



Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics



Yan Jin ^{a,b,*}, Yonggang Shi ^b, Liang Zhan ^{a,b}, Boris A. Gutman ^{a,b}, Greig I. de Zubicaray ^d, Katie L. McMahon ^e, Margaret J. Wright ^f, Arthur W. Toga ^b, Paul M. Thompson ^{a,b,c}

^a Imaging Genetics Center, University of Southern California, Los Angeles, CA 90089, USA

^b Institute for Neuroimaging & Informatics, Departments of Neurology, University of Southern California, Los Angeles, CA 90089, USA

^c Department of Psychiatry, Pediatrics, Engineering, Radiology, and Ophthalmology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

^d School of Psychology, University of Queensland, St. Lucia, QLD 4072, Australia

^e Centre for Advanced Imaging, University of Queensland, St. Lucia, QLD 4072, Australia

^f QIMR Berghofer Institute of Medical Research, Herston, QLD 4029, Australia

ARTICLE INFO

Article history:

Accepted 12 April 2014

Available online 9 May 2014

Keywords:

HARDI

Tractography

Fiber clustering

Label fusion

Genetic heritability

ABSTRACT

To understand factors that affect brain connectivity and integrity, it is beneficial to automatically cluster white matter (WM) fibers into anatomically recognizable tracts. Whole brain tractography, based on diffusion-weighted MRI, generates vast sets of fibers throughout the brain; clustering them into consistent and recognizable bundles can be difficult as there are wide individual variations in the trajectory and shape of WM pathways. Here we introduce a novel automated tract clustering algorithm based on label fusion – a concept from traditional intensity-based segmentation. Streamline tractography generates many incorrect fibers, so our top-down approach extracts tracts consistent with known anatomy, by mapping multiple hand-labeled atlases into a new dataset. We fuse clustering results from different atlases, using a mean distance fusion scheme. We reliably extracted the major tracts from 105-gradient high angular resolution diffusion images (HARDI) of 198 young normal twins. To compute population statistics, we use a pointwise correspondence method to match, compare, and average WM tracts across subjects. We illustrate our method in a genetic study of white matter tract heritability in twins.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Diffusion-weighted magnetic resonance imaging (DT-MRI) (Basser et al., 1994) is a powerful non-invasive brain imaging technique introduced in LeBihan et al. (1986), Merboldt et al. (1985), Taylor and Bushell (1985). DT-MRI measures water diffusion in tissues, and provides biologically and clinically relevant information on white matter (WM) integrity and connectivity not available from other imaging modalities. It is increasingly used to study pathology and connectivity of WM pathways in the living brain (Daianu et al., 2013; Jahanshad et al., 2012; Thomason and Thompson, 2011).

Recently, diffusion MRI has been extended to more sophisticated models of local diffusion, such as high angular resolution diffusion imaging (HARDI) (Tuch, 2004), diffusion spectrum imaging (Wedeen et al., 2005), or even hybrid imaging where large numbers of angular samples

are collected at several diffusion weightings (Zhan et al., 2011). With these imaging protocols, we can more accurately reconstruct fibers that mix and cross.

Tractography is a method to reconstruct the pathways of major WM fiber bundles, by fitting a curved path through the directional diffusion data at each voxel. Deterministic tractography (Basser et al., 2000; Conturo et al., 1999; Mori et al., 1999) recovers fibers emanating from a seed voxel by following the principal direction of the diffusion tensor or the dominant direction of the diffusion orientation distribution function (ODF). However, deterministic tractography has limitations: it depends on the choice of initial seed points and can be sensitive to the estimated principal directions. To overcome those drawbacks, probabilistic tractography methods have been proposed (Aganj et al., 2011; Behrens et al., 2003; Parker and Alexander, 2003). They can be computationally more intensive but can be more robust to partial volume averaging effects and uncertainties in the underlying fiber direction, which are inevitable due to imaging noise.

Several approaches have been developed to study brain connectivity using whole-brain tractography. Jahanshad et al. (2011) computed a whole-brain connectivity matrix based on streamline tractography and anatomical parcellation. Network-based analysis of this matrix can identify factors that affect the interconnectedness of regions in the

* Corresponding author.

E-mail addresses: yjin@ucla.edu (Y. Jin), yonggang@usc.edu (Y. Shi), zhan.liang@gmail.com (L. Zhan), bgutman@gmail.com (B.A. Gutman), greig.dezubicaray@uq.edu.au (G.I. de Zubicaray), katie.mcmahon@cai.uq.edu.au (K.L. McMahon), margie.wright@qimr.edu.au (M.J. Wright), toga@usc.edu (A.W. Toga), pthomp@usc.edu (P.M. Thompson).

brain. For example, Ingalhalikar et al. (2014) revealed connectivity pattern differences between males and females. Prasad et al. (2011) applied a probabilistic WM atlas to extract major fiber bundles and represented them by using a “maximum density” path. A mean curve was used to represent each bundle in each subject. Fractional anisotropy (FA) values, and other indices of diffusion, can be compared along this path across a population, using ‘along-tract’ statistics (Colby et al., 2011; Corouge et al., 2006).

Obviously it is important to accurately identify WM structures and fibers from whole-brain tractography. If fibers are grouped into bundles, the results can offer valuable insight on how disease affects the integrity of particular WM tracts (Price et al., 2007, 2008). Clustering methods can group fibers obtained from tractography into organized bundles or tracts, enabling large population studies of disease and genetic effects on tract integrity, or even tract shapes. One simple yet practical strategy selects anatomically well-known WM tracts that interconnect anatomical regions of interest (ROI) (Wakana et al., 2007; Zhang et al., 2010). 3D models of tracts can facilitate large-scale population studies (Brouwer et al., 2010; Yushkevich et al., 2008). Even so, the final results often need substantial manual intervention to help screen out false positive fibers.

Automatic fiber clustering would accelerate and empower population studies, so long as the results are accurate and reliable. A typical framework for fiber clustering defines a pairwise similarity/distance between each pair of fibers in a large set of candidate fibers, to group them into separate and distinct tracts. Many different fiber similarity metrics have been proposed, such as the mean vector and the covariance matrix of fiber points (Brun et al., 2004), the number of points shared within the same voxel (Jonasson et al., 2004), an associativity vector (Wang et al., 2012), the average mean distance (Gerig et al., 2004; O'Donnell et al., 2006; Xia et al., 2005), Hausdorff distance (Gerig et al., 2004; Xia et al., 2005), and Mahalanobis distance (Maddah et al., 2008). Also, various clustering algorithms have been advocated, such as hierarchical clustering (Gerig et al., 2004; Visser et al., 2011; Xia et al., 2005), expectation-maximization (Wang et al., 2012), fuzzy *c*-means (Li et al., 2010), *k*-nearest neighbors (Ding et al., 2003), normalized cuts (Brun et al., 2004), dual rooted graphs (Tsai et al., 2007), and spectral clustering (O'Donnell and Westin, 2007; Wassermann et al., 2008).

If clustering algorithms have no anatomical information to guide them, tracts may not correspond to any anatomically familiar subdivisions. There is also no guarantee that the same basic sets of bundles will be generated again in datasets from new subjects, making it hard to compare results from one study to the next. Also, a user typically needs to specify the number of clusters or a threshold to decide when to stop merging or splitting clusters. Clustering results can vary drastically when different numbers of clusters are specified. “Bottom-up” methods cluster fibers into larger groups until major tracts are aggregated, but they may not efficiently filter out erroneous fibers buried in the large number of streamlines (100,000–1,000,000) generated by whole-brain tractography.

Recent hybrid approaches extract the well-known WM tracts by using a combination of prior information from an anatomically-labeled atlas and similarity-based clustering. Wassermann et al. (2010) proposed a Gaussian process framework to generate a fiber ‘denrogram’ and selected which ones to merge through a query system based on parcellated volumetric information. Li et al. (2010) clustered tracts via anatomical ROI guidance, and then passed them through similarity-based fuzzy *c*-means clustering. Guevara et al. (2012) implemented a two-level (intra-subject and inter-subject) centroid-based average-link hierarchical clustering. The resulting clusters were manually labeled to form a multi-subject WM atlas. A new tractography data set was similarly segmented and the clusters were labeled using a supervised classification based on the atlas.

The large number of false positive fibers produced by streamline-based tractography hinders large population studies. An atlas-based top-down clustering method resolves this, by requiring that all subjects’

WM tracts fall within a pre-defined set of shapes or regions. Even so, an atlas based on one individual subject's anatomy is not sufficient to capture the variability of individual WM tracts. One classical solution is called multi-atlas labeling or *label fusion*. This has commonly been applied to label brain structures on standard anatomical MRI (Chou et al., 2007, 2008, 2009; Heckemann et al., 2006; Lötjönen et al., 2010; Rohlfing et al., 2004; Sabuncu et al., 2010).

In traditional image segmentation, a deformable atlas may be used, in which an atlas is non-rigidly registered to the image to be labeled. The resulting deformation may then be used to map the training labels onto the new image. Multiple atlases and registrations may also be used to transfer multiple training labels to the new subject's space. The final labeling can be obtained by applying a weighting approach to the labels transferred from different atlases. Label fusion has two advantages: 1) it is easier to accommodate large individual variations in anatomy if one does not have to rely on a single atlas; 2) multiple registrations improve robustness against occasional registration failures and non-global minima of the registration cost function. The same idea can also improve voxel-based or tensor-based morphometry (Leporé et al., 2008).

Here we extend label fusion to fiber clustering and introduce a multi-atlas framework to automatically extract anatomically meaningful WM tracts. Based on the ROIs from a publicly available parcellated WM atlas (Oishi et al., 2009), we first manually construct a number of WM fiber tract atlases, consisting of several major WM tracts. In contrast to prior “bottom-up” methods, we use the WM tracts in multiple hand-labeled atlases as prior anatomical information. Our “top-down” approach transfers tract labels by selecting only fibers similar to the corresponding tracts in the atlases, based on a similarity measure. This eliminates many false positive fibers hidden in the ~1,000,000 fibers per subject produced by streamline tractography. Multiple atlases adapt to the variability of tract shapes in new subjects. This reduces the number of outliers and picks fibers that can be incorrectly omitted when registering a single atlas to the whole-brain tractography in a new subject. Finally, we use label fusion to combine the clustered results from individual atlases.

In the second part of the paper, we illustrate our method to study tract heritability based on the clustering results from our algorithm. Voxel-wise genetic analyses of DT-MRI show that many diffusivity measures, including FA, are heritable (Chiang et al., 2011; Jahanshad et al., 2013, 2014; Kochunov et al., 2011; Lee et al., 2010), but it is not yet well-understood which tracts are genetically influenced.

As individual WM fiber tracts are highly variable in shape, it can be difficult to find corresponding fibers that belong to the same tract across a population. Recent studies examined the skeleton of tracts, with methods such as tract-based spatial statistics (TBSS) (Bodini et al., 2009; Smith et al., 2007) or the average fiber tracts (Brouwer et al., 2010; Prasad et al., 2011) to perform statistical analyses of diffusion parameters in a large population. Nevertheless, these approaches do not always retain the full 3D profile of information from the tracts. To address this, we use a pointwise tract correspondence method to study clustered tract parameters in 3D. Finally, we calculate heritability statistics from corresponding tract points to understand genetic influences on the brain's tracts, and to demonstrate a practical use of our entire workflow.

Label fusion clustering framework

Here we summarize our clustering algorithm before explaining each step in detail, and the datasets used to test it.

Overview

First, whole-brain tractography needs to be generated, as a basis to perform clustering (Tractography section). To help with clustering fibers in new datasets, we manually reconstructed multiple

WM atlases to represent anatomically well-defined tracts. The ROIs of a single-subject standard template were warped to our manual atlases through registration and the tracts were extracted by using a look-up table and manually edited (Tract atlas construction section). To cluster tracts in a new subject, the corresponding tracts from our manual atlases were warped to that subject's space and a multi-level fiber clustering scheme was applied to label the tracts (Multi-level fiber clustering section). Finally, labels were fused to synthesize the individual clustering results from different manually-created atlases (Fiber label fusion section). These steps are summarized in Fig. 1.

Subjects and image acquisition

We analyzed a dataset from 198 healthy young adult twins (mean age: 23.2 ± 2.1SD) from 99 families in Australia. All twins were right-handed. No subjects had any major medical condition or psychiatric illness. All subjects were evaluated to exclude any pathology known to affect brain structure. Diffusion imaging was available in 99 complete pairs – 62 monozygotic pairs (21 male-only pairs) and 37 same-sex dizygotic twin pairs (12 male-only pairs).

