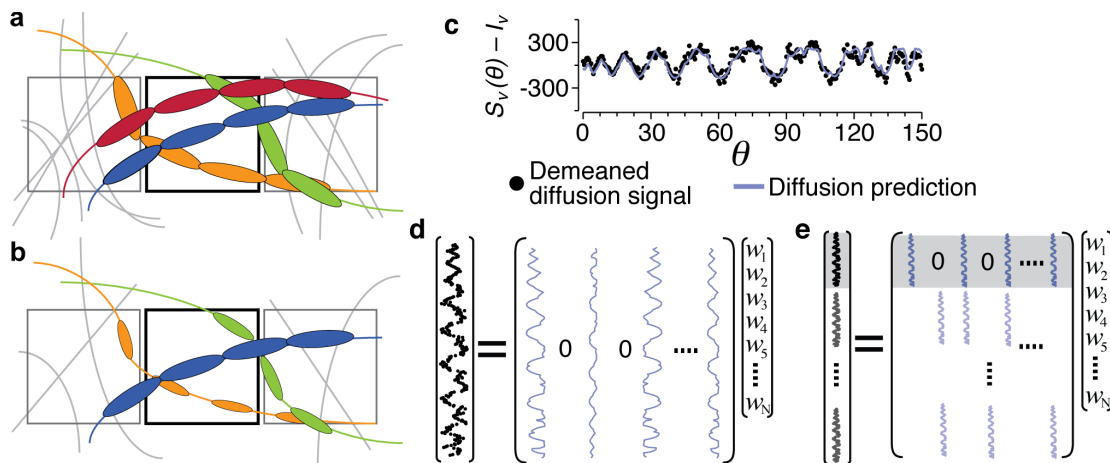**Supplementary Figure 1**

**Tract trajectory estimates and cortical projection zones for different tractography algorithms.**

**a. Deterministic tractography.** Sagittal view of the right Superior Lateral Fasciculus (SLF; purple) and right Arcuate Fasciculus (gold) in a candidate connectome generated using deterministic tractography (Tournier et al., 2012). Right panel, lateral view of the right cortical surface showing the cortical projection zones of the SLF. The color overlay depicts the density of fiber projections (yellow: higher density).

**b. Probabilistic tractography.** As in part (a) except the candidate connectome was generated using probabilistic tractography. Right panel, SLF cortical projection zones.

**Supplementary Figure 2**

**The linear fascicle evaluation (LiFE) method.**

a, b. Three adjacent voxels and multiple fascicle estimates are illustrated. For clarity, only fascicles passing through the middle voxel are colored. The ellipsoids suggest the predicted diffusion at each fascicle node, separated by 1 mm. **a. The candidate connectome.** The candidate contains many fascicles that were estimated independently using tractography. **b. The optimized connectome.** LiFE solves a set of simultaneous linear equations in order to evaluate how well the full candidate connectome predicts the diffusion data (see Methods Eq. 11). The solution yields a weight for each fascicle illustrated by the size of the ellipsoids. The weight describes the fascicle's contribution in predicting the diffusion data. In the example, the red fascicle receives a zero weight and is eliminated. The blue, green and orange fascicles weights are positive, but differ from one another. LiFE evaluates all the fascicles in all the voxels simultaneously, and thus it is a global connectome evaluation method.

**c. The predicted and observed demeaned diffusion signal in a typical voxel.** The x-axis represents the 150 diffusion directions measured for this voxel. The y-axis shows the demeaned diffusion signal. **d. The matrix tableau of the LiFE linear model for a single voxel.** A voxel containing four fascicles is shown. The vector on the left is the demeaned signal across orientations. The columns of the matrix on the right contain the canonical prediction for the four fascicles, each with its own orientation. Columns of fascicles that do not pass through the voxel are all zeros. **e. A matrix tableau of the LiFE linear model for all voxels**. The vector on the left contains the demeaned signal of every voxel stacked vertically. The matrix is the prediction matrices stacked vertically. A candidate whole brain LiFE model generally begins with about 500,000 columns (fascicles) and 7,000,000 rows (voxels X directions). The optimized connectome will have fewer fascicles (columns). The gray-shaded region corresponds to the example voxel in d. The matrix is sparse because most fascicles do not pass through most voxels.

