.g. eigenvalues and eigenvectors) and local fiber orientation along each point in a streamline based on the previous point, allowing the algorithm to perform well in the presence of noise and uncertainty (Fillard et al., 2011; Malcolm et al., 2010, 2009) as well as in the presence of crossing fibers (Baumgartner et al., 2012; Fillard et al., 2011).

T1-weighted images were processed with FreeSurfer (Fischl, 2012), a publicly available software package (http://www.martinos.org/freesurfer), in order to parcellate brain gray and white matter into distinct anatomical regions. FreeSurfer's gray and white matter parcellation is based on Fischl et al.'s (2004, 2002) automated neuroanatomical structure segmentation. FreeSurfer anatomical labelmaps were registered to diffusion space using ANTS by performing a rigid registration of the T1 to the T2, followed by a non-linear registration of the T2 to a diffusion b=0 volume, and applying a composite transformation to the FreeSurfer labelmap.

2.2. White Matter Tract Delineation Methods

Three tractography post-processing white matter tract (fiber tract) delineation methods were applied to data from the 10 subjects. The arcuate fasciculus (AF), the middle longitudinal fasciculus (MdLF) and the uncinate fasciculus (UF) were extracted from the whole-brain two-tensor tractography using the following methods: 1) Manual Multiple-ROI Labeling (Manual), in which ROIs are drawn around the stems of fiber bundles, 2) the White Matter Query Language (WMQL), a fully automated method that defines white matter tracts based on their endpoints in anatomically parcellated gray and white matter, and 3) Data-Driven Groupwise Fiber Clustering (Clustering), a semi-automated method that groups streamlines from whole-brain tractography into groups of fibers ("fiber clusters") based on fiber similarity properties.

2.2.1. Manual Multiple-ROI Labeling (Manual)—White matter ROIs were identified and drawn manually on color-by-orientation (directionally encoded color) DTI maps using 3D Slicer. Inclusion and exclusion ROIs were drawn for the AF, the MdLF, and the UF. ROI placement was based on current anatomical knowledge of tract trajectories. Color-by-orientation DTI maps follow an RGB (red-green-blue) scheme (Makris et al., 1997) to indicate the major eigenvector direction, which aligns with the direction of major white matter tracts (Beaulieu, 2002) (red: left-right, green: anterior-posterior, blue: dorsal-ventral). Anatomical ROI definitions for the AF and the UF were derived from the procedures specified by Catani and Thiebaut de Schotten (2008a). ROI definitions for the MdLF were based on criteria described in Makris et al. (2013), as this tract is not included in Catani and Thiebaut de Schotten (2008a). Streamlines identified using the whole-brain two-tensor tracking algorithm that penetrated the manually defined inclusion ROIs (and did not

penetrate exclusion ROIs) were assigned to the white matter tract associated with those ROIs.

The AF is an arcing fiber bundle that connects the inferior frontal gyrus to the middle temporal gyrus, the posterior portion of the superior temporal gyrus, and portions of the lateral temporo-occipital transition regions (Catani et al., 2005; Makris et al., 2005). The stem of the AF was delineated with 2 inclusion ROIs drawn in the axial plane, and 2 exclusion ROIs drawn in coronal and sagittal planes (Figure 1A).

The MdLF is a long, horizontally-oriented fiber bundle beginning in the temporal pole that runs within the superior temporal gyrus and bifUrcates caudally to connect to the angular gyrus (Makris and Pandya, 2009) and the superior parietal lobule (Makris et al., 2013; Wang et al., 2013). The stem of the MdLF was identified with 4 inclusion ROIs and 2 exclusion ROIs drawn in the coronal plane (Figure 1B).

The UF is a hook-shaped fiber bundle that connects the anterior temporal lobe and the frontal lobe, where it spans medial and lateral regions, extending into orbital regions and the middle and inferior frontal gyri (Makris et al., 1997). The stem of the UF was delineated using 3 inclusion ROIs and 2 exclusion ROIs drawn in the coronal plane (Figure 1C).

Intra-rater reliability was calculated for the extraction of white matter tracts using the Manual Multiple-ROI Labeling method via a two-way random Intraclass Correlation analysis with measures of absolute agreement. Intra-rater reliability was excellent for tract volume (AF: ICC= 0.92, MdLF: ICC= 0.81, UF: ICC= 0.94) and FA (AF: ICC= 0.99, MdLF: ICC= 0.90, UF: ICC= 0.99).

2.2.2. White Matter Query Language (WMQL)—WMQL (Wassermann et al., 2016) is an automated fiber tract delineation method in which white matter tracts are defined, from a theoretical standpoint, by the regions of brain gray matter they are understood to connect. White matter tracts are delineated in WMQL by identifying streamlines that connect specified ROIs in the FreeSurfer atlas, namely, FreeSurfer ROIs that are included in a given tract's anatomical "definition". Tract definitions in WMQL are written as queries (created by expert anatomist NM), which inform the software which streamlines to extract from whole-brain tractography. Table 1 contains the WMQL query definitions used to extract the AF, the MdLF and the UF. The queries are an updated version of those originally presented in Wassermann et al. (2016); updates reflect improvements following query testing and modification by NM and colleagues. It should be noted that the anatomical regions specified in the queries in Table 1 include both the corresponding cortical region and the adjacent white matter region. For example, "temporalpole.side" includes both the FreeSurfer region "ctx-*h-temporalpole" and the region "wm-*h-temporalpole".

2.2.3. Data-Driven Groupwise Fiber Clustering (Clustering)—Spectral fiber clustering was implemented using the freely available pipeline first described in O'Donnell and Westin (2007) (https://github.com/SlicerDMRI/whitematteranalysis). First, a groupwise entropy-based registration was performed in order to register whole-brain tractography from each of the 10 subjects into a common space (referred to as group space) (O'Donnell et al.,

2012). A high-dimensional, study-specific, multi-subject clustering atlas was then generated by grouping subsets of fibers from all 10 subjects into 800 distinct clusters. Clusters were created based on fiber distance, fiber shape, and cluster commonality across hemispheres and across subjects, following the method described in O'Donnell and Westin (2007). Four iterations of group outlier detection were performed during the generation of the multi-subject clustering atlas, by iteratively clustering and identifying outlier fibers that had a low probability (more than 2 standard deviations away from cluster mean probability) of belonging to a given cluster. The multi-subject clustering atlas was then used to group tractography from each individual subject into 800 homologous white matter clusters; only streamlines with a fiber length of greater than 50 mm were clustered. Finally, subject-specific outlier removal was performed to remove potentially aberrant fibers from subject clusters, again with a threshold of 2 standard deviations.

All 800 clusters were visually examined for each of the 10 subjects in order to determine candidate clusters for the AF, the MdLF, and the UF. Clusters with fibers that were primarily part of a given tract in at least 70% of subjects were included in the finalized tract. Decisions were guided by an anatomical expert (NM). The AF was best characterized by a subset of 15 clusters in all subjects, the MdLF by 9 clusters, and the UF by 8 clusters. Tracts were extracted and registered from group space back to subject space using an inverse transform.

2.3. Qualitative Comparisons of Intra- and Inter-Method Overlap: Method-Specific Tract Overlay Maps and Tract Heatmap Generation

2.3.1. Method-Specific Tract Overlay Maps for Intra-Method Overlap—To allow for a qualitative comparison of the three methods, method-specific tract overlay maps were generated, for each method, to visually represent the extent of tract overlap across subjects. To generate the method-specific tract overlay maps, the 90 subject-specific fiber tracts were converted into binary subject-specific tract masks by assigning a label value of 1 to all voxels that contained tract streamlines. The resultant subject-specific tract masks were registered to the MNI152 T1-weighted standard brain, to allow for overlap quantification. A series of registration steps was carried out with ANTS, as follows: a nonlinear transformation from subject diffusion space to subject T2 space, a rigid transformation from subject T2 space to subject T1 space, a nonlinear transformation from subject T1 space to the MNI152 T1 template. Following registration, binary method-specific tract overlay maps with an inclusion threshold of 50% of subjects were created. Specifically, for each method/tract, voxels that were labeled as "1" in at least 50% of subjects were assigned a value of 1, and voxels that were labeled as "1" in less than 50% of subjects were assigned a value of 0. The resulting method-specific tract overlay maps thereby display the degree of intra-method tract overlap. The 50% inclusion threshold was chosen in order to identify voxels that had streamlines in a majority of subjects (and thus represented a consistent anatomical tract feature), while still conserving a level of between-method variability in overlap. Volume rendering was performed on the binary method-specific tract overlay maps to generate 3D models for visualization.

2.3.2. Heatmaps for Inter-Method Overlap—The binary method-specific tract overlay maps generated for each method for the AF, the MdLF and the UF were overlaid on

the MNI152 template to produce heatmaps that display spatial agreement across the three methods. Heatmaps were used to enable visualization of tracts areas where the three methods have high and low degrees of overlap.

2.4. Quantitative Comparisons

The following measures were used to quantitatively evaluate similarities and differences across the three methods: a) Volumes of the Method-Specific Tract Overlay Maps (see 2.3.1); b) Intraclass Correlation Coefficients; c) Normalized Tract Volume; d) Tract Length; e) FA; f) The Jaccard Coefficient; g) Simultaneous Truth and Performance Level Estimation (STAPLE) Algorithm sensitivity and specificity (Warfield et al., 2004).

2.4.1. The Jaccard Coefficient (JC)—The JC is a similarity measure that evaluates the degree of agreement between two datasets. It is defined as the size of the intersection over the size of the union, $JC = |X \cap Y|/|X \cup Y|$, where X and Y are two sets of voxels obtained from binary subject-specific tract masks produced by methods “X” and “Y”. A JC was computed for each method pair (i.e., Manual and WMQL, Manual and Clustering, and Clustering and WMQL) by computing the JC for each method pair for each subject, and taking the average across subjects. Additionally, a JC was computed to measure agreement among the three methods, $JC = (|X \cap Y \cap Z|/|X \cup Y \cup Z|)$, by computing subject-specific JCs and taking the average across all ten subjects. In the three method JC, the numerator represents the number of voxels labeled as 1 in subject-specific tract masks in all three methods, and the denominator represents the total number of voxels labeled as 1 in any of the three methods.

2.4.2. The Simultaneous Truth and Performance Level Estimation Algorithm —The STAPLE algorithm (Warfield et al., 2004) utilizes information from segmentations produced by multiple methods in order to estimate an optimal combination of the segmentations based on common agreement across inputs. Each input segmentation is assigned a weight depending on its estimated performance level. The probabilistic optimal segmentation and method-specific performance levels are estimated iteratively via an Expectation-Maximization (EM) algorithm. Performance level measures for each method with respect to estimated common agreement are expressed in terms of sensitivity (“p”) and specificity (“q”), which, here, were calculated for each tract generated by each method. The sensitivity measure (p) referenced throughout this paper represents the “true positive rate”, and is equal to the percentage of voxels that are correctly identified as being part of a tract, and the specificity measure (q) described here represents the “true negative rate” and reflects the percentage of voxels that are correctly identified as not being part of a tract. These measures will be affected by both between and within method variability; methods that are internally consistent and that have greater spatial overlap with the other methods under consideration will generally have better performance (higher p and q) than methods that are internally variable or that have lower spatial agreement with other methods.