HARDI scans were acquired with a 4T Siemens Bruker Medspec MRI scanner, using single-shot echo planar imaging with parameters: TR/TE = 6090/91.7 ms, 23 cm FOV, and a 128 × 128 acquisition matrix. Each 3D volume consisted of 55 2-mm axial slices, with no gap, and 1.79 × 1.79 mm² in-plane resolution. 105 image volumes were acquired per subject: 11 with no diffusion sensitization, i.e., T2-weighted b_0 volumes, and 94 diffusion-weighted volumes ($b = 1159$ s/mm²). The raw HARDI images were corrected for eddy-current induced distortions with FSL (www.fmrib.ox.ac.uk/fsl/). The gradient table that represented the diffusion scanning angles was adjusted accordingly.

Tractography

We performed whole-brain tractography with Camino (<http://cmic.cs.ucl.ac.uk/camino/>), an open source software package that uses either streamline or probabilistic methods to reconstruct fiber paths. It uses a spherical harmonic (SH) representation to represent the diffusion ODF; this can be more robust to noise and may even be more accurate for detecting fiber crossings than the original

numerical q -ball reconstruction method (Descoteaux et al., 2007). Explicitly, the SH basis may be expressed as follows:

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\varphi} \quad (1)$$

where l denotes the order, m denotes the phase factor, $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi]$, and P_l^m is an associated Legendre polynomial. Signal at each gradient direction may be approximated as a linear combination of a modified version of this SH basis. We used the 6th order ($l = 6$) SH series to reconstruct orientation distribution functions (ODF) for our HARDI data and a maximum of 3 local ODF maxima (where fibers mix or cross) were set to be detected at each voxel.

Next, we performed fiber tracking with a probabilistic algorithm, called the 'Probabilistic Index of Connectivity' (PiCo) (Parker et al., 2003), in Camino. At first, we created a simulated data set for diffusion probability density function (PDF) calibration, based on the signal-to-noise ratio of our dataset. Then, based on different diffusion distribution uncertainty models (Bingham or Watson), simulated data was reconstructed to generate a look-up table, which was in turn used to produce the PDF estimates of the actual data from the derived local ODF maxima. Seed points were chosen at those voxels whose FA values were greater than 0.4. Monte Carlo simulation was used to generate fibers proceeding from the seed points throughout the entire brain. Streamline fiber tracking followed the voxel-wise PDF profile with the Euler interpolation method for 10 iterations per each seed point. The maximum fiber turning angle was set to 40°/voxel, and tracing stopped at any voxel whose FA was less than 0.2.

Tract atlas construction

We randomly selected five subjects (none were members of the same twin pair, three males and two females) from our HARDI data set and constructed WM tract atlases. In this context, atlas means a hand labeled representation of the fibers in a subject. The FA images of all the atlases were registered to a single-subject template in the ICBM-152 space called the "Type II Eve Atlas" (a 32-year-old healthy female) (Oishi et al., 2009). The entire brain of the "Eve" template was parcellated using 130 bilateral ROIs (Zhang et al., 2010).

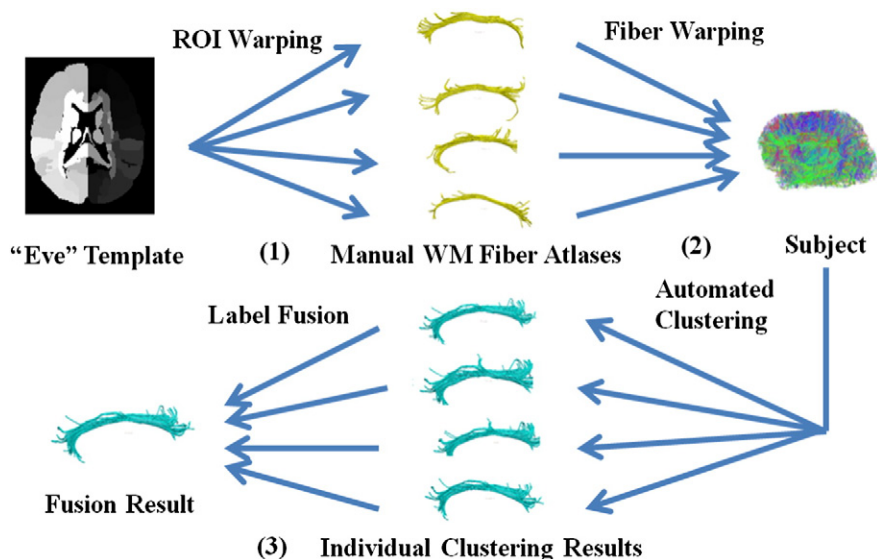


Fig. 1. Flow chart of steps in our fiber clustering framework. (1) Manual WM atlases were constructed through warped ROI extraction from a standard template; (2) a multi-level fiber clustering scheme was used to label tracts in new subjects; (3) final results were created by fusing individual results from step (2).

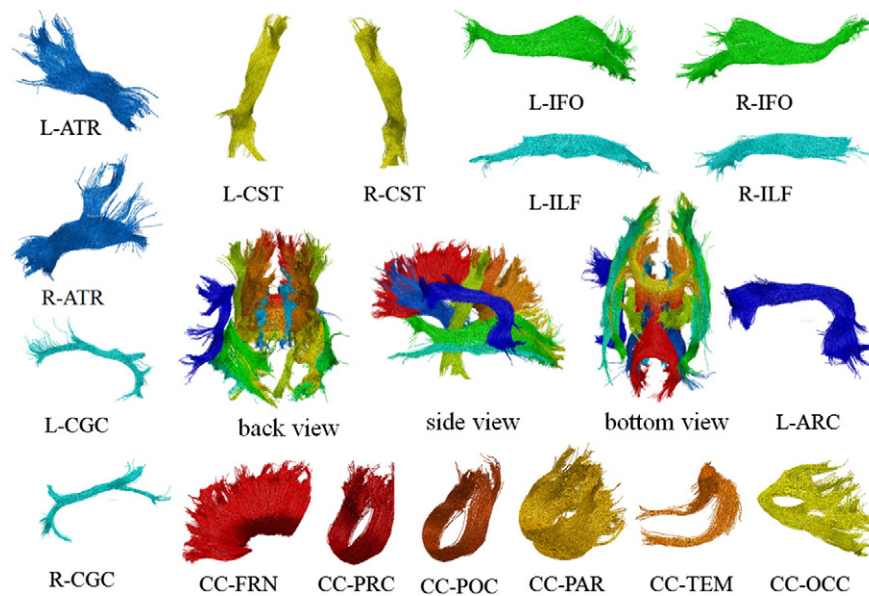


Fig. 2. A representative WM fiber atlas computed, and manually edited, from 4-Tesla 105-gradient HARDI data, showing the 17 major tracts. We created these, with manual editing, in 5 subjects and propagated the tracts into new subjects. The tracts on the left side and the corpus callosum segments are viewed from the left, while the tracts on the right are viewed from the right. *Back, left side, and bottom views* of tract overlays are shown in the middle of the figure.

The labeled template ROIs were re-assigned to the five registered atlases, respectively, by warping them with the deformation fields generated by Advanced Neuroimaging Tools (ANTs) (Avants et al., 2008, <http://picsl.upenn.edu/software/ants/>). Fibers that traversed the ROIs were extracted according to the look-up table in Zhang et al. (2010). For example, the corticospinal tract was extracted from fibers passing between the precentral gyrus and the cerebral peduncle. Finally, each tract was manually edited to remove visible outliers. We rounded the floating point coordinate locations along a fiber to the location of the closest voxel. For a given tract, there is a certain set of ROIs that it is expected to intersect – this may be two or more. If any of the fiber's discrete lattice points fell into the ROI, we would consider that this fiber traversed the ROI; otherwise, it didn't. A fiber must traverse all the required ROIs for a given tract to be considered as a candidate member of that tract, or it will be discarded. There was no ambiguity in deciding whether a fiber belonged to a tract or not, because it was counted as traversing the set of expected ROIs or not.

Currently, each atlas comprised 17 major WM tracts: left (L)/right (R) corticospinal tract (CST), L/R anterior thalamic radiation (ATR), L/R cingulum (CGC), L/R inferior fronto-occipital fasciculus (IFO), L/R inferior longitudinal fasciculus (ILF), L arcuate fasciculus (part of the superior longitudinal fasciculus) (ARC), and six segments of the corpus callosum – projecting to both frontal lobes (CC-FRN), precentral gyri (CC-PRC), postcentral gyri (CC-POC), superior parietal lobes (CC-PAR), temporal lobes (CC-TEM), and occipital lobes (CC-OCC) (Hofer and

Frahm, 2006). We did not include the right arcuate fasciculus as not all subjects had this tract in its entirety (Catani et al., 2007), and that would have made statistical analysis difficult. Fig. 2 shows all 17 WM tracts that we created and views of overlaid tracts from different angles (*back, left side, and bottom views*).

Multi-level fiber clustering

For each test subject (i.e., each new data set to be labeled), whole-brain tractography was extracted by using Camino as well. We designed a multi-stage fiber clustering scheme to pick out appropriate candidate fibers for each tract.

Length-based clustering

Image noise and limits in image resolution cause whole-brain tractography to generate many implausible short-length streamlines. All of our target fiber bundles are major tracts, so it was reasonable to assume that no component fiber is shorter than 35 mm. Streamlines shorter than this were removed. Of course this leads to some arbitrary limit in the types of tracts surveyed, and we have to concede that some true short fibers may be suppressed as well as many false ones.

ROI-based clustering

The labeled “Eve” template ROIs were re-assigned to the subjects in our data set with the ANTs registration. Fibers that did not traverse the

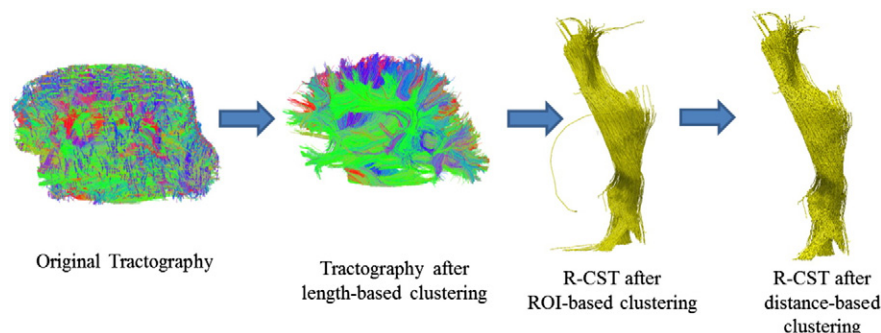


Fig. 3. Here we show how many of the streamlines generated by the original tractography are filtered to form the corticospinal tract, step by step, through multi-level fiber clustering.

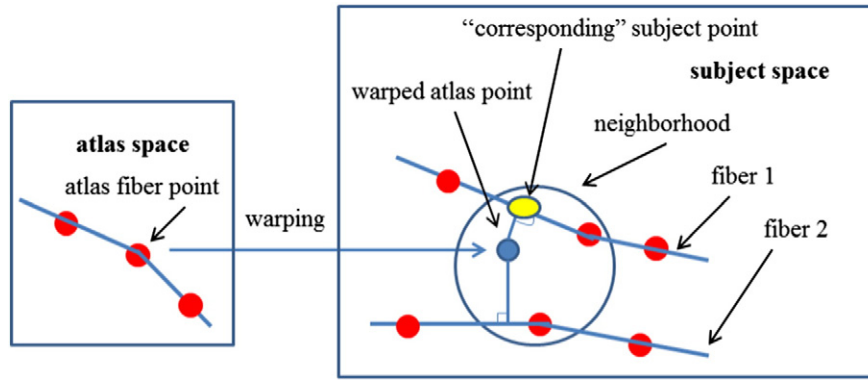


Fig. 4. An illustration of tract projection.

ROIs for a particular tract were removed. This reduced the number of detected fibers from around a million to a few hundreds or thousands of fibers of interest.

A few previous studies showed reliable and reproducible clustering results with either manual ROI placement (Wakana et al., 2007) or automated placement (Zhang et al., 2010). In our workflow, this step was also helpful to ease the computation when further refining the results in the next step (Distance-based clustering section). With the automated ROI placement, we could effectively reduce the number of fibers of interest for a particular tract. Since it was a standard procedure consistently applied to all the subjects in our dataset, it established a basis to compare subjects for population studies.

