



Published in final edited form as:

Neuroimage. 2021 November 15; 242: 118451. doi:10.1016/j.neuroimage.2021.118451.

Fiber tractography bundle segmentation depends on scanner effects, vendor effects, acquisition resolution, diffusion sampling scheme, diffusion sensitization, and bundle segmentation workflow

Kurt G. Schilling^{a,b}, Chantal M.W. Tax^c, Francois Rheault^d, Colin Hansen^d, Qi Yang^d, Fang-Cheng Yeh^e, Leon Cai^d, Adam W. Anderson^{a,b,f}, Bennett A. Landman^{a,b,d,f}

^aDepartment of Radiology & Radiological Science, Vanderbilt University Medical Center, Nashville, TN, United States

^bVanderbilt Institute of Imaging Science, Vanderbilt University Medical Center, Nashville, TN, United States

^cCardiff University Brain Research Imaging Centre (CUBRIC), Cardiff University, Cardiff, United Kingdom

^dDepartment of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, United States

^eDepartment of Neurological Surgery, University of Pittsburgh, United States

^fDepartment of Biomedical Engineering, Vanderbilt University, Nashville, TN, United States

Abstract

When investigating connectivity and microstructure of white matter pathways of the brain using diffusion tractography bundle segmentation, it is important to understand potential confounds and sources of variation in the process. While cross-scanner and cross-protocol effects on diffusion microstructure measures are well described (in particular fractional anisotropy and mean diffusivity), it is unknown how potential sources of variation effect bundle segmentation results, which features of the bundle are most affected, where variability occurs, nor how these sources of variation depend upon the method used to reconstruct and segment bundles. In this study, we investigate six potential sources of variation, or confounds, for bundle segmentation: variation

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

kurt.g.schilling.1@vumc.org (K.G. Schilling).

Credit Roles

KS: Conceptualization; Formal analysis; Investigation; Methodology; Writing;

CT: Conceptualization; Data curation; Writing

FR: Investigation; Methodology; Software

CH: Investigation; Methodology; Data curation

QY: Investigation; Methodology

FY: Conceptualization; Investigation; Methodology; Writing; Software

LC: Investigation; Methodology; Data curation

AA: Conceptualization; Methodology

BL: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision

(1) across scan repeats, (2) across scanners, (3) across vendors (4) across acquisition resolution, (5) across diffusion schemes, and (6) across diffusion sensitization. We employ four different bundle segmentation workflows on two benchmark multi-subject cross-scanner and cross-protocol databases, and investigate reproducibility and biases in volume overlap, shape geometry features of fiber pathways, and microstructure features within the pathways. We find that the effects of acquisition protocol, in particular acquisition resolution, result in the lowest reproducibility of tractography and largest variation of features, followed by vendor-effects, scanner-effects, and finally diffusion scheme and b-value effects which had similar reproducibility as scan-rescan variation. However, confounds varied both across pathways and across segmentation workflows, with some bundle segmentation workflows more (or less) robust to sources of variation. Despite variability, bundle dissection is consistently able to recover the same location of pathways in the deep white matter, with variation at the gray matter/ white matter interface. Next, we show that differences due to the choice of bundle segmentation workflows are larger than any other studied confound, with low-to-moderate overlap of the same intended pathway when segmented using different methods. Finally, quantifying microstructure features within a pathway, we show that tractography adds variability over-and-above that which exists due to noise, scanner effects, and acquisition effects. Overall, these confounds need to be considered when harmonizing diffusion datasets, interpreting or combining data across sites, and when attempting to understand the successes and limitations of different methodologies in the design and development of new tractography or bundle segmentation methods.

Keywords

Tractography; Bundle segmentation; White matter; Reproducibility; Harmonization

1. Introduction

Diffusion-weighted magnetic resonance imaging (dMRI) has proven valuable to characterize tissue microstructure in health and disease (Alexander et al., 2019; Jones et al., 2018; Novikov et al., 2018). Moreover, the use of dMRI fiber tractography to virtually dissect fiber pathways (Jeurissen et al., 2019) is increasingly used to localize microstructure measurements to specific white matter bundles (Raffelt et al., 2017; Chamberland et al., 2019; Yeatman et al., 2012), and to study the connections and shapes of pathways (Jeurissen et al., 2019; Maffei et al., 2019; Forkel et al., 2014; Hau et al., 2017; Hau et al., 2016; Sarubbo et al., 2019; Sarubbo et al., 2013; Neubert et al., 2015; Neubert et al., 2014; Mars et al., 2012). Despite promises of noninvasive measurements of white matter features, variability may exist in measurements due to inherent variability within scanners and across scanners, differences in acquisition protocol parameters, and differences due to processing pipelines, amongst others. These sources of variance challenge the quantitative nature of derived measures of microstructure and connectivity, and hinder the ability to interpret different findings or combine different datasets.

These effects have been intensively studied for tissue microstructure features, specifically diffusion tensor imaging (DTI) (Pierpaoli et al., 1996) indices of fractional anisotropy (FA) and mean diffusivity (MD). Numerous studies have characterized intra-scanner and

inter-scanner DTI variability (AK Prohl et al., 2019; Mirzaalian et al., 2016; VA Magnotta et al., 2012; Landman et al., 2011; Teipel et al., 2011; Vollmar et al., 2010; Farrell et al., 2007; Landman et al., 2007; Heiervang et al., 2006; Pfefferbaum et al., 2003; Jones, 2003; Lori et al., 2002), differences due to acquisition parameters (Farrell et al., 2007; Landman et al., 2007; Jones, 2003; Jones et al., 2020; Papinutto et al., 2013; Jones, 2004; Jones and Basser, 2004) including image resolution, number of diffusion images, and diffusion sensitization (i.e. the b-value), and differences due to processing and algorithmic choices (Jones et al., 2007; Chang et al., 2005). These have paved the way towards recommendations and guidelines for reliable and reproducible DTI (Jones et al., 2013; Jones, 2010; Jones and Cercignani, 2010; DK Jones et al., 1999); however, a standardized universal dMRI protocol does not exist, and differences are expected across sites and studies (Fig. 1) (L Ning et al., 2020; CM Tax et al., 2019). Yet, there is significant interest in combining data from different sites to increase statistical power and benefit from multi-center recruitment abilities (Mirzaalian et al., 2016; L Ning et al., 2020; Zhong et al., 2020; Cetin Karayumak et al., 2019; KM Huynh et al., 2019; Vishwesh Nath et al., 2018; Yu et al., 2018; Mirzaalian et al., 2018; Fortin et al., 2017), and it is clear that these differences need to be accounted for, or removed, prior to data aggregation or joint statistical analysis.

Notwithstanding the increased awareness and improved characterization of dMRI microstructural measures, very little work has been performed to characterize and understand reproducibility of tractography-derived features across scanners, across protocols, and across different tractography bundle segmentation algorithms (Pestilli et al., 2014; Nath et al., 2019). Variability in tractography estimates of fiber pathways will further increase variability in connectivity analyses and impact microstructural characterization, e.g. when tractography is used to define ROIs or to perform along-tract profiling. While few studies do exist, they are often limited to a single pathway (Chamberland et al., 2018; F Rheault et al., 2020), a single dissection protocol (Vazquez et al., 2020; Guevara et al., 2017), or a single source of potential variation (Schilling et al., 2020), such as test-retest or population-based reproducibility (Guevara et al., 2017; F Zhang et al., 2019; Guevara et al., 2012). Additionally, they do not investigate where in the brain or along the pathway that this variability occurs, and are often limited to characterizing only microstructural features of these pathways (i.e., the FA or MD along or within the pathway) (Heiervang et al., 2006; Wakana et al., 2007). Thus, we currently do not know which sources of variation impact tractography bundle segmentation the most, which features of the bundle are most affected, where variability occurs, nor how these questions are dependent upon the workflow used to dissect fiber bundles. Thus, for the first time, we combine, assess, and rank all previously studied sources of potential variation in the same study, with a focus on tractography rather than just DTI measures.

Here, we investigate and compare the reproducibility of tractography across six confounds, or sources of variation: intrinsic variability across scan repeats, differences across scanners, across vendors, across different acquisition spatial resolution and acquisition angular resolution, and across different diffusion sensitizations (b-values). We employ and examine four fully-automated and commonly utilized bundle reconstruction workflows on two cross-scanner cross-protocol benchmark datasets. We first investigate how these confounds affect not only the overlap and location of pathways, but also evaluate variability in topological

measures of the bundle including length, area, shape, and volume features. We ask which pathways, which bundle segmentation workflow, and which features are most reproducible? And what source of variation is most significant for each method? Second, we visualize *where in the brain, and where within a pathway*, tractography is most variable (and most robust) and investigate if sources of variation effect this in different ways. Third, we quantify and visualize differences in tractography that result when using different bundle segmentation workflows. Finally, we analyze traditional DTI measures and quantify differences due to these sources of variation as well as the *added* variance introduced by the tractography process over and above that inherent across scanners and across acquisition protocols.

2. Methods

2.1. Datasets

Here we utilize two open-sourced multi-subject, multi-scanner, and multi-protocol benchmark databases: the MASiVar (Cai et al., 2020) and MUSHAC datasets (L Ning et al., 2020; CM Tax et al., 2019). We note that other multi-site databases exist (see Discussion), although they are often limited to investigating differences across subjects and scanners, whereas the two chosen datasets together allow investigation of repeats, scanners, vendors, and acquisition protocols (resolution, direction, b-values).

2.1.1. MUSHAC dataset—The MUSHAC database will allow investigation of cross-scanner, cross-protocol, and cross b-value effects (L Ning et al., 2020; CM Tax et al., 2019). This database was part of the 2018 and 2019 MICCAI Harmonization challenge. Here, we utilize the data acquired from 10 healthy subjects used as training data in the challenge, and described in (L Ning et al., 2020; CM Tax et al., 2019). Each subject has 4 unique datasets. This work focuses on the data acquired on two scanners with different gradient strengths: a) 3T Siemens Prisma (80 mT/m), and b) 3T Siemens Connectom (300 mT/m). Two types of protocols were acquired from each scanner: 1) a ‘standard’ protocol with acquisition parameters matched to a typical clinical protocol; and 2) a more advanced or ‘state-of-the-art’ protocol where the superior hardware and software specifications were utilized to increase the number of acquisitions and spatial resolution per unit time. The ‘standard’ protocol from both scanners is matched as closely as possible, with an isotropic resolution of 2.4 mm, TE = 89 ms and TR = 7.2 s, and 30 diffusion-weighted directions acquired at two b-values: $b = 1200, 3000$ s/mm² (scan time ~7.5 min). On the other hand, the Prisma ‘state-of-the-art’ data has a higher isotropic resolution of 1.5 mm, TE = 80 ms, TR = 7.1 s and 60 directions at the same b-values (~14.5 min). While the Connectom ‘state-of-the-art’ data has the highest resolution of 1.2 mm with TE = 68 ms, TR = 5.4 s and 60 directions (~11 min). All data was corrected for distortions, motion, eddy currents (Andersson et al., 2003), and gradient nonlinearity distortions (Glasser et al., 2013). For each subject, the Prisma standard-acquisition dataset was used as a reference space and all additional datasets were affinely registered to this space using the corresponding FA maps with FSL Flirt with appropriate b-vector rotation.

2.1.2. MASiVar dataset—The MASiVar database will allow investigation of scan-rescan and cross-scanner effects. Here we used a subset of Cohort II of this database described in (Cai et al., 2020), which consisted of 5 healthy subjects with 6 unique “datasets”. Each subject was scanned on four scanners: a) 3T Philips Achieva (80 mT/m) and b) a different 3T Philips Achieva (60mT/m) at the same site, c) a 3T General Electric Discovery MR750 Scanner at a different site, and d) a 3T Siemens Skyra scanner at a different site. These acquisitions were matched as closely as possible and are similar to that of the standard-protocol described above: with an isotropic resolution of 2.5 mm, TE = 55 ms and TR = 6.2 s (7.0 s for scanner-b), and 32 diffusion-weighted directions acquired at $b = 1000\text{s/mm}^2$ (scan time ~ 3.5 min). Additionally, the subjects were scanned twice on the first scanner, and also had an acquisition that consisted of a 96-direction $b = 1000$ dataset, both of which were also utilized in the current study. We note that one subject did not have a repeat scan on the first scanner (a) and one subject did not have a scan on the GE Scanner (b).

All data were corrected for distortions, motion, and eddy currents (Andersson et al., 2003; Cai et al., 2021). For each subject, the first session on scanner-a was used as a reference space and all additional datasets were affinely registered to this space using the corresponding FA maps with FSL Flirt (Jenkinson et al., 2012) with appropriate b-vector rotation.

2.2. Sources of variation

We investigate several possible sources of variation in the bundle segmentation process.

RESCAN: the effects of repeating a scan on the same scanner (i.e. scan-rescan) in a different session, but with a matched acquisition. This effect is quantified using the repeated acquisitions from the MASiVar database.

SCAN1: inter-scanner (cross-scanner) effects, with a matched acquisition and of the same vendor. SCAN1 is quantified using the matched acquisitions from the MASiVar database acquired on different Philips scanners (both Philips Achieva).

SCAN2: inter-scanner (cross-scanner) effects, with a matched acquisition and of the same vendor. SCAN2 is quantified using the matched standard acquisitions from the MUSHAC acquired on different Siemens scanners (Siemens Connectom and Siemens Prisma).

VEN1: inter-vendor (cross-vendor) effects, with a matched acquisition. VEN1 is quantified using the matched acquisitions from the MASiVar database, but acquired on scanners from different vendors (Philips Achieva and General Electric Discovery).

VEN2: inter-vendor (cross-vendor) effects, with a matched acquisition. VEN2 is quantified using the matched acquisitions from the MASiVar database, but acquired on scanners from different vendors (Philips Achieva and Siemens Skyra).

RES1: effects of spatial resolution, with matched scanner, diffusion directions, and b-value. RES1 is quantified by using the MUSHAC acquisitions from the Prisma standard-acquisition and from the Prisma state-of-the-art acquisition but with only 30 uniformly

distributed directions utilized (to match the standard-acquisition). This represents differences between a 2.4 mm isotropic and 1.5 mm isotropic acquisition.

RES2: effects of spatial resolution, with matched scanner, diffusion directions, and b-value. RES2 is quantified by using the MUSHAC acquisitions from the Connectom standard-acquisition and from the Connectom state-of-the-art acquisition but with only 30 uniformly distributed directions utilized (to match the standard-acquisition). This represents differences between a 2.4 mm isotropic and 1.2 mm isotropic acquisition.

DIR1: effects of number of diffusion-weighted directions, with matched scanner, resolution, and b-value. DIR1 is quantified using the MASIvar acquisitions from the first scanner at 32 directions and the acquisition on the same scanner at 96 directions.

DIR2: effects of number of diffusion-weighted directions, with matched scanner, resolution, and b-value. DIR2 is quantified using the MUSHAC acquisitions from the state-of-the-art Prisma acquisition with only 30 uniformly distributed directions utilized and the full state-of-the-art acquisition which consists of 60 directions.

BVAL: effects of changing the b-value, on the MUSHAC Prisma scanner with the ‘standard’ protocol, from $b = 1200$ to $b = 3000$, within the same acquisition.

We note that we also investigated a second effect of b-value (within the state-of-the-art Prisma protocol, with no statistically significant differences, and for figure simplicity only show the above-mentioned b-value analysis). Previous version of this manuscript (and preprint) included an ACQ1 and ACQ2 (from state-of-the-art to standard-acquisition) that were isolated into both effects of directions (DIR1 and DIR2) and resolution (RES1 and RES2).

A final source of variation investigated is that caused by the use of different bundle reconstruction workflows. Because all workflows segment different numbers of, and sets of, fiber pathways (see below), for this analysis, we investigated only those fiber pathways which are common to all algorithms. In this case, we identified 7 (bilateral) pathways which are segmented by all automated methods.

2.3. Tractography bundle dissection

We utilized four common, well-validated, and fully-automated fiber bundle reconstruction workflows, all implemented using standard and/or recommended settings. It is important to highlight that each workflow included differences in local fiber-direction estimation, fiber tractography, and bundle segmentation algorithms, and our attempt was to implement the entire workflow as would be done in a typical scientific study (see Discussion on limitations of confounds due to differences in bundle segmentation process). While there are dozens of bundle segmentation algorithms, we have chosen these to be representative of common approaches, utilizing regions of interest, atlases, machine learning, templates, etc. (see Discussion and Limitations).

2.3.1. TractSeg—TractSeg is based on convolutional neural networks and performs bundle-specific tractography based on a field of estimated fiber orientations (Wasserthal

et al., 2019; J Wasserthal et al., 2018; J Wasserthal et al., 2018). We implemented the dockerized version at (<https://github.com/MIC-DKFZ/TractSeg>), which generates fiber orientations using constrained spherical deconvolution with the MRtrix3 software (Tournier et al., 2019). We note that different reconstruction methods could have been chosen to generate fiber orientations. This method dissects 72 bundles.