2.5. Streamline Endpoint Analysis

To compare tract anatomical connectivity across methods, a streamline endpoint analysis was conducted using FreeSurfer anatomical labelmaps registered to diffusion images. For

each region in the FreeSurfer atlas, the percentage of a given tract's streamlines with endpoints in that region was computed. Percentages for each region were averaged over all subjects for each method and tract, and regions containing at least 5% of AF, MdLF or UF streamline endpoints were identified.

2.6. Statistical Analyses

Statistical analyses were conducted with SPSS Version 23. Mean and standard deviations for tract volume, tract length, and FA were obtained. Separate ANOVAs were conducted to compare mean tract volume, mean tract length, and mean FA values for the AF, the MdLF and the UF (1 ANOVA for each measure and each tract). Tract volume and FA were entered as the dependent variables, and 'hemisphere' and 'method' as the independent variables. If a significant main effect of 'method' was found, post-hoc Bonferroni-corrected t-tests were conducted.

3. Results

3.1. Qualitative Comparisons

3.1.1. 3D Streamline Visualization of Individual Subject White Matter Tracts Across Methods

3.1.2. Method-Specific Tract Overlay Maps: Intra-Method Overlap—Method-specific tract overlay maps were generated for the right and left AF, MdLF, and UF, to evaluate the degree of tract overlap within each of the three methods. Figure 3 displays the 3D volume renderings of these maps. Groupwise Fiber Clustering (Clustering) maps are displayed in green, Manual Multiple-ROI Labeling (Manual) maps in blue, and White Matter Query Language (WMQL) maps in red. The 3D volume models reveal tracts of similar shape and size across methods, however several differences should be noted. Across the AF models, each method generates a distinct anatomical feature not seen in the other two methods. For Clustering, it takes the form of a posterior projection off of the dorsal portion of the tract's arch. For the Manual method, the AF extends further anteriorly towards the frontal pole. For WMQL, the AF includes fibers connecting the dorsal and ventral portions of the tract. Across the MDLF models, Clustering generates a tract with a larger stem and spray, and across the UF models, WMQL generates a tract with a smaller stem and spray.

3.1.3. Heatmaps: Inter-Method Overlap—To evaluate the degree to which the three methods overlap spatially when delineating the AF, the MdLF and the UF, heatmaps were created for each tract (Figure 4).

3.2. Quantitative Comparisons

3.2.1. Method-Specific Tract Overlay Map Volumes—Volumes of the method-specific tract overlay maps (displayed in Figure 3) are provided in Table 2. Volumes are largest for the tracts produced by Clustering. Manual volumes are larger than WMQL volumes for the AF and the UF, but smaller for the MdLF.

3.2.2. Intraclass Correlation Analysis—Inter-method reliability was evaluated with an ICC analysis. Two-way mixed ICCs with measures of absolute agreement were calculated

for each fiber tract, based on tract volume. The degree of consistency among the three methods was highest for the UF (ICC=0.727, Cronbach's alpha=0.863), followed by the AF (ICC=0.616, Cronbach's alpha=0.743) and the MdLF (ICC=0.348, Cronbach's alpha=0.505), representing excellent (UF and AF) and fair (MdLF) agreement.

3.2.3. Tract Volume, Tract Length, and Fractional Anisotropy—Tract Volume, Tract Length, and FA values (Table 3) were compared between methods. Volume and length measurements were largely variable across methods, with no apparent pattern across methods or tracts. FA values were consistent across methods for the AF; differences were within .01. FA values were slightly more variable for the MdLF and the UF, with Clustering exhibiting the highest FA, followed by Manual, and then WMQL.

ANOVAs were conducted for each fiber tract (AF, MdLF, UF) and measure (tract volume, tract length, and FA). The analysis of tract volume showed a significant main effect of method for all three fiber tracts (AF: $F_{(2,57)}$=8.646, p=.001; MdLF: $F_{(2,57)}$=14.304, p<0.001; UF: $F_{(2,57)}$=7.776, p=.001). Post-hoc analyses revealed differences between Clustering and WMQL for the AF (p < .0001), between Clustering and Manual (p < .0001) and WMQL and Manual (p= .001) for the MDLF, and between Clustering and WMQL (p= .001) and WMQL and Manual (p=.013) for the UF (see Figure 5A). The analysis of tract length showed a significant main effect of method for the MdLF ($F_{(2,57)}$=53.815, p<0.001) and the UF ($F_{(2,57)}$=23.527, p<0.001). Tract length was longer for Clustering than for Manual (p<0.001) and WMQL (p<0.001) in both tracts (Figure 5B). Likewise, the analysis of FA revealed a significant main effect of method for the MdLF ($F_{(2,57)}$=10.218, p<0.001) and the UF ($F_{(2,57)}$=15.282, p<0.001) only, and not for the AF ($F_{(2,57)}$=1.091, p=0.343). For both the MdLF and the UF, Clustering exhibited significantly greater FA than Manual (MDLF: p=002, UF: p=004) and WMQL (p<0.001 for MDLF and UF) (Figure 5C).

3.2.4. Jaccard Coefficients (JC)—Table 4 illustrates the degree of spatial overlap between methods, based on the JC. Overlap across all three methods was highest for the UF (JC=0.307), followed by the AF (JC=0.286) and the MdLF (JC=0.163) (Table 4). For the two-method comparisons, JCs were highest for Clustering and Manual, followed by Clustering and WMQL, and lowest for WMQL and Manual.

3.2.5. Sensitivity and Specificity Measures—Table 5 displays STAPLE sensitivity and specificity values for the three tract delineation methods. Sensitivity was highest for the Clustering method (mean=.860), followed by the Manual method (mean=.766) and WMQL (mean=.612). Sensitivity values (p) measure how often each method correctly identifies fibers that should belong to a given tract (true positives). Specificity (q) was high for all three tracts and methods (q > 0.99). Specificity values measure how often each method correctly identifies brain regions that should not be considered part of a given tract (true negatives).

3.3. Streamline Endpoint Analysis

To compare anatomical connectivity across methods, a streamline endpoint analysis was conducted. Table 6 presents anatomical regions containing at least 5% of a given tract's streamline endpoints by tract and method.

4. Discussion

This head to head comparison of three conceptually distinct approaches to white matter tract delineation—a manual multiple ROI-based approach, an atlas-based and a groupwise fiber clustering approach—corroborates that different approaches can extract similar white matter tracts from whole brain tractography. Differences between approaches in tract volume and length, FA, and intra-method tract overlap were, however, observed, as were incongruities in tract architecture and anatomical connectivity. Though some inter-method differences arise in an unsystematic manner, most observed dissimilarities occur systematically and can be explained by underlying differences in delineation strategy and methodology. Additionally, spatial agreement and metric consistency across approaches appear to be, to an extent, contingent upon the architecture of the tract of interest.

The qualitative output (Figures 2 and 3) substantiates that white matter tracts generated by the three methods are comparable in overall architecture, however differences in size, exemplified best by the MdLF and the UF, and alterations in tract anatomy, as seen with the AF, do occur. Heatmaps (Figure 4) depicting the degree of spatial overlap across methods for each tract illustrate that inter-method agreement is high in the stem of white matter tracts, though agreement decreases, and topological variability increases, at the spray and extreme periphery of tracts. Spatial agreement as measured by the Jaccard Coefficient was greatest between Manual and Clustering methods, most likely because the Manual method identifies fibers passing through the stems of white matter tracts, which tend to have a consistent shape, and Clustering is largely shape-based. Moreover, both of these methods are white matter-centric, i.e., based largely on white matter architecture. WMQL, on the other hand, is a gray matter-centric approach that depends primarily on an individual's cortical architecture.

Results from the STAPLE algorithm, a more advanced approach to investigating spatial correspondence across methods, suggest that all three methods were adept at excluding tractography streamlines that almost certainly do not belong to a given tract (true negatives). Clustering had the highest STAPLE sensitivity (true positive) value, signifying that it has high internal consistency, and that it most reliably identified streamline-containing voxels that were included in the output generated by the other two methods. The fact that tracts generated by Clustering encompass the majority of streamlines included in Manual and WMQL tracts, as well as additional streamlines, likely arises because in the Clustering method, streamlines are not constrained to passing through or ending in specific ROIs. Accordingly, the Clustering method will extract streamlines that get excluded from the other methods, for example streamlines on tract peripheries that do not pass through the stem (which may be excluded from the Manual method), streamlines that pass through or end in anatomically "incorrect" regions due to either misregistration or to the tract definition (which may be excluded from WMQL), or shorter streamlines that do not span all ROIs

(which may be excluded from both methods). The absence of precise inclusion/exclusion criteria in Clustering can, however, lead to output that is overinclusive: streamlines not strictly considered to belong to a tract of interest (based on the accepted understanding of the tract's anatomical location and the brain regions it connects) may be included in clustering methods if the streamlines share a similar shape with, and are located proximally to, the tract of interest (e.g. consider fibers of the inferior longitudinal fasciculus and the inferior fronto-occipital fasciculus).

The lower STAPLE sensitivity value for WMQL is driven by a combination of lower spatial agreement with the other methods (corroborated by the Jaccard Coefficient results) and greater tract variability across subjects. This greater inter-subject variability in tract location and shape is expected for a primarily gray matter-centric approach when compared to approaches that are white matter-centric, as gray matter architecture (especially cortical architecture) is more highly variable across individuals, whereas core white matter morphology is generally more consistent. It must additionally be noted that STAPLE metrics are entirely dependent on the tract segmentations input into the algorithm, and therefore output measures are biased to the most consistent spatial representation of a tract, which is not necessarily the most anatomically accurate one.

A more relevant question therefore may be, do the different methods extract white matter tracts with the same anatomical connectivity profiles? Results from the streamline endpoint analysis (Table 6) reveal that connectivity is variable across the methods, with Clustering being more anatomically expansive than the other two methods. Again, this may point to either (or more likely, to both) an overinclusiveness for tracts produced with Clustering, or to the use of inclusionary ROIs that are too restrictive by the other two methods. This analysis emphasizes the importance of understanding which regions of the brain a tract of interest connects, as this has direct implications for anatomical, functional, and clinical interpretation of diffusion MRI tractography results, as well as for the replicability of different studies analyzing the same white matter tract.

