

## On the cortical connectivity in the macaque brain: A comparison of diffusion tractography and histological tracing data

Gabriel Girard<sup>a,b,c,\*</sup>, Roberto Caminiti<sup>d</sup>, Alexandra Battaglia-Mayer<sup>e</sup>, Etienne St-Onge<sup>f</sup>, Karen S. Ambrosen<sup>g,h</sup>, Simon F. Eskildsen<sup>i</sup>, Kristine Krug<sup>j,k,1</sup>, Tim B. Dyrby<sup>h,g</sup>, Maxime Descoteaux<sup>f</sup>, Jean-Philippe Thiran<sup>a,b,c,1</sup>, Giorgio M. Innocenti<sup>c,m,n,1</sup>

<sup>a</sup> Radiology Department, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

<sup>b</sup> Center for BioMedical Imaging, Lausanne, Switzerland

<sup>c</sup> Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

<sup>d</sup> Neuroscience and Behavior Laboratory, Istituto Italiano di Tecnologia, Rome, Italy

<sup>e</sup> Department of Physiology and Pharmacology, University of Rome SAPIENZA, Rome, Italy

<sup>f</sup> Sherbrooke Connectivity Imaging Lab, Computer Science Department, Faculty of Science, Université de Sherbrooke, Sherbrooke, Canada

<sup>g</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark

<sup>h</sup> Danish Research Centre for Magnetic Resonance, Center for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

<sup>i</sup> Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

<sup>j</sup> Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom

<sup>k</sup> Institute of Biology, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany

<sup>l</sup> Leibniz-Institute for Neurobiology, Magdeburg, Germany

<sup>m</sup> Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

<sup>n</sup> Brain and Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

### ARTICLE INFO

#### Keywords:

Tractography  
White matter  
Diffusion MRI  
Histological tracing  
Ex-vivo  
Macaque monkey  
Cortico-cortical  
Connectivity  
Parieto-frontal network

### ABSTRACT

Diffusion-weighted magnetic resonance imaging (DW-MRI) tractography is a non-invasive tool to probe neural connections and the structure of the white matter. It has been applied successfully in studies of neurological disorders and normal connectivity. Recent work has revealed that tractography produces a high incidence of false-positive connections, often from “bottleneck” white matter configurations. The rich literature in histological connectivity analysis studies in the macaque monkey enables quantitative evaluation of the performance of tractography algorithms. In this study, we use the intricate connections of frontal, cingulate, and parietal areas, well established by the anatomical literature, to derive a symmetrical histological connectivity matrix composed of 59 cortical areas. We evaluate the performance of fifteen diffusion tractography algorithms, including global, deterministic, and probabilistic state-of-the-art methods for the connectivity predictions of 1711 distinct pairs of areas, among which 680 are reported connected by the literature. The diffusion connectivity analysis was performed on a different ex-vivo macaque brain, acquired using multi-shell DW-MRI protocol, at high spatial and angular resolutions. Across all tested algorithms, the true-positive and true-negative connections were dominant over false-positive and false-negative connections, respectively. Moreover, three-quarters of streamlines had endpoints location in agreement with histological data, on average. Furthermore, probabilistic streamline tractography algorithms show the best performances in predicting which areas are connected. Altogether, we propose a method for quantitative evaluation of tractography algorithms, which aims at improving the sensitivity and the specificity of diffusion-based connectivity analysis. Overall, those results confirm the usefulness of tractography in predicting connectivity, although errors are produced. Many of the errors result from bottleneck white matter configurations near the cortical grey matter and should be the target of future implementation of methods.

### 1. Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI) tractography is a non-invasive tool to probe neural connections and the orga-

nization of the central nervous system tissues (Basser et al., 2000; Jones et al., 1999; Mori and van Zijl, 2002). Tractography uses the anisotropy in the orientation sensitive measurements of DW-MRI to estimate the coarse trajectory of a group of axons, termed streamline. A collection

\* Corresponding author at: Radiology Department, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland.  
E-mail address: [gabriel.girard@epfl.ch](mailto:gabriel.girard@epfl.ch) (G. Girard).

<sup>1</sup> These authors contributed equally.

of streamlines delineates a pathway coursing in the brain's white matter. Streamlines are generally obtained by iteratively following the estimated axonal orientations until a stopping criterion is reached, such as terminating in the grey matter or exceeding a curvature constraint. See [Jeurissen et al. \(2017\)](#) for a detailed review of diffusion tractography methods.

Tractography has been applied in studies of neurological disorders (e.g. [Catani, 2006](#); [Ciccarelli et al., 2008](#); [Lin et al., 2005](#); [Lo et al., 2010](#); [Pannek et al., 2011](#); [Shu et al., 2011](#)), such as in stroke, to quantify the disconnections of fronto-parietal and fronto-occipital pathways ([Urbanski et al., 2011](#)) and the changes in the spatial distribution of the motor connections ([Newton et al., 2006](#)). It was also used in the identification of the spatial extent of tumors for neurosurgical intervention planning (e.g. [Fortin et al., 2012](#); [Golby et al., 2011](#); [Romano et al., 2009](#); [Yu et al., 2005](#)), in the identification of pathways involved in language (e.g. [Finkl et al., 2019](#); [Glasser and Rilling, 2008](#); [Sarubbo et al., 2019](#)), in the study of the topological organization of white matter pathways (e.g. [Chenot et al., 2019](#); [Hau et al., 2017](#); [Sarubbo et al., 2019](#)) and in white matter network organization (e.g. [Bassett et al., 2011](#); [Bullmore and Sporns, 2009](#); [Hagmann et al., 2008](#); [Johansen-Berg et al., 2004](#); [Mars et al., 2011](#); [2018](#); [Thiebaut de Schotten et al., 2012](#)). Despite the successful applications, recent studies have shown that tractography recovers existing white matter connections but produces a high incidence of false-positive connections when compared to ground truth connections ([Côté et al., 2013](#); [Jbabdi et al., 2015](#); [Maier-Hein et al., 2017](#); [Thomas et al., 2014](#)). One source of false-positive connections is white matter “bottleneck” configurations, first discussed in [Mangin et al. \(1996\)](#). They refer to white matter sites where multiple pathways converge, and their trajectory is constrained by surrounding grey matter. Bottleneck configurations were shown to be a source of spurious streamlines for tractography algorithms in [Guevara et al. \(2011\)](#). Recently, [Maier-Hein et al. \(2017\)](#) showed their effect in estimating the connectivity of a complex simulation dataset with known ground truth. They showed that even using ground truth white matter orientations and high spatial resolution, the bottleneck configurations lead tractography algorithms to incorrectly connect pairs of areas. [Girard et al. \(2017\)](#) showed that the ambiguities between the white matter fascicles crossings at the voxel level or bending within the voxel (kissing configurations) cannot be resolved using orientation alone at the millimeter scale. Despite the false-positives, the reproducibility of tractography has been shown to be high ([Dyrby et al., 2007](#)). This raises the question of sensitivity/specificity trade-off in the estimation of white matter network reconstructed using tractography. In addition to local bottleneck configurations, pathway characteristics affect their successful reconstruction using tractography, leading to false-negative connections and/or creating non-existing, false-positives, connections. We have grouped sources of bias in the tractography reconstruction in four categories: the shape, the location, the gyral, and the seeding biases.

First, tractography algorithms tend to more successfully reconstruct shorter, larger and straighter pathways, as opposed to long, narrow or bending ones ([Côté et al., 2013](#); [Daducci et al., 2016](#); [Descoteaux et al., 2009](#); [Girard et al., 2014](#); [Jbabdi and Johansen-Berg, 2011](#); [Jeurissen et al., 2017](#); [Jones, 2010](#); [Jones and Cercignani, 2010](#); [Jones et al., 2013](#); [Rheault et al., 2020](#); [Thomas et al., 2014](#); [Tournier et al., 2011](#); [Yeh et al., 2016](#)). Thus, this results in tractography under-representing some pathways with particular curvature, size, and length. This is due to the accumulation of local errors in the propagation directions of streamlines, selected by the algorithms. The sources of these errors are, among others, the presence of noise and artifacts in the DW-MRI data ([Coupé et al., 2006](#); [Manjón et al., 2010](#); [Sotiropoulos et al., 2013](#); [Veraart et al., 2016](#)), the limited spatial and angular resolutions of the diffusion measurements ([Heidemann et al., 2012](#); [Schilling et al., 2017](#); [Smith et al., 2015b](#); [Sotiropoulos et al., 2013](#); [Ugurbil et al., 2013](#)), and the assumptions made in the modeling or estimation of fiber orientations ([Jeurissen et al., 2014](#); [Schilling et al., 2019a](#)).

Second, white matter pathways bordering grey matter or cerebrospinal fluid voxels tend to be underrepresented by tractography, due to partial volume effects (voxels with two or more tissue compartments), which increase both the angular error and the detection of spurious orientations in the estimation of the local fiber orientation distributions ([Alexander et al., 2001](#); [Roine et al., 2014](#)). These errors cause tractography to deviate, giving rise to spurious streamlines, or to incorrectly stop before reaching the targeted grey matter, creating incomplete segments of pathways ([Côté et al., 2013](#)). Additionally, tractography is more likely to incorrectly terminate in areas bordering the white matter volume, due to errors in the stopping constraints ([Girard et al., 2014](#); [Smith et al., 2012](#)), such as a tractography mask blocking the streamline propagation due to its low spatial resolution (i.e. the staircase effect).

Third, most tractography algorithms suffer from the so-called, gyral bias ([Donahue et al., 2016](#); [Reveley et al., 2015](#); [Rheault et al., 2020](#); [Schilling et al., 2018a](#); [St-Onge et al., 2018](#); [Teillac et al., 2017](#); [Van Essen et al., 2014](#)), which designates an overrepresentation of pathways connecting the gyral crown while missing most of the fanning organization of the gyral walls. This is partly due to the low spatial resolution of DW-MRI images, where only a few voxels separate the sulcal walls. Additionally, the gyral bias is emphasized by inadequate curvature constraints imposed to help propagating streamlines in the white matter. This constraint is generally fixed for the whole reconstruction and is typically too restrictive to fully capture the local fanning organization of fibers entering the gyrus and connecting the walls of the sulci ([Reveley et al., 2015](#); [Van Essen et al., 2014](#)).

Fourth, tractography is affected by the seeding (i.e. initialization) technique used ([Girard et al., 2014](#); [Li et al., 2012](#); [Rheault et al., 2020](#); [Smith et al., 2012](#); [St-Onge et al., 2018](#)). Seeding from all white matter voxels will artificially over-represent long pathways by placing more seeds (i.e. initial streamline positions) within the voxels they cover, compared to short pathways ([Donahue et al., 2016](#); [Li et al., 2012](#); [Liptrot et al., 2014](#)). Seeding from the white matter and grey matter interface reduces this bias but makes the tractography more susceptible to spurious initial orientations, due to partial volume effects ([Roine et al., 2014](#)) and has difficulties in reconstructing narrow pathways, with few voxels at the white matter and grey matter interface, such as in the fornix and the posterior commissure. Nevertheless, [Ambrosen et al. \(2020\)](#) have obtained similar results at the structural connectome level using both white matter and interface seeding strategies.

Those biases will affect the overall streamline distribution, and thus the structural connectivity matrix to a various extent for each algorithm and pathway. In some cases, those biases have opposite effects that can cancel out, such as the long pathways under-represented due to the accumulation of error, but over-represented by white matter seeding. Nevertheless, algorithms should target the reduction of bias in the distribution of streamlines across white matter pathways.

