Histological Validation of DW-MRI Tractography in Human Postmortem Tissue

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Histological Validation of DW-MRI Tractography in Human Postmortem Tissue

Arne K. Seehaus[1,2], Alard Roebroeck[2], Oriana Chiry[1], Dae-Shik Kim[3], Itamar Ronen[4], Hansjürgen Bratzke[5], Rainer Goebel[2,6] and Ralf A. W. Galuske[1,7]

[1]Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany, [2]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, the Netherlands, [3]Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea, [4]Department of Radiology, C. J. Gorter Center for High Field MRI, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands, [5]Institute of Forensic Medicine, Johann Wolfgang Goethe-University of Frankfurt Medical School, 60596 Frankfurt am Main, Germany, [6]Department of Neuroimaging and Neuromodeling, Netherlands Institute for Neuroscience, an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), 1105 BA Amsterdam, the Netherlands and [7]Department of Neurophysiology, Max-Planck-Institute for Brain Research, 60528 Frankfurt am Main, Germany

Arne Seehaus and Alard Roebroeck have contributed equally to this work

Address correspondence to Arne Seehaus, Department of Biology, Technische Universität Darmstadt, Schnittspahnstrasse 3, 64287 Darmstadt, Germany. Email: seehaus@bio.tu-darmstadt.de.

Despite several previous attempts, histological validation of diffusion-weighted magnetic resonance imaging (DW-MRI)-based tractography as true axonal fiber pathways remains difficult. In the present study, we establish a method to compare histological and tractography data precisely enough for statements on the level of single tractography pathways. To this end, we used carbocyanine dyes to trace connections in human postmortem tissue and aligned them to high-resolution DW-MRI of the same tissue processed within the diffusion tensor imaging (DTI) formalism. We provide robust definitions of sensitivity (true positives) and specificity (true negatives) for DTI tractography and characterize tractography paths in terms of receiver operating characteristics. With sensitivity and specificity rates of approximately 80%, we could show a clear correspondence between histological and inferred tracts. Furthermore, we investigated the effect of fractional anisotropy (FA) thresholds for the tractography and identified FA values between 0.02 and 0.08 as optimal in our study. Last, we validated the course of entire tractography curves to move beyond correctness determination based on pairs of single points on a tract. Thus, histological techniques, in conjunction with alignment and processing tools, may serve as an important validation method of DW-MRI on the level of inferred tractography projections between brain areas.

Keywords: diffusion tensor imaging, fluorescent dyes, human brain, receiver operating characteristics, single pathways

Introduction

Understanding the function of the brain is not possible without precise knowledge about the connectivity between its different parts. In the last century, several methods have been developed to study the fiber connections in tissue probes, starting with simple dissection approaches and ending with highly sophisticated staining techniques. However, all of these methods require animal experiments or postmortem tissue. As a consequence, they are not suitable for questions involving structural brain connectivity in the living human. In the last 2 decades, magnetic resonance imaging (MRI) has provided important new tools to address this problem. In particular, diffusion-weighted imaging techniques, such as diffusion tensor imaging (DTI), are capable of probing neuronal fibers in vivo. Since the introduction of DTI (Basser et al. 1994), it has been applied to a wide range of topics, especially in clinical studies. For example, the role of white matter characteristics in schizophrenia (Kanaan et al. 2005; Kubicki et al. 2005), major depression (Maller et al. 2010; Tham et al. 2010), Alzheimer's disease (Medina and Gaviria 2008; Hess 2009), and epilepsy (Richardson 2010) has been investigated using DTI. Beyond the mapping of scalar tensor measures, as applied in the above-mentioned studies, DTI can also support tractography of large fiber bundles (Conturo et al. 1999; Mori et al. 1999; Basser et al. 2000).

However, previous studies have shown that DTI has several limitations for tractography. First, it is known to perform poorly in brain areas of high fiber curvature or with fibers crossing (X) or kissing (> <). In addition, DTI cannot resolve whether fibers are converging or diverging at Y-shaped structures because directional information cannot be provided (Roebroeck et al. 2008). Second, DTI studies are principally limited to the investigation of white matter, as the fractional anisotropy (FA), the basic information on which DTI tracking algorithms perform, is too low in gray matter. In other words, DTI needs clear diffusion profiles as given by white matter fiber bundles to perform well. Likewise, it is difficult to recognize very small fiber bundles with DTI (Conturo et al. 1999). This, however, is based on the spatial resolution of the respective scans and might be improved using more powerful MRI hardware.

Given these confounds and the uncertainty about the structural origin of the diffusion signals (Beaulieu 2002), validation methods are necessary which provide quantitative measures of how DTI performs under certain circumstances. Previous studies have addressed this issue by different methodological combinations, among them studies involving DTI phantoms (Pullens et al. 2010), MRI contrast agents (Lin et al. 2001), or anatomy atlases (Catani et al. 2002). The most commonly used approach was to compare DTI data with actual histological data, which visualized white matter architecture or neuronal connectivity by selective or unselective staining techniques (Kaufman et al. 2005; Dauguet et al. 2007; Leergaard et al. 2010). One of the processing steps in these studies was to reconstruct a 3D volume from the histological slices that is aligned to the volume obtained through the DTI measurement. This task is affected by several considerations. First, during histological processing, the tissue almost inevitably gets distorted, at least by a global scaling factor in one or both of its

planar directions. Second, creating a histological slice usually reduces all information from the tissue to 2D because the depth information within individual slices is lost. In consequence, it is difficult to identify the depth orientation of stained fiber material, which is useful for a 3D reconstruction of whole fibers. Third, histological and DTI data are substantially different: With an inherently 2D high-resolution (single microns) signal on the one hand and a 3D low-resolution (hundreds of microns) signal on the other hand, it is difficult to obtain a quantitative valid comparison of these 2 data sources. So far, it has mostly been done by downsampling the histological data to the resolution of the MRI and applying a comparison sectionwise (Kaufman et al. 2005; Dauguet et al. 2007). However, in such approaches, a significant amount of the information contained in the histology is lost. Most importantly, this approach precludes validating tractography algorithms using actual axonal tracts labeled from their origin or termination in the gray matter.

Therefore, the aim of our study was to implement a new approach of validating DTI-based streamline tractography on the level of single tractography pathways. To this end, we used fluorescent carbocyanine dyes to trace connections in human postmortem tissue and compared these to DTI tractography results on the same tissue. DiI and DiA are fluorescent lipophilic tracers that are able to selectively stain fibers projecting from or to a crystal implantation location. Placing the dye in the gray matter, this property enables us to obtain information about the origin of white matter fiber paths, which is very difficult to obtain otherwise. We use this property to define true positives (sensitivity), true negatives (specificity), and resulting receiver operating characteristic (ROC) curves for DTI tractography in the framework of signal detection theory (SDT). After investigation of tractography performance in terms of pairs of single points on a tract, we analyze the correspondence of entire tract curves with the carbocyanine dye staining. This approach allows for the evaluation of the congruence of DTI tractography and actual structural connectivity not only for the white matter fiber "highways" but also for the fiber terminations that are far below the resolution of current MR hardware.

