Limits to anatomical accuracy of diffusion tractography using modern approaches

Schilling, Kurt G.; Nath, Vishwesh; Hansen, Colin; Parvathaneni, Prasanna; Blaber, Justin; Gao, Yurui; Neher, Peter; Aydogan, Dogu Baran; Shi, Yonggang; Ocampo-Pineda, Mario

Total number of authors:
42

Published in:
NeuroImage

Link to article, DOI:
10.1016/j.neuroimage.2018.10.029

Publication date:
2019

Document Version
Peer reviewed version

Citation (APA):
Schilling, K. G., Nath, V., Hansen, C., Parvathaneni, P., Blaber, J., Gao, Y., Neher, P., Aydogan, D. B., Shi, Y., Ocampo-Pineda, M., Schiavi, S., Daducci, A., Girard, G., Barakovic, M., Rafael-Patino, J., Romascano, D., Rensonnet, G., Pizzolato, M., Bates, A., ... Landman, B. A. (2019). Limits to anatomical accuracy of diffusion tractography using modern approaches. *NeuroImage*, *185*, 1-11. https://doi.org/10.1016/j.neuroimage.2018.10.029

NeuroImage 185 (2019) 1–11

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Limits to anatomical accuracy of diffusion tractography using modern approaches

Kurt G. Schilling [a,*], Vishwesh Nath [b], Colin Hansen [c], Prasanna Parvathaneni [c], Justin Blaber [c], Yurui Gao [a], Peter Neher [d], Dogu Baran Aydogan [e], Yonggang Shi [e], Mario Ocampo-Pineda [f], Simona Schiavi [f], Alessandro Daducci [f], Gabriel Girard [g], Muhamed Barakovic [g], Jonathan Rafael-Patino [g], David Romascano [g], Gaëtan Rensonnet [g], Marco Pizzolato [g], Alice Bates [g], Elda Fischi [g], Jean-Philippe Thiran [g,h], Erick J. Canales-Rodríguez [g,h], Chao Huang [i], Hongtu Zhu [i], Liming Zhong [j], Ryan Cabeen [e], Arthur W. Toga [e], Francois Rheault [k], Guillaume Theaud [k], Jean-Christophe Houde [k], Jasmeen Sidhu [k], Maxime Chamberland [l], Carl-Fredrik Westin [m], Tim B. Dyrby [n,o], Ragini Verma [p], Yogesh Rathi [q], M. Okan Irfanoglu [r], Cibu Thomas [s], Carlo Pierpaoli [r], Maxime Descoteaux [k], Adam W. Anderson [a,t], Bennett A. Landman [a,t,c]

[a] Vanderbilt University Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA
[b] Electrical Engineering & Computer Science, Vanderbilt University, Nashville, TN, USA
[c] Department of Electrical Engineering, Vanderbilt University, Nashville, TN, USA
[d] Division of Medical Image Computing, German Cancer Research Center, Heidelberg, Germany
[e] Laboratory of Neuro Imaging (LONI), USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[f] Computer Science Department, University of Verona, Verona, Italy
[g] Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[h] Radiology Department, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland
[i] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[j] Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou, China
[k] Sherbrooke Connectivity Imaging Lab (SCIL), Computer Science Department, Université de Sherbrooke, Canada
[l] Cardiff University, Brain Research Imaging Centre, School of Psychology, Cardiff, UK
[m] Department of Radiology, Harvard Medical School, Boston, MA, USA
[n] Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark
[o] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
[p] Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA
[q] Brigham and Women's Hospital, Harvard Medical School, USA
[r] National Institute of Biomedical Imaging and Bioengineering, NIH, Bethesda, MD, USA
[s] Section on Learning and Plasticity, Laboratory of Brain and Cognition, NIMH, Bethesda, MD, USA
[t] Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Keywords:
Diffusion MRI
Tractography
White matter
Tracer
Validation
Phantom
Connectivity

ABSTRACT

Diffusion MRI fiber tractography is widely used to probe the structural connectivity of the brain, with a range of applications in both clinical and basic neuroscience. Despite widespread use, tractography has well-known pitfalls that limits the anatomical accuracy of this technique. Numerous modern methods have been developed to address these shortcomings through advances in acquisition, modeling, and computation. To test whether these advances improve tractography accuracy, we organized the 3-D Validation of Tractography with Experimental MRI (3D-VoTEM) challenge at the ISBI 2018 conference. We made available three unique independent tractography validation datasets – a physical phantom and two ex vivo brain specimens - resulting in 176 distinct submissions from 9 research groups. By comparing results over a wide range of fiber complexities and algorithmic strategies,

* Corresponding author.
E-mail address: kurt.g.schilling.1@vumc.org (K.G. Schilling).

https://doi.org/10.1016/j.neuroimage.2018.10.029
Received 13 August 2018; Received in revised form 14 September 2018; Accepted 9 October 2018
Available online 11 October 2018

this challenge provides a more comprehensive assessment of tractography's inherent limitations than has been reported previously. The central results were consistent across all sub-challenges in that, despite advances in tractography methods, the anatomical accuracy of tractography has not dramatically improved in recent years. Taken together, our results independently confirm findings from decades of tractography validation studies, demonstrate inherent limitations in reconstructing white matter pathways using diffusion MRI data alone, and highlight the need for alternative or combinatorial strategies to accurately map the fiber pathways of the brain.

1. Introduction

Mapping the detailed structural connectivity of the human brain has been a major scientific goal for decades. Currently, the only safe, non-invasive method to map the white matter connections in the living brain is called diffusion MRI tractography (Conturo et al., 1999), which uses information about the displacement of water molecules in the brain (Le Bihan et al., 1986) to map fiber pathways. For nearly two decades, tractography has been used to probe both the spatial extent (or trajectory) of white matter pathways, as well as the region-to-region (cortical-cortical) connectivity of the brain. These techniques have been applied not only by neuroscientists in order to elucidate fundamental insights about brain function, cognition, and development, as well as neurological diseases and disorders, but also by neurosurgeons for surgery planning (Essayed et al., 2017; Jones, 2010). Thus, the anatomical accuracy of tractography is critical for sound scientific conclusions or effective surgical outcomes. Specifically, tractography must be able to classify the presence or absence of connections in the brain (i.e. have high specificity and sensitivity), as well as precisely delineate the full spatial extent of the fiber pathways.

A number of validation studies have been carried out with the aim of determining the reliability of tractography - typically utilizing numerical simulations, physical phantoms, histological tracers, or comparisons against prior anatomical knowledge (Alexander and Barker, 2005; Alexander et al., 2002; Cote et al., 2013; Daducci et al., 2014; Donahue et al., 2016; Dyrby et al., 2007; Girard et al., 2014; Irfanoglu et al., 2012; Jones, 2003; Jones and Cercignani, 2010; Knosche et al., 2015; Leergaard et al., 2010; Maier-Hein et al., 2017; Ning et al., 2015; Reveley et al., 2015; Schilling et al., 2018b, 2018d; Thomas et al., 2014; Tournier, 2010; Wheeler-Kingshott and Cercignani, 2009). Together, this collection of studies have revealed pitfalls, uncertainties, and sources of error in the tractography process that may limit anatomical accuracy. For example, the sources of error can emerge during any stage of the tracking process: image acquisition, local voxel-wise reconstruction, and/or tracking streamlines from voxel to voxel. Specifically, with regard to image acquisition, it is well known that diffusion MRI (particularly with echo planar imaging (EPI)) is noisy, and prone to artifacts due to susceptibility gradients affecting EPI acquisitions, head motion, and eddy currents. These artifacts can lead to uncertainty in orientation estimates (Alexander et al., 2002; Jones, 2003), biases in diffusion indices (Wheeler-Kingshott and Cercignani, 2009), geometric distortion in pathways (Irfanoglu et al., 2012), all of which can result in anatomically incorrect tractography (Jones and Cercignani, 2010). Another source of error involves drawing inferences about local fiber orientation from the diffusion displacement profile. MRI voxels are typically on the scale of millimeters, and can contain hundreds of thousands of axons with a large number of potentially complex geometric configurations (see (Dyrby et al., 2018) for a review on diffusion validation and its relationship to basic brain anatomy). In particular, fibers with crossing, kissing, fanning, and curving configurations have been a subject of concern for many diffusion reconstruction algorithms (Leergaard et al., 2010; Ning et al., 2015; Tournier, 2010), resulting in incorrect and ambiguous estimates of fiber orientation (Daducci et al., 2014). In addition, these reconstructions have been shown to be dependent on data acquisition conditions (including signal-to-noise ratio, amount of diffusion weighting, and number of diffusion encoding directions), as well as axonal geometry (for example, the crossing fiber angle) (Alexander and Barker, 2005; Schilling et al., 2018d). Finally, the tracking process itself is known to be subject to biases or inaccuracies due to lengths of streamlines (Donahue et al., 2016), shape and size of pathways (Girard et al., 2014), cortical folding patterns (Reveley et al., 2015; Schilling et al., 2018b), ambiguity in pathways selection (Maier-Hein et al., 2017), and choices of tracking parameters (e.g., seeding and stopping criteria, step size, curvature thresholds) (Dyrby et al., 2007; Knosche et al., 2015). Together, these difficulties have limited the anatomical accuracy of past tractography algorithms (Cote et al., 2013; Thomas et al., 2014). Some authors even argued that the anatomical accuracy of diffusion MRI tractography is inherently limited because inferring fiber direction information from a water diffusion displacement profile is fundamentally a complex, underdetermined inverse problem (Thomas et al., 2014).