Recently, algorithms using anatomical priors have been introduced to better constrain the propagation of streamlines in the white matter and produce more plausible pathways. [Smith et al. \(2012\)](#) and [Girard et al. \(2014\)](#) proposed respectively the ACT (Anatomically-constrained tractography) and the PFT (Particle Filtering Tractography) methods using tissue information derived from an anatomical T1-weighted image to constrained the propagation of streamlines in the white matter. [Savadjiev et al. \(2014\)](#) and [St-Onge et al. \(2018\)](#) proposed geometrical approaches based on the brain morphology to guide the tractography near the cortex. [Aydogan and Shi \(2016, 2018\)](#) use a geometric prior to guide the streamline propagation in the white matter and improve the regularity of streamlines. Other tractography methods use prior knowledge on fascicles position and shape ([Chamberland et al., 2017](#); [Rheault et al., 2019](#)).

Other methods were proposed to reconstruct the whole white matter at once, with so-called global tractography algorithms ([Christiaens et al., 2015](#); [Fillard et al., 2009](#); [Jbabdi et al., 2007](#); [Kreher et al., 2008](#); [Mangin et al., 2002](#); [Reisert et al., 2014](#); [2011](#); [Teillac et al., 2017](#)). These methods show improvement in the reconstruction of white mat-

ter pathways compared to streamline propagation methods in regions of crossing configurations (Christiaens et al., 2015; Fillard et al., 2009; Reisert et al., 2014), and at the white matter and grey matter interface (Teillac et al., 2017). Although the results are promising, the optimization procedure remains challenging and requires the selection of several parameters.

Post-processing approaches, such as MicroTrack (Sherbondy et al., 2010), COMMIT (Daducci et al., 2013; 2014), SIFT (Smith et al., 2013; 2015a) and LiFE (Pestilli et al., 2014), use modeling of the DW-MRI signal and streamlines to remove or to penalize streamlines with a poor agreement with the diffusion measurements. This has been shown to improve the reproducibility of connectivity matrices and reduce streamline density bias (Daducci et al., 2014; Smith et al., 2015b). Other post-processing methods use geometrical approaches to identify and remove streamlines with unrealistic trajectories (Aydogan and Shi, 2015; Garyfallidis et al., 2012; 2018; Jordan et al., 2018; Wang et al., 2018). These approaches use the topographic regularity of white matter pathways to identify spurious trajectories from a set of streamlines.

The rich and extensive literature in histological tracing studies in animal models (see Averbeck and Seo, 2008; Calabrese et al., 2015a; 2015b; Caminiti et al., 2017; Dyrby et al., 2007; Knösche et al., 2015; Markov et al., 2014, and references therein) offer a tool to evaluate the performance of tractography (Ambrosen et al., 2020; Aydogan et al., 2018; Azadbakht et al., 2015; Crosson et al., 2005; Dauguet et al., 2007; Deletre et al., 2019; Donahue et al., 2016; Hagmann et al., 2008; van den Heuvel et al., 2015; Jbabdi et al., 2013; Leergaard et al., 2010; Liu et al., 2020; Markov et al., 2014; Sarubbo et al., 2019; Schilling et al., 2017; 2019a; 2018b; 2019b; Thomas et al., 2014; Van Essen et al., 2014). Indeed, neural pathways delineated with axonally transported tracers provide the strongest ground-truth to evaluate the performance of tractography (Dyrby et al., 2018). Nevertheless, structural connectivity derived from tracing studies might fail to demonstrate pathways documented by tractography, in specific cases, calling for other types of validation or new anatomical experiments (Dyrby et al., 2018; Knösche et al., 2015).

In this study, we use the ipsilateral cortical connections among frontal, cingulate, and parietal areas of the macaque monkey brain, well established by the anatomical tracing literature. We use the histological connectivity to compare various tractography algorithms in their ability to correctly identify connected and unconnected cortical areas among 1711 distinct pairs of areas. The tractography analysis is performed on high spatial and angular resolutions multi-shell ex-vivo DW-MRI measurements. We compare fifteen tractography algorithms, including global, deterministic and probabilistic state-of-the-art methods, with and without anatomical constraints and post-processing filtering. We especially investigated the effect of the streamline propagation method on the connectivity estimation, analyzing the patterns of incorrectly connected areas (i.e. false-positive connections) and incorrectly unconnected areas (i.e. false-negative connections) each method produces. Moreover, we report the effect of bottleneck and crossing white matter configurations on the structural connectivity estimation. Overall, we propose a quantitative method for the evaluation of tractography algorithms, with the aim of improving the sensitivity and the specificity of diffusion-based connectivity analysis.

## 2. Materials and methods

### 2.1. Diffusion data processing

One ex-vivo monkey brain, available from Ambrosen et al. (2020), was used (M105, *Macaca mulatta*, male, 4 years and 10 months old, 10.1kg body weight, see Ahmed et al. (2012) for the perfusion protocol). The brain was prepared for scanning as described by Dyrby et al. (2011). DW-MRI data was acquired at 0.5mm isotropic resolution on three shells of 180 uniformly distributed gradient directions with  $b = [1.477, 4.102, 8.040]$  ms/um<sup>2</sup>,  $G = [150, 250, 350]$  mT/m,  $\delta = 8$  ms,  $\Delta = 17$  ms,  $TE = 30$  ms, and 9  $b = 0$  ms/um<sup>2</sup> (b0) images (see Ambrosen et al. (2020) for

more details). The DW-MRI acquisition was done twice and averaged for further processing. Visual inspection revealed no need for pre-processing and state-of-the-art fiber orientation distribution functions estimation was directly performed using the Multi-Shell Multi-Tissue Constrained Spherical Deconvolution algorithm (Jeurissen et al., 2014; Tournier et al., 2019). Partial volume estimates for the white matter, grey matter, and cerebrospinal fluid were obtained from the averaged b0 image using FSL Fast (Zhang et al., 2001). We used the white matter surface and the midcortical surface from Ambrosen et al. (2020), which were generated by the Fast Accurate Cortex Extraction algorithm (Eskildsen and Østergaard, 2006).

### 2.2. Cortical parcellation and histological data

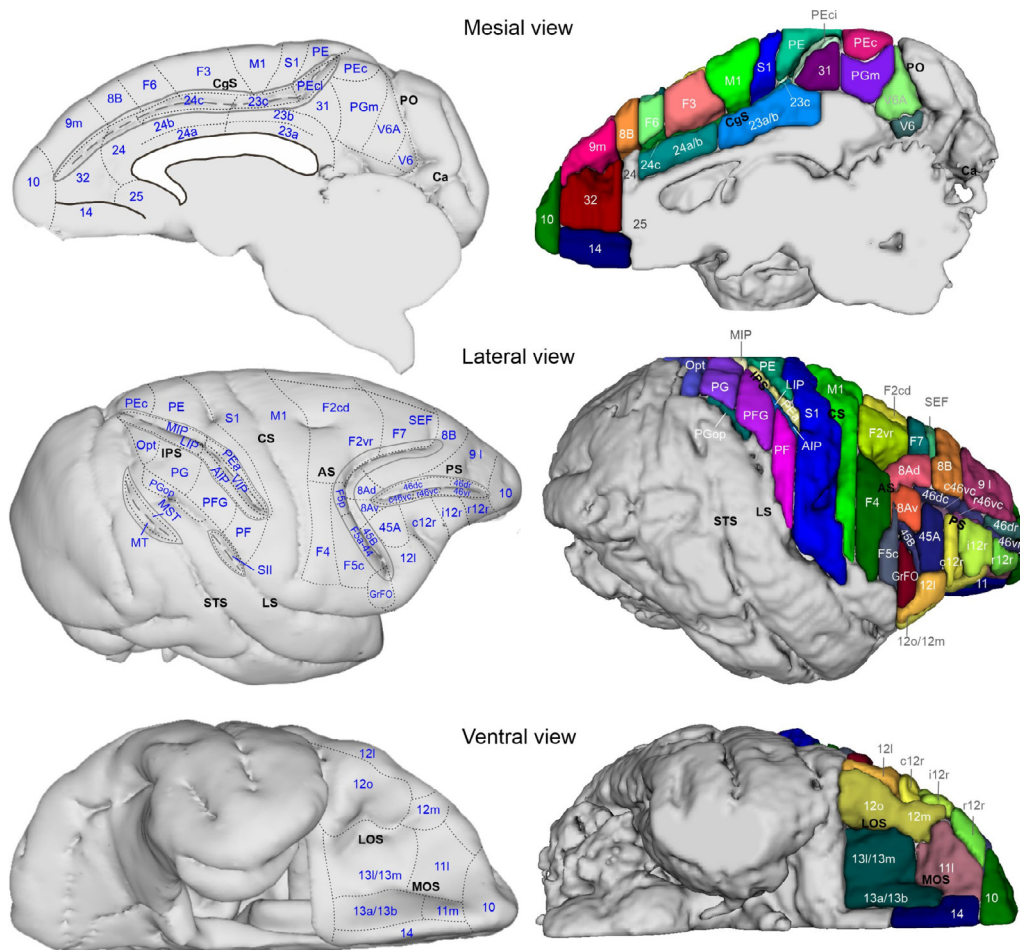
A symmetrical histological connectivity matrix for 59 cortical regions of interest (ROIs) was obtained from a systematic retrograde tracer literature review (see Caminiti et al., 2017, and references therein). The location in the brain of the cortical areas taken into consideration is shown in Fig. 1 (left). All 59 studied cortical areas were injected with retrograde tracers. A literature review of the reported connectivity for each area to each other area was performed. The histological analysis resulted in a symmetrical connectivity matrix, showed in Fig. 2, containing 1711 distinct pairs of ROIs, among which 680 are connected (light green) and 1031 are unconnected (dark green). A pair of ROIs was labeled connected by histology if any number of labeled cells was reported in an ROI, for an injection site in the other ROI, and unconnected otherwise. Hence, the histological connectivity matrix is constructed from several brains, each injected with a retrograde tracer in one or a few locations. The brain used for the diffusion data acquisition was not used for histological tracing. See Section 4.1 for a discussion on the limitations of histological tracing for connectivity analysis.

The average b0 MRI image was used to manually parcellate the 59 ROIs of the right hemisphere of the macaque brain, shown in Fig. 1 (right). The parcellation was performed using the itk-SNAP software (Yushkevich et al., 2006) following Caminiti et al. (2017) and atlases of the rhesus monkey brain (Paxinos et al., 2000; Saleem and Logothetis, 2012). A detailed description of the parcellated areas is given in Appendix. The cortical parcellation of the 59 ROIs and the corresponding histological connectivity matrix is available at <http://www.drctr.dk/rhesus-macaque-brain>.