Materials and Methods

Tissue
The tissue probes were obtained from the brain of a single subject (male, 33 years old, no history of neurological disorders or brain injury). After a postmortem delay of 12 h, the upper temporal lobe was dissected and immersion fixated for 4 days in a solution containing 2.6% paraformaldehyde, 0.8% iodoacetic acid, 0.8% sodium periodate, and 0.1 M D-L-lysine in 0.1 M phosphate buffer at pH 7.4.

Carbocyanine Tracing
DiI and DiA are fluorescent lipophilic tracers that progress along lipid membranes by diffusion and have for long been tested as anterograde and retrograde neuronal connection tracers (Honig and Hume 1986; Godement et al. 1987; Honig and Hume 1989). More recently, these dyes were confirmed to work also in adult human postmortem tissue (Burkhalter et al. 1993; Galuske et al. 2000). Four carbocyanine crystals (2 DiI and 2 DiA) were implanted into the upper surface of the block at a depth of 500–1000 μm. The distance between each crystal of the same dye was about 3 cm. After 48 months, the tissue was sliced at a thickness of 70 μm and stained fiber material of every fifth slice (i.e., every 350 μm) was graphically documented using fluorescent microscopy at ×37 enlargement factor and excitation wavelengths of 546 nm (DiI) and 440–490 nm (DiA), respectively. One of the

implantation sites had to be excluded from further analysis because its projections were mainly outside the scope of the MR scans of the respective tissue.

Diffusion-Weighted Magnetic Resonance Imaging
Seventeen months after fixation, a diffusion-weighted MRI (DW-MRI) scan was performed. Scanning procedures were similar to those described in Roebroeck et al. (2008): Scanning was done using a 9.4-T/31-cm MR scanner (Magnex Scientific, UK) equipped with 0.003 T/cm gradients (11 cm ID, 300 μs rise time; Magnex Scientific) and driven by a UnityINOVA console (Varian, Walnut Creek, CA) at the Center for Magnetic Resonance Research (CMRR), University of Minnesota Medical Center. The RF coil was a homebuilt quadrature surface coil, composed of two 3.3 cm diameter partially overlapping coils. DW-MRIs were acquired at room temperature using a multi-shot pseudo-3D double spin-echo echo-planar imaging (SE-EPI) sequence with 2 phase encoding directions and 4 segments per pseudoslice. The following parameters were used: time repetition/time echo = 2000 ms/35 ms, $|g| = 0.0017$ T/cm, $\delta = 6$ ms, $\Delta = 15$ ms ($b = 1584$ s/mm^2 for each direction). Six gradient combinations according to the direction scheme $(X, Y, 0)$, $(X, 0, Z)$, $(0, Y, Z)$, $(X, -Y, 0)$, $(0, Y, -Z)$, $(-X, 0, Z)$ and one image with minimal gradient strength on 3 directions was acquired as the unweighted image. The field of view was $5 \times 5 \times 6.4$ cm^3; the matrix size was $128 \times 128 \times 64$ (giving a nominal resolution of 391×391 μm in-plane with a slice thickness of 1 mm). Shimming for field inhomogeneities was done using FASTMAP to include high-order shims in an analytical manner. A set of images showing raw T_2-weighted (i.e., without diffusion weighting, b_0) apparent diffusion coefficient (ADC) and FA contrast is shown in Figure 1.

DTI Tractography
DW-MRI data were further processed and visualized in custom written software in C++ and using OpenGL—for more details, see Roebroeck et al. (2008). Diffusion tensors were calculated for the diffusion profile of each voxel as a 3D Gaussian using a standard linear regression approach. Subsequently, tractography was performed by the Euler stepping solution for streamlines locally corresponding to the resulting tensor field (e.g., Basser et al. 2000). That is, we used a line propagation algorithm that starts from a specified seed point and subsequently takes steps of stepsize α in a direction determined by the local tensor and the previous tracking direction, terminating when the local FA falls below the FA threshold or the curvature (the angle between actual and previous direction) exceeds the angle threshold. The tractography parameters used were a stepsize of 100 μm, an angle threshold of 90°, and 45 seed points per source voxel. We used various FA thresholds of 0.01, 0.02, 0.04, 0.06, 0.08, 0.10, 0.15, 0.20, and 0.25, respectively. A low FA threshold lets the algorithm draw more and longer tracts. Thus, in our approach, FA thresholds serve as the detection criterion for ROC analysis: a low threshold makes tractography more sensitive; a high threshold makes it more specific.