Distance-based clustering

After applying ROI constraints and the filters above, most of the short and obviously erroneous fibers were removed. To eliminate the remaining false fibers (see Fig. 3) we implemented a geometry-based clustering algorithm to select only those streamlines whose shapes and locations were similar to those of manually constructed WM atlases in Tract atlas construction section.

First, we registered the test subject's FA image to the FA image of each tract atlas by using ANTs. Each atlas's tracts were then warped to

the subject space with the corresponding deformation fields generated from the FA registration.

We defined a fiber distance metric to decide the subject's fibers that should be included in any individual warped atlas tract. For any pair of fibers γ_i and γ_j , we defined the symmetric Hausdorff distance (Gerig et al., 2004):

$$d_H(\gamma_i, \gamma_j) = \max(d_{H'}(\gamma_i, \gamma_j), d_{H'}(\gamma_j, \gamma_i)) \quad (2)$$

where $d_{H'}$ is the asymmetric Hausdorff distance.

$$d_{H'}(\gamma_i, \gamma_j) = \max_{x \in \gamma_i} \min_{y \in \gamma_j} \|x - y\| \quad (3)$$

$\| \cdot \|$ is the Euclidean norm and the ordered pair (γ_i, γ_j) indicates an asymmetric distance from γ_i to γ_j . x 's and y 's are the coordinate points along fibers γ_i and γ_j , respectively.

Initial fuzzy fiber labeling based on each individual atlas was performed simply by aggregating fibers passing within a neighborhood of each atlas fiber, and setting the appropriate label. Fig. 3 illustrates how the right corticospinal tract is parsed out from the original whole-brain tractography following the multi-level fiber clustering steps.

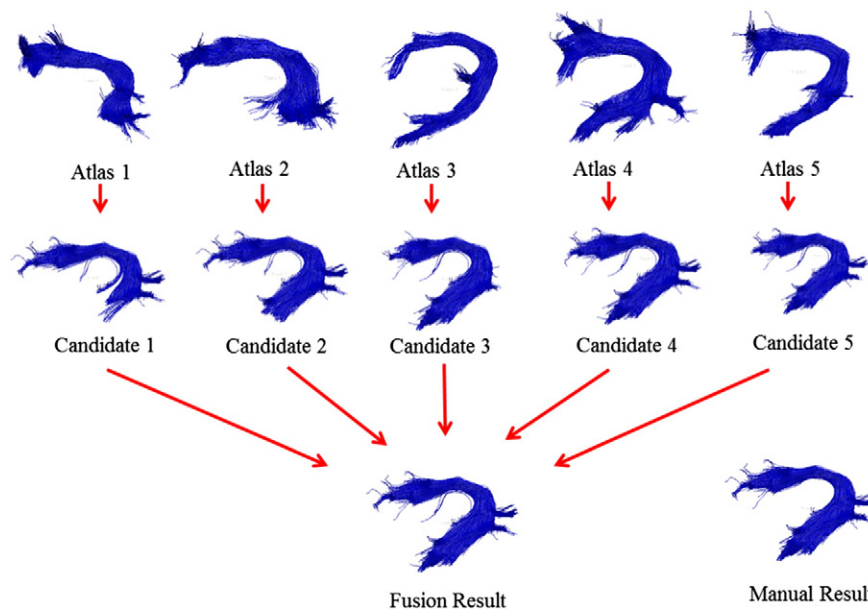


Fig. 5. Label fusion result obtained from the five manually labeled atlases for the left arcuate fasciculus (in blue) in a test subject (viewed from the left). A manual segmentation is included for comparison at the bottom right.

Fiber label fusion

We chose the Hausdorff distance metric for the fiber clustering phase to select only streamlines with similar geometric shapes, and lying in the same region as the atlas tracts. However, due to the WM variability of individual atlases, different atlases may 'nominate' different candidates. We extended the well-established label fusion concept from traditional structural image segmentation to the WM fiber space. We defined a mean fiber distance to rank the fibers nominated by individual atlases. For each fiber, the mean fiber distance was defined as:

$$d_{mean} = \frac{\sum_{d_i < d_{cutoff}} d_i + \sum_{d_i > d_{cutoff}} d_{sup}}{n} \quad (4)$$

where d_i is the Hausdorff distance between an unlabeled subject's fiber and the i -th atlas, d_{cutoff} is the empirical cutoff threshold chosen in [Distance-based clustering](#) section, d_{sup} is the upper bound Hausdorff distance within which a subject fiber can be possibly considered a candidate

for a given tract, and n is the number of atlases. We ranked all the candidate fibers from different atlases based on their d_{mean} 's. The smaller its d_{mean} , the higher its rank. For each particular tract, a fusion percentage was defined to include fibers whose d_{mean} 's were among the top specified percentage. For example, if we set the fusion percentage as 90%, this means that we keep 90% fibers whose ranks were among the top 90% according to their d_{mean} 's and throw away the rest.

Statistical analysis of heritability

Fiber matching

To perform group studies, we first need to establish some kind of correspondence between fibers of the segmented tracts across the population. For each tract, we chose a representative sample among our manually constructed atlases. The representative sample was then projected into individual new subject space as follows. After applying the registration warp ([ROI-based clustering](#) section) to the representative

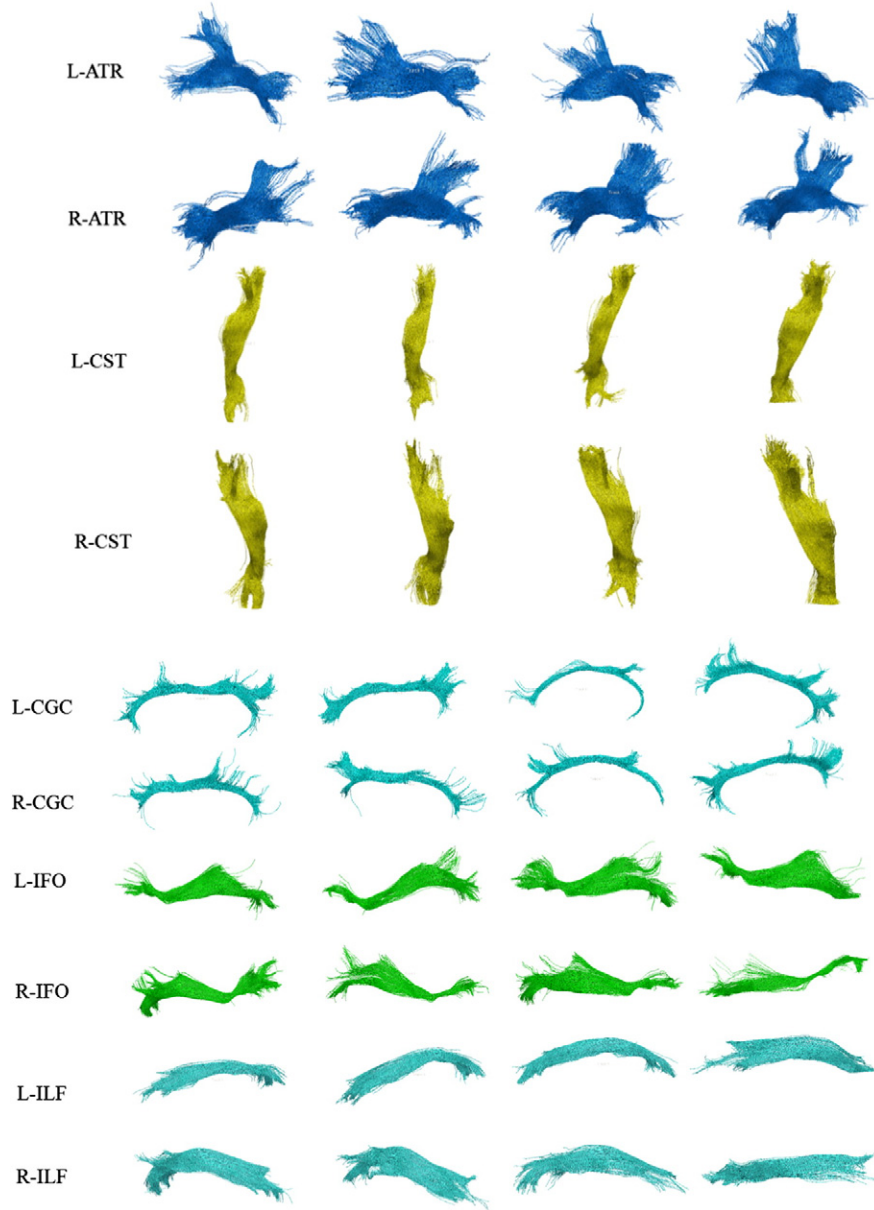


Fig. 6. Clustering results in four randomly selected subjects for the left and right anterior thalamic radiations (L/R-ATR), the left and right corticospinal tracts (L/R-CST), the left and right cingulum (L/R-CGC), the left and right inferior fronto-occipital fasciculi (L/R-IFO), and the left and right inferior longitudinal fasciculi (L/R-ILF). Tracts on the left side are viewed from the left, while the tracts on the right are viewed from the right.

tract, we defined the corresponding point in the new subject space for each point in the sample tract. The corresponding point is defined as the point in the new subject tract closest to the warped sample point, if such a point exists within a given neighborhood, or the original sample projection, if not. An illustration of fiber matching is shown in Fig. 4.

Genetic analysis

Monozygotic (MZ) twins share 100% of their genetic variants whereas dizygotic (DZ) twins share, on average, 50% of their genes. A simple and widely-used estimate of heritability in twin studies is to assess how much the intra-class correlation for MZ twin pairs (r_{MZ}) exceeds the DZ twin correlation (r_{DZ}). Falconer's heritability statistic (Falconer and Mackay, 1996) is defined as:

$$h^2 = 2(r_{MZ} - r_{DZ}) \quad (5)$$

The statistic estimates the proportion of the overall variance due to genetic differences among individuals. Of course more sophisticated structural equation models can be used, but here the main purpose was just to give an example of a statistical manipulation of the tracts, to compute a statistic of interest. Jin et al. (2011b)

used the mean and the standard deviation of the Euclidean distances between individual fibers and the mean curve for a particular tract as metrics, where the mean distance is related to the tract thickness. Here, interpolated FA values at corresponding tract points (from Fiber matching section) were used to calculate intra-class correlations r_{MZ} and r_{DZ} .

To account for the multiple comparisons problem that arises when testing a statistical hypothesis at every fiber point, the false discovery rate (FDR) was used (Benjamini and Hochberg, 1995) at the conventional 5% level to provide corrected critical p values for the maps of r_{MZ} and r_{DZ} in Falconer's heritability statistic. r_{MZ} and r_{DZ} whose associated p values were larger than 0.05 were set to 0. The Falconer's heritability statistic h^2 was computed from r_{MZ} and r_{DZ} at points whose p -value passed FDR.