#### 2.2.1. Diffusion tractography algorithms

Deterministic and probabilistic streamline tractography algorithms from the DIPY software package (*DIPY Probabilistic*, *DIPY Deterministic*, *DIPY EuDX*; Garyfallidis et al., 2014) and the MRtrix3 software package (*MRtrix3 iFOD2*, *MRtrix3 SDSTREAM*; Tournier et al., 2010, 2019) were used to reconstruct the white matter pathways. Additionally, we reconstructed the pathways using Anatomically-Constrained Tractography (*MRtrix3 iFOD2 ACT*; Smith et al., 2012), Particle Filtering Tractography (*DIPY Probabilistic PFT*; Girard et al., 2014) and Surface-Enhanced Tractography (*DIPY Probabilistic PFT SET*; St-Onge et al., 2018) methods. Microstructure-informed tractography (*DIPY Probabilistic PFT COMMIT*, *MRtrix3 iFOD2 ACT SIFT2*; Daducci et al., 2014; Smith et al., 2015a) and the Cluster Confidence Index (*DIPY Probabilistic CCI*; Jordan et al., 2018) streamlines filtering/weighting methods were also used. The Stick-Zeppelin-Ball microstructure model was used for COMMIT (Daducci et al., 2014; Panagiotaki et al., 2012). The Cluster Confidence Index was computed using a maximum streamline distance of 2.5mm; streamlines with  $CCI < 1$  were removed.

The fiber orientation distributions were provided as input to all previously mentioned tractography algorithms. Moreover, diffusion tensor-based tractography algorithm (*MRtrix3 Tensor Det*; Basser et al., 2000; Tournier et al., 2019) and global tractography algorithm (*MRtrix3 Global*; Christiaens et al., 2015) were included in the analysis. For the *MRtrix3 Global* method, we combined the output of eight runs of the method to increase the number of generated streamlines.



**Fig. 1.** Mesial, lateral and ventral views of the of the macaque brain showing the parcellation of the cortical areas used for this study (left), and their 3D manual reconstruction on the MRI images used for tractography (right). On the mesial view, Ca indicates the calcarine fissure, PO and CgS parieto-occipital and cingulate sulcus, respectively; on the lateral view, IPS, STS, CS, LS, AS and PS indicate in the order intraparietal, superior temporal, central, lateral, arcuate, and principal sulci; on the ventral view, LOS and MOS indicate, respectively, the lateral and medial orbitofrontal sulci.

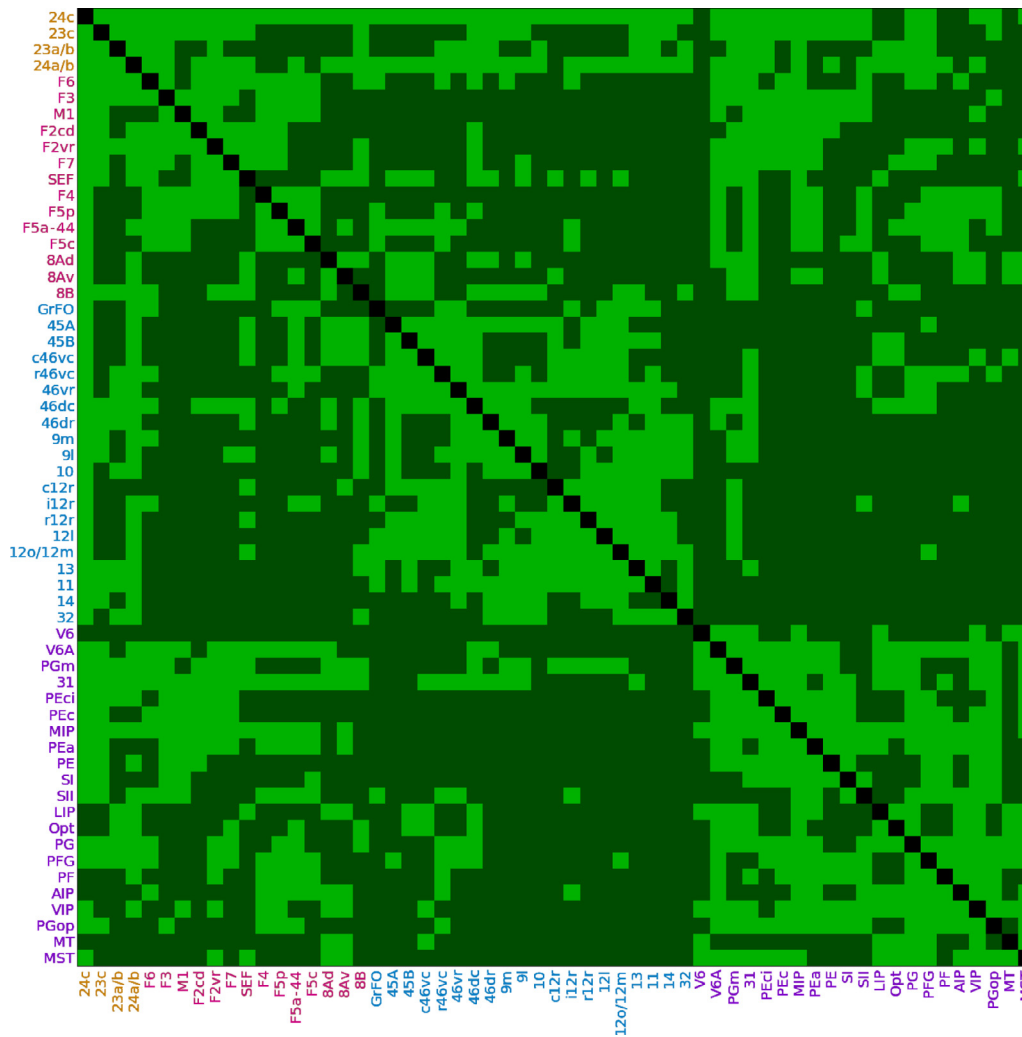
Whole-brain tractography was performed in the white matter mask obtained from non-zero white matter partial volume estimate voxels (ACT and PFT methods used the full partial volume estimate maps). The tractography seeding was done in the white matter mask for all previously mentioned methods. Additionally, we tested two methods with tractography seeding done at the white matter and grey matter interface (*DIPY Probabilistic PFT gmwmi* and *MRtrix3 iFOD2 ACT gmwmi*). The interface mask was obtained by selecting all voxels with partial volume estimates of both white matter and grey matter higher than 0.1 (Girard et al., 2014). Each method was run such that the total number of streamlines connecting any pair of ROIs of the matrix was higher than 100,000. The tractography step size, maximum angle, and fiber orientation distribution threshold were chosen following recommendations from authors of the individual algorithms. No minimum streamline length threshold was used (i.e. 0mm). Supplementary Figure S1 shows the white matter surface used as input for the SET algorithm in white, the surface after the modified mean-curvature flow algorithm in green, and a subset of streamlines generated with the PFT tractography algorithm connecting the green surface. The SET algorithm iteratively follows the inward normal direction to the surface in regions of positive curvature, forming the streamline segment connecting the two surfaces.

### 2.3. Tractography connectivity analysis

Streamlines with endpoints in ROIs were identified using the *MRtrix3 tck2connectome* (Tournier et al., 2019) command, using a radial search

of 1mm (streamlines with endpoints within 1mm distance from an ROI voxel were assigned to it). All streamlines without both endpoints in our cortical parcellation were discarded from the connectivity analysis (e.g. incomplete streamlines terminating in the white matter, streamlines of the left hemisphere, the corpus callosum and the corticospinal tracts). The tractography connectivity matrix was constructed using the number of streamlines between each pair of ROIs divided by the total number of streamlines connecting those ROIs with any other one, to account for ROI volume differences (Delettre et al., 2019; Donahue et al., 2016). For *DIPY Probabilistic PFT COMMIT* and *MRtrix3 iFOD2 ACT SIFT2* algorithms, the connectivity matrix was constructed using the contribution weights estimated for each streamline (Daducci et al., 2014; Smith et al., 2015a).

We report the Receiver Operating Characteristic (ROC) and the corresponding Area Under Curve (AUC; Fawcett, 2006), which describes the overall classification performances of the tested tractography algorithms (classifying each pair of ROIs as connected or unconnected by histological tracing). To build the ROC curves, we iteratively threshold the tractography connectivity matrices, considering only values above the threshold as connected. We report the True-Positives (TP) as pairs of ROIs labeled as connected by both tractography and histology, and the True-Negatives (TN) as unconnected pairs by both methods. False-Positives (FP) are reported as pairs of ROIs connected by tractography and unconnected by histology, and conversely for False-Negatives (FN). Additionally, for each tractography algorithm, we report the sensitiv-



**Fig. 2.** Symmetrical connectivity matrix of 59 cortical regions of interest (ROIs) from histological studies based on retrograde transport of neural tracers (Caminiti et al., 2017). Connected pairs of ROIs are shown in light green and unconnected pairs of ROIs in dark green. The matrix shows 1711 distinct pairs of ROIs, 680 connected and 1031 unconnected. ROIs are grouped in four sub-networks: cingulate (orange), motor, premotor and oculomotor (red), prefrontal (blue) and parietal (purple). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ity ( $\frac{TP}{TP+FN}$ ), i.e. the proportion of correctly identified connected pair of ROIs among all pairs labeled connected, on the basis of histological evidence, and the specificity ( $\frac{TN}{FP+TN}$ ), i.e. the proportion of correctly identified unconnected pair of ROIs among all pairs labeled as unconnected in the available histological data. The ideal streamline fraction threshold of each tractography algorithm was selected by maximizing the Youden's index (defined as  $sensitivity + specificity - 1$ ; Thomas et al., 2014; Youden, 1950), which varies from 0 to 1, and indicates the overall tractography classification performance in reproducing histological connectivity. Finally, we report the classification performances for the algorithm-specific threshold maximizing the Youden's index, and for three fixed threshold values of 0.01, 0.001 and 0.

To quantify the percentage of streamlines with endpoints in ROIs labeled as connected by histological studies, but with erroneous trajectories, we used the streamline's length median absolute deviation (Leys et al., 2013),

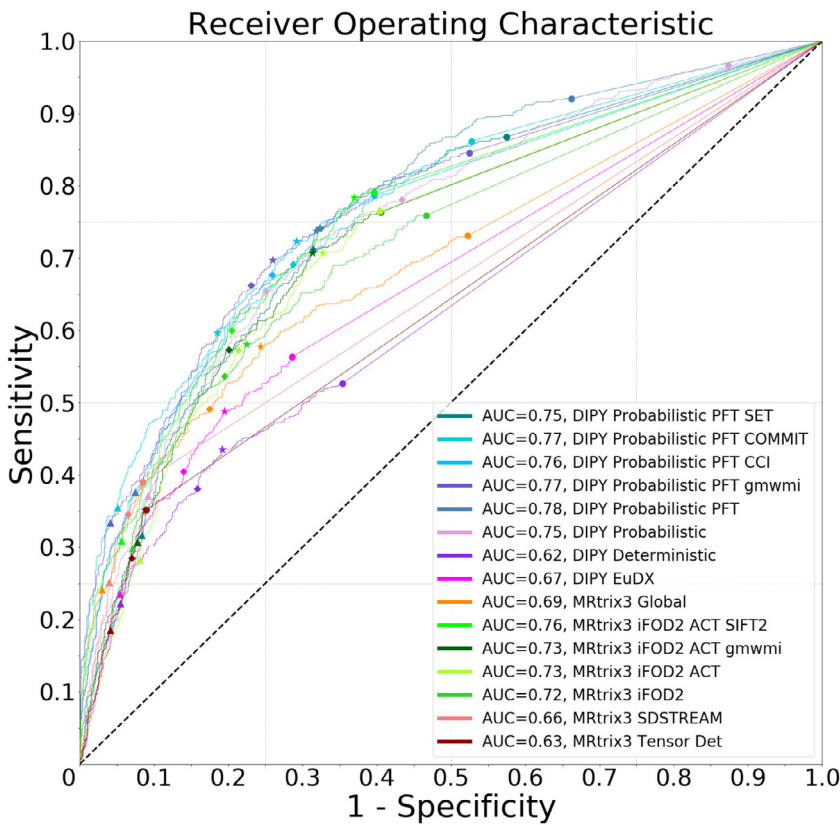
$$MAD = b * Median_i(|l_i - Median_j(l_j)|),$$

where  $l$  is the length of the  $n^{th}$  streamline and  $b = 1.4826$  is a normality constant (Rousseeuw and Croux, 1993). For each pair of ROIs labeled as connected by histology, streamlines with length superior to 2.5 times the MAD of the length of all streamlines estimated for that connection

were considered outliers ( $\frac{l - Median_j(l_j)}{MAD} > 2.5$ ; Leys et al., 2013; Miller, 1991).