Recently, several advancements in image acquisition, diffusion modeling, computational strategies, and tracking algorithms have been achieved with the aim of addressing these tractography limitations. To test whether these developments improve tracking accuracy, we organized the 3-D Validation of Tractography with Experimental MRI (3D-VoTEM) challenge that took place at the 2013 IEEE International Symposium on Biomedical Engineering (ISBI) conference, which advances tractography validation using three different validation datasets: [1] a macaque dataset with a histological map of known tracer connections (Thomas et al., 2014), [2] a squirrel monkey dataset with registered histological sections of the same sample (Schilling et al., 2018a, 2019), and [3] a 3D physical fiber phantom with manually traced ground-truth pathways (Synaptive Medical, Toronto, ON).

This challenge differs from the conventional methods of validating tractography – rather than a researcher proposing a novel method or algorithm and evaluating this technique on proprietary datasets which can vary in a number of aspects, 3D-VoTEM provides image data and a reference standard to a number of independent research groups who can implement, parameterize, and optimize their choice of algorithms. Thus, this challenge serves as a platform to compare algorithms and results on the same data, and in a fair manner. Providing the community with three well-characterized, curated diffusion MRI and corresponding ground truth data, allows groups that may not have the resources or abilities to carry out animal experiments, histological processing, phantom construction, or MRI acquisitions to test their methodologies. In this way, this challenge facilitates validation from research groups that one group, acting alone, may be unable to perform due to limited resources, expertise, or hardware. In addition, tractography is performed by research groups that have tuned their setup for optimal performance, given their knowledge and experiences, rather than an individual research group evaluating many methods by simply evaluating an entire parameter space for one optimal solution of parameters. Past diffusion MRI challenges have utilized similar frameworks to assess reproducibility of tractography and fiber orientation reconstruction (Daducci et al., 2014). In other community challenges, the performance of tractography has been assessed qualitatively on neurosurgical datasets (Pujol et al., 2015), and quantitatively on simulated human images (Maier-Hein et al., 2017) and 2D phantoms (Cote et al., 2013; Neher et al., 2014), revealing the successes and limitations of a number of past reconstruction strategies and tracking algorithms. Expanding upon these, 3D-VoTEM, utilizes three independent sub-challenges which allows us to test the conclusions that individual research groups, validation studies, and tractography challenges have shown in the past. By evaluating the results of these three sub-challenges, each providing insights into the same

problems, we sought to characterize the anatomical accuracy of the current state-of-the-art of diffusion tractography methods. In addition, by comparing results across a range of validation strategies, fiber complexities, and algorithmic strategies, the results from this challenge confirm the pitfalls of tractography revealed by independent research groups, as well as provide a more comprehensive assessment of tractography's inherent limitations and successes than has been demonstrated previously.

2. Materials and Methods

2.1. Data and ground truth

The sub-challenges vary in both data acquisition and definition of ground truth. Example data and ground truth volumes are shown in Fig. 1. The first sub-challenge consisted of a high quality - high resolution, high signal-to-noise ratio, and high angular sampling - ex vivo macaque dataset (Fig. 1A) featured in previous validation studies (Reveley et al., 2015; Thomas et al., 2014). The two ground truth pathways were derived from anterograde tracer injections placed in the precentral gyrus (PCG) (Fig. 1A, red) and the ventral part of visual area V4 (V4v) (Fig. 1A, yellow), as described and characterized in (Schmahmann and Pandya, 2009). Gray and white matter regions of interest were manually delineated on the data in order to assess agreement between tracer and tractography. This dataset allows validation of region-to-region connectivity. The second sub-challenge is performed on an ex vivo squirrel monkey dataset (Schilling et al., 2017a, 2017b, 2018a), acquired at a coarser resolution (relative to brain volumes), a lower SNR, and fewer sampling directions (31 versus 114 for the macaque). The ground truth, is defined based on an anterograde and retrograde tracer injection in the primary motor cortex (M1) of the same brain. Image processing on histological slices allows extraction of the ground truth fiber pathways on a voxel-by-voxel basis (Fig. 1B) as well as the creation of a binary "ground-truth" fiber pathway (Fig. 1C). Gray and white matter regions of interest are defined based on additional histological stains. This sub-challenge allows validation of both region-to-region connectivity as well as voxel-wise spatial overlap between tractogram (tractography streamlines) and tracer. The final sub-challenge consists of data acquired on a biomimetic anisotropic diffusion phantom (Synaptive Medical, Toronto, ON) containing 16 separate fiber bundles (Fig. 1D). Image acquisition consists of an overnight scan on two different scanners (scanner "A" and scanner "B") in the same imaging facility, with multiple

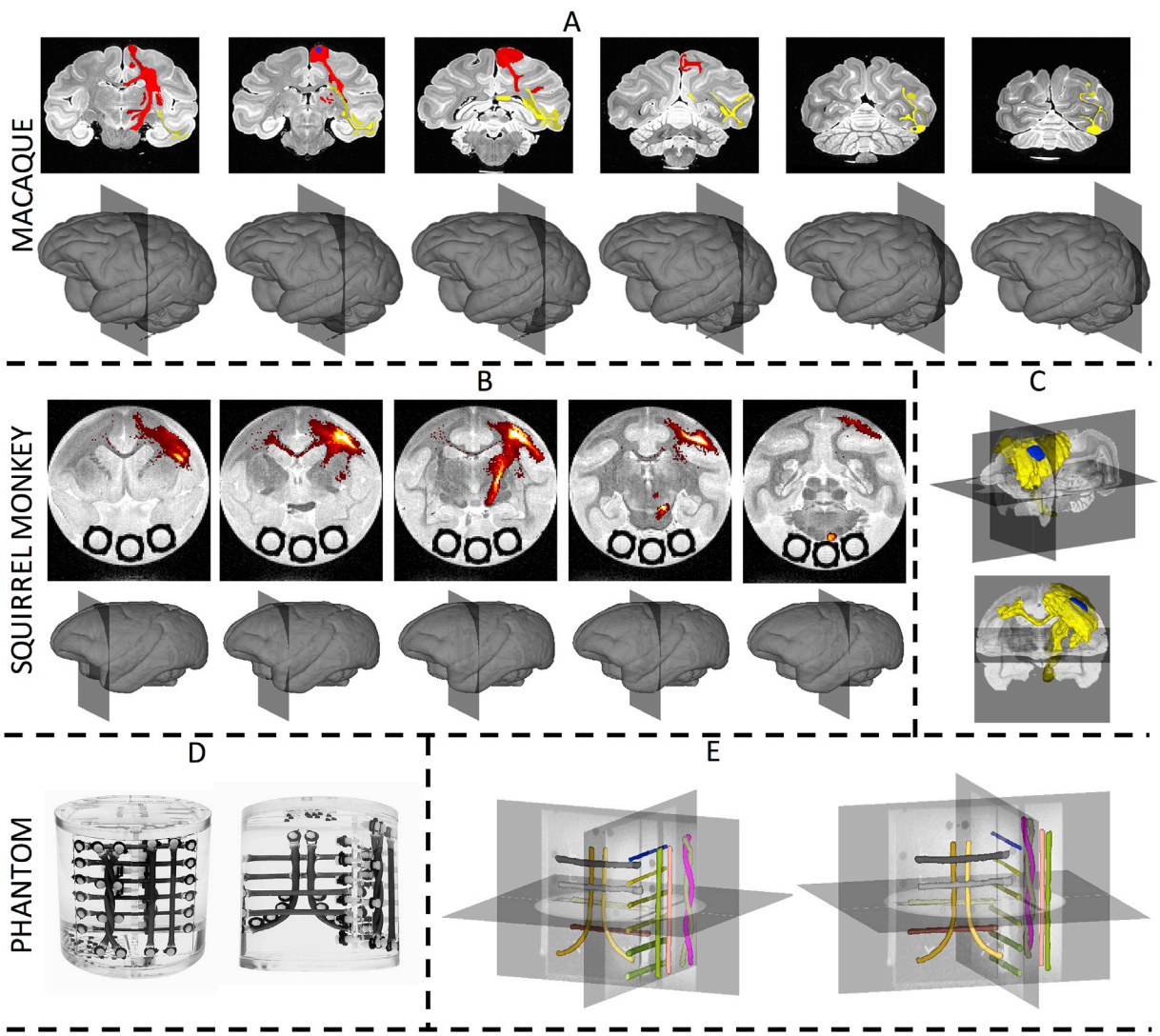

Fig. 1. Ground truth fiber pathways for all sub-challenges. The macaque challenge ground truth (A) is derived from tracer studies for both the PCG (red labels) and V4v (yellow labels) pathways, with seed region for PCG shown as a blue label. The squirrel monkey ground truth is derived from histological tracer density maps from the same subject (B), and visualized as a 3D volume rendering (C), with seed/injection regions shown in blue. The phantom (D) ground truth for all bundles is derived from manual tracing on a high resolution T1 weighted image (E).

diffusion weightings (b = 1000 and 2000 s/mm2), a large number of sampling directions (96 per b-value), and seven repetitions. The ground truth is manually defined on a high resolution T1-weighted image for all 16 bundles, and registered to dMRI space for a voxel-wise comparison of the spatial overlap between tractography and ground truth bundles (Fig. 1E). Details regarding the acquisition and processing procedures, as well as ground truth creation, are described below. All animal procedures followed the *Guide for the Care and Use of Laboratory Animals* and were approved by appropriate Animal Care and Use Committees.