2.3.2. Automatic fiber tractography (ATK)—ATK was performed in DSI Studio software using batch automated fiber tracking (Yeh, 2020). Data were reconstructed using generalized q-sampling imaging (Yeh et al., 2010) with a diffusion sampling length ratio of 1.25. 20 white matter pathways were automatically reconstructed using seeding regions defined in the HCP842 tractography atlas (Yeh et al., 2018), randomly generated tracking parameters of anisotropy threshold, angular threshold, step size, and subsequent segmentation and pruning. The Dockerized source code is available at <http://dsi-studio.labsolver.org>.

2.3.3. Recobundles (RECO)—Recobundles (Garyfallidis et al., 2018) segments streamlines based on their shape-similarity to a dictionary of expertly delineated model bundles (Yeh et al., 2018). Recobundles was run using DIPY (Garyfallidis et al., 2014) software (<https://dipy.org>) after performing whole-brain tractography using spherical deconvolution and DIPY LocalTracking algorithm. The bundle-dictionary contains 80 bundles, but only 44 were selected to be included in this study after consulting with the algorithm developers based on internal quality assurance (for example, removing cranial nerves which are often not used in brain imaging). Of note, Recobundles is a method to automatically extract and recognize bundles of streamlines using prior bundle models, and the implementation we chose uses the DIPY bundle dictionary (Yeh et al., 2018) for extraction, although others can be used, as well as alternative shape-similarity filtering criteria.

2.3.4. Xtract—Xtract (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/XTRACT>) is a recent automated method for probabilistic tractography based on carefully selected inclusion, exclusion, and seed regions, defined in a standard space (Warrington et al., 2020). Xtract used the ball-and-stick model of diffusion from FSL's bedpostx algorithm (Jenkinson et al., 2012), in combination with a probabilistic tractography algorithm probtrackx, to reconstruct 42 white matter pathways. In contrast to the preceding methods, which result in streamlines, this method results in visitation count maps for each pathway.

A list of all segmentations generated from each method and corresponding acronyms is given in the appendix. The 7 pathways identified to be common to all tractography bundle segmentation techniques includes: arcuate fasciculus (AF), corticospinal tract (CST), inferior fronto-occipital fasciculus (IFO), inferior longitudinal fasciculus (ILF), middle longitudinal fasciculus (MdLF), optic radiations (OR), and uncinata fasciculus (UF), all of which are bilateral including left (_L) and right (_R) hemisphere pathways.

A thorough quality control was performed for all subjects, and for all pathways. This included first visualization and verification of adequate alignment of all FA maps (to ensure appropriate quantification of overlap measures). Second, all pathways, for a subjects, were

visualized in mosaic form using tools from the SCILPY tool-box (<https://github.com/scilus/scilpy>), and pathways were visually assessed and removed from analysis if deemed in the incorrect location or shape. Finally, individual bundles were removed from analysis if the number of segmented streamlines was less than 3 standard deviations away from the mean number (for each pathway), or if the total number of streamlines was below 200 (indicating failure of tractography), and subjects were removed from analysis (for a given algorithm) if > 20% of pathways failed QC.

2.4. Feature extraction

A number of features were extracted from each bundle segmented. First, for simple comparisons of the volume occupied by each pathway, all bundles (from all methods) were binarized and resampled at 1 mm isotropic resolution. For methods generating streamlines (Tractseg, ATK, and RECO) this is equivalent to binarizing based on a streamline density of 1. Because Xtract output is in the form of a normalized probability distribution, where a threshold of $2.5E-4$ was chosen based on (Warrington et al., 2020). The binarized segmentation was used for measures of Dice overlap (described below).

Second, several descriptors of the shape and geometry of the bundles were extracted. Shape analysis was performed using DSI Studio, and made available as matlab code (https://github.com/dmitrishastin/tractography_shapes/), based on (Yeh, 2020), to derive length, area, volume, and shape metrics of a bundle. Briefly, length features include mean length, span, diameter, and average radius of end regions. Area features include total surface area and the total area of end regions. Volume features include total volume, trunk volume, and branch volume. Shape features include pathway curl, elongation, and irregularity.

Finally, microstructure measures of FA and MD (calculated using iteratively reweighted linear least squares estimator) within pathways were extracted. In all cases, a simple measure of the average value within the binary volume was performed, although we note that these measures can also be weighted by certainty or streamline density. To isolate the added variation due to tractography from that of the existing sources of variation, these measures were extracted in two ways. First, using the binary regions defined in the reference scan-space only (i.e., the Prisma standard-acquisition and first session on scanner-a for MUSHAC and MASiVar datasets respectively) were used as the *same* region-of-interest across all effects, in order to isolate each source of variation while keeping ROIs constant. Second, the binary region defined by tractography for each specific dataset was used to extract the average FA (or MD), which includes both variation due to the effect under investigation and the variation due to tractography differences.

2.5. Reproducibility evaluation

Reproducibility was evaluated using several metrics, and across each source of variation. First, the Dice overlap was calculated for each pair of bundles as an overall measure of similarity of volumes. The Dice overlap is calculated as two times the intersection divided by the sum of the volumes of each dataset. Results were displayed across all fiber pathways for a given source of variation, and differences between effects were calculated using the nonparametric paired (i.e. same subject, different effect) Wilcoxon signed rank tests.

Differences in scalar shape features are calculated as the mean absolute percentage error (MAPE), sometimes referred to as the mean absolute percentage deviation. For two different scans, this measure is calculated as the difference divided by the mean, and can be converted to a percentage error by multiplication by 100. This measure was calculated over all subjects, and results were displayed across all fiber pathways for a given source of variation. Differences between effects were again calculated using the nonparametric paired (i.e. same subject, different effect) Wilcoxon signed rank tests.

For visual comparisons only, all subjects were nonlinearly registered to MNI space, using the 1 mm isotropic FA template and the corresponding FA maps with FSL FLIRT + FNIRT. Streamlines were directly warped to this space for visualization of agreement/disagreement across the cohort. Note that quantification of shape features was performed in native space prior to warping.

For all statistical analysis, thresholds were corrected for multiple comparisons. For example, when investigating differences in effects of DICE/MAPE, etc., we tested differences between 10 effects, resulting in 55 tests performed for each analysis.

3. Results

3.1. Qualitative variation

Fig. 1 shows FA maps of the same subject, but acquired on different scanners and with different protocols. In agreement with the literature (L Ning et al., 2020; CM Tax et al., 2019), differences in magnitude, contrast, and signal-to-noise ratios are readily apparent, and dMRI measures qualitatively vary due to scanner and acquisition effects.

Fig. 2 shows tractography bundle segmentation results for an example pathway (the arcuate fasciculus; AF) on a single subject, for two scanners, two protocols, two b-values, and all four reconstruction methods. For a given bundle segmentation method, minor differences are observed in individual gyri and at regions of low streamline density. However, bundles are visually very similar across scanners and protocols, with similar shapes, locations, curvatures, and connections. Most notably, and as expected (Schilling et al., 2020), the biggest differences are observed when comparing the same pathway across different bundle segmentation methods.

3.2. Quantitative variation due to rescan, scanner, vendor, resolution, directions, and b-value effects

The effects of RESCAN, SCAN1, SCAN2, VEN1, VEN2, DIR1, DIR2, and BVAL on Dice overlap coefficient is shown in Fig. 3 for fourteen selected pathways common to all bundle segmentation methods. Notably, reproducibility is most dependent on the bundle dissection method, with TractSeg consistently resulting in high reproducibility for all sources of variation. Within a method, most pathways show similar patterns of reproducibility. For example, for TractSeg and Xtract all pathways indicate high RESCAN, DIR(1 and 2) and BVAL reproducibility, but are most sensitive to RES, with RES2 showing more variation than RES1. Additionally, Dice overlap shows some variation across pathways, for example

CST and UF generally have higher overlap than OR, IFO, and AF, although trends are different for different workflows.

The results of the Dice overlap coefficient-analysis for each method is shown in Fig. 4, but condensed across all pathways within a given bundle segmentation method. Similar trends are observed as in Fig. 3, with TractSeg consistently indicating the highest Dice overlap, and all methods indicating moderate-to-good overall overlap for most pathways. In general, the largest differences are observed when changing resolution, with changes due to RES2 resulting in larger differences than RES1. Following this, differences across vendors (VEN1 more different than VEN2 comparisons) are greater than across scanners (for both SCAN1 and SCAN2), which are greater than the inherently stochastic nature of RESCAN variability. Finally, differences caused by DIR (1 and 2) and BVAL are on the level of, or even less than, those caused by RESCAN, with the notable exception of ATK, which utilizes a reconstruction method and tractography propagation inherently dependent on diffusion sensitization.

3.3. Localization of variation

Fig. 5 visualizes locations of tractography bundle segmentation agreement (or consistency), and where it disagrees (variability) as hot and cold colormaps, respectively. Agreement and disagreement are averaged across all subjects and shown for all sources of variation. For display, we have chosen an example pathway that is highly reproducible (the AF from TractSeg) and one which displayed lower reproducibility (the SLFII from Xtract). For the highly reproducible pathway, all sources of variation show very similar results. The agreement is very high throughout the entire pathway (hot colors), and percent-disagreement remains fairly low (black and dark blue colors). This means that when two bundles disagree, the disagreement is largely randomly distributed, rather than a *consistent* localized bias introduced by a certain source of variation – an effect which would show up as a consistent disagreement (i.e. a high percent-disagreement). Disagreement tends to occur at the periphery, or boundaries, of the pathway, in particular at the gray-white matter junction, and within individual gyri.

For the less reproducible pathway, the agreement is moderate to high in the dense core, or center, of the pathway in the deep white matter. Again, disagreements are at the edges, and prominent at the white matter and gray matter boundary. However, even though disagreement is more noticeable, the percent-disagreement remains low, indicating random disagreement as opposed to a consistent bias in the spatial location of this pathway. In this case, sources of variation from SCAN2 and RES2 and VEN1 are more noticeable as a larger source of variation, in agreement with quantitative results.

3.4. Variation of shape features

Fig. 6 shows the RESCAN reproducibility of shape features as measured by MAPE, for all features and all pathways, visualized in decreasing reproducibility. In agreement with Dice, TractSeg has higher overall reproducibility, with most features and most pathways below 10% MAPE. Similarly, ATK and Reco are able to reproducibly characterize most features of most pathways with high consistency. In general, reproducibility of features

follows similar order across all methods, with features of Curl, Length, Span, and Diameter highly reproducible, and those of surface area, volume, and end area less so. Additionally, reproducibility is highly dependent on pathway, with clear variation depending upon the bundle being analyzed.

Fig. 7 summarizes the MAPE of different features across different sources of variation. Again, Curl, Length, and Span are highly reproducible across all effects, with MAPE always below 10%, and surface area and volume result in higher MAPE. Trends are the same as those observed for Dice overlap, with generally larger differences due to resolution and vendor acquisition effects (RES 1 and 2, VEN 1 and 2), followed by scanner effects (SCAN1 showing the largest variation).

To look for systematic differences introduced in the quantification of features, we calculate the mean percent variation (i.e., the signed value of MAPE), across all sources of variation, for all features (across all bundles). Fig. 8 shows that most effects do not significantly bias bundle shape measures. For example, nearly all features derived from TractSeg are within a 10% variation and largely centered on 0. However, RES2 and VEN2 do introduce a small, but consistent, bias, in measures of surface area, end area, and volume (in this case, the higher resolution results in smaller values). Similarly, for ATK, a bias is observed in the opposite direction for the same features for effects of acquisition resolution. Additionally, b-value introduces a significant bias for ATK, with the higher b-value scan resulting in larger quantitative values for these features. Reco, in agreement with previous figures, has a much wider range of variation, and larger effects due to acquisition for features of Diameter, Surface Area, End Areas and Volume. Thus, different sources of variation may bias quantitative extraction of shape features, and bias them differently for different bundle segmentation methods.

3.5. Variation across bundle segmentation methods

Next, we compared the agreement of the same bundle, but across different bundle segmentation methods. Fig. 9 shows the Dice overlap for 14 common bundles, comparing each method to every other. There is a low-to-moderate agreement, with Dice overlap values between 0.1–0.5 for all pathways. In general, ATK was most similar to TractSeg and Reco for most bundles (with some exceptions), while Xtract was most dissimilar to all others. The AF, ILF, and MDLF, were the most dissimilar across methods.

Fig. 10 visualizes where agreement and disagreement occurs across bundle segmentation methods, with example-pathways AF and OR. Here, while most of the core agrees across methods, there is also a *consistent* disagreement across methods, particularly in the thickness of the bundle and in the regions of the temporal lobe for the AF and connections in the occipital lobe for the OR. Thus, instead of random differences due to noise, differences across methods are reproducible disagreement, likely caused by fundamental differences in the segmentation technique and structure to be segmented.

3.6. Variation in diffusion MRI microstructure measures

We next investigate reproducibility of microstructure measures due to the aforementioned sources of variation, and tractography variation. Fig. 11 shows the MAPE of FA for all

four bundle segmentation methods. In all cases, the standard-color boxplots are variations due to the queried source of variation alone, whereas the darker-shaded boxplots are due to the source of variation *and* the added variation of tractography variation. Most notably, the MAPE due to RESCAN, SCAN, VEN, DIR, and BVAL alone are highly similar for all segmentation methods, with only minor differences due to the slightly different representations of the pathways (Fig. 9). These results are in line with the literature, with variation < 3% for SCAN rescan (Farrell et al., 2007; Landman et al., 2007; Wakana et al., 2007), with 5–15% due to scanner and vendor effects (L Ning et al., 2020; CM Tax et al., 2019), and as much as 10% due to differences in acquisition and diffusion sensitization (Jones and Bassler, 2004; L Ning et al., 2020; Tax et al., 2020). Notably, the added variation due to tractography does indeed increase differences in FA (as indicated by a solid horizontal line) in many cases, although the% increase in variation is on average < 5%.

Fig. 12 shows the MAPE of MD for different sources of variation. Most noticeable, MD is highly different when calculated using two different b-values, as expected (Novikov et al., 2018; Landman et al., 2007; Jones, 2004; De Luca et al., 2021; DK Jones et al., 1999), followed by differences due to vendors. Differences across RESCAN, SCAN, RES, and DIR are typically < 5%. Again, the use of tractography adds to this variance, although on 3% or less on average.

4. Discussion

The primary focus of this work was to study variability of diffusion fiber tractography bundle segmentation, performing the same analysis on different datasets on different scanners or with different acquisition protocols. For the databases investigated here, we have shown that the process of tractography bundle segmentation shows significant variation across different acquisition resolution and across different vendors, with less, albeit significant, variation across scanners and across diffusion sensitization. Variation is indeed expected when scanning the same subject twice, with all other experimental parameters constant, due to imaging noise and the stochastic nature of the tractography process, however, these additional sources of variation add potential confounds to tractography analysis that may bias measurements, limit aggregation of datasets, and hinder direct interpretation and meta-analysis of different results across studies. While the primary focus was on variation due to vendor and scanner effects, acquisition effects, and b-value effects, we also show the most bundle segmentation workflows are highly reproducible when running the same analysis on data acquired in different sessions, but with the same scanner and protocol.

It is well-known that microstructural features at different sites and with different protocols are not immediately comparable, and in fact significantly biased due to various effects. However, the process of tractography is largely dependent upon fiber orientation estimates, rather than features of the signal magnitude directly (i.e., MD/FA), and it is not immediately intuitive that differences in scanners, acquisitions, and b-values may lead to significantly different results. The results of this work suggest that, indeed, the results of tractography and across sites adds variability that must be considered in the interpretation of both microstructural and shape features of these pathways.

4.1. Do we need to harmonize tractography?

“Harmonization” can be considered any effort at reducing variability in quantitative metrics between different databases, scanners, and studies. We have known that the voxel-wise signal varies across sites, scanners, and acquisitions (as evidenced by the multitude of efforts in the literature to study effects on DTI-indices (Alexander et al., 2019; Jones et al., 2018; Novikov et al., 2018; Jeurissen et al., 2019; Raffelt et al., 2017; Chamberland et al., 2019; Yeatman et al., 2012; Maffei et al., 2019; Forkel et al., 2014; Hau et al., 2017; Hau et al., 2016; Sarubbo et al., 2019; Sarubbo et al., 2013; Neubert et al., 2015; Neubert et al., 2014)) and now confirm that the tractography process itself does as well, and have quantified the extent that tractography contributes to variability. The question becomes “do we need to harmonize tractography?”. The short answer is “yes”, the long answer is: harmonizing likely entails both harmonizing the signal (e.g., FA, MD, RISH measures), harmonizing orientation, reducing effects of resolution, and combining the strengths of different bundle segmentation approaches.

The field of diffusion MRI harmonization has grown in recent years, with significant efforts to make diffusion microstructural measures comparable across sites and scanners (Mirzaalian et al., 2016; L Ning et al., 2020; Zhong et al., 2020; Cetin Karayumak et al., 2019; KM Huynh et al., 2019; Mirzaalian et al., 2018; Fortin et al., 2017). Yet, these endeavors have traditionally not considered variability of tractography, which is ultimately influenced at both the local scale of individual voxels and voxel-wise reconstruction as well as a global scale of connecting discrete orientation estimates across the brain.