Results

Cluster visualization

Fig. 5 shows how we obtained one example tract – the left arcuate fasciculus – in a test subject – a new subject whose scan was not used to create the fiber atlases. The first row shows the atlas (hand-labeled)

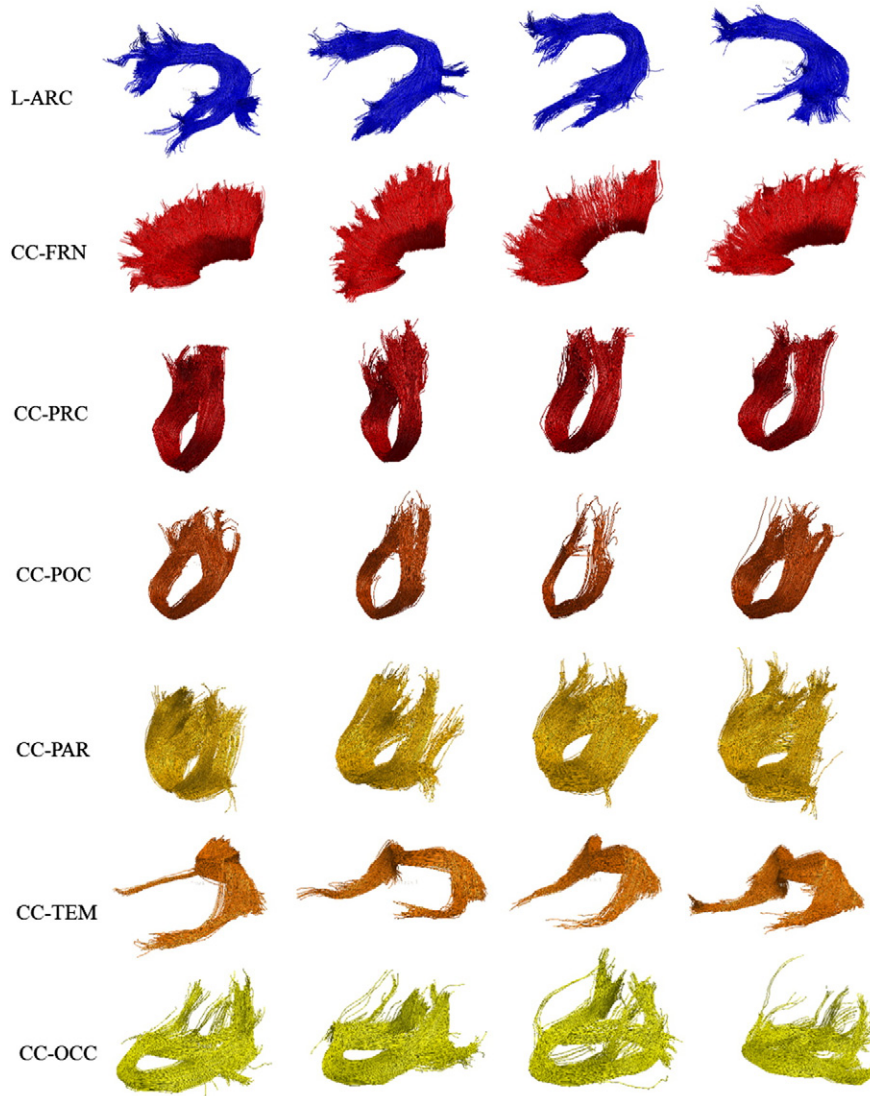


Fig. 7. Clustering results – for the same four subjects as in Fig. 6 – for the left arcuate fasciculus (L-ARC) and six segments of the corpus callosum – projecting to both frontal lobes (CC-FRN), precentral gyri (CC-PRC), postcentral gyri (CC-POC), superior parietal lobes (CC-PAR), temporal lobes (CC-TEM), and occipital lobes (CC-OCC). All the tracts are viewed from the left.

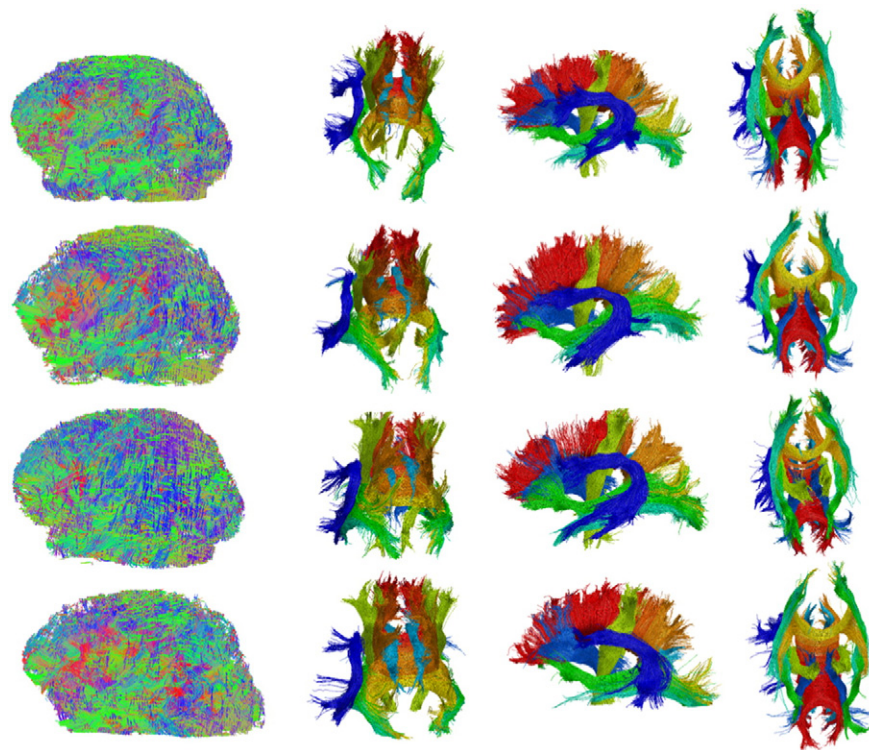


Fig. 8. Back, left side, and bottom views of the same four subjects' (in Figs. 6 and 7) compositional fiber clustering results are shown. The original whole-brain tractography (the leftmost column) is included for comparison, clearly showing the utility of the data reduction.

versions of the tract. The second row shows the different candidates for this tract in the same test subject, based on using each atlas to decide which fibers it should contain (Distance-based clustering section). The final result for this tract was obtained by applying the label fusion scheme in Fiber label fusion section. It is not hard to see that the label fusion process can help to eliminate outliers, and it can also add missing fibers to a single candidate labeling of the tract. A manually edited segmentation result is also included for comparison (see the right bottom panel).

Figs. 6 and 7 show the label fusion results for the 17 segmented tracts in four randomly selected subjects. Despite individual variations, the overall tract shapes are consistent across the population. Fig. 8 shows the combined WM fiber clustering results for the four test subjects. The types of tracts and their colors are as in Fig. 2. The average fiber number in our full set of clustering results is ~40,000 per subject, or roughly 1/10th of the fibers from the initial tractography. There are three factors that affect how many fibers are included in the final results. First, in this work, we mainly focused on 17 major anatomically well-known white matter tracts. Therefore, only those tracts are shown in Fig. 8. Many other less-known tracts are not shown and could be added in future work, although it might be more challenging to reliably find smaller tracts in the mix of all the other major pathways. Second, streamline whole-brain tractography generates large numbers of false positive fibers and those need to be removed for our ultimate goal – population studies. Last, fiber clustering may show enormous individual variation when applied across a population. However, to perform an

effective population study, we only included fibers whose shape shares the most common characteristics throughout the population for each tract. This was our intent when we built our manually constructed atlases. Clearly we would need to admit that some clinically interesting variation is missed by focusing on a set of standard tracts. But finding additional consistent tracts across subjects is challenging and runs the risk of including false positives.

Quantitative validation

To quantitatively evaluate the proposed framework, we converted each of the fiber tracts to a binary image, where voxels that the tracts cross were marked as 1, and 0 otherwise. Then we used the Dice coefficient to assessing the overlap or agreement between two tracts, defined as:

$$D(a, b) = \frac{2 * (V(a) \cap V(b))}{V(a) + V(b)} \quad (6)$$

where $V()$ is the volume of the region that the tract penetrates.

Due to the wide variability between different tracts, we need to tune the parameters of our algorithm to optimize its performance. We have two key parameters to adjust. One is the Hausdorff distance threshold used to select fibers for each tract per atlas (d_{cutoff} in Eq. (4)), and the other is the percentage of fibers included in the final label fusion stage described in Fiber label fusion section.

Table 1
The values of the clustering distance threshold and the fusion percentage for each tract described in Tract atlas construction section that were used to cluster the 198 subjects in our data set.

Tract name	L/R-ATR	L/R-CGC	L/R-CST	L/R-IFO	L/R-ILF	L-ARC	CC-FRN	CC-PRC	CC-POC	CC-PAR	CC-TEM	CC-OCC
d_{cutoff} (mm)	12	12	12	12	12	12	12	12	12	12	12	12
Fusion pct. (%)	100	85	95	95	70	95	90	95	100	95	45	100

In addition to the 5 subjects that we used for our manually constructed atlases, we randomly selected another 7 subjects (non-twin pairs, three males and four females) from our data set (12 in total) and manually segmented the 17 tracts mentioned in [Tract atlas construction](#) section. Initially, we tuned d_{cutoff} and the fusion percentage with the leave-one-out method using the 5 manually labeled atlases. We first used a loose Hausdorff distance bound ($d_{\text{cutoff}} = d_{\text{sup}}$ in Eq. (4)), 15 mm, to select the candidate fibers for each tract per atlas so that all the true fibers were included without introducing too many false fibers. Then we optimized the percentage threshold for inclusion in label fusion from 20% to 100% (in increments of 5%) to obtain the optimal percentage (the best Dice coefficients against manual segmentation). Next, we varied the Hausdorff distance threshold (d_{cutoff}) from 3 mm to 15 mm (in increments of 1 mm) to decide the optimal distance, while the optimal fusion percentage was used from the previous step. The optimal parameters for each tract are shown in [Table 1](#). The optimized parameters were then applied to the seven randomly selected subjects as the test data. [Fig. 9](#) compares the average Dice coefficients of all clustered tracts between the training data set and the test data. The error bars stand for the standard deviation of the Dice coefficients for each tract. Not surprisingly, the training data outperformed the test data for almost all the tracts, except for the left fronto-occipital fasciculus (L-IFO). [Fig. 10](#) shows the average Dice coefficients for all tracts with our label fusion method and ROI-only clustering (based on the look-up table in [Zhang et al., 2010](#)), against manual segmentation for the test subjects only. Overall, our algorithm outperformed the ROI method for every tract, and also gave a smaller variance, especially for those tracts that have unclear or loose ROI constraints (CGC, ILF, and CC-TEM).

Genetic analysis

For each tract, we selected a representative example among our 12 subjects with manually segmented tracts in [Quantitative validation](#)

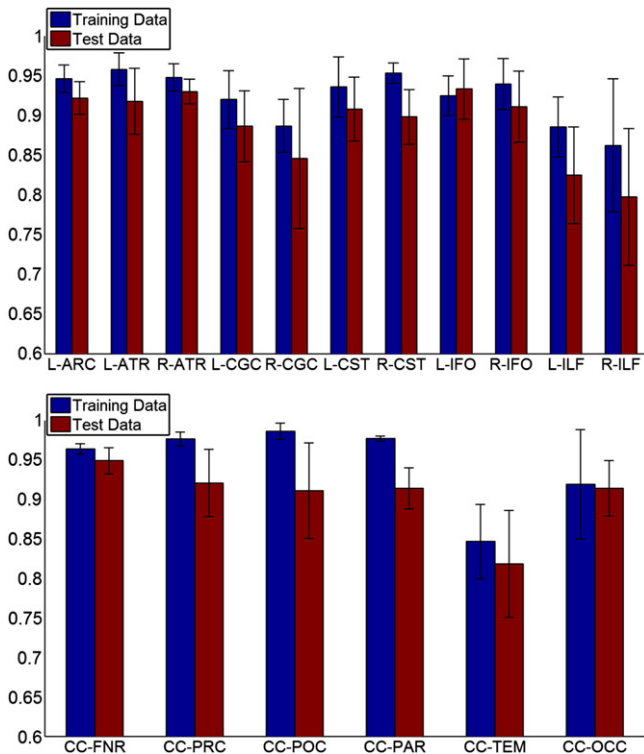


Fig. 9. Average Dice coefficients of all the tracts described in Tract atlas construction section for the training data set (five manually constructed atlases by the leave-one-out test) and the test data set (the average of seven randomly selected subjects from our data set) using our label fusion method against manual segmentation. The general pattern of coefficients above 0.8 indicates good agreement of automatically segmented and hand-segmented tracts.

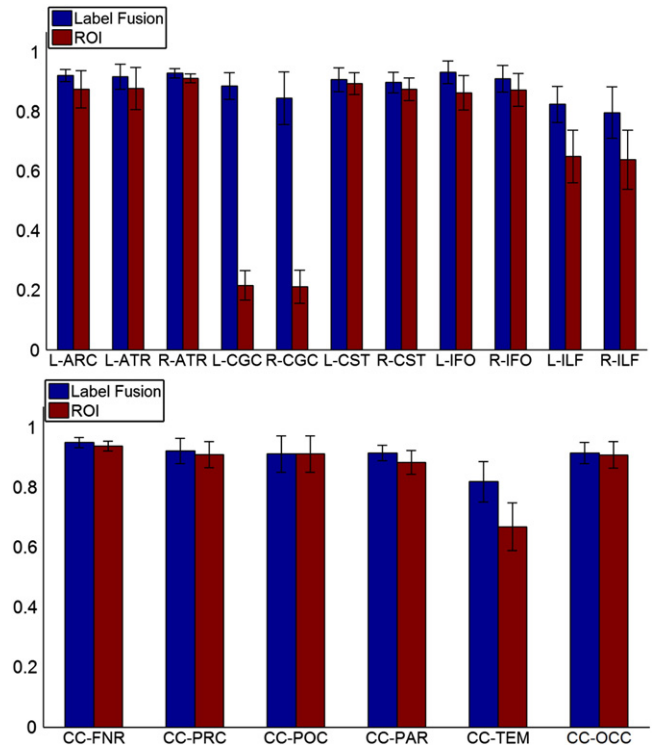


Fig. 10. The average Dice coefficients of all the tracts described in Tract atlas construction section for our label fusion method and the ROI method against manual segmentation. The label fusion method universally performs very well (dark blue bars), even when the ROI method (red) performs poorly.

section. We projected the tract representation to the remaining 197 subjects, following [Fiber matching](#) section. The search range was a 1-cm radius sphere. The FA values at sub-voxel fiber points were used to calculate r_{MZ} and r_{DZ} for a particular tract. Falconer's heritability statistic was computed from Eq. (5). To make the computation easier, we uniformly resampled each fiber at 15 equidistant points and reduced the number of fibers in the representative tract, ensuring that the region enclosing the original tract was still covered entirely by the remaining fibers.

