SCIENTIFIC REPORTS

OPEN

Multidimensional encoding of brain connectomes

Cesar F. Caiafa [1,8,9] & Franco Pestilli [1,2,3,4,5,6,7]

The ability to map brain networks in living individuals is fundamental in efforts to chart the relation between human behavior, health and disease. Advances in network neuroscience may benefit from developing new frameworks for mapping brain connectomes. We present a framework to encode structural brain connectomes and diffusion-weighted magnetic resonance (dMRI) data using multidimensional arrays. The framework integrates the relation between connectome nodes, edges, white matter fascicles and diffusion data. We demonstrate the utility of the framework for *in vivo* white matter mapping and anatomical computing by evaluating 1,490 connectomes, thirteen tractography methods, and three data sets. The framework dramatically reduces storage requirements for connectome evaluation methods, with up to 40x compression factors. Evaluation of multiple, diverse datasets demonstrates the importance of spatial resolution in dMRI. We measured large increases in connectome resolution as function of data spatial resolution (up to 52%). Moreover, we demonstrate that the framework allows performing anatomical manipulations on white matter tracts for statistical inference and to study the white matter geometrical organization. Finally, we provide open-source software implementing the method and data to reproduce the results.

A fundamental goal of neuroscience is to develop methods to understand how brain networks support function and behavior in individuals across human populations[1–4]. The recent increase in availability of neuroimaging data and large scale projects has the potential to empower new ways of discovery by studying large populations of human brains[5–23]. Exploiting these large-scale data sets will require convergent efforts in advancing measurement methods, data representation frameworks, as well as computational algorithms and theory[24, 25].

Recent advances in measurement methods and computational algorithms are shifting the study of the white matter and brain networks beyond qualitative characterization (such as camera lucida drawings), toward structural and functional quantification[26–31]. Tractography and diffusion-weighted magnetic resonance imaging (dMRI) are the primary methods for mapping structural brain networks and white matter tissue properties in living human brains. Using these methods we have learned much about the macrostructural organization of the human brain, such that network neuroscience has become one of the fastest-growing scientific fields[3, 27, 28, 30, 32–40].

Tractography algorithms use dMRI data to estimate the three-dimensional trajectory of neuronal axons bundles wrapped by myelin sheaths–the white matter fascicles. Fascicles are normally represented as sets of brain coordinates, with coordinates segments spanning anything between 0.01 to 1 *mm* in length (Fig. 1a **top**). Fascicles have historically been clustered into anatomically cohesive groups called white matter tracts. The largest of these tracts have associated names–such as the corticospinal tract (CST) and the arcuate fasciculus (Fig. 1b **top**[41, 42]). White matter tracts communicate between cytoarchitectonically and functionally distinct areas–such as Broca's or Wernicke's areas involved in human language processing (Fig. 1c **top**[43–45]). White matter tracts and brain areas together compose a large-scale network called the connectome[46]. Within this network, white-matter tracts represent communication pathways (the edges; Fig. 1b **top**) and brain areas units of information processing (the nodes; Fig. 1c **top**).

We propose a connectome encoding framework that integrates models of white matter fascicles anatomy, microstructural tissue properties as well as the dMRI measurements. The framework encodes altogether

[1]Department of Psychological and Brain Sciences Indiana University Bloomington, IN, 47405, USA. [2]Department of Intelligent Systems Engineering Indiana University Bloomington, IN, 47405, USA. [3]Department of Computer Science Indiana University Bloomington, IN, 47405, USA. [4]Program in Neuroscience Indiana University Bloomington, IN, 47405, USA. [5]Program in Cognitive Science Indiana University Bloomington, IN, 47405, USA. [6]School of Optometry Indiana University Bloomington, IN, 47405, USA. [7]Indiana Network Science Institute Indiana University Bloomington, IN, 47405, USA. [8]Instituto Argentino de Radioastronomía (IAR), CONICET CCT - La Plata Villa Elisa, 1894, Argentina. [9]Facultad de Ingeniería - Departamento de Computación - UBA Buenos Aires, C1063ACV, Argentina. Correspondence and requests for materials should be addressed to F.P. (email: franpest@indiana.edu)
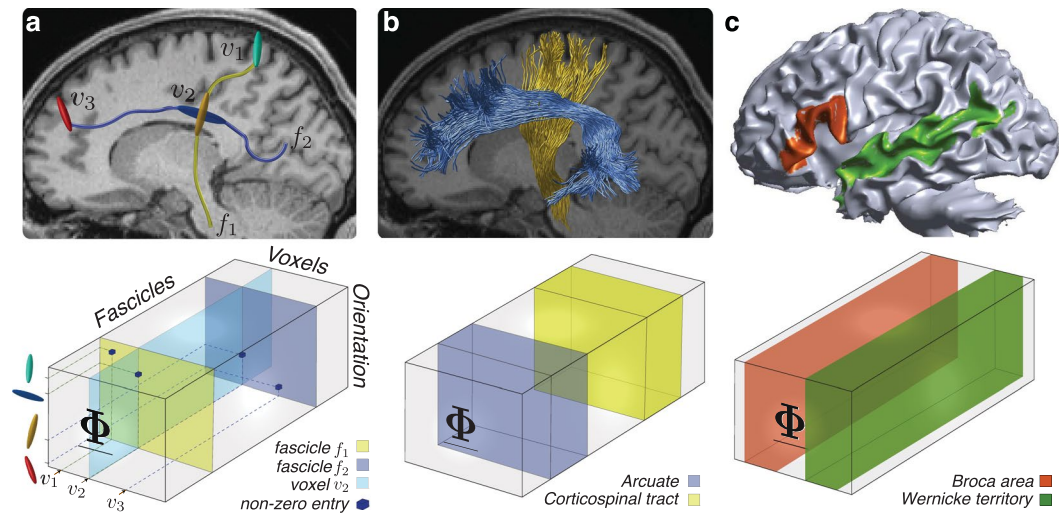
Figure 1. Connectome encoding using multidimensional arrays. (**a**) **Top**. Two white matter fascicles (f_1 and f_2) and three voxels (v_1, v_2 and v_3). **Bottom**. Tensor encoding of fascicles' spatial and geometrical properties. Yellow f_1, dark blue f_2, cyan v_2. Non-zero entries in $\underline{\Phi}$ indicate fascicles orientation (1st mode), position (voxel, 2nd mode) and identity (3rd mode). (**b**) **Top**. Two major human white matter tracts (connectome edges). The corticospinal tract and Arcuate fasciculus. **Bottom**. Tensor encoding of connectome edges. The corticospinal tract and Arcuate fasciculus are encoded as collections of frontal slices–blue and yellow subtensors. (**c**) **Top**. Two human cortical areas (connectome nodes). Wernicke's territory and Broca's area. **Bottom**. Tensor encoding of connectome nodes. We show examples of a large temporal area comprising also Wernicke's territory and Broca's area encoded as collections of lateral slices–red and green subtensors (areas defined using Freesurfer[43–45, 47]).

connectome edges, nodes as well as the associated dMRI data using multidimensional arrays–also called tensors[48–51]. Below, we introduce the framework and show four applications. First, we use the framework to implement efficiently methods for connectome evaluation. Second, we use the framework to perform a large scale tractography evaluation (13 tracking algorithms, 1,490 brain connectomes, three different data sources[52–56]). Finally, we present two additional applications by describing how the framework can be used to perform efficiently statistical inferences on brain connections and white matter tracts using the recently introduced virtual lesion method[53, 57] and to chart the reliability and reproducibility in the estimates of the geometrical organization of the human white matter[58–60]. We provide open source software[61, 62] implementing the encoding framework at http://www.github.com/brain-life/encode and data to reproduce the analyses at doi:10.5967/K81J97NM.

Results

We present a method to encode the anatomical properties of connectome edges and nodes into multidimensional arrays[48] (see Supplementary Methods, section 1.1). The encoding scheme maps fascicles into the three dimensions of a sparse array $\underline{\Phi}$ (Fig. 1a **bottom**). The first dimension of $\underline{\Phi}$ (1st mode) encodes fascicles orientation along their trajectory. Where single nodes in a fascicle are encoded as non-zero entries of the sparse array (see dark-blue cubes in Fig. 1a **bottom**), and full fascicles as complete frontal slices (yellow and blue in Fig. 1a **bottom**). The second dimension (2nd mode) encodes spatial position within the brain, that is voxels. Slices in this dimension represent single voxels (see cyan slice in Fig. 1a **bottom**). The third dimension (3rd mode) encodes fascicles, or better the indices of each fascicle within the connectome. We show that connectome edges (a white matter tract) is an ensemble of fascicles that can be represented by a set of frontal slices in (Fig. 1b **bottom**). Once these slices are reorganized (permuted) they come together to represent a white matter tract. Blue and yellow frontal volumes (technically called sub-tensors) in Fig. 1b **bottom** correspond to the encoded representation of the Arcuate Fasciculus and Corticospinal Tract, reproduced in Fig. 1b **top** in their natural brain space. Also, connectome nodes (an ensemble of voxels) are encoded in $\underline{\Phi}$. For example, Fig. 1c **bottom** shows the lateral sub-volumes encoding the voxels for Broca's area (red) and Wernicke's territory (green; regions also reproduced in their natural brain space in Fig. 1c **top**.)