It is unclear what “harmonizing” tractography may entail. Clearly, consistent orientation estimates are key, but also streamline generation algorithms robust to voxel-sizes, and also segmentation algorithms that are consistently able to identify streamlines belonging to a pathway-of-interest. With the vast array of options to reconstruct orientation, generate streamlines, and segment bundles, it may be impossible to harmonize data in a way that is appropriate for all methods. Some effort has been performed to harmonize fiber orientation estimation specifically across time or across scanners (Vishwesh Nath et al., 2018; Chen et al., 2016; Moyer et al., 2020; KM Huynh et al., 2019). It may be possible that harmonizing the microstructural measures themselves may remove some possible confounds (i.e., if FA is used as a stopping criteria). Similarly, it is possible that the application and process of tractography in a standard space (as performed for XTRACT), or at a standard resolution may remove confounds associated with image resolution. Alternatively, various multi-site methods used for scalar microstructure features, instead of harmonizing bundles of streamlines directly, may be utilized to harmonize features extracted from bundles. Finally, even while there is significant variation, large agreement occurs in the core of reconstructed white matter pathways, and weighting all derived measures and features by tract density, or isolating the trunk of the bundle (Yeatman et al., 2012), may remove sources of variation.

Reassuringly, the automated methods considered are fairly robust to these studied sources of variation. Visually, the pathways look remarkably similar across scanners, acquisition, and protocols (Fig. 2), for all methods. Quantitatively, methods such as TractSeg, which utilize orientation estimates alone, in combination with machine learning techniques in order to map out tract orientation maps, endpoints, and binary segmentations are highly

reproducible. Similarly, the other methods, while quantitatively having moderately larger variation, show similar shapes, locations, and connectivity across all effects. A final possible harmonization approach may be to combine the strengths of the various algorithms, rethinking the process of bundle segmentation to possibly utilize some combination of machine learning (TractSeg), and a volume-based extraction prior to streamline generation, followed by atlas-based (ATK, Xtract), or shape-based filtering (Reco) in order to delineate bundles consistently across potential confounds.

4.2. Which confounds impact tractography the most?

It is important to emphasize that we are purposefully not attempting to “rank” algorithms, or suggest that ones are better than others. Even the methods with apparent lower reproducibility of features and shapes are still moderately robust, and different implementations of these algorithms may have yielded different quantitative values. For example, different thresholding could have been applied to both density-based (Xtract) or streamline-based (all others) methods to increase specificity (or vice-versa, specificity), or different whole-brain tractography could have been applied prior to bundle dissection using Recobundles. However, regardless of implementation and choices of hyperparameters, we expect methods to show similar dependencies to the investigated sources of variation.

To our knowledge, this is the first time that multiple sources of variation of tractography have been investigated together. Reproducibility across raters, across algorithms, and across scanners have previously been investigated. Our results allow comparison of the relative impact of changes across sites or scanners, and suggest that, in general spatial resolution leads to the most dramatic differences in resulting tractograms. Less tissue-based partial volume effects within the white matter may facilitate delineation of white matter bundles (F Rheault et al., 2020). Additionally, when quantifying volume overlap and shape features, voxel-wise partial volume effects may cause a higher (or lower) estimate due to the representation of the bundle as a binary volume at the given spatial resolution. Finally, orientation-based partial volume effects are observed with different spatial resolution (Jones et al., 2020; Schilling et al., 2017), leading to differences in accuracy of fiber orientation distributions, as well as fundamental differences in common diffusion measures such as FA (which are often used in the tracking process).

The second biggest contributor to variability was vendor differences. Differences across scanners are known to introduce variability due to factors of maximum gradient strength (and hence echo times and repetition times), field strengths, gradient nonlinearities, receive coil sensitivities, software version, and system calibration (Mirzaalian et al., 2016). Here, we show that differences in vendors are typically greater than that due to different scanners (yet same vendor) alone. Over and above scanner differences, vendors themselves may variations in algorithm choices, algorithms for acquisition, reconstruction, background noise reduction, multi-coil fusion (Griswold et al., 2002; Pruessmann et al., 1999), and pulse sequence implementation. Here, we have shown that in addition to inconsistencies in DTI measures across vendors consistently shown in previous studies (AK Prohl et al., 2019; VA Magnotta et al., 2012; Min et al., 2018) there is also a large inconsistency in tractography volumes and locations due to differences in vendors.

Reassuringly, variation of b-value and number of diffusion directions led to relatively consistent tractography. While it is well-known that angular resolution affects the ability to reconstruct fiber orientations (Jones et al., 2020; Schilling et al., 2018; Canales-Rodriguez et al., 2018; Tournier et al., 2013; Tournier et al., 2008; Prckovska et al., 2008), most reconstruction methods are robust with as few as 30 directions (or less). Similarly, while reconstruction algorithms are dependent on diffusion sensitization (Schilling et al., 2018; Daducci et al., 2014), the b-value did not significantly affect tractography results (although does affect quantitative metrics association with DTI).

It is also interesting that the relative magnitude of sources of variation depend on the bundle dissection method. While variability generally decreases from RESCAN, DIR, BVAL, SCAN, then VEN and RES, several notable exceptions occur. ATK is highly sensitive to the b-value. This is likely due to the fact that this automated tractography is reconstructed using Generalized Q-ball Imaging (Yeh et al., 2010), and tracking thresholds are determined by the normalized quantitative anisotropy, which is known to be highly dependent on b-value (Yeh et al., 2013). In contrast, XTRACT is a probabilistic method based largely on fiber orientation (and its dispersion) alone (from the ball-and-stick model (Sotiropoulos et al., 2012)), and different b-values give highly similar results of orientation (although dispersion will vary). XTRACT is also most sensitive to drastic change in resolution, likely caused by the probabilistic nature of the tractography process and subsequent thresholding for segmentation.

4.3. Shape variation and location of variation

This is to the best of our knowledge also the first time that reproducibility of different shape features of tractography has been investigated. While the variation across and within subjects has previously been studied (Yeh, 2020), it is important to understand cross-protocol and cross-scanner effects if these features are to be potential biomarkers in health and disease. These shape measures show similar patterns of variability, largest across resolution, vendors, and scanners, and smallest variation across repeats, directions, and b-values. More than variation, different resolutions and b-values can significantly bias measures, for example consistently overestimating volume and surface areas at lower resolutions where more partial volume effects are expected. Depending on tractography method, many features are remarkably robust, with MAPE below 5%, in line with that of microstructure features.

We also investigated locations of differences and similarities by visualizing where there was consistent agreement and disagreement. Importantly, even with differences in acquisition and scanners, methods are able to consistently reproduce the major shape and location of the intended pathway, with differences most frequently occurring at the periphery, or edges, of the pathway, and along the white matter and gray matter interface. While features of shape and geometry may be biased due to sources of variation, these differences do not consistently occur at any one location or place along the pathway.

4.4. Different workflows

Over and above the typically studied sources of variation, we found that differences due to the choice of bundle segmentation workflows are most pronounced. For any given pathway, overlap from one workflow to another was low-to-moderate. This is in part due to the inherent sensitivity/specificity of different algorithms – for example Recobundles will look for clusters exhibiting a certain shape, while Tractseg is based on deep-learned segmentation, and Xtract will be highly dependent on the chosen threshold – but more importantly due to fundamental differences in how the pathway is dissected or defined (Schilling et al., 2020; Mandonnet et al., 2018). For example, the definition of a pathway by one method may be entirely different from another method, including choices in the presence or absence of connections to entire lobes or lobules, or differences in estimated spatial extent of pathways. While differences across methods were larger, they were importantly *consistently* different, meaning that comparing findings using different methods may result in differing conclusions on connectivity or microstructure. Differences between bundle segmentation workflows are also confounded by differences in the entire process of tractography, including differences in modeling, generation of streamlines (i.e., tractography), and bundle segmentation or filtering. Thus, it is intuitive that major differences exist when implementing different standard workflows to study the brain.

4.5. Microstructure variation

Finally, we looked at how much the variation in tractography contribute to the already existing cross-protocol and cross-scanner variation in dMRI measures. For FA, difference across scanners are known to be as much as 5–15% (L Ning et al., 2020; CM Tax et al., 2019), and differences are expected due to different b-values, while scan-rescan reproducibility is high (< 5%). The variation in tractography segmentations does indeed statistically significantly increase this variation for most effects, although the increase is typically very small and < 5%. Similar results are observed for MD, although most changes are most pronounced for MD across different scanners. Thus, while tractography has the benefit of added specificity over simply propagating atlas-derived regions to subject-space, it does potentially increase variability in these measurements. Although methods such as tract-based spatial statistics (Smith et al., 2006) have been developed to mitigate these effects, we lose the added benefit of characterizing an index of interest along or within the full trajectory of the pathway.

4.6. Future studies and limitations

Future studies should investigate additional sources of variation. Manual dissection of fiber bundles gives the dissector the ability to interactively manipulate pathways to their liking (F Rheault et al., 2020), and it remains to be seen how this is influenced by scanner and site given the flexibility of this approach. Further, it is unknown whether these variabilities will matter in a clinical setting (Vanderweyen et al., 2020; Mancini et al., 2019; Fekonja et al., 2019; Essayed et al., 2017), although with the importance of determining pathway boundaries, we hypothesize that the partial volume effects due to acquisition resolution will possibly influence decision making. It is worth investigating the potentially large array of automated bundle segmentation methods that exist, as some are likely more/less

appropriate when comparing or combining datasets with different confounds. Additionally, as alternative segmentation methods, or even whole-brain connectome analysis pipelines, are proposed, the use of open-source multi-site multi-subject datasets (Koller et al., 2021; Avesani et al., 2019; Jack et al., 2008) should be encouraged to investigate the successes and limitations of new approaches. Many algorithms for reconstruction and tractography are now able to utilize multiple diffusion shells, and the change in variability and precision of tractography using these techniques compared to isolated diffusion weightings should be compared, but is outside the scope of this work. As along-fiber quantification (Chamberland et al., 2019; Yeatman et al., 2012) has proven valuable in the research setting, it would be worthwhile to perform investigations which parallel the current study in order to ask how and where along the bundle differences occur due to different effects. This has been previously investigated, but is largely limited to scan-rescan analysis (Yeatman et al., 2012; Koller et al., 2021; Chandio et al., 2020), while the tract-averaged indices are still commonly utilized in neuroimaging studies.

A major limitation of the current study is the limited sample sizes of both datasets due to challenges associated with scanning the same subjects on different scanners and with different protocols. However, there are few multi-site multi-subject databases, and fewer still with varied protocols on the same subjects, whereas here we are able to remove effects across subjects by analyzing only the same subject with different protocols. It is expected that more datasets will become available as big-data and multi-site collaborations become more important to the neuroimaging community, and traveling subjects become common place in order to harmonize across sites. Exemplar open-sourced datasets include that of (Tong et al., 2020) with $N=3$ subjects at 20 sites with Prisma scanners and a multi-shell dataset (allowing analysis of RESCAN, SCAN, BVAL, DIR), the traveling human phantom dataset (VA Magnotta et al., 2012) with $N=5$ subjects at 8 center (SCAN, VEN, DIR), or consortiums such as Pharmacog (Galluzzi et al., 2016), ADNI (Jack et al., 2008), HCP (Glasser et al., 2013), or OASIS (Marcus et al., 2007), all with large sample size and repeat scans, but typically limited to RESCAN analysis only or without matched subjects across scanners/vendors/protocols. Because of this, for simplicity, we have chosen two datasets in this study which allow incorporation of all intended sources of variation without compromising readability. While we have looked at a wider range of variability factors than previous studies, we emphasize that these results are based only on two specific databases, and analysis should be reproduced on other (and new) databases in future work to show generalizability.

Finally, while the primary focus of our study was on variation due to scanner-effects, acquisition-effects, and b-value-effects, our analysis was limited to studying these effects on only four bundle segmentation workflows. We did not implement all existing automated bundle reconstruction pipelines or workflows (Yeatman et al., 2012; Vazquez et al., 2020; F Zhang et al., 2019; Guevara et al., 2012; Wakana et al., 2007; Wasserthal et al., 2019; F Zhang et al., 2020; F Zhang et al., 2020; F Zhang et al., 2019; O'Donnell and Westin, 2007; Wassermann et al., 2016; Ros et al., 2013; Zöllei et al., 2019; Schilling et al., 2021), however, our selection captures a variety of techniques used to reconstruction bundles, including differences in the use of atlases or regions-of-interest, those based on shape and/or orientation features, machine learning techniques, and differences in the generation of

streamlines – a wide variety of vastly different approaches that we consider a strength of this study. To create a tractable parameter space, we have chosen only these four representatives of the wide variety of possible approaches.

Finally, we did not directly perform harmonization techniques in this study. There are dozens of methods available to do this (see (L Ning et al., 2020; CM Tax et al., 2019)), and understanding and characterizing harmonization results across several algorithms would take away from the main focus of this study – which is characterization and ranking of variability across confounds. Further, harmonization would only affect a subset of results (i.e., those looking at FA/MD) as most harmonization approaches leave orientation untouched.

5. Conclusion

When investigating connectivity and microstructure of the white matter pathways of the brain using tractography, it is important to understand potential confounds and sources of variation in the process. Here, we find that tractography bundle segmentation results are influenced by the use of different vendors and scanners, and different acquisition choices of resolution, diffusion directions, and diffusion sensitizations, thus results may not be directly comparable when combining data or results across studies. Additionally, different bundle segmentation protocols have different successes/limitations when dealing with sources of variation, and the use of different protocols for bundle segmentation may result in different representations of the same intended pathway. These confounds need to be considered when designing or developing new tractography or bundle dissection algorithms, and when interpreting or combining data across sites.

6. Code

Multi-site, multi-scanner, multi-protocol, and multi-subject databases are available for MASIvar (<https://openneuro.org/datasets/ds003416>) and for MUSHAC (by request). Tractography pipelines are implemented as described by each software package using default parameters for TractSeg (Release 2.3; <https://github.com/MIC-DKFZ/TractSeg>), ATK (Lct 17 2020 build; <http://dsi-studio.labsolver.org>), RECO (Dipy 1.2.0; <https://dipy.org>), and XTRACT (FSL 6.0.3; <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/XTRACT>). Shape analysis is available in DSI Studio, as Matlab Code (https://github.com/dmitrishastin/tractography_shapes/).

Acknowledgements

This work was supported by the National Science Foundation Career Award #1452485, the National Institutes of Health under award numbers R01EB017230, and T32EB001628, and in part by ViSE/VICTR VR3029 and the National Center for Research Resources, Grant UL1 RR024975-01. CMWT was supported by a Sir Henry Wellcome Fellowship (215944/Z/19/Z) and a Veni grant (17331) from the Dutch Research Council (NWO).

Appendix

The bundles resulting from each segmentation pipeline are given as a list below, with acronyms used in the text.

Recobundles:

Anterior Commissure (AC); Arcuate Fasciculus left (AF_L); Arcuate Fasciculus right (AF_R); Cerebellum left (CB_L); Cerebellum right (CB_R); Cingulum left (C_L); Cingulum right (C_R); Corpus Callosum (CC); Corticospinal Tract left (CST_L); Corticospinal Tract Right (CST_R); Corticostriatal Pathway left (CS_L); Corticostriatal Pathway right (CS_R); Central Tegmental Tract left (CT_L); Central Tegmental Tract right (CT_R); Extreme Capsule left (EMC_L); Extreme Capsule right (EMC_R); Fornix left (F_L); Fornix right (F_R); Frontal Aslant Tract left (FAT_L); Frontal Aslant Tract right (FAT_R); Fronto-pontine tract left (FPT_L); Fronto-pontine tract right (FPT_R); Inferior Cerebellar Peduncle left (ICP_L); Inferior Cerebellar Peduncle right (ICP_R); Inferior Fronto-occipital Fasciculus left (IFOF_L); Inferior Fronto-occipital Fasciculus right (IFOF_R); Inferior Longitudinal Fasciculus left (ILF_L); Inferior Longitudinal Fasciculus right (ILF_R); Middle Cerebellar Peduncle (MCP); Middle Longitudinal Fasciculus left (MdLF_L); Middle Longitudinal Fasciculus right (MdLF_R); Medial Lemniscus left (ML_L); Medial Lemniscus right (ML_R); Occipito Pontine Tract left (OPT_L); Occipito Pontine Tract right (OPT_R); Optic Radiation left (OR_L); Optic Radiation right (OR_R); Parieto Pontine Tract left (PPT_L); Parieto Pontine Tract right (PPT_R); Superior Cerebellar Peduncle (SCP); Superior longitudinal fasciculus left (SLF_L); Superior longitudinal fasciculus right (SLF_R); Uncinate Fasciculus left (UF_L); Uncinate Fasciculus right (UF_R);

TractSeg:

Arcuate fascicle left (AF_L); Arcuate fascicle right (AF_R); Anterior Thalamic Radiation left (ATR_L); Thalamic Radiation right; (ATR_R); Commissure Anterior (CA); Rostrum (CC_1); Genu (CC_2); Rostral body (Premotor) (CC_3); Anterior midbody (Primary Motor) (CC_4); Posterior midbody (Primary Somatosensory) (CC_5); Isthmus (CC_6); Splenium (CC_7); Corpus Callosum – all (CC); Cingulum left (CG_L); Cingulum right (CG_R); Corticospinal tract left (CST_L); Corticospinal tract right (CST_R); Fronto-pontine tract left (FPT_L); Fronto-pontine tract right (FPT_R); Fornix left (FX_L); Fornix right (FX_R); Inferior cerebellar peduncle left (ICP_L); Inferior cerebellar peduncle right (ICP_R); Inferior occipito-frontal fascicle left (IFO_L); Inferior occipito-frontal fascicle right (IFO_R); Inferior longitudinal fascicle left (ILF_L); Inferior longitudinal fascicle right (ILF_R); Middle cerebellar peduncle (MCP); Middle longitudinal fascicle left (MLF_L); Middle longitudinal fascicle right (MLF_R); Optic radiation left (OR_L); Optic radiation right (OR_R); Parieto-occipital pontine left (POPT_L); Parieto-occipital pontine right (POPT_R); Superior cerebellar peduncle left (SCP_L); Superior cerebellar peduncle right (SCP_R); Superior longitudinal fascicle III left (SLF_III_L); Superior longitudinal fascicle III right (SLF_III_R); Superior longitudinal fascicle II left (SLF_II_L); Superior longitudinal fascicle II right (SLF_II_R); Superior longitudinal fascicle I left (SLF_I_L); Superior longitudinal fascicle I right (SLF_I_R); Striato-fronto-orbital left (ST_FO_L); Striato-fronto-orbital right (ST_FO_R); Striato-occipital left (ST_OCC_L); Striato-occipital right (ST_OCC_R); Striato-parietal left (ST_PAR_L); Striato-parietal right (ST_PAR_R); Striato-postcentral left (ST_POSTC_L); Striato-postcentral right (ST_POSTC_R); Striato-precentral left (ST_PREC_L); Striato-precentral right (ST_PREC_R); Striato-prefrontal left (ST_PREF_L); Striato-prefrontal

right (ST_PREF_R); Striato-premotor left (ST_PREM_L); Striato-premotor right (ST_PREM_R); Thalamo-occipital left (T_OCC_L); Thalamo-occipital right (T_OCC_R); Thalamo-parietal left (T_PAR_L); Thalamo-parietal right (T_PAR_R); Thalamo-postcentral left (T_POSTC_L); Thalamo-postcentral right (T_POSTC_R); Thalamo-precentral left (T_PREC_L); Thalamo-precentral right (T_PREC_R); Thalamo-prefrontal left (T_PREF_L); Thalamo-prefrontal right (T_PREF_R); Thalamo-premotor left (T_PREM_L); Thalamo-premotor right (T_PREM_R); Uncinate fascicle left (UF_L); Uncinate fascicle right (UF_R).