[Figs. 11, 12, and 13](#) show Falconer's heritability statistics on FA after correcting for multiple comparisons with FDR. As heritability must be positive and lies between 0 and 1, Falconer's heritability statistics were set to 0 if their estimator was negative and 1 if it was greater than 1. Locations in red show greater genetic influence than those in blue. The percentages of points with high genetic influence (set arbitrarily to $h^2 > 0.7$) for each tract are also listed in [Table 2](#). Genetic factors tend to have greater influence on the tracts on the left side than the right side (ATR, CGC, CST, and ILF), except for the IFOs.

Clustering parameter selection

To rationalize our choice of the two key parameters, the fusion percentage and the clustering distance threshold (d_{cutoff}), here we examined how the clustering performance varied with these two parameters. [Fig. 14\(a–c\)](#) shows the changes in the average Dice coefficients over the seven test subjects used for each tract, respectively, when the fusion percentages increase from 20% to 100%, that is, choosing the top 20% to 100% ranked candidate fibers obtained from the five manually constructed atlases, based on the mean clustering distance defined in Eq. (4), where $d_{\text{cutoff}} = d_{\text{sup}} = 15$ mm.

A high fusion percentage leads to fewer missing or false negative fibers from the candidates obtained from other manually constructed atlases. However, a high fusion percentage will also include more false

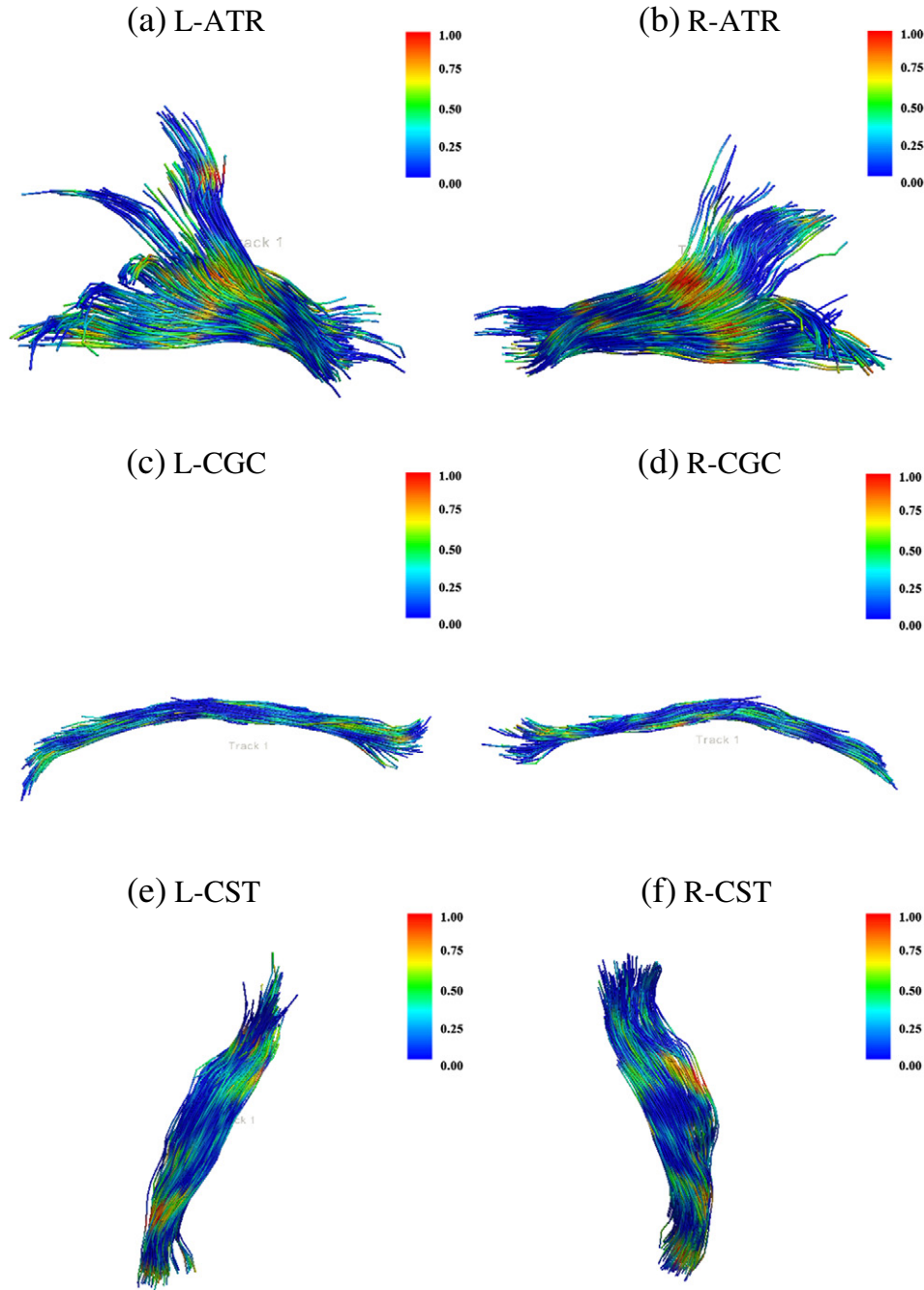


Fig. 11. Color maps of Falconer's heritability statistic on FA for (a) the left anterior thalamic radiation (L-ATR), (b) the right anterior thalamic radiation (R-ATR), (c) the left cingulum (L-CGC), (d) the right cingulum (R-CGC), (e) the left corticospinal tract (L-CST), and (f) the right corticospinal tract (R-CST). Warmer colors show regions with higher genetic influence (h^2 closer to 1). Tracts on the left side are viewed from *the left*, while the tracts on the right are viewed from *the right*.

positive fibers from all the candidates. Tracts that have helpful constraints based on ROIs (ATR, CGC, CST, IFO, ARC, CC-FRN, CC-PRC, CC-POC, CC-PAR, and CC-OCC) usually have fewer false positives in each candidate. Therefore, the benefit of reducing the false negatives tends to overwhelm the gain in false positives when the fusion percentage goes up. Dice coefficients increase until they hit plateaus between 85% and 100%; however, for those tracts whose ROI constraints are relatively loose (IFO and CC-TEM), the plateaus come much sooner when the fusion percentages are still relatively low (less than 70%). This is because eliminating more false positive fibers in each candidate is more critical for maintaining a high Dice coefficient. The percentages listed in Table 1 were therefore adopted for all the subjects in our data set.

Fig. 15(a–c) shows the change of the average Dice coefficients versus the cutoff clustering distances d_{cutoff} , when the fusion percentages are fixed at the values listed in Table 1. It seems that the average Dice coefficients become stable after $d_{cutoff} > 10$ mm. This is probably because a smaller d_{cutoff} is too conservative and excludes too many true fibers. Hence, we chose $d_{cutoff} = 12$ mm uniformly for all the tracts, which corresponds to the percentage of fibers eliminated during manual labeling of the atlases for the given tract.

Label fusion vs. single atlas

Label fusion-based segmentation methods take advantage of the information of multiple atlases and generally outperform methods

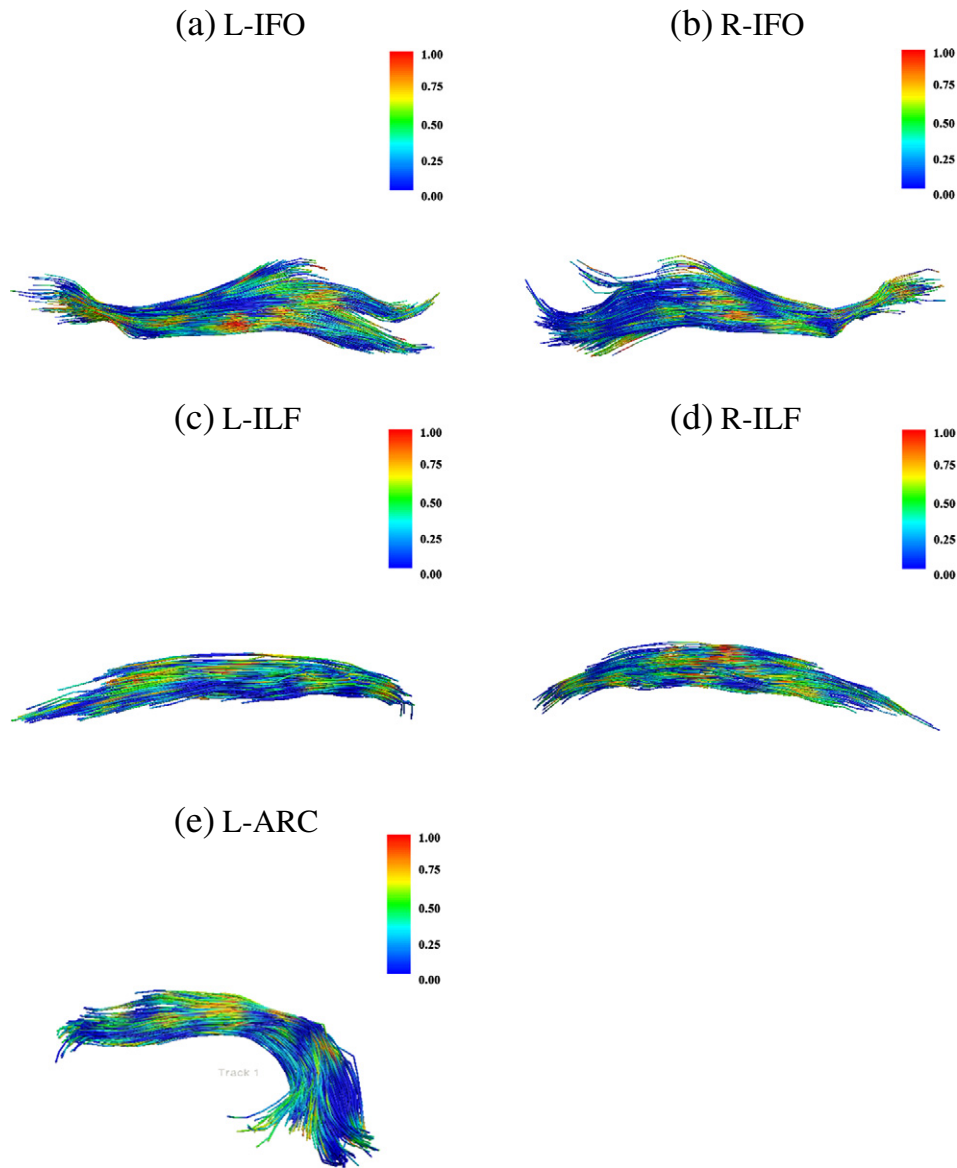


Fig. 12. Color maps show Falconer's heritability statistic on FA for (a) the left inferior fronto-occipital fasciculus (L-IFO), (b) the right inferior fronto-occipital fasciculus (R-IFO), (c) the left inferior longitudinal fasciculus (L-ILF), (d) the right inferior longitudinal fasciculus (R-ILF), and (e) the left arcuate fasciculus (L-ARC). Warmer colors show regions with higher genetic influence ($h^2 \sim 1$). Tracts on the left side are viewed from the left, while the tracts on the right are viewed from the right.

that only use a single atlas. We listed the mean Dice coefficients against manual segmentation between label fusion clustering and single atlas clustering in Table 3. The parameters of label fusion clustering were chosen as in Table 1. The mean Dice coefficients were computed over the 7 test subjects chosen in Quantitative validation section. For comparison, we calculated Dice coefficients using a single atlas out of the five manually constructed atlases with $d_{\text{cutoff}} = 12$ mm. The single-atlas mean Dice coefficients were averaged over all the five atlases and all the seven test subjects (35 combinations) per tract. Table 3 justifies fusing multiple atlas results in cases where the performance of a single atlas is well below the average performance of all the five atlases.