### 3. Results

To evaluate the performance of the fifteen tested tractography algorithms, we compare their ability to classify each pair of ROIs (connected or unconnected). Fig. 3 shows the corresponding ROCs curves for a decreasing threshold on the tractography connectivity value compared to the histological connectivity matrix. For each algorithm, the star marker indicates the point on the ROC curve obtained for the threshold that maximizes the Youden's index. The triangle, diamond, and circle markers indicate the performances of each algorithm for fixed threshold values of 0.01, 0.001, and 0, respectively. The Area Under Curve (AUC) of each algorithm is reported in the legend. The prediction performances for the connectivity of the 1711 distinct pairs of ROIs for all tested tractography algorithms are shown in Supplementary Figure S3. The corresponding sensitivity, specificity, and Youden's index are shown in Fig. 4. Probabilistic tractography algorithms (*DIPY Probabilistic* and *MR-trix3 iFOD2* variants) have higher AUC compared to other algorithms. They also show higher Youden's index for most thresholding parameters and algorithms.



**Fig. 3.** Receiver Operating Characteristic (ROC) curves of the classification of pairs of ROIs (connected or unconnected) of the fifteen tested diffusion tractography algorithms for a decreasing threshold of streamline fractions. The legend lists the Area Under Curve (AUC) measures of classification performances. The star markers indicate the algorithm-specific ideal threshold maximizing the Youden's index. The triangles, diamonds, and circle markers indicate the classification performances for three fixed thresholds of 0.1, 0.01, and 0 (without thresholding), respectively. A perfect classification performance would produce a curve reaching the top left corner (*sensitivity* = 1 and *specificity* = 1) with an AUC of 1. The diagonal black dashed line represents the performance of a random classification of the pairs of ROIs.

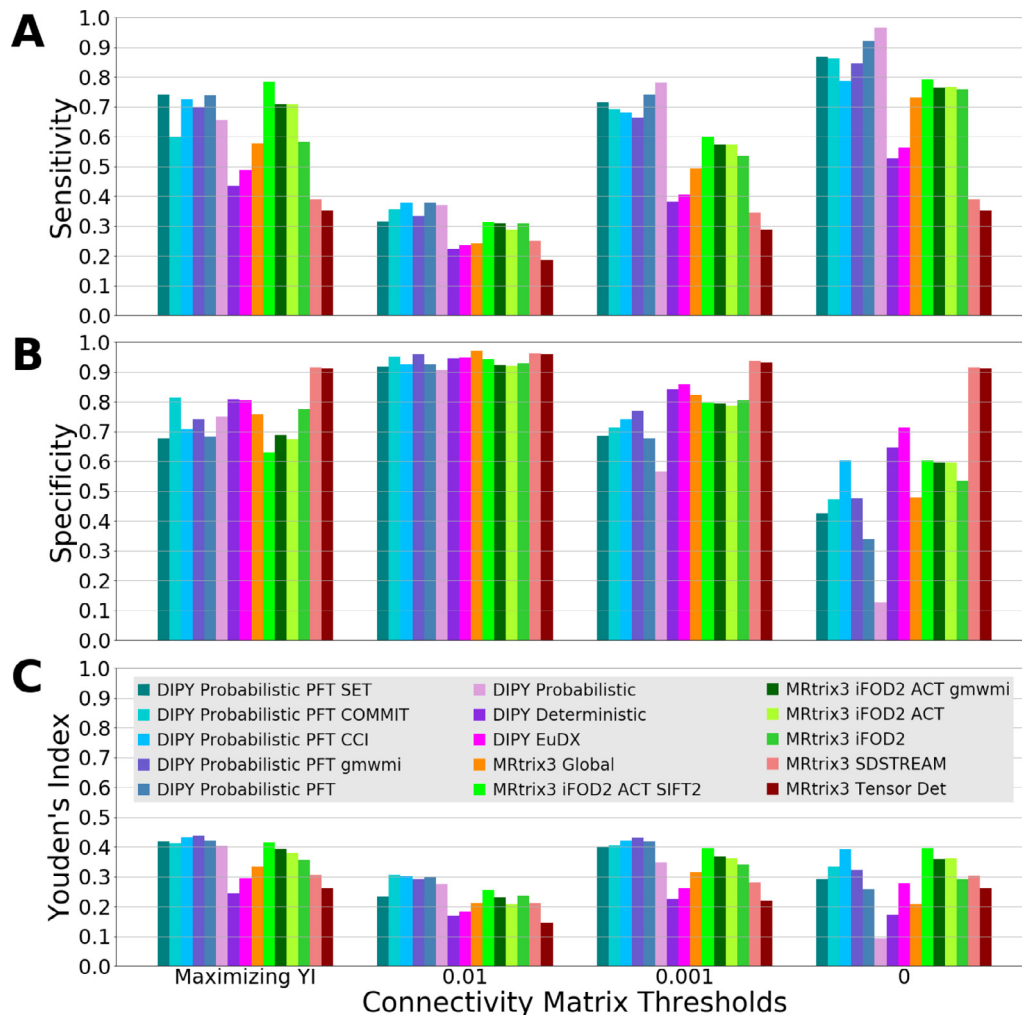
**Table 1**

For each tractography algorithm, the table lists the number of generated streamlines for the whole-brain tractogram and the number of connecting streamlines, ending in any pairs of ROIs (59 ROIs of the right hemisphere). Among the connecting streamlines, the table lists the percentage of streamlines ending in pairs of ROIs labeled connected by histological data (valid streamline), the percentage of valid streamlines with length more than 2.5 time the median absolute deviation (MAD) of their corresponding connection, and the algorithm-specific threshold maximizing the Youden's index. For *DIPY Probabilistic PFT COMMIT* and *MRtrix3 iFOD2 ACT SIFT2* algorithms, the table shows the streamline weighted percentages.

Tractography Algorithm	Number of Generated Streamlines	Number of Connecting Streamlines	Percentage of Valid Streamlines	Percentage of Valid Streamlines with MAD > 2.5	Threshold Maximizing Youden's Index
<i>DIPY Probabilistic PFT SET</i>	3,885,500	110,479	70%	2.9%	8.92e-4
<i>DIPY Probabilistic PFT COMMIT</i>	3,885,500	187,266	82%	4.2%	2.17e-3
<i>DIPY Probabilistic PFT CCI</i>	3,885,500	177,029	75%	2.8%	6.02e-4
<i>DIPY Probabilistic PFT gmwmi</i>	3,284,260	120,279	83%	7.5%	7.48e-4
<i>DIPY Probabilistic PFT</i>	3,885,500	187,266	74%	3.5%	1.03e-3
<i>DIPY Probabilistic</i>	3,885,500	216,811	69%	3.1%	2.42e-3
<i>DIPY Deterministic</i>	3,885,500	248,884	70%	5.0%	5.31e-4
<i>DIPY EuDX</i>	7,770,000	268,790	73%	2.4%	2.53e-4
<i>MRtrix3 Global</i>	2,114,940	116,300	89%	-	5.23e-4
<i>MRtrix3 iFOD2 ACT SIFT2</i>	3,000,000	217,834	73%	2.9%	1.10e-4
<i>MRtrix3 iFOD2 ACT gmwmi</i>	3,000,000	112,775	81%	5.5%	1.61e-4
<i>MRtrix3 iFOD2 ACT</i>	3,000,000	217,834	70%	3.0%	1.41e-4
<i>MRtrix3 iFOD2</i>	5,000,000	193,382	76%	3.0%	6.97e-4
<i>MRtrix3 SDSTREAM</i>	8,000,000	225,945	80%	-	2.72e-5
<i>MRtrix3 Tensor Det</i>	5,000,000	260,679	73%	-	4.77e-5

Table 1 lists the number of streamlines generated for each tractography algorithm, the number of streamlines connecting any pairs of ROIs, the percentage of streamlines with endpoints in pairs of connected ROIs, and the percentage of outlier streamlines having a length more than 2.5 times the median absolute deviation (MAD) compared to the length other streamlines of the same connection. The percentage of valid streamlines is the number of streamlines in all pairs of ROIs labeled connected by histology over the total number of streamlines connecting

any pairs of ROIs. Across the tested algorithms, the average percentage of valid streamlines is 76%. The *MRtrix3 Global* algorithm has the highest percentage of valid streamlines with 89%. Across all tested algorithms, up to 7.5% of streamlines with valid endpoints location had a length deviating from the length of other streamlines connecting the same ROIs. Three algorithms, *MRtrix3 Global*, *MRtrix3 SDSTREAM*, and *MRtrix3 Tensor Det*, did not consistently show spurious streamline trajectories, hence we did not apply outlier removal derived from the median



**Fig. 4.** Classification performance of diffusion tractography algorithms for ROIs connectivity using the algorithm-specific threshold maximizing the Youden's Index (YI) and three fixed threshold values: 0.01, 0.001 and 0 (without thresholding). The sub-figures (A-C) show the sensitivity, the specificity and the Youden's index, respectively.

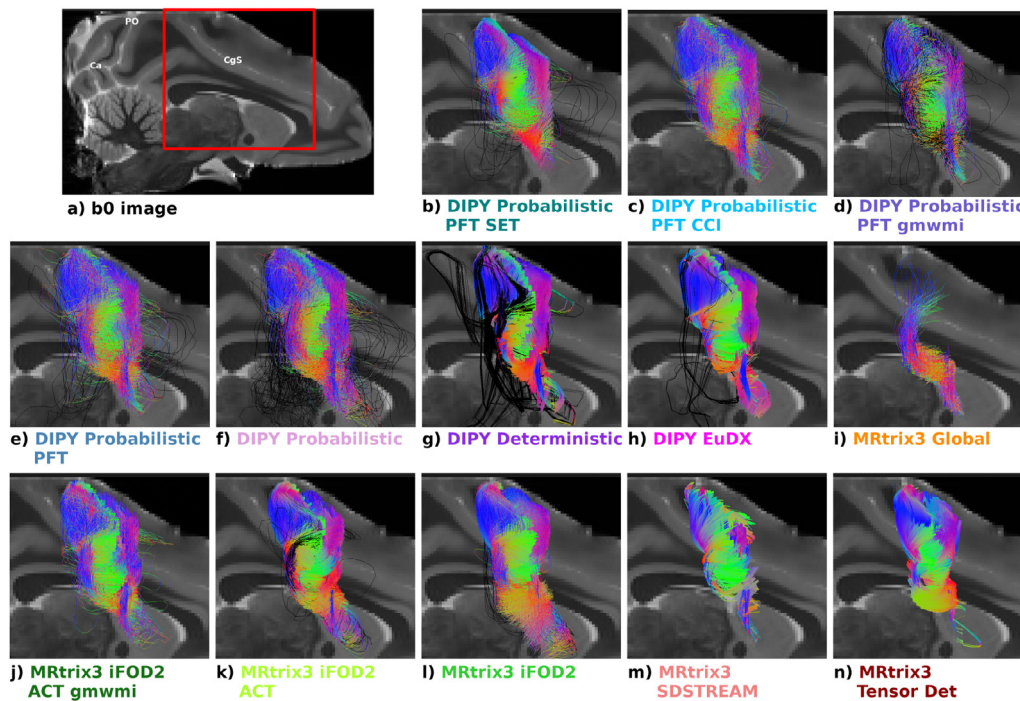
absolute deviation (*MAD*). This is qualitatively shown in Fig. 5 on the streamlines of the true-positive connection of primary somatosensory cortex (SI) and primary motor cortex (M1/F1). Streamlines with an outlier length are shown in black, whereas other streamlines are colored using the orientation of segments following the red-green-blue convention (red: left-right, green: anterior-posterior, blue: superior-inferior).

To further investigate tractography errors, we selected the *DIPY Probabilistic PFT SET* algorithm, which has both a high Youden's index score and a reported increase in cortical surface coverage properties. The classification performance for the threshold maximizing the Youden's index is shown in Fig. 6 (TP: light green, TN: dark green, FP: light red, FN: dark red). The tractography connectivity matrix before thresholding is shown in Supplementary Figure S2 (see Supplementary Video for a visualization of the classification performance at various thresholds). The sensitivity, specificity, and Youden's index of each ROI (column of the matrix) projected on the surface of the middle cortical layer are reported in Fig. 7 (see Supplementary Figures S4, S5 and S6 for the surface projections of the classification performances for all tested algorithms using the algorithm-specific ideal threshold.).

The structural connectivity of some ROIs is well predicted by tractography with both high specificity and sensitivity (yellow). However, tractography underperformed for other ROIs, showing lower specificity and low sensitivity (red and orange), with a mean across all ROIs of 0.67 (standard deviation of 0.20) and 0.74 (standard deviation 0.19), re-

spectively. On the lateral aspect of the hemisphere, sensitivity is higher around the central sulcus, and decays slightly moving rostrally toward the frontal pole and caudally toward the superior temporal sulcus and the parieto-occipital junction (Fig. 7A, left). On the mesial aspect of the hemisphere (Fig. 7A, right), sensitivity is higher across most of the antero-posterior extent of the cortex located dorsal to the cingulate sulcus and within its banks, as well as around area 32. On the contrary, it is low around the frontal pole, including orbito-frontal areas. An opposite pattern is observed for specificity (Fig. 7B), both on the lateral and mesial aspects of the hemisphere. The corresponding Youden's index is shown in Fig. 7C, where areas with the highest score are V6 (0.84), MIP (0.80) and 32 (0.74), and areas with the lowest are 45B (0.04), 23a/b (0.07) and 8Av (0.07).

Large false-positive connections reconstructed using *DIPY Probabilistic PFT SET* tractography algorithm are shown in Fig. 8 (red streamlines). The second and third columns show true-positive connections in overlapping white matter areas in blue and green, respectively. The last column shows the false-positive connection with the streamline's segments colored in blue if the voxel is shared with the true-positive of column 2, in green if the voxel is shared with the true-positive of column 3 or in cyan for both. Each false-positive connection shares a large portion of its volume with one or both true-positive connections. In Fig. 8A,B,C, the tractography incorrectly follows the direction associated with other connections that form a merging, i.e. a bottleneck, configuration below



**Fig. 5.** Streamlines reconstructed using tractography for the true-positive connection between area SI (primary somatosensory cortex) and area M1 (primary motor area / F1). Streamlines are superimposed on a mesial sagittal MRI section. Streamlines segments are colored using the red-green-blue convention (red:left-right, green:anterior-posterior, blue: superior-inferior; Calamante et al., 2012; Pajevic and Pierpaoli, 1999). Streamlines with length superior to 2.5 the median absolute deviation (MAD) are shown in black. See Fig. 1 for anatomical references. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the principal sulcus. Similarly, Fig. 8D shows a false-positive connection constructed from segments of two true-positive connections forming a bottleneck below the precentral gyrus, and Fig. 8E shows a bottleneck configuration below the arcuate sulcus. Those five false-positive connections are shown for all algorithms, using the same coloring approach, in Supplementary Figures S7, S8, S9, S10 and S11. Although, for some algorithms those false-positives are small, they tend to show a similar overlap with the selected true-positives.

The few streamlines reconstructed using *DIPY Probabilistic PFT SET* tractography algorithm connecting orbitofrontal and medial area 12 to inferior parietal area PFG, identified as a false-negative connection, are shown in Fig. 9. This connection was incorrectly rejected using tractography because not enough streamlines were reconstructed, i.e. the connectivity score ( $7.82e - 4$ ) was lower than the threshold maximizing the Youden's index ( $8.92e - 4$ ). Although only a few streamlines were reconstructed (shown in red), those streamlines share an ending area in a bottleneck configuration with well reconstructed true-positive connections (shown in blue and green). Among the fifteen tested algorithms, *MRtrix3 iFOD2* and *MRtrix3 Global* were the only ones correctly reporting this connection with a diffusion score high than their optimal threshold (See Supplementary Figure S12).

#### 4. Discussion

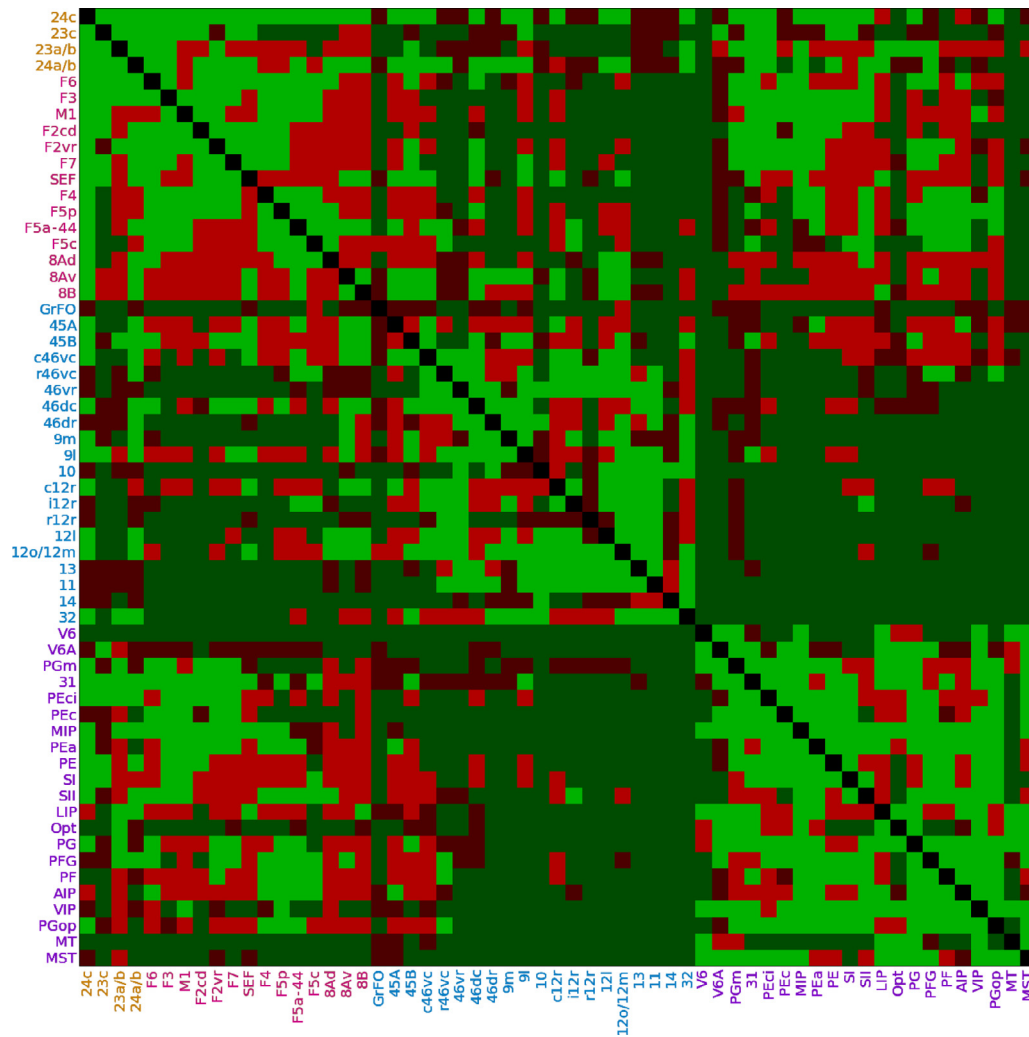
In this study, we have chosen to challenge the performance of fifteen tractography methods with histological data from several laboratories, characterized by the particularly intricate connectivity of a network of cortical areas in the macaque cerebral cortex identified by a combination of anatomical and functional criteria. The results confirm the predictive power of tractography methods, producing more true-positive/negative connections than false-positive/negative connections, with an average sensitivity of 0.61 and a specificity of 0.76 using optimal thresholding (0.56 and 0.78 respectively, using a fixed threshold of 0.001). We have identified a number of tractography errors, false-

positive and false-negative connections, which highlight the need for further validation, as discussed previously by Dyrby et al. (2018), and should be the target of future methodological development.

##### 4.1. The histological approach to brain connections

The data which provided the ground-truth for the present study where almost entirely derived from histological work using the retrograde transport of horseradish peroxidase (HRP), introduced by Kristensson and Olsson (1971), as well as of fluorescent tracers (Bentivoglio et al., 1980). The results of histological studies undoubtedly provide the best ground truth to challenge DW-MRI tractography on real white matter pathways. Despite their enormous success, all the histological tracing methods share a number of limitations or drawbacks, not always recognized. The methods involve injecting the substances directly into the nervous tissue. Since the site and location of the injections are revealed only after sacrificing the animal, the precise region where the substance was taken up by axon terminals can only be determined a posteriori. In the time elapsed between injection and the animal's sacrifice, the injection site has undergone modification and usually presents a core containing a high density of the tracer injected. This is where uptake and transport were presumably optimal. Furthermore, there exists a lighter halo of diffusion where uptake and transport of the substance may have been taken up albeit less effectively. The possibility that some cell bodies or axon terminals may not take up the substance is still unknown for most tracers albeit this possibility was eventually disregarded in the early HRP studies. Moreover, the visualization of the tracer, after it has been transported, depends both on the histochemical (when it applies) reaction used and on the carefulness by which the labeled neurons or axon terminal have been identified. Finally, and most importantly, histological tracer injections rarely covers the entire extent of an area, especially when it is buried in the banks of a sulcus, such as areas of the intraparietal sulcus. This could result in a connectivity pattern not fully revealed, thus producing false-negative results.





**Fig. 6.** DIPY Probabilistic PFT SET tractography algorithm classification performance using the threshold maximizing the Youden's index. The true-positive connections are shown in light green, the true-negatives in dark green, the false-positives in light red, and the false-negatives in dark red. ROIs are grouped as in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Because of the limitations listed above, it is unsafe to assume that all the differences between the results of histological studies and tractography should be ascribed to failures of the latter. Therefore, discrepancies between the two approaches must be carefully evaluated by exploring the limitations of both and the surviving differences might spur further anatomical studies, thus conferring to tractography a predictive power for histological analysis.

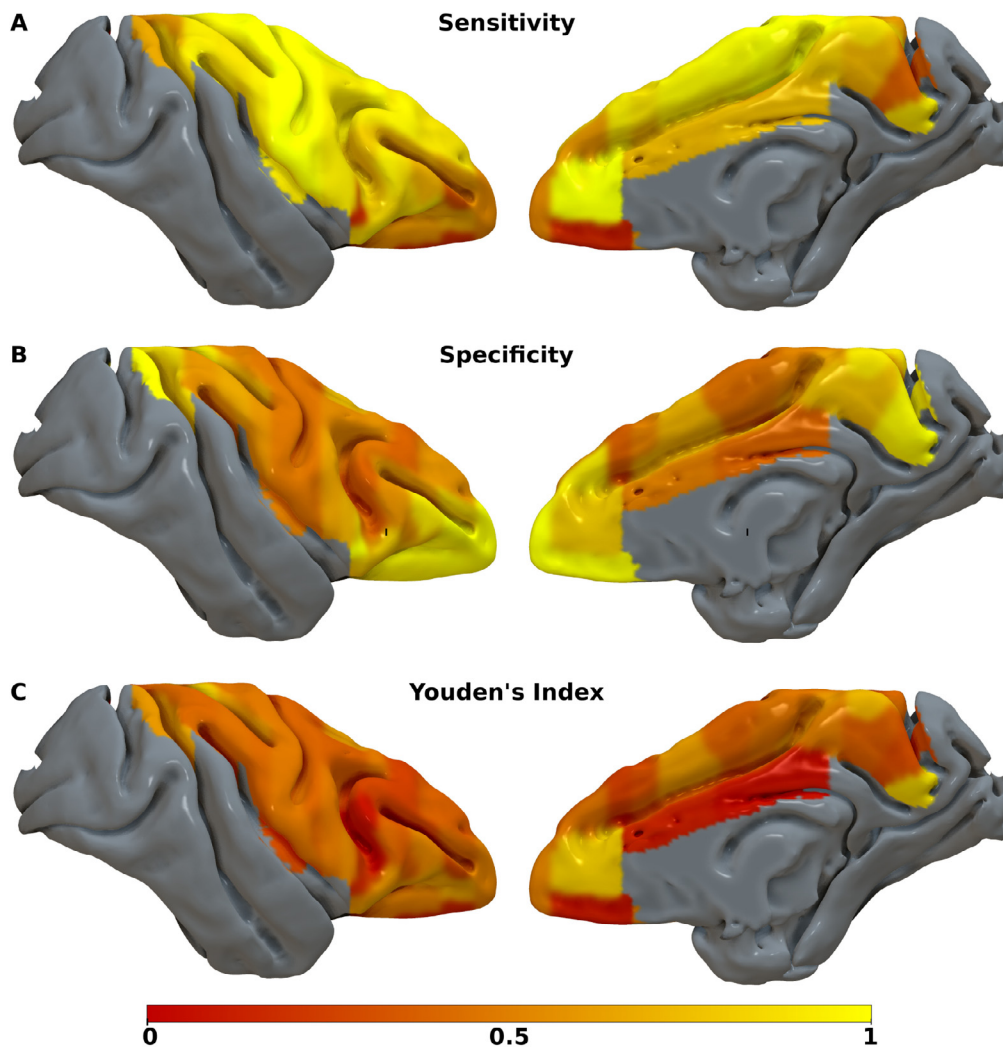
#### 4.2. Accuracy of connectivity with tractography algorithms

This work demonstrates the ability of various tractography algorithms to identify the structural network of the frontal, cingulate and parietal areas of the macaque monkey. Across all algorithms, 76% of the streamlines agreed with the connections found by histological tracers and hence only 24% of the streamlines were forming erroneous connections (percentage of streamlines associated with true-negatives of the histological connectivity matrix shown in Fig. 2). This is lower than the 36% (standard deviation of 17%) of streamlines reported by Maier-Hein et al. (2017), on simulated data. However, we only used the location of the streamline endpoints to identify connections, contrarily as in Maier-Hein et al. (2017), where the whole trajectory was used. Thus, streamlines with valid endpoints but erroneous pathways were considered in agreement with the histological connectivity. We have estimated that up to 7.5% of streamlines have valid endpoints, but have

outlier length compared to other streamlines of the connections. The histological tracing data we used do not have the information to support a complete streamline trajectory validation because it is based on retrograde transport of tracers that labels only the soma of cells whose axons terminate in the injection site. This needs to be further investigated in future work on specific brain connections.

Across all algorithms, the number of FP connections is on average 61% of the number the TP connections with ideal thresholding (31%, 59%, 92% using 0.01, 0.001, 0 fixed thresholds, respectively). This contrast with Maier-Hein et al. (2017), which reported four times the number of FP connections than the number of TP connections. Similarly to the previous results, the FP connections in this study are limited to streamline endpoints rather than the full trajectories, reducing the number of FP we report for each algorithm. Moreover, the histological data used in this work form a network of 59 areas, with 680 connections and 1031 unconnected areas. This network is more densely connected than the simulated data used in Maier-Hein et al. (2017), which reduces potential errors for tractography.

In this study, we focused on the selection of the propagation directions (deterministic vs probabilistic), the masking and seeding strategies, and streamline post-processing methods. For most tractography algorithm, we fixed the local reconstruction to multi-shell, multi-tissue constrained spherical deconvolution (Jeurissen et al., 2014; Tournier et al., 2019), the tractography step sizes, and maximum bending angles.



**Fig. 7.** Spatial distribution of the sensitivity (A), specificity (B) and Youden's index (C) over the 59 ROIs using the threshold maximizing the overall Youden's index of the *DIPY Probabilistic PFT SET* tractography algorithm, shown on the middle cortical layers surface. The structural connectivity of some ROIs is well predicted with tractography with both high specificity and sensitivity (yellow). Tractography underperformed for other ROIs, showing low sensitivity and low specificity (red and orange). See Fig. 1 for anatomical references. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We also used high angular and high spatial resolution DW-MRI data. It would be interesting to specifically assess the effect of multi-shell data on tractography performances, compared to single shell data. Similar to the analysis performed by Ambrosen et al. (2020), it would be interesting to study the effect of reducing the angular and/or spatial resolution of DW-MRI data on the various tractography algorithms performance, on the estimation of the structural network connectivity presented in this work. In future work, it would be of interest to use the tractography evaluation method presented in this work to analyze the effect of specific parameters on tractography performances, such as the parameters used in the fiber orientation distribution estimation (e.g. Canales-Rodríguez et al., 2019; Cheng et al., 2014; Descoteaux et al., 2009; Jeurissen et al., 2014; Tournier et al., 2007; Yeh et al., 2011) or the presence of noise in the DW-MRI images when estimating the fiber orientation distributions (e.g. Manjón et al., 2010; Sotiropoulos et al., 2013; Veraart et al., 2016; Wiest-Daesslé et al., 2008).

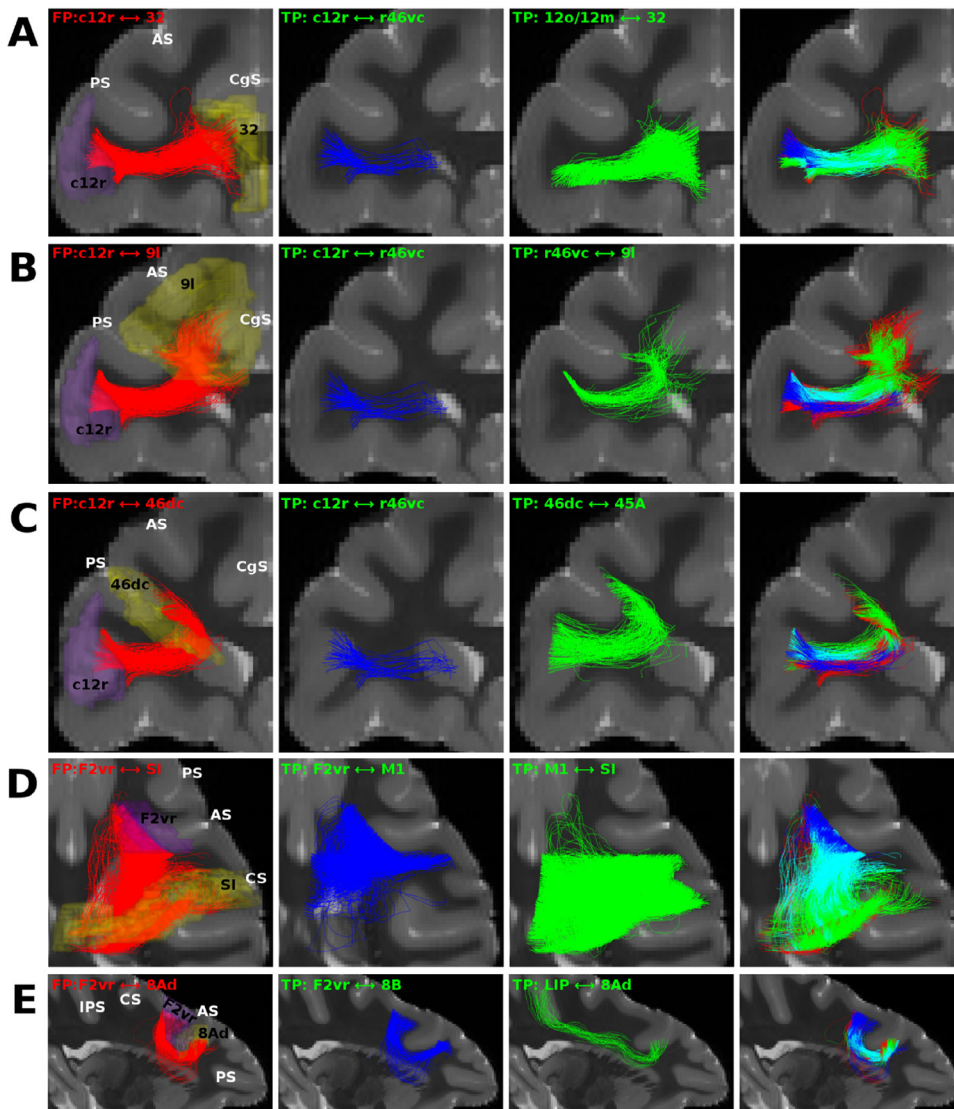
#### 4.2.1. Streamline propagation methods

Probabilistic tractography algorithms (*DIPY Probabilistic* and *MRtrix3 iFOD2* variants) showed the overall highest AUC (Fig. 3), ranging from 0.72 to 0.78. Deterministic tractography algorithms (*DIPY Deterministic*,

*DIPY EuDX*, *MRtrix3 SDSTREAM* and *MRtrix3 Tensor Det*) show the lowest AUC ranging from 0.62 to 0.67. The *MRtrix3 Global* algorithm shows intermediate performances compared to other tested tractography algorithms with an AUC of 0.69. Although it has a high percentage of valid streamlines (89%, see Table 1), the algorithm shows a partial coverage of the connection between area SI (primary somatosensory cortex) and area M1 (primary motor cortex / F1) (Fig. 5i), compared to other algorithms. However, it does not show streamlines with dubious trajectories. This is also the case for *MRtrix3 Tensor Det* and *MRtrix3 SDSTREAM* algorithms. This highlights the trade-off between the generation of spurious streamlines and classification performances when selecting a tractography algorithm.