LiFE estimates the fascicle weights that best predict the diffusion signal. The relationship between the demeaned diffusion signal, Eq. 7, the diffusion modulation function, $O_f(\theta)$ and the fascicle weights, $w_f$, can be written in matrix tableau form as

$$\begin{pmatrix} \vdots \\ \vdots \\ S(\theta) - I_v \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ O_{f1}(\theta) & O_{f2}(\theta) & \cdots & O_{fN}(\theta) \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}$$

Each term in the Tableau is defined in Methods. The Tableau describes the predictions for a single voxel based on the fascicles passing through that voxel. The matrices for many voxels are combined to form a larger matrix that describes the linear equation used to estimate the weights for the complete connectome (Supplemental Figure 2e).
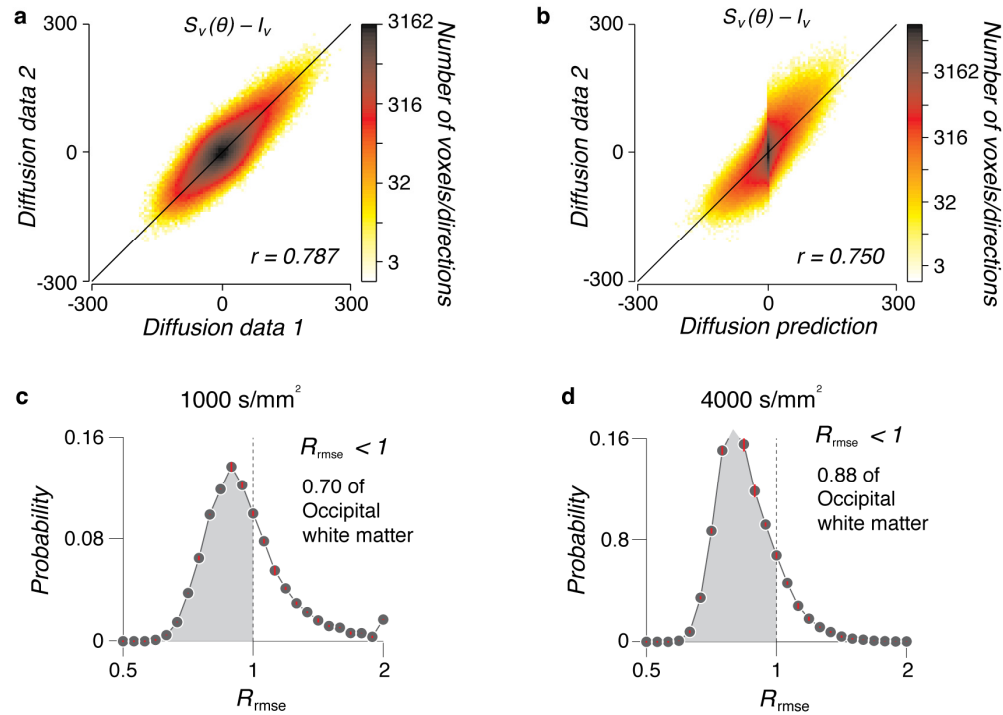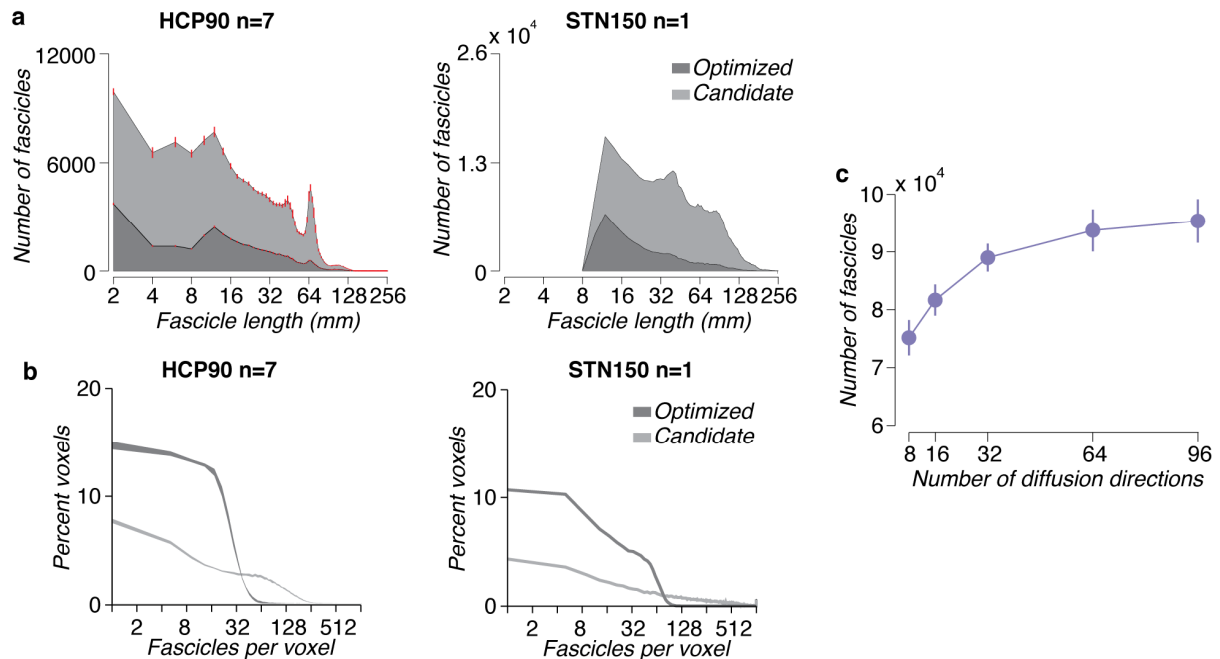
**Supplementary Figure 3**