Alignment
In order to obtain a common reference frame for both the histology and the tractography data, first, we fused the manual drawings of traces of different dyes with a set of high-resolution histology photographs (Fig. 2A). The resulting histological slices were then aligned to the DW-MRI volume with a separate 8-parameter affine transformation (3 translations, 3 rotations, and 2 scales) for each slice. The transformations were conducted manually by matching the gray–white matter borders as seen in the histology and a FA map (Fig. 2B) of the DTI volume. As indicated in Figure 2A,B, the gray–white matter boundary is a robust and useful alignment criterion in both modalities, histology and MR, that is unaffected by noise and nonmatching contrast variations within tissue compartments as well as outside the tissue. Visual judgment of alignment was aided by an interactive fusion of both modalities, such that each transformation of a histological slice led to an immediate update of the FA map. In this way, each slice could be successively transformed toward an
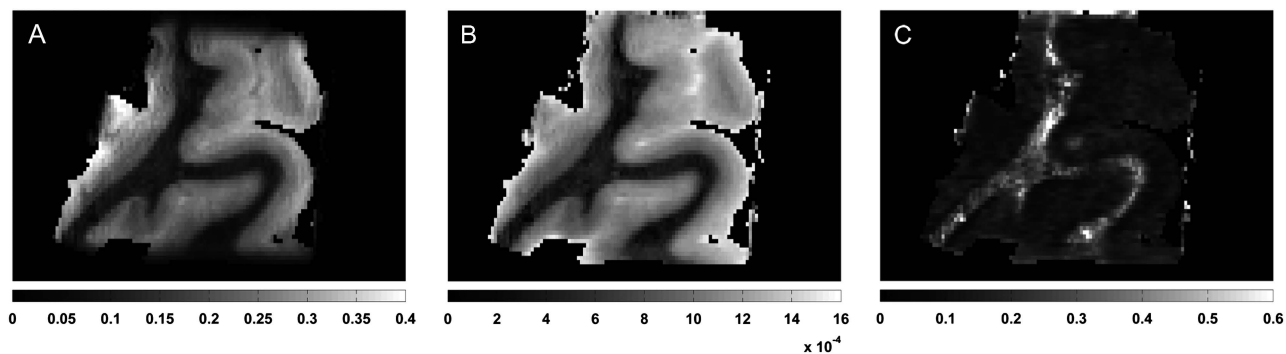
Figure 1. Typical images and calculated maps shown for one slice of the specimen used in the experiments. (*A*) T_2-weighted image ($b = 0$). The scale is given in arbitrary units. (*B*) ADC map. The units are square millimeters per second. The ADC value for white matter was between 3 and 5×10^{-4} mm^2/s. The b value used in the experiment ($b = 1936$ mm^2/s) is close to optimal. (*C*) FA map. Typical values for white matter range between 0.25 and 0.5. Please note the sharp ADC differences between gray and white matter, typically not observed in in vivo DTI measurements. Similar observations can be seen in other postmortem DW-MRI studies (e.g., D'Arceuil et al. 2008).
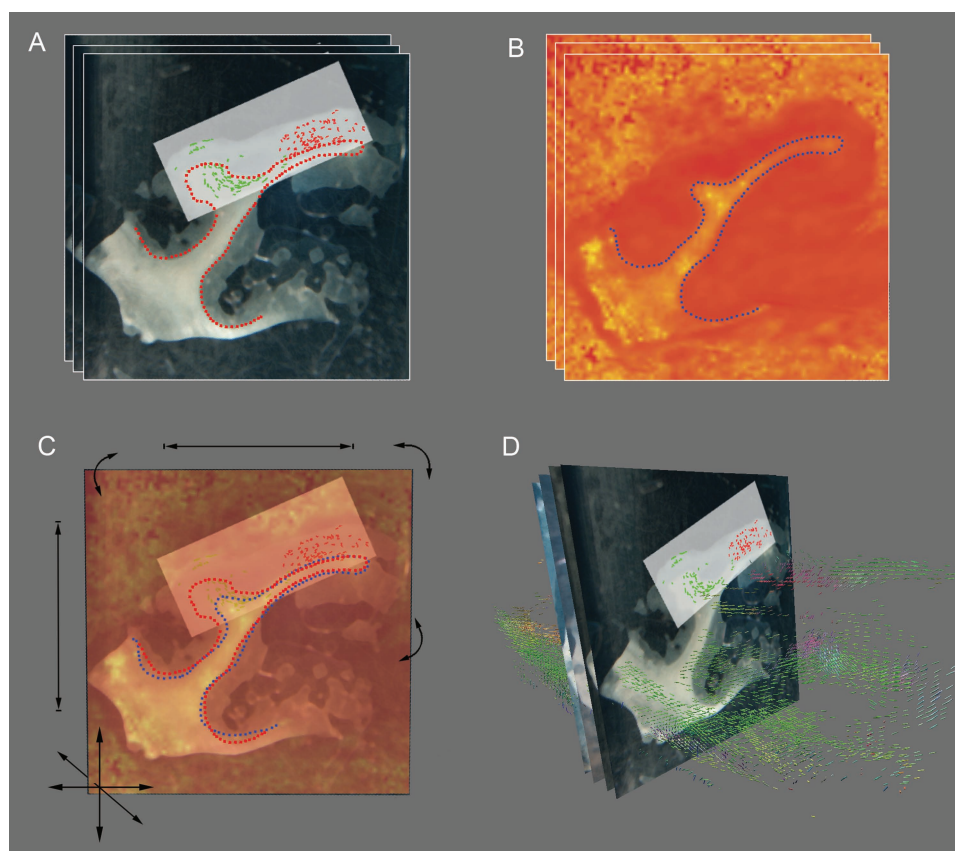

Figure 2. An illustration of the procedure and criteria for alignment of histology and MR diffusion data. (*A*) Example of the high-resolution 2D photographs of our set of coplanar histological slices. The manual drawings of traces of different dyes (in red and green) were fused with the photographs. The red dotted outline illustrates the gray–white matter boundary, which was used as a criterion for alignment. (*B*) An arbitrary 2D plane through the 3D FA map derived from the diffusion MR data, close to the optimal alignment with the histological slice in *A*, with its gray–white matter boundary illustrated with the blue dotted line. (*C*) A fusion of the histology and 2D plane through the FA map which was used for alignment. Interactive and online manual changes to 3 translations, 3 rotation, and 2 scale parameters (illustrated by arrows) lead to an immediate update of the FA map on the plane. These parameters were adjusted slice-by-slice to visually match the gray–white matter boundary between the 2 modalities and align histology to diffusion MR in 3D. (*D*) Illustration of the alignment result for 4 of our histological slices rendered in the 3D space of the DTI data. Voxels with high FA values are indicated as colored cylinders, with the color coding for the respective directionality, in order to illustrate 3D fiber direction information.

optimal fitting of the gray–white matter boundaries (Fig. 2*C*). Individual single slice alignment was constrained by the necessary parallelism of the histological slices, which was also used to provide an initial interpolated position for all middle slices after alignment of a few of the outer slices.

The alignment procedure resulted in a full 3D spatial coregistration of histological slices and diffusion tensors such that pixel locations in histological slices and tract coordinates in the DTI volumes could be directly compared for proximity (Fig. 2*D*).

Sensitivity and Specificity Analysis

We evaluated the DTI tractography by answering the following 2 questions:

1. To what extent are real axons stained with DiI recognizable as DTI tractography pathways? (sensitivity)
2. To what extent can DTI tractography paths be trusted not to appear when there are no corresponding axons? (specificity)

To accomplish this, we conducted the following steps:

1. MR voxels within a radius of approximately 0.3–1.5 cm around the injection site of the dyes and with FA values of at least 0.1 were classified as either containing ("true") or lacking ("false") dye staining. To correct for possible inaccuracies of the alignment and to avoid ambiguous classifications, we defined stricter voxel classifications for the sensitivity and specificity analysis (Fig. 3A). We defined "strictly true" ("strictly false") voxels as true (false) voxels for which all 26 nearest neighbors are also classified as true (false). This procedure lead to a total sample of $n = 13381$ source voxels, from which tractography was performed.
2. Originating from each of these voxels, we performed fiber tractography with the afore-mentioned algorithm. Subsequently, the voxels were classified as positive if any one of their fibers hit the corresponding implantation site, or negative, if not (Fig. 3B).
3. Combining step 1 and step 2, voxels could now be divided into hits (true positive), misses (true negative), correct rejections (false negative), and false alarms (false positive).

An illustration of the entire classification approach is given in Figure 3. A potential confound in this approach are unequal numbers of true and false voxels or voxels, which are near to or far from the implantation site (as voxels near to the injection site should be more likely to hit it, a priori).

Consequently, voxel counts were balanced for these confounds using the following procedure for each of the implantations:

1. Cut the sets T and F of true and false voxels such that both have the same range of distance to the injection site. This is determined by the smaller distance range found in either of the sets.
2. Divide the smaller set (let's assume: T) into 3 equally sized distance ranges, T_{near}, T_{medium}, and T_{far}.
3. Divide the bigger set F into subsets F_{near}, F_{medium}, and F_{far} according to the same distance ranges.
4. For analyses within a certain distance range, reduce the bigger true/false subset of the respective distance range at random to the size of the smaller subset. For overall analyses, reduce all subsets to match the size of the smallest subset.