#### 4.2.2. Connectivity matrix thresholding

Probabilistic tractography algorithms show good sensitivity and specificity performances. This can be observed in Fig. 4 where the Youden's index of probabilistic tractography variants is generally higher than for other algorithms. However, a threshold set to 0 tends to produce low specificity as most pair of ROIs are connected by at least one streamline, especially as the number of streamline increases. This can be observed in Fig. 4C, where the average Youden's index for a thresh-



**Fig. 8.** Large false-positives (first column, red streamlines) and true-positives (second and third columns, blue and green streamlines) in overlapping white matter areas generated using *DIPY Probabilistic PFT SET* tractography algorithm. The last column shows the false-positives with the streamlines points color in blue or green if the voxel is shared with the corresponding true-positive or in cyan for both. Subfigures show streamlines in coronal anterior view (A,B,C), axial superior view (D), and sagittal right view (E), superimposed on a MRI section. See Fig. 1 for anatomical references. Abbreviations: c12r - part of caudal area 12 rostral, 32 - area 32, r46vc - rostral area 46 ventrocaudal, 12o/12m - orbitofrontal and medial area 12, 9l - lateral area 9, 46dc - area 46 dorso-caudal, 45A - area 45A, F2vr - frontal area F2 ventrorostral, SI - primary somatosensory cortex, M1 - primary motor area (F1), 8Ad - area 8A dorsal, 8B - area 8B, LIP - lateral intraparietal area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

old of 0 is 0.29 and increases to 0.35 using a threshold of 0.001, and to 0.37 using the ideal threshold. However, a threshold too high reduces the sensitivity of connectivity estimations. This can be observed using a threshold of 0.01, where the average Youden's index decreases to 0.24. It is not possible to compute the ideal threshold for a network without known ground truth connectivity, but we advise using a threshold value higher than 0, in particular for probabilistic tractography algorithms. For this specific network, connectivity matrix construction and tested tractography algorithms, the ideal threshold maximizing the Youden's index varies from  $2.42e - 3$  to  $2.72e - 5$  (see Table 1).

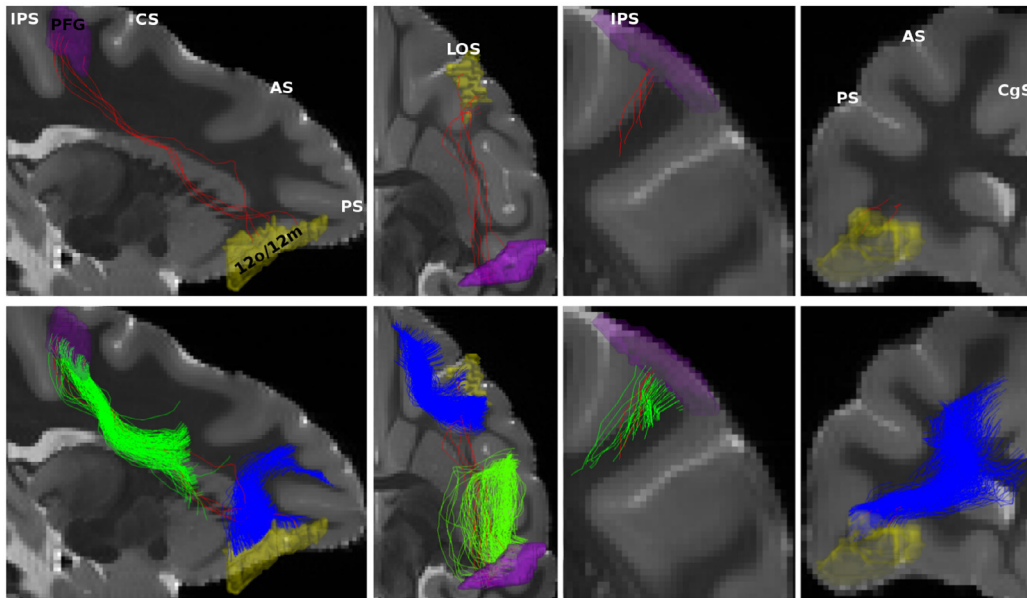
#### 4.2.3. Masking and seeding strategies

The use of partial volume estimate maps for tractography with ACT and PFT shows minor performance increase compared to the same probabilistic algorithms using a conventional masking approach. The AUC scores increased by 0.01 using ACT and 0.03 using PFT. The percentage of valid streamlines increased from 69% to 74% using PFT, but decreased from 76% to 70% using ACT, seeding from the white matter. However, these methods are particularly suited for interface seeding (Girard et al., 2014; Smith et al., 2012) as they used tissue partial volume estimations to improve the streamline propagation. Although seeding from the white matter mask bias the streamlines density in favor of long connections, both seeding methods tested (white matter or in-

terface) showed similar classification performances for both *DIPY Probabilistic PFT* and *MRtrix3 iFOD2 ACT* (see Figs. 3 and 4). The interface seeding strategy increased the percentage of valid streamlines from 74% to 83% for *DIPY Probabilistic PFT gmwmi* and from 70% to 81% for *MRtrix3 iFOD2 ACT gmwmi*. In future work, it would be interesting to further investigate the seeding bias on a brain network with longer pathways, such as fronto-occipital association fibers and commissural fibers.

#### 4.2.4. Streamline post-processing methods

The microstructure-informed tractography method *MRtrix3 iFOD2 ACT SIFT2* show an overall improvement of performances, increasing the AUC from 0.73 to 0.76 and the percentage of valid streamlines from 70% to 73%, compared to *MRtrix3 iFOD2 ACT*. On the other hand, *DIPY Probabilistic PFT COMMIT* decreased the AUC by 0.01 to 0.77 compared to *DIPY Probabilistic PFT*, but increase the percentage of valid streamlines from 74% to 82%. It is important to note that for SIFT2 and COMMIT, the percentage of valid streamlines is computed as the sum of weights associated to valid streamlines over the sum of all weights (all streamlines connecting pairs of ROIs). This suggests that the microstructure models used in SIFT2 and COMMIT penalized more spurious than valid streamlines. The Cluster Confidence Index filtering approach used, *DIPY Probabilistic PFT CCI*, showed a decrease in AUC by 0.02 compared to *DIPY Probabilistic PFT* and an increase in the percentage of



**Fig. 9.** False-negative connection obtained using *DIPY Probabilistic PFT SET* tractography algorithm. All streamlines correctly end in ROIs identified as connected by histological data, but this connection would be rejected using the diffusion threshold maximizing the Youden's index because its connectivity scores is too low ( $7.82e - 4$ ), creating a false-negative. The figure shows the false-negative 12o/12m $\leftrightarrow$ PFG in red, with the ending ROIs in yellow (12o/12m) and purple (PFG). True-positive connections for 12o12m $\leftrightarrow$ 9l and F4 $\leftrightarrow$ PFG are shown in blue and green, respectively. From left to right, the viewpoints are: sagittal right, axial superior, coronal posterior, coronal anterior. Abbreviations: 9l - lateral area 9, 12o/12m - orbitofrontal and medial area 12, F4 - frontal area F4, PFG - inferior parietal area PFG. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

valid streamlines by 1% to 75%. Thus, streamlines with valid endpoints but with trajectories identified as outliers by our parametrization of the Cluster Confidence Index do not greatly affect the matrix weights and consequently the derived binary connectivity estimation of the studied network. Nonetheless, removing outlier trajectories is necessary to improve the volumetric analysis of white matter fascicles (Aydogan et al., 2018; Schilling et al., 2019b). Surface-Enhanced Tractography (*DIPY Probabilistic PFT SET*) showed a decreased AUC (0.75) compared as *DIPY Probabilistic PFT* (0.78) and a decrease in percentage of valid streamlines by 4% to 70%. Although the SET algorithm did not improve the binary connectivity estimation of the studied network, the algorithm has shown improved uniform cortical surface coverage (St-Onge et al., 2018).

#### 4.3. Connectivity estimation for each region of interests

The structural connectivity of each ROI is not equally estimated, as shown on the surfaces of Fig. 7 for *DIPY Probabilistic PFT SET* using the ideal threshold. This results in a varying Youden's index for the structural connectivity of each ROI (mean of 0.41 and standard deviation of 0.17). The reported Youden's index was also variable between sites in Thomas et al. (2014), using the default tractography parameters. This highlights the variability in tractography performances for various white matter structures and strengthen the need for tractography algorithms with parameters that can adapt to the local white matter configurations (Aydogan and Shi, 2016; Dyrby et al., 2018; Girard et al., 2017; Innocenti et al., 2019; Reisert et al., 2014; Rheault et al., 2019; St-Onge et al., 2018; Teillac et al., 2017) or the combined use of multiple tractography algorithms, each targeting the reconstruction of a subset of white matter structures (Daducci et al., 2014; Smith et al., 2015a; Takemura et al., 2016).

#### 4.4. The bottleneck problem on ex-vivo data

We observed the effect of white matter bottleneck configurations (Guevara et al., 2011; Maier-Hein et al., 2017; Mangin et al., 1996) on an ex-vivo macaque brain, where multiple pathways overlap and share

similar local orientations. Tractography algorithms following those local orientations will naturally connect all ROIs sending axons to those overlapping regions. Thus, tractography will likely find existing white matter structures, but also incorrectly merge segments to form non-existing white matter structures. In Fig. 8, most of the volume of the false-positive connections is shared between one or both of the true-positive connections shown. This is in addition to other true-positive connections sharing this white matter area, such as commissural connections. Three large false-positive connections (Fig. 8A,B,C) all share an ending ROI (area c12r) with the true-positive connecting this ROI to area r46vc. While bottleneck configurations can increase the number of false-positives, they can also decrease the true-positives count. Some of the existing histological connections are not found with enough streamlines to survive the thresholding step, creating false-negatives connections. The fraction of streamlines that should have been reconstructed for those connections ended in other true-positive connections (and false-positive connections). Although false-negatives are particularly difficult to visualize, we can appreciate in Fig. 9 that a portion of the false-negatives seems to share white matter volume with dominating true-positive connections. Those pathways are going locally in a similar orientation, making it hard for tractography to consistently follow the minor directions associated with those connections. This highlights the needs for more information, in addition to the orientation, to guide and improve tractography in challenging white matter configurations, such as the intra-voxel compartment diffusivities and volume fractions (Reisert et al., 2014), the index of axon diameter (Barakovic et al., 2018; Girard et al., 2017), axonal dispersion (Rowe et al., 2013) and the myelin water fraction (Schiavi et al., 2019). Additionally, the topographic organization of neurons in the cortex (Patel et al., 2014) could be used to guide the propagation of streamlines in the white matter. This was suggested by Sarubbo et al. (2019) after observing a laminar organization of the white matter pathways in the external capsule, on both tractography and anatomical dissections. Moreover, the use of topographic priors was implemented in tractography by Aydogan and Shi (2016, 2018) to improve streamline regularity and preserve the spatial relationship. Future work should target further validation and development of methods

using such priors, in particular, paired with anatomically-constrained tractography and microstructure-informed tractography methods.