White matter tracts are most frequently extracted for the purpose of studying quantitative tract-specific measures, for example structural or scalar measures. Here it is apparent that connectivity variability co-occurs with considerable between-method variability in measures of tract volume and tract length. Significant differences in these measures were pervasive, with no definitive emergent pattern. FA tended to be marginally more reliable across the three methods, with no significant differences arising between FA values derived from Manual and WMQL tracts. Tracts produced by Clustering, however, tended to have significantly higher FA values, potentially due to outlier removal steps that remove irregularly shaped or cluster-inconsistent streamlines (which are most likely to occur on tract peripheries), and due to Clustering's ability to consistently extract all streamlines that comprise the core of white matter bundles (i.e., streamlines with high FA).

These results demonstrate the various ways in which the choice of white matter delineation method used can affect the tract that is produced and the measures that are derived from it, even when the same input tractography is used. Reciprocally, choices pertaining to which

white matter tract to study and which tractography method to employ can also affect the performance of these white matter delineation approaches.

4.1. Impact of white matter tract architecture, image resolution and tractography method on white matter delineation

In this study, spatial and volumetric congruence across methods, assessed by Jaccard Coefficients and Intraclass Correlation Coefficients respectively, were greatest for the UF, followed by the AF, and then the MdLF. Agreement among distinctive white matter delineation methods thereby appears to be influenced by tract shape. Methods may thus more consistently define white matter tracts with unique shapes for which true positive streamlines are more readily distinguished from surrounding linear streamlines—for example the UF, the AF, the cingulum bundle and the fornix—than white matter tracts with long, straight, and radiating or bifurcating fibers—for example the MdLF and the inferior longitudinal fasciculus, the inferior fronto-occipital fasciculus and the corticospinal tract.

Studying tracts with highly variable shape versus variable anatomical connectivity across subjects will also affect individual methods differently. Whereas groupwise fiber clustering methods will potentially perform more poorly when tracts are inconsistently shaped across subjects, especially if larger clusters are used, the accuracy of atlas-based methods will be more affected by diversity in anatomical connectivity as tract definitions may need to be overly inclusive to extract appropriate streamlines from all subjects, or else, to avoid false positives, overly restrictive. Due to their subject specific nature and reliance on identifying bundle cores, manual ROI-based methods should not be affected by such inter-subject variability.

Changes in image acquisition and tractography method will also affect white matter delineation and extraction. Acquiring diffusion images with better spatial resolution and more gradient directions should, in theory, improve the output of all tract delineation methods via improvements in tractography performance. Generating whole-brain tractography with less noise and fewer errors should, for example, increase between-subject consistency, which will improve the implementation of groupwise fiber clustering algorithms. Improved image spatial resolution will also allow for more accurate and specific manual ROI placements, as well as for the tracing of much smaller white matter tracts. Atlas-based methods may derive the greatest benefit from improvements in image quality as in addition to better input tractography, this will allow for enhanced anatomical parcellation accuracy, for parcellations with smaller segmentation regions, and for better atlas to image registration. Additionally, because some atlas-based approaches are conditional on identifying streamlines with endpoints in gray matter ROIs, they are susceptible to failures of tractography to trace streamlines into regions of gray matter. While this concern is largely mitigated by the use of lenient stopping criteria that allow for tracing into gray matter (e.g., minimum FA of 0.15 and GFA of 0.1 (Truong et al., 2014)) and a tractography algorithm that performs well in areas of low anisotropy and high uncertainty, acquisition improvements that allow for greater accuracy at the gray-white boundary will be important for gray matter-centric methods.

The present study cannot provide direct evidence as to how the application of different tractography methodologies would affect results. We postulate however that the use of single rather than multi-tensor tractography would be equally disadvantageous to the three approaches utilized here. Correspondingly, the use of high angular resolution diffusion-weighted imaging (Tuch et al., 2002) based tractography, which has been shown to perform similarly to the algorithm we employed (Baumgartner et al., 2012; Fillard et al., 2011), would most likely yield results from which the same general trends with regards to spatial overlap, tract volume, FA, connectivity, and intra-method consistency would be observed. There are additionally white matter delineation approaches that are not streamline-based, such as probabilistic tractography methods that generate probability heat maps of fiber tract connectivity pathways between brain regions (Behrens et al., 2003; Jbabdi et al., 2007; Parker and Alexander, 2005; Parker et al., 2003), as well as tract segmentation methods that generate 3-dimensional geometric models of white matter structures (Jonasson et al., 2007; Nazem-Zadeh et al., 2011; Rousson et al., 2004; Zhukov, Leonid et al., 2003). Though outside the scope of this study, future studies should resolve whether white matter tract derived measures systematically differ across streamline-based, probabilistic, and segmentation approaches.

Given the non-trivial differences that can arise between conceptually distinct white matter delineation approaches, we offer some suggestions as to how to select an appropriate approach to employ, based on how the strengths and weaknesses of the approach relate to different scientific and clinical contexts.

4.2. Manual Region of Interest-Based Approaches

Past research has demonstrated that manually imposing two or more anatomically specific constraints on tractography input can greatly enhance the precision of white matter tract output (Catani et al., 2002; Huang et al., 2004). This subject-specific approach eliminates the need for registrations, data smoothing, and other data normalization steps that can introduce errors, and the precise drawing of ROIs on a case to case basis ensures that the streamlines being extracted belong to the white matter tract of interest. Consequently, these methods are regarded as the most anatomically accurate, and are used as a basis of comparison during new method validation. In the present study, results (tract volume, tract length, FA, STAPLE sensitivity) derived from the Manual method were generally intermediate between the other two methods, thus manual multiple-ROI methods do appear to serve as a reliable basis for comparison. Nonetheless, these methods have limitations. First, the more conservative use of ROIs in this method can result in false negatives. And second, they require considerable anatomical knowledge and a substantial amount of time, and the placement of ROIs drawn by experimenters can introduce errors or biases. However, a potential alternative to the conventional drawing of manual ROIs is to instead manually place seeds along a white matter tract of interest in order to generate seed-based tractography (e.g. Nazem-Zadeh et al., 2012); this is thought to reduce the time investment and is less dependent on the operator’s anatomical knowledge.

Given the elevated potential for anatomical precision, manual multiple-ROI methods are most apposite for subject-specific analyses (Lazar et al., 2006) and pre-surgical planning

(Berman, 2009; Chen et al., 2009; Golby et al., 2011; Pujol et al., 2015; Spena et al., 2010; Yogarajah et al., 2009), wherein delineating an exact representation of a tract is critical. Manual methods should also be utilized in studies involving individuals with deformed or atrophied brains (e.g., Rijken et al., 2015) in which anatomical atlases will not correctly register with diffusion images, and tractography fibers will not be consistently organized across subjects. Studies that include a smaller number of participants and aim to analyze a select number of white matter tracts may consider employing a manual multiple-ROI approach when feasible, particularly those interested in examining stem (core) white matter fibers and their corresponding anatomical features (e.g., tract volume, fiber length, or fiber dispersion (Sotiropoulos et al., 2012)). Previous studies have successfully utilized manual methods to identify abnormalities in white matter tracts in patient populations, for example in the AF in schizophrenia patients experiencing positive symptoms (Psomiades et al., 2016; Seitz et al., 2016), and in the cingulum bundle in individuals with bipolar disorder (Wang et al., 2008). Robust structure-function relationships have also been discovered for white matter tracts extracted with manual methods, for instance, the role of the AF in language-related functions (Maffei et al., 2015).

4.3. Groupwise Fiber Clustering Approaches

Groupwise fiber clustering methods enable the identification and analysis of homologous white matter structures across subjects. Three findings from the present study support that groupwise clustering approaches do extract the most consistent white matter tracts across subjects. First, Clustering consistently produced method-specific tract overlay maps with the greatest volumes (Table 2), signifying the most extensive overlap of tracts across different subjects. Although this may partially result from the fact that the majority of individual tracts produced by Clustering had greater volumes, this finding held for the MdLF, wherein individual Clustering tracts were smallest. Second, Clustering had the highest STAPLE sensitivity value, indicating high within-method spatial consistency. And third, the standard deviations for tract volume, tract length, and FA were lowest for Clustering.

Because groupwise fiber clustering approaches are initially entirely data driven, they avoid *a priori* anatomical constraints and operator-dependent ROI selection (Voineskos et al., 2009). As seen here, this allows for the inclusion of fibers that ROI-based methods may miss, decreasing the rate of false negatives, but potentially leading to overinclusive results as well. Similar to the lack of ROI use, the use of outlier removal (Guevara et al., 2011; O'Donnell et al., 2016) has both benefits and potential disadvantages. Removing streamlines with low resemblance to nearby clusters (i.e., streamlines that are irregularly shaped, unusual, or uncommon) from the final clustered output reduces the presence of false positives and tractography-associated noise. However, stringent outlier removal may additionally eliminate true positive fibers that deviate from the normal tract trajectory. In other words, seemingly abnormal streamlines may be real, and more aesthetically pleasing tracts may not always be the most anatomically accurate.

Overall, groupwise fiber clustering approaches have the potential to generate very clean bundles, and they facilitate comparisons of relatively uniformly delineated tracts across subjects. Yet, they may limit the ability to identify subtle group differences in structural

connectivity. These approaches may therefore be more apposite for studies aiming to minimize variability in tract architecture and volume in order to compare dMRI-derived scalar measures from tracts that are defined consistently across subjects or populations. Groupwise fiber clustering also offers an advanced and unique approach to systematically classifying the entirety of whole-brain tractography, which can be used in whole-brain analyses aimed at studying global white matter features. For example, (Zhang et al., 2017) were able to predict whether an individual had Autism Spectrum Disorder with 78% accuracy by extracting FA and mean diffusivity features from 800 fiber clusters generated from whole-brain tractography.

4.4. Gray and White Matter Atlas-Based Approaches

Atlas-based methods facilitate the extraction of a large number of white matter tracts with clear anatomical connectivity interpretations. The public availability of these methods reduces the need for researchers to have robust neuroanatomical knowledge, thereby increasing the accessibility of performing tract-based dMRI research. Furthermore, the same atlas and tract definitions can be utilized across different studies, increasing reproducibility and the feasibility of comparing results across datasets and across study populations.

Advantageously, automated atlas-based ROI selection eliminates experimenter bias in the drawing of ROIs, yet the ROIs that can be used are limited to those that exist in the atlas. Given that most atlases have relatively large ROIs (e.g. encompassing an entire gyrus), extracting smaller, more anatomically-localized tracts is not feasible unless additional manual segmentations are performed. Large ROIs can furthermore lead to an increased rate of false positives (seen in the present study, for example, between dorsal and ventral portions of the AF). This susceptibility to false positives is further compounded by the possibility of atlas-to- subject misregistrations, which can introduce spurious fibers or variability in tract output. Nevertheless, these methods often make it easy to modify tract extraction instructions when needed (e.g., by changing or constructing new queries in WMQL) so as to produce high quality output, and enhanced registrations can be achieved with the correct combination of registration method (rigid, affine and/or nonlinear) and MR image type (T1-weighted or diffusion images, FA maps and/or DTIs), or by using an anatomical atlas composed of multiple reference anatomies (Suarez et al., 2012).