With this procedure, we obtained the sample sizes illustrated in Table 1. In these samples, it was ensured that the subsets of true and false voxels were of equal size and had an equal distribution of distances to the injection site.

Based on these data and with the FA threshold of the tracking serving as a detection threshold parameter, we calculated the following: sensitivity (#hits/#true), specificity (#correct rejections/#false), and

ROC curves (plotting sensitivity over 1 specificity for each of the 9 different FA values). As a measurement for the overall goodness of a sensitivity/specificity tradeoff, we used the Euclidean distance of its ROC curve point to the point of perfect discrimination (0/1). In the theoretically optimal case, where perfect discrimination is actually achieved, D is 0. The apparently worst case is $D = $ sqrt(2) for 0 sensitivity and specificity. This value would, however, express just the same perfect discrimination, with only positive and negative responses systematically switched. The actual worst case lies on the main diagonal of the respective ROC curve, the so-called "line of no discrimination," and results in a D between 1 and sqrt(2)/2.

Entire Tract Analysis

As noted above, sensitivity and specificity only depend on the connections of pairs of points on a tractography path, whereas the actual course of the entire tractography pathway is not considered. More precisely, perfect sensitivity would be concluded if every tractography path originating from a true voxel is somehow connected to the implantation site, even if by a highly unlikely path. In contrast, DTI tractography paths that are a perfect image of actual fibers should be accompanied by stained fiber material on their whole way from source voxel to implantation site. Consequently, we expected that the majority of the voxels passed by true pathways are actually true voxels (whereas the opposite should hold for false pathways). This hypothesis was statistically tested with a sample of 30 true and 30 false voxels (10 voxels for each distance range), which was randomly chosen within one implantation site. Originating from each of these voxels, one single fiber was tracked with the same set of FA thresholds as in the main analysis. Subsequently, the numbers of true voxels passed by these pathways were calculated and compared using unpaired t-tests.

Results

Sensitivity/Specificity

We report detailed results from 1 of the 3 implantation sites (the "red" site). Results from the 2 other sites are reported and discussed generally. The best sensitivity and specificity (sens./spec.) pairs for the fixated postmortem tissue scanned

Table 1
Numbers of voxels analyzed

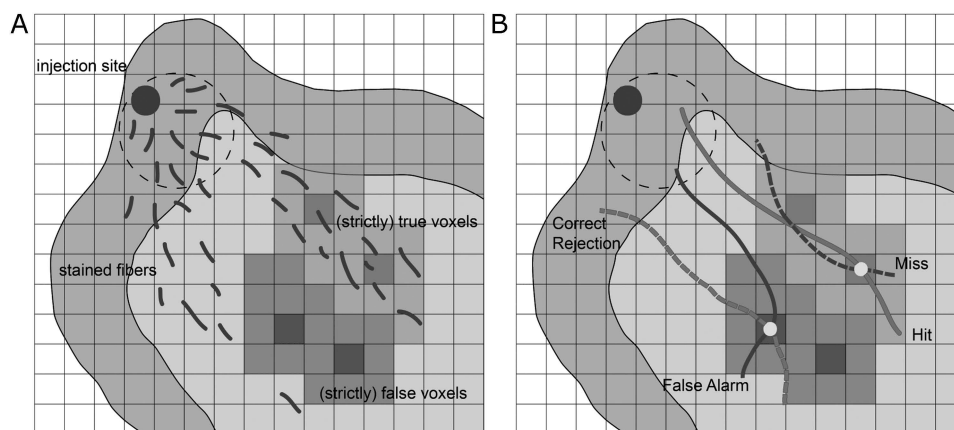| Implant	Overall	Near		Medium		Far	
1 (Blue)	1446	(3–8.3 mm)	1064	(8.3–10.8 mm)	1064	(10.8–12.9 mm)	482
2 (Green)	720	(3–6.4 mm)	240	(6.4–8.2 mm)	978	(8.2–11.9 mm)	978
3 (Red)	384	(3–4.5 mm)	128	(4.5–5.5 mm)	230	(5.5–9 mm)	308

Figure 3. (A) Definition of true and false voxels based on DiI staining. Strictly true (strictly false) voxels are those for which none of their 26 nearest neighbors is a false (true) voxel. (B) Definition of hit, miss, false alarm, and correct rejection.

at room temperature were found for FA values between 0.02 and 0.08. With this setting, sensitivity ranged from 70% to 78%, specificity from 79% to 87%. As described above, the conjoint optimality of these values was measured as the Euclidean distance D from the sensitivity/specificity tradeoff achieved to the optimal value of 100% for both sensitivity and specificity. With respect to D, the best single value observed was 0.78/0.79 for an FA threshold of 0.02. Altogether, there is a clear discrimination for FA thresholds between 0.02 and 0.08. There was a significant decrease for FA values of 0.1 or higher. For extremely low FA values, the loss of specificity adumbrates the increase of sensitivity, consequently resulting in a degrading discrimination as well (Table 2).

Table 3 shows the sensitivity/specificity values and optimality distance D obtained for the different distance ranges. The best decision quality was achieved for near voxels, 0.82/0.89 being the best single result obtained at an FA threshold of 0.08. Similarly to the overall result, there was an optimal range of FA thresholds between 0.02 and 0.08. Within this range, the goodness of the decision was relatively stable. Likewise, the results drop off rapidly for higher FA thresholds.

ROC Curves

Figure 4 shows the ROC curves as obtained from the sensitivity/specificity values for each FA threshold. The curves start at high FA values with 100% specificity and sensitivity slightly above 0 and rise with decreasing FA to a maximum sensitivity of 70–90% combined with specificity values between 60% and 80%. This indicates the crucial role of FA values for the manner in which the line propagation algorithm works: Low FA values lead to many long tractography paths, thus resulting in many hits (high sensitivity) but also many false alarms (low specificity). However, high values produce a more restrictive tracking with less hits (low sensitivity) but more correct rejections (high specificity). The overall goodness of an ROC curve is often assessed by the area under

the curve (Fawcett 2006), which ranges between 0 and 1. Similarly to D, a value of 1 means perfect discrimination and 0.5 is the theoretical worst value. To determine this value, our ROC data had to be extrapolated toward the extreme point of no specificity, which has never been reached in our experiments. Using a hypothetical extreme point of 0 specificity and 90% sensitivity, we got an area of 0.80 under the overall ROC curve, indicating a strong discrimination.