**Test-retest reliability, signal prediction and quality of prediction.**

**a,b. Scatter-density plots of the diffusion modulation for each voxel and direction, STN150. a. Test-retest reliability.** The first measurement (Diffusion data 1) predicts a second measurement (Diffusion data 2) with a Pearson correlation (r) of 0.787. **b.** The optimized connectome model (Diffusion prediction) predicts a second measurement with a correlation of r=0.75.

**c, d.The optimized probabilistic connectome predicts diffusion as well as test-retest reliability across b-values and repetitions of the tractography process, STN150. c. Relative error histogram at b=1000.** Histogram of the probability of observing a certain $R_{rmse}$ in the entire white-matter volume for a data set with a b-value of 1000 s/mm$^2$. The voxel-wise $R_{rmse}$ model predicts better than data (< 1) in about 70% of the white-matter volume. **d. Relative error histogram at b=4000.** As in (h) but for a b-value of 4000 s/mm$^2$. The voxel-wise $R_{rmse}$ model predicts better than test-retest reliability (< 1) in about 80% of the volume. Error bars show ±2 standard deviations across repetitions of the entire process. We obtained estimates of the median $R_{rmse}$ in each data set by repeating the following steps three times: 1. Constrained spherical deconvolution modeling of the diffusion signal, 2. Random seeding for fiber generation. 3. Whole-brain tracking, 4. LiFE model building and fitting. Stochasticity in the process is introduced at steps 2 and 3.
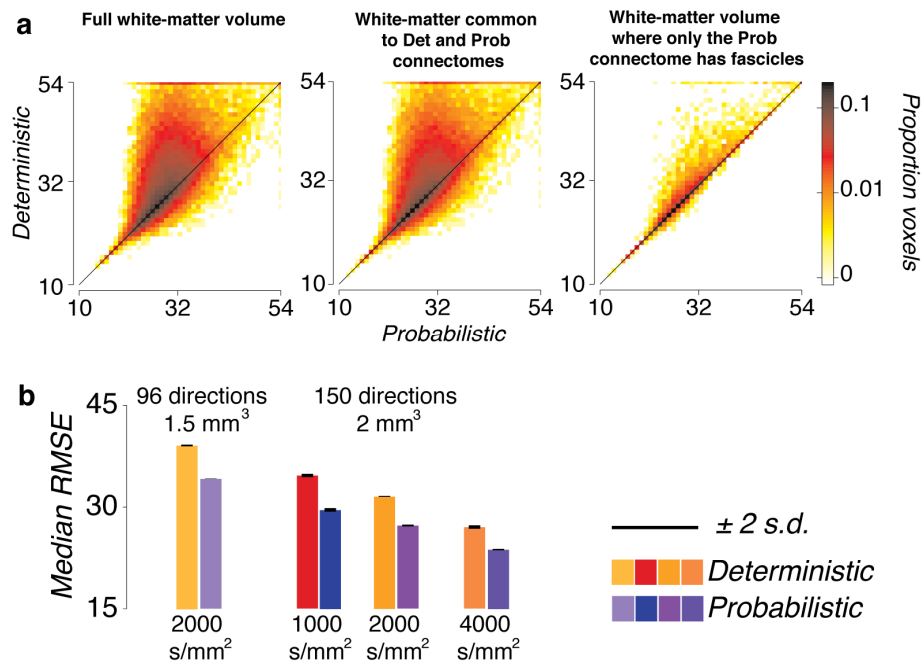
**Supplementary Figure 4**

**Fiber density and length.**

**a. Fascicle length distribution.** Histogram of fascicles length for candidate and optimized connectomes from data sets HCP90 and STN150 (Probabilistic CSD $L_{max}$=10).

**b. Candidate and optimized connectome fascicle density histograms.** Fiber density histogram for the candidate (light gray) and optimized (dark gray) connectomes for data set HCP90 (a) and STN150 (b). Line-width in a indicates ±1 standard deviations across seven brains. Dataset HCP90 is provided by the Human Connectome Project (HCP) at a spatial resolution of 1.25 mm isotropic (see Methods). These data provide measurements at multiple b-values, and we selected the 90 directions that matched the gradient strength in the other data sets (b=2000 s/mm$^2$). For this data set we analyzed large portions of the white-matter comprising the left and right occipital lobes, part of the temporal and parietal lobes and the cerebellum. Results are similar to STN96, which has identical b-value and similar number of directions (90 versus 96) and spatial resolution (1.25 versus 1.5 mm$^3$).

**c. Number of supported fascicles as function of number of diffusion directions.** We analyzed the effect of reducing the number of directions in STN96. We used LiFE to calculate 6 whole-brain optimized connectomes from the STN96 data set and counted the number of fascicles. We then eliminated a subset of the diffusion directions (choosing the retained directions along surface of a sphere using the electro-static repulsion algorithm of Jones, Horsfield, & Simmons, 1999), and recomputed the optimized connectome. The graph shows the relationship between the number of directions and the number of fascicles. Error bars are ±1 s.e.m. across six brains.

**a**

Full white-matter volume

White-matter common to Det and Prob connectomes

White-matter volume where only the Prob connectome has fascicles

*Deterministic*

54

32

10

10    32    54

*Probabilistic*

0.1

0.01

0

*Proportion voxels*

**b**

*Median RMSE*

45

30

15

96 directions
1.5 mm$^3$

150 directions
2 mm$^3$

2000
s/mm$^2$

1000
s/mm$^2$

2000
s/mm$^2$

4000
s/mm$^2$

—— ± 2 s.d.