Number of atlases

How many atlases should be used in the label fusion scheme is always an open question. Among other examples, Chou et al. (2007, 2008, 2009) studied this for the case of lateral ventricular segmentation on standard anatomical MRI. In our case, we performed an experimental

analysis to explore the choice of atlas number. We used the twelve subjects (including the five atlases) in Quantitative validation section in our analysis. The five atlases plus two more additional subjects formed our atlas pool. We calculated the average dice coefficients between the clustering results of the rest of the five subjects against their manual segmentations while we used 1, 2, 3, 5, and 7 atlases from our pool. The atlas selection was random. The clustering parameters were chosen as the same as those in Table 1. The results are shown in Table 4. The Dice coefficients of L-ATR and R-ATR increased 8%–9% from a single atlas to two atlases and continued going up as more atlases were used. For R-CGC, L-ILF, CC-TEM, and CC-OCC, the Dice coefficients seemed to be benefited with more atlases, too. Overall, the Dice coefficients approaches stable after three atlases, while more atlases didn't decrease the Dice coefficients. However, by considering our small test sample size and balancing the tradeoff between stability and computation cost, five atlases may be suitable to be representative of our data set. Based on the validation results in Quantitative validation section, the five atlases that we selected could effectively cluster the tracts for new subjects in our data set. (See Table 4.)

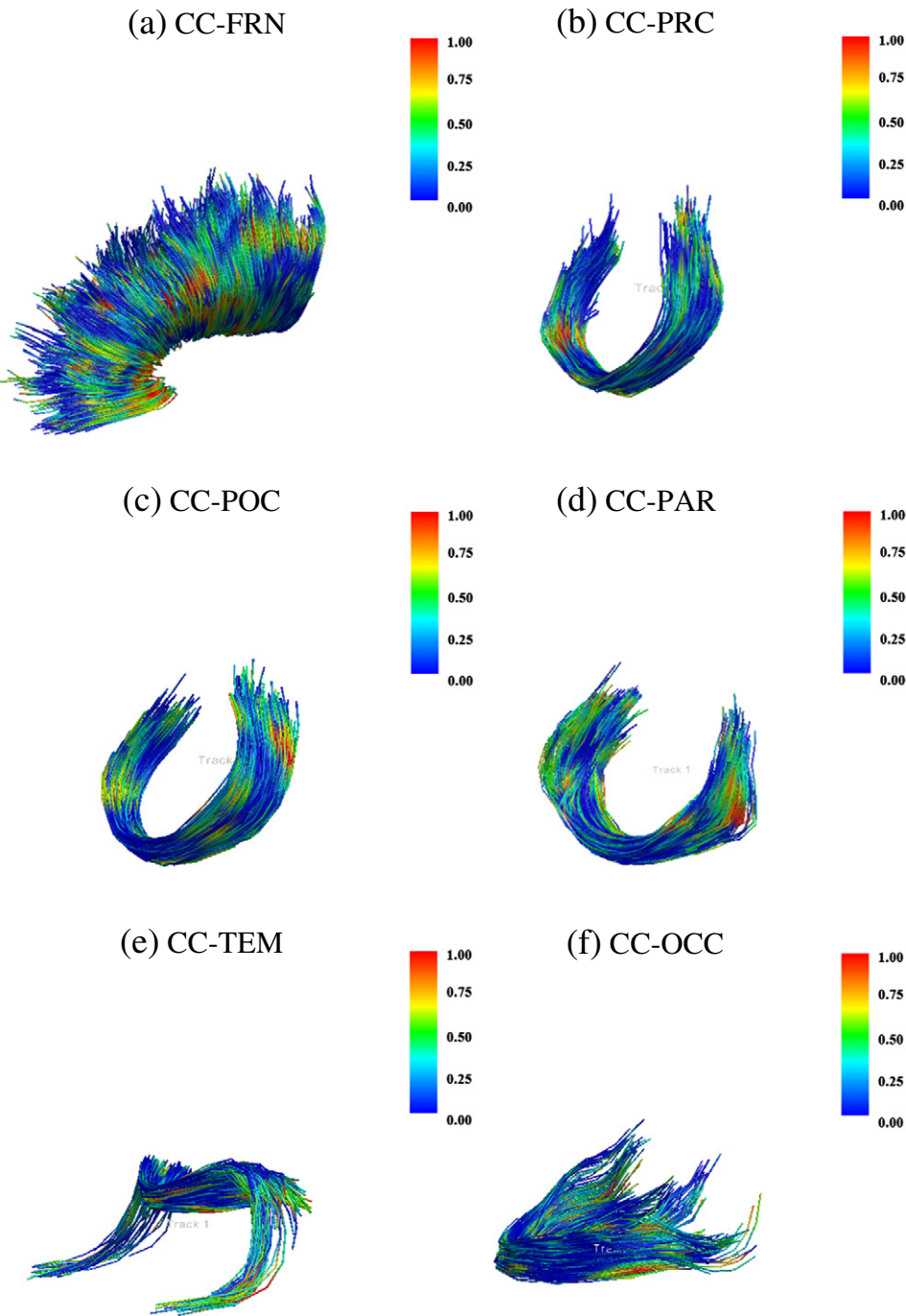


Fig. 13. Color maps of Falconer's heritability statistic on FA for the six segments of the corpus callosum – projecting to both (a) frontal lobes (CC-FRN), (b) precentral gyri (CC-PRC), (c) postcentral gyri (CC-POC), (d) superior parietal lobes (CC-PAR), (e) temporal lobes (CC-TEM), and (f) occipital lobes (CC-OCC). Warmer colors show regions with higher genetic influence ($h^2 \sim 1$). All of the corpus callosum segments are viewed from the left.

Table 2
The percentage of fiber points in all clustered tracts where high genetic influence is detected (arbitrarily set to $h^2 > 0.7$). The high genetic influence detected in the frontal regions of the corpus callosum (CC-FRN) may reflect its generally higher FA, which tends to lead to higher SNR for statistical analysis.

Tract name	L-ATR	R-ATR	L-CGC	R-CGC	L-CST	R-CST	L-IFO	R-IFO	L-ARC
Pct. (%)	8.6	7.1	4.2	1.2	7.8	4.5	15.4	13.6	7.8
Tract name	L-ILF	R-ILF	CC-FRN	CC-PRC	CC-POC	CC-PAR	CC-TEM	CC-OCC	
Pct. (%)	7.7	11.2	17.5	7.9	6.2	4.6	5.6	6.8	

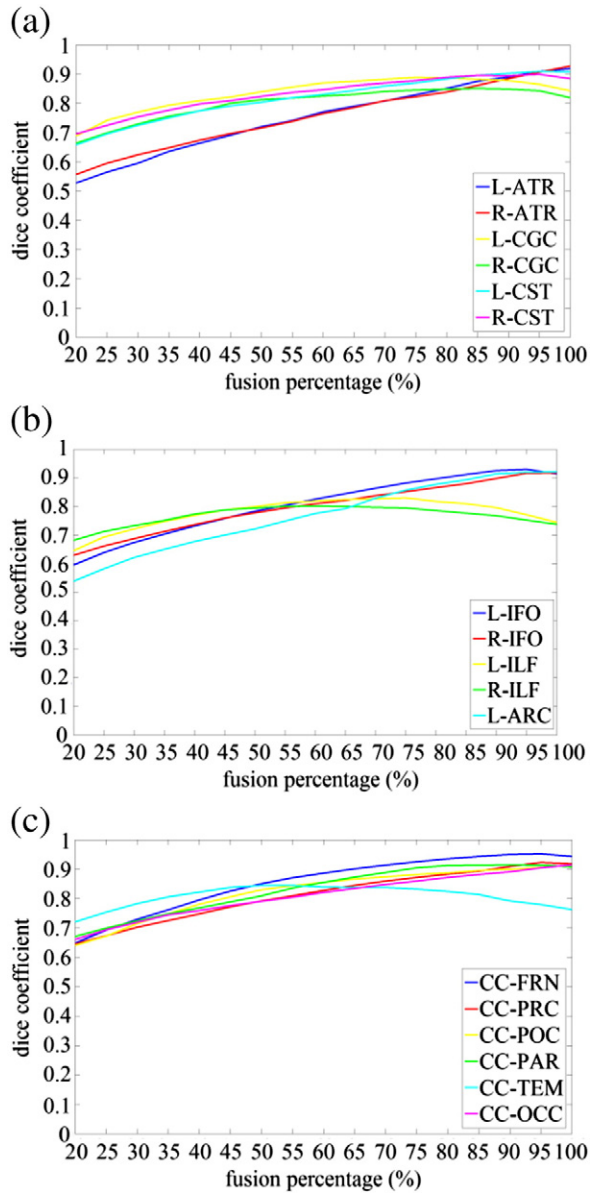


Fig. 14. Changes in the average Dice coefficients are shown, over the seven test subjects against the fusion percentages that were applied in the label fusion stage (Fiber label fusion section) for all the tracts described in Tract atlas construction section.

Genetic stability analysis

Falconer's heritability statistic is defined as twice the difference between the MZ intra-class correlation and the DZ intra-class correlation as described in Eq. (5). The intra-class correlations are associated with the variances between and across members of pairs of twins in each group. One important consideration is that the estimated heritability (or any other statistic) needs to be robust to the details of parameter selection – for example, it may change as the interpolated FA values of the corresponding fiber points change if different radii of sphere are chosen to find corresponding points for each tract across the population. We evaluated the stability of Falconer's heritability statistic for each tract by computing the percentage of fiber points whose Falconer's heritability statistics change by less than 0.2 (this is admittedly arbitrary) when changing the radius of the sphere used to search for corresponding points across the population (those familiar with the TBSS method will note that a similar dependency may arise for the cross-subject correspondences used in TBSS). Table 5 lists the stability of Falconer's

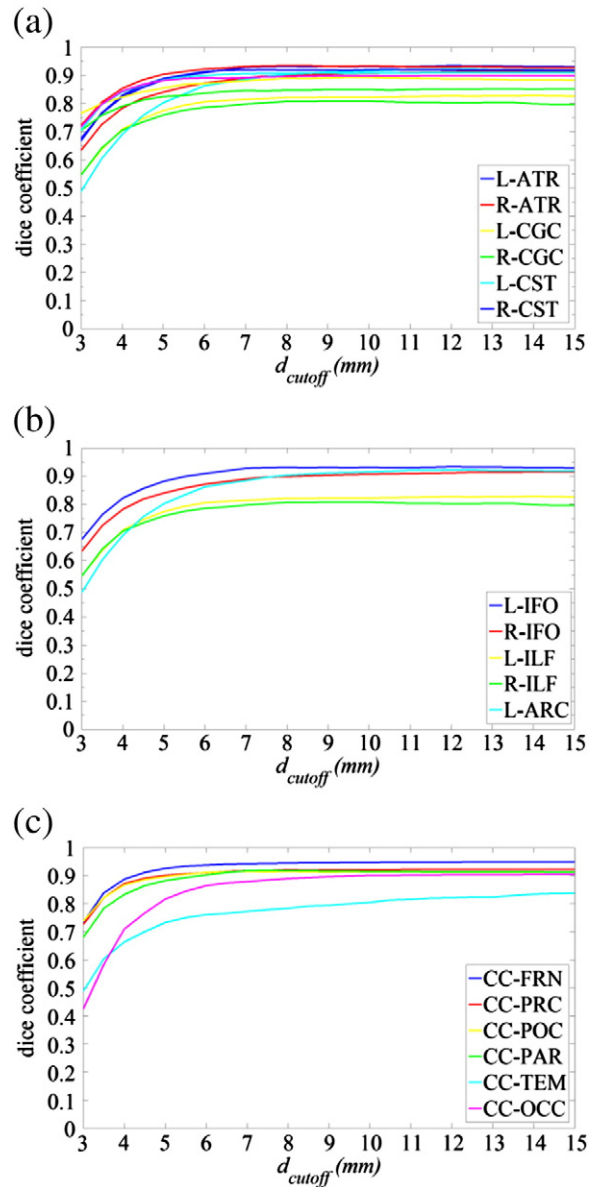


Fig. 15. The changes of the average Dice coefficient in seven test subjects versus the clustering distance threshold (d_{cutoff}) (Distance-based clustering section) for all the tracts described in Tract atlas construction section. Here, the optimal values of the fusion percentage (shown in Table 1) were selected.

heritability statistics for all clustered tracts when the radius of the search sphere changes from 0.5 cm to 1 cm to 1.5 cm. Overall, Falconer's heritability statistic is stable for most tracts except CC-TEM. Therefore, the results in Genetic analysis section do provide somewhat robust information on genetic influences for each tract. As the number of twin pairs increases, the Falconer's heritability should have tighter confidence limits. The relatively low stability of CC-TEM arises because the clustering result for the CC-TEM bundle has a large variance (its Dice coefficient in Quantitative validation section is also the lowest). This is perhaps due to the large variations in the tractography results in that area.