As for the different distance ranges (Fig. 4B), best general discrimination was achieved for voxels close to the implantation site ("near") with an area of 0.86 under the respective ROC curve. The more distant "medium" and "far" curves enclosed slightly smaller areas of 0.75 and 0.76, respectively. Thus, the principal differences in discrimination behavior for different distances are in agreement with the expectation that "tractography over short distances is easier."

The remaining 2 implantation locations ("blue" and "green") showed similar specificity values of generally more than 66%, yet at a specificity that was even for the lowest FA values not much higher than 60%. This result was probably due to the shape of the tracing, as will be discussed below.

Entire Tract Analysis

Fiber tracts originating from true voxels pass significantly more true voxels than pathways tracked from false voxels. This result could be obtained for all FA values used, but it was most pronounced for the lower half of the FA spectrum. Table 4 shows the average numbers of true voxels passed by true and false pathways for each FA value. The most significant result was obtained for a FA threshold of 0.08. For this value, Figure 5 shows the distribution of true and false voxels passed for both true and false fibers as defined above. It is obvious that all true pathways contain many true voxels on their way (Fig. 6A), while most of the false pathways contain very few or no true voxels, with a few outliers (Fig. 6B). FA thresholds of 0.04 and lower yield the same result in this analysis, as the respective pathways have reached their maximal length within the area of the staining. In addition, these pathways tend to reach far beyond this area and are mostly too long and too curved as to represent real fiber bundles (Fig. 6C).

Discussion

DTI Tractography

We report sensitivity/specificity values of up to 0.78/0.79 over all tractography and an area under the ROC curve of 0.80. This is a convincing accuracy, especially as compared with previous

Table 2

Overall result for one injection

FA threshold	Sensitivity	Specificity	D
0.01	0.8229	0.7083	0.3412
0.02	0.7813	0.7917	0.3021
0.04	0.7656	0.7969	0.3101
0.06	0.7292	0.8333	0.318
0.08	0.7031	0.8698	0.3242
0.1	0.6042	0.9219	0.4035
0.15	0.3281	0.9792	0.6722
0.2	0.1875	0.9896	0.8126
0.25	0.1094	1	0.8906

Table 3

Sensitivity/specificity results within 3 distance ranges

| FA threshold	Near (3–4.5 mm)			Medium (4.5–5.5 mm)			Far (5.5–9 mm)		
	Sensitivity	Specificity	D	Sensitivity	Specificity	D	Sensitivity	Specificity	D
0.01	0.875	0.625	0.3953	0.713	0.8	0.3498	0.8247	0.6688	0.3747
0.02	0.8594	0.8125	0.2344	0.6609	0.8087	0.3894	0.7727	0.7597	0.3307
0.04	0.8438	0.8281	0.2323	0.6435	0.8087	0.4046	0.7662	0.7597	0.3352
0.06	0.8438	0.8438	0.221	0.6261	0.8696	0.396	0.7143	0.7987	0.3495
0.08	0.8281	0.8906	0.2037	0.5913	0.913	0.4178	0.6558	0.8247	0.3862
0.1	0.75	0.9531	0.2544	0.5043	0.9217	0.5018	0.487	0.8896	0.5247
0.15	0.3281	1	0.6719	0.3565	0.9739	0.644	0.2662	0.9545	0.7352
0.2	0.1875	1	0.8125	0.2087	0.9826	0.7915	0.1234	0.9805	0.8768
0.25	0.1406	1	0.8594	0.1217	1	0.8783	0.0065	1	0.9935
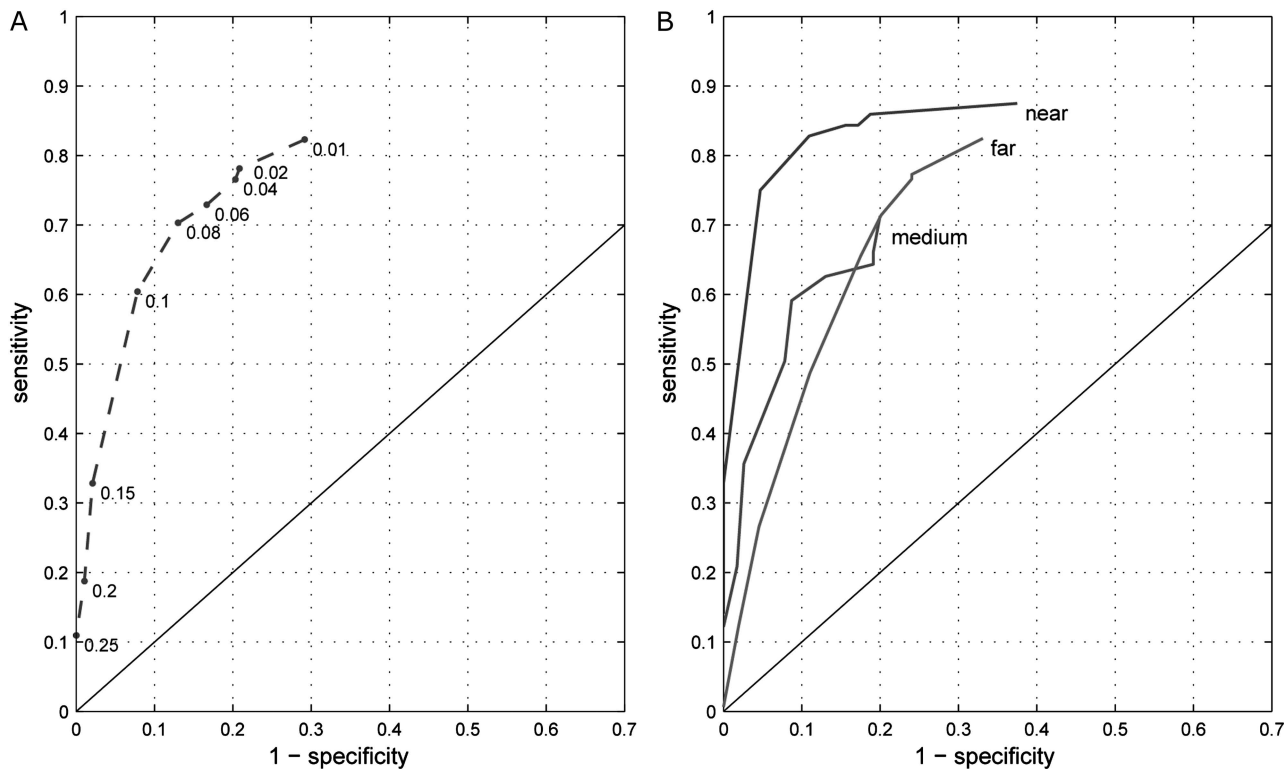
Figure 4. ROC curves obtained from a single dye implantation. (*A*) Overall result (*n* = 384) balanced for true and false voxels over distance. The numbers next to the curve give the respective FA thresholds. (*B*) Distinct curves for the different voxel groups: near (3–4.5 mm, *n* = 128), medium (4.5–5.5 mm, *n* = 230), and far (5.5–9 mm, *n* = 308) distance.