*Deterministic*

*Probabilistic*

**Supplementary Figure 5**

**Comparison of probabilistic and deterministic optimized connectomes in different data sets.**

**a.** Each white matter voxel has a cross-validated RMSE for both the optimized probabilistic and deterministic connectomes. The scatter-density plot compares these two values. **Left panel.** Voxel-wise prediction error estimated in the whole white-matter volume. This plot compares the two probabilistic and deterministic connectomes in the entirety of the white-matter voxels. Because the deterministic model fills up only about 80% of the total white-matter volume, to create this plot we set the prediction of the deterministic connectome for voxels where it has no fascicles to the isotropic prediction. This plot is equivalent to the one in Fig 5a in the main article, except that the axes are scaled differently. **Center panel.** Voxel-wise prediction error estimated in the subset of white-matter voxels common to both deterministic and probabilistic connectomes. This plot avoids setting the prediction of the deterministic model to zero in voxels where the model has no fascicles. **Right panel.** Voxel-wise prediction error estimated in the subset of white-matter voxels where only the probabilistic connectome has fascicles. This plot compares the prediction error of the probabilistic connectome with a model that predicts isotropic signal in voxels where the deterministic connectome has no fascicles and hence the prediction is zero.

**b.** Median RMSE from 24 example connectomes estimated with deterministic and probabilistic tractography in four different data sets (STN96 and STN150 at 1000, 2000, 4000 s/mm$^2$). Each bar shows the median RMSE for within the white matter of the right occipital lobe of an individual. Warm colors show the RMSE for the deterministic connectome. The cool colors show the median RMSE for the probabilistic connectome. Results are shown for four different data sets with different spatial and angular resolutions and with different b-values. Error bars show ±2 standard deviations across three repetitions of the entire process. We obtained estimates of the median RMSE in each data set by repeating the following steps three times: 1. Constrained spherical deconvolution modeling of the diffusion signal, 2. Random seeding for fiber generation. 3. Whole-brain tracking, 4. LiFE model building and fitting. Stochasticity in the process is introduced at steps 2 and 3.