Discussion

It is obvious that tract clustering quality is influenced by the quality of tractography. A better tractography algorithm that produces fewer false fibers makes clustering easier and more accurate. The whole-brain tractography methods place seeds throughout the brain and

Table 3
The average Dice coefficients and their standard deviations over the seven test subjects against manual segmentation for label fusion on the five manually labeled atlases and only a single atlas being used (the average was computed over the five atlases).

Mean Dice coefficients \pm standard deviation						
Tract name	L-ATR	R-ATR	L-CGC	R-CGC	L-CST	R-CST
Label fusion	0.92 \pm 0.04	0.93 \pm 0.02	0.88 \pm 0.04	0.85 \pm 0.09	0.91 \pm 0.04	0.90 \pm 0.03
Single atlas avg.	0.88 \pm 0.05	0.90 \pm 0.05	0.88 \pm 0.04	0.84 \pm 0.08	0.91 \pm 0.04	0.89 \pm 0.04
Tract name	L-IFO	R-IFO	L-ILF	R-ILF	L-ARC	
Label fusion	0.93 \pm 0.04	0.91 \pm 0.04	0.82 \pm 0.06	0.80 \pm 0.09	0.92 \pm 0.02	
Single atlas avg.	0.92 \pm 0.04	0.89 \pm 0.06	0.80 \pm 0.04	0.78 \pm 0.09	0.86 \pm 0.06	
Tract name	CC-FRN	CC-PRC	CC-POC	CC-PAR	CC-TEM	CC-OCC
Label fusion	0.95 \pm 0.02	0.92 \pm 0.04	0.91 \pm 0.06	0.91 \pm 0.03	0.81 \pm 0.07	0.91 \pm 0.04
Single atlas avg.	0.95 \pm 0.02	0.92 \pm 0.05	0.90 \pm 0.06	0.91 \pm 0.03	0.81 \pm 0.06	0.89 \pm 0.05

produce streamlines in a short period of time. However, many false fibers are created, for example, in the left and right inferior longitudinal fasciculi and the temporal segment of the corpus callosum, as there are multiple fiber crossings in the regions traversed by those tracts. Since our purpose was to perform large scale population studies, we were inclined to be more conservative and keep only fibers that form a consistent tract shape across the population. When we picked a representative sample for a particular tract, we followed the same principle. For example, the cingulum has many branches along their way from the frontal lobe to the temporal lobe. We removed those branches in our representative tract samples for population studies. It is not practical to find corresponding fibers for those branches across the population. We only studied the common areas across most subjects.

While our methods achieved robust performance in fiber clustering, it is important to note that there are many parameters that could be tuned in whole-brain tractography, which no doubt could potentially affect the final clustering results. First, tractography is based on mathematical input models, such as tensor, ODF, and fiber orientation density (Tournier et al., 2004); then, tractography can be deterministic or probabilistic (e.g. PICO); next, in terms of the tracking algorithm, it can be fiber assignment by continuous tracking (FACT) (Mori et al., 1999), Euler (Basser et al., 2000), Runge-Kutta (Basser et al., 2000), tensor deflection (TEND) (Lazar et al., 2003), etc.; as to interpolation algorithms, it may be nearest-neighbor, linear, TEND, etc.; finally, multiple stopping criteria can be chosen, such as masking, fiber maximum turning angle, the lowest anisotropy to

terminate tracking, etc. Furthermore, there are many tractography software packages available for brain imaging research (DTI Studio, Diffusion Toolkit, Camino, MRtrix, etc.). Each package might produce slightly different tractography results. Because of the lack of ground truth, the problem of selecting the parameters and software packages for optimal clustering performance needs more dedicated exploration and will be an important future work.

In our multi-level fiber clustering algorithm (Multi-level fiber clustering section), non-linear FA registration was used to align fibers between the atlases and the subjects. Ideally, an ODF-based registration method might be used to reorient the fibers between different spaces. However, such a registration scheme would have much larger cost in terms of computing resources and time (a few hours per registration) if it were performed on a large scale, as in the label fusion scheme. In contrast, FA registration takes only around 5 min per registration on our data set. Moreover, we have found that fiber alignment is indeed improved significantly with FA-based elastic registration (Jin et al., 2011a).

It is always difficult to compare different clustering algorithms quantitatively as we lack ground truth for both the tractography and the white matter segmentation. Even though their relative accuracy is hard to assess, it is easier to point to some conceptual similarities and differences among the various approaches that have been proposed for tract clustering. Many “bottom-up” methods (e.g., Gerig et al., 2004; Maddah et al., 2008; O'Donnell and Westin, 2007; Visser et al., 2011; Xia et al., 2005) use the relationship between neighboring fibers in one single subject and apply standard clustering algorithms to segment the tracts individually. These methods can assign each and every extracted fiber to a class, which means all the fibers are retained and represent white matter variability in an individual. Even so, they do not necessarily yield a method to match tracts across subjects – a step commonly needed for large-scale population studies. In other words, if clustering is applied independently to data from numerous individuals, there may be no easy way to match the tracts across subjects or find correspondences. On the other hand, “top-down” methods (Li et al., 2010; Wakana et al., 2007; Zhang et al., 2010) impose ROI constraints to effectively filter out a massive number of false positive streamlines generated by tractography. The tracts segmented may also have a clearer anatomical interpretation, as known ROIs are used to define them. This then offers a common set of tracts, enabling population studies. Still other methods (Guevara et al., 2012; Maddah et al., 2008; Wang et al., 2012; Wassermann et al., 2010) were driven by elegant mathematical/statistical models. However, many of these methods have not been used or widely tested in clinical research.

The contribution of our paper is to take account of individual variability by constructing multiple white matter atlases. A top-down method is also used to extract anatomically meaningful tracts. We also proposed a novel fiber correspondence scheme to show how to use our clustering results to answer a biological question.

Table 4
The average Dice coefficients over five test subjects against manual segmentation for different numbers of atlases applied.

Mean Dice coefficients					
# of Atlases	1	2	3	5	7
L-ARC	0.901	0.897	0.906	0.914	0.920
L-ATR	0.833	0.908	0.910	0.919	0.919
R-ATR	0.829	0.898	0.927	0.928	0.927
L-CGC	0.872	0.883	0.875	0.878	0.878
R-CGC	0.859	0.880	0.876	0.870	0.877
L-CST	0.921	0.917	0.914	0.911	0.911
R-CST	0.899	0.903	0.902	0.903	0.902
L-IFO	0.933	0.940	0.938	0.934	0.930
R-IFO	0.911	0.910	0.910	0.902	0.906
L-ILF	0.797	0.797	0.815	0.811	0.823
R-ILF	0.800	0.786	0.789	0.794	0.797
CC-FRN	0.943	0.947	0.946	0.946	0.945
CC-PRC	0.916	0.914	0.913	0.912	0.913
CC-POC	0.903	0.901	0.903	0.903	0.903
CC-PAR	0.908	0.902	0.906	0.908	0.907
CC-TEM	0.782	0.787	0.796	0.797	0.829
CC-OCC	0.897	0.911	0.911	0.912	0.911

Table 5

The stability of Falconer's heritability statistic (the percentage of points in the tract whose Falconer's heritability statistic changes by less than 0.2) with changes in the spherical neighborhoods (radius 0.5–1.5 mm) used to find corresponding points for each tract over the individuals in our dataset.

Stability (%)	L-ATR	R-ATR	L-CGC	R-CGC	L-CST	R-CST	L-IFO	R-IFO	L-ARC
0.5 mm vs. 1 mm	89	90	91	90	89	86	81	81	87
1.5 mm vs. 1 mm	93	94	96	94	87	87	90	87	91
	L-ILF	R-ILF	CC-FRN	CC-PRC	CC-POC	CC-PAR	CC-TEM	CC-OCC	
0.5 mm vs. 1 mm	77	84	90	80	84	78	61	83	
1.5 mm vs. 1 mm	87	93	93	82	88	81	66	92	

Conclusion

Here we presented an automatic fiber clustering workflow that uses anatomical information from multiple manually made atlases. The top-down approach helps to suppress effects of false positive streamlines by placing constraints on locations and shapes of tracts through their Hausdorff distance to warped hand-labeled atlases. The candidates from multiple atlases are combined with a fusion strategy. The clustering results are illustrated visually and quantitatively validated for a randomly selected sample data set. The clustering results from our method were more accurate than those from the ROI-only method. We also showed an example of how to perform a group statistical analysis (a heritability study) by using the sub-voxel fiber diffusion information mapped onto the clustered tracts. The complete workflow provides us with a practical tool for future large population studies that may reveal how the brain is affected by genetic factors, and by a variety of psychiatric or neurological disorders such as Alzheimer's disease.

Acknowledgments

This study was supported by grant R01 HD050735 from the National Institutes of Health and by grant 496682 from the National Health and Medical Research Council, Australia. G.I. de Zubicaray is supported by an Australian Research Council Future Fellowship.