Atlas-based methods may be generally favorable for the extraction of a large number of white matter tracts from large N datasets, as well as for conducting exploratory connectivity studies, given the ability to create new atlas-based tract extraction instructions. The WMQL method, for example, has been utilized to reveal cortico-striatal degeneration in 191 individuals with prodromal Huntington’s Disease (Shaffer et al., 2017), and to identify tract-based alterations that maximally discriminated individuals with 22q11.2 deletion syndrome from control subjects (Tylee et al., 2017). These methods may furthermore be beneficial for conducting multi-modal analyses that aim to relate dMRI structural connectivity findings to structural, functional, or positron emission tomography MRI findings, as the same anatomical atlas can be utilized across different imaging modalities to localize and relate findings.

5. Conclusions

This comparison of three representative post-tractography white matter tract delineation methods demonstrates that although distinct approaches are able to identify and extract relatively consistent forms of white matter tracts, important differences in tract architecture, connectivity, volume, and mean scalar measures can arise, even when methods rely on the same input tractography and neuroanatomical knowledge. The level of agreement between methods will be influenced by both the specific methodologies being employed, and by the white matter tracts being studied. Careful consideration should therefore be given to decisions regarding which white matter delineation approach to utilize in both research and clinical settings. The future of white matter tract delineation should likely involve methodologies that combine multiple approaches, for example methods that apply multiple manual ROI segmentations to generate multi-reference template atlases (Suarez et al., 2012), or methods that combine initial fiber clustering with subsequent atlas-based white matter classification (Ros et al., 2013; Wassermann et al., 2010; Xu et al., 2013).

Given the present findings, additional future work should be undertaken to investigate the impact that varied white matter delineation approaches have on detecting differences in diffusion measures between groups in typical case-control studies. Elucidating how different approaches ultimately affect the detection of significant results is crucial to study replication and for understanding “case”-associated abnormalities.

6. Limitations

One limitation of this study is that three representative methods were used to exemplify, in turn, a manual multiple ROI-based, an atlas-based, and a groupwise fiber clustering approach, yet not all implementations of these three major approaches will be identical. Nonetheless, methods that can be classified methodologically to one of these three conceptual categories will likely be comparable to the representative methods used in this report. A second limitation arises from the lack of a validated ground truth with regard to white matter fiber tract anatomy. Because the true anatomy of white matter can only be confirmed ex-vivo and post-mortem, we cannot unequivocally determine which method best approximated the true anatomical properties of the white matter tracts under study. Lastly, this analysis was performed using data from 10 healthy participants, thus results may differ in studies of patient populations, especially when pathology interferes with tract anatomy.

Acknowledgements

This work was supported by the National Institutes of Health (T32MH016259-35 to AEL, R01AG042512 to MK and NM, R01MH102377 to MK, K24MH110807 to MK, R01MH112748 to SB, MK and NM, U01CA199459 to LO and FZ, P41EB015902 to LO and FZ, R01MH074794 to LO and FZ, and P41EB015898 to LO and FZ) and a European Research Council Starting Grant (ERC-StG NeuroLang to DW).

Abbreviations:

(ANTS)	Advanced Normalization Tools
(AF)	Arcuate Fasciculus

(dMRI)	Diffusion Magnetic Resonance Imaging
(DTI)	Diffusion Tensor Image
(TE)	Echo Time
(FA)	Fractional Anisotropy
(ICC)	Intraclass Correlation Coefficient
(TI)	Inversion Time
(JC)	Jaccard Coefficient
(MR)	Magnetic Resonance
(MdLF)	Middle Longitudinal Fasciculus
(ROI)	Region of Interest
(ROIs)	Regions of Interest
(TR)	Repetition Time
(STAPLE)	Simultaneous Truth and Performance Level Estimation
(S.D.)	Standard Deviation
(UF)	Uncinate Fasciculus
(UKF)	Unscented Kalman Filter
(WMQL)	White Matter Query Language

References

Avants BB, Epstein CL, Grossman M, Gee JC, 2008 Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12, 26–41. 10.1016/j.media.2007.06.004 [PubMed: 17659998]

Baumgartner C, Michailovich O, Levitt J, Pasternak O, Bouix S, Westin CF, Rathi Y, 2012 A unified tractography framework for comparing diffusion models on clinical scans, in: Computational Diffusion MRI Workshop. Presented at the MICCAI, Nice, pp27–32.

Beaulieu C, 2002 The basis of anisotropic water diffusion in the nervous system - a technical review. NMR Biomed 15, 435–455. 10.1002/nbm.782 [PubMed: 12489094]

Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MFS, Woolrich MW, 2007 Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34, 144–155. 10.1016/j.neuroimage.2006.09.018 [PubMed: 17070705]

Behrens TEJ, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM, 2003 Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 50, 1077–1088. 10.1002/mrm.10609 [PubMed: 14587019]

Berman J, 2009 Diffusion MR Tractography As a Tool for Surgical Planning. Magnetic Resonance Imaging Clinics of North America, Clinical Applications of MR Diffusion and Perfusion Imaging 17, 205–214. 10.1016/j.mric.2009.02.002

Brun A, Knutsson H, Park H-J, Shenton ME, Westin C-F, 2004 Clustering Fiber Traces Using Normalized Cuts, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004, Lecture Notes in Computer Science Presented at the International Conference on Medical

Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, pp. 368–375. 10.1007/978-3-540-30135-6_45

Bürgel U, Mädler B, Honey CR, Thron A, Gilsbach J, Coenen VA, 2009 Fiber Tracking with Distinct Software Tools Results in a Clear Diversity in Anatomical Fiber Tract Portrayal. Cen Eur Neurosurg 70, 27–35. 10.1055/s-0028-1087212

Catani M, Howard RJ, Pajevic S, Jones DK, 2002 Virtual in vivo interactive dissection of white matter fasciculi in the human brain. Neuroimage 17, 77–94. 10.1006/nimg.2002.1136 [PubMed: 12482069]

Catani M, Jones DK, Ffytche DH, 2005 Perisylvian language networks of the human brain. Annals of Neurology 57, 8–16. 10.1002/ana.20319 [PubMed: 15597383]

Catani M, Thiebaut de Schotten M, 2008a A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 44, 1105–1132. 10.1016/j.cortex.2008.05.004 [PubMed: 18619589]

Catani M, Thiebaut de Schotten M, 2008b A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 44, 1105–1132. 10.1016/j.cortex.2008.05.004 [PubMed: 18619589]

Chen X, Weigel D, Ganslandt O, Buchfelder M, Nimsky C, 2009 Prediction of visual field deficits by diffusion tensor imaging in temporal lobe epilepsy surgery. Neuroimage 45, 286–297. 10.1016/j.neuroimage.2008.11.038 [PubMed: 19135156]

Chua TC, Wen W, Slavin MJ, Sachdev PS, 2008 Diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease: a review. Curr. Opin. Neurol. 21, 83–92. 10.1097/WCO.0b013e3282f4594b [PubMed: 18180656]

Conturo TE, Lori NF, Cull TS, Akbudak E, Snyder AZ, Shimony JS, McKinstry RC, Burton H, Raichle ME, 1999 Tracking neuronal fiber pathways in the living human brain. Proc Natl Acad Sci U S A 96, 10422–10427. [PubMed: 10468624]

Dauguet J, Peled S, Berezovskii V, Delzescaux T, Warfield SK, Born R, Westin C-F, 2007. Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. Neuroimage 37, 530–538. 10.1016/j.neuroimage.2007.04.067 [PubMed: 17604650]

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ, 2006 An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980. 10.1016/j.neuroimage.2006.01.021 [PubMed: 16530430]

Ding Z, Gore JC, Anderson AW, 2003 Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. Magn Reson Med 49, 716–721. 10.1002/mrm.10415 [PubMed: 12652543]

Donahue CJ, Sotiropoulos SN, Jbabdi S, Hernandez-Fernandez M, Behrens TE, Dyrby TB, Coalson T, Kennedy H, Knoblauch K, Van Essen DC, Glasser MF, 2016 Using Diffusion Tractography to Predict Cortical Connection Strength and Distance: A Quantitative Comparison with Tracers in the Monkey. J. Neurosci. 36, 6758–6770. 10.1523/JNEUROSCI.0493-16.2016 [PubMed: 27335406]

Dyrby TB, Søgaard LV, Parker GJ, Alexander DC, Lind NM, Baaré WFC, Hay-Schmidt A, Eriksen N, Pakkenberg B, Paulson OB, Jelsing J, 2007 Validation of in vitro probabilistic tractography. Neuroimage 37, 1267–1277. 10.1016/j.neuroimage.2007.06.022 [PubMed: 17706434]

Essayed WI, zhang F, Unadkat P, Cosgrove GR, Golby AJ, O'Donnell LJ, 2017 White matter tractography for neurosurgical planning: A topography-based review of the current state of the art. Neuroimage: Clinical 15, 659–672. 10.1016/j.nicl.2017.06.011 [PubMed: 28664037]

Fillard P, Descoteaux M, Goh A, Gouttard S, Jeurissen B, Malcolm J, Ramirez-Manzanares A, Reisert M, Sakaie K, Tensaouti F, Yo T, Mangin J-F, Poupon C, 2011 Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. Neuroimage 56, 220–234. 10.1016/j.neuroimage.2011.01.032 [PubMed: 21256221]

Fischl B, 2012 FreeSurfer. Neuroimage, 20 YEARS OF fMRI 62, 774–781. 10.1016/j.neuroimage.2012.01.021

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM, 2002 Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355. 10.1016/S0896-6273(02)00569-X [PubMed: 11832223]

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM, 2004 Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22. https://doi-org.ezproxyhost.library.tmc.edu/10.1093/cercor/bhg087 [PubMed: 14654453]

Fritzsche KH, Laun FB, Meinzer H-P, Stieltjes B, 2010 Opportunities and pitfalls in the quantification of fiber integrity: what can we gain from Q-ball imaging? Neuroimage 51, 242–251. 10.1016/j.neuroimage.2010.02.007 [PubMed: 20149879]

Garyfallidis E, Brett M, Correia MM, Williams GB, Nimmo-Smith I, 2012 QuickBundles, a Method for Tractography Simplification. Front Neurosci 6 10.3389/fnins.2012.00175

Golby AJ, Kindlmann G, Norton I, Yarmarkovich A, Pieper S, Kikinis R, 2011 Interactive Diffusion Tensor Tractography Visualization for Neurosurgical Planning. Neurosurgery 68, 496–505. 10.1227/NEU.0b013e3182061ebb [PubMed: 21135713]

Gong G, He Y, Concha L, Lebel C, Gross DW, Evans AC, Beaulieu C, 2009 Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. Cereb. Cortex 19, 524–536. 10.1093/cercor/bhn102 [PubMed: 18567609]