**a**

**b**
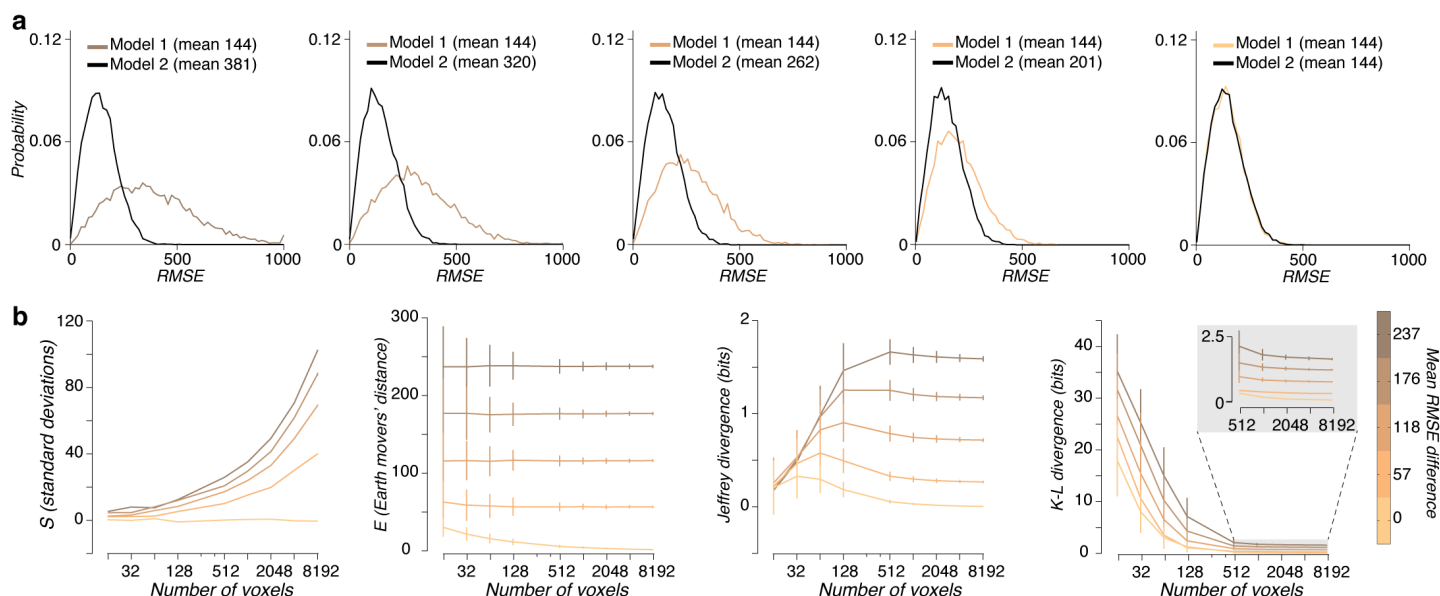
**Supplementary Figure 6**

**Evaluation of metrics to compare r.m.s. error distributions.**

**a. Simulated distributions of root-mean-squared-error (RMSE) with five different effect sizes.** RMSE distributions were simulated using a non-negative (Rician) distribution which approximates the RMSE distributions. One distribution was fixed at a low RMSE level (black $v$=80; $n$=100); the spread parameter ($n$) of the second distribution was varied (brown colors; $v$=80; n=100, 150, 200, 250, 300) to simulate the RMSE distribution observed at different effect sizes (Fig. 4 and 5).