References

- Aganj, I., Lenglet, C., Jahanshad, N., Yacoub, E., Harel, N., Thompson, P.M., Sapiro, G., 2011. A Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography. *Med. Image Anal.* 15 (4), 414–425.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12 (1), 26–41.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66 (1), 259–267.
- Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., 2000. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44 (4), 625–632.
- Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50 (5), 1077–1088.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B* 57 (1), 289–300.
- Bodini, B., Khaleeli, Z., Cercignani, M., Miller, D.H., Thompson, A.J., Ciccarelli, O., 2009. Exploring the relationship between white matter and gray matter damage in early primary progressive multiple sclerosis: an in vivo study with TBSS and VBM. *Hum. Brain Mapp.* 30 (9), 2852–2861.
- Brouwer, R.M., Mandl, R.C.W., Peper, J.S., van Baal, G.C.M., Kahn, R.S., Boomsma, D.I., Hulshoff Pol, H.E., 2010. Heritability of DTI and MTR in nine-year-old children. *Neuroimage* 53 (3), 1085–1092.
- Brun, A., Knutsson, H., Park, H.J., Shenton, M.E., Westin, C.F., 2004. Clustering fiber traces using normalized cuts. 7th International Conference on Medical Image Computing and Computer-Assisted Intervention LNCS 3216, pp. 368–375.
- Catani, M., Allin, M.P., Husain, M., Pugliese, L., Mesulam, M.M., Murray, R.M., Jones, D.K., 2007. Symmetries in human brain language pathways correlate with verbal recall. *Proc. Natl. Acad. Sci.* 104 (43), 17163–17168.
- Chiang, M.C., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Hickie, I., Toga, A.W., Wright, M.J., Thompson, P.M., 2011. Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. *Neuroimage* 54 (3), 2308–2317.
- Chou, Y.Y., Lepore, N., de Zubicaray, G.I., Rose, S.E., Carmichael, O.T., Becker, J.T., Toga, A.W., Thompson, P.M., 2007. Automatic 3D mapping & analysis of the lateral ventricles using fluid registration of multiple labeled atlases. 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington, D.C., USA, pp. 1288–1291.
- Chou, Y.Y., Lepore, N., de Zubicaray, G.I., Carmichael, O.T., Becker, J.T., Toga, A.W., Thompson, P.M., 2008. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. *Neuroimage* 40 (2), 615–630.
- Chou, Y.Y., Lepore, N., Chiang, M.C., Avedissian, C., Barysheva, M., McMahon, K.L., de Zubicaray, G.I., Meredith, M., Wright, M.J., Toga, A.W., Thompson, P.M., 2009. Mapping genetic influences on ventricular structure in twins. *Neuroimage* 44 (4), 1312–1323.
- Colby, J.B., Soderberg, L., Lebel, C., Dinov, I.D., Thompson, P.M., Sowell, E.R., 2011. Along-tract statistics allow for enhanced tractography analysis. *Neuroimage* 59 (4), 3227–3342.
- Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E., 1999. Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci.* 96 (18), 10422–10427.
- Corouge, I., Fletcher, P.T., Joshi, S., Gouttard, S., Gerig, G., 2006. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Med. Image Anal.* 10 (5), 786–798.
- Daianu, M., Jahanshad, N., Nir, T.M., Toga, A.W., Jack Jr., C.R., Weiner, M.W., Thompson, P.M., 2013. Alzheimer's disease neuroimaging initiative. Breakdown of brain connectivity between normal aging and Alzheimer's disease: a structural k-core network analysis. *Brain Connect.* 3 (4), 407–422.
- Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R., 2007. Regularized, fast, and robust analytical Q-ball imaging. *Magn. Reson. Med.* 58 (3), 497–510.
- Ding, Z., Gore, J.C., Anderson, A.W., 2003. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Magn. Reson. Med.* 49 (4), 716–721.
- Falconer, D., Mackay, T.F., 1996. Introduction to Quantitative Genetics, 4th ed., Benjamin Cummings.
- Gerig, G., Gouttard, S., Corouge, S., 2004. Analysis of brain white matter via fiber tract modeling. 26th Annual International Conference of IEEE Engineering in Medicine and Biology Society, 6, pp. 4421–4424.
- 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 (4), 1083–1099.
- Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* 33 (1), 115–126.
- Hofer, S., Frahm, J., 2006. Topography of the human corpus callosum revisited – Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *Neuroimage* 32, 989–994.
- Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakonarson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural connectome of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 111 (2), 823–828.
- Jahanshad, N., Aganj, I., Lenglet, C., Joshi, A., Jin, Y., Barysheva, M., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Wright, M.J., Toga, A.W., Sapiro, G., Thompson, P.M., 2011. Sex differences in the human connectome: 4-Tesla high angular resolution diffusion imaging (HARDI) tractography in 234 young adult twins. 14th Congress of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, pp. 939–943.
- Jahanshad, N., Valcour, V.G., Nir, T.M., Kohannim, O., Busovaca, E., Nicolas, K., Thompson, P.M., 2012. Disrupted brain networks in the aging HIV+ population. *Brain Connect.* 2 (6), 335–344.
- Jahanshad, N., Kochunov, P.V., Sprooten, E., Mandl, R.C., Nichols, T.E., Almasy, L., Blangero, J., Brouwer, R.M., Curran, J.E., de Zubicaray, G.I., Duggirala, R., Fox, P.T., Hong, L.E., Landman, B.A., Martin, N.G., McMahon, K.L., Medland, S.E., Mitchell, B.D., Olvera, R.L., Peterson, C.P., Starr, J.M., Sussmann, J.E., Toga, A.W., Wardlaw, J.M., Wright, M.J., Hulshoff Pol, H.E., Bastin, M.E., McIntosh, A.M., Deary, I.J., Thompson, P.M., Glahn, D.C., 2013. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage* 81, 455–469.
- Jahanshad, N., Kochunov, P., Nichols, T.E., Sprooten, E., Mandl, R.C., Almasy, L., Brouwer, R.M., Curran, J.E., de Zubicaray, G.I., Dimitrova, R., Fox, P.T., Hong, L.E., Landman, B.A., Lemaitre, H., Lopez, L., Martin, N.G., McMahon, K.L., Mitchell, B.D., Olvera, R.L., Peterson, C.P., Starr, J.M., Sussmann, J.E., Toga, A.W., Wardlaw, J.M., Wright, M.J., Wright, S.N., Bastin, M.E., McIntosh, A.M., Boomsma, D.I., Kahn, R.S., Braber, A., Deary, I.J., Hulshoff Pol, H.E., Williamson, D., Blangero, J., van 't Ent, D., Glahn, D.C., Thompson, P.M., 2014. Combining meta- and mega-analytic approaches for multi-site diffusion imaging based genetic studies: from the ENIGMA-DTI working group.

- 11th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Beijing, China, pp. 1234–1238.
- Jin, Y., Shi, Y., Jahanshad, N., Aganj, I., Sapiro, G., Toga, A.W., Thompson, P.M., 2011a. 3D elastic registration improves HARDI-derived fiber alignment and automated tract clustering. 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, pp. 822–826.
- Jin, Y., Shi, Y., Joshi, S.H., Jahanshad, N., Zhan, L., de Zubicaray, G.I., McMahon, K.L., Martin, N.G., Wright, M.J., Toga, A.W., Thompson, P.M., 2011b. Heritability of white matter fiber tract shapes: a HARDI study of 198 twins. *Multimodal Brain Image Analysis Workshop at 14th International Conference on Medical Image Computing and Computer-Assisted Intervention LNCS 7012*, pp. 35–43.
- Jonasson, L., Hagmann, P., Thiran, J.P., Wedeen, V.J., 2004. Fiber tracts of high angular resolution diffusion MRI are easily segmented with spectral clustering. 13th Annual Meeting of International Society for Magnetic Resonance in Medicine, Miami, FL, USA.
- Kochunov, P., Glahn, D.C., Nichols, T.E., Winkler, A.M., Hong, E.L., Holcomb, H.H., Stein, J.L., Thompson, P.M., Curran, P.E., Carless, M.A., Olvera, R.L., Johnson, M.P., Cole, S.A., Kochunov, V., Kent, J., Blangero, J., 2011. Genetic analysis of cortical thickness and fractional anisotropy of water diffusion in the brain. *Front. Neurosci.* 5 (article 120).
- Lazar, M., Weinstein, D.M., Tsuruda, J.S., Hasan, K.M., Arfanakis, K., Meverand, M.E., Badie, B., Rowley, H.A., Haughton, V., Field, A., Alexander, A.L., 2003. White matter tractography using diffusion tensor deflection. *Hum. Brain Mapp.* 18 (4), 306–321.
- LeBihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* 161 (2), 401–407.
- Lee, A.D., Lepore, N., de Leeuw, J., Brun, C.C., Barysheva, M., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Wright, M.J., Thompson, P.M., 2010. Multivariate variance-components analysis in DTI. 13th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, The Netherlands, pp. 1157–1160.
- Lepore, N., Brun, C.C., Chou, Y.Y., Lee, A.D., Barysheva, M., de Zubicaray, G.I., Meredith, M., McMahon, K., Wright, M.J., Toga, A.W., Thompson, P.M., 2008. Multi-atlas tensor-based morphometry and its application to a genetic study of 92 twins. *Mathematical Foundations of Computational Anatomy at 11th International Conference on Medical Image Computing and Computer-Assisted Intervention*, New York, NY, USA.
- Li, H., Xue, Z., Guo, L., Liu, T., Hunter, J., Wong, S.T., 2010. A hybrid approach to automatic clustering of white matter fibers. *Neuroimage* 49 (2), 1249–1258.
- Lötjönen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., Alzheimer's Disease Neuroimaging Initiative, 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49 (3), 2352–2365.
- Maddah, M., Grimson, W.E.L., Warfield, S.K., Wells, W.M., 2008. A unified framework for clustering and quantitative analysis of white matter fiber tracts. *Med. Image Anal.* 12 (2), 191–202.
- Merboldt, K.D., Hanicke, W., Frahm, J., 1985. Self-diffusion NMR imaging using stimulated echoes. *J. Magn. Reson.* 64 (3), 479–486.
- Mori, S., Crain, B.J., Chacko, V.P., Van Zijl, P.C.M., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45 (2), 265–269.
- O'Donnell, L.J., Westin, C.F., 2007. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging* 26 (11), 1562–1575.
- O'Donnell, L.J., Kubicki, M., Shenton, M.E., Dreusicke, M.H., Grimson, W.E., Westin, C.F., 2006. A method for clustering white matter fiber tracts. *Am. J. Neuroradiol.* 25 (5), 1032–1036.
- Oishi, K., Faria, A., Jiang, H., Li, X., Akhter, K., Zhang, J., Hsu, J.T., Miller, M.I., van Zijl, P.C., Albert, M., Lyketsos, C.G., Woods, R., Toga, A.W., Pike, G.B., 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 (2), 486–499.
- Parker, G.J.M., Alexander, D.C., 2003. Probabilistic Monte Carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. *Inf. Process. Med. Imaging LNCS 2732* 684–695.
- Parker, G.J.M., Haroon, H.A., Wheeler-Kingshott, C.A.M., 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 (2), 242–254.
- Prasad, G., Jahanshad, N., Aganj, I., Lenglet, C., Sapiro, G., Toga, A.W., Thompson, P.M., 2011. Atlas-based fiber clustering for multi-subject analysis of high angular resolution diffusion imaging tractography. 14th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, pp. 276–280.
- Price, G., Cercignani, M., Parker, G.J., Altmann, D.R., Barnes, T.R., Barker, G.J., Joyce, E.M., Ron, M.A., 2007. Abnormal brain connectivity in first-episode psychosis: a diffusion MRI tractography study of the corpus callosum. *Neuroimage* 35 (2), 458–466.
- Price, G., Cercignani, M., Parker, G.J., Altmann, D.R., Barnes, T.R., Barker, G.J., Joyce, E.M., Ron, M.A., 2008. White matter tracts in first-episode psychosis: a DTI tractography study of the uncinate fasciculus. *Neuroimage* 39 (3), 949–955.
- Rohlfing, T., Russakoff, D.B., Maurer Jr., C.R., 2004. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans. Med. Imaging* 23 (8), 983–994.
- Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imaging* 29 (10), 1714–1729.
- Smith, S.M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T.E., Miller, K.L., Robson, M.D., Jones, D.K., Klein, J.C., Bartsch, A.J., Behrens, T.E., 2007. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nat. Protoc.* 2 (3), 499–503.
- Taylor, D.G., Bushell, M.C., 1985. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Phys. Med. Biol.* 30 (4), 345–349.
- Thomson, M.E., Thompson, P.M., 2011. Diffusion imaging, white matter, and psychopathology. *Annu. Rev. Clin. Psychol.* 7, 63–85.
- Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 23 (3), 1176–1185.
- Tsai, A., Westin, C.F., Hero, A.O., Willsky, A.S., 2007. Fiber tract clustering on manifolds with dual rooted-graphs. *IEEE International Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, pp. 1–6.
- Tuch, D., 2004. Q-ball imaging. *Magn. Reson. Med.* 52 (6), 1358–1372.
- Visser, E., Nijhuis, E.H., Buitelaar, J.K., Zwiers, M.P., 2011. Partition-based mass clustering of tractography streamlines. *Neuroimage* 54 (1), 303–312.
- Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., 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 (3), 630–644.
- Wang, Q., Yap, P.T., Wu, G., Shen, D., 2012. Application of neuroanatomical features to tractography clustering. *Hum. Brain Mapp.* <http://dx.doi.org/10.1002/hbm.22051>.
- Wassermann, D., Descoteaux, M., Deriche, R., 2008. Diffusion maps clustering for magnetic resonance q-ball imaging segmentation. *Int. J. Biomed. Imaging* 2008, 526906. <http://dx.doi.org/10.1155/2008/526906>.
- 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 (1), 228–241.
- Wedeen, V.J., Hagmann, P., Tseng, W.Y., Reese, T.G., Weisskoff, R.M., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn. Reson. Med.* 54 (6), 1377–1386.
- Xia, Y., Turken, A.U., Whitfield-Gabrieli, S.L., Gabrieli, J.D., 2005. Knowledge-based classification of neuronal fibers in entire brain. 8th International Conference on Medical Image Computing and Computer-Assisted Intervention LNCS 3749, pp. 205–212.
- Yushkevich, P.A., Zhang, H., Simon, T.J., Gee, J.C., 2008. Structure-specific statistical mapping of white matter tracts. *Neuroimage* 41 (2), 448–461.
- Zhan, L., Leow, A.D., Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Harel, N., Toga, A.W., Thompson, P.M., 2011. Differential information content in staggered multiple shell HARDI measured by the tensor distribution function. 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, pp. 305–309.
- Zhang, Y., Zhang, J., Oishi, K., Faria, A.V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G.B., Evans, A., Toga, A.W., Woods, R., Mazziotta, J.C., Miller, M.I., van Zijl, P.C.M., Mori, S., 2010. Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *Neuroimage* 52 (4), 1289–1301.