Guevara P, Duclap D, Poupon C, Marrakchi-Kacem L, Fillard P, Le Bihan D, Leboyer M, Houenou J, Mangin J-F, 2012 Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. Neuroimage 61, 1083–1099. 10.1016/j.neuroimage.2012.02.071 [PubMed: 22414992]

Guevara P, Poupon C, Riviere D, Cointepas Y, Descoteaux M, Thirion B, Mangin J-F, 2011 Robust clustering of massive tractography datasets. Neuroimage 54, 1975–1993. 10.1016/j.neuroimage.2010.10.028 [PubMed: 20965259]

Horsfield MA, Jones DK, 2002 Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases - a review. NMR Biomed 15, 570–577. 10.1002/nbm.787 [PubMed: 12489103]

Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM, Mori S, 2008 Tract Probability Maps in Stereotaxic Spaces: Analyses of White Matter Anatomy and Tract-Specific Quantification. Neuroimage 39, 336–347. 10.1016/j.neuroimage.2007.07.053 [PubMed: 17931890]

Huang H, Zhang J, van Zijl PCM, Mori S, 2004 Analysis of noise effects on DTI-based tractography using the brute-force and multi-ROI approach. Magn Reson Med 52, 559–565. 10.1002/mrm.20147 [PubMed: 15334575]

Jbabdi S, Woolrich MW, Andersson JLR, Behrens TEJ, 2007 A Bayesian framework for global tractography. Neuroimage 37, 116–129. 10.1016/j.neuroimage.2007.04.039 [PubMed: 17543543]

Jolles D, Wassermann D, Chokhani R, Richardson J, Tenison C, Bammer R, Fuchs L, Supekar K, Menon V, 2016 Plasticity of left perisylvian white-matter tracts is associated with individual differences in math learning. Brain Struct Funct 221, 1337–1351. 10.1007/s00429-014-0975-6 [PubMed: 25604464]

Jonasson L, Bresson X, Thiran JP, Wedeen VJ, Hagmann P, 2007 Representing Diffusion MRI in 5-D Simplifies Regularization and Segmentation of White Matter Tracts. IEEE Transactions on Medical Imaging 26, 1547–1554. 10.1109/TMI.2007.899168 [PubMed: 18041269]

Jones DK, 2008 Studying connections in the living human brain with diffusion MRI. Cortex 44, 936–952. 10.1016/j.cortex.2008.05.002 [PubMed: 18635164]

Jones DK, Simmons A, Williams SC, Horsfield MA, 1999 Non-invasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI. Magn Reson Med 42, 37–41. https://doi-org.ezproxyhost.library.tmc.edu/10.1002/(SICI)1522-2594(199907)42:1<37::AID-MRM7>3.0.CO;2-O [PubMed: 10398948]

Khalsa S, Mayhew SD, Chechlacz M, Bagary M, Bagshaw AP, 2014 The structural and functional connectivity of the posterior cingulate cortex: comparison between deterministic and probabilistic tractography for the investigation of structure-function relationships. Neuroimage 102 Pt 1, 118–127. 10.1016/j.neuroimage.2013.12.022 [PubMed: 24365673]

Knösche TR, Anwander A, Liptrot M, Dyrby TB, 2015 Validation of tractography: Comparison with manganese tracing. Hum Brain Mapp 36, 4116–4134. 10.1002/hbm.22902 [PubMed: 26178765]

Kubicki M, McCarley R, Westin C-F, Park H-J, Maier S, Kikinis R, Jolesz FA, Shenton ME, 2007 A review of diffusion tensor imaging studies in schizophrenia. J Psychiatr Res 41, 15–30. 10.1016/j.jpsychires.2005.05.005 [PubMed: 16023676]

Landis JR, Koch GG, 1977 The measurement of observer agreement for categorical data. Biometrics 33, 159–174. https://10.1148/radiol.2282011860 [PubMed: 843571]

Lawes INC, Barrick TR, Murugam V, Spierings N, Evans DR, Song M, Clark CA, 2008 Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection. Neuroimage 39, 62–79. 10.1016/j.neuroimage.2007.06.041 [PubMed: 17919935]

Lazar M, Alexander AL, Thottakara PJ, Badie B, Field AS, 2006 White Matter Reorganization After Surgical Resection of Brain Tumors and Vascular Malformations. American Journal of Neuroradiology 27, 1258–1271. [PubMed: 16775277]

Lazar M, Weinstein DM, Tsuruda JS, Hasan KM, Arfanakis K, Meyerand ME, Badie B, Rowley HA, Haughton V, Field A, Alexander AL, 2003 White matter tractography using diffusion tensor deflection. Hum Brain Mapp 18, 306–321. 10.1002/hbm.10102 [PubMed: 12632468]

Lee S-H, Kubicki M, Asami T, Seidman LJ, Goldstein JM, Mesholam-Gately RI, McCarley RW, Shenton ME, 2013 Extensive white matter abnormalities in patients with first-episode schizophrenia: a Diffusion Tensor Iimaging (DTI) study. Schizophr. Res. 143, 231–238. 10.1016/j.schres.2012.11.029 [PubMed: 23290268]

Maffei C, Soria G, Prats-Galino A, Catani M, 2015 Imaging white-matter pathways of the auditory system with diffusion imaging tractography. Handbook of Clinical Neurology, The Human Auditory System 129, 277–288. 10.1016/B978-0-444-62630-1.00016-0

Maier-Hein KH, Neher PF, Houde J-C, Côté M-A, Garyfallidis E, Zhong J, Chamberland M, Yeh F-C, Lin Y-C, Ji Q, Reddick WE, Glass JO, Chen DQ, Feng Y, Gao C, Wu Y, Ma J, Renjie H, Li Q, Westin C-F, Deslauriers-Gauthier S, González JOO, Paquette M, St-Jean S, Girard G, Rheault F, Sidhu J, Tax CMW, Guo F, Mesri HY, Dávid S, Froeling M, Heemskerk AM, Leemans A, Boré A, Pinsard B, Bedetti C, Desrosiers M, Brambati S, Doyon J, Sarica A, Vasta R, Cerasa A, Quattrone A, Yeatman J, Khan AR, Hodges W, Alexander S, Romascano D, Barakovic M, Auría A, Esteban O, Lemkaddem A, Thiran J-P, Cetingul HE, Odry BL, Mailhe B, Nadar MS, Pizzagalli F, Prasad G, Villalon-Reina JE, Galvis J, Thompson PM, Requejo FDS, Laguna PL, Lacerda LM, Barrett R, Dell'Acqua F, Catani M, Petit L, Caruyer E, Daducci A, Dyrby TB, Holland-Letz T, Hilgetag CC, Stieltjes B, Descoteaux M, 2017 The challenge of mapping the human connectome based on diffusion tractography. Nature Communications 8, 1349 10.1038/s41467-017-01285-x

Makris N, Kennedy DN, Mclnerney S, Sorensen AG, Wang R, Caviness VS, Pandya DN, 2005 Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. Cereb. Cortex 15, 854–869. 10.1093/cercor/bhh186 [PubMed: 15590909]

Makris N, Pandya DN, 2009 The extreme capsule in humans and rethinking of the language circuitry. Brain Struct Funct 213, 343–358. 10.1007/s00429-008-0199-8 [PubMed: 19104833]

Makris N, Pandya DN, Normandin JJ, Papadimitriou GM, Rauch SL, Caviness VS, Kennedy DN, 2002 Quantitative DT-MRI Investigations of the Human Cingulum Bundle. CNS Spectrums 7, 522–528. 10.1017/S1092852900018071

Makris N, Papadimitriou GM, Kaiser JR, Sorg S, Kennedy DN, Pandya DN, 2009 Delineation of the Middle Longitudinal Fascicle in Humans: A Quantitative, In Vivo, DT-MRI Study. Cereb Cortex 19, 777–785. 10.1093/cercor/bhn124 [PubMed: 18669591]

Makris N, Papadimitriou GM, Sorg S, Kennedy DN, Caviness VS, Pandya DN, 2007 The occipitofrontal fascicle in humans: a quantitative, in vivo, DT-MRI study. Neuroimage 37, 1100–1111. 10.1016/j.neuroimage.2007.05.042 [PubMed: 17681797]

Makris N, Preti MG, Asami T, Pelavin P, Campbell B, Papadimitriou GM, Kaiser J, Baselli G, Westin CF, Shenton ME, Kubicki M, 2013a Human middle longitudinal fascicle: variations in patterns of anatomical connections. Brain Struct Funct 218, 951–968. 10.1007/s00429-012-0441-2 [PubMed: 22782432]

Makris N, Preti MG, Wassermann D, Rathi Y, Papadimitriou GM, Yergatian C, Dickerson BC, Shenton ME, Kubicki M, 2013b Human middle longitudinal fascicle: segregation and behavioral-clinical implications of two distinct fiber connections linking temporal pole and superior temporal

gyrus with the angular gyrus or superior parietal lobule using multi-tensor tractography. Brain Imaging Behav 7, 335–352. 10.1007/s11682-013-9235-2 [PubMed: 23686576]

Makris N, Worth AJ, Sorensen AG, Papadimitriou GM, Wu O, Reese TG, Wedeen VJ, Davis TL, Stakes JW, Caviness VS, Kaplan E, Rosen BR, Pandya DN, Kennedy DN, 1997 Morphometry of in vivo human white matter association pathways with diffusion-weighted magnetic resonance imaging. Ann. Neurol. 42, 951–962. 10.1002/ana.410420617 [PubMed: 9403488]

Malcolm JG, Shenton ME, Rathi Y, 2010 Filtered multi-tensor tractography. IEEE Transactions on Medical Imaging 29, 1664–1675. 10.1109/TMI.2010.2048121.Filtered [PubMed: 20805043]

Malcolm JG, Shenton ME, Rathi Y, 2009 Neural tractography using an unscented Kalman filter. Inf Process Med Imaging 21, 126–138. [PubMed: 19694258]

Malykhin N, Concha L, Seres P, Beaulieu C, Coupland NJ, 2008 Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. Psychiatry Res 164, 132–142. 10.1016/j.pscychresns.2007.11.007 [PubMed: 18945599]

Mori S, Crain BJ, Chacko VP, van Zijl PC, 1999 Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann. Neurol. 45, 265–269. [PubMed: 9989633]

Mori S, Kaufmann WE, Davatzikos C, Stieltjes B, Amodei L, Fredericksen K, Pearlson GD, Melhem ER, Solaiyappan M, Raymond GV, Moser HW, van Zijl PCM, 2002 Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. Magn Reson Med 47, 215–223. https://doi-org.ezproxyhost.library.tmc.edu/10.1002/mrm.10074 [PubMed: 11810663]

Mori S, C M van Zijl P, Oishi K, Faria AV, 2011 MRI Atlas of Human White Matter, 2nd ed Academic Press.