**b. Evaluation of four measures that compare RMSE distributions as function of effect size (color) and number of voxels (x-axis).** $S$ increases systematically with number of voxels and effect size. Error bars indicate ±1 standard deviation across 10 Monte Carlo repeats of the whole simulation process. The mean $E$ measure is relatively invariant to the number of voxels and increases systematically with effect size. The *Jeffrey's* and *K-L* divergence measures are biased for small number of voxels. See Rubner, Tomasi and Guibas (2000) for an extensive treatment of these three measures.
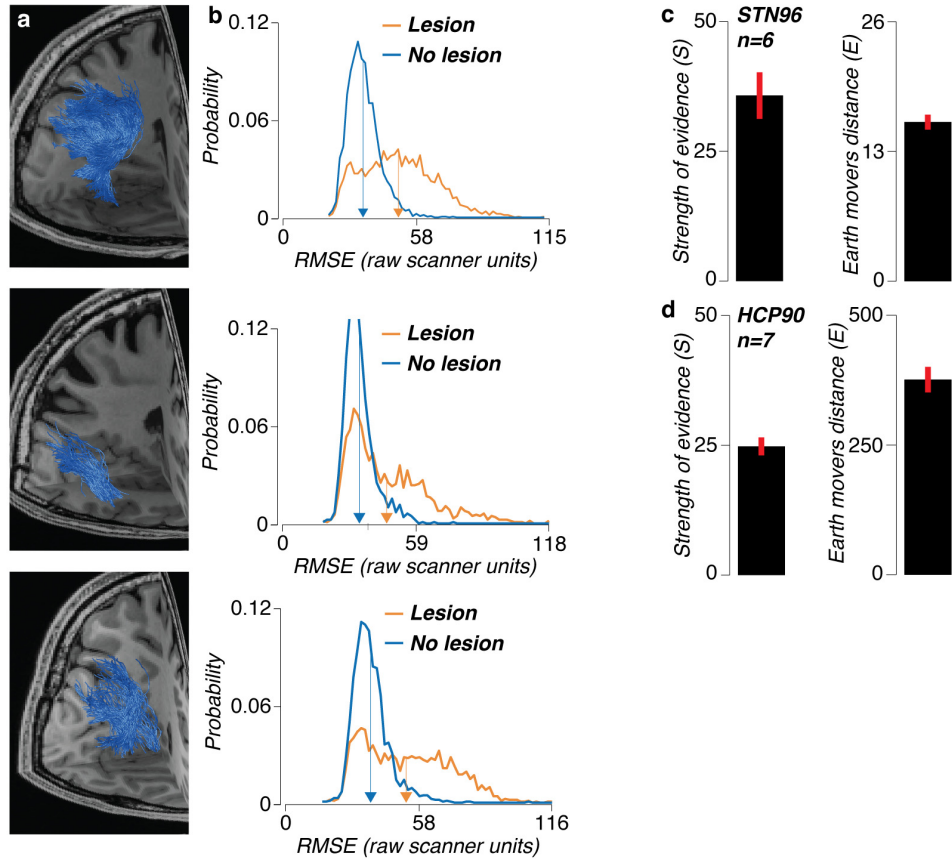
**Supplementary Figure 7**