2.2. Sub-challenge 1 – ex vivo macaque

2.2.1. Data description

The provided dataset is the one used and described in detail in (Thomas et al., 2014). Briefly, the images were acquired from an ex-vivo fixed macaque brain at 0.25 mm isotropic resolution. The diffusion weighted-images (DWIs) contain 7 vol with b = 0 s/mm^2 and 114 vol with b = 4900 s/mm^2 (with small variations due to the effects of the imaging gradients) (scanning time≈71 h, SNR≈40). The data were preprocessed using the TORTOISE software package (Pierpaoli et al., 2010) and were corrected for eddy current distortions and motion-like artifacts caused by frequency drifts.

2.2.2. Ground truth pathways

Two ground truth pathways were derived from the anterograde tracer injections placed in (i) the precentral gyrus corresponding to the foot region of the primary motor cortex and (ii) rostroventral part of the occipital region corresponding to the ventral part of area V4 (V4v) and the adjacent ventral area V3 – as described and characterized in (Schmahmann and Pandya, 2009). The tracer-labeled regions of interest were transferred to the same space as the diffusion data. In addition, gray matter and white matter regions of interest were manually delineated on the high-resolution data in order to assess agreement between tracer and tractography results.

2.3. Sub-challenge 2 – ex vivo squirrel monkey

2.3.1. Tracer injection

The histological ground truth data is acquired on a squirrel monkey brain. Here we utilize a commonly used neuroanatomical tracer for studying neuronal pathways, biotinylated dextrane amine (BDA). Because it is transported both anterograde and retrograde, BDA yields sensitive and detailed labeling of both axons and terminals, as well as neuronal cell bodies. This tracer relies on axonal transport systems; thus BDA injection is performed prior to ex vivo imaging. Under general anesthesia using aseptic techniques, BDA was injected into left hemisphere M1 cortex. Eight injections were made in order to cover a large M1 region representing the forearm as identified by intracortical microstimulation. After surgery, the monkey was allowed to recover from the procedure, giving the tracer sufficient time to be transported along axons to all regions connected to M1.

2.3.2. MRI imaging

For ex vivo scanning, the brain was perfusion fixed with 4% paraformaldehyde preceded by rinse with physiological saline. The brain was removed from the skull and stored in buffered saline overnight. The next day, the brain was scanned on a 9.4 T Varian scanner. Diffusion-weighted imaging was performed using a pulsed gradient spin echo multi-shot spinwarp imaging sequence with full brain coverage (TR = 5.2s, TE = 26 ms, number of diffusion gradient directions = 31, b = 0, 1200s/mm^2, voxel size = 300 × 300 × 300 μm^3, data matrix = 128 × 128 × 192, number of acquisitions = 10, SNR≈25, scanning time≈50 h). The b value used in this experiment was lower than is optimal for diffusion studies in fixed tissue, due to hardware limitations. A low b value decreases the available diffusion contrast-to-noise ratio (CNR) in the image data, which has the same effect as higher image noise. To compensate for this shortcoming, we extended the scan time to 50 h, which yielded a CNR comparable to in vivo human studies (equivalent to an in vivo study with mean diffusivity = 0.7 × 10−3 mm^2/s and SNR≈20). All sub-challenge data will be distributed and analyzed directly in the space in which diffusion data were acquired.

2.3.3. Histological acquisition

Following ex vivo MRI scanning, the brain was frozen and cut serially on a microtome in the coronal plane at 50 μm thickness. Prior to cutting every third section (i.e., at 150 mm intervals), the surface of the frozen tissue block was photographed using a Canon digital camera (image resolution = 50 μm/pixel, image size = 3330 × 4000 pixels, number of images per brain ~280), mounted above the microtome. Every 6th section (approximately the size of an MR voxel) is processed for BDA to trace pathways associated with the M1 cortex. Whole-slide Brightfield microscopy was performed using a Leica SCN400 Slide Scanner at 20x magnification, resulting in a maximum in-plane resolution of 0.5 μm/pixel.

2.3.4. Ground truth M1 connectivity

The "ground truth" connectivity of the injection area was determined by the presence of BDA-labeled axons in our high-resolution histology, which displayed as brown in the digital images. BDA-labeled fibers were segmented and counted following a series of morphological processes: top-hat filtering was performed to correct uneven illumination, global thresholding to extract fibers (segmenting brown [r/g/b = 165/42/42] using the "colorseg" function available on MathWorks File Exchange), and morphological operations to remove non-fiber objects (objects less than 11 pixels, empirically chosen) and to remove branch points of overlapping fibers. Histological images were down-sampled to the resolution of the MRI-data (300 μm isotropic), and the number of BDA fibers per voxel was counted, resulting in BDA density maps. These BDA density maps represent the ground truth "strength of connections" to the M1 injection area.

A total of 71 gray matter and white matter regions of interest were defined in MRI-space, using both histological and MRI-derived information, as described in (Gao et al., 2014, 2016; Schilling et al., 2017a), and retrieved from the squirrel monkey brain atlas (Schilling et al., 2017b), in order to assess connectivity agreement between tracer and tractography.

2.3.5. Registration

The multi-step registration utilized here is very similar to the registration procedure validated in an earlier study (Choe et al., 2011), which showed that the accuracy of the overall registration was approximately one MRI voxel (~0.3 mm). From the Leica image file, the TIFF image stored at 128 μm/pixel (down-sample factor 256) was extracted and registered to the down-sampled photograph (256 × 256 pixels at a resolution of approximately 128 μm/pixel) of the corresponding tissue block using a 2D affine transformation followed by a 2D non-rigid transformation, semi-automatically calculated via the Thin-Plate Spline algorithm (Bookstein, 1989). Next, all down-sampled block face photographs were assembled into a 3D block volume and registered to the corresponding 3D MRI volume using a 3D affine transformation followed by a non-rigid transformation automatically calculated via the Adaptive Bases Algorithm (Rohde et al., 2003). The deformation fields produced by all registration steps were applied to processed histological data in order to transfer the ground truth histological pathways into the diffusion space for comparisons with tractography.

2.4. Sub-challenge 3 – anisotropic fiber phantom

2.4.1. Phantom construction

The Anisotropic Diffusion Phantom (Synaptive Medical, Toronto, ON) is a physical phantom containing complex geometries of anisotropic fibers that mimic the tissues of the brain. The phantom contains 16 flexible

fiber bundles, each containing as many as 4.4 million proprietary solid-core fibers held in place with a flexible casing. Pathways are aligned in orthogonal planes, as well as in curved (both 90° and helical curving), and kissing geometries to mimic complex nerve fibers of the brain, with bundle dimensions of magnitudes comparable to major white matter pathways in the human brain, ranging from 2 mm up to 6 mm diameter bundles. The phantom is filled with distilled water.

2.4.2. MRI imaging

MR scans were performed on two scanners, both Philips 3.0T systems. The 16 cm diameter phantom was imaged for both structural and diffusion contrasts. The structural scan utilized a 3D MPRAGE sequence to acquire a T1 contrast (TE/TR = 3.6/8 ms, Matrix = 256 * 256, Resolution = 0.88*0.88 mm, slice thickness = 1.0 mm). A low-resolution diffusion contrast was acquired using a 2D EPI diffusion weighted sequence (TE/TR = 75 ms/9.65s, Matrix = 72*72, resolution = 2.25*2.25 mm, slice thickness = 2.5 mm). 96 diffusion directions were acquired, uniformly sampled over a sphere, at b-values of 1000 s/mm2 and 2000 s/mm2. Non-diffusion weighted images were acquired between every 8 diffusion-weighted images. Sampling was performed with phase encoding both anterior to posterior, and repeated posterior to anterior, in order to allow pre-processing for motion, eddy currents, and susceptibility distortions. This series of scans (2 b-values, 96 uniformly distributed directions, with two phase encoding directions each) was repeated 7 times on each scanner.

2.4.3. MRI data processing

Diffusion MRI pre-processing was performed in the coordinate system that the data were acquired in. Steps included correction for movement, susceptibility induced distortions, and eddy currents using FSLs *topup* and *eddy* algorithms [5]. The gradient tables were rotated based on the transformations obtained from the corrections.

2.4.4. Ground truth

Ground Truth was manually delineated for each bundle on the T1-weighted high resolution image, separately for each scanner, using ITK Snap (www.itksnap.org, v 2.4.0). For each scanner, the T1-weighted image was registered to the average non-diffusion weighted image using 3D affine followed by a 3D non-rigid registration (FSL Software Library v5.0 (Jenkinson et al., 2012)). Ground truth labels were individually transformed to diffusion space using nearest-neighbor interpolation.

2.5. Anatomical accuracy measures

Measures were calculated which describe the anatomical fidelity of the resulting tractograms, several of which have been previously employed in the validation literature. Here, measures are divided into ROI-based fidelity metrics and voxel-wise fidelity metrics. All metrics, both ROI-based and voxel-wise, are computed for all algorithms.

2.5.1. ROI-based measures

For both squirrel monkey and macaque sub-challenges, the ROI-based connectivity to seed regions was assessed using the white matter and gray matter regions of interest. Anatomical fidelity metrics of sensitivity, specificity, and Youden index were derived for all tractograms.