Multidimensional encoding of connectomes provides a variety of computational opportunities. This is because direct array operations can be applied globally to connectomes. For example, fascicle search, area to area mapping, charting brain connections or fascicles crossing angles become trivial operations, such as finding indices in the array $\underline{\Phi}$. Below, we demonstrate four applications involving such operations. Section 2.3.1 of the Supplementary Material describes in more detail advantages and disadvantages of the encoding method.

First application: Efficient connectome evaluation. It has been recognized that estimates of brain connectomes can differ substantially depending on the tracking method and data type[53, 59, 60, 63]. Such differences motivated measuring accuracy for brain connectomes in individual brains in order to identify the best fitting connectome model before further studying its properties[53, 63].
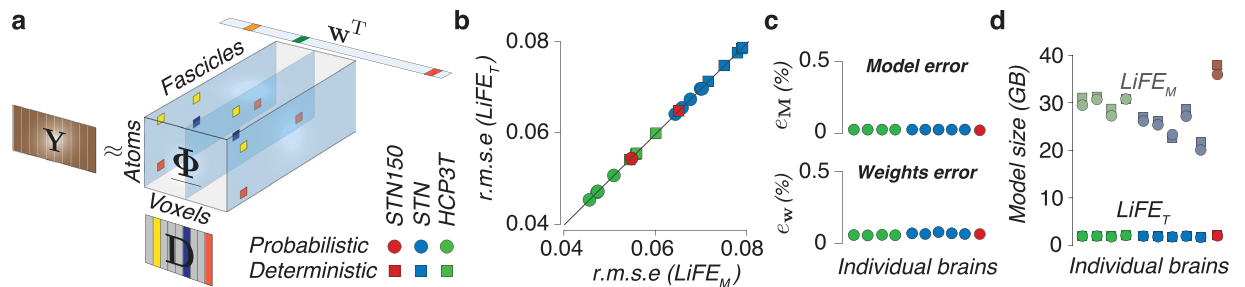
Figure 2. Tensor decomposition of the Linear Fascicle Evaluation method. (**a**) The tensor decomposition model, LiFE$_T$ (see Supplementary Section 2.1 for details). LiFE$_T$ uses a dictionary (matrix **D**) of precomputed diffusion predictions in combination with the sparse tensor, $\underline{\Phi}$, and a vector of fascicles weights (**w**) to model the measured dMRI signal (matrix **Y**). (**b**) **Comparison of the error in predicting diffusion**. Scatter plot of the global r.m.s error ($\bar{e}_{rms}$; equation (8), **Methods**) in predicting diffusion measurements for LiFE$_M$[53] and LiFE$_T$ in ten brains, three dataset (HCP3T, STN and STN150) and two tracking methods (tensor-based deterministic and probabilistic tractography). The r.m.s is virtually identical. (**c**) **top** LiFE$_T$ error in approximating the LiFE$_M$ matrix (e_M; Equation (9); **Methods**) computed for ten brains (HCP3T, STN and STN150 datasets, probabilistic tractography, $L_{max} = 10$). **Bottom**. Error (e_w; Equation (10); **Methods**) of LiFE$_T$ in recovering the fascicle contributions (vector **w**) assigned by LiFE$_M$. ($N = 10$, probabilistic tractography $L_{max} = 10$) (**d**) **Model compression**. Measured size of LiFE$_M$ (matrix **M**) and the decomposed model, LiFE$_T$, (tensor $\underline{\Phi}$ and matrix **D**; $N = 20$). Matrices and tensors all stored using double floating-point precision avoiding zero entries[66, 67].

A few methods to evaluate connectomes and compute errors have been proposed recently[53, 64, 65]. One of these methods, the Linear Fascicle Evaluation algorithm, or LiFE[53], computes the error of a connectome in predicting the diffusion signal. LiFE takes as input the set of white-matter fascicles generated using tractography and returns as output the subset of fascicles that predict the dMRI measurements with smallest error (see[53] and **Methods**). LiFE predicts diffusion measurements (vector **y**, Equation 4) in individual brains by combining the diffusion prediction from individual fascicles in a connectome (columns of matrix **M**, Equation 4) as described in Supplementary Fig. 2c. The LiFE model is fit to the data by assigning weights to the fascicles in the connectome (entries in vector **w**, Equation 4) via a non-negative least-squares method. We show that the LiFE model based on matrix **M** (hereafter referred to as LiFE$_M$), can be accurately approximated using tensor decomposition and the framework introduced in Fig. 1 (see Supplementary Results). Hereafter, we refer to the LiFE model represented by multidimensional arrays as LiFE$_T$.

Figure 2a depicts LiFE$_T$, where the diffusion measurement (matrix **Y**, Equation (20) in Supplementary material) is factorized into: (1) a dictionary matrix **D** in which each atom (column) represents the precomputed diffusion prediction for a specific fascicle orientation, evaluated at all gradient directions (θ, see Equation (17) in Supplementary material), (2) the sparse array $\underline{\Phi}$ (Fig. 1a **bottom**.) and (3) a vector of fascicle weights **w**. Supplementary Results, Section 2.1 provides additional details on the decomposition method.

We measured the accuracy of LiFE$_T$ in approximating LiFE$_M$ using three publicly available data sets: STN, STN150 and HCP3T[53–56, 68]. To do so, we built connectomes in ten individual brains using both, probabilistic[69, 70] (CSD, $L_{max} = 10$ and deterministic[71, 72] tractography, see **Methods**). We report three main results showing that given a sufficient number of dictionary atoms ($L > 360$ in **D**; Supplementary Fig. 2d): (1) the global r.m.s. error (Equation 8) in predicting diffusion is virtually identical between LiFE$_M$ and LiFE$_T$ (Fig. 2b). (2) LIFE$_T$ approximates the LiFE$_M$ matrix (**M**) accurately. Specifically, the Frobenius norm-based relative error, e_M, is less than 0.1% (Fig. 2c **top**; **Methods**, Equation 9). (3) The fascicles weights assigned by LiFE$_M$ and LIFE$_T$ are virtually identical (Fig. 2c **bottom**, <0.1%). The relative error between weights estimated by LiFE$_M$ and LiFE$_T$, was computed using the l_2-norm (**Methods**, Equation 10). We also show that by increasing decomposition resolution (L) the difference in r.m.s., as well as e_M and e_w decrease (Supplementary Fig. 2g,h and i).

Importantly, LiFE$_T$ requires a fraction of the memory used by LiFE$_M$. To show this, we measured the size of the computer memory used by matrix **M** in the LiFE$_M$ model (**Methods**, Equation 4) and compared that to the total memory used by arrays $\underline{\Phi}$ and **D** in the LiFE$_T$ model (Equation (20) in Supplementary material). Figure 2d shows measurements in gigabytes for 20 connectomes (500,000 fascicles each, two tracking methods) in ten subjects from the three data sets. Whereas LiFE$_M$ can require up to 40 GB per connectome, the decomposed model LiFE$_T$ requires less than 1 GB, a 40x compression factor. All calculations were performed using double precision floating point and sparse data format[66, 67]. See Supplementary Fig. 2 and Supplementary Results, Section 2.6 for details on the effect of the number of gradient directions (N_θ) and connectome fascicles (N_f) on memory consumption.

Second application: Large-scale analysis of quality and reproducibility of tractography.
The availability of multiple tracking methods and data types can be both an opportunity or a burden for investigators interested in using them as biomarkers for health and disease[7, 13, 14, 16, 23, 73, 74]. In an ideal world, a single tracking method or data type would supersede all others. In practice, a single algorithm or data type superior to all the rest has not been identified. Yet multiple algorithms or data can help depending on study goals and available measurements infrastructure. For example, when measuring patient populations or in developmental or ageing studies it might be necessary to measure at lower resolution given time constraints. In principle, higher directional and

spatial resolution should be preferred to lower resolution one. Yet, to date we do not have computations to relate data quality and resolution or tractography quality and flexibility to what it is possible to map of the human connectome.

We used $LiFE_T$ to perform a large-scale evaluation of the reproducibility of connectome estimates in individual brains to identify the degree to which estimates depend on data quality and tractography. To do so, we generated a total of 1,490 connectomes using thirteen different combinations of tracking methods and parameters on data from twelve individual brains and three sources. Specifically, we used data from (a) HCP3T (4 subjects, 1.25 mm isotropic spatial resolution, 90 diffusion directions[22], (b) HCP7T (5 subjects, 1.05 mm isotropic spatial resolution, 60 diffusion directions[56] and (c) STN (4 subjects, 1.5 mm isotropic spatial resolution, 96 diffusion directions[53].