Table 4
Average number of true and false voxels passed by true and false voxels (standard deviation in brackets) and the significance for an unpaired *t*-test for the true–false difference

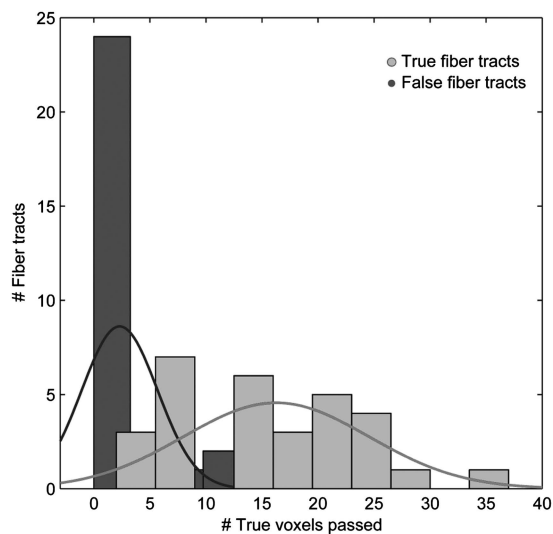| FA threshold	0.01	0.02	0.04	0.06	0.08	0.1	0.15	0.2	0.25
True voxel	17.533 (8.801)	17.533 (8.801)	17.533 (8.801)	17.533 (8.801)	16.267 (8.124)	14.2 (7.373)	9.933 (7.572)	3.467 (4.493)	1.333 (2.371)
False voxel	3.367 (4.223)	3.367 (4.223)	3.367 (4.223)	3.333 (4.237)	2.3 (3.328)	1.9 (3.187)	0.6 (1.2)	0.067 (0.359)	0 (0)
P	$<10^{-7}$	$<10^{-7}$	$<10^{-7}$	$<10^{-7}$	$<10^{-8}$	$<10^{-7}$	$<10^{-6}$	<0.001	<0.01

Figure 5. Histogram of numbers of true voxels passed by true tractography pathways (light gray) and false tractography pathways (dark gray). Many more true voxels are passed by true pathways than by false pathways.

ROC analyses of DTI measurements (Iturria-Medina et al. 2011).

This result has been obtained in full only for 1 of the 3 injections, in the other 2, we detected a high specificity but a low sensitivity. This observation is most likely due to the different pathways of fiber tracts arising from the different injections (Fig. 7). While in the case which offered both, high sensitivity and specificity, a discrete bundle of fibers was stained in the white matter, the other 2 injections provided longer (up to 13 instead of 9 mm) and more diffuse projections, which were directed in several directions.

Thus, the low sensitivity in these cases may reflect the weakness of DTI to track pathways through areas with many crossing pathways. Moreover, this shows a methodological limitation in the validation approach because there may be many voxels that count as true voxels due to just a few single axonal structures, which in fact do not represent DTI tractography pathways. Therefore, these true voxels lead to a high number of misses and, thus, strongly interfere with the sensitivity. Consequently, this validation method currently works most accurately and unambiguously when areas are chosen where staining provides unambiguous directions of the
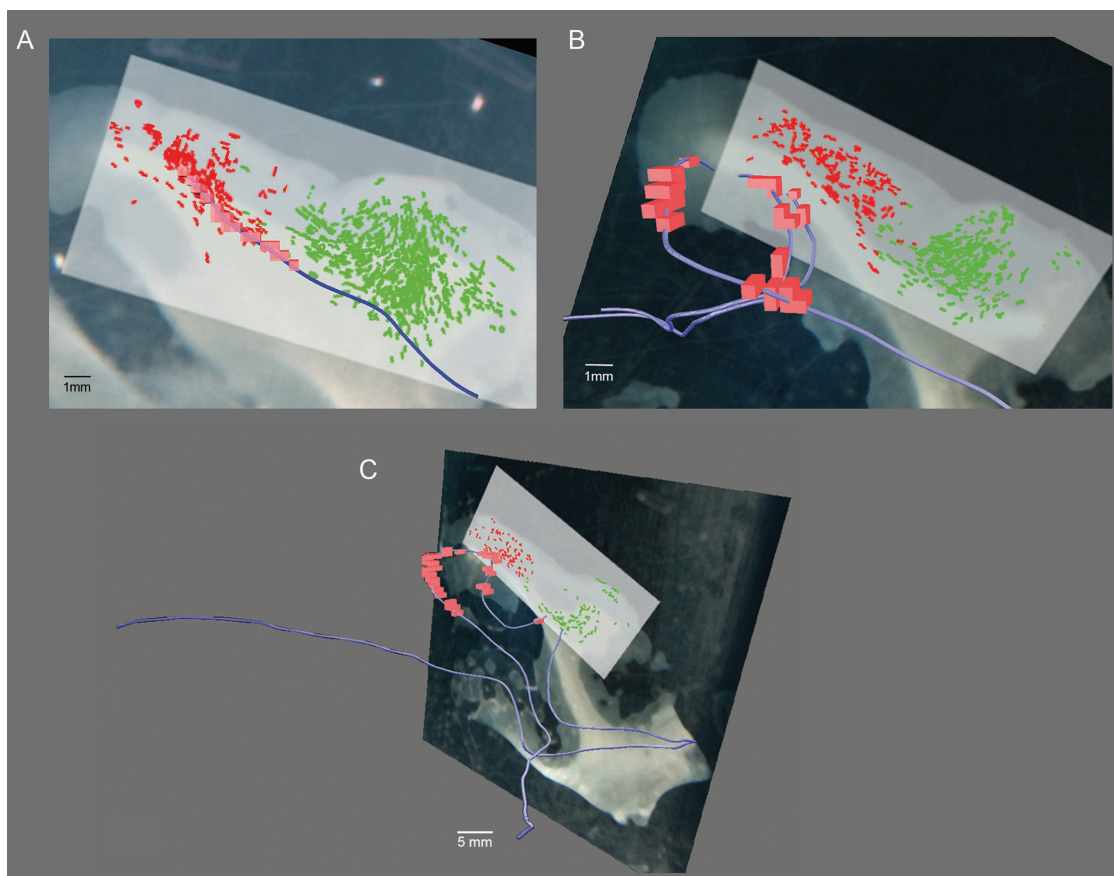
Figure 6. (*A*) True positive DW-MRI tractography pathway passing through many true voxels (transparent red). Red and green: different fluorescent crystal tracings. (*B*) False fibers that nevertheless contain many true voxels. (*C*) True fiber coursing through many true voxels but having unrealistically long and tangled shape.
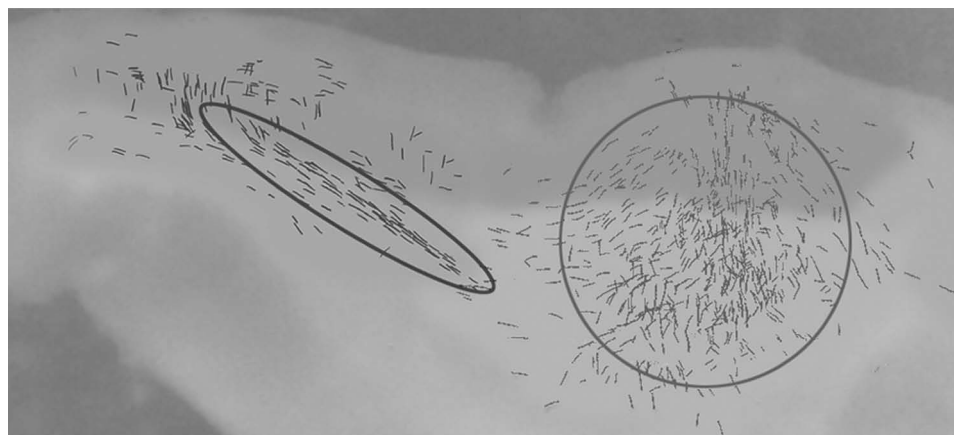