- Sensitivity – True positive rate; measures the proportion of positives (regions that are occupied by ground truth) that are correctly identified as such (using tractography). Sensitivity measures the ability to correctly detect all connections of the seed region.
- Specificity – True negative rate; measures the proportion of negatives (regions that do not contain ground truth) that are correctly identified as such (do not contain streamlines). Specificity measures the ability to correctly identify voxels that do not have connections with the seed region.
- Youden's J statistic – Sensitivity+Specificity-1; a statistic that captures the performance of a diagnostic test, and estimates the probability of an informed decision, ranging from −1 to 1. A value of 1 indicates a perfect test with no false positives or false negatives.

2.5.2. Voxel-wise measures

Voxel-wise measures were calculated for the phantom and squirrel monkey sub-challenges, because the ground truth volumes are defined voxel-wise. In the following, the Ground Truth volume is represented by G_j (j = 1,2, …, m) and tractography volume represented by T_i (i = 1,2, …, n).

- Bundle Overlap (OL) (Cote et al., 2013): The proportion of voxels that contain the ground truth volume that are traversed by at least one streamline. The OL describes how well tractography is able to describe the volume occupied by the ground truth and is defined as:

$$OL = \frac{|T_i \cap G_j|}{|G_j|} \tag{1}$$

where |•| denotes cardinality.

- Bundle Overreach (OR) (Cote et al., 2013): the number of voxels containing streamlines that are outside of the ground truth volume divided by the total number of voxels within the ground truth bundle:

$$OR = \frac{|T_i \setminus G_j|}{|G_j|} \tag{2}$$

where operator \ denotes relative complement operation.

- Dice Overlap Coefficient (D): measures the overall similarity between ground truth and tractography volume by taking twice the shared information (intersection) over the sum of the cardinalities:

$$D = \frac{2|T_i \cap G_j|}{|T_i| + |G_j|} \tag{3}$$

3. Results

3.1. Submissions

Although the submission site remains open (https://my.vanderbilt.edu/votem/submissions/), the data in this study includes only those submitted before the ISBI 2018 conference (April 4, 2018). Overall, 176 unique submissions were submitted across the challenges (58 for the macaque, 62 for squirrel monkey, and 56 for the phantom) from nine international research groups. Submissions ranged in complexity from open-sourced software, diffusion tensor based tractography with default software configurations to that of complex, multi-shell, in-house algorithms with extensive post-processing – with most featuring either reconstruction or tracking strategies developed in the last few years. Details of each submission are provided in Supplementary Tables 1–3. The most common reconstruction methods were some form of spherical deconvolution or multi-compartment models. Both deterministic and probabilistic algorithms were employed, with most utilizing some form of constraint on fractional anisotropy (FA), curvature, or anatomical mask. The seed regions (where tractography is initiated) provided along with the datasets were used as both true seeds as well as regions of interest after whole-brain tractography was performed. Standard pre-processing for susceptibility distortions, motion, and eddy currents was performed for all datasets, but very few groups used additional pre-processing steps (with the exception of denoising techniques), and post-processing included various filtering techniques, track grouping, and manual track selection. The measures of anatomical accuracy for each method and dataset are provided in Supplementary Tables 4–6,

allowing the algorithms to be compared for specific reconstruction and algorithm parameter choices. However, we do not attempt to declare an overall winner of the challenge (or sub-challenges), since this would require making arbitrary choices about the relative importance of different metrics and validation datasets.

3.2. Qualitative results

The tractography streamlines for randomly selected submissions are shown in Fig. 2 for the three sub-challenges. Qualitatively, there is large variability in the resulting connectivity profiles and pathways represented. Specifically, for the macaque and squirrel monkey, visualizing submitted streamlines shows a range in spatial extent from only connectivity nearby the seed region, to covering large expanses of the entire hemisphere. The phantom submissions generally capture the correct shape, position, and orientation of all 16 bundles, with most noticeable differences in sparsity of streamlines and thickness of pathways.

3.3. Region-to-region connectivity validation – sensitivity and specificity

For the macaque and squirrel monkey datasets the agreement between tracer and tractography results are evaluated using sensitivity and specificity measures, validating the ability of tractography to accurately map region-to-region (or seed-to-region, see Materials and Methods section) connectivity. Additionally, to identify the best combination of sensitivity and specificity, the Youden index (J) (Specificity + Sensitivity – 1) is computed, where a value of 1 indicates a perfect test and a value of zero indicates no predictive value. The results across all submissions are

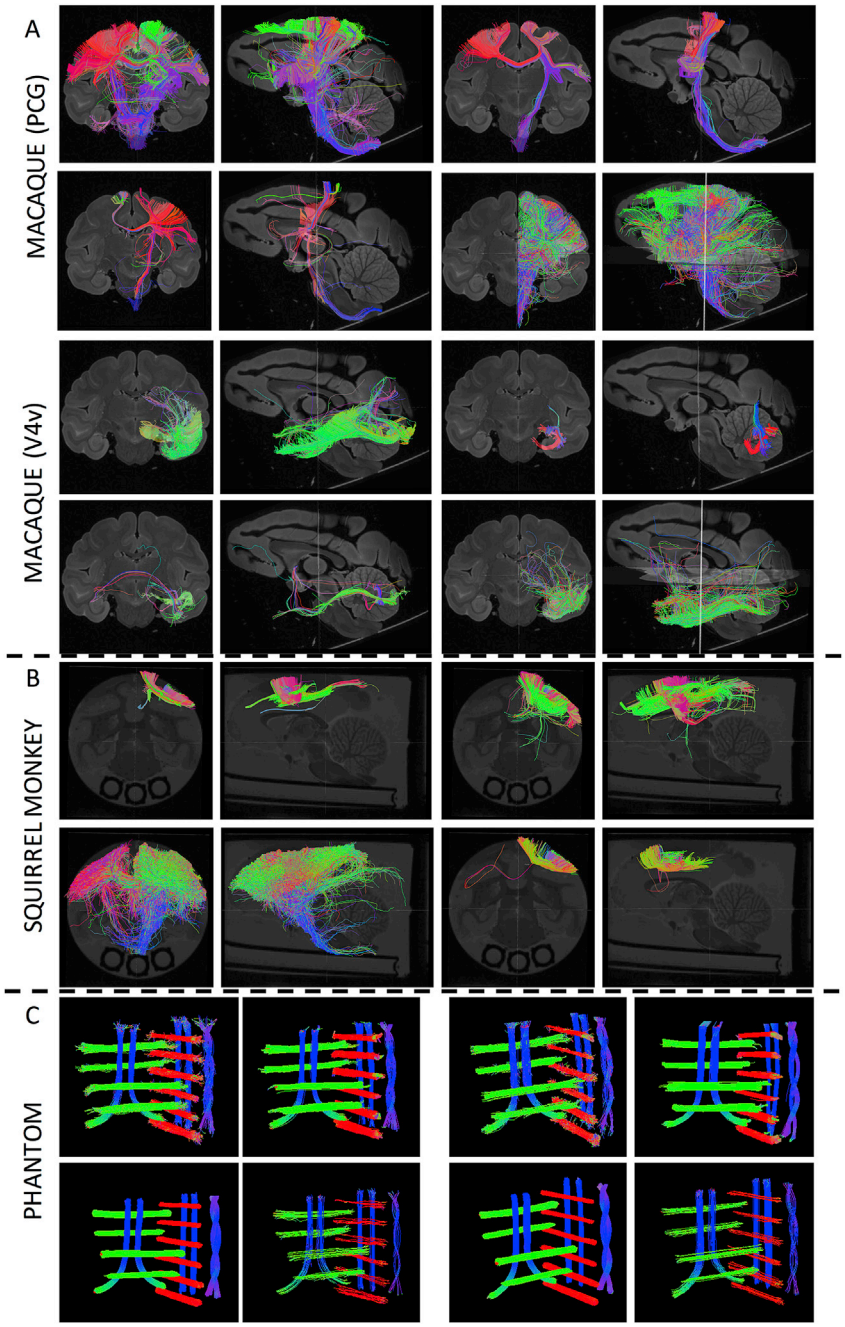

Fig. 2. Diffusion tractograms for randomly selected submissions. Tractography is shown in the coronal and sagittal planes, for both macaque pathways (A), the squirrel monkey pathway (B), and all 16 phantom bundles (C).

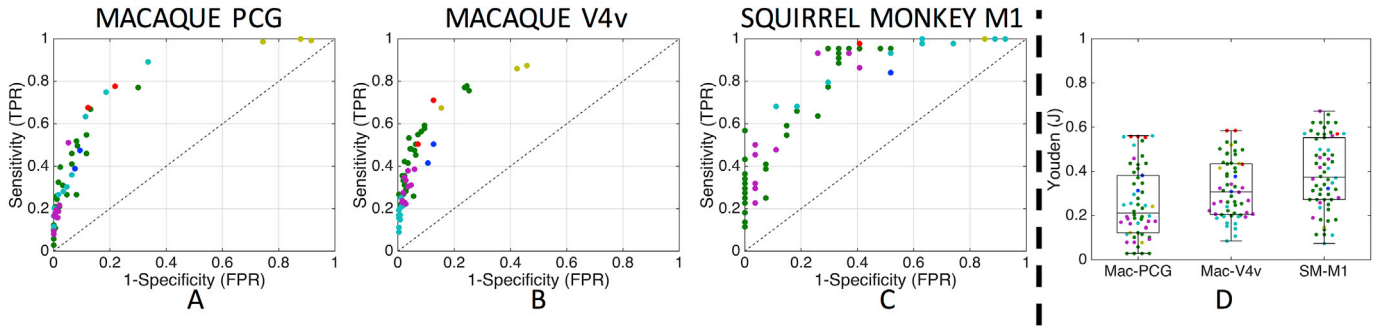

Fig. 3. Region-to-region connectivity validation. Results are shown as ROC curves on the macaque PCG (A) and V4v (B) pathways, and squirrel monkey M1 pathways (C). Boxplots of corresponding Youden values (D) are shown for both challenges. One marker is shown per submission, with marker colors indicating unique research groups.

shown as ROC curves in Fig. 3, where marker color indicates unique research groups. In both macaque and the squirrel monkey datasets, the main finding is that no algorithm or submission consistently identifies true positive pathways without also generating a large number of false positive pathways, and none consistently identify true non-connections without suffering a low true positive rate (i.e., an increase in sensitivity comes at the cost of a decrease in specificity, and vice-versa). For the macaque, most submissions result in high specificity values (with a large number of false negative connections), while the squirrel monkey algorithms typically lie at the extremes of the ROC plots.