To test the quality and reproducibility of connectome estimates we generated ten connectomes for each individual brain and tracking method. We used both, probabilistic and deterministic tracking, based on either constrained spherical deconvolution (CSD) or the tensor model[70, 72] and generated 500,000 candidate fascicles. We also varied tracking parameters by estimating fiber orientation distribution functions using a range of CSD parameter values ($L_{max} = 2, 4, 6, 8, 10, 12$). Each one of these 1,490 candidate connectomes was then processed using $LiFE_T$. $LiFE_T$ identified optimized connectomes, that is, the subset of fascicles with non-zero weight[53] and computed connectomes error in predicting the diffusion signal (r.m.s., Equation 7). We used this large set of statistically validated, repeated-measures connectomes to test the reproducibility of connectome estimates in individual subjects, as function of tracking method and data type (spatial resolution, signal-to-noise ratio (SNR), and number of diffusion directions).

We assessed quality using multiple measures. Connectome quality can be assessed in several ways. For example, the error of the connectome in predicting the diffusion signal can be measured to establish connectome quality[53, 64, 65]. In addition, connectome resolution, the number of fascicles supported by the data can also inform about connectome quality. Finally, the accuracy of the connectome fascicles can be estimated qualitatively by comparing the anatomical variability of known major white matter tracts estimated from the connectomes using atlases[42]. We established the reproducibility of these three measures across repeated connectome estimates within individual brains and across tracking methods, parameters and data types.

Figure 3a plots mean optimized connectome error and number of found fascicles (± 5 standard error of the mean, s.e.m) for the three datasets: STN, HCP3T and HCP7T (1,490 connectomes). The plot shows a series of informative findings. First, data sets naturally cluster into groups, an effect mostly driven by the connectome error, the abscissa. Second, individual brains are nearly separable (along diagonals) both within and between datasets, such separation is largely independent of tracking method or parameters. Third, the number of found fascicles (connectome resolution) increases with the number of CSD parameters (L_{max}), this is true in each data set, for both deterministic and probabilistic tracking but the effect is accentuated with deterministic methods (Fig. 3a inset). Fourth, connectome resolution and error are both extremely reliable. $LiFE_T$ returns an almost identical number of found fascicles and connectome error across repeated tracking for a given set of parameter and tracking method (error bars are very small compared to the mean values). Fifth, probabilistic methods consistently show lower error in fitting the dMRI data and higher number of fascicles than deterministic models, this confirms previously reported results[53].

Our results show that increasing dMRI data spatial resolution increases connectome resolution, despite differences in number of measured diffusion directions. To evaluate the impact of spatial resolution on the number of fascicles supported by the data, we first compared the number of fascicles assigned a non-zero weight by $LiFE_T$ in a single subject between the 1.25 mm^2 and 1.05 mm^3 resolution (HCP3T and HCP7T respectively; blue and orange color in Fig. 3a). Results show a 46% (± 5% s.e.m.) increase in number of fascicles with the higher resolution data set. An even larger increase in connectome resolution was measured across all subjects by comparing connectome resolution blocked by data resolution and averaged either across probabilistic (52% $\pm$ 3% s.e.m. across L_{max}) or deterministic (50% $\pm$ 6% s.e.m. across L_{max}) models. We computed this average by comparing only models common across data sets (i.e., L_{max} 2, 4, 6 and 8). Such 52% increase is well supported by the 68.7% increase in data volumetric resolution, and it is measured despite the decrease in number of measured diffusion directions in the higher resolution data (HCP7T: 60 directions, HCP3T: 90 directions). This demonstrates a profound impact of spatial resolution in mapping brain connectomes that goes beyond improvements due to directional resolution[56].

We further performed a qualitative evaluation of the degree to which connectomes generated using different tracking methods and optimized with $LiFE_T$ show reliable anatomical features. To do so, we segmented twenty major human white matter tracts using standard methods and atlases[42, 75]. Figure 3b shows two examples of repeated tracts identified in one subject (HCP3T), using probabilistic (top) and deterministic (bottom) tracking. Results show high degree of anatomical similarity for tracts in $LiFE_T$ optimized connectomes when using a single tracking method–compare left and right in the top or bottom panels. Conversely, results show anatomical differences within a single individual across tracking parameters–the $LiFE_T$ optimization cannot change this result–compare top and bottom tracts. This reproduces previous results[53]. Figure 3c shows similar results for a different subject in the HCP3T data set. Importantly, by comparing two different subjects in Fig. 3b and Fig. 3c it is clearly possible to discriminate between brains based on the anatomical features of the connectomes. Supplementary Fig. 3b shows additional examples of major tracts anatomy estimated in individual subjects using repeated connectome measures. These plots allow to appreciate the degree of anatomical similarity within subjects given a single tracking method. Supplementary Fig. 3c shows multiple examples of major tracts anatomy estimated in individual subjects using different tracking methods and parameter sets. These plots also allow to appreciate the anatomical variability that different tracking methods introduced even within the same subject and data set by using different number of parameters for tracking.
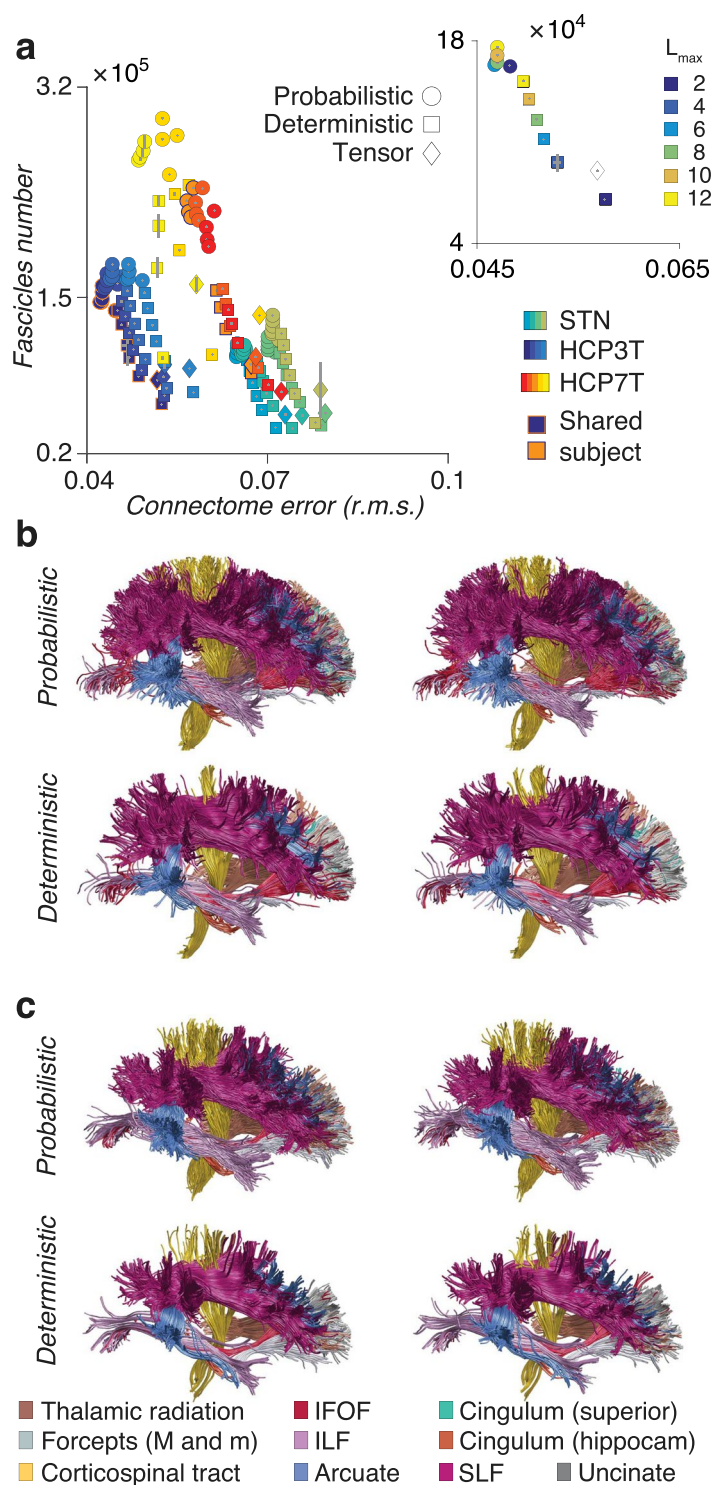
Figure 3. Connectome resolution and anatomical reliability as function of data and method. (**a**) Scatter plot of number of found fascicles and global r.m.s error in LiFE$_T$ optimized connectomes (mean ± 5 standard error of the mean, s.e.m., $N = 1{,}490$, $n = 12$ subjects, $m = 10$ repeated tracking, using either 13 or 9 different L_{max} values for either STN, HCP3T or HCP7T). Inset shows the relation between the number of found fascicles (ordinate) and r.m.s. error (abscissa) and L_{max} (color) in one subject from the HCP3T dataset. (**b**) Reproducibility of connectome anatomy. Twenty major human white matter tracts, two repeated estimates in a single subject probabilistic (top) and deterministic (bottom) tracking, HCP3T dataset. Tracts anatomy is very similar between repeated estimates when using a single tracking method (compare between columns, top and bottom). Estimated tracts anatomy differs within a single subject when the different tracking methods are used (compare between rows, left or right). (**c**) A different subject from the HCP3T dataset.
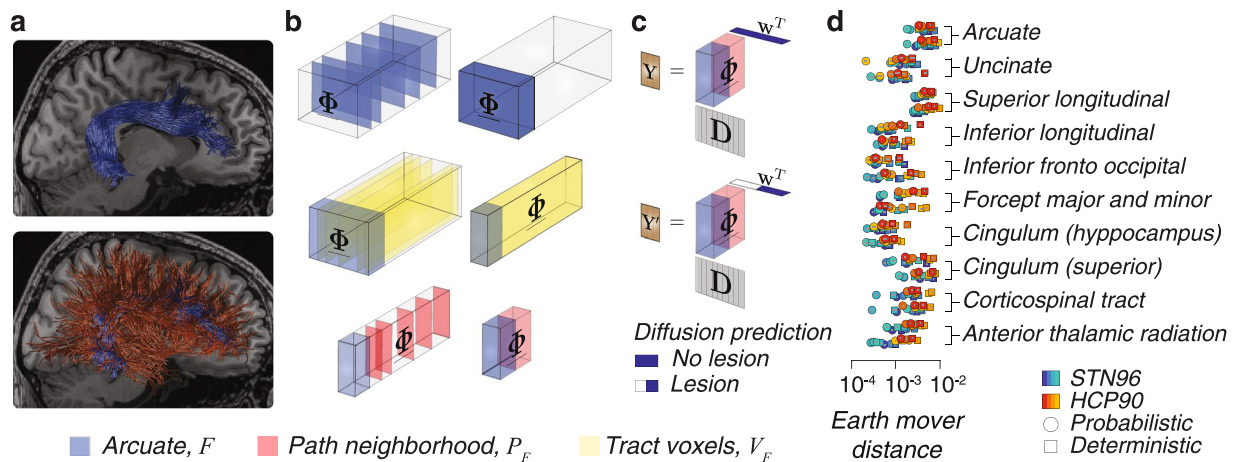
Figure 4. Virtual lesion of white matter tracts using the tensor encoding framework. (**a**) Anatomical representation of the arcuate fasciculus and its path-neighborhood, blue and red respectively. (**b**) Identification of the arcuate fasciculus and its path-neighborhood. **Top**. Arcuate fascicles encoded as frontal slices collated by a permutation (*F*, blue). **Middle**. Ensemble of all voxels touched by the arcuate (lateral tensor slices, yellow) collated by a permutation. **Bottom**. The path-neighborhood (*P_F*, red) contained in the non-empty frontal slices of *V_F*. (**c**) The virtual lesion using the encoding framework. **Top**. Diffusion prediction (**Y**) in the arcuate voxels by the arcuate and its path-neighborhood. **Bottom**. Diffusion prediction (**Y′**) associated to *P_F*, (arcuate fasciculus weights are set to zero, white). (**d**) Statistical evidence for twenty human major white matter tracts[42] established using the sparse tensor encoding framework. Error bars show ±1 s.e.m.