Figure 7. Projections of 2 injections, the left one with a discrete pathway and the right one with several staining directions.

projections. Note that this is also exactly the case in which DTI-based tractography is expected to work a priori: single fiber bundles without complex fiber architecture, such as strong intravoxel curvature or fiber crossing. Thus, the current validation approach is able to robustly quantify the sensitivity and specificity of DTI-based tractography on an entire pathway level away from problematic complex fiber architecture regions. One might even reverse the argument and infer complex fiber architectures from low sens./spec. patterns. Therefore, a detailed analysis of error sources in this validation

approach is needed in order to obtain a better view of the effect of fiber architecture on sensitivity and specificity.

Another very interesting perspective for future research would be to extend this validation approach to validate tractography based on intravoxel diffusion models that can represent complex white matter architecture, such as diffusion spectrum imaging (Wedeen et al. 2008) and constrained spherical deconvolution (Tournier et al. 2007). In its most advanced form, this would require delineation from histology of individual or very small bundles of axons along larger parts of

their path through white matter. It is an important open question whether the type of histological processing that will best support this is a form of carbocyanine dye diffusion, as used here, a more classical axon stain photographed at high resolution or a technique that is sensitive to the local direction of axons in white matter at very high resolution, such as polarized light imaging (Axer et al. 2011). Irrespective of which approach may turn out to be the best, the sensitivity and specificity definition and approach to ROC curve analysis for entire tractography pathways as demonstrated in the present study can be transferred to other validation approaches.

Alignment of histology and diffusion MRI data is an important concern in any histological validation effort. In the present study, a manual slice-by-slice adjustment was used which was based on a match of the white–gray matter boundaries between the 2 modalities. Although tedious, this process led to a precise alignment, which was undisturbed by irrelevant image features. For future efforts, it is desirable to perform automatic alignment of histological slices and MR images, in particular when dealing with a higher amount of data. This could for instance be guided by a mutual information criterion and might be complemented with edge-detecting image filtering techniques. Nevertheless, recent work shows that manual intervention is often still required with automated procedures (Choe et al. 2011). In principle, such procedures can also allow for more complex nonglobal or nonlinear alignment transformations. However, the application of these techniques to the specific alignment problem here requires considerable efforts in fine-tuning the respective parameters. These include immunity to the large differences in spatial resolution and elementary contrast in the 2 modalities and focusing on important image regions (such as tissue boundaries) while ignoring irrelevant ones.

SDT has been an important and widely used validation tool in medical and psychophysical fields for several decades (for review, see Zou 2002) and perfectly suits our approach. The main advantage of SDT for our studies is that it does not compute a direct comparison of 2 data sets but rather evaluates whether their outcome is the same. Therefore, it allows to compare data as different as a DTI volume and a series of histological slices. Particularly, a SDT analysis is relatively independent of the resolution and does not lose the microscopic information obtained from histology. Moreover, it is possible to determine sensitivity and specificity in their full dependence on the free parameters of the tractography algorithm, for example, FA or curvature thresholds. The resulting ROC curves give insight into the effect of those parameters on DTI tractography and can effectively be used as tuning curves. Thus, this approach does not only provide a statement about the overall goodness of inferred DTI tracts but also helps to adjust tractography algorithms optimally. This has been exemplified by optimizing the FA threshold of the line propagation algorithm in our study.

The ROC curves achieved here have typical shapes and clearly show the role of the FA threshold as the main detection threshold parameter of the line propagation algorithm. That is, pathways can be tracked more conservatively for high or more permissively for low FA thresholds. The optimal sens./spec. tradeoff was obtained for FA values between 0.02 and 0.08. However, values greater than 0.1 should not be chosen because the tracking algorithm is likely to be overly restrictive under

these conditions and consequently performs poorly. We evaluated the plausibility of these findings by examination of the whole pathway of selected tractography. This analysis revealed an optimal FA setting at 0.08 for 2 reasons. First, at this value, true pathways showed large amounts of true (i.e., stained) voxels on their pathway. Second, these pathways were biologically plausible, as opposed to pathways tracked with lower FA thresholds. Combining both results, FA thresholds of around 0.08 are likely to yield the best concordance of inferred and real fiber paths. This finding is in good accordance with a previous study, which identified FA values around 0.10 as optimal (Dauguet et al. 2007).

It is important to note that the effect of FA threshold on tractography results and its optimal value is obtained for paraformaldehyde-fixated postmortem tissue, scanned at room temperature (rather than body temperature). Although water diffusion directional anisotropy has been shown to be preserved in fixated tissue (Guilfoyle et al. 2003; Sun et al. 2003, 2005), both the chemical fixation process and the lower temperature are expected to lower the absolute value of ADCs and FA. Thus, the FA thresholds for tractography in this tissue are difficult to be transferred to in vivo measurements.

Evaluation Method

There have been several approaches to evaluate diffusion based imaging techniques with histological methods. In the majority of these studies, 3D images were reconstructed from a series of histological slices and consequently the overlap of DTI and the histological volumes was evaluated (e.g., Dauguet et al. 2007). Approaches of this kind bear some inherent limitations. Most notably, they are very vulnerable to inaccuracies in the registration process and therefore require either high tolerance thresholds or very simply structured tissue probes to obtain acceptable results. An alternative approach, which is more robust against distortions, is the receiver operator approach applied here which has been successfully used in recent studies (Iturria-Medina et al. 2011). Moreover, an ROC curve has the additional advantage over Pearson correlation–based approaches that it provides insight into the dynamic effect of any experimental parameter. That is, rather than just detecting the optimal setting of the parameters, it can precisely reveal the effect of the parameters in terms of sensitivity and specificity.

Conclusion

In summary, we can conclude that 1) DTI can principally reflect the shape and orientation of nerve pathways even on a small spatial scale and 2) that using our approach of diffusive dye staining that selectively marks entire single axons and ROC curve analysis, it is feasible to validate DTI on this level. Medium FA values of about 0.08 seem to be the best choice for DTI-based tractography on fixed postmortem tissue at room temperature when considering ROC curve analysis and tractography pathway analysis together. Given our results of approximately 80% sensitivity and specificity, we have been able to clearly demonstrate that this evaluation method reflects the concordance of DTI and histological measurements in noncomplex fiber architecture regions and is suitable for further evaluation studies of DW-MRI tractography approaches that may address issues, such as crossing and sharp intravoxel curvature.

Notes

References

Axer M, Amunts K, Grässel D, Palm C, Dammers J, Axer H, Pietrzyk U, Zilles K. 2011. A novel approach to the human connectome: ultra-high resolution mapping of fiber tracts in the brain. Neuroimage. 54:1091–1101.

Basser PJ, Mattiello J, LeBihan D. 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. J Magn Reson B. 103:247–254.

Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A. 2000. In vivo fiber tractography using DT-MRI data. Magn Reson Med. 44:625–632.

Beaulieu C. 2002. The basis of anisotropic water diffusion in the nervous system—a technical review. NMR Biomed. 15:435–455.

Burkhalter A, Bernardo KL, Charles V. 1993. Development of local circuits in human visual cortex. J Neurosci. 13:1916–1931.

Catani M, Howard RJ, Pajevic S, Jones DK. 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. Neuroimage. 17:77–94.

Choe AS, Gao Y, Li X, Compton KB, Stepniewska I, Anderson AW. 2011. Accuracy of image registration between MRI and light microscopy in the ex vivo brain. Magn Reson Imaging. 29:683–692.

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

D'Arceuil H, Liu C, Levitt P, Thompson B, Kosofsky B, de Crespigny A. 2008. Three-dimensional high-resolution diffusion tensor imaging and tractography of the developing rabbit brain. Dev Neurosci. 30:262–275.

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

Fawcett T. 2006. An introduction to ROC analysis. Pattern Recogn Lett. 27:861–874.

Galuske R, Schlote W, Bratzke H, Singer W. 2000. Interhemispheric asymmetries of the modular structure in human temporal cortex. Science. 289:1946–1949.

Godement P, Vanselow J, Thanos S, Bonhoeffer F. 1987. A study in developing visual systems with a new method of staining neurones and their processes in fixed tissue. Development. 101:697–713.

Guilfoyle DN, Helpern JA, Lim KO. 2003. Diffusion tensor imaging in fixed brain tissue at 7.0 T. NMR Biomed. 16:77–81.

Hess CP. 2009. Update on diffusion tensor imaging in Alzheimer's disease. Magn Reson Imaging Clin N Am. 17:215–224.

Honig MG, Hume RI. 1986. Fluorescent carbocyanine dyes allow living neurons of identified origin to be studied in long-term cultures. J Cell Biol. 103:171–187.

Honig MG, Hume RI. 1989. DiI and diO: versatile fluorescent dyes for neuronal labelling and pathway tracing. Trends Neurosci. 12:333–341.

Iturria-Medina Y, Perez Fernandez A, Morris DM, Canales-Rodriguez EJ, Haroon HA, García Penton L, Augath M, Galan García L, Logothetis N, Parker GJM, et al. 2011. Brain hemispheric structural efficiency and interconnectivity rightward asymmetry in human and nonhuman primates. Cereb Cortex. 21:56–67.

Kanaan RAA, Kim J, Kaufmann WE, Pearlson GD, Barker GJ, McGuire PK. 2005. Diffusion tensor imaging in schizophrenia. Biol Psychiatry. 58:921–929.

Kaufman JA, Ahrens ET, Laidlaw DH, Zhang S, Allman JM. 2005. Anatomical analysis of an aye-aye brain (Daubentonia madagascariensis, primates: Prosimii) combining histology, structural magnetic resonance imaging, and diffusion-tensor imaging. Anat Rec A Discov Mol Cell Evol Biol. 287:1026–1037.

Kubicki M, Westin CF, McCarley RW, Shenton M. 2005. The application of DTI to Investigate white matter abnormalities in schizophrenia. Ann N Y Acad Sci. 1064:134–148.

Leergaard TB, White NS, de Crespigny A, Bolstad I, D'Arceuil H, Bjaalie JG, Dale AM. 2010. Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain. PLoS One. 5:e8595.

Lin CP, Tseng WY, Cheng HC, Chen JH. 2001. Validation of diffusion tensor magnetic resonance axonal fiber imaging with registered manganese-enhanced optic tracts. Neuroimage. 14:1035–1047.

Maller JJ, Thompson RHS, Lewis PM, Rose SE, Pannek K, Fitzgerald PB. 2010. Traumatic brain injury, major depression, and diffusion tensor imaging: making connections. Brain Res Rev. 64:213–240.

Medina DA, Gaviria M. 2008. Diffusion tensor imaging investigations in Alzheimer's disease: the resurgence of white matter compromise in the cortical dysfunction of the aging brain. Neuropsychiatr Dis Treat. 4:737–742.

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

Pullens P, Roebroeck A, Goebel R. 2010. Ground truth hardware phantoms for validation of diffusion-weighted MRI applications. J Magn Reson Imaging. 32:482–488.

Richardson M. 2010. Current themes in neuroimaging of epilepsy: brain networks, dynamic phenomena, and clinical relevance. Clin Neurophysiol. 121:1153–1175.

Roebroeck A, Galuske R, Formisano E, Chiry O, Bratzke H, Ronen I, Kim D, Goebel R. 2008. High-resolution diffusion tensor imaging and tractography of the human optic chiasm at 9.4 T. Neuroimage. 39:157–168.

Sun SW, Neil JJ, Liang HF, He YY, Schmidt RE, Hsu CY, Song SK. 2005. Formalin fixation alters water diffusion coefficient magnitude but not anisotropy in infarcted brain. Magn Reson Med. 53:1447–1451.

Sun SW, Neil JJ, Song SK. 2003. Relative indices of water diffusion anisotropy are equivalent in live and formalin-fixed mouse brains. Magn Reson Med. 50:743–748.

Tham MW, Woon PS, Sum MY, Lee TS, Sim K. 2010. White matter abnormalities in major depression: evidence from post-mortem, neuroimaging and genetic studies. J Affect Disord. 132:26–36.

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

Wedeen VJ, Wang RP, Schmahmann JD, Benner T, Tseng WYI, Dai G, Pandya DN, Hagmann P, D'Arceuil H, de Crespigny AJ. 2008. Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. Neuroimage. 41:1267–1277.

Zou KH. 2002. Receiver operating characteristic (ROC) literature research [Internet]. Boston (MA): Harvard Medical School. [cited 2012 Feb 6]. Online bibliography available from: http://www.spl.harvard.edu/archive/spl-pre2007/pages/ppl/zou/roc.html.