Most submissions have relatively low predictive value, with median Youden indices of 0.21, 0.30, and 0.37 for macaque PCG, macaque V4v, and squirrel monkey M1 pathways, respectively (Fig. 3D). The highest Youden values for each pathway are only 0.56, 0.58, and 0.67. Thus, even the anatomical accuracy of the most predictive algorithms are suboptimal. The squirrel monkey results have a statistically significant (1-way ANOVA, $p < 0.01$) higher population mean Youden value than the macaque results – thus, in general, the algorithms provide slightly more anatomically accurate tracts on the squirrel monkey than macaque.

3.4. Spatial overlap validation – bundle overlap and overreach

A voxel-wise measure of spatial agreement between tracer and tractography is possible for the squirrel monkey with binary tracer data and phantom datasets with manually drawn tracts, because the ground truths are established in the same animal/phantom, making voxel-by-voxel comparisons possible. In these sub-challenges, we compute the bundle overlap: the proportion of voxels that contain the ground truth that are traversed by a streamline – and bundle overreach: the number of voxels containing streamlines outside the ground truth divided by the total number of voxels within the ground truth. In short, the overlap is a measure of the true positive rate (i.e., sensitivity) while the overreach is related to the false positive rate (i.e., specificity).

Plots of overlap and overreach for the squirrel monkey and both phantom scans (Fig. 4, A-C) show very similar results as the regional connectivity accuracy: algorithms that are successful at identifying the full extent of the pathways (high overlap) suffer from high overreach. In the squirrel monkey, algorithms that did not suffer from a significant overreach (<10%), often had very low overlap values, identifying less than 25% of the full histologically defined ground truth volume. While the phantom had significantly improved overlap values, many algorithms that recover the full bundle volumes can suffer from overreach as much as 1.5–5x the actual ground truth volume.

The Dice overlap coefficient (Fig. 4, D) has median values of 0.34, 0.46, and 0.51 for the squirrel monkey, phantom on scanner A, and phantom on scanner B, respectively, with maximum Dice coefficients reach 0.51, 0.63, and 0.72. The phantom submissions have statistically significant (1-way ANOVA, $p < 0.01$) higher Dice coefficients than that of the squirrel monkey, indicating an overall better voxel-wise accuracy.

4. Discussion

The 3D-VoTEM challenge combines and presents three separate tractography validation strategies, inviting ideas and algorithms from researchers from around the world, with the primary objective to determine whether recent technical advancements in diffusion MRI tractography can deliver anatomically accurate maps of the brain structural connectivity. More specifically, given the known limitations of these techniques, we asked if advances in algorithms, acquisition, and methodologies utilized in modern tractography techniques have improved anatomical accuracy. The key finding is that, despite a better understanding of limitations and pitfalls of these techniques, and considerable effort leading to advances in these algorithms, the anatomical accuracy of modern tractography approaches is still limited. Importantly, the limited anatomical accuracy is observed in three independent sub-challenges, each with algorithms created, developed, and

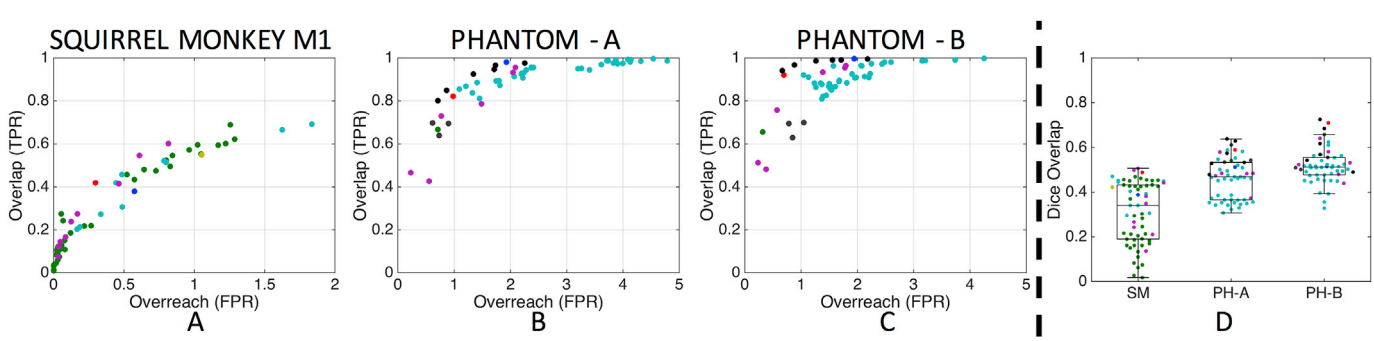

Fig. 4. Voxel-wise spatial overlap validation. Plots of overlap versus overreach are shown for the squirrel monkey (A) and phantom datasets on scanner A (B) and scanner B (C). Boxplots of the corresponding Dice overlap coefficients are shown for both challenges (D). One marker is shown per submission, with marker colors indicating unique research groups.

optimized by leading research groups in the field. These findings support the results and conclusions demonstrated over the last decade of validation studies, across species and phantoms, performed by individual research groups. Advances in the accuracy and reliability of tractography reconstructions will likely depend on the availability of shared validation and experimental datasets, standardized processing pipelines, and incorporating new information in the tracking process including better priors and alternative sources of tissue contrast.

4.1. Limits to accuracy

Importantly, we find consistent results across a diverse range of validation approaches. The sub-challenges vary in not only the systems under investigation (phantom versus non-human primates), but also acquisition (voxel size, angular resolution, SNR, diffusion weightings), complexity of pathways, and definition of ground truth. In all cases, algorithms that succeeded in recovering the true connections (high sensitivity or high overlap) consistently generated a large number of false positive connections (low specificity or high overreach), and no algorithm was highly informative or highly similar to the ground truth (high Youden or high Dice). In fact, most algorithms had surprisingly low connectivity predictive value and low spatial overlap with the true pathways. Thus, accuracy in tractography is not only hampered by a false positive problem (Maier-Hein et al., 2017), but many algorithms appear to be dominated by false negative connections (Aydogan et al., 2018).

While the accuracy tradeoffs have been consistent across challenges, differences in tractography performance between the challenges are apparent. This is expected, as tractography, and especially local reconstruction, are known to be heavily affected by the quality of the diffusion MRI acquisition. For example, it is generally assumed that many failures of tractography will be mitigated through improved angular and spatial resolution data. However, tractography in the macaque system resulted in less accurate connectivity measurements than tractography in the squirrel monkey system, despite significantly improved resolution, SNR, and diffusion sensitivity. Thus, differences in accuracy likely depend on the complexity of the pathway of interest, rather than acquisition quality alone. It should also be considered, however, that the ground truth data for the macaque brain were obtained from tracer studies performed in different animals so that interindividual variability in brain connections may have slightly lowered the accuracy value that could be reached with that dataset. Similarly, the phantom, with relatively sparse, well-defined, and less-complicated pathways resulted in significantly higher overlap agreement than that of the squirrel monkey.

We consider all submissions, in all challenges, to be “modern” algorithms. In most cases, investigators implemented reconstruction or tractography techniques developed only recently, with many specifically created to address one or more known limitations. Most importantly, these algorithms were tuned based on the collective knowledge and experience of the research lab, with the aim to optimize the accuracy of their results. Other implemented algorithms were proposed as early as 2001 (for example, using the tensor with a low-order streamline integration), and while they may be considered rudimentary or basic, because they are still in use today – sometimes as the default algorithm in many open source software packages – they are considered modern. Thus, the observed plateau or limits in anatomical accuracy applies to not only the state of the art approaches, but also to the techniques of the past, on which the bulk of current knowledge of structural connectivity in the human brain is based upon.

The trend in many of the more recently developed algorithms and pipelines is to include some variation of informed post-processing. This includes track grouping or clustering (Garyfallidis et al., 2012), streamline filtering based on the diffusion signal or track densities (Smith et al., 2013), globally fitting streamlines to microstructural models (Daducci et al., 2015), and even manual delineation of regions of interest or streamlines. In all, 67 submissions (33 phantom, 11 squirrel monkey, 23 macaque) used some form of either anatomically-, globally-, or microstructurally-informed post-processing. Although the false positive rate was reduced in many of these (increased specificity, decreased overreach), no statistically significant difference was observed between these and submissions not utilizing post-processing – although there is a diverse range of alternative confounding factors across algorithms, including pre-processing, reconstruction methods, algorithms, constraints and number of streamlines. It would be informative to compare tractograms to ground truth both before and after post-processing to confirm increased accuracy and reduced false positives. In addition to new post-processing, several teams used recently developed reconstruction methods (most a variant of spherical deconvolution or multi-compartment models), software packages (Dipy, MI-BRAIN, Quantitative Imaging Toolkit, dMRITool, MRTrix, FiberNavigator), and streamline algorithms.