Third application: Statistical inference on white matter tracts. The concept of virtual lesion has been utilized in several contexts[57, 76–79]. More recently, virtual lesions have been used to compute statistical evidence for white matter tracts by measuring the impact of removing entire sets of fascicles from individual whole-brain connectomes[53].

The LiFE method requires fascicles in an optimized connectome to contribute to the diffusion prediction by assigning non-zero weights to successful fascicles. Because of this, lesioning fascicles from the model (by setting their weights to zero) increases the prediction error, r.m.s. More specifically, if a set of fascicles, *F*, passes through the set of voxels *V_F*, their path-neighborhood, *P_F*, is defined as all fascicles passing through *V_F* excluded *F*. The full signal prediction in *V_F* depends on *F* ∪ *P_F*. The lesioned model instead, predicts the signal in *V_F* only using *P_F*. The two models of the signal in *V_F*, the lesioned (*P_F*) and unlesioned (*F* ∪ *P_F*) model generate two distributions of r.m.s. error among voxels in *V_F*. These two distributions can be compared using various measures to establish the statistical evidence for given the data[53].

To date, the virtual lesion method has been employed to establish the statistical evidence for brain tracts and connections[37, 39, 63, 80]. The operations necessary to perform virtual lesions using data represented directly in the brain natural anatomical space require multiple mappings between fascicles coordinates, voxel indices and the corresponding entries in the LiFE model (matrix **M** columns and associated weights). The computational complexity of these operations becomes trivial after encoding connectomes in the multidimensional framework. We show a visualization of the virtual lesion of the right arcuate fasciculus in a single individual (Fig. 4a,b). Given the arcuate fasciculus, *F* (Fig. 4a,b, blue), the identification of *V_F* and *P_F*, can be achieved in a computationally efficient way using the encoding framework. *V_F* is the set of lateral slices with non-zero entries within the subtensor identified by *F* (Fig. 4b, yellow) and *P_F* is the set of fascicles (frontal slices) not in *F* but touching *V_F* (Fig. 4b, red). Computing the signal prediction with and without lesion is then reduced to evaluate the sparse tensor decomposition and consider the tract weights zero (with lesion) or non-zero (without lesion), as shown in Fig. 4c.

Figure 4d and Supplementary Fig. 4 shows the statistical strength of evidence for twenty major human white matter computed with 19,200 virtual lesions (in all connectomes in Fig. 3) measured as the earth mover distance[53, 81] and strength of evidence[53]. These results are important because they reproduce previous findings[53] and show large scale reliability of the *in vivo* statistical evidence of major human white matter tracts validated *post mortem*[41, 42].

Fourth application: Estimates of white matter geometrical organization. Clarifying the geometrical organization of the brain white matter is emerging as an important opportunity given recent improvements in both, measurement and mapping methods[30, 31, 58, 82–84]. Hereafter, we utilize the encoding framework and 160 statistically validated connectomes to quantify the distribution of angles between white matter fascicles associated with pairs of white matter tracts or between tracts and their path-neighborhood[58–60].

The corticospinal tract (CST), arcuate fasciculus (Arc) and superior lateral fasciculus (SLF) were segmented in the right and left hemispheres of 160 connectomes estimated using either probabilistic or deterministic tractography in eight brains (STN *n* = 4; HCP3T *n* = 4, *L_max* = 10, ten repeated tracking per brain) and standard atlases[42, 75]. Angles between pairs of fascicles within a voxel were estimated by operating on the connectome encoding framework (Fig. 5a–d). We performed three experiments to establish the dependence of fascicle angles
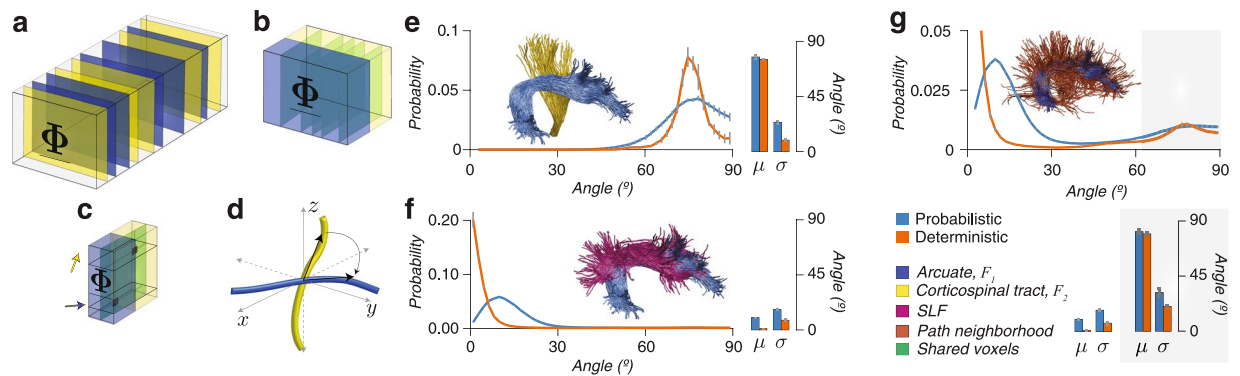
Figure 5. Quantifying variability of estimates for angles of incidence between fascicles in the human white matter. (**a**) Arcuate (Arc, blue) and corticospinal tract (CST, yellow) fascicles identified in frontal slices of $\underline{\mathbf{\Phi}}$. (**b**) Voxels shared between Arc and CST located by finding lateral slices in $\underline{\mathbf{\Phi}}$ (green) with non-zero entries in the yellow and blue subtensors. (**c**) Measurement of the angle of incidence in the voxels shared by Arc and CST (green). Angles are determined by finding the indices in the first dimension of $\underline{\mathbf{\Phi}}$ (1st mode). (**d**) Depiction of angles being computed in brain space. (**e**) Distribution of crossing angles between Arc and CST. (**f**) Distribution of angles incidence between Arc and SLF. (**g**) Distribution of crossing angles between Arc and its neighborhood. Angles computed on Probabilistic (blue) and Deterministic (orange) connectomes ($L_{max} = 10$, STN and HCP3T). Analyses based only on fascicles with positive weight. Histograms show mean across subjects ($n = 8$). Bar plots show peak angle (μ) and width-at-half height (σ). Error bars ± 1 standard error of the mean, s.e.m, across subjects ($n = 8$).

on the tracking method and measured the distribution of angles between fascicles in tracts and neighborhoods. We measured: (a) Crossing angles between fascicles in the Arc and CST at voxels of overlap between the tracts. These fascicles were expected to cross with non-zero degree angle. (b) Angles between fascicles in the Arc and SFL. These fascicles were expected to bypass each other with expected angle near zero degrees. (c) Angles between the Arc and its path-neighborhood. The expected angle of crossing between tracts and path neighborhoods has generated important debates[58–60, 82].