Nazem-Zadeh M-R, Chapman CH, Lawrence TL, Tsien CI, Cao Y, 2012 Radiation therapy effects on white matter fiber tracts of the limbic circuit. Med Phys 39, 5603–5613. 10.1118/1.4745560 [PubMed: 22957626]

Nazem-Zadeh M-R, Davoodi-Bojd E, Soltanian-Zadeh H, 2011 Atlas-based fiber bundle segmentation using principal diffusion directions and spherical harmonic coefficients. Neuroimage 54 Suppl 1, S146–164. 10.1016/j.neuroimage.2010.09.035 [PubMed: 20869453]

Norton I, Essayed WI, Zhang F, Pujol S, Yarmarkovich A, Golby AJ, Kindlmann G, Wassermann D, Estepar RSJ, Rathi Y, Pieper S, Kikinis R, Johnson HJ, Westin C-F, O'Donnell LJ, 2017 SlicerDMRI: Open Source Diffusion MRI Software for Brain Cancer Research. Cancer Res. 77, e101–e103. 10.1158/0008-5472.CAN-17-0332 [PubMed: 29092950]

O'Donnell LJ, Golby AJ, Westin C-F, 2013 Fiber clustering versus the parcellation-based connectome. Neuroimage 80, 283–289. 10.1016/j.neuroimage.2013.04.066 [PubMed: 23631987]

O'Donnell LJ, Suter Y, Rigolo L, Kahali P, Zhang F, Norton I, Albi A, Olubiyi O, Meola A, Essayed WI, Unadkat P, Ciris PA, Wells WM, Rathi Y, Westin C-F, Golby AJ, 2016 Automated white matter fiber tract identification in patients with brain tumors. Neuroimage Clin 13, 138–153. 10.1016/j.nicl.2016.11.023 [PubMed: 27981029]

O'Donnell LJ, Wells WM, Golby AJ, Westin C-F, 2012 Unbiased groupwise registration of white matter tractography. Med Image Comput Comput Assist Interv 15, 123–130. [PubMed: 23286122]

O'Donnell LJ, Westin CF, 2007 Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas. IEEE Transactions on Medical Imaging 26, 1562–1575. 10.1109/TMI.2007.906785 [PubMed: 18041271]

Oishi K, Faria A, Jiang H, Li X, Akhter K, Zhang J, Hsu JT, Miller MI, van Zijl PCM, Albert M, Lyketsos CG, Woods R, Toga AW, Pike GB, Rosa-Neto P, Evans A, Mazziotta J, Mori S, 2009 Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. Neuroimage 46, 486–499. [PubMed: 19385016]

Parker GJ., Alexander DC, 2005 Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. Philos Trans R Soc Lond B Biol Sci 360, 893–902. 10.1098/rstb.2005.1639 [PubMed: 16087434]

Parker GJM, Haroon HA, Wheeler-Kingshott CAM, 2003 A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. J Magn Reson Imaging 18, 242–254. 10.1002/jmri.10350 [PubMed: 12884338]

Poupon C, Rieul B, Kezele I, Perrin M, Poupon F, Mangin J-F, 2008 New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models. Magn Reson Med 60, 1276–1283. 10.1002/mrm.21789 [PubMed: 19030160]

Psomiades M, Fonteneau C, Mondino M, Luck D, Haesebaert F, Suaud-Chagny M-F, Brunelin J, 2016 Integrity of the arcuate fasciculus in patients with schizophrenia with auditory verbal hallucinations: A DTI-tractography study. NeuroImage: Clinical 12, 970–975. 10.1016/j.nicl.2016.04.013 [PubMed: 27995063]

Pujol S, Wells W, Pierpaoli C, Brun C, Gee J, Cheng G, Vemuri B, Commowick O, Prima S, Stamm A, Goubran M, Khan A, Peters T, Neher P, Maier-Hein KH, Shi Y, Tristan-Vega A, Veni G, Whitaker R, Styner M, Westin C-F, Gouttard S, Norton I, Chauvin L, Mamata H, Gerig G, Nabavi A, Golby A, Kikinis R, 2015 The DTI Challenge: Toward Standardized Evaluation of Diffusion Tensor Imaging Tractography for Neurosurgery. J Neuroimaging 25, 875–882. https://doi.org/10.nn/jon.12283 [PubMed: 26259925]

Pullens P, Roebroeck A, Goebel R, 2010 Ground truth hardware phantoms for validation of diffusion-weighted MRI applications. J Magn Reson Imaging 32, 482–488. 10.1002/jmri.22243 [PubMed: 20677281]

Rijken BFM, Leemans A, Lucas Y, Montfort K van, Mathijssen IMJ, Lequin MH, 2015 Diffusion Tensor Imaging and Fiber Tractography in Children with Craniosynostosis Syndromes. American Journal of Neuroradiology 36, 1558–1564. 10.3174/ajnr.A4301 [PubMed: 25953762]

Rojkova K, Volle E, Urbanski M, Humbert F, Dell'Acqua F, Thiebaut de Schotten M, 2016 Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study. Brain Struct Funct 221, 1751–1766. 10.1007/s00429-015-1001-3 [PubMed: 25682261]

Ros C, Güllmar D, Stenzel M, Mentzel H-J, Reichenbach JR, 2013 Atlas-Guided Cluster Analysis of Large Tractography Datasets. PLOS ONE 8, e83847 10.1371/journal.pone.0083847 [PubMed: 24386292]

Rousson M, Lenglet C, Deriche R, 2004 Level Set and Region Based Surface Propagation for Diffusion Tensor MRI Segmentation, in: Sonka M, Kakadiaris IA, Kybic J (Eds.), Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis. Springer Berlin Heidelberg, pp. 123–134.

Seehaus AK, Roebroeck A, Chiry O, Kim D-S, Ronen I, Bratzke H, Goebel R, Galuske RAW, 2013 Histological validation of DW-MRI tractography in human postmortem tissue. Cereb. Cortex 23, 442–450. 10.1093/cercor/bhs036 [PubMed: 22345356]

Seitz J, Zuo JX, Lyall AE, Makris N, Kikinis Z, Bouix S, Pasternak O, Fredman E, Duskin J, Goldstein JM, Petryshen TL, Mesholam-Gately RI, Wojcik J, McCarley RW, Seidman LJ, Shenton ME, Koerte IK, Kubicki M, 2016 Tractography Analysis of 5 White Matter Bundles and Their Clinical and Cognitive Correlates in Early-Course Schizophrenia. Schizophr Bull 42, 762–771. 10.1093/schbul/sbv171 [PubMed: 27009248]

Shaffer JJ, Ghayoor A, Long JD, Kim RE-Y, Lourens S, O'Donnell LJ, Westin C-F, Rathi Y, Magnotta V, Paulsen JS, Johnson HJ, 2017 Longitudinal Diffusion Changes in Prodromal and Early HD: Evidence of White-Matter Tract Deterioration. Hum Brain Mapp 38, 1460–1477. 10.1002/hbm.23465 [PubMed: 28045213]

Shenton ME, Hamoda HM, Schneiderman JS, Bouix S, Pasternak O, Rathi Y, Vu M-A, Purohit MP, Helmer K, Koerte I, Lin AP, Westin C-F, Kikinis R, Kubicki M, Stern RA, Zafonte R, 2012 A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. Brain Imaging Behav 6, 137–192. 10.1007/s11682-012-9156-5 [PubMed: 22438191]

Sherbondy A, Akers D, Mackenzie R, Dougherty R, Wandell B, 2005 Exploring connectivity of the brain's white matter with dynamic queries. IEEE Transactions on Visualization and Computer Graphics 11, 419–430. 10.1109/TVCG.2005.59 [PubMed: 16138552]

Sotiropoulos SN, Behrens TEJ, Jbabdi S, 2012 Ball and rackets: Inferring fiber fanning from diffusion-weighted MRI. NeuroImage 60, 1412–1425. 10.1016/j.neuroimage.2012.01.056 [PubMed: 22270351]

Spena G, Nava A, Cassini F, Pepoli A, Bruno M, D'Agata F, Cauda F, Sacco K, Duca S, Barletta L, Versari P, 2010 Preoperative and intraoperative brain mapping for the resection of eloquent-area tumors. A prospective analysis of methodology, correlation, and usefulness based on clinical

outcomes. Acta Neurochir (Wien) 152, 1835–1846. 10.1007/s00701-010-0764-9 [PubMed: 20730457]

Suarez RO, Commowick O, Prabhu SP, Warfield SK, 2012 Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. Neuroimage 59, 3690–3700. 10.1016/j.neuroimage.2011.11.043 [PubMed: 22155046]

Sullivan EV, Pfefferbaum A, 2006 Diffusion tensor imaging and aging. Neurosci Biobehav Rev 30, 749–761. 10.1016/j.neubiorev.2006.06.002 [PubMed: 16887187]

Tensaouti F, Lahlou I, Clarisse P, Lotterie JA, Berry I, 2011 Quantitative and reproducibility study of four tractography algorithms used in clinical routine. J Magn Reson Imaging 34, 165–172. 10.1002/jmri.22584 [PubMed: 21618329]

Tournier J-Donald, Calamante Fernando, Connelly Alan, 2012 MRtrix: Diffusion tractography in crossing fiber regions. International Journal of Imaging Systems and Technology 22, 53–66. 10.1002/ima.22005

Truong T-K, Guidon A, Song AW, 2014 Cortical Depth Dependence of the Diffusion Anisotropy in the Human Cortical Gray Matter In Vivo. PLoS One 9 10.1371/journal.pone.0091424

Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ, 2002 High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn Reson Med 48, 577–582. 10.1002/mrm.10268 [PubMed: 12353272]

Tylee DS, Kikinis Z, Quinn TP, Antshel KM, Fremont W, Tahir MA, Zhu A, Gong X, Glatt SJ, Coman IL, Shenton ME, Kates WR, Makris N, 2017 Machine-learning classification of 22q11.2 deletion syndrome: A diffusion tensor imaging study. NeuroImage: Clinical 15, 832–842. 10.1016/j.nicl.2017.04.029 [PubMed: 28761808]

Varentsova A, Zhang S, Arfanakis K, 2014 Development of a high angular resolution diffusion imaging human brain template. Neuroimage 91, 177–186. 10.1016/j.neuroimage.2014.01.009 [PubMed: 24440528]

Voineskos AN, O'Donnell LJ, Lobaugh NJ, Markant D, Ameis SH, Niethammer M, Mulsant BH, Pollock BG, Kennedy JL, Westin CF, Shenton ME, 2009 Quantitative examination of a novel clustering method using magnetic resonance diffusion tensor tractography. Neuroimage 45, 370–376. 10.1016/j.neuroimage.2008.12.028 [PubMed: 19159690]

Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, Hua K, Zhang J, Jiang H, Dubey P, Blitz A, van Zijl P, Mori S, 2007 Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage 36, 630–644. 10.1016/j.neuroimage.2007.02.049 [PubMed: 17481925]

Wakana S, Jiang H, Nagae-Poetscher LM, van Zijl PCM, Mori S, 2004 Fiber tract-based atlas of human white matter anatomy. Radiology 230, 77–87. 10.1148/radiol.2301021640 [PubMed: 14645885]

Wang F, Jackowski M, Kalmar JH, Chepenik LG, Tie K, Qiu M, Gong G, Pittman BP, Jones MM, Shah MP, Spencer L, Papademetris X, Constable RT, Blumberg HP, 2008 Abnormal anterior cingulum integrity in bipolar disorder determined through diffusion tensor imaging. Br J Psychiatry 193, 126–129. 10.1192/bjp.bp.107.048793 [PubMed: 18669996]

Wang Y, Fernández-Miranda JC, Verstynen T, Pathak S, Schneider W, Yeh F-C, 2013 Rethinking the role of the middle longitudinal fascicle in language and auditory pathways. Cereb. Cortex 23, 2347–2356. 10.1093/cercor/bhs225 [PubMed: 22875865]

Warfield SK, Zou KH, Wells WM, 2004 Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23, 903–921. 10.1109/TMI.2004.828354 [PubMed: 15250643]

Wassermann D, Bloy L, Kanterakis E, Verma R, Deriche R, 2010. Unsupervised White Matter Fiber Clustering and Tract Probability Map Generation: Applications of a Gaussian Process framework for white matter fibers. Neuroimage 51, 228–241. 10.1016/j.neuroimage.2010.01.004 [PubMed: 20079439]

Wassermann D, Makris N, Rathi Y, Shenton M, Kikinis R, Kubicki M, Westin C-F, 2016 The white matter query language: a novel approach for describing human white matter anatomy. Brain Struct Funct 221, 4705–4721. 10.1007/s00429-015-1179-4 [PubMed: 26754839]

Xu Q, Anderson AW, Gore JC, Ding Z, 2013 Gray matter parcellation constrained full brain fiber bundling with diffusion tensor imaging. Med Phys 40, 072301 10.1118/1.4811155 [PubMed: 23822449]

Yendiki A, Panneck P, Srinivasan P, Stevens A, Zöllei L, Augustinack J, Wang R, Salat D, Ehrlich S, Behrens T, Jbabdi S, Gollub R, Fischl B, 2011 Automated Probabilistic Reconstruction of White-Matter Pathways in Health and Disease Using an Atlas of the Underlying Anatomy. Front Neuroinform 5 10.3389/fninf.20n.00023

Yogarajah M, Focke NK, Bonelli S, Cercignani M, Acheson J, Parker GJM, Alexander DC, McEvoy AW, Symms MR, Koepp MJ, Duncan JS, 2009 Defining Meyer's loop–temporal lobe resections, visual field deficits and diffusion tensor tractography. Brain 132, 1656–1668. 10.1093/brain/awp114 [PubMed: 19460796]

Yoo SW, Guevara P, Jeong Y, Yoo K, Shin JS, Mangin J-F, Seong J-K, 2015 An Example-Based Multi-Atlas Approach to Automatic Labeling of White Matter Tracts. PLoS ONE 10, e0133337 10.1371/journal.pone.0133337 [PubMed: 26225419]

Zhang F, Savadjiev P, Cai W, Song Y, Rathi Y, Tunç B, Parker D, Kapur T, Schultz RT, Makris N, Verma R, O'Donnell LJ, 2017 Whole brain white matter connectivity analysis using machine learning: An application to autism. Neuroimage. 10.1016/j.neuroimage.2017.10.029

Zhang H, Awatea SP, Das SR, Woo JH, Melhem ER, Gee JC, Yushkevich PA, 2010 A tract-specific framework for white matter morphometry combining macroscopic and microscopic tract features. Med Image Anal 14, 666–673. 10.1016/j.media.2010.05.002 [PubMed: 20547469]

Zhang W, Olivi A, Hertig SJ, van Zijl P, Mori S, 2008 Automated fiber tracking of human brain white matter using diffusion tensor imaging. Neuroimage 42, 771–777. 10.1016/j.neuroimage.2008.04.241 [PubMed: 18554930]

Zhukov Leonid, Museth Ken, Breen David E., Whitaker Ross T., Barr Alan H., 2003 Level set modeling and segmentation of diffusion tensor magnetic resonance imaging brain data. Journal of Electronic Imaging 12 10.1117/1.1527628

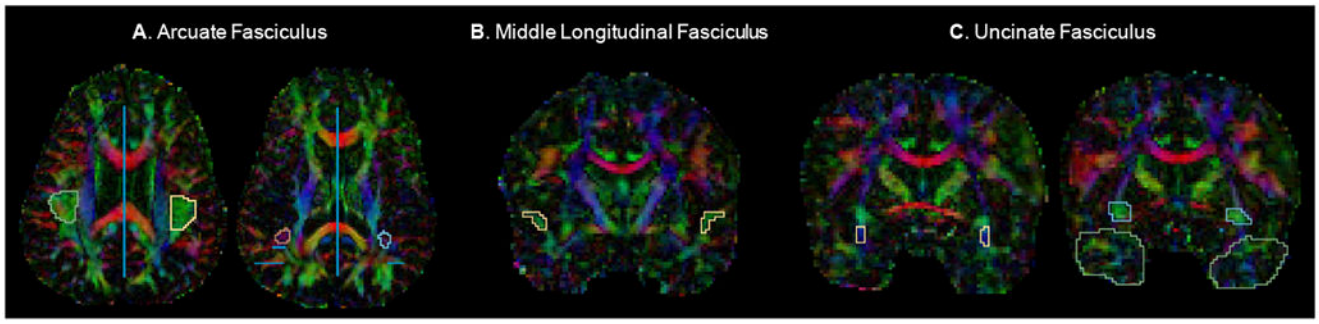

Figure 1.

Example regions of interest (ROIs) drawn for manual delineation of the A. arcuate fasciculus (AF), B. middle longitudinal fasciculus (MdLF), and C. uncinate fasciculus (UF) displayed on diffusion tensor color maps. The 2 circular inclusion ROIs can be seen for the right and left AF. Blue lines represent AF exclusion ROIs. One of four inclusion ROIs can be seen for both the right and left MdLF. All 3 inclusion ROIs can be seen for both the right and left UF. Exclusion ROIs are not visible in these slices for the MdLF and the UF.

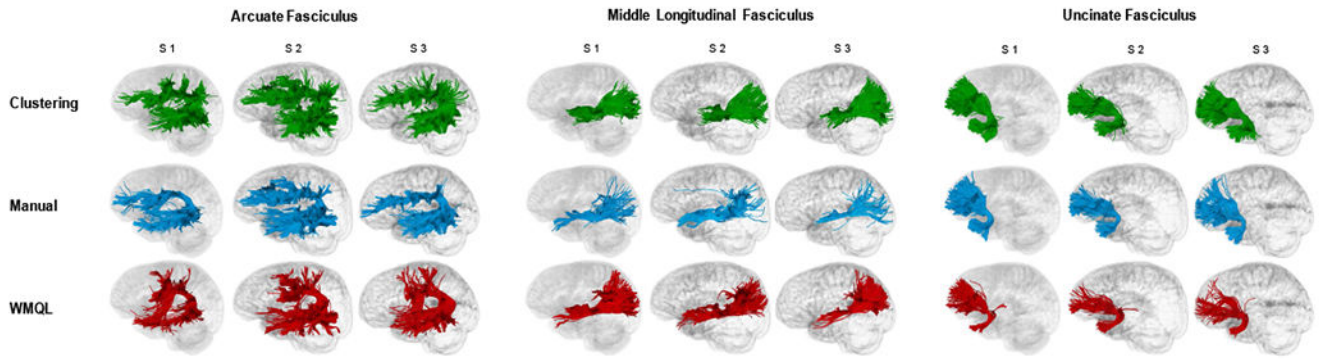

Figure 2.

Individual subjects' white matter tracts displayed with streamline tube visualization in three representative subjects. Data-Driven Groupwise Fiber Clustering tracts are displayed in green, Manual Multiple-ROI Labeling tracts in blue, and White Matter Query Language tracts in red.

S1= Subject 1, S2= Subject 2, S3= Subject 3.

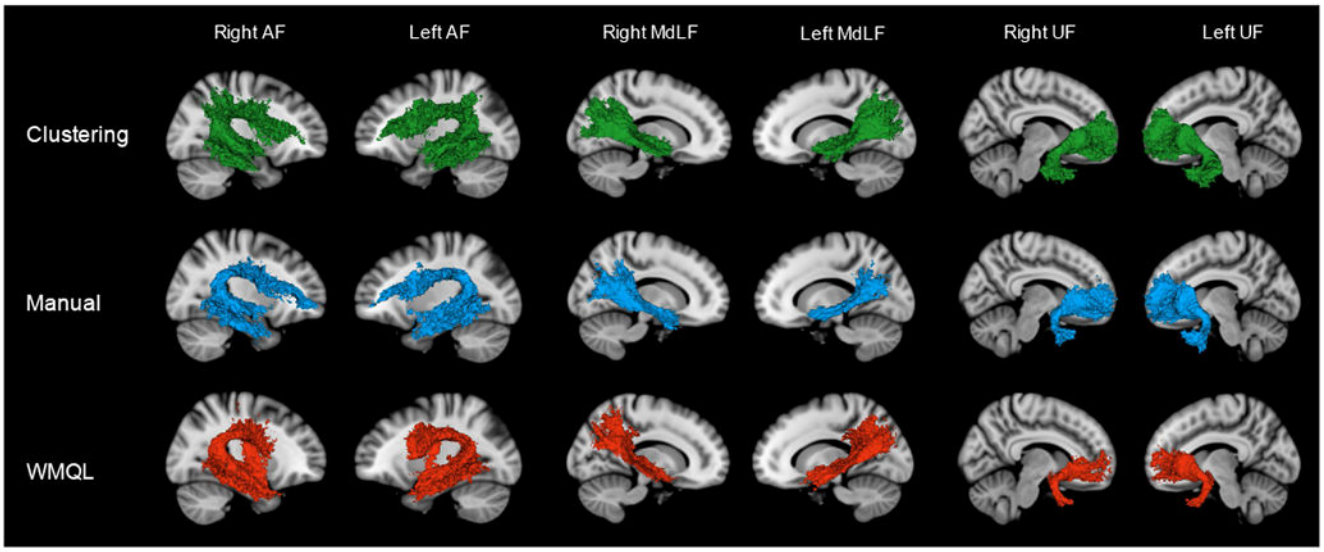

Figure 3.

3D Models of the method-specific tract overlay maps generated by Data-Driven Groupwise Fiber Clustering (green), Manual Multiple-ROI Labeling (blue) and the White Matter Query Language (red). Method-specific tract overlay maps were obtained by overlaying binarized tract masks from all 10 subjects in MNI152 space and labeling voxels that were present in at least 50% of subjects, thus capturing the degree of intra-method tract overlap.