The results of the 3D VoTEM challenge confirm and expand upon the limitations and shortcomings demonstrated over the last decade in validation literature. Importantly, the algorithms submitted in this challenge are run and optimized (and often developed) by the contestants themselves, rather than run as off-the shelf algorithms typically implemented in validation literature. Submitted algorithms are compared and benchmarked on the same dataset, using the same evaluation criteria. In the past, both white matter pathways and long-range connections have been assessed using either histological validations (Azadbakht et al., 2015; Calabrese et al., 2015; Dauguet et al., 2006; Donahue et al., 2016; Knosche et al., 2015), simulated datasets (Close et al., 2009), or physical phantoms (Cote et al., 2013). Past studies have demonstrated that DTI tractography has difficulties when tracts cross or divide (Dauguet et al., 2006), highlighting the importance of the crossing fiber problem. However, DTI tractography is strongly correlated with true connectivity on the scale of major cortical regions, but is less reliable at measuring voxel-wise connectivity (Gao et al., 2013). The current challenge confirms this, not only for DTI, but for a range of reconstruction techniques (Both DTI and higher order models) and tracking strategies. Cortical-cortical connection strengths of tractography have been shown to be modestly informative predictions of tracer connections (Donahue et al., 2016) with biases dependent on path lengths and connections strengths. Tractography is also capable of finding the spatial extent of major pathways (Knosche et al., 2015), however, it was found not possible to achieve high specificity and sensitivity at the same time, with only moderate ability to detect true positive (~0.35–0.85 true positive rate) and true negative (~0.05–0.4 true negative rate) connections. General conclusions across all studies were that tractography was informative, but that accuracy would be improved through improvements in acquisition, newer algorithms, high quality data. Towards this end, in 2013, Thomas et al. (2014) acquired an ex vivo macaque dataset with high angular and spatial resolution – estimated to be equivalent to an in vivo acquisition requiring thousands of hours of scan time. Using standard algorithms at the time, they find that despite exceptional data, accurate tractography still remains an elusive goal. In comparison, even with new and improved algorithms in the current macaque sub-challenge, only minor improvements are made (an increase in Youden value of 0.05 for the optimal algorithm) in accuracy compared to those from nearly five years ago – suggesting that the ROC curves have not shifted dramatically in the last few years.

4.2. Advancements needed

While our phantom and ex vivo validations result in similar trends and findings across a range of ground truth geometries, acquisition settings, and image qualities, the ultimate goal is to accurately map the in vivo human brain. Although tractography on a human cannot be directly validated, the accuracy of tractography based on these non-human validation paradigms has largely plateaued in recent years, which likely reflects similar sensitivity/specificity limitations of the process in a human brain. These specific datasets should require a dedicated processing

pipeline, tuned and optimized for it. Most existing tools and software packages were developed and tuned based on the field's understanding of human anatomy. While some exist (or are easily adaptable), tools for small animals, or larger animals e.g. monkeys, or any non-human tractography need to be improved to create better masks, better labels, or better priors so that modern, and future, tractography developments can be leveraged. These tools and resources will not only be applicable to validation studies, but any research on the structural connectivity of the non-human brain.

"Solving" the tractography problems in these phantoms and animal models does not necessarily guarantee perfect reconstructions in the human brain. However, better understanding mistakes relative to the ground truth will certainly spur improvements and innovations in these techniques. Advances will be made through a process well-described by Dyrby et al. (2018) where we must "loop until our method's results agree with the gold standard, and/or until the updated knowledge of ground truth can explain the discrepancies observed." This includes continually updating theory and implementation of methods, validation against gold standards, understanding deviations from the truth, followed by further modifications to theory and implementation, etc. Consequently, there is a need for more advanced and sophisticated gold standards, and a need for validation across a range of spatial scales. In the past (and in the current study), validation is done as an overall assessment in sensitivity and specificity (or overlap and overreach). Future studies should not only explore accuracy at assessing connections and overlap, but also voxel-wise and microstructural features of the datasets. For example, validation strategies could include multiple histological stains or phantoms with varying fiber densities/diameters/volume fractions, in order to evaluate both connectivity and microstructural features simultaneously. The multi-modal or multi-scale strategies could lend insight into individual steps of the tracking process in order to better understand where tractography "first" goes wrong - whether it is assumptions about microstructural features, axonal orientations, or simply tractography decision making.

When validating tractography it is important to clearly define what we hope to map with tractography, and more importantly, how well the ground truth represents this. The goal could be to validate microstructural features of specific pathways (fiber densities, fiber orientations), the course of white matter pathways, the presence or absence of connections between regions, or some measure of connectivity between regions (number of connecting axons, proportion of axons reaching a region, conductivity between regions). While the challenges in this study focused on the course of the pathway (phantom and squirrel monkey) and presence or absence of connections (macaque and squirrel monkey), they are not without their limitations in representing true tissue structures (Dyrby et al., 2018). Several factors limit the accuracy of the gold standard in ex vivo validations, including changes in tissue due to extraction and fixation (D'Arceuil and de Crespigny, 2007), imperfect registrations between histology and MRI, and tracer uptake and visualization. As mentioned above, the macaque MRI and tracer injection was performed on different subjects. While the squirrel monkey experiments were all on the same subject, the acquisition was sub-optimal for ex vivo imaging (Dyrby et al., 2011), and included only a single pathway of interest. The phantom is limited by its simplicity, with a simple geometry on the macroscopic scale. Potential opportunities involve including more adjacent bundles (crossing, kissing, fanning) where partial volume occurs on the scale of individual voxels, as well as features that better mimic the in vivo brain (cortex, varying diffusion compartments, fiber dispersion). Future validation approaches should continually strive for improvements in creation or construction of the ground truth, aim for innovation in validation approaches and strategies, and aim to minimize deviations of the "ground truth" from the true tissue properties by accurately extracting the feature of interest.

This stresses the need for sharing and distribution of validation datasets and ground truths, and tackling the validation problem from a number of perspectives is critical. However, these datasets are time consuming to acquire, expensive, and often require expertise in various niche fields (i.e. histology or phantom creation). While the current challenge was the first to combine separate datasets with very different validation strategies, there are a large number of existing datasets that have lent their own, unique, insight into interpreting tractography (see above for examples). However, it is important to not only validate tractography on different spatial scales (i.e. microscopic versus macroscopic), diverse datasets, and various representations of ground truth, but also necessary to make these open source for valid comparisons of existing and future algorithms and approaches. An online tractography validation tool (much like the "Tractometer" tool for the FiberCup physical phantom (Cote et al., 2013)) combining a large repository of validation datasets would make it easier for neuroscientists, computer scientists, and physicians to submit and test new algorithms, datasets, and methods. Current neuroimaging validation databases do exist (for example, the White Matter Microscopy Database: https://osf.io/yp4qg/), containing largely microstructural validation datasets – but tractography is just modeling microstructure at a macroscopic length scale. Thus, we recommend this, or similar, databases to collect and distribute tractography validation data. This, in combination with more sophisticated algorithms, will almost certainly lead to advances in tractography, and allow us to gain better insights into trends and limitations of these techniques.

While it seems that the results of this study paint a pessimistic view of tractography, there are several positive takeaways. First, some algorithms are indeed able to recover the full spatial extent of pathways, while others have a specificity high enough to make confident predictions about the presence of pathways. Finally, reassuringly, there will almost always be human involvement in this process, especially if tractography is used for surgical planning. A surgeon may not be interested in sparse, stray tracts, or may only care about streamlines in specific locations (i.e. peri-tumoral), and perfect sensitivity/specificity may not be a concern. Alternatively, interaction with the tracking software (and subsequent parameters, ROIs, etc.) allows the surgeon to fine tune based on his or her prior knowledge. This, in combination with the large variability in reconstructions, makes it critical to educate tractography users that the process as it stands is more akin to an art, than an absolute representation of the brains fiber pathways.

In a typical use of tractography, an investigator uses estimated orientation information to ask which brain region is connected to another, as well as the shape, size, route, and strength of this connection. Similarly, in these challenges, the only information given to the investigator is in the form of fiber orientation information (the diffusion signal), and the beginning of the pathway (the seed region). Results from our current study as well as the seminal work of Maier-Hein et al. (2017) clearly shows that having ***only this information, i.e. the local orientation and seed, is not enough!*** Tractography needs more information to overcome the specificity-sensitivity curse of current methods. Potential solutions are appearing such as i) including better and more priors based on known neuroanatomy (Chamberland et al., 2017; Rheault et al., 2018), ii) including microstructural information along local orientations to better trace-out orientations that belong to the same connection from end-to-end (Daducci et al., 2016; Girard et al., 2017; Grinberg et al., 2018), iii) machine learning techniques that could learn from all submissions, from challenges with ground truth, the local and global structure of valid and invalid connections (Neher et al., 2017), and iv) information from other modalities such as myelin markers (Stikov et al., 2015)and functional imaging contrasts (Frank and Galinsky, 2016; Galinsky and Frank, 2017; Schilling et al., 2018c) that could help reduce the number of invalid connection and increase the number of valid connections (Deslauriers-Gauthier et al., 2016, 2017; Schurr et al., 2018).

Better priors from hundreds of years of neuroanatomy research as well as functional imaging could bring novel information about the 'where' and 'how' streamlines should start and end, as well as traverse complex crossing and bottleneck regions. Microstructural information

from dMRI or other modalities could add a vector of features along each fiber orientation to help connect orientations that belong to the same structure, that have the same properties (axon diameter, intra/extracellular space, myelin volume, etc). Moreover, with the terabytes of streamlines generated by state-of-the-art techniques in numerous challenges organized internationally as well as initiatives such as Tractometer (Cote et al., 2013), there is a great potential for having a deep learning algorithm learn the easy-to-track and hard-to-track parts of the brain, both locally and globally, and potentially highlight the untrackable regions and locations of errors.

While no submission was consistently successful in every tracking fidelity metric, the results of our study do not invalidate tractography as a useful biomedical tool, as many were fairly predictive of connectivity, or had moderate to good ability to delineate spatial pathways. Instead, the results of our study emphasize that given current state of the art approaches, pathway reconstruction increasingly appears to be a problem that is unlikely to be wholly solved using only local orientation estimates, and it may be necessary to incorporate other information, other modalities, or new tracking strategies, to successfully resolve tractography's known limitations.

Acknowledgments

This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. We thank Synaptive Medical for providing the anisotropic diffusion phantom, and for expertise on its use in research. This work was supported by the National Institutes of Health under award numbers R01EB017230 (Landman), and T32EB001628. This project was supported in part by ViSE/VICTR VR3029 and the National Center for Research Resources, Grant UL1 RR024975-01. All data, for all three challenges, are freely available at the challenge website (https://my.vanderbilt.edu/votem/). Challenge submissions remain open, and will be processed upon submission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2018.10.029.

References

Alexander, D.C., Barker, G.J., 2005. Optimal imaging parameters for fiber-orientation estimation in diffusion MRI. Neuroimage 27, 357–367.
Alexander, D.C., Barker, G.J., Arridge, S.R., 2002. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magn. Reson. Med. 48, 331–340.
Aydogan, D.B., Jacobs, R., Dulawa, S., Thompson, S.L., Francois, M.C., Toga, A.W., Dong, H., Knowles, J.A., Shi, Y., 2018. When tractography meets tracer injections: a systematic study of trends and variation sources of diffusion-based connectivity. Brain Struct. Funct. 223, 2841–2858.
Azadbakht, H., Parkes, L.M., Haroon, H.A., Augath, M., Logothetis, N.K., de Crespigny, A., D'Arceuil, H.E., Parker, G.J., 2015. Validation of high-resolution tractography against in vivo tracing in the macaque visual cortex. Cerebr. Cortex 25, 4299–4309.
Bookstein, F.L., 1989. Principal warps - Thin-Plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11, 567–585.
Calabrese, E., Badea, A., Cofer, G., Qi, Y., Johnson, G.A., 2015. A diffusion MRI tractography connectome of the mouse brain and comparison with neuronal tracer data. Cerebr. Cortex 25, 4628–4637.
Chamberland, M., Scherrer, B., Prabhu, S.P., Madsen, J., Fortin, D., Whittingstall, K., Descoteaux, M., Warfield, S.K., 2017. Active delineation of Meyer's loop using oriented priors through MAGNEtic tractography (MAGNET). Hum. Brain Mapp. 38, 509–527.
Choe, A.S., Gao, Y., Li, X., Compton, K.B., Stepniewska, I., Anderson, A.W., 2011. Accuracy of image registration between MRI and light microscopy in the ex vivo brain. Magn. Reson. Imaging 29, 683–692.
Close, T.G., Tournier, J.D., Calamante, F., Johnston, L.A., Mareels, I., Connelly, A., 2009. A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. Neuroimage 47, 1288–1300.
Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E., 1999. Tracking neuronal fiber pathways in the living human brain. Proc. Natl. Acad. Sci. U. S. A. 96, 10422–10427.
Cote, M.A., Girard, G., Bore, A., Garyfallidis, E., Houde, J.C., Descoteaux, M., 2013. Tractometer: towards validation of tractography pipelines. Med. Image Anal. 17, 844–857.
D'Arceuil, H., de Crespigny, A., 2007. The effects of brain tissue decomposition on diffusion tensor imaging and tractography. Neuroimage 36, 64–68.
Daducci, A., Canales-Rodriguez, E.J., Descoteaux, M., Garyfallidis, E., Gur, Y., Lin, Y.C., Mani, M., Merlet, S., Paquette, M., Ramirez-Manzanares, A., Reisert, M., Reis Rodrigues, P., Sepehrband, F., Caruyer, E., Choupan, J., Deriche, R., Jacob, M., Menegaz, G., Prckovska, V., Rivera, M., Wiaux, Y., Thiran, J.P., 2014. Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. IEEE Trans. Med. Imag. 33, 384–399.
Daducci, A., Dal Palu, A., Descoteaux, M., Thiran, J.P., 2016. Microstructure informed tractography: pitfalls and open challenges. Front. Neurosci. 10, 247.
Daducci, A., Dal Palu, A., Lemkaddem, A., Thiran, J.P., 2015. COMMIT: convex optimization modeling for microstructure informed tractography. IEEE Trans. Med. Imag. 34, 246–257.
Dauguet, J., Peled, S., Berezovskii, V., Delzescaux, T., Warfield, S.K., Born, R., Westin, C.F., 2006. 3D histological reconstruction of fiber tracts and direct comparison with diffusion tensor MRI tractography. Med Image Comput Comput Assist Interv 9, 109–116.
Deslauriers-Gauthier, S., Lina, J.-M., Butler, R., Bernier, P.-M., Whittingstall, K., Deriche, R., Descoteaux, M., 2017. Inference and Visualization of Information Flow in the Visual Pathway Using DMRI and EEG. Springer International Publishing, Cham, pp. 505–512.
Deslauriers-Gauthier, S., Lina, J.-M., Butler, R., Whittingstall, K., Bernier, P.-M., Descoteaux, M., 2016. Fibre directionality and information flow through the white matter: preliminary results on the fusion of diffusion MRI and EEG. In: Proceedings of International Society of Magnetic Resonance in Medicine. ISMRM, Singapore.
Donahue, C.J., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Behrens, T.E., Dyrby, T.B., Coalson, T., Kennedy, H., Knoblauch, K., Van Essen, D.C., Glasser, M.F., 2016. Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey. J. Neurosci. 36, 6758–6770.
Dyrby, T.B., Baare, W.F., Alexander, D.C., Jelsing, J., Garde, E., Sogaard, L.V., 2011. An ex vivo imaging pipeline for producing high-quality and high-resolution diffusion-weighted imaging datasets. Hum. Brain Mapp. 32, 544–563.
Dyrby, T.B., Innocenti, G., Bech, M., Lundell, H., 2018 Nov 15. Validation strategies for the interpretation of microstructure imaging using diffusion MRI. Neuroimage 182, 62–79. https://doi.org/10.1016/j.neuroimage.2018.06.049 [Epub 2018 Jun 18].
Dyrby, T.B., Sogaard, L.V., Parker, G.J., Alexander, D.C., Lind, N.M., Baare, W.F., Hay-Schmidt, A., Eriksen, N., Pakkenberg, B., Paulson, O.B., Jelsing, J., 2007. Validation of in vitro probabilistic tractography. Neuroimage 37, 1267–1277.
Essayed, W.I., Zhang, F., Unadkat, P., Cosgrove, G.R., Golby, A.J., O'Donnell, L.J., 2017. White matter tractography for neurosurgical planning: a topography-based review of the current state of the art. Neuroimage Clin 15, 659–672.
Frank, L.R., Galinsky, V.L., 2016. Dynamic multiscale modes of resting state brain activity detected by entropy field decomposition. Neural Comput. 28, 1769–1811.
Galinsky, V.L., Frank, L.R., 2017. A unified theory of neuro-MRI data shows scale-free nature of connectivity modes. Neural Comput. 29, 1441–1467.
Gao, Y., Choe, A.S., Stepniewska, I., Li, X., Avison, M.J., Anderson, A.W., 2013. Validation of DTI tractography-based measures of primary motor area connectivity in the squirrel monkey brain. PloS One 8, e75065.
Gao, Y., Khare, S.P., Panda, S., Choe, A.S., Stepniewska, I., Li, X., Ding, Z., Anderson, A., Landman, B.A., 2014. A brain MRI atlas of the common squirrel monkey, Saimiri sciureus. In: Proc SPIE Int Soc Opt Eng, p. 90380C.
Gao, Y., Parvathaneni, P., Schilling, K., Zu, Z., Choe, A., Stepniewska, I., Ding, Z., Landman, B.A., Anderson, A.W., 2016. A 3D high resolution ex vivo white matter atlas of the common squirrel monkey (Saimiri sciureus) based on diffusion tensor imaging. In: Proceedings of the SPIE Medical Imaging Conference. California, San Diego.
Garyfallidis, E., Brett, M., Correia, M.M., Williams, G.B., Nimmo-Smith, I., 2012. QuickBundles, a method for tractography simplification. Front. Neurosci. 6, 175.
Girard, G., Daducci, A., Petit, L., Thiran, J.P., Whittingstall, K., Deriche, R., Wassermann, D., Descoteaux, M., 2017. AxTract: toward microstructure informed tractography. Hum. Brain Mapp. 38, 5485–5500.
Girard, G., Whittingstall, K., Deriche, R., Descoteaux, M., 2014. Towards quantitative connectivity analysis: reducing tractography biases. Neuroimage 98, 266–278.
Grinberg, F., Maximov II, Farrher, E., Shah, N.J., 2018. Microstructure-informed slow diffusion tractography in humans enhances visualisation of fibre pathways. Magn. Reson. Imaging 45, 7–17.
Irfanoglu, M.O., Walker, L., Sarlls, J., Marenco, S., Pierpaoli, C., 2012. Effects of image distortions originating from susceptibility variations and concomitant fields on diffusion MRI tractography results. Neuroimage 61, 275–288.
Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782–790.
Jones, D.K., 2003. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. Magn. Reson. Med. 49, 7–12.
Jones, D.K., 2010. Diffusion MRI: Theory, Methods, and Application. Oxford University Press, Oxford; New York.
Jones, D.K., Cercignani, M., 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed. 23, 803–820.
Knosche, T.R., Anwander, A., Liptrot, M., Dyrby, T.B., 2015. Validation of tractography: comparison with manganese tracing. Hum. Brain Mapp. 36, 4116–4134.
Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 161, 401–407.