We performed three experiments to measure the dependence of angles between white matter fascicles as function of different tracking methods. In the first experiment, we computed pairwise angles between fascicles associated with either of two tracts, F_1 and F_2, the Arc and CST respectively. We began by identifying the fascicles associated with tracts using the frontal slices of $\underline{\mathbf{\Phi}}$ (3$_{\text{rd}}$ mode; Fig. 5a). F_1 and F_2 identify two subtensors, Fig. 5b, blue and yellow respectively. Voxels containing both F_1 and F_2 were selected by finding the lateral slices of $\underline{\mathbf{\Phi}}$ with non-zero entries in both subtensors (Fig. 5b, green slices, 2nd mode). Finally, we computed all pairwise angles between fascicles in F_1 and F_2 by identifying the atoms (indices in 1st mode) corresponding to the non-zero entries in those lateral slices of $\underline{\mathbf{\Phi}}$ (Fig. 5c,d).

Using the operations described above, we collected distributions of crossing angles, and computed peak distribution (μ) as well as width-at-half-max (σ, Fig. 5e). Importantly, we computed approximately 76,000,000 crossing-angles using fascicles validated statistically (fascicles with positive LiFE weights). Crossing angles distributions between Arc and CST peaked approximated at 75° and 78° for deterministic and probabilistic connectomes, respectively (μ, Fig. 5e). The measured σ was almost three-fold smaller for deterministic than probabilistic connectomes, 9° and 24°, respectively. These results must be put into context by considering the difference in quality of fit of the two connectomes; where probabilistic connectomes on average have a 4.4% lower error (s.d. 1.4%) and 16.2% higher number of supported fascicles (s.d. 1.1%) than deterministic ones (see Fig. 3a, datasets STN and HCP3T). Supplementary Fig. 5a shows the same analyses repeated with a different pair of tracts, the CST and SLF. Results are similar for these tracts with distribution peaking (μ) approximately at 78.1° and 86.4° for deterministic and probabilistic connectomes, respectively. Measured was almost two-fold smaller for deterministic than probabilistic connectomes, 17.1° and 31.5°, respectively.

In a second experiment, we measured μ and σ for the distribution of angles between fascicles within two tracts travelling approximately parallel across the axial plane of the human brain; the Arc and SLF (Fig. 5f). We computed angles distributions for both, probabilistic and deterministic connectomes. The peak distribution (μ) was approximately 0° and 15° for deterministic and probabilistic connectomes, respectively. The estimated σ were 8.1° and 16.6°, respectively, a 2x increase in variability.

In a final experiment we estimated the distribution of angles between fascicles in a tract, Arc, and its path neighborhood as function of tractography algorithm. Estimates of crossing angles between white matter tracts and path-neighborhoods have been debated[58–60]. We report μ and σ for crossing angles between the Arc and its path neighborhood using 8 subjects on STN and HCP3T data sets with probabilistic and deterministic ($L_{max} = 10$) tracking methods. For each subject, we identified the Arc and its path-neighborhood by using tensorial operations similar to the ones described in Fig. 5a–d. Results show characteristic bimodal distributions (Fig. 5g). A majority of the path-neighborhood fascicles show angles between 0° and 20° with tract fascicles (μ, 9° and 0° for probabilistic and deterministic tracking, respectively) and around 80° (μ, 81° and 80° for probabilistic and

deterministic tracking, respectively). The estimated σ for μ peaking at around 80° were 20.5° and 31.7° for deterministic and probabilistic connectomes, respectively, a 1.5x increase in variability.

Considering that probabilistic connectomes predict the diffusion measurement better than deterministic ones, these results demonstrate substantial variability in the estimates of crossing angles that can be obtained using neuroimaging methods and that the estimates will depend on the data and analysis methods[58–60]. This result shows a degree of variability of the estimates consistent with recent reports[82, 85].

Discussion

We presented a connectome encoding framework that provides investigators with an integrated multidimensional relationship between connectome nodes, edges and the associated measurements. We showed the utility of the encoding framework with four applications.

The recent increase in availability, quantity and quality of neuroimaging data and mapping methods poses new opportunities as well as challenges for mapping the human connectome[6–23, 55, 86]. Technological advances in dMRI data acquisition have permitted reduction of measurement time by factors up to 8-fold[87–89] and increase in spatial resolution up to 13-fold–when comparing volumetric resolution between clinical and high-field dMRI data[56] (e.g., 2.5 *mm* and 1.05 *mm* linear resolution respectively). Firstly, increased data quality and resolution also means increased size. Secondly, increased availability and diversity of data accompanied by the established variability in results from tractography, makes it difficult to identify a single tracking algorithm, parameter set or data type valid for every study[53, 58, 60, 63, 90, 91]. For this reason, developing principled methods for evaluating data quality and tractography routinely in their relation to the connectome estimates has become paramount.

The current practice in mapping connectomes is to choose a single tractography method and data resolution. Yet, multiple reports have been made highlighting many methodological limitations as well as the dependency of results on data and algorithm[90, 92–96]. As a result, we now understand that no single tracking method nor data set is likely to solve all problems or provide the ultimate quality. Instead, data and models will need to be improved and carefully evaluated. Routine statistical evaluation of brain connectomes can become standard practice in the process of connectome mapping[31, 53, 64, 65, 97]. The proposal is to build predictive models of the measured dMRI signal from the structure of brain connectomes[53, 63, 97] and compare the model prediction to the data by using statistical methods such as cross-validation[98]. The statistical evaluation approach complements the work on tractography validation based on either synthetic or post-mortem preparations[91, 99, 100]. Previous work evaluated model accuracy, namely how well a tractography method predicts independent dMRI measurements[53]. The present work advances by measuring model precision, how similar connectome estimates are when using a single tractography method repeatedly.

Multidimensional decomposition methods have been used to help investigators make sense of large multimodal datasets[50, 101]. Yet to date these methods have found only a few applications in neuroscience, such as performing multi-subjects, clustering and electroencephalography analyses[49, 102–107]. Generally, decomposition methods have been used to find compact representations of complex data by estimating the combination of a limited number of common meaningful factors that best fit the data[51, 101, 108]. We propose a new application that instead of using the decomposition to estimate latent factors, it encodes the structure of the problem explicitly. This innovative application in neuroscience can open new avenues of investigation in mapping brain and behavior using multivariate methods[109, 110] and to allow improving future generations of models of connectomics, tractography evaluation and microstructure[53, 63–65]. Improving these models will allow going beyond current limitations of the state of the art methods[93]. For example, extensions of the proposed framework would allow building more complex relationships between connectome matrices, edges and nodes without the loss of information of dMRI data and fascicles properties inherent to current methods for connectomics[4].

The field of network neuroscience[4] and the study of white matter[31, 111, 112] are striving to improve methods for mapping connectomes using modern large-scale data sets from living human brains. Our results show that connectome evaluation can be applied on such data sets with thousands of brain. In addition, the results show a profound effect of dMRI data spatial resolution on the number of brain connections that can be mapped. The effect of spatial resolution goes even beyond that of directional resolution that is lower for the HCP7T than the HCP3T data set used here[56]. This is particularly important because of its implications in guiding future study design.

Advances in frameworks to integrate computations on fascicles, brain areas as well as dMRI data, can profoundly improve efforts in clarifying the properties of human brain macroscopic connectivity[1, 29, 30, 109] and white matter microstructure[113–118]. Data representation frameworks such as the one proposed here have the potential to become fundamental in advancing the application of machine learning algorithms to mapping the functional, structural properties of the human connectome to capture brain individuality and variability in health and disease[21, 35, 74, 82, 119–123]. To contribute advancing scientific understanding and reproducibility, we provide an open source implementation of the encoding method and files to reproduce figures at http://www.github.com/brain-life/encode.

Methods

Diffusion-weighted MRI datasets. We use diffusion-weighted Magnetic Resonance Imaging data (dMRI) from three publicly available sources[22, 52, 53, 55, 56]. Dataset are available online at http://purl.stanford.edu/rt034xr8593, http://purl.stanford.edu/ng782rw8378 and https://www.humanconnectome.org/data/.

Stanford datasets. *STN*, 96 *gradient directions*, 1.5 *mm isotropic resolution*. dMRI dataset were collected in five males subjects (age 37–39) at the Stanford Center for Cognitive and Neurobiological Imaging using a 3 T General Electric Discovery 750 (General Electric Healthcare) equipped with a 32-channel head coil (Nova Medical). dMRI datasets with whole-brain volume coverage were acquired using a dual-spin echo diffusion-weighted sequence. Water-proton diffusion was measured using 96 directions chosen using the electrostatic repulsion

algorithm[124]. Diffusion-weighting gradient strength was set to 2,000 s/mm^2 ($TE = 96.8\,ms$). Data were acquired at 1.5 mm isotropic spatial resolution. Individual datasets were acquired twice and averaged in k-space ($NEX = 2$). Ten non-diffusion-weighted ($b = 0$) images were acquired at the beginning of each scan. Data acquisition and preprocessing steps are described in[53].

STN150, 150 *gradient directions*, 2.0 *mm isotropic resolution*. dMRI data were acquired in one subject using 150 directions, 2 mm isotropic spatial resolution and b value of 2,000 s/mm^2 ($TE = 83.1, 93.6$, and $106.9\,mm$).

Data acquisition and preprocessing steps are described in[53].

Human Connectome Project datasets.   *HCP3T*, 90 *gradient directions*, 1.25 *mm isotropic resolution*. Data of four subjects, part of the Human Connectome Project[55], acquired using a Siemens 3 T "Connectome" scanner were used. Measurements from the 2,000 s/mm^2 shell were extracted from the original dataset and used for all analyses. Processing methods are described in[22].

HCP7T, 60 *gradient directions*, 1.05 *mm isotropic resolution*. Five subjects part of the Human Connectome 7-Tesla (7 T) dataset were used. Data were collected a Siemens 7 T scanner[56]. Measurements from the 2,000 s/mm^2 shell were extracted from the original data and were used for further analyses.

Whole-brain connectomes generation.   Tractography was performed using the MRtrix 0.2 toolbox[70]. White-matter tissue was identified from the cortical segmentation performed on the T1-weighted images and resampled at the resolution of the dMRI data. Only white-matter voxels were used to seed fiber tracking. We used three tracking methods: (i) tensor-based deterministic tracking[70–72], (ii) CSD-based deterministic tracking[69, 70], and (iiI) CSD-based probabilistic tracking[69, 70, 125, 126]. Maximum harmonic orders (L_{max}) of 2, 4, 6, 8, 10 and 12 were used as long as the number of directions is larger than the number of parameters $N_p = 0.5(L_{max} + 1)$ $(L_{max} + 2)$[69, 127]. The following parameter values were used for all tracking: step size: 0.2 mm; minimum radius of curvature, 1 mm; maximum length, 200 mm; minimum length, 10 mm; and the fibers orientation distribution function (f_{ODF}) amplitude cutoff, was set to 0.1.

We created 10 candidate whole-brain connectomes by repeating tracking using 500,000 fascicles in each individual brain dataset (fourteen), tractography method (three) and parameter L_{max} (six).

A total number of 1,490 connectomes were generated in this work. For each connectome, fascicles of the twenty major human were identified using Automatic Fiber Quantification - AFQ[75].

The Linear Fascicle Evaluation (LiFE) method.   Here we introduce the linear model used in[53] to predict diffusion signals based on a multi-compartment voxel model[128, 129]. We refer to Supplementary section 1.2 for an introduction to magnetic resonance diffusion signals.

For a given sensitization strength b and gradient direction θ, the diffusion signal $S(\theta, v)$ measured at a location within a brain (voxel v) can be estimated by using the following Equation:

$$S(\theta, v) \approx S_0(v)\left(w_0 e^{-A_0} + \sum_{f \in v} w_f e^{-b\theta^T \mathbf{Q}_{f,v}\theta}\right),$$
(1)

where f is the index of the candidate white-matter fascicles within the voxel, $S(\theta,v)$ is the diffusion-weighted signal, $S_0(v)$ is the non diffusion-weighted signal ($b = 0$), A_0 is the isotropic apparent diffusion (diffusion in all directions) and $\mathbf{Q}_{f,v}$ is the diffusion tensor matrix (see Supplementary section 1.2).

LiFE predicts the demeaned diffusion signal defined as $\overline{S}(\theta, v) = S(\theta, v) - I_v$, where $I_v = \frac{1}{N_\theta}\sum_\theta S(\theta, v)$ is the mean and N_θ is the number of gradient directions[52, 53]. Using this definition and Equation (1) we arrive at:

$$\overline{S}(\theta, v) \approx \sum_{f \in v} w_f S_0(v) O_f(\theta, v),$$
(2)

where and $O_f(\theta,v)$ is the orientation distribution function specific to each fascicle, i.e. the anisotropic modulation of the diffusion signal around its mean and it is defined as follows:

$$O_f(\theta, v) = e^{-b\theta^T \mathbf{Q}_{f,v}\theta} - \frac{1}{N_\theta}\sum_\theta e^{-b\theta^T \mathbf{Q}_{f,v}\theta}.$$
(3)

The right-hand side of Equation (2) is the prediction model (see Supplementary Fig. 2a,b). The LiFE model extends from the single voxel to all white-matter voxels in the following way (see Supplementary Fig. 2c):

$$\mathbf{y} \approx \mathbf{Mw},$$
(4)

where $\mathbf{y} \in \mathbb{R}^{N_\theta N_v}$ is a vector containing the demeaned signal for all white-matter voxels v and across all gradient directions θ, i.e. $y_i = \overline{S}(\theta_i, v_i)$. The matrix $\mathbf{M} \in \mathbb{R}^{N_\theta N_v \times N_f}$ contains at column f the signal contribution given by fascicle f at all voxels across all gradient directions, i.e., $\mathbf{M}(i,f) = S_0(v_i)O_f(\theta_i)$, and $\mathbf{w} \in \mathbb{R}^{N_f}$ contains the weights for each fascicle in the connectome.

The vector of weights $\mathbf{w}$ in Equation (4) and Supplementary Fig. 2c is computed by solving a convex optimization problem[53, 64]. More specifically we solve a non-negative least-square (NNLS) problem, defined as follows:

$$\min_{\mathbf{w}}\left(\frac{1}{2}\|\mathbf{y} - \mathbf{Mw}\|^2\right) \text{ subject to } w_f \geq 0, \forall f.$$
(5)

9

Commonly, the size of the matrix **M** is very large (around 30 *GB* or 40 *GB* for the datasets used here, see Fig. 2d). Because of this reason, we use NNLS algorithms suitable for large scale problems, such as the BB-NNLS developed in[130].

Connectome model prediction error. LiFE predicts the measured (demeaned) diffusion signal using the right-hand side of Equation (2). Thus, we can assess the ability of LiFE to model the measured diffusion signal by computing the prediction error in each white-matter voxel. In order to make errors relatively independent of scanner parameters, we compute them on the relative diffusion signal (also referred to as diffusion attenuation), defined as follows:

$$\bar{S}_r(\boldsymbol{\theta}, v) = \bar{S}(\boldsymbol{\theta}, v)/S_0(v). \tag{6}$$

The root mean squared (r.m.s) error in voxel *v* is defined as follows:

$$e_{rms}(v) = \sqrt{\frac{1}{N_\theta}\sum_{\boldsymbol{\theta}}\left(\bar{S}_r(\boldsymbol{\theta}, v) - \sum_{f \in v}w_f O_f(\boldsymbol{\theta}, v)\right)^2}. \tag{7}$$

The r.m.s error (Equation 7) can be used to compare alternative connectome models. A global r.m.s error $\bar{e}_{rms}$ can be computed by averaging $e_{rms}(v)$ over all voxels:

$$\bar{e}_{rms} = \frac{1}{N_v}\sum_v e_{rms}(v). \tag{8}$$

LiFE models comparison. We compare a LiFE$_M$ model matrix **M** (see Equation 4) and its approximated version $\hat{\mathbf{M}}$ using the relative error:

$$e_{\mathbf{M}} = \|\mathbf{M} - \hat{\mathbf{M}}\|_F / \|\mathbf{M}\|_F, \tag{9}$$

where $\|\mathbf{M}\|_F = \sqrt{\sum_{i,j}\mathbf{M}^2(i, j)}$ is the Frobenius matrix norm.

Similarly, we compare a vector of LiFE$_M$ weights **w** and its approximated version $\hat{\mathbf{w}}$ using the relative error defined as follows:

$$e_{\mathbf{w}} = \|\mathbf{w} - \hat{\mathbf{w}}\| / \|\mathbf{w}\|, \tag{10}$$

where $\|\mathbf{w}\| = \sqrt{\sum_f w_f^2}$ is the Euclidean vector-norm.