Xtract:

Anterior Commissure (AC); Arcuate Fascicle left (AF_L); Arcuate Fascicle right (AF_R); Acoustic Radiation left (AR_L); Acoustic Radiation right (AR_R); Anterior Thalamic Radiation left (ATR_L); Anterior Thalamic Radiation right (ATR_R); Cingulum Bundle Dorsal left (CBD_L); Cingulum Bundle Dorsal right (CBD_R); Cingulum Bundle Parahippocampal left (CBP_L); Cingulum Bundle Parahippocampal right (CBP_R); Cingulum Bundle Temporal left (CBT_L); Cingulum Bundle Temporal right (CBT_R); Corticospinal Tract left (CST_L); Corticospinal Tract right (CST_R); Frontal Aslant left (FA_L); Frontal Aslant right (FA_R); Forceps Major (FMA); Forceps Minor (FMI); Fornix left (FX_L); Fornix right (FX_R); Inferior Fronto-occipital Fasciculus left (IFO_L); Inferior Fronto-occipital Fasciculus right (IFO_R); Inferior Longitudinal Fasciculus left (ILF_L); Inferior Longitudinal Fasciculus right (ILF_R); Middle Cerebellar Peduncle (MCP); Medio-Dorsal Longitudinal Fasciculus left (MDLF_L); Medio-Dorsal Longitudinal Fasciculus right (MDLF_R); Optic Radiation left (OR_L); Optic Radiation right (OR_R); Superior Longitudinal Fasciculus 1 left (SLF1_L); Superior Longitudinal Fasciculus 1 right (SLF1_R); Superior Longitudinal Fasciculus 2 left (SLF2_L); Superior Longitudinal Fasciculus 2 right (SLF2_R); Superior Longitudinal Fasciculus 3 left (SLF3_L); Superior Longitudinal Fasciculus 3 right (SLF3_R); Superior Thalamic Radiation left (STR_L); Superior Thalamic Radiation right (STR_R); Uncinate Fasciculus left (UF_L); Uncinate Fasciculus right (UF_R); Vertical Occipital Fasciculus left (VOF_L); Vertical Occipital Fasciculus right (VOF_R).

ATK:

Arcuate_Fasciculus_L (AF_L); Arcuate Fasciculus R (AF_R); Cortico Spinal Tract L (CST_L); Cortico Spinal Tract R (CST_R); Cortico Striatal Pathway L (CS_L); Cortico Striatal Pathway R (CS_R); Corticobulbar Tract L (CBT_L); Corticobulbar Tract R (CBT_R); Corticopontine Tract L (CPT_L); Corticopontine Tract R (CPT_R); Corticothalamic Pathway L (CTP_L); Corticothalamic Pathway R (CTP_R); Inferior Cerebellar Peduncle L (ICP_L); Inferior Cerebellar Peduncle R (ICP_R); Inferior Fronto Occipital Fasciculus L (IFOF_L); Inferior Fronto Occipital Fasciculus R (IFOF_R); Inferior Longitudinal Fasciculus L (ILF_L); Inferior Longitudinal Fasciculus R (ILF_R); Optic Radiation L (OR_L); Optic Radiation R (OR_R); Middle Longitudinal Fasciculus L (Mdlf_L); Middle Longitudinal Fasciculus R (Mdlf_R); Uncinate Fasciculus L (UF_L); Uncinate Fasciculus R (UF_R).

Data And Code Availability

The data for this study were selected from two open-sourced multi-subject, multi-scanner, and multi-protocol benchmark databases: the MASiVar (Alexander et al., 2019) and MUSHAC datasets (Jones et al., 2018; Novikov et al., 2018) – thus the data is freely available as found in associated references.

References

- Alexander DC, Dyrby TB, Nilsson M, Zhang H, 2019. Imaging brain microstructure with diffusion MRI: practicality and applications. *NMR Biomed.* 32 (4), e3841. doi: 10.1002/nbm.3841, Epub 2017/11/29PubMed [PubMed: 29193413]
- Andersson JL, Skare S, Ashburner J, 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20 (2), 870–888. doi: 10.1016/S1053-8119(03)00336-7. [PubMed: 14568458]
- Avesani P, McPherson B, Hayashi S, Caiafa CF, Henschel R, Garyfallidis E, et al. , 2019. The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Sci. Data* 6 (1), 69. doi: 10.1038/s41597-019-0073-y. [PubMed: 31123325]
- Cai LY, Yang Q, Hansen CB, Nath V, Ramadass K, Johnson GW, et al. , 2021. PreQual: an automated pipeline for integrated preprocessing and quality assurance of diffusion weighted MRI images. *Magn. Reson. Med* doi: 10.1002/mrm.28678, n/a(n/a)<https://doi.org/>.
- Cai LY, Yang Q, Kanakaraj P, Nath V, Newton AT, Edmonson HA, et al. , 2020. MASiVar: multisite, Multiscanner, and Multisubject Acquisitions for Studying Variability in Diffusion Weighted Magnetic Resonance Imaging. *bioRxiv* doi: 10.1101/2020.12.03.408567, 2020.12.03.408567.
- Canales-Rodríguez EJ, Legarreta JH, Pizzolato M, Rensonnet G, Girard G, Patino JR, et al. , 2018. Sparse wars: a survey and comparative study of spherical deconvolution algorithms for diffusion MRI. *Neuroimage* 184, 140–160. doi: 10.1016/j.neuroimage.2018.08.071. [PubMed: 30193974]
- Cetin Karayumak S, Bouix S, Ning L, James A, Crow T, Shenton M, et al. , 2019. Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. *Neuroimage* 184, 180–200. doi: 10.1016/j.neuroimage.2018.08.073. [PubMed: 30205206]
- Chamberland M, Raven EP, Genc S, Duffy K, Descoteaux M, Parker GD, et al. , 2019. Dimensionality reduction of diffusion MRI measures for improved tractometry of the human brain. *Neuroimage* 200, 89–100. doi: 10.1016/j.neuroimage.2019.06.020, Epub 2019/06/20PubMed. [PubMed: 31228638]
- Chamberland M, Tax CMW, Jones DK, 2018. Meyer’s loop tractography for image-guided surgery depends on imaging protocol and hardware. *Neuroimage Clin* 20, 458–465. doi: 10.1016/j.nicl.2018.08.021, Epub 2018/08/13PubMed. [PubMed: 30128284]
- Chandio BQ, Risacher SL, Pestilli F, Bullock D, Yeh FC, Koudoro S, et al. , 2020. Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations. *Sci. Rep* 10 (1), 17149. doi: 10.1038/s41598-020-74054-4, Epub 2020/10/15.
- Chang LC, Jones DK, Pierpaoli C, 2005. RESTORE: robust estimation of tensors by outlier rejection. *Magn. Reson. Med* 53 (5), 1088–1095. doi: 10.1002/mrm.20426. [PubMed: 15844157]
- Chen G, Zhang P, Li K, Wee C-Y, Wu Y, Shen D, et al. , 2016. Improving Estimation of Fiber Orientations in Diffusion MRI Using Inter-Subject Information Sharing. *Sci. Rep* 6 (1), 37847. doi: 10.1038/srep37847.
- Daducci A, Canales-Rodríguez EJ, Descoteaux M, Garyfallidis E, Gur Y, Lin YC, et al. , 2014. Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. *IEEE Trans. Med. Imaging* 33 (2), 384–399. doi: 10.1109/TMI.2013.2285500, Epub 2013/10/11. [PubMed: 24132007]
- De Luca A, Ianus A, Leemans A, Palombo M, Shemesh N, Zhang H, et al. , 2021. On the generalizability of diffusion MRI signal representations across acquisition parameters,

sequences and tissue types: chronicles of the MEMENTO challenge. bioRxiv doi: 10.1101/2021.03.02.433228, 2021.03.02.433228.

- 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 Clin* 15, 659–672. doi: 10.1016/j.nicl.2017.06.011. [PubMed: 28664037]
- Farrell JA, Landman BA, Jones CK, Smith SA, Prince JL, van Zijl PC, et al. , 2007. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *J. Magn. Reson. Imaging* 26 (3), 756–767. doi: 10.1002/jmri.21053, Epub 2007/08/31. [PubMed: 17729339]
- Fekonja L, Wang Z, Bährend I, Rosenstock T, Rösler J, Wallmeroth L, et al. , 2019. Manual for clinical language tractography. *Acta Neurochir. (Wien)* 161 (6), 1125–1137. doi: 10.1007/s00701-019-03899-0. [PubMed: 31004240]
- Forkel SJ, Thiebaut de Schotten M, Kawadler JM, Dell'Acqua F, Danek A, Catani M, 2014. The anatomy of fronto-occipital connections from early blunt dissections to contemporary tractography. *Cortex* 56, 73–84. doi: 10.1016/j.cortex.2012.09.005. [PubMed: 23137651]
- Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. , 2017. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161, 149–170. doi: 10.1016/j.neuroimage.2017.08.047, Epub 2017/08/18. [PubMed: 28826946]
- Galluzzi S, Marizzoni M, Babiloni C, Albani D, Antelmi L, Bagnoli C, et al. , 2016. Clinical and biomarker profiling of prodromal Alzheimer's disease in workpackage 5 of the Innovative Medicines Initiative PharmaCog project: a 'European ADNI study'. *J. Intern. Med* 279 (6), 576–591. doi: 10.1111/joim.12482, Epub 2016/03/05. [PubMed: 26940242]
- Garyfallidis E, Brett M, Amirbekian B, Rokem A, van der Walt S, Descoteaux M, et al. , 2014. Dipy, a library for the analysis of diffusion MRI data. *Front Neuroinform* 8, 8. doi: 10.3389/fninf.2014.00008. [PubMed: 24600385]
- Garyfallidis E, Cote MA, Rheault F, Sidhu J, Hau J, Petit L, et al. , 2018. Recognition of white matter bundles using local and global streamline-based registration and clustering. *Neuroimage* 170, 283–295. doi: 10.1016/j.neuroimage.2017.07.015. [PubMed: 28712994]
- Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. , 2013. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127. [PubMed: 23668970]
- Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. , 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med* 47 (6), 1202–1210. doi: 10.1002/mrm.10171. [PubMed: 12111967]
- Guevara M, Roman C, Houenou J, Duclap D, Poupon C, Mangin JF, et al. , 2017. Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography. *Neuroimage* 147, 703–725. doi: 10.1016/j.neuroimage.2016.11.066. [PubMed: 28034765]
- Guevara P, Duclap D, Poupon C, Marrakchi-Kacem L, Fillard P, Le Bihan D, et al. , 2012. Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *Neuroimage* 61 (4), 1083–1099. doi: 10.1016/j.neuroimage.2012.02.071, Epub 2012/03/15. [PubMed: 22414992]
- Hau J, Sarubbo S, Houde JC, Corsini F, Girard G, Deledalle C, et al. Revisiting the human uncinate fasciculus, its subcomponents and asymmetries with stem-based tractography and microdissection validation. *Brain structure & function*. 2017;222(4):1645–62. doi: 10.1007/s00429-016-1298-6. [PubMed: 27581617]
- Hau J, Sarubbo S, Perchey G, Crivello F, Zago L, Mellet E, et al. , 2016. Cortical Terminations of the Inferior Fronto-Occipital and Uncinate Fasciculi: anatomical Stem-Based Virtual Dissection. *Front. Neuroanat* 10, 58. doi: 10.3389/fnana.2016.00058. [PubMed: 27252628]
- Heiervang E, Behrens TE, Mackay CE, Robson MD, Johansen-Berg H, 2006. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage* 33 (3), 867–877. doi: 10.1016/j.neuroimage.2006.07.037, Epub 2006/09/27. [PubMed: 17000119]

- Huynh KM, Kim J, Chen G, Wu Y, Shen D, Yap P-T (Eds.), 2019a. Longitudinal Harmonization for Improving Tractography in Baby Diffusion MRI. Computational Diffusion MRI Springer International Publishing, Cham. 2019//.
- Huynh KM, Chen G, Wu Y, Shen D, Yap PT, 2019b. Multi-Site Harmonization of Diffusion MRI Data via Method of Moments. *IEEE Trans. Med. Imaging* 38 (7), 1599–1609. doi: 10.1109/TMI.2019.2895020. [PubMed: 30676953]
- Jack CR Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. , 2008. The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27 (4), 685–691. doi: 10.1002/jmri.21049. [PubMed: 18302232]
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM, 2012. *Fsl. Neuroimage* 62 (2), 782–790. doi: 10.1016/j.neuroimage.2011.09.015. [PubMed: 21979382]
- Jeurissen B, Descoteaux M, Mori S, Leemans A, 2019. Diffusion MRI fiber tractography of the brain. *NMR Biomed.* 32 (4), e3785. doi: 10.1002/nbm.3785, Epub 2017/09/25. [PubMed: 28945294]
- Jones DK, Alexander DC, Bowtell R, Cercignani M, Dell’Acqua F, McHugh DJ, et al. , 2018. Microstructural imaging of the human brain with a ‘super-scanner’: 10 key advantages of ultra-strong gradients for diffusion MRI. *Neuroimage* 182, 8–38. doi: 10.1016/j.neuroimage.2018.05.047, Epub 2018/05/22. [PubMed: 29793061]
- Jones DK, Basser PJ, 2004. “Squashing peanuts and smashing pumpkins”: how noise distorts diffusion-weighted MR data. *Magn. Reson. Med. : Off. J. Soc. Magn. Reson. Med. /Soc. Magn. Reson. Med.* 52 (5), 979–993. doi: 10.1002/mrm.20283.
- Jones DK, Cercignani M, 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed.* 23 (7), 803–820. doi: 10.1002/nbm.1543. [PubMed: 20886566]
- Jones DK, Chitnis XA, Job D, Khong PL, Leung LT, Marengo S, et al., 2007. What happens when nine different groups analyze the same DT-MRI data set using voxel-based methods.
- Jones DK, Horsfield MA, Simmons A, 1999a. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med* 42 (3), 515–525. [PubMed: 10467296]
- Jones DK, Horsfield MA, Simmons A, 1999b. Optimal Strategies for Measuring Diffusion in Anisotropic Systems by Magnetic Resonance Imaging. *Magn. Reson. Med* 42, 515–525. [PubMed: 10467296]
- Jones DK, Knosche TR, Turner R, 2013. White matter integrity, fiber count, and other fallacies: the do’s and don’ts of diffusion MRI. *Neuroimage* 73, 239–254. doi: 10.1016/j.neuroimage.2012.06.081, Epub 2012/08/01. [PubMed: 22846632]
- Jones DK, 2003. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn. Reson. Med. : Off. J. Soc. Magn. Reson. Med. /Soc. Magn. Reson. Med* 49 (1), 7–12. doi: 10.1002/mrm.10331.
- Jones DK, 2004. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. *Magn. Reson. Med. : Off. J. Soc. Magn. Reson. Med. /Soc. Magn. Reson. Med* 51 (4), 807–815. doi: 10.1002/mrm.20033, Epub 2004/04/06.
- Jones DK, 2010. Precision and accuracy in diffusion tensor magnetic resonance imaging. *Top. Magn. Reson. Imaging* 21 (2), 87–99. doi: 10.1097/RMR.0b013e31821e56ac. [PubMed: 21613874]
- Jones R, Grisot G, Augustinack J, Magnain C, Boas DA, Fischl B, et al. , 2020. Insight into the fundamental trade-offs of diffusion MRI from polarization-sensitive optical coherence tomography in ex vivo human brain. *Neuroimage*, 116704 doi: 10.1016/j.neuroimage.2020.116704.
- Koller K, Rudrapatna U, Chamberland M, Raven EP, Parker GD, Tax CMW, et al. , 2021. MICRA: microstructural image compilation with repeated acquisitions. *Neuroimage* 225, 117406. doi: 10.1016/j.neuroimage.2020.117406.
- Landman BA, Farrell JA, Jones CK, Smith SA, Prince JL, Mori S, 2007. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Neuroimage* 36 (4), 1123–1138. doi: 10.1016/j.neuroimage.2007.02.056, Epub 2007/05/29. [PubMed: 17532649]
- Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IA, Farrell JA, et al. , 2011. Multi-parametric neuroimaging reproducibility: a 3-T resource study. *Neuroimage* 54 (4), 2854–2866. doi:10.1016/j.neuroimage.2010.11.047. [PubMed: 21094686]