**A novel pathway connecting hMT+ and the superior parietal gyrus (SPG), additional subjects.**

**a.** White matter fascicles intersecting hMT+ and superior parietal gyrus (SPG) in the right hemisphere of three brains. Dataset STN96.

**b.** Distribution of RMSE for lesioned and unlesioned models. Dataset STN96.

**c and d.** Strength of evidence (S) and Earth Movers Distance (E) for the novel pathway connecting hMT+ and the SPG for datasets STN96 (**c**) and HCP90 (**d**). Error bars indicate ±1 s.e.m. across brains.
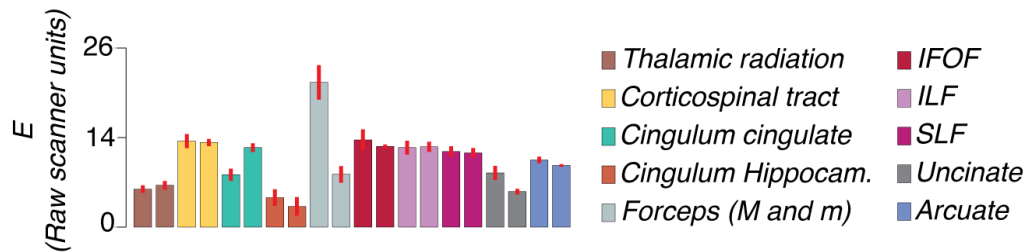
**Supplementary Figure 8**

**Major white-matter fascicles are supported by LiFE.**

Twenty major human white-matter tracts were identified in the optimized connectome (STN96, CSD probabilistic $L_{max}$=10). The bar plot shows the mean Earth Movers Distance ($E$) in support the existence of these tracts. Each bar is the average $E$ (N=5) computed from the left and hemispheres. Error bars are ±s.e.m. across brains. Each bar color represents one tract. The first bar of one color (left-hand) corresponds to the left hemisphere. The second bar of one color (right-hand) corresponds to the right hemisphere. IFOF: Inferior Fronto-occipital Fasciculus. ILF: Inferior Lateral Fasciculus. SLF: Superior Lateral Fasciculus. $E$ values for Forceps Major (M, left-hand bar) and Minor (m, right-hand bar) are presented next to each other.

Putative connection between V1 and the anterior commissure
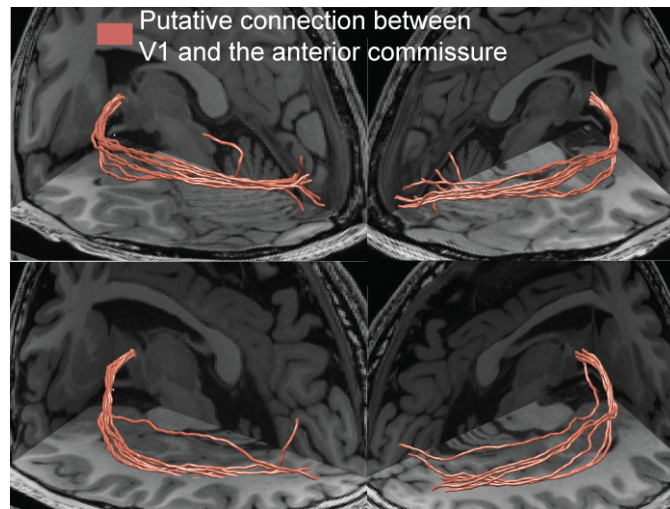
**Supplementary Figure 9**