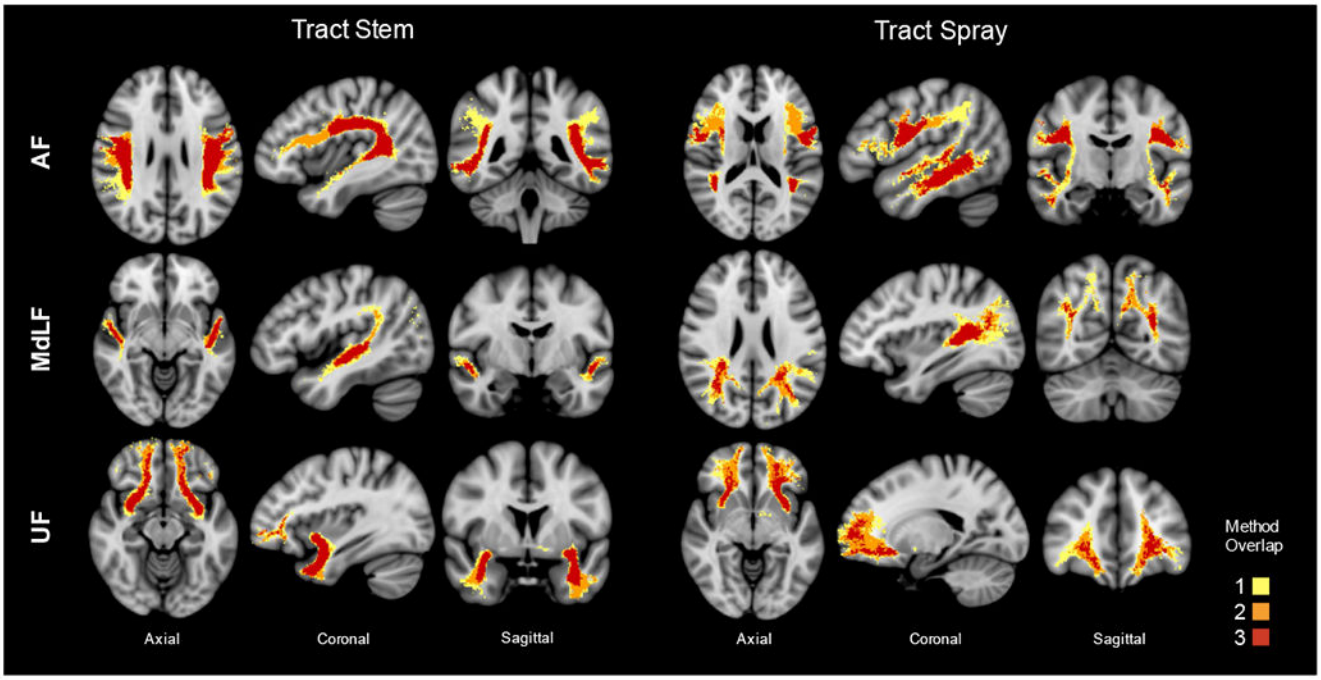

Figure 4.

Representative axial, coronal, and sagittal views of tract stem and tract spray regions of the AF, the MdLF, and the UF are shown. Binarized method-specific tract overlay maps generated for each method and tract were overlaid on the MNI152 T1 template to produce heatmaps that display spatial agreement across the three methods. Red areas depict voxels where streamlines were present from all three methods. Orange areas depict voxels where streamlines were present from two out of three methods. Yellow areas depict voxels where streamlines were present from one method only.

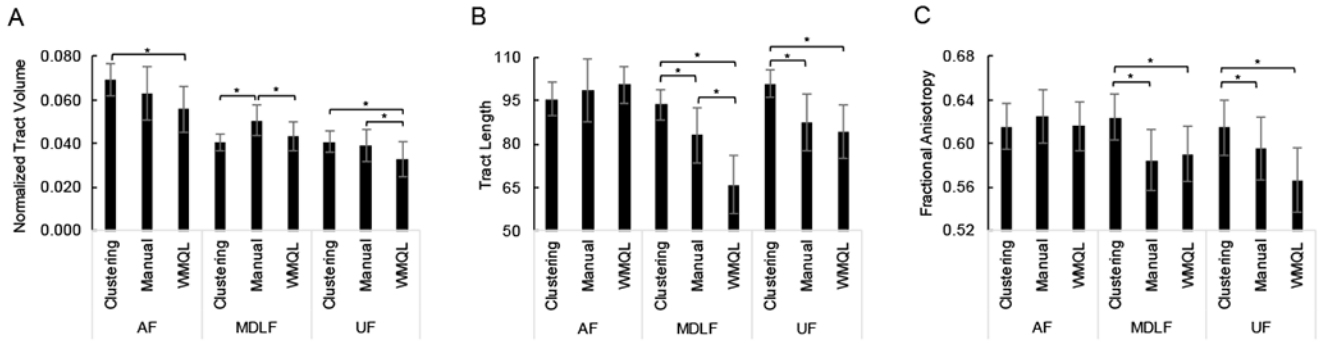

Figure 5.

Mean tract volume, tract length, and fractional anisotropy measures across methods.

* indicates significance

Table 1.

White Matter Query Language Tract Definitions.

White Matter Tract	WMQL Query Definition
Arcuate Fasciculus	AF.side = ((inferior_frontal_gyrus.side or middle_frontal_gyrus.side or precentral.side) and (superiortemporal.side or middletemporal.side) not in hemisphere.opposite, not in medial_of(supramarginal.side), not in ilf.side, not in ioff.side, not in ex.side, not in mdlf.side, not in mdlf_probable_sections.side, not in temporalpole.side, not in frontalpole.side, not in subcortical.side, not in rostralmiddlefrontal.side, not in lateralorbitofrontal.side, not in parstriangularis.side, not in superfrontal.side, not in parsopercularis.side)
Middle Longitudinal Fasciculus	MdLF.side = only((temporalpole.side or superiortemporal.side) and (inferiorparietal.side or superiorparietal.side or supramarginal.side or precuneus.side or (UnsegmentedWhiteMatter.side and superiorparietal.side) or (UnsegmentedWhiteMatter.side and inferiorparietal.side)))
Uncinate Fasciculus	UF.side = ((orbitofrontalgyrus.side or inferior_frontal_gyrus.side) and endpoints_in(temporalpole.side) and insula.side) not in occipital.side, not in parietal.side, not in cingular.side, not in posterior_of(putamen.side), not in hemisphere.opposite)

Table 2.

Volumes of the Method-Specific Tract Overlap Maps.

Tract overlay maps (displayed in Figure 3) were obtained by overlaying binarized tract masks from all 10 subjects in MNI152 space and labeling voxels that were present in at least 50% of subjects. This table displays the volume of the region with greater than 50% intra-method overlap.

White Matter Tract	Method	Volume (mm^3)
Arcuate Fasciculus	Clustering	40,850
	Manual	29,996
	WMQL	24,179
Middle Longitudinal Fasciculus	Clustering	19,770
	Manual	10,341
	WMQL	17,715
Uncinate Fasciculus	Clustering	27,600
	Manual	21,231
	WMQL	9,455

Table 3.

Tract-Derived Measures.

Mean (S.D.) normalized tract volume (absolute tract volume divided by total intracranial volume), tract length, and fractional anisotropy measures obtained from individual subject white matter tracts.

White Matter Tract	Method	Normalized Tract Volume	Tract Length	FA
Arcuate Fasciculus	Clustering	.069 (.008)	95.6 (6.0)	0.615 (0.02)
	Manual	.063 (.012)	98.5 (10.8)	0.625 (0.02)
	WMQL	.056 (.011)	100.5 (6.3)	0.616 (0.02)
Middle Longitudinal Fasciculus	Clustering	.040 (.004)	93.6 (5.2)	0.624 (0.02)
	Manual	.050 (.007)	83.1 (9.4)	0.596 (0.03)
	WMQL	.043 (.007)	65.9 (10.1)	0.590 (0.03)
Uncinate Fasciculus	Clustering	.041 (.005)	100.8 (4.7)	0.614 (0.03)
	Manual	.039 (.007)	87.3 (9.7)	0.585 (0.03)
	WMQL	.033 (.008)	84.1 (9.2)	0.566 (0.03)

Table 4.

Jaccard Coefficients.

Mean (S.D.) across- and between-method Jaccard Coefficients. JC= Jaccard Coefficient

White Matter Tract	3 Method JC	Clustering + Manual JC	Clustering + WMQL JC	Manual + WMQL JC
Arcuate Fasciculus	0.286 (0.070)	0.514 (0.078)	0.418 (0.075)	0.401 (0.082)
Middle Longitudinal Fasciculus	0.163 (0.036)	0.321 (0.061)	0.366 (0.066)	0.294 (0.047)
Uncinate Fasciculus	0.307 (0.067)	0.632 (0.076)	0.355 (0.068)	0.429 (0.079)
		Mean = 0.489	Mean = 0.379	Mean = 0.374

Table 5.

STAPLE Performance Level Measures.

Sensitivity and specificity values for each method computed by the STAPLE algorithm.

Method	Sensitivity (p)	Specificity (q)
Clustering	.860 (.102)	.998 (.0007)
Manual	.766 (.166)	.999 (.0005)
WMQL	.612 (.126)	.999 (.0008)

Table 6.
Tract-Specific Anatomical Connectivity Across Methods.

An analysis of streamline endpoints based on FreeSurfer atlas regions. Regions displayed contained endpoints from at least 5% of a given white matter tract's streamlines.

White Matter Tract	Method	Cortical FreeSurfer Regions with >= 5% of Streamline Endpoints	White Matter FreeSurfer Regions with >= 5% of Streamline Endpoints
Arcuate Fasciculus	Clustering	precentral gyrus +++ middle temporal gyrus +++ inferior temporal gyrus ++ supramarginal gyrus	middle temporal wm +++ supramarginal wm
	Manual	precentral gyrus +++ middle temporal gyrus +++ inferior temporal gyrus ++	middle temporal wm +++ inferior temporal wm
	WMQL	precentral gyrus +++ superior temporal gyrus middle temporal gyrus +++	precentral wm middle temporal wm +++
Middle Longitudinal Fasciculus	Clustering	superior temporal gyrus +++ superior parietal lobule ++ inferior parietal lobule +++	superior temporal wm +++ inferior parietal wm
	Manual	superior temporal gyrus +++ inferior parietal +++	superior temporal wm +++
	WMQL	superior temporal gyrus +++ superior parietal lobule ++ inferior parietal lobule +++supramarginal gyrus	superior temporal wm +++ superior parietal wm supramarginal wm
Uncinate Fasciculus	Clustering	superior frontal gyrus rostral middle frontal gyrus middle temporal gyrus ++ inferior temporal gyrus temporal pole gyrus +++	
	Manual	middle temporal gyrus ++ temporal pole gyrus +++	lateral orbitofrontal wm ++
	WMQL	temporal pole gyrus +++	lateral orbitofrontal wm ++ temporal pole wm

+++Region appears in all three methods

++Region appears in two methods