- Lori NF, Akbudak E, Shimony JS, Cull TS, Snyder AZ, Guillory RK, et al. , 2002. Diffusion tensor fiber tracking of human brain connectivity: acquisition methods, reliability analysis and biological results. *NMR Biomed.* 15 (7–8), 494–515. doi: 10.1002/nbm.779. [PubMed: 12489098]
- Maffei C, Sarubbo S, Jovicich J, 2019. Diffusion-based tractography atlas of the human acoustic radiation. *Sci. Rep* 9 (1), 4046. doi: 10.1038/s41598-019-40666-8. [PubMed: 30858451]
- Magnotta VA, Matsui JT, Liu D, Johnson HJ, Long JD, Bolster BD, et al. , 2012a. Multicenter reliability of diffusion tensor imaging. *Brain Connect* 2 (6), 345–355. doi: 10.1089/brain.2012.0112. [PubMed: 23075313]
- Magnotta VA, Matsui JT, Liu D, Johnson HJ, Long JD, Bolster BD Jr., et al. , 2012b. Multicenter reliability of diffusion tensor imaging. *Brain Connect* 2 (6), 345–355. doi: 10.1089/brain.2012.0112. [PubMed: 23075313]
- Mancini M, Vos SB, Vakharia VN, O’Keeffe AG, Trimmel K, Barkhof F, et al. , 2019. Automated fiber tract reconstruction for surgery planning: extensive validation in language-related white matter tracts. *Neuroimage Clin.* 23, 101883. doi: 10.1016/j.nicl.2019.101883.
- Mandonnet E, Sarubbo S, Petit L, 2018. The Nomenclature of Human White Matter Association Pathways: proposal for a Systematic Taxonomic Anatomical Classification. *Front. Neuroanat* 12, 94. doi: 10.3389/fnana.2018.00094, Epub 2018/11/06. [PubMed: 30459566]
- Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL, 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci* 19 (9), 1498–1507. doi: 10.1162/jocn.2007.19.9.1498, Epub 2007/08/24. [PubMed: 17714011]
- Mars RB, Sallet J, Schuffelgen U, Jbabdi S, Toni I, Rushworth MF, 2012. Connectivity-based subdivisions of the human right “temporoparietal junction area”: evidence for different areas participating in different cortical networks. *Cereb. Cortex* 22 (8), 1894–1903. doi: 10.1093/cercor/bhr268. [PubMed: 21955921]
- Min J, Park M, Choi JW, Jahng GH, Moon WJ, 2018. Inter-Vendor and Inter-Session Reliability of Diffusion Tensor Imaging: implications for Multicenter Clinical Imaging Studies. *Korean J. Radiol* 19 (4), 777–782. doi: 10.3348/kjr.2018.19.4.777, Epub 2018/07/03. [PubMed: 29962884]
- Mirzaalian H, Ning L, Savadjiev P, Pasternak O, Bouix S, Michailovich O, et al. , 2016. Inter-site and inter-scanner diffusion MRI data harmonization. *Neuroimage* 135, 311–323. doi: 10.1016/j.neuroimage.2016.04.041, Epub 2016/04/30. [PubMed: 27138209]
- Mirzaalian H, Ning L, Savadjiev P, Pasternak O, Bouix S, Michailovich O, et al. , 2018. Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging Behav.* 12 (1), 284–295. doi: 10.1007/s11682-016-9670-y. [PubMed: 28176263]
- Moyer D, Ver Steeg G, Tax CMW, Thompson PM, 2020. Scanner invariant representations for diffusion MRI harmonization. *Magn. Reson. Med* 84 (4), 2174–2189. doi: 10.1002/mrm.28243. [PubMed: 32250475]
- Nath V, Schilling KG, Parvathaneni P, Huo Y, Blaber JA, Hainline AE, et al. , 2019. Tractography reproducibility challenge with empirical data (TraCED): the 2017 ISMRM diffusion study group challenge. *J. Magn. Reson. Imaging* doi: 10.1002/jmri.26794, Epub 2019/06/09.
- Neubert FX, Mars RB, Sallet J, Rushworth MF, 2015. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proc. Natl. Acad. Sci. U.S.A* 112 (20), E2695–E2704. doi: 10.1073/pnas.1410767112. [PubMed: 25947150]
- Neubert FX, Mars RB, Thomas AG, Sallet J, Rushworth MF, 2014. Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81 (3), 700–713. doi: 10.1016/j.neuron.2013.11.012. [PubMed: 24485097]
- Ning L, Bonet-Carne E, Grussu F, Seppehrband F, Kaden E, Veraart J, et al. , 2020a. Cross-scanner and cross-protocol multi-shell diffusion MRI data harmonization: algorithms and results. *Neuroimage* 221, 117128. doi: 10.1016/j.neuroimage.2020.117128, Epub 2020/07/17.
- Ning L, Bonet-Carne E, Grussu F, Seppehrband F, Kaden E, Veraart J, et al. , 2020b. Cross-scanner and cross-protocol multi-shell diffusion MRI data harmonization: algorithms and results. *Neuroimage* 221, 117128. doi: 10.1016/j.neuroimage.2020.117128.

- Novikov DS, Fieremans E, Jespersen SN, Kiselev VG, 2018. Quantifying brain microstructure with diffusion MRI: theory and parameter estimation. *NMR Biomed.* e3998. doi: 10.1002/nbm.3998, Epub 2018/10/15. [PubMed: 30321478]
- O'Donnell LJ, Westin CF, 2007. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging* 26 (11), 1562–1575. doi: 10.1109/TMI.2007.906785, Epub 2007/11/29. [PubMed: 18041271]
- Papinutto ND, Maule F, Jovicich J, 2013. Reproducibility and biases in high field brain diffusion MRI: an evaluation of acquisition and analysis variables. *Magn. Reson. Imaging* 31 (6), 827–839. doi: 10.1016/j.mri.2013.03.004, Epub 2013/04/30. [PubMed: 23623031]
- Pestilli F, Yeatman JD, Rokem A, Kay KN, Wandell BA, 2014. Evaluation and statistical inference for human connectomes. *Nat. Methods* 11 (10), 1058–1063. doi: 10.1038/nmeth.3098. [PubMed: 25194848]
- Pfefferbaum A, Adalsteinsson E, Sullivan EV, 2003. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J. Magn. Reson. Imaging* 18 (4), 427–433. doi: 10.1002/jmri.10377, Epub 2003/09/26. [PubMed: 14508779]
- Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Chiro GD, 1996. Diffusion Tensor MR Imaging of the Human Brain. *Radiology* 201, 637–648 Epub December 1996. [PubMed: 8939209]
- Prckovska V, Roebroek AF, Pullens WL, Vilanova A, ter Haar Romeny BM, 2008. Optimal acquisition schemes in high angular resolution diffusion weighted imaging. *Med. Image Comput. Comput. Assist. Interv* 11 (Pt 2), 9–17.
- Prohl AK, Scherrer B, Tomas-Fernandez X, Filip-Dhima R, Kapur K, Velasco-Annis C, et al. , 2019a. Reproducibility of Structural and Diffusion Tensor Imaging in the TACERN Multi-Center Study. *Front. Integr. Neurosci* 13, 24. doi: 10.3389/fnint.2019.00024, Epub 2019/07/17. [PubMed: 31417372]
- Prohl AK, Scherrer B, Tomas-Fernandez X, Filip-Dhima R, Kapur K, Velasco-Annis C, et al. , 2019b. Reproducibility of Structural and Diffusion Tensor Imaging in the TACERN Multi-Center Study. *Front. Integr. Neurosci* 13. doi: 10.3389/fnint.2019.00024, 24.
- Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P, 1999. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med* 42 (5), 952–962 Epub 1999/11/05. [PubMed: 10542355]
- Raffelt DA, Tournier JD, Smith RE, Vaughan DN, Jackson G, Ridgway GR, et al. , 2017. Investigating white matter fibre density and morphology using fixel-based analysis. *Neuroimage* 144 (PtA), 58–73. doi: 10.1016/j.neuroimage.2016.09.029, Epub 2016/09/14. [PubMed: 27639350]
- Rheault F, De Benedictis A, Daducci A, Maffei C, Tax CMW, Romascano D, et al. , 2020a. Tractostorm: the what, why, and how of tractography dissection reproducibility. *Hum. Brain Mapp* doi: 10.1002/hbm.24917.
- Rheault F, Poulin P, Valcourt Caron A, St-Onge E, Descoteaux M, 2020b. Common misconceptions, hidden biases and modern challenges of dMRI tractography. *J. Neural Eng* 17 (1), 011001. doi: 10.1088/1741-2552/ab6aad.
- Ros C, Gullmar D, Stenzel M, Mentzel HJ, Reichenbach JR, 2013. Atlas-guided cluster analysis of large tractography datasets. *PLoS One* 8 (12), e83847. doi: 10.1371/journal.pone.0083847, Epub 2014/01/05. [PubMed: 24386292]
- Sarubbo S, De Benedictis A, Maldonado IL, Basso G, Duffau H, 2013. Frontal terminations for the inferior fronto-occipital fascicle: anatomical dissection, DTI study and functional considerations on a multi-component bundle. *Brain Struct. Funct* 218 (1), 21–37. doi: 10.1007/s00429-011-0372-3. [PubMed: 22200882]
- Sarubbo S, Petit L, De Benedictis A, Chioffi F, Pfitz M, Dyrby TB, 2019. Uncovering the inferior fronto-occipital fascicle and its topological organization in non-human primates: the missing connection for language evolution. *Brain Struct. Funct* 224 (4), 1553–1567. doi: 10.1007/s00429-019-01856-2. [PubMed: 30847641]
- Schilling K, Gao Y, Janve V, Stepniewska I, Landman BA, Anderson AW, 2017. Can increased spatial resolution solve the crossing fiber problem for diffusion MRI? *NMR Biomed.* 30 (12). doi: 10.1002/nbm.3787, Epub 2017/09/15.

- Schilling KG, Janve V, Gao Y, Stepniewska I, Landman BA, Anderson AW, 2018. Histological validation of diffusion MRI fiber orientation distributions and dispersion. *Neuroimage* 165, 200–221. doi: 10.1016/j.neuroimage.2017.10.046, Epub 2017/10/23. [PubMed: 29074279]
- Schilling KG, Rheault F, Petit L, Hansen CB, Nath V, Yeh F–C, et al. , 2020. Tractography dissection variability: what happens when 42 groups dissect 14 white matter bundles on the same dataset? *bioRxiv* doi: 10.1101/2020.10.07.321083, 2020.10.07.321083.
- Schilling KG, Rheault F, Petit L, Hansen CB, Nath V, Yeh F–C, et al. , 2021. Tractography dissection variability: what happens when 42 groups dissect 14 white matter bundles on the same dataset? *bioRxiv* doi: 10.1101/2020.10.07.321083, 2020.10.07.321083.
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. , 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31 (4), 1487–1505. doi: 10.1016/j.neuroimage.2006.02.024. [PubMed: 16624579]
- Sotiropoulos SN, Behrens TE, Jbabdi S, 2012. Ball and rackets: inferring fiber fanning from diffusion-weighted MRI. *Neuroimage* 60 (2), 1412–1425. doi: 10.1016/j.neuroimage.2012.01.056, Epub 2012/01/25. [PubMed: 22270351]
- Tax CM, Grussu F, Kaden E, Ning L, Rudrapatna U, Evans J, et al. , 2019a. Cross-scanner and cross-protocol diffusion MRI data harmonisation: a benchmark database and evaluation of algorithms. *Neuroimage* doi: 10.1016/j.neuroimage.2019.01.077, Epub 2019/02/01.
- Tax CM, Grussu F, Kaden E, Ning L, Rudrapatna U, John Evans C, et al. , 2019b. Cross-scanner and cross-protocol diffusion MRI data harmonisation: a benchmark database and evaluation of algorithms. *Neuroimage* 195, 285–299. doi: 10.1016/j.neuroimage.2019.01.077, Epub 2019/02/01. [PubMed: 30716459]
- Tax CMW, Szczepankiewicz F, Nilsson M, Jones DK, 2020. The dot-compartment revealed? Diffusion MRI with ultra-strong gradients and spherical tensor encoding in the living human brain. *Neuroimage* 210, 116534. doi: 10.1016/j.neuroimage.2020.116534.
- Teipel SJ, Reuter S, Stieltjes B, Acosta-Cabronero J, Ernemann U, Fellgiebel A, et al. , 2011. Multicenter stability of diffusion tensor imaging measures: a European clinical and physical phantom study. *Psychiatry Res.: Neuroimaging* 194 (3), 363–371. doi: 10.1016/j.psychres.2011.05.012, <https://doi.org/>.
- Tong Q, He H, Gong T, Li C, Liang P, Qian T, et al. , 2020. Multicenter dataset of multi-shell diffusion MRI in healthy traveling adults with identical settings. *Sci. Data* 7 (1), 157. doi:10.1038/s41597-020-0493-8, Epub 2020/05/29. [PubMed: 32461581]
- Tournier JD, Calamante F, Connelly A, 2013. Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR Biomed.* 26 (12), 1775–1786. doi: 10.1002/nbm.3017. [PubMed: 24038308]
- Tournier JD, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, et al. , 2019. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* 202, 116137. doi: 10.1016/j.neuroimage.2019.116137.
- Tournier JD, Yeh CH, Calamante F, Cho KH, Connelly A, Lin CP, 2008. Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* 42 (2), 617–625. doi: 10.1016/j.neuroimage.2008.05.002. [PubMed: 18583153]
- Vanderweyen DC, Theaud G, Sidhu J, Rheault F, Sarubbo S, Descoteaux M, et al. The role of diffusion tractography in refining glial tumor resection. *Brain Structure and Function.* 2020;225(4):1413–36. doi: 10.1007/s00429-020-02056-z. [PubMed: 32180019]
- Vazquez A, Lopez-Lopez N, Houenou J, Poupon C, Mangin JF, Ladra S, et al. , 2020. Automatic group-wise whole-brain short association fiber bundle labeling based on clustering and cortical surface information. *Biomed. Eng Online* 19 (1), 42. doi: 10.1186/s12938-020-00786-z.
- Vishwesh Nath KGS, Parvathaneni, Prasanna, Hansen, Colin B, Hainline, Allison E, Bermudez, Camilo, Remedios, Samuel, et al. (Eds.), 2018. Inter-Scanner Harmonization of High Angular Resolution DW-MRI Using Null Space Deep Learning MIC-CAI-CDMRI, Granada, Spain.
- Vollmar C, O’Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, et al. , 2010. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *Neuroimage* 51 (4), 1384–1394. doi: 10.1016/j.neuroimage.2010.03.046, <https://doi.org/>. [PubMed: 20338248]

- Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, et al. , 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36 (3), 630–644. doi: 10.1016/j.neuroimage.2007.02.049, Epub 2007/03/20. [PubMed: 17481925]
- Warrington S, Bryant KL, Khrapitchev AA, Sallet J, Charquero-Ballester M, Douaud G, et al. , 2020. XTRACT - Standardised protocols for automated tractography in the human and macaque brain. *Neuroimage* 217, 116923. doi: 10.1016/j.neuroimage.2020.116923.
- Wassermann D, Makris N, Rathi Y, Shenton M, Kikinis R, Kubicki M, et al. , 2016. The white matter query language: a novel approach for describing human white matter anatomy. *Brain Struct. Funct* 221 (9), 4705–4721. doi: 10.1007/s00429-015-1179-4, Epub 2016/01/13. [PubMed: 26754839]
- Wassermann D, Makris N, Rathi Y, Shenton M, Kikinis R, Kubicki M, et al., 2018a. Tract Orientation Mapping for Bundle-Specific Tractography. In: Wasserthal J, Neher PF, Maier-Hein KH (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI*. Springer International Publishing, Cham 2018 2018//.
- Wasserthal J, Neher P, Maier-Hein KH, 2018b. TractSeg - Fast and accurate white matter tract segmentation. *Neuroimage* 183, 239–253. doi: 10.1016/j.neuroimage.2018.07.070. [PubMed: 30086412]
- Wasserthal J, Neher PF, Hirjak D, Maier-Hein KH, 2019. Combined tract segmentation and orientation mapping for bundle-specific tractography. *Med. Image Anal* 58, 101559. doi: 10.1016/j.media.2019.101559.
- Yeatman JD, Dougherty RF, Myall NJ, Wandell BA, Feldman HM, 2012. Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS One* 7 (11), e49790. doi: 10.1371/journal.pone.0049790, Epub 2012/11/14. [PubMed: 23166771]
- Yeh FC, Panesar S, Fernandes D, Meola A, Yoshino M, Fernandez-Miranda JC, et al. , 2018. Population-averaged atlas of the macroscale human structural connectome and its network topology. *Neuroimage* 178, 57–68. doi: 10.1016/j.neuroimage.2018.05.027. [PubMed: 29758339]
- Yeh FC, Verstynen TD, Wang Y, Fernandez-Miranda JC, Tseng WY, 2013. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS One* 8 (11), e80713. doi: 10.1371/journal.pone.0080713, Epub 2013/12/19. [PubMed: 24348913]
- Yeh FC, Wedeen VJ, Tseng WY, 2010. Generalized q-sampling imaging. *IEEE Trans. Med. Imaging* 29 (9), 1626–1635. doi: 10.1109/TMI.2010.2045126, Epub 2010/03/23. [PubMed: 20304721]
- Yeh F–C, 2020. Shape analysis of the human association pathways. *Neuroimage* 223, 117329. doi: 10.1016/j.neuroimage.2020.117329.
- Yu M, Linn KA, Cook PA, Phillips ML, McInnis M, Fava M, et al. , 2018. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum. Brain Mapp* 39 (11), 4213–4227. doi: 10.1002/hbm.24241, Epub 2018/07/01. [PubMed: 29962049]
- Zhang F, Cetin Karayumak S, Hoffmann N, Rathi Y, Golby AJ, O'Donnell LJ, 2020a. Deep white matter analysis (DeepWMA): fast and consistent tractography segmentation. *Med. Image Anal* 65, 101761. doi: 10.1016/j.media.2020.101761, Epub 2020/07/06.
- Zhang F, Hoffmann N, Karayumak SC, Rathi Y, Golby AJ, O'Donnell LJ, 2019a. Deep white matter analysis: fast, consistent tractography segmentation across populations and dMRI acquisitions. *Med. Image Comput. Comput. Assist. Interv* 11766, 599–608. doi: 10.1007/978-3-030-32248-9_67, Epub 2020/06/20. [PubMed: 32558816]
- Zhang F, Noh T, Juvekar P, Frisken SF, Rigolo L, Norton I, et al. , 2020b. SlicerDMRI: diffusion MRI and Tractography Research Software for Brain Cancer Surgery Planning and Visualization. *JCO Clin. Cancer Inform* 4, 299–309. doi: 10.1200/CCI.19.00141, Epub 2020/03/29. [PubMed: 32216636]
- Zhang F, Wu Y, Norton I, Rathi Y, Golby AJ, O'Donnell LJ, 2019b. Test–retest reproducibility of white matter parcellation using diffusion MRI tractography fiber clustering. *Hum. Brain Mapp* 40 (10), 3041–3057. doi: 10.1002/hbm.24579. [PubMed: 30875144]
- Zhong J, Wang Y, Li J, Xue X, Liu S, Wang M, et al. , 2020. Inter-site harmonization based on dual generative adversarial networks for diffusion tensor imaging: application to neonatal white

matter development. *Biomed. Eng. Online* 19 (1), 4. doi: 10.1186/s12938-020-0748-9. [PubMed: 31941515]

Zöllei L, Jaimes C, Saliba E, Grant PE, Yendiki A, 2019. TRActs constrained by UnderLying INfant anatomy (TRACULInA): an automated probabilistic tractography tool with anatomical priors for use in the newborn brain. *Neuroimage* 199, 1–17. doi: 10.1016/j.neuroimage.2019.05.051, Epub 2019/05/24. [PubMed: 31132451]

1. Cai LY, et al. , MASiVar: Multisite, Multiscanner, and Multisubject Acquisitions for Studying Variability in Diffusion Weighted Magnetic Resonance Imaging. *bioRxiv*, 2020: p. 2020.12.03.408567.
2. Ning L, et al. , Cross-scanner and cross-protocol multi-shell diffusion MRI data harmonization: Algorithms and results. *NeuroImage*, 2020. 221 : p. 117128. [PubMed: 32673745]
3. Tax CM, et al. , Cross-scanner and cross-protocol diffusion MRI data harmonization: A benchmark database and evaluation of algorithms. *Neuroimage*, 2019. 195 : p. 285–299. [PubMed: 30716459]

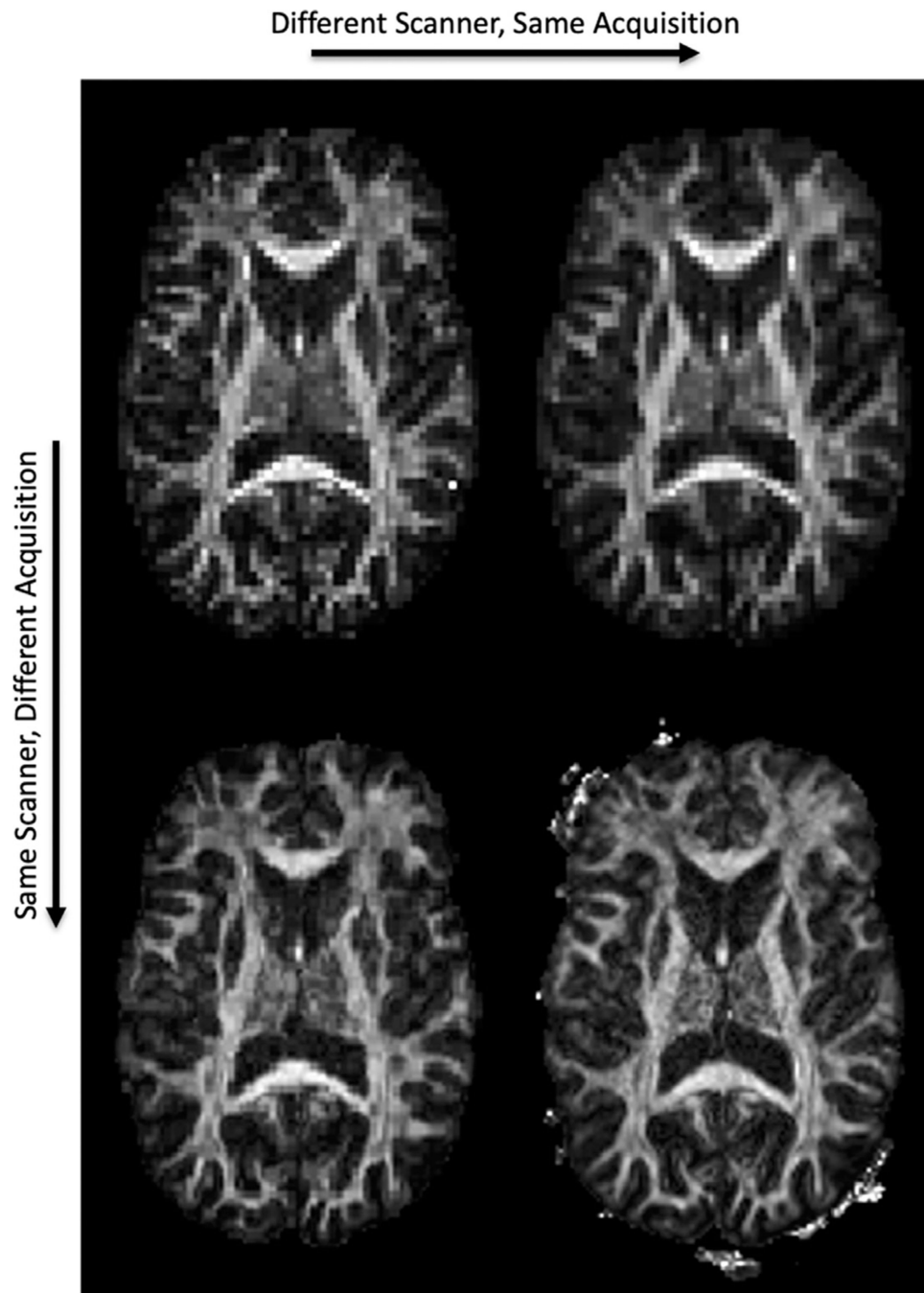


Fig. 1. Microstructure varies across scanners and across acquisitions. An FA map is shown, derived from the same subject, on two scanners (Siemens Prisma, left; Siemens Connectom, right) and two acquisitions (standard acquisition, top; state-of-the-art acquisition, bottom). See Methods for scanner and acquisition details.

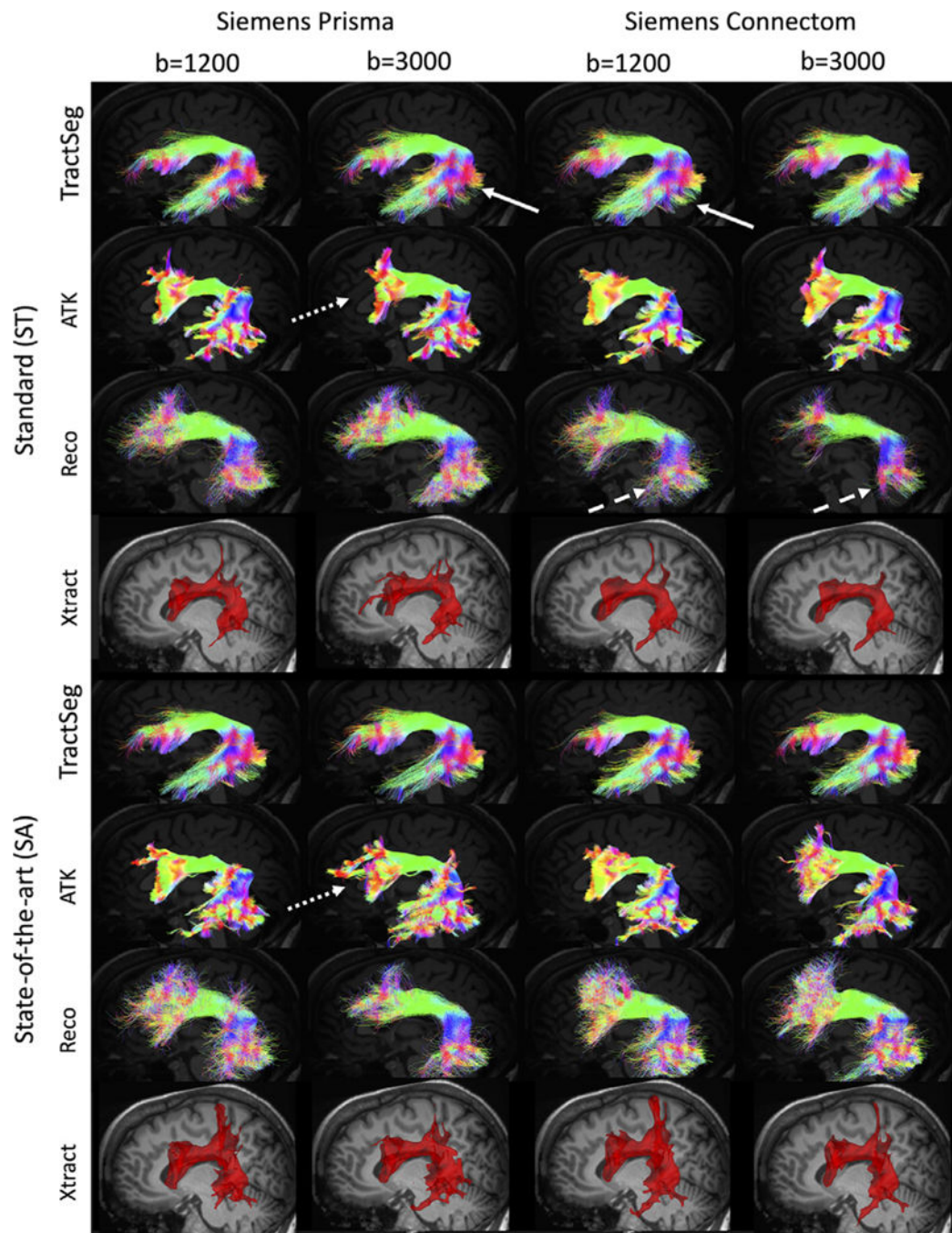


Fig. 2. Tractography varies across scanners, acquisitions, b-values, and bundle segmentation methods. On the same subject, the arcuate fasciculus is shown for each of the 4 bundle segmentation methods, for two scanners and two acquisitions. Note that the pathway is visualized as streamlines for TractSeg, ATK, and Reco but a probability density map for Xtract. Arrows highlight visible examples of differences in streamlines across scanners (solid arrows), across acquisition (dotted arrows), and across b-values (dashed arrows).

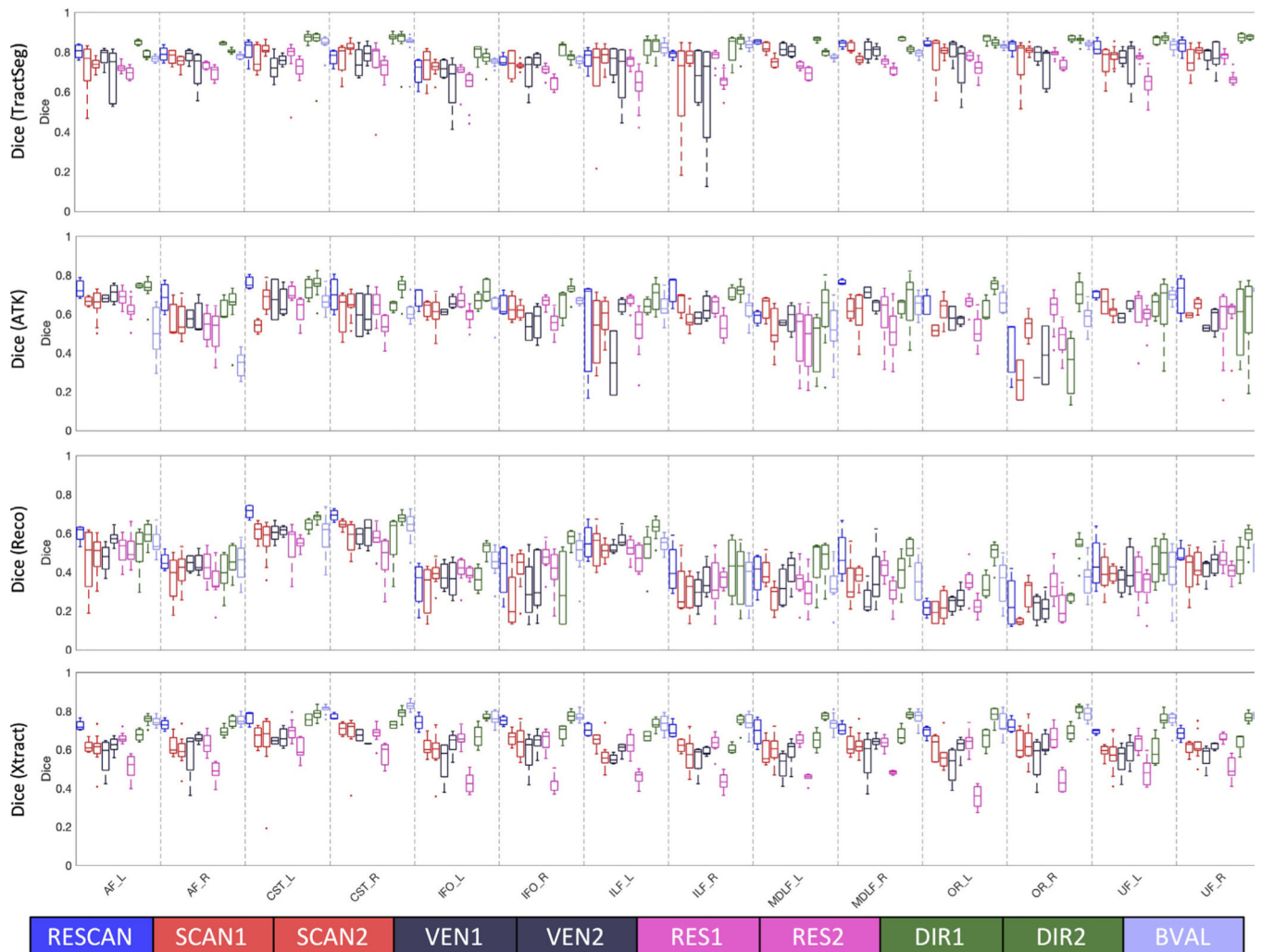


Fig. 3. Reproducibility is dependent on all investigated effects, and varies by pathway and by dissection method. Effects of scan-rescan (RESCAN; blue), scanners (SCAN1, SCAN2; red), vendor (VEN1, VEN2; dark purple), resolution (RES1, RES2; pink), diffusion directions (DIR1, DIR2; green) and b-value (BVAL; light purple) on dice overlap coefficient for individual bundles. Results are shown for 14 fiber bundles that are common to each tractography workflow. Please see Appendix for bundle abbreviations.