#### 4.5. Limitations of the connectivity analysis

This study focused on binary endpoints connectivity, not on the white matter volume, histological cell counts, or trajectories. Although this limits some aspect of the analysis, at the same time it offers the possibility to use a wide range of histological tracing studies, which did not specifically target the volume assessment of the white matter pathways or reported quantitative labeled cell counts. Moreover, the diffusion data were acquired on a brain not used for the histological tracing, and the histological data were collected from several brains, which cannot account for specimen connectivity variability. This allowed us to study a large network composed of 1711 distinct pairs areas, which would not be possible to study in a single specimen. Nonetheless, studying bi-directional connections, using both retrograde and anterograde tracers, on the same brain as the diffusion data would be of high interest in future studies on tractography validation. The accuracy of the pathways estimated with tractography for connected ROIs cannot be assessed in this experimental setup. Nevertheless, the qualitative results presented show bottleneck configurations using ex-vivo data, similar to what was observed on simulated data (Maier-Hein et al., 2017). This paves the way for future work on ex-vivo quantitative white matter volume estimation using tractography targeted at white matter bottleneck configurations.

## 5. Conclusion

In conclusion, tractography performed on ex-vivo macaque brain shows a good agreement with connections of the frontal, parietal and cingulate areas identified with histological studies based on retrograde transport of tracers. On average, the false-positive and false-negative connections estimated with tractography represent 30% of all predictions, across all algorithms, using a fixed threshold of 0.001. They seem to be related in part to bottleneck configurations, especially near the cortical grey matter. Disentangling the bottlenecks could lead to more sensitive and more specific results. The proposed method enables a quantitative evaluation of tractography algorithm performances and, potentially fosters the development of new ones with improved sensitivity and specificity. Moreover, disagreements between diffusion tractography and histological studies could spark targeted tract-tracing analysis which might lead to a refined ground truth of the cortical connectivity of the frontal, parietal and cingulate areas of the macaque monkey.

### CRedit authorship contribution statement

**Gabriel Girard:** Conceptualization, Methodology, Investigation, Writing - original draft. **Roberto Caminiti:** Conceptualization, Methodology, Investigation, Writing - original draft. **Alexandra Battaglia-Mayer:** Conceptualization, Methodology, Investigation, Writing - original draft. **Etienne St-Onge:** Investigation, Writing - review & editing. **Karen S. Ambrosen:** Resources, Writing - review & editing. **Simon F. Eskildsen:** Resources, Writing - review & editing. **Kristine Krug:** Resources, Writing - review & editing. **Tim B. Dyrby:** Methodology, Resources, Writing - review & editing. **Maxime Descoteaux:** Methodology, Writing - review & editing. **Jean-Philippe Thiran:** Conceptualization, Writing - review & editing, Funding acquisition. **Giorgio M. Innocenti:** Conceptualization, Methodology, Writing - original draft.

### Acknowledgements

This work is supported in part by the Center for Biomedical Imaging (CIBM) of the Geneva-Lausanne Universities and the EPFL, as well as the foundations Leenaards and Louis-Jeantet. Roberto Caminiti acknowledges the support of the *Istituto Italiano di Tecnologia* (contract n.

000000247) and of **MIUR** of Italy (PRIN 2015, grant n. [AWSW2Y\\_002](#)). Alexandra Battaglia-Mayer acknowledges the support of MIUR of Italy (PRIN 2017, grant n. [94KEER\\_002](#)). Kristine Krug is supported by the **Deutsche Forschungsgemeinschaft** (DFG) through a Heisenberg Professorship ([406269671](#)).

## Appendix A. Cortical area subdivisions

The cingulate cortex has been subdivided according to **Brodmann (1909)** into a caudal granular area 23 and a rostral agranular area 24. Both these areas were subdivided into 23a/b and 24a/b, located on the cingulate gyrus, and 23c and 24c, within the cingulate sulcus (**Vogt et al., 1987**). Area 24c, includes 24d (**Matelli et al., 1991; Vogt et al., 2005**), and is coextensive with the rostral cingulate motor area (CMar; **Dum and Strick, 1991**), whereas area 23c largely corresponds to the dorsal (CMAd) and the ventral (CMAv) cingulate motor areas (**Dum and Strick, 1991**), which are also collectively referred to as caudal cingulate motor area (CMAc).

The caudal part of the frontal lobe corresponds to the classical agranular frontal cortex, for which we have adopted the subdivision and nomenclature proposed by **Matelli et al. (1985, 1991)**. Furthermore, we have incorporated data on a distinct oculomotor field, the supplementary eye field (SEF) located in the dorso-rostral part of area F7 (**Schlag and Schlag-Rey, 1987**), the subdivision of area F2 into a ventro-rostral and a dorsal "pre-central dimple" sectors (**Matelli et al., 1998**), and the subdivision of area F5 into three fields (**Belmalih et al., 2009**). The dysgranular area 44 along the fundus of the lateral limb of the arcuate sulcus has been considered together with area F5a, as these two areas share several connective features. The cortex located in the pre-arcuate convexity has been combined with that of the anterior bank of the arcuate sulcus (**Saleem et al., 2014**), where the Frontal Eye Fields (FEF) have been identified (**Stanton et al., 1989**), into a single area labeled as 8A. To distinguish between the dorsal and the ventral part of the FEF, where large-amplitude and small amplitude saccades are, respectively, represented (**Bruce et al., 1985**), area 8A was subdivided into a dorsal and a ventral part.

The subdivision adopted for the lateral prefrontal cortex is very similar to that used by **Saleem et al. (2014)**, which is derived from **Walker (1940)**, **Preuss and Goldman-Rakic (1991)**, and **Petrides and Pandya (1999)**. Based on connectivity studies (**Gerbella et al., 2013; Petrides and Pandya, 1999; Saleem et al., 2014**) both the dorsal (46d) and the ventral (46v) parts of area 46 have been further subdivided into a rostral and a caudal sector, as done for caudal area 46v (46vc). Area 12r was subdivided into caudal, intermediate, and rostral sectors, based on different connectivity profiles (**Borra et al., 2011**). The work by **Carmichael and Price (1994)** guided the delineation of the medial prefrontal and orbitofrontal cortex. However, based on similarities in connectivity patterns, areas 12o and 12m were combined, as done for areas 13l and 13m. Areas 13a/b were considered together with the subdivisions of area 14, while areas 25 and 11m have not been considered, for the paucity of information on their connectivity.

The nomenclature proposed by **Pandya and Seltzer (1982)** has been primarily used for the posterior parietal cortex (PPC). Furthermore, we have also incorporated data from later studies concerning parietal areas V6A and MIP (**Colby and Duhamel, 1991; Galletti et al., 1996; Johnson et al., 1996; Luppino et al., 2005; Marconi et al., 2001**), located in the caudal part of the superior parietal lobule (SPL), VIP along the fundus of the intraparietal sulcus (IPS) (**Colby and Duhamel, 1991; Lewis and Van Essen, 2000**), as well as areas LIP and AIP (**Blatt et al., 1990; Borra et al., 2007; Taira et al., 1990**) in the caudal and rostral parts of the lateral bank of the IPS. There is not yet a comprehensive picture of the connectivity of the different subdivisions of the primary somatosensory cortex of the macaque, especially for areas 3a, 3b, and 1, therefore for this study these have been taken a single area, referred to as SI. This approach was used also for the second somatosensory area (SII). In the









- 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 (46), 16574–16579. doi:10.1073/pnas.1405672111.
- Tournier, J.-D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* 35 (4), 1459–1472.
- Tournier, J.-D., Calamante, F., Connelly, A., 2010. Improved probabilistic streamlines tractography by 2 nd order integration over fibre orientation distributions. In: *Proceedings of the International Symposium on Magnetic Resonance in Medicine*. Stockholm, Sweden.
- Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.H., Connelly, A., 2019. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* 202, 116137. doi:10.1016/j.neuroimage.2019.116137.
- Tournier, J.-D., Mori, S., Leemans, A., 2011. Diffusion tensor imaging and beyond. *Magn. Reson. Med.* 65 (6), 1532–1556.
- Ugurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., de Moortele, P.F.V., Strupp, J., Sapiro, G., Martino, F.D., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Essen, D.C.V., Yacoub, E., Ugurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., de Moortele, P.F.V., Strupp, J., Sapiro, G., Martino, F.D., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., et al., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. *Neuroimage* 80, 80–104. doi:10.1016/j.neuroimage.2013.05.012.
- Urbanski, M., Thiebaut De Schotten, M., Rodrigo, S., Oppenheim, C., Touzé, E., Méder, J.-F., Moreau, K., Loeper-Jeny, C., Dubois, B., Bartolomeo, P., 2011. DTI-MR Tractography of white matter damage in stroke patients with neglect. *Exp. Brain Res.* 208, 491–505. doi:10.1007/s00221-010-2496-8.
- Van Essen, D.C., Jbabdi, S., Sotiropoulos, S.N., Chen, C., Dikranian, K., Coalson, T., Harwell, J., Glasser, M.F., 2014. Mapping connections in humans and non-Human primates: aspirations and challenges for diffusion imaging. *Diffus. MRI* 337–358. doi:10.1016/B978-0-12-396460-1.00016-0.
- Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016. Denoising of diffusion MRI using random matrix theory. *Neuroimage* 142, 394–406. doi:10.1016/j.neuroimage.2016.08.016.
- Vogt, B.A., Pandya, D.N., Rosene, D.L., 1987. Cingulate cortex of the rhesus monkey: I. cytoarchitecture and thalamic afferents. *J. Comp. Neurol.* 262 (2), 256–270. doi:10.1002/cne.902620207.
- Vogt, B.A., Vogt, L., Farber, N.B., Bush, G., 2005. Architecture and neurocytology of monkey cingulate gyrus. *J. Compar. Neurol.* 485 (3), 218–239. doi:10.1002/cne.20512.
- Walker, A.E., 1940. A cytoarchitectural study of the prefrontal area of the macaque monkey. *J. Comp. Neurol.* 73 (1), 59–86. doi:10.1002/cne.900730106.
- Wang, J., Aydogan, D.B., Varma, R., Toga, A.W., Shi, Y., 2018. Modeling topographic regularity in structural brain connectivity with application to tractogram filtering. *Neuroimage* 183, 87–98. doi:10.1016/j.neuroimage.2018.07.068.
- Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S.P., Barillot, C., 2008. Rician Noise Removal by Non-Local Means Filtering for Low Signal-to-Noise Ratio MRI: Applications to DT-MRI, pp. 171–179. doi:10.1007/978-3-540-85990-1\_21.
- Yeh, C.H., Smith, R.E., Liang, X., Calamante, F., Connelly, A., 2016. Correction for diffusion MRI fibre tracking biases: the consequences for structural connectomic metrics. *Neuroimage* 142, 150–162. doi:10.1016/j.neuroimage.2016.05.047.
- Yeh, F.C., Wedeen, V.J., Tseng, W.Y.I., 2011. Estimation of fiber orientation and spin density distribution by diffusion deconvolution. *Neuroimage* 55 (3), 1054–1062. doi:10.1016/j.neuroimage.2010.11.087.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer* 3 (1), 32–35. doi:10.1002/1097-0142(1950)3:1 <32::AID-CNCR2820030106 >3.0.CO;2-3.
- Yu, C.S., Li, K.C., Xuan, Y., Ji, X.M., Qin, W., 2005. Diffusion tensor tractography in patients with cerebral tumors: a helpful technique for neurosurgical planning and postoperative assessment. *Eur. J. Radiol.* 56 (2), 197–204. doi:10.1016/J.EJRAD.2005.04.010.
- Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G., 2006. User-Guided {3D} active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31 (3), 1116–1128.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20 (1), 45–57. doi:10.1109/42.906424.