References

1. Sporns, O. & Betzel, R. F. Modular Brain Networks. *Annual Review of Psychology* **67**, 613–640 (2016).
2. Dubois, J. & Adolphs, R. Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences* **20**, 425–443 (2016).
3. Van den Heuvel, M. P., Bullmore, E. T. & Sporns, O. Comparative Connectomics. *Trends in Cognitive Sciences* **20**, 345–361 (2016).
4. Bassett, D. S. & Sporns, O. Network neuroscience. *Nature Neuroscience* **20**, 353–364 (2017).
5. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiology study. *Nature Neuroscience* 1–18 (2016).
6. Di Martino, A. *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**, 659–667 (2014).
7. Jernigan, T. L. *et al.* The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *Human Brain Mapping Journal* **124**, 1149–1154 (2016).
8. Taylor, J. R. *et al.* The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *Human Brain Mapping Journal* (2015).
9. Nooner, K. B. *et al.* The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Frontiers in Neuroscience* **6** (2012).
10. Jack, C. R. *et al.* The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of magnetic resonance imaging: JMRI* **27**, 685–691 (2008).
11. Biswal, B. B. *et al.* Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4734–4739 (2010).
12. Zuo, X.-N. *et al.* An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific Data* **1**, 140049 (2014).
13. Thompson, P. M., Hibar, D. P., Stein, J. L. & Jahanshad, N. Imaging Genomics and ENIGMA. In *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry* 101–115 (Elsevier, 2016).
14. Thompson, P. M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior* **8**, 1–30 (2014).
15. Holmes, A. J. *et al.* Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Scientific Data* **2**, 150031–16 (2015).
16. Sowell, E. R. *et al.* Mapping cortical change across the human life span. *Nature Neuroscience* **6**, 309–315 (2003).
17. Tavor, I. *et al.* Task-free MRI predicts individual differences in brain activity during task performance. *Science* **352**, 216–220 (2016).
18. Poldrack, R. A. *et al.* Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* **6**, 8885 (2015).
19. Laumann, T. O. *et al.* Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* **87**, 657––670 (2015).
20. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Publishing Group* **18**, 1664–1671 (2015).
21. Smith, S. M. *et al.* A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Publishing Group* **18**, 1565–1567 (2015).
22. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79 (2013).

23. Thomason, M. E. & Thompson, P. M. Diffusion imaging, white matter, and psychopathology. *Clinical Psychology* **7**, 63–85 (2011).
24. Sejnowski, T. J., Churchland, P. S. & Movshon, J. A. Putting big data to good use in neuroscience. *Nature Neuroscience* **17**, 1440–1441 (2014).
25. Dinov, I. D. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *GigaScience* 1–15 (2016).
26. Dell'Acqua, F. & Catani, M. Structural human brain networks. *Current Opinion in Neurology* 1 (2012).
27. Sporns, O. Making sense of brain network data. *Nature Methods* **10**, 491–493 (2013).
28. Van den Heuvel, M. P. & Sporns, O. Rich-Club Organization of the Human Connectome. *Journal of Neuroscience* **31**, 15775–15786 (2011).
29. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews: Neuroscience* **10**, 186–198 (2009).
30. Jbabdi, S., Sotiropoulos, S. N., Haber, S. N., Van Essen, D. C. & Behrens, T. E. Measuring macroscopic brain connections *in vivo*. *Nature Publishing Group* **18**, 1546–1555 (2015).
31. Wandell, B. A. Clarifying Human White Matter. *Annual Review of Neuroscience* **39**, 103–128 (2016).
32. Gu, S. *et al.* Controllability of structural brain networks. *Nat. Commun.* **6**, 8414 (2015).
33. Khambhati, A. N. & Bassett, D. S. A Powerful DREADD: Revealing Structural Drivers of Functional Dynamics. *Neuron* **91**, 213–215 (2016).
34. Davison, E. N. *et al.* Individual Differences in Dynamic Functional Brain Connectivity across the Human Lifespan. *PLoS Computational Biology* **12**, e1005178–29 (2016).
35. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature Publishing Group* **536**, 171–178 (2016).
36. Donahue, C. J. *et al.* Using Diffusion Tractography to Predict Cortical Connection Strength and Distance: A Quantitative Comparison with Tracers in the Monkey. *Journal of Neuroscience* **36**, 6758–6770 (2016).
37. Gomez, J. *et al.* Functionally defined white matter reveals segregated pathways in human ventral temporal cortex associated with category-specific processing. *Neuron* **85**, 216–227 (2015).
38. Yeatman, J. D. *et al.* The vertical occipital fasciculus: A century of controversy resolved by *in vivo* measurements. *Proceedings of the National Academy of Sciences* **111**, E5214–E5223 (2014).
39. Leong, J. K., Pestilli, F., Wu, C. C., Samanez-Larkin, G. R. & Knutson, B. White-Matter Tract Connecting Anterior Insula to Nucleus Accumbens Correlates with Reduced Preference for Positively Skewed Gambles. *Neuron* **89**, 63–69 (2016).
40. Atasoy, S., Donnelly, I. & Pearson, J. Human brain networks function in connectome-specific harmonic waves. *Nat. Commun.* **7**, 10340 (2016).
41. Catani, M. & de Schotten, M. T. *Atlas of Human Brain Connections* (Oxford University Press, 2012).
42. Zhang, Y. *et al.* Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage* **52**, 1289–1301 (2010).
43. Bürgel, U. *et al.* White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage* **29**, 1092–1105 (2006).
44. Jacobs, B. & Scheibel, A. B. A quantitative dendritic analysis of Wernicke's area in humans. I. Lifespan changes. *The Journal of comparative neurology* **327**, 83–96 (1993).
45. Amunts, K. *et al.* Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology* **412**, 319–341 (1999).
46. Sporns, O., Tononi, G. & Kötter, R. The human connectome: A structural description of the human brain. *PLoS Computational Biology* **1**, e42–e42 (2005).
47. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).
48. Comon, P. Tensors: A brief introduction. *IEEE Signal Processing Magazine* **31**, 44–53 (2014).
49. Beckmann, C. F. & Smith, S. M. Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage* **25**, 294–311 (2005).
50. Cichocki, A. *et al.* Tensor decompositions for signal processing applications: from two-way to multiway component analysis. *IEEE Signal Processing Magazine* **32**, 145–163 (2015).
51. Kolda, T. & Bader, B. Tensor decompositions and applications. *SIAM Review* **51**, 455–500 (2009).
52. Rokem, A. *et al.* Evaluating the accuracy of diffusion MRI models in white matter. *PLoS ONE* **10**, e0123272 (2015).
53. Pestilli, F., Yeatman, J. D., Rokem, A., Kay, K. N. & Wandell, B. A. Evaluation and statistical inference for human connectomes. *Nature Methods* **11**, 1058–1063 (2014).
54. Van Essen, D. C. *et al.* The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62**, 2222–2231 (2012).
55. Sotiropoulos, S. N. *et al.* Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *Human Brain Mapping Journal* **80**, 125–143 (2013).
56. Vu, A. T. *et al.* High resolution whole brain diffusion imaging at 7 T for the Human Connectome Project. *Human Brain Mapping Journal* **122**, 318–331 (2015).
57. Honey, C. J. & Sporns, O. Dynamical consequences of lesions in cortical networks. *Human Brain Mapping* **29**, 802–809 (2008).
58. Wedeen, V. J. *et al.* The geometric structure of the brain fiber pathways. *Science* **335**, 1628–1634 (2012).
59. Wedeen, V. J. *et al.* Response to comment on "The geometric structure of the brain fiber pathways". *Science* **337**, 1605–1605 (2012).
60. Catani, M., Bodi, I. & Dell'Acqua, F. Comment on "The geometric structure of the brain fiber pathways". *Science* **337**, 1605 (2012).
61. Poline, J.-B. *et al.* Data sharing in neuroimaging research. *Frontiers in Neuroinformatics* **6**, 9 (2012).
62. Donoho, D. L. An invitation to reproducible computational research. *Biostatistics* **11**, 385–388 (2010).
63. Takemura, H., Caiafa, C. F., Wandell, B. A. & Pestilli, F. Ensemble Tractography. *PLoS Computational Biology* **12**, e1004692 (2016).
64. Daducci, A., Palù, A. D., Lemkaddem, A. & Thiran, J.-P. COMMIT: Convex optimization modeling for microstructure informed tractography. *Medical Imaging, IEEE Transactions on* **34**, 246–257 (2015).
65. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *Human Brain Mapping Journal* **119**, 338–351 (2015).
66. Bader, B. W. & Kolda, T. G. Efficient MATLAB Computations with Sparse and Factored Tensors. *SIAM J. SCI. COMPUT.* **30**, 205–231 (2008).
67. Gilbert, J. R., Moler, C. & Schreiber, R. Sparse matrices in matlab: design and implementation. *SIAM Journal on Matrix Analysis and Applications* **13**, 333–356 (1992).
68. Ugurbil, K. *et al.* Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Human Brain Mapping Journal* **80**, 80–104 (2013).
69. Descoteaux, M., Deriche, R., Knosche, T. R. & Anwander, A. Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *Medical Imaging, IEEE Transactions on* **28**, 269–286 (2009).
70. Tournier, J.-D., Calamante, F. & Connelly, A. MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* **22**, 53–66 (2012).
71. Lazar, M. *et al.* White matter tractography using diffusion tensor deflection. *Human Brain Mapping* **18**, 306–321 (2003).
72. Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J. & Aldroubi, A. *In vivo* fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* **44**, 625–632 (2000).

73. Risacher, S. L. & Saykin, A. J. Neuroimaging and Other Biomarkers for Alzheimer's Disease: The Changing Landscape of Early Detection. *Clinical Psychology* **9**, 621–648 (2013).

74. Alexander, D. C. *et al*. Image quality transfer and applications in diffusion MRI. *Human Brain Mapping Journal* 1–65 (2017).

75. Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A. & Feldman, H. M. Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS ONE* **7**, e49790 (2012).

76. Pascual-Leone, A., Bartres-Faz, D. & Keenan, J. P. Transcranial magnetic stimulation: studying the brain-behaviour relationship by induction of 'virtual lesions'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **354**, 1229–1238 (1999).

77. Pascual-Leone, A. Transcranial magnetic stimulation in cognitive neuroscience – virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology* **10**, 232–237 (2000).

78. Pascual-Leone, A., Amedi, A., Fregni, F. & Merabet, L. B. The plastic human brain cortex. *Annual Review of Neuroscience* **28**, 377–401 (2005).

79. Amedi, A., Floel, A., Knecht, S., Zohary, E. & Cohen, L. G. Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nature Neuroscience* **7**, 1266–1270 (2004).

80. Takemura, H. *et al*. A Major Human White Matter Pathway Between Dorsal and Ventral Visual Cortex. *Cerebral cortex (New York, NY: 1991)* **26**, 2205–2214 (2016).

81. Rubner, Y., Tomasi, C. & Guibas, L. J. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision* **40**, 99–121 (2000).

82. Tax, C. M. W. *et al*. Sheet Probability Index (SPI): Characterizing the geometrical organization of the white matter with diffusion MRI. *Human Brain Mapping Journal* 1–53 (2016).

83. Tax, C. M. W. *et al*. Quantifying the brain's sheet structure with normalized convolution. *Medical Image Analysis* 1–36 (2017).

84. Fields, R. D. A new mechanism of nervous system plasticity: activity-dependent myelination. *Nature Reviews: Neuroscience* **16**, 756–767 (2015).

85. De Santis, S., Assaf, Y., Jeurissen, B., Jones, D. K. & Roebroeck, A. T1 relaxometry of crossing fibres in the human brain. *Human Brain Mapping Journal* **141**, 133–142 (2016).

86. Sudlow, C. *et al*. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779–10 (2015).

87. Breuer, F. A. *et al*. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magnetic Resonance in Medicine* **55**, 549–556 (2006).

88. Moeller, S. *et al*. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine* **63**, 1144–1153 (2010).

89. Feinberg, D. A. *et al*. Correction: Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLoS ONE* **6** (2011).

90. Bastiani, M., Shah, N. J., Goebel, R. & Roebroeck, A. Human cortical connectome reconstruction from diffusion weighted MRI: the effect of tractography algorithm. *Human Brain Mapping Journal* **62**, 1732–1749 (2012).

91. Seehaus, A. *et al*. Histological validation of high-resolution DTI in human post mortem tissue. *Front. Neuroanat.* **9**, 98 (2015).

92. Daducci, A. *et al*. Quantitative Comparison of Reconstruction Methods for Intra-Voxel Fiber Recovery From Diffusion MRI. *Medical Imaging, IEEE Transactions on* **33**, 384–399 (2014).

93. Daducci, A., Dal Palu, A., Descoteaux, M. & Thiran, J.-P. Microstructure Informed Tractography: Pitfalls and Open Challenges. *Frontiers in Neuroscience* **10**, 1374–13 (2016).

94. Zalesky, A. *et al*. Whole-brain anatomical networks: does the choice of nodes matter? *Human Brain Mapping Journal* **50**, 970–983 (2010).

95. Bonilha, L. *et al*. Reproducibility of the Structural Brain Connectome Derived from Diffusion Tensor Imaging. *PLoS ONE* **10**, e0135247 (2015).

96. Bassett, D. S., Brown, J. A., Deshpande, V., Carlson, J. M. & Grafton, S. T. Conserved and variable architecture of human white matter connectivity. *Human Brain Mapping Journal* **54**, 1262–1279 (2011).

97. Pestilli, F. Test-retest measurements and digital validation for *in vivo* neuroscience. *Scientific Data* **2**, 140057 (2015).

98. Efron, B. & Gong, G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician* **37**, 36 (1983).

99. Seehaus, A. K. *et al*. Histological validation of DW-MRI tractography in human postmortem tissue. *Cerebral Cortex* **23**, 442–450 (2013).

100. Knösche, T. R., Anwander, A., Liptrot, M. & Dyrby, T. B. Validation of tractography: Comparison with manganese tracing. *Human Brain Mapping* **36**, 4116–4134 (2015).

101. Mørup, M. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1**, 24–40 (2011).

102. Zhao, Q. *et al*. Higher Order Partial Least Squares (HOPLS): A Generalized Multilinear Regression Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1660–1673 (2013).

103. Yu, Y., Jin, J., Liu, F. & Crozier, S. Multidimensional Compressed Sensing MRI Using Tensor Decomposition-Based Sparsifying Transform. *PLoS ONE* **9**, e98441 (2014).

104. Barnathan, M., Megalooikonomou, V., Faloutsos, C., Faro, S. & Mohamed, F. B. TWave: High-order analysis of functional MRI. *Human Brain Mapping Journal* **58**, 537–548 (2011).

105. Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J. & Arnfred, S. M. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. *Human Brain Mapping Journal* **29**, 938–947 (2006).

106. Miwakeichi, F. *et al*. Decomposing EEG Data into Space–time–frequency Components using Parallel Factor Analysis. *NeuroImage* **22**, 1035–1045 (2004).

107. Cong, F. *et al*. Tensor decomposition of EEG signals: a brief review. *Journal of neuroscience methods* **248**, 59–69 (2015).

108. Kroonenberg, P. M. *Applied Multiway Data Analysis* (John Wiley & Sons, 2008).

109. Misic, B. & Sporns, O. From regions to connections and networks: new bridges between brain and behavior. *Current Opinion in Neurobiology* **40**, 1–7 (2016).

110. McIntosh, A. R. & Misic, B. Multivariate Statistical Analyses for Neuroimaging Data. *Annual Review of Psychology* **64**, 499–525 (2013).

111. Rokem, A. *et al*. The visual white matter: The application of diffusion MRI and fiber tractography to vision science. *Journal of Vision* **17**, 4 (2017).

112. Shi, Y. & Toga, A. W. Connectome imaging for mapping human brain pathways. *Mol. Psychiatry* **340**, 1234 (2017).

113. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: Practical *in vivo* neurite orientation dispersion and density imaging of the human brain. *Human Brain Mapping. Journal* **61**, 1000–1016 (2012).

114. Farooq, H. *et al*. Microstructure Imaging of Crossing (MIX) White Matter Fibers from diffusion MRI. *Nature Publishing Group* **6**, 38927 (2016).

115. Assaf, Y. & Basser, P. J. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage* **27**, 48–58 (2005).

116. Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y. & Basser, P. J. Axcaliber: A method for measuring axon diameter distribution from diffusion MRI. *Magnetic Resonance in Medicine* **59**, 1347–1354 (2008).

117. Dyrby, T. B., S gaard, L. V., Hall, M. G., Ptito, M. & Alexander, D. C. Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI. *Magnetic Resonance in Medicine* **70**, 711–721 (2012).
118. Ferizi, U. *et al.* White matter compartment models for *in vivo* diffusion MRI at 300mT/m. *Human Brain Mapping. Journal* **118**, 468–483 (2015).
119. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine* 1–16 (2016).
120. Hazlett, H. C. *et al.* Early brain development in infants at high risk for autism spectrum disorder. *Nature Publishing Group* **542**, 348–351 (2017).
121. Neher, P. F., Cote, M.-A., Houde, J.-C., Descoteaux, M. & Maier-Hein, K. H. Fiber tractography using machine learning. *bioRxiv* 1–20 (2017).
122. Zhu, D., Jahanshad, N., Riedel, B. C. & Zhan, L. Population learning of structural connectivity by white matter encoding and decoding. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 554–558 (IEEE, 2016).
123. Nedjati-Gilani, G. L. *et al.* Machine learning based compartment models with permeability for white matter microstructure imaging. *Human Brain Mapping Journal* **150**, 119–135 (2017).
124. Jones, D. K., Horsfield, M. A. & Simmons, A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine* **42**, 515–525 (1999).
125. Behrens, T. *et al.* Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience* **6**, 750–757 (2003).
126. Parker, G. J. M., Haroon, H. A. & Wheeler-Kingshott, C. A. M. A framework for a streamline-based Probabilistic Index of Connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging* **18**, 242–254 (2003).
127. Tournier, J.-D., Calamante, F., Gadian, D. G. & Connelly, A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* **23**, 1176–1185 (2004).
128. Frank, L. R. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine* **47**, 1083–1099 (2002).
129. Behrens, T. E. J. *et al.* Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine* **50**, 1077–1088 (2003).
130. Kim, D., Sra, S. & Dhillon, I. S. A non-monotonic method for large-scale non-negative least squares. *Optimization Methods and Software* **28**, 1012–1039 (2013).

Acknowledgements

Author Contributions

F.P. and C.F.C. conceived the study. C.F.C. developed the tensor decomposition model. C.F.C. and F.P. designed and performed experiments. C.F.C. and F.P. wrote paper. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-09250-w

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.