**LiFE alters the matrix of connections in the candidate connectome.**

The candidate connectome contains fascicles projecting from the anterior commissure to V1 (orange). LiFE eliminates these fascicles from the optimized connectome; they receive zero weights. Because of these zero-weights, the error distributions of the connectome with and without these connections are identical (not shown), indicating that the measured diffusion data provides no support for this connection.

Many investigators use tractography to derive the matrix of connections between brain structures. Here, we illustrate a case in which LiFE alters the matrix of connections in the candidate connectome. The case we illustrate concerns an interesting potential connection between primary visual cortex (V1) and the anterior commissure.

Human anatomical measurements show that the anterior commissure principally connects with most of the temporal lobe, amygdala, and auditory cortex (Eran et al., 2003; Commentary 2.1, page 48). Post-mortem studies of lesioned human brains, which are rare and of relatively low precision, suggest that individuals with lesions around V1 have reduced volume in the anterior commissure. This indirect measure suggest that there is a projection from V1 into the anterior commissure (DiVirgilio et al., 1999). Against this hypothesis, resecting the anterior in macaque does not affect responses of V1 neurons (see Eran et al., 2003; Chapter 8, Page 209).

In ten hemispheres (STN96), we find that probabilistic tractography candidate connectomes include tracts between V1 and the anterior commissure (Supplemental Figure 5b). These tracts are about 10 cm long with mean fractional anisotropy of $0.42\pm0.2$ and mean diffusivity of $0.9\pm0.5$ um$^2$/msec. These are reasonable values, and investigators who rely only on tractography might decide to accept these tracts. But in each of the ten hemispheres, the optimized connectome does not include these tracts. The weights assigned to the tracts are zero and the error distributions for connectomes that do or do not contain these tracts are identical. This example shows that LiFE does more than eliminate redundant tracts. Rather, it can reject false positives. In this example, LiFE eliminates an entire pathway for which the diffusion data provide inadequate support.

Eran, Z. & Iacoboni, M. *The Parallel Brain: The Cognitive Neuroscience of the Corpus Callosum*. (MIT Press, 2003).

Di Virgilio Gabrielle, Clarke, S., Pizzolato, G. & Schaffner, T. Cortical regions contributing to the anterior commissure in man. *Experimental Brain Research* **124,** 1–7 (1999).

| Toolbox | Reference | Web reference |
|---------|-----------|---------------|
| MRtrix | Tournier et al. (2012) | http://www.nitrc.org/projects/mrtrix/ |
| TrackVis | Wang and Wedeen (2007) | http://www.trackvis.org/ |
| Camino | Cook et al., (2008) | http://cmic.cs.ucl.ac.uk/camino/ |
| ExploreDTI | Leemans et al. (2009) | http://www.exploredti.com/generalinfo.htm |
| ConTrack | Sherbondy et al. (2008) | http://vistalab.stanford.edu/newlm/index.php/ConTrack |
| FDT | Behrens et al. (2003) | http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT |
| DTIStudio | Jiang and Mori (2003) | https://www.mristudio.org/ |
| Tracula | Yendiki et al. (2011) | http://ftp.nmr.mgh.harvard.edu/fswiki/Tracula |

**Supplementary Table 1. Publicly available software to generate connectomes.**

A collection of publicly available software to generate, analyze and visualize white-matter connectomes.