Leergaard, T.B., White, N.S., de Crespigny, A., Bolstad, I., D'Arceuil, H., Bjaalie, J.G., Dale, A.M., 2010. Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain. PloS One 5 e8595.

Maier-Hein, K.H., Neher, P.F., Houde, J.C., Cote, M.A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.C., Lin, Y.C., Ji, Q., Reddick, W.E., Glass, J.O., Chen, D.Q., Feng, Y., Gao, C., Wu, Y., Ma, J., Renjie, H., Li, Q., Westin, C.F., Deslauriers-Gauthier, S., Gonzalez, J.O.O., Paquette, M., St-Jean, S., Girard, G., Rheault, F., Sidhu, J., Tax, C.M.W., Guo, F., Mesri, H.Y., David, S., Froeling, M., Heemskerk, A.M., Leemans, A., Bore, A., Pinsard, B., Bedetti, C., Desrosiers, M., Brambati, S., Doyon, J., Sarica, A., Vasta, R., Cerasa, A., Quattrone, A., Yeatman, J., Khan, A.R., Hodges, W., Alexander, S., Romascano, D., Barakovic, M., Auria, A., Esteban, O., Lemkaddem, A., Thiran, J.P., Cetingul, H.E., Odry, B.L., Mailhe, B., Nadar, M.S., Pizzagalli, F., Prasad, G., Villalon-Reina, J.E., Galvis, J., Thompson, P.M., Requejo, F.S., Laguna, P.L., Lacerda, L.M., Barrett, R., Dell'Acqua, F., Catani, M., Petit, L., Caruyer, E., Daducci, A., Dyrby, T.B., Holland-Letz, T., Hilgetag, C.C., Stieltjes, B., Descoteaux, M., 2017. The challenge of mapping the human connectome based on diffusion tractography. Nat. Commun. 8, 1349.

Neher, P.F., Cote, M.A., Houde, J.C., Descoteaux, M., Maier-Hein, K.H., 2017. Fiber tractography using machine learning. Neuroimage 158, 417–429.

Neher, P.F., Laun, F.B., Stieltjes, B., Maier-Hein, K.H., 2014. Fiberfox: facilitating the creation of realistic white matter software phantoms. Magn. Reson. Med. 72, 1460–1470.

Ning, L., Laun, F., Gur, Y., DiBella, E.V., Deslauriers-Gauthier, S., Megherbi, T., Ghosh, A., Zucchelli, M., Menegaz, G., Fick, R., St-Jean, S., Paquette, M., Aranda, R., Descoteaux, M., Deriche, R., O'Donnell, L., Rathi, Y., 2015. Sparse Reconstruction Challenge for diffusion MRI: validation on a physical phantom to determine which acquisition scheme and analysis method to use? Med. Image Anal. 26, 316–331.

Pierpaoli, C., Walker, L., Irfanoglu, O., Barnett, A., Basser, P., Chang, L.-C., Koay, C., Pajevic, S., Rohde, G., Sarlls, J., Wu, M., 2010. TORTOISE: an integrated software package for processing of diffusion MRI data. In: 18th Scientific Meeting of the International Society for Magnetic Resonance in Medicine, p. 1597.

Pujol, S., Wells, W., Pierpaoli, C., Brun, C., Gee, J., Cheng, G., Vemuri, B., Commowick, O., Prima, S., Stamm, A., Goubran, M., Khan, A., Peters, T., Neher, P., Maier-Hein, K.H., Shi, Y., Tristan-Vega, A., Veni, G., Whitaker, R., Styner, M., Westin, C.F., Gouttard, S., Norton, I., Chauvin, L., Mamata, H., Gerig, G., Nabavi, A., Golby, A., Kikinis, R., 2015. The DTI challenge: toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery. J. Neuroimaging 25, 875–882.

Reveley, C., Seth, A.K., Pierpaoli, C., Silva, A.C., Yu, D., Saunders, R.C., Leopold, D.A., Ye, F.Q., 2015. Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. Proc. Natl. Acad. Sci. U. S. A. 112, E2820–E2828.

Rheault, F., St-Onge, E., Sidhu, J., Chenot, Q., Petit, L., Descoteaux, M., 2018. Bundle-specific Tractography. Springer International Publishing, Cham, pp. 129–139.

Rohde, G.K., Aldroubi, A., Dawant, B.M., 2003. The adaptive bases algorithm for intensity-based nonrigid image registration. IEEE Trans. Med. Imag. 22, 1470–1479.

Schilling, K., Gao, Y., Stepniewska, I., Choe, A.S., Landman, B.A., Anderson, A.W., 2017a. Reproducibility and variation of diffusion measures in the squirrel monkey brain, in vivo and ex vivo. Magn. Reson. Imaging 35, 29–38.

Schilling, K.G., Gao, Y., Christian, M., Janve, V., Stepniewska, I., Landman, B.A., Anderson, A.W., 2018a Jul 13. A web-based atlas combining MRI and histology of the squirrel monkey brain. Neuroinformatics. https://doi.org/10.1007/s12021-018-9391-z [Epub ahead of print].

Schilling, K., Gao, Y., Janve, V., Stepniewska, I., Landman, B.A., Anderson, A.W., 2018b Mar. Confirmation of a gyral bias in diffusion MRI fiber tractography. Hum. Brain Mapp. 39 (3), 1449–1466. https://doi.org/10.1002/hbm.23936 [Epub 2017 Dec 19].

Schilling, K.G., Gao, Y., Li, M., Wu, T.-L., Blaber, J., Landman, B.A., Anderson, A.W., Ding, Z., Gore, J.C., 2018c Sep 18. Functional tractography of white matter by high angular resolution functional-correlation imaging (HARFI). Magn. Reson. Med. https://doi.org/10.1002/mrm.27512 [Epub ahead of print].

Schilling, K.G., Gao, Y., Stepniewska, I., Janve, V., Landman, B.A., Anderson, A.W., 2019. Anatomical accuracy of standard-practice tractography algorithms in the motor system - a histological validation in the squirrel monkey brain. Magn. Reson. Imaging 55, 7–25.

Schilling, K.G., Gao, Y., Stepniewska, I., Wu, T.L., Wang, F., Landman, B.A., Gore, J.C., Chen, L.M., Anderson, A.W., 2017b Oct. The VALiDATe29 MRI based multi-channel atlas of the squirrel monkey brain. Neuroinformatics 15 (4), 321–331. https://doi.org/10.1007/s12021-017-9334-0.

Schilling, K.G., Janve, V., Gao, Y., Stepniewska, I., Landman, B.A., Anderson, A.W., 2018d. Histological validation of diffusion MRI fiber orientation distributions and dispersion. Neuroimage 165, 200–221.

Schmahmann, J.D., Pandya, D., 2009. Fiber Pathways of the Brain. OUP, USA.

Schurr, R., Duan, Y., Norcia, A.M., Ogawa, S., Yeatman, J.D., Mezer, A.A., 2018. Tractography optimization using quantitative T1 mapping in the human optic radiation. Neuroimage 181, 645–658.

Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2013. SIFT: spherical-deconvolution informed filtering of tractograms. Neuroimage 67, 298–312.

Stikov, N., Campbell, J.S., Stroh, T., Lavelee, M., Frey, S., Novek, J., Nuara, S., Ho, M.K., Bedell, B.J., Dougherty, R.F., Leppert, I.R., Boudreau, M., Narayanan, S., Duval, T., Cohen-Adad, J., Picard, P.A., Gasecka, A., Cote, D., Pike, G.B., 2015. In vivo histology of the myelin g-ratio with magnetic resonance imaging. Neuroimage 118, 397–405.

Thomas, C., Ye, F.Q., Irfanoglu, M.O., Modi, P., Saleem, K.S., Leopold, D.A., Pierpaoli, C., 2014. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. Proc. Natl. Acad. Sci. U. S. A. 111, 16574–16579.

Tournier, J.D., 2010. The biophysics of crossing fibers. In: Jones, D.K. (Ed.), Diffusion MRI: Theory, Methods, and Application. Oxford University Press, Oxford; New York, pp. 465–481.

Wheeler-Kingshott, C.A., Cercignani, M., 2009. About "axial" and "radial" diffusivities. Magn. Reson. Med. 61, 1255–1260.