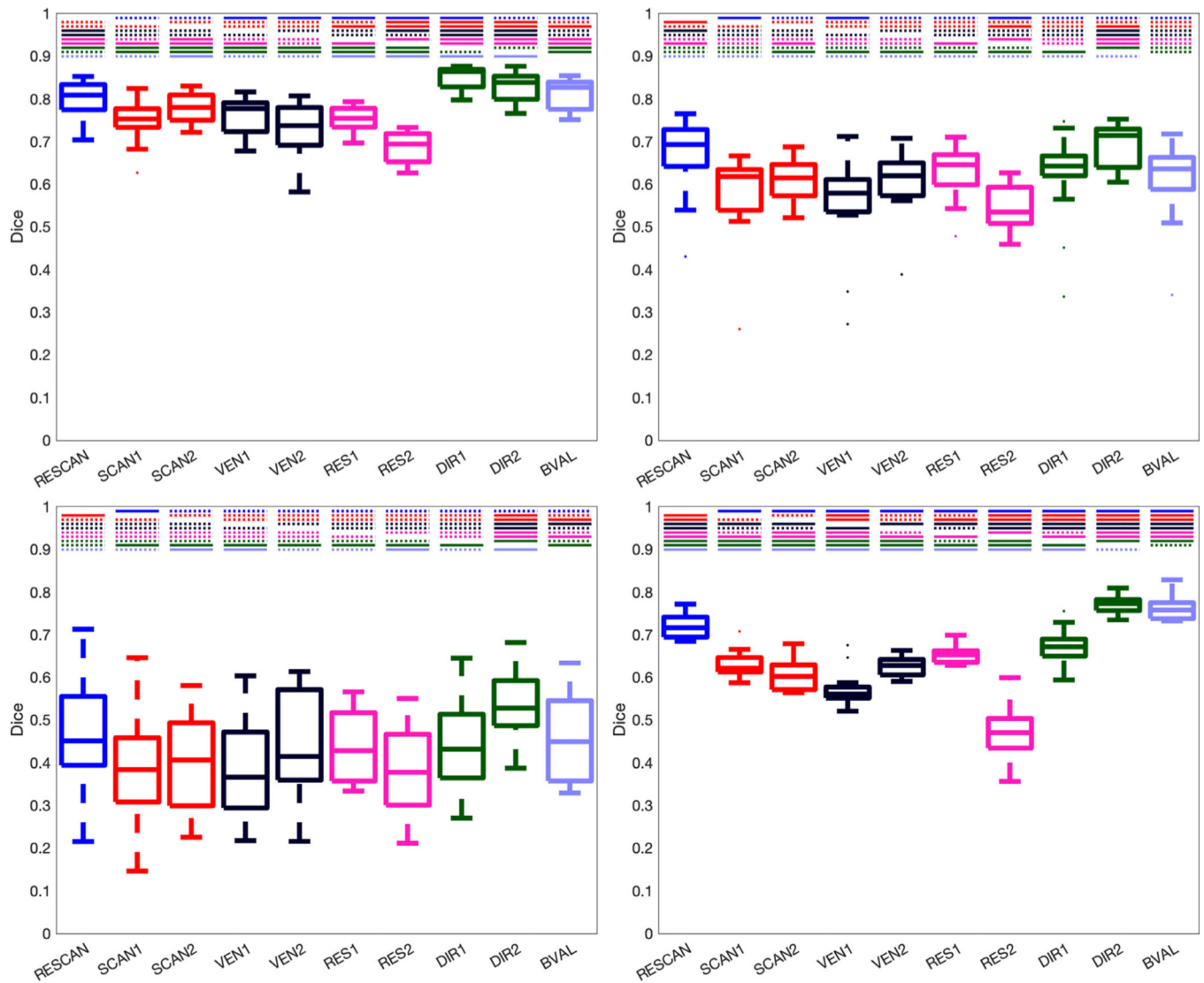


Fig. 4.

Reproducibility is dependent upon all investigated effects, and each bundle segmentation methods is affected differently. Effects of scan-rescan (RESCAN; blue), scanners (SCAN1, SCAN2; red), vendor (VEN1, VEN2; dark purple), resolution (RES1, RES2; pink), diffusion directions (DIR1, DIR2; green) and b-value (BVAL; light purple) on dice overlap coefficient for all fiber bundles dissected using each technique. For each, a Wilcoxon signed rank test is performed to investigate differences in effects. Statistically significant results ($p < .05/45/4$ comparisons) are shown as a solid line, and those not reaching statistical significance are shown as dashed line. Tractseg (top-left), ATK (top-right), Reco (bottom-left), and Xtract (bottom-right).

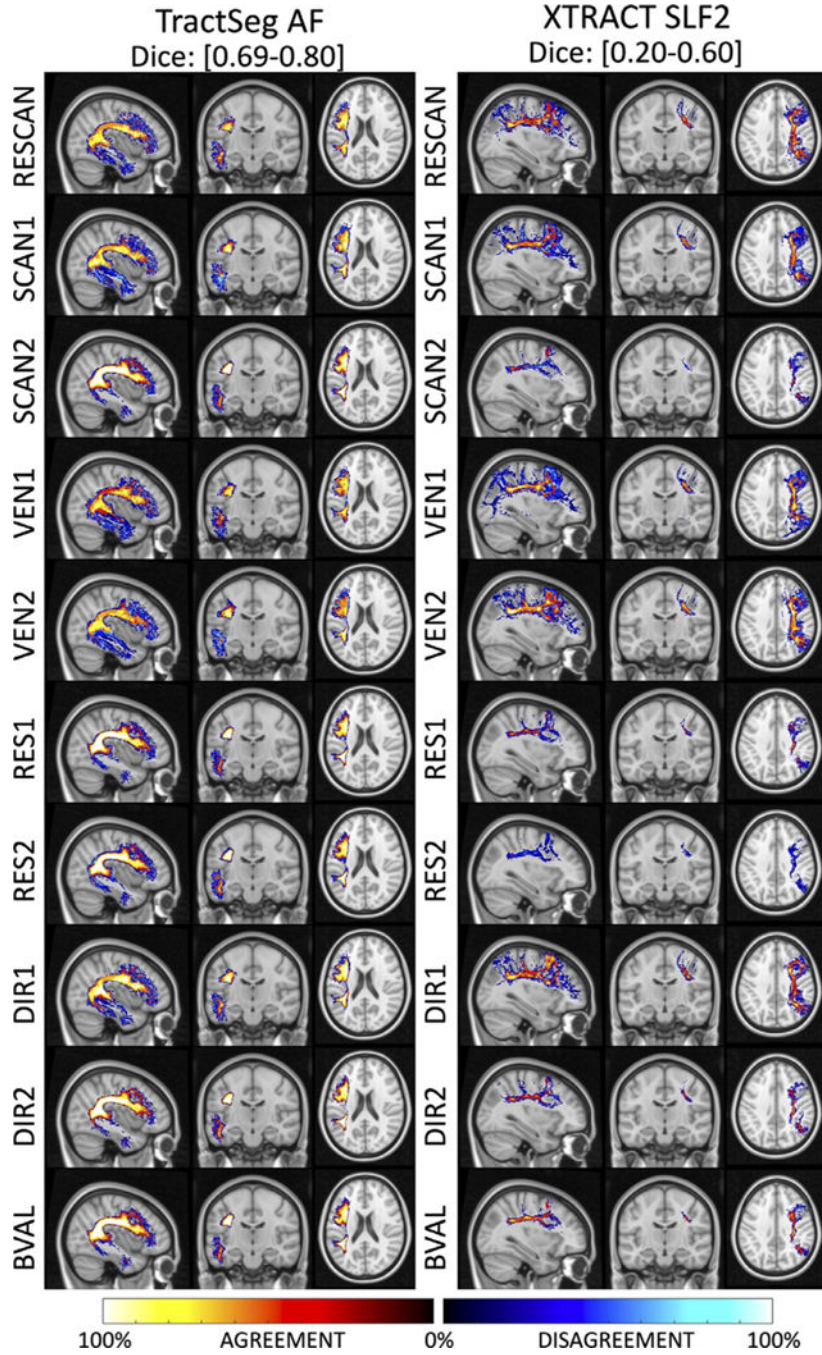


Fig. 5. Locations of agreement and disagreement across effects. Maps are computed by overlaying (for each source of variation), maps of where there is overlap (i.e. agreement) and non-overlap (disagreement), averaged across all subjects. For each effect, the percent agreement indicates areas where a pathway is consistently located and is shown using a “hot” colormap, while the percent disagreement indicates areas without consistent overlap and is shown using a “cold” colormap. Results are shown for a highly reproducible pathway (AF_L dissected using TractSeg) and for a less reproducible pathway (SLF2 dissected using XTRACT). Note that even though disagreement is abundant, it does

not consistently occur (i.e.,% disagreement remains low; black and dark blue) suggesting no systematic bias due to effects, and disagreements are largely attributed to the stochastic nature of the tractography and dissection process.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

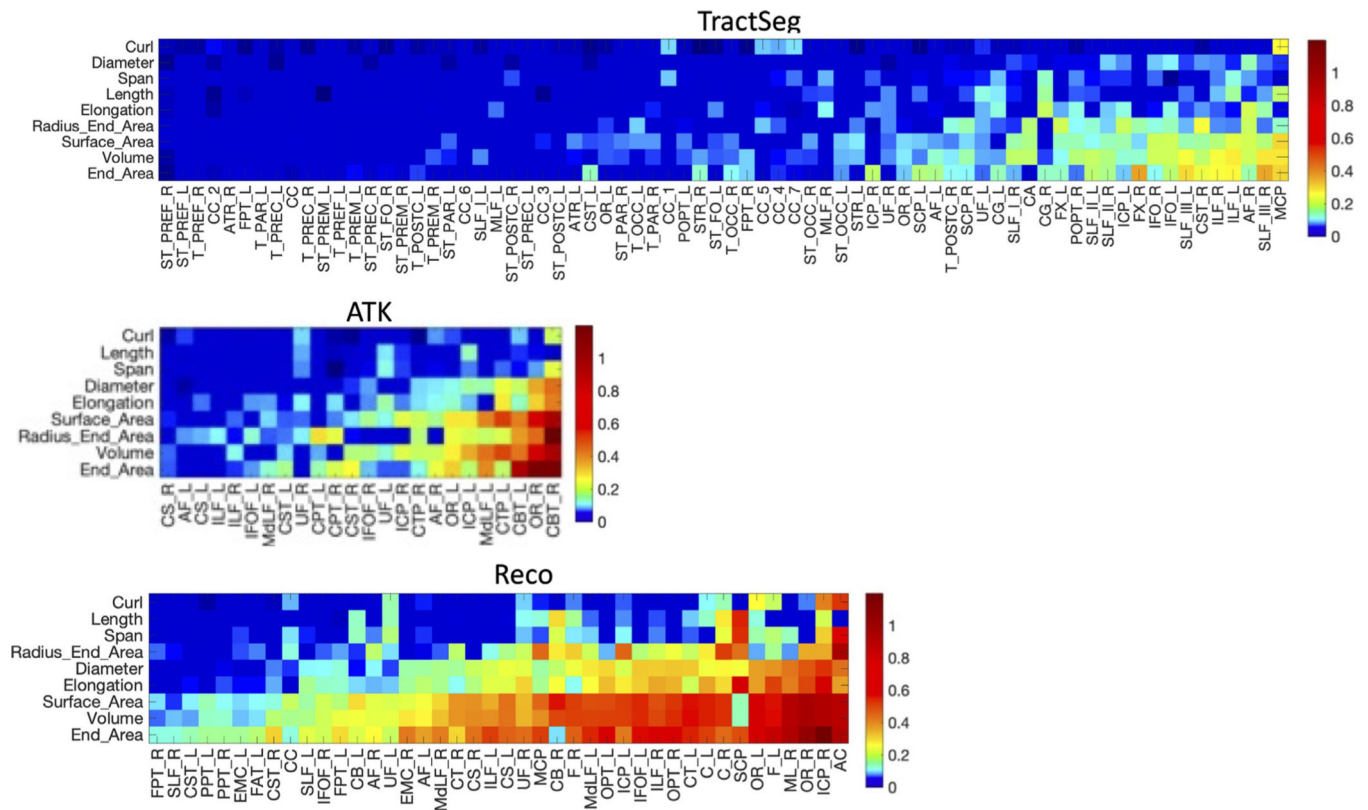
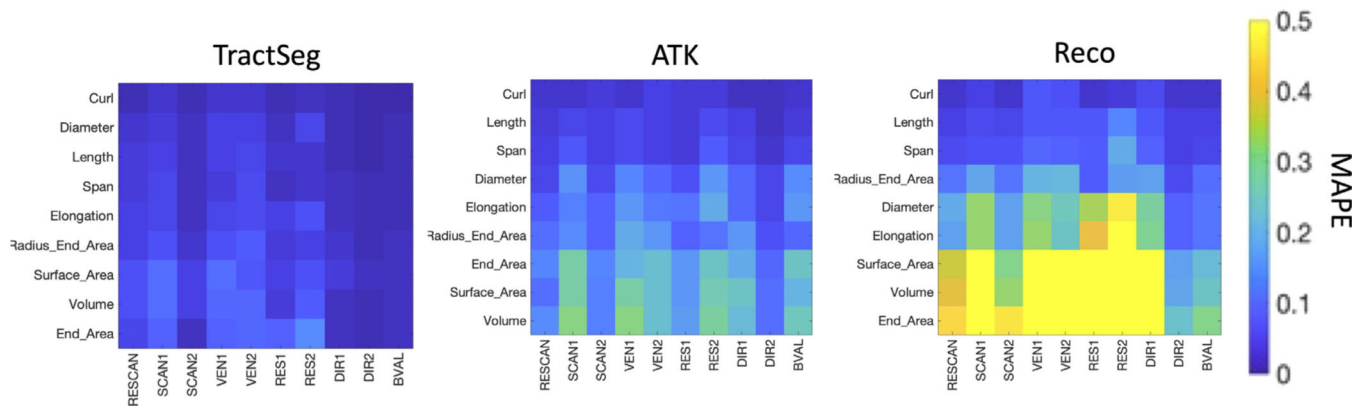


Fig. 6. Reproducibility of pathway shape features depends on pathway and bundle dissection method.

Reproducibility is shown as a MAPE for each tractography segmentation method. For each method, the features are ordered (from top to bottom) from lowest to highest average MAPE, and pathways are similarly ordered (from left to right) from lowest to highest average MAPE. Note that the colormap is nonlinear to better highlight MAPE between 0 and 0.10. Many shape features are highly reproducible, and with differences across pathways and bundle dissection methods. Please see Appendix for bundle abbreviations.

**Fig. 7.**

Variability of shape features is influenced by scanner, vendor, acquisition, and b-value. Variability is shown as MAPE for each TractSeg, ATK, and Reco methods, for scan-rescan (RESCAN), scanners (SCAN1, SCAN2), vendor (VEN1, VEN2), resolution (RES1, RES2), diffusion directions (DIR1, DIR2) and b-value (BVAL). Values shown are averaged across all pathways within a bundle dissection method. Shape features are ordered (from top to bottom) from lowest to highest average MAPE. Many shape features are highly reproducible, and MAPE is influenced by all effects investigated.

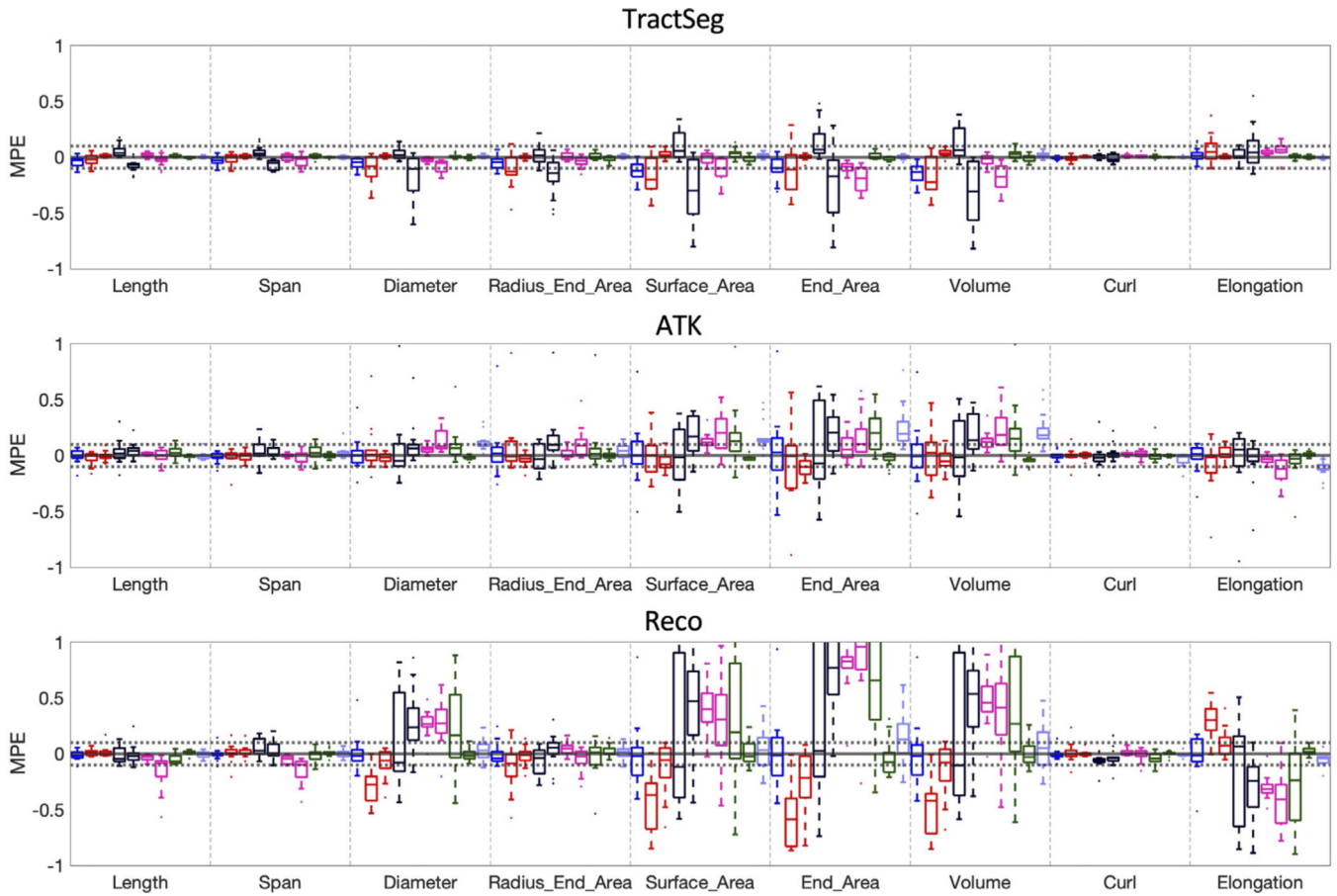


Fig. 8.

Sources of variation may introduce bias in shape features. The mean percent variation (MPV), i.e., the signed MAPE, is shown for each bundle segmentation method, for all features, with the distribution across fiber pathways. A distribution not centered on 0 suggests systematic differences introduced by the given effect. For interpretation, RESCAN (repeat 2 – repeat 1), SCAN1 (Philips Achieva scanner 2 – Philips Achieva scanner 1), SCAN2 (Siemens Connectome standard acquisition – Siemens Prisma standard acquisition), VEN1 (GE Discovery - Philips Achieva), VEN2 (Siemens Skyra - Philips Achieva), RES1 (Prisma state-of-the-art 30 directions - Prisma standard acquisition), RES2 (Connectom state-of-the-art 30 directions - Connectom standard acquisition), DIR1 (Philips Achieva 96 directions – Philips Achieva 32 directions), DIR2 (Prisma state-of-the-art 60 directions - Prisma state-of-the-art 30 directions), BVAL (Prisma standard-acquisition $b = 3000$ – Prisma standard-acquisition $b = 1000$).

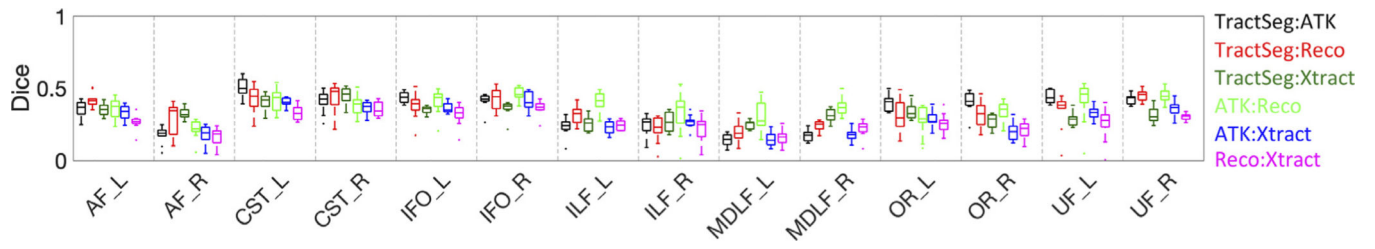


Fig. 9.

Different workflows result in low-to-moderate Dice overlap of the same pathways. Dice overlap coefficients for individual bundles, when measuring agreement between different bundle dissection methods. Please see Appendix for bundle abbreviations.

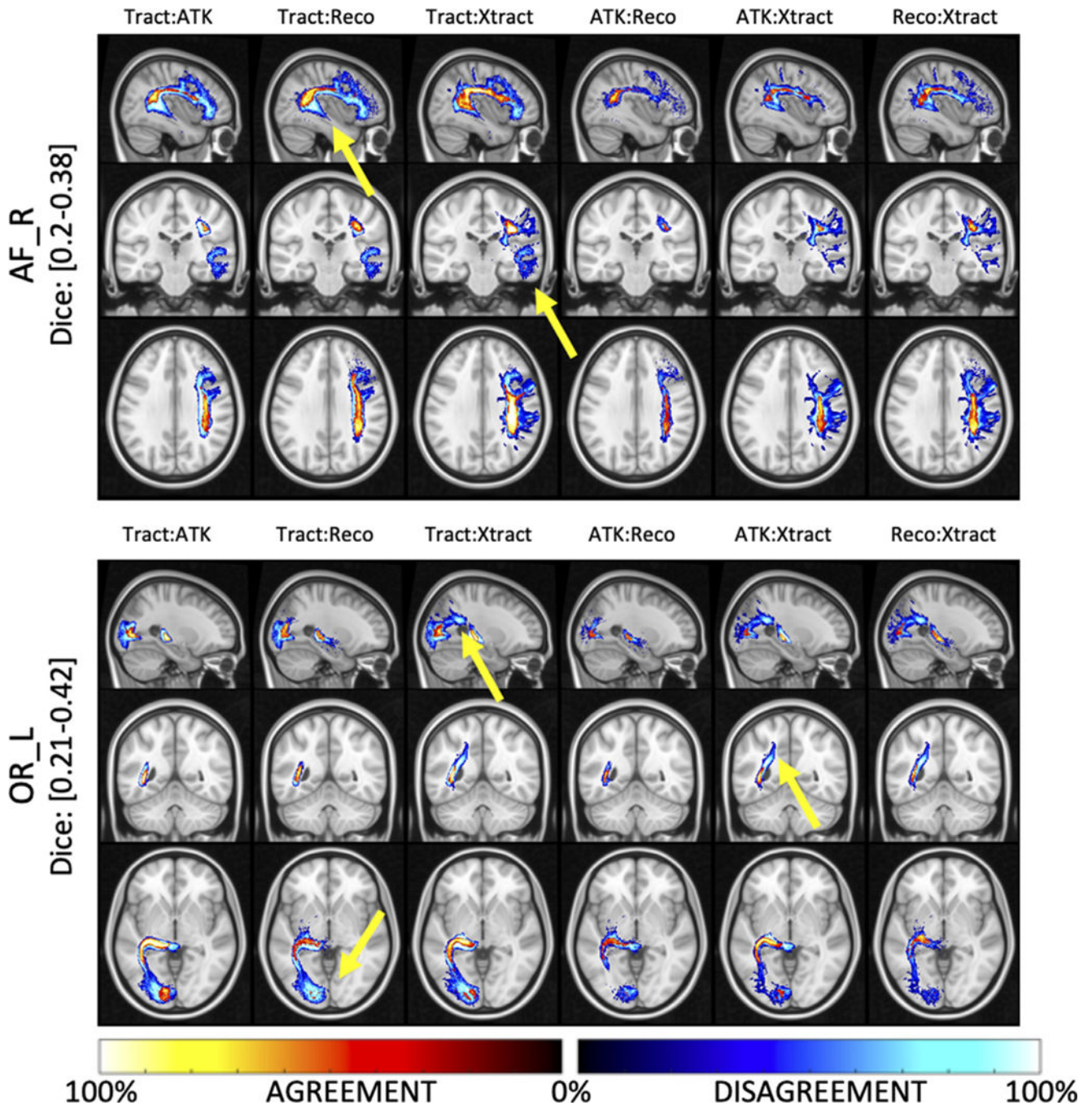


Fig. 10.

Locations of agreement and disagreement across bundle dissection methods. For each comparison, percent agreement indicates areas where methods agree in space and is shown using a “hot” colormap, while percent disagreement indicates areas where disagreement occurs and is shown using a “cold” colormap. Results are shown for two example pathways (AF_R and OR_L). Here, there are areas of high% disagreement between methods, indicating a consistent and reproducible difference between bundle dissection methods (highlighted by yellow arrows).

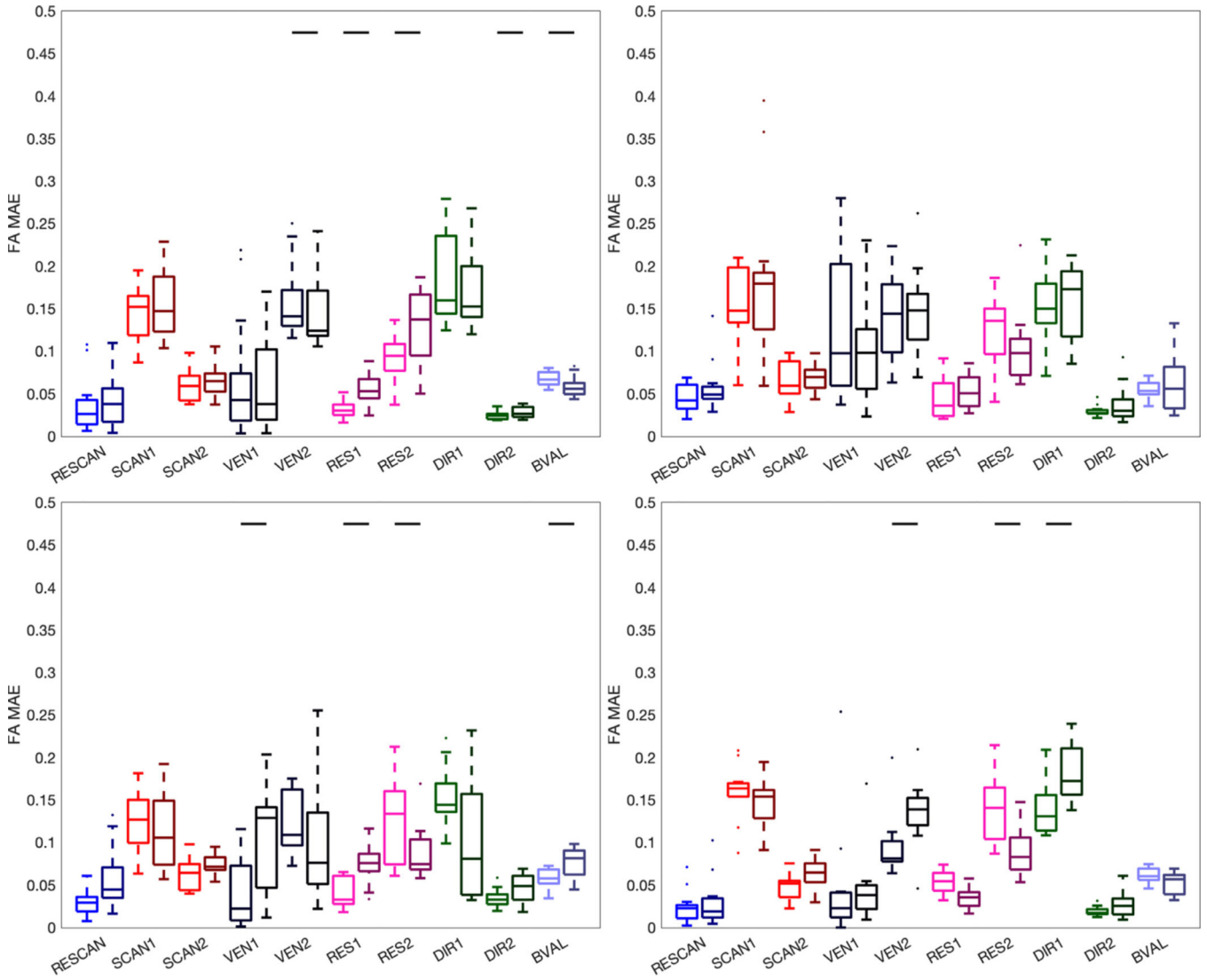


Fig. 11. Variation of FA. Effects of scan-rescan (RESCAN; blue), scanners (SCAN1, SCAN2; red), vendor (VEN1, VEN2; dark purple), resolution (RES1, RES2; pink), diffusion directions (DIR1, DIR2; green) and b-value (BVAL; light purple) on MAPE of the FA for all fiber bundles dissected using each technique. The left boxplots are indicative of the variability inherent due to each effect, whereas the darker-hued (right) boxplots indicate the added variability due to differences in tractograms. For each, a Wilcoxon signed rank test is performed to investigate whether tractography adds to (or removes) significant variance to this metric, and statistical significance is indicated by a solid black line.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

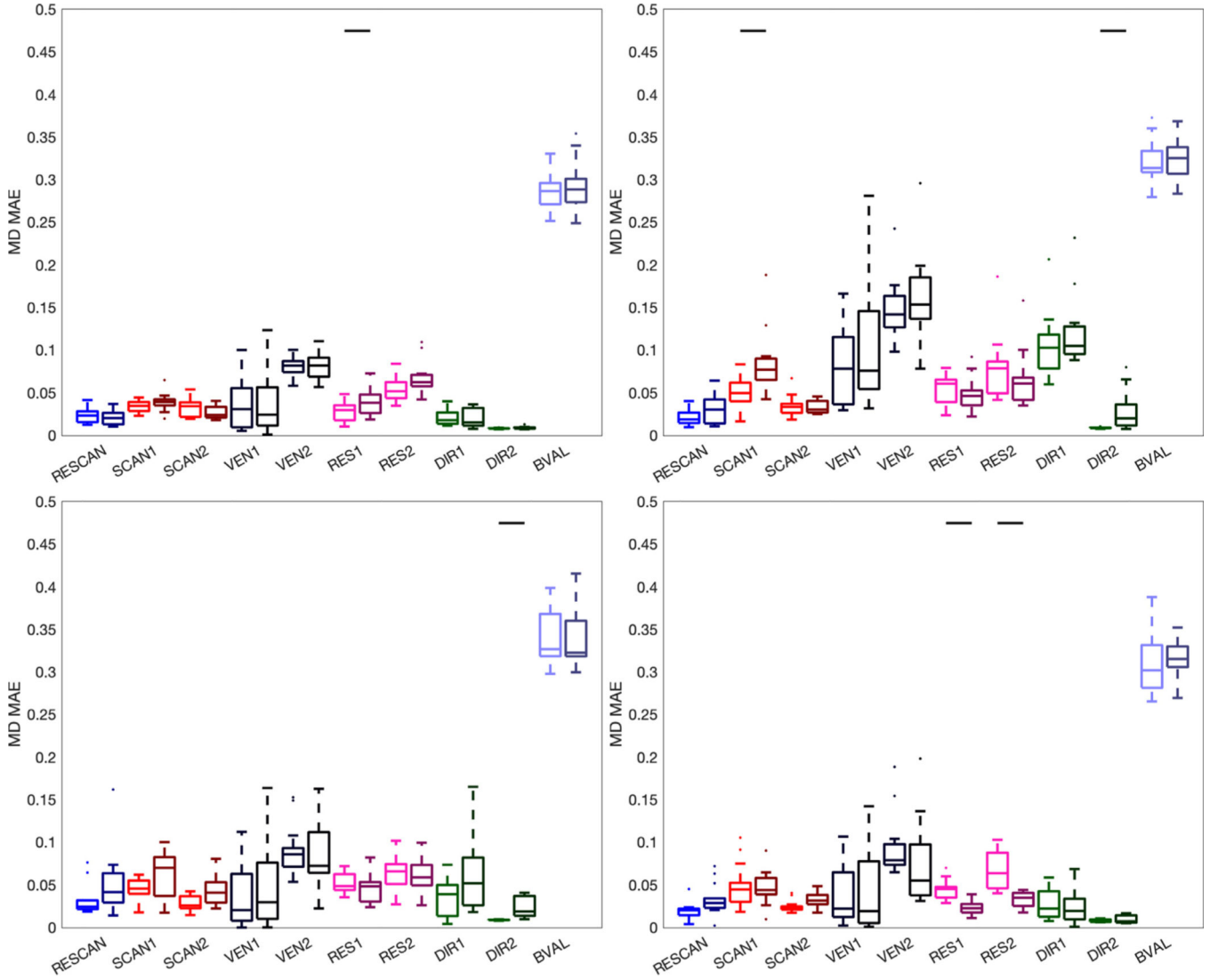


Fig. 12. Variation of MD. Effects of scan-rescan (RESCAN; blue), scanners (SCAN1, SCAN2; red), vendor (VEN1, VEN2; dark purple), resolution (RES1, RES2; pink), diffusion directions (DIR1, DIR2; green) and b-value (BVAL; light purple) on MAPE of the MD for all fiber bundles dissected using each technique. The left boxplots are indicative of the variability inherent due to each effect, whereas the darker-hued (right) boxplots indicate the added variability due to differences in tractograms. For each, a Wilcoxon signed rank test is performed to investigate whether tractography adds to (or removes) significant variance to this metric, and statistical significance is indicated by a solid black line.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript