

Population-based tract-to-region connectome of the human brain and its hierarchical topology

Received: 15 November 2021

Accepted: 5 August 2022

Published online: 22 August 2022

 Check for updates

Fang-Cheng Yeh ^{1,2} 

Connectome maps region-to-region connectivities but does not inform which white matter pathways form the connections. Here we constructed a population-based tract-to-region connectome to fill this information gap. The constructed connectome quantifies the population probability of a white matter tract innervating a cortical region. The results show that ~85% of the tract-to-region connectome entries are consistent across individuals, whereas the remaining (~15%) have substantial individual differences requiring individualized mapping. Further hierarchical clustering on cortical regions revealed dorsal, ventral, and limbic networks based on the tract-to-region connective patterns. The clustering results on white matter bundles revealed the categorization of fiber bundle systems in the association pathways. This tract-to-region connectome provides insights into the connective topology between cortical regions and white matter bundles. The derived hierarchical relation further offers a categorization of gray and white matter structures.


Mapping the human connectome is the key to understanding how brain structure gives rise to functions and how brain diseases cause dysfunctions^{1,2}. Studies have used structural or functional connectivity to quantify the region-to-region connectivity as the connectome^{3,4} and delineate the network topology of the nervous system. The network topology revealed by the brain connectome further informed the functional implications of cortical regions and enabled graphical theoretical analysis⁵. However, the conventional region-to-region connectome is agnostic of the role played by white matter pathways and does not indicate which pathways form the cortical connections. Consequently, for many neuroscience studies investigating region-to-region connectivity, the white matter is still a black box with much unknown that needs further exploration.

Here we mapped the tract-to-region connectome to address this information gap. The connection probability between white matter pathways and cortical regions was evaluated on 1065 young adult subjects. For m brain regions and n white matter bundles, the tract-to-region connectome can be quantified by an m -by- n matrix, where each matrix entry records the corresponding population probability of a white matter pathway innervating a cortical region.

Several technical advances were used to construct the tract-to-region connectome (Fig. 1). The white matter bundles of the 1065 young adults were mapped using automated tractography. Although many automated tractography methods are available^{6–11}, most have used cortical parcellation to recognize tracts. These region-based methods could lead to circular analysis in the tract-to-region connectome. Therefore, this study used trajectory-based recognition¹² and did not filter tractogram by brain regions. After trajectory-based recognition, connections substantially deviated from the expert-verified tractography atlas were removed.

Four tract-to-region connectome matrices were quantified using the Brodmann parcellation, Kleist parcellation¹³, the Human Connectome Project's multimodal parcellations (HCP-MMP)¹⁴, and a random parcellation, respectively. The tract-to-region information provided by this approach could complete the circuit diagram for many structure-function models and inform the likelihood of a white matter lesion causing a functional deficit in the dysconnectome studies. Further hierarchical clustering was applied to the tract-to-region connectome. The clustering results revealed the hierarchical relation of cortical regions and white matter pathways that informed their categorization.

¹Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA. ²Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA.

 e-mail: frank.yeh@pitt.edu

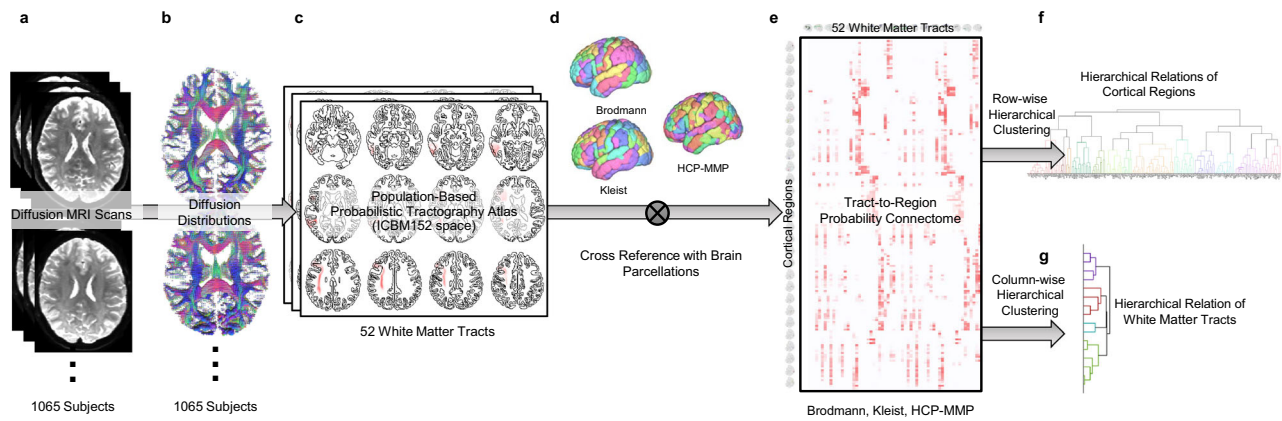


Fig. 1 | The processing flow to construct a population-based tractography connectome and derive its hierarchical relation. **a** The diffusion MRI data of 1065 subjects were used. **b** The data were reconstructed to calculate the diffusion distribution for fiber tracking. **c** For each subject, 52 white matter bundles were mapped using automated tractography. The track recognition was based on trajectory similarity with a tractography atlas without using the cortical parcellations. The tracking results were aggregated to construct a population-based probability

atlas of 52 white matter pathways. **d** Cortical regions from cortical parcellations and the white matter trajectories of each subject were used to derive the connectome matrix. **e** The results from each subject were accumulated to construct the tract-region connectome based on population probability. **f** Hierarchical clustering was applied to the row vectors of the connectome to derive the hierarchical relation of cortical regions. **g** Hierarchical clustering was applied to the column vectors to derive the hierarchical relation of white matter bundles.

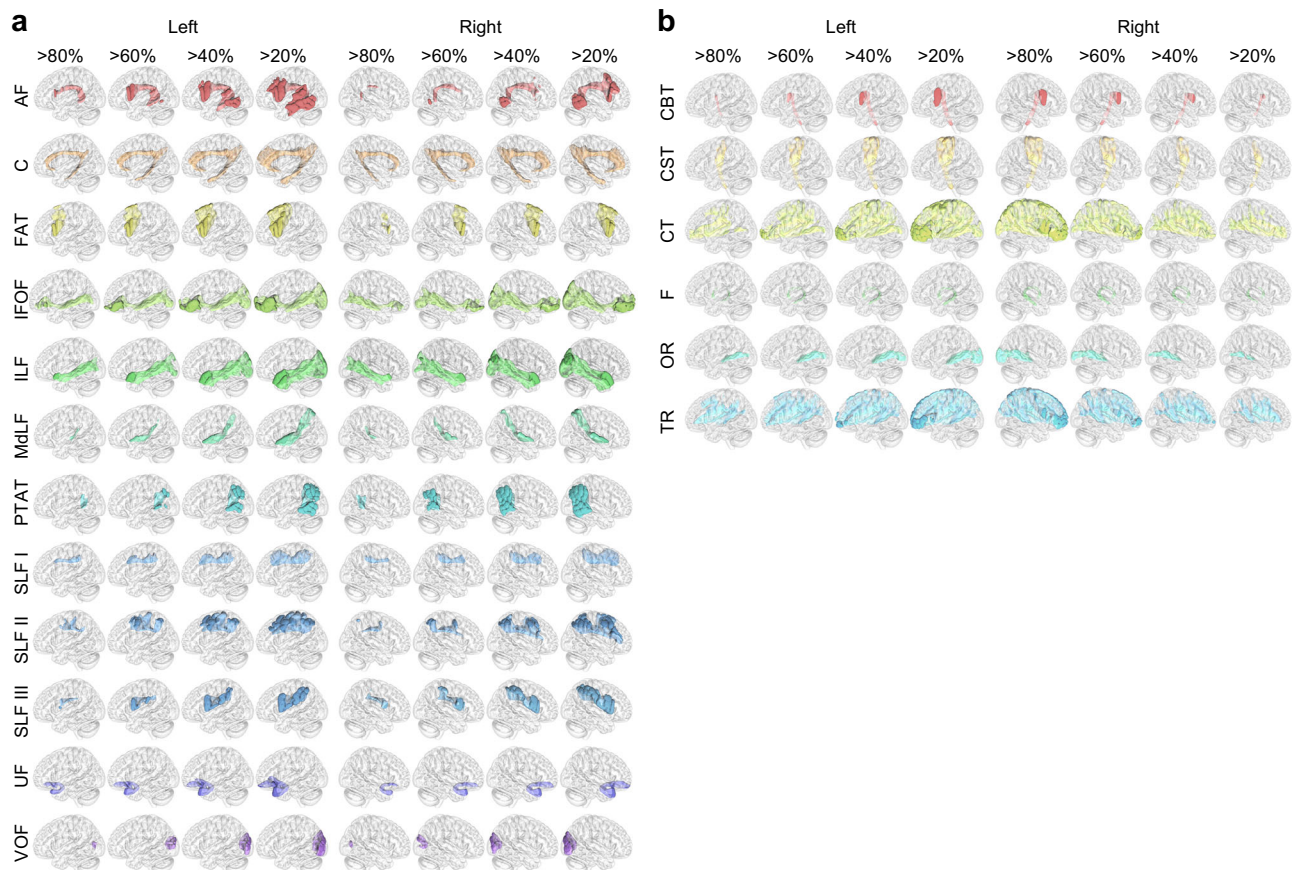


Fig. 2 | Probabilistic tractography atlas of white matter pathways visualized at population probability of 20, 40, 60, and 80%. **a** Association pathways are

visualized at different population probabilities. **b** Projection pathways are visualized at different population probabilities.

Results

Population-based tractography of young adults

We examined the tractography of 1065 subjects in the ICBM152 space. Figure 2a shows the voxel-wise probability of the association pathways, whereas Fig. 2b shows the projection pathways. Each white matter tract is visualized by a population

probability of 20, 40, 60, and 80%, respectively. The probability was quantified by the percentage of subjects with the white matter bundle passing the ICBM152 space voxels. The abbreviations of white matter bundles are listed in Supplementary Table 1. The tractography results shown in Fig. 2 are consistent with known neuroanatomy¹⁵ and existing tractography results^{9,16,17}. The lateralization of left arcuate

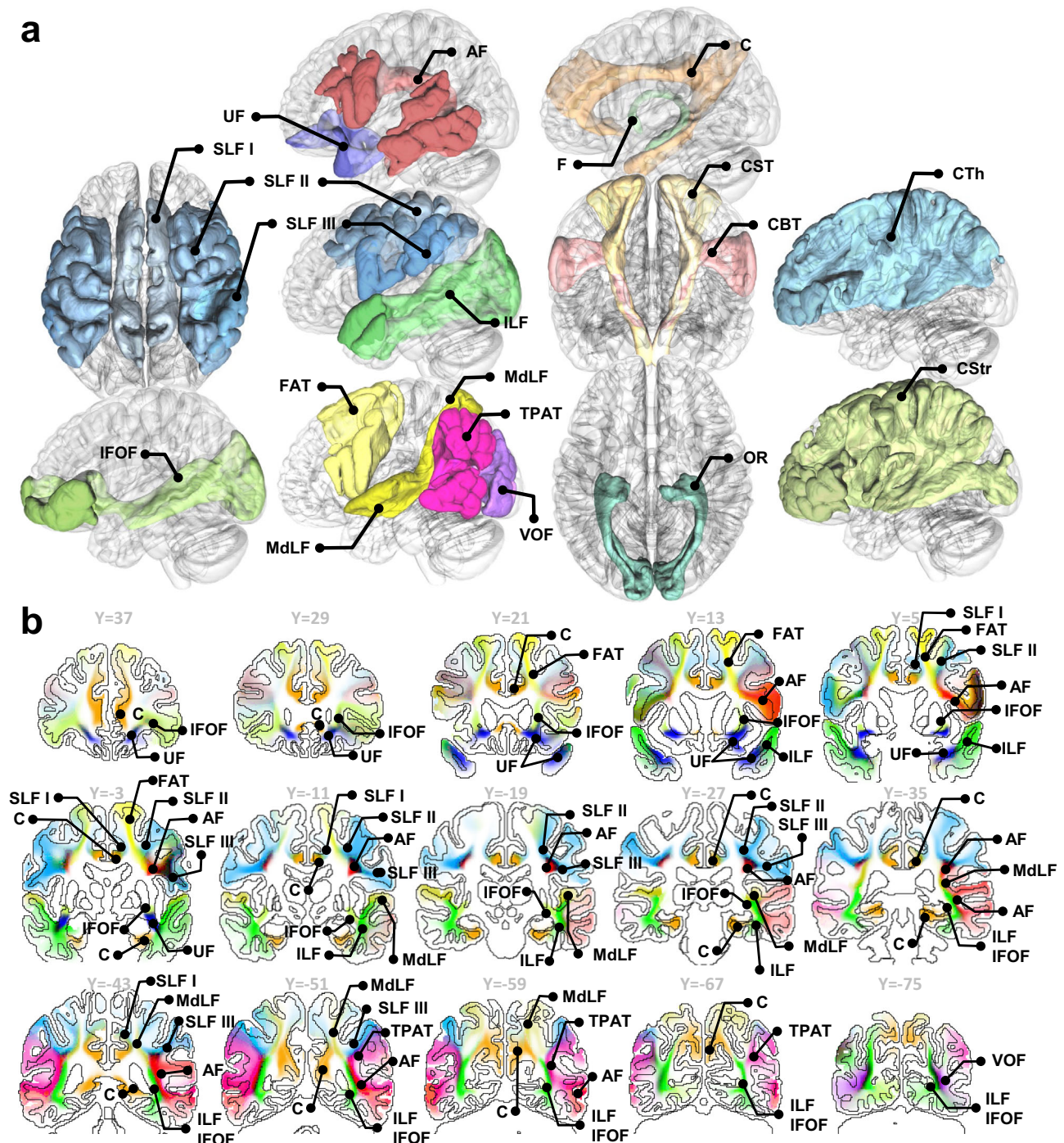


Fig. 3 | Overview of the population-based tractography atlas in 3D rendering and slice-wise coronal sections. **a** White matter pathways are visualized using 20% population probability. **b** Coronal sections of association pathways in the

ICBM152 space. The color intensity scales with the population probability of the white matter bundles in the young adult population.

fasciculus (AF) can be readily observed by its substantially larger volume.

Figure 3a visualizes all white matter bundles rendered by an isosurface of 20% population probability in the ICBM152 space. The AF and superior longitudinal fasciculus (SLF) in Fig. 3 show broader coverage than those of the projection pathways such as corticospinal tract (CST), corticobulbar tract (CBT), optic radiation (OR), and fornix (F). This result can be explained by higher between-subject differences in AF and SLF¹². Figure 3b further shows coronal sections of tract probability. The maximum color saturation corresponds to 100% population probability, whereas white color corresponds to 0% population

probability. The probabilities of association pathways are visualized to illustrate their anatomical relation and relative location. The results are consistent with a recent population-based tractography¹⁷.

Tract-to-region connectome

The tract-to-region connectome based on Brodmann areas and Kleist parcellations are shown in Fig. 4a and Fig. 4b, respectively, whereas the one based on HCP-MMP parcellations is shown in Fig. 5. The Brodmann, Kleist, and HCP-MMP parcellations have 39, 49, and 180 cortical regions. The resulting connectivity matrices have 39-by-52, 49-by-52, and 180-by-52 entries for Brodmann, Kleist, and HCP-MMP

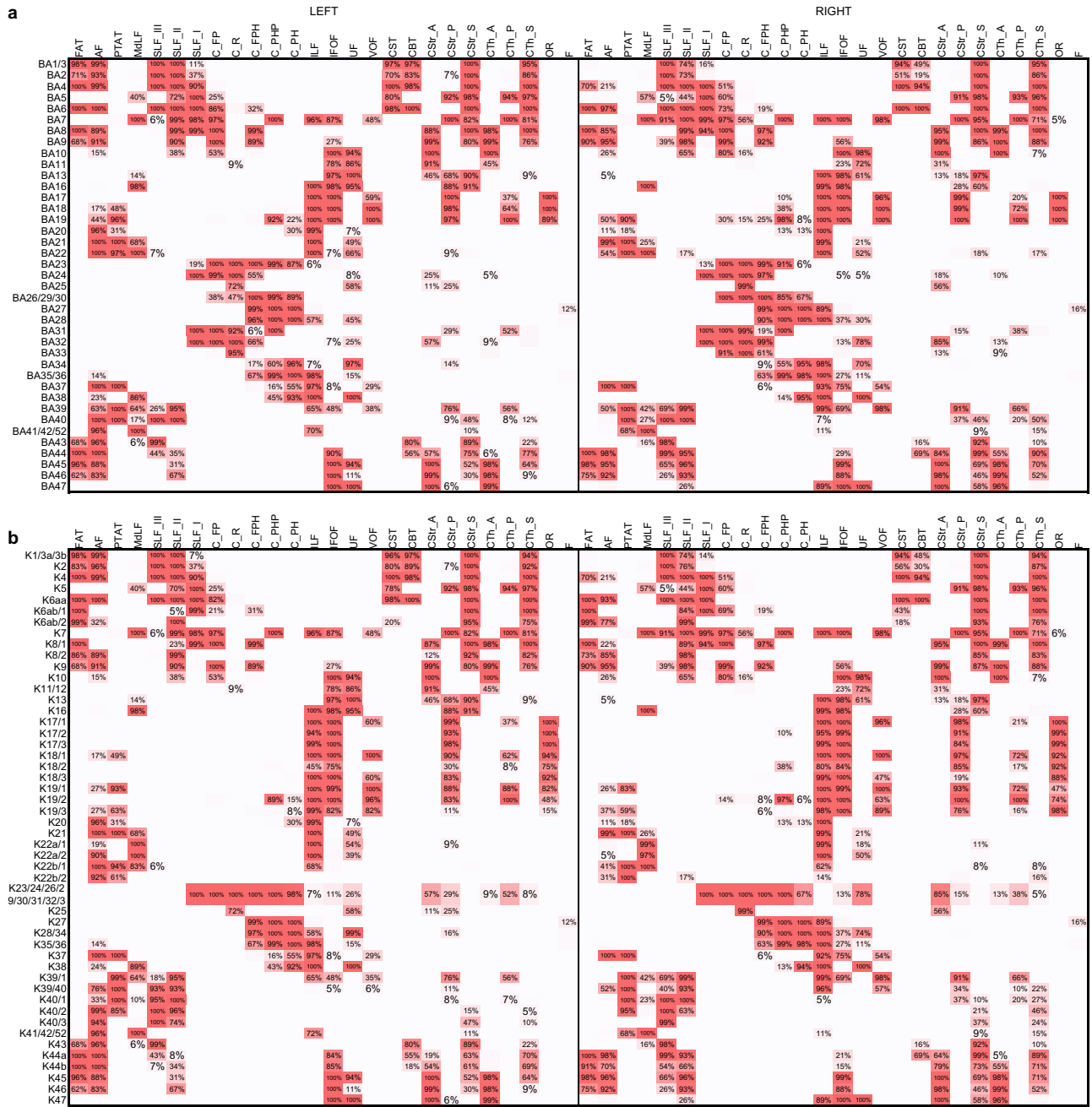


Fig. 4 | The tract-to-region connectome matrices derived from the Brodmann and Kleist brain parcellations, respectively. a The tract-to-region connectome using the Brodmann parcellation. **b** The tract-to-region connectome using the Kleist parcellations. The rows of the matrices correspond to each brain region defined by cortical parcellations, whereas the columns correspond to each white

matter bundle. The tract-to-region connectome matrices show the population probability quantified from 1065 young adults. Probability values lower than 5% were left blank to facilitate inspection. Source data are provided as a Source Data file.

parcellations. Each row of the matrices corresponds to a cortical region, and each column corresponds to a white matter tract. The population probability was quantified by checking the corresponding cortical region and white matter tract intersection in the ICBM152 space. The left half of the matrix is the connection probabilities in the left hemisphere, whereas the right half is those in the right hemisphere. The population probabilities are color-coded by red colors: the highest saturation corresponds to the highest probability (100%), while the white color corresponds to the lowest probability (0%). The entries with less than 5% connection probability are left blank to facilitate visualization.

Most matrix entries in Brodmann (1733 out of 2028 entries, 85.45%) and Kleist parcellations (2164 out of 2548 entries, 84.93%) have population probabilities greater than 95% or smaller than 5%, meaning that these tract-to-region pairs are either connected or not connected in the majority of the study population. Around 15% of the matrix entries in Brodmann and Kleist parcellations show probabilities between 5 and 95% due to substantial individual variation. Interestingly, although HCP-MMP have more parcellation regions (180 regions) than Brodmann and Kleist parcellations (39 and 49 regions), it gives a remarkably similar figure. A total of 86.50% of its matrix entries (8096 out of 9360 entries) also have probability values greater than 95% or smaller than 5%. A

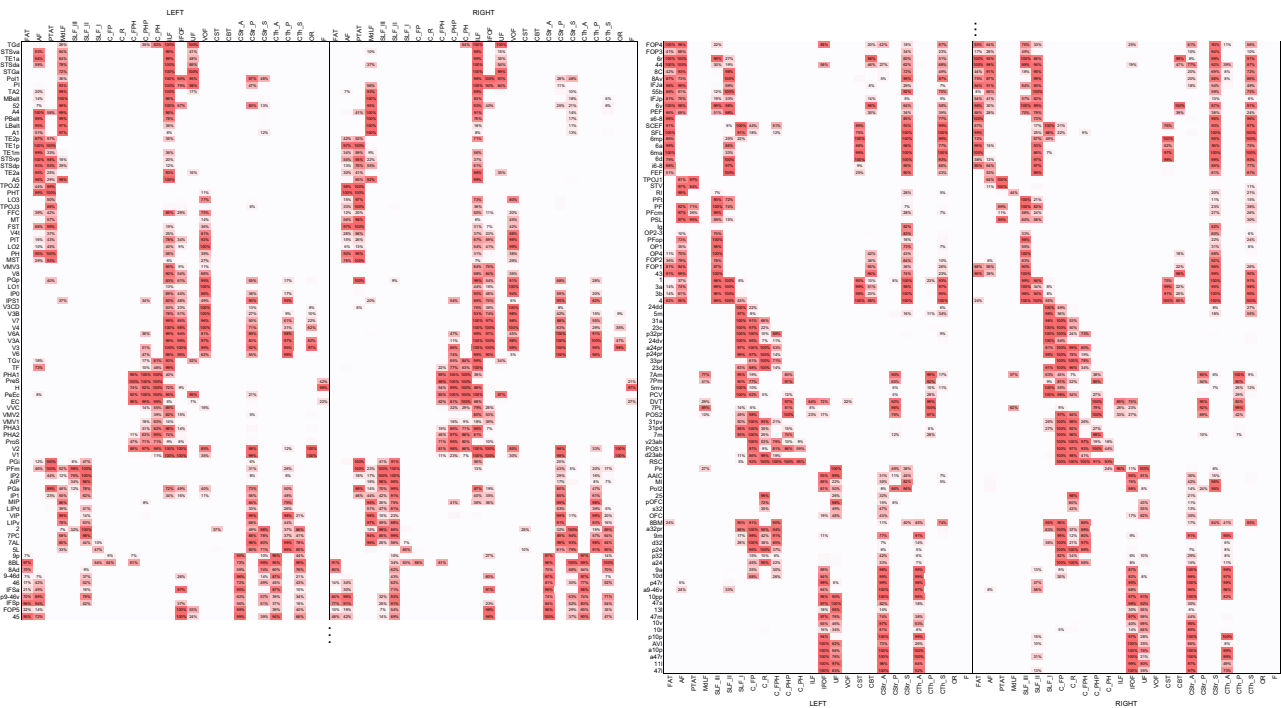


Fig. 5 | The probabilistic tract-to-region connectome matrices derived from the HCP multimodal parcellation. The 180 rows of the matrices correspond to each brain region defined by the HCP cortical parcellations, whereas the columns

correspond to each white matter bundle. Source data are provided as a Source Data file.

random parcellation with 360 regions derived from Craddock's random parcellations¹⁸ also showed a similar result (Supplementary Fig. 1): 87.09% of the matrix entries (8152 out of 9360) had probability values ranging between 95 and 5%. Overall, the tract-to-region connectome showed that the young adult population shares a similar connective pattern in ~85% of the tract-to-region entries. The remaining ~15% entries have substantial individual variations with population probability between 5 and 95%, thus warranting individualized mapping.

We further examined AF connections in the tract-to-region connectome using the Sankey flow diagrams shown in Fig. 6. The diagram is based on HCP-MMP, and the color saturation scales with the population probability. The connective pattern shown in Fig. 6 is consistent with the conventional view that the left AF connects Wernicke's area in the superior temporal regions (red) and Broca's area in the inferior frontal cortex (orange). Furthermore, the diagrams also show more detailed connections of left AF to the caudal dorsolateral prefrontal cortex and the inferior parietal lobule^{19–23} as well as premotor/motor regions¹⁹. The lateralization of AF to frontoparietal (yellow), angular (green), and superior temporal regions (red) can also be seen by comparing Fig. 6a and Fig. 6b.

Hierarchical relation of cortical regions

Figure 7 shows the similarity matrices and the derived hierarchical relations of the cortical regions defined by Brodmann (Fig. 7a) and Kleist parcellations (Fig. 7b), whereas the results for HCP-MMP are shown in Fig. 8. The column and row positions of the matrices are reordered based on clustering results to facilitate inspection. The dendrograms on the top of Figs. 7 and 8 show the hierarchical relation of the cortical regions and the vertical distance scales by the cost for merging. Overall, Figs. 7 and 8 show a consistent result, with cortical regions categorized into dorsal, ventral, and limbic networks. Although differences can be observed at each parcellation, the dorsal network includes most frontal (excluding prefrontal) and parietal regions, whereas the ventral network includes temporal and occipital regions. The limbic network comprises the prefrontal, insula, and

upper cingulum regions. In Brodmann parcellation (Fig. 7a), its dorsal network further includes the superior temporal gyrus. In contrast, the dorsal network in Kleist (Fig. 7b) and HCP-MMP (Fig. 8) only include a small posterior section of the superior temporal gyrus. The discrepancy is likely due to more detailed parcellation in Kleist and HCP-MMP at areas 22, 39, and 40. The detailed parcellation in HCP-MMP allows for revealing the subnetworks under the dorsal network, including frontal (orange colored), inferior parietal (yellowish and light green), and superior parietal (cyan) subnetworks (Supplementary Fig. 2). Similarly, HCP-MMP shows subnetworks under the ventral networks, including occipital (purple and light blue), inferior temporal (magenta), superior temporal (light red) subnetworks (Supplementary Fig. 3). The limbic networks are primarily consistent across Brodmann, Kleist, and HCP-MMP. The subnetworks cover prefrontal regions and cingulum, bridging the dorsal and ventral networks. More detailed subnetworks are shown in Supplementary Fig. 4.

Hierarchical relation of white matter bundles

Figure 9 further shows the similarity matrix between the association pathways (Fig. 9a), and the dendrogram illustrates the hierarchical clustering results (Fig. 9b) based on the HCP-MMP tract-to-region connectome. The left and right hemispheres show highly similar hierarchical relations that group association pathways into four systems, including the arcuate system (purple), anterior ventral system (red), posterior ventral system (cyan), and cingulum system (green). The first system includes AF, SLF II, SLF III, and FAT. These pathways all connect to Broca's area and have correlated with language functions shown by several studies (detailed in the Discussion section). The second system includes MdLF, TPAT, VOF, and ILF. TPAP has several alternative naming, such as the posterior AF, posterior SLF, or SLF-TP (Supplementary Table 1). The third system includes UF and IFOF, and both are characterized by their frontal connection from the temporal and occipital lobes, respectively. The fourth system includes all cingulum pathways and SLF I, likely because the SLF I is closely adjacent to the cingulum at ($Y = -3$ and $Y = -11$) and entirely separated from SLF II

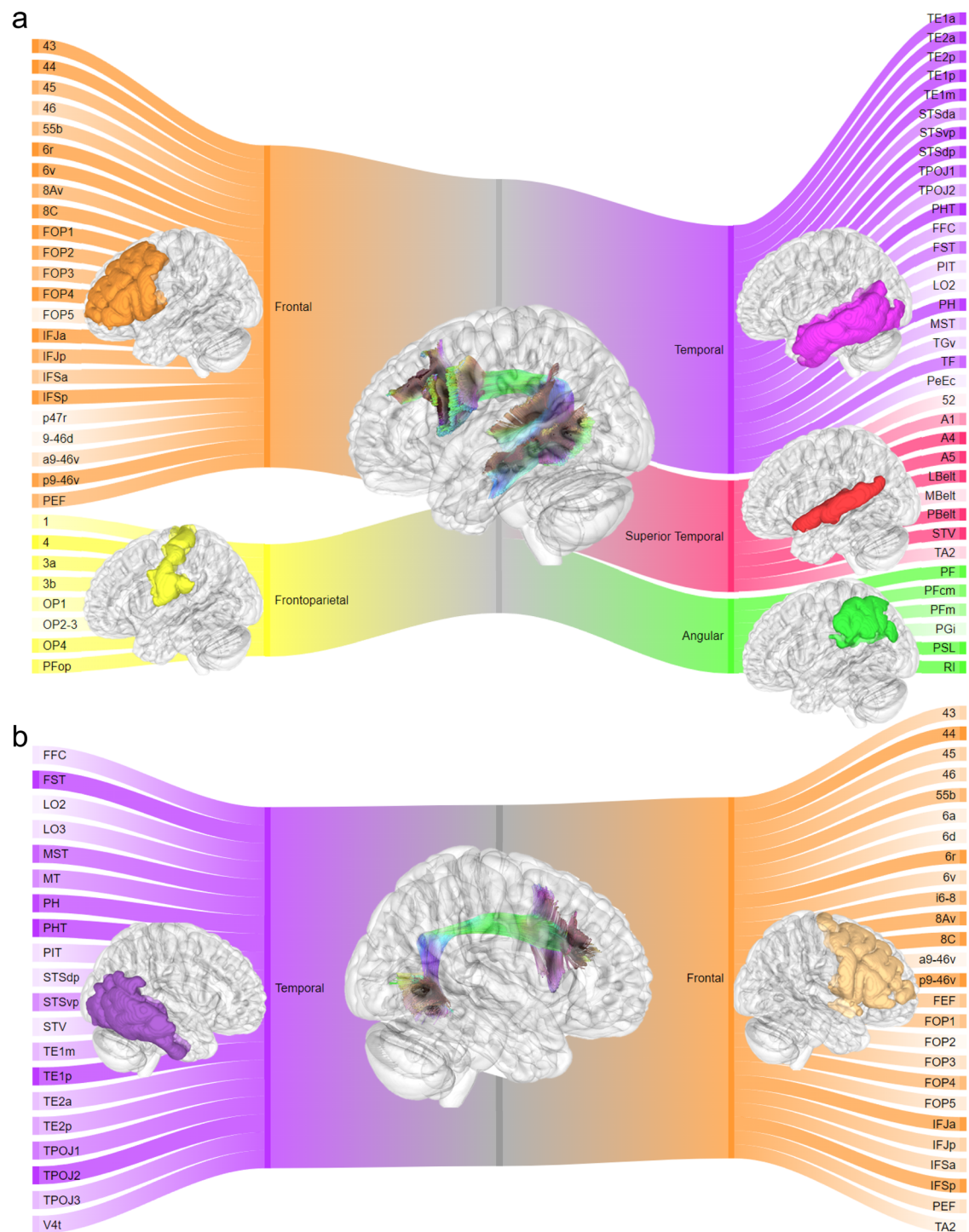


Fig. 6 | The tract-to-region connective pattern of the arcuate fasciculus shown by Sankey flow diagrams. a The connective pattern of the left arcuate fasciculus. **b** The connective pattern of the right arcuate fasciculus. The diagrams are based on the population probability calculated from the tract-to-region connectome in Fig. 5.

The color saturation scales with the connection probability. The left arcuate fasciculus shows substantially lateralized connections to frontal-parietal (yellow), angular (green), and superior temporal regions (red).

and III by FAT (Fig. 3b). The above data-driven clustering results showed the relation between white matter pathways based on their similarity in the tract-to-region connectome.

Discussion

Here we quantify the tract-to-region connectome in the young adult population. The constructed matrices provide a resource for both neuroscience and clinical studies to evaluate the probability of a white

matter tract connecting to a cortical region. Based on this tract-to-region connectome, we further applied hierarchical clustering between cortical regions and between white matter association pathways to understand their relations. The results in cortical regions revealed dorsal, ventral, and limbic networks, especially using more detailed parcellations such as HCP-MMP.

The dorsal, ventral, and limbic networks can be applied to many existing functional models. The dorsal system includes most frontal

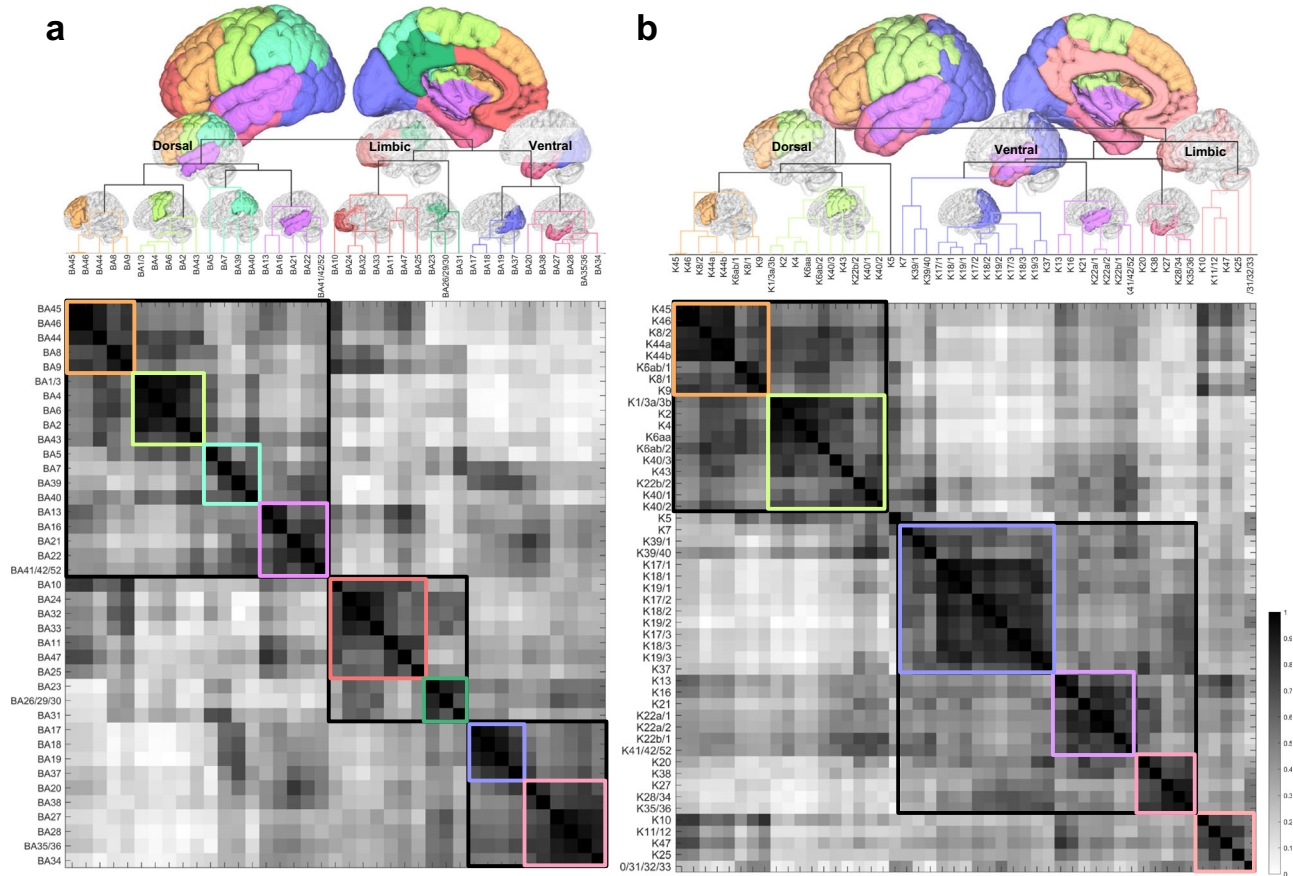


Fig. 7 | Similarity matrices between cortical regions and derived dendrograms based on Brodmann and Kleist parcellations, respectively. a The similarity matrix based on Brodmann parcellations. **b** The similarity matrix based on Kleist parcellations. The similarity matrices were calculated by nonparametric Spearman correlation between the row vectors of the connectome matrices. The hierarchical relation of cortical areas is then visualized using dendrograms computed by

hierarchical clustering. The vertical distances in the dendrograms are scaled with the clustering cost. Both dendrograms show grossly consistent results revealing the limbic, dorsal, and ventral networks. Source data are provided as a Source Data file. MATLAB (MathWorks®, <https://www.mathworks.com/>) was used to create the diagram and matrix.

lobe (excluding prefrontal) and parietal lobe, whereas the ventral system includes the temporal lobe and occipital lobe. The dorsal and ventral subnetworks shared many similarities with the existing dual-stream models in the language and visual functions^{24–27}. The tract-to-region connectome matrices and all intermediate data, such as tractogram of each bundle, read-to-track subject data, and source code, are publicly available at <https://brain.labsolver.org>. The shared data could also construct the conventional region-to-region connectome for each white matter pathway, as illustrated in our previous connectome study²⁸.

Neuroanatomical evidence has shown that brain regions in both human and non-human primates could be connected through more than one route²⁹. Existing studies have predominately focused on region-to-region connections and simplified the role of white matter bundles as “edges” in the network model. Consequently, the region-to-region connectome falls short of illustrating the association between cortical regions and white matter pathways. This limitation becomes obvious in lesion-symptom mapping studies of aphasia: damaging Broca’s area does not necessarily lead to Broca’s aphasia^{30,31}, but lesions involving the anterior segment of the left AF are a strong symptom predictor³². Since cortical regions themselves may not be sufficient to explain functional deficits³³, the role of white matter pathways should be considered in dysconnectome studies^{34–36}. For these studies, the tract-to-region connectome could provide a population-based reference to understand the relation between cortical regions and white matter bundles.

The tract-to-region connectome can be utilized in various scenarios in which the white matter tracts are the targets of interest. For clinical cases involving a lesion in deep white matter structures, the population probability quantified by the tract-to-region connectome can provide the likelihood of an affected white matter pathway leading to functional deficits in a cortical region. Conversely, in fMRI, EEG, or SEEG studies identifying a cortical region of interest, the tract-to-region connectome can translate the findings to their corresponding white matter pathways based on population probabilities. Subsequently, the information can delineate the circuit mechanism behind a cognitive model or verify the structure-function hypotheses. Both of them may further help explore white matter targets for neurological modulations using deep brain stimulation, focused ultrasound ablation, or laser ablation.

The tract-to-region and region-to-region connectome provide different perspectives on the organization of brain networks, as shown by their different clustering results. The clustering on the region-to-region connectome concerns whether two cortical regions are closely connected^{37,38}, but clustering on the tract-to-region connectome concerns whether two cortical regions share similar white matter connections. As a result, the region-to-region connectome tends to group frontal and prefrontal regions due to their strong connections through short association pathways^{38,39}. In comparison, the results from the tract-to-region connectome using three differential parcellations unanimously separated prefrontal and frontal regions due to their distinctly different connections to the limbic and dorsal networks. The

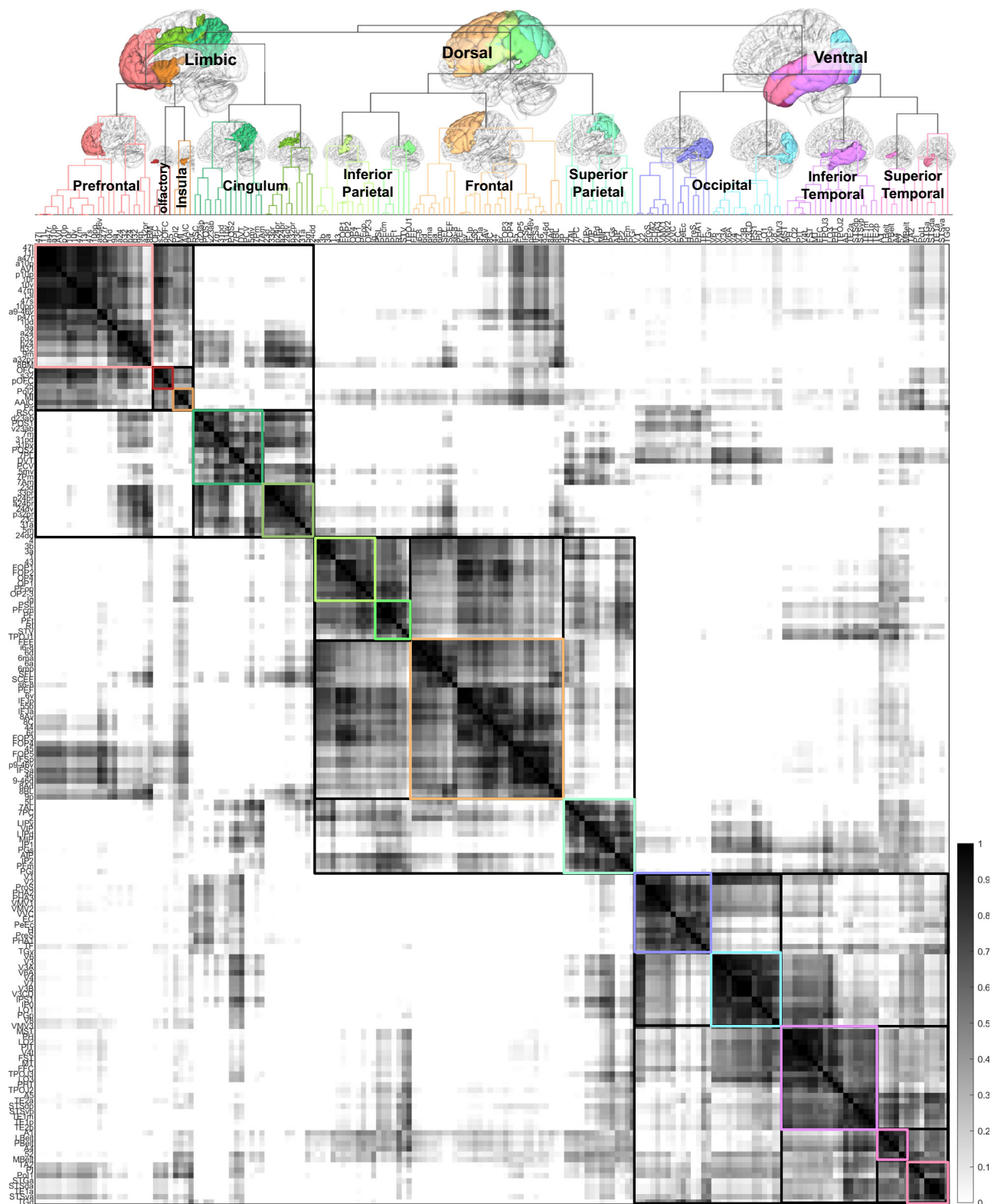


Fig. 8 | Similarity matrix between cortical regions and its derived dendrogram based on HCP multimodal parcellation. Consistent with the previous figure's results derived from Brodmann and Kleist parcellations, the cortical regions can be clustered into limbic, dorsal, and ventral networks. The limbic network includes the limbic system, prefrontal cortex, olfactory cortex, and insula. The dorsal network includes most of the remaining frontal lobe, parietal lobe, and part of the superior

temporal gyrus, whereas the ventral network includes most of the temporal and occipital lobe. Each network has its downstream hierarchical structures of the subnetworks. Source data are provided as a Source Data file. MATLAB (MathWorks®, <https://www.mathworks.com/>) was used to create the diagram and matrix.

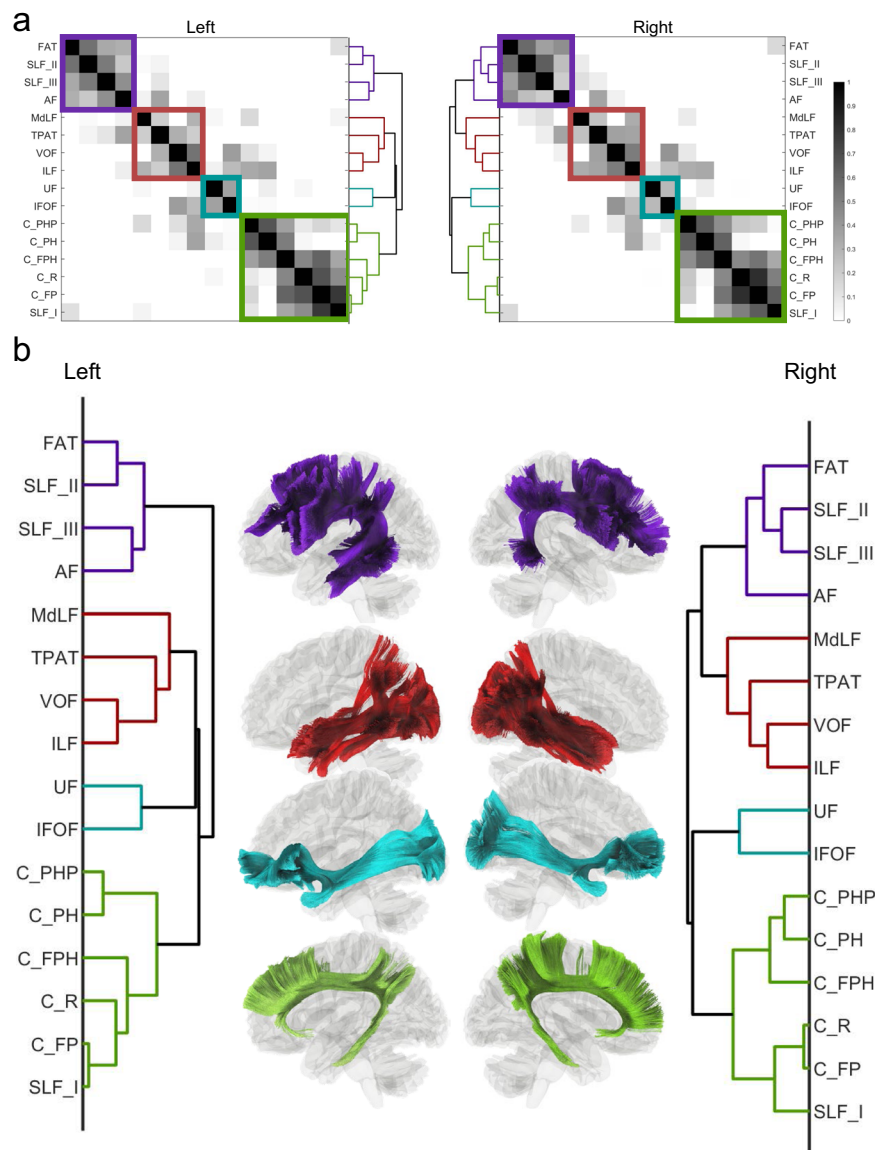


Fig. 9 | Similarity matrices between association pathways and their derived dendrograms based on HCP multimodal parcellation. **a** The similarity matrices were calculated by nonparametric Spearman correlation between the column vectors of the connectome matrix. **b** The hierarchical relation is then visualized using dendrograms computed by using hierarchical clustering. The horizontal distance in the dendrograms scales with the clustering cost. Both dendrograms

show four categories of association pathways on both hemispheres, including the cingulum system (green), posterior ventral system (cyan), anterior ventral system (red), and arcuate system (purple). Source data are provided as a Source Data file. MATLAB (MathWorks®, <https://www.mathworks.com/>) was used to create the diagram and matrix.

prefrontal regions are closely connected with the limbic network through the cingulum, whereas frontal regions are closely connected with the dorsal network through SLF II and SLF III. The difference in clustering context will lead to entirely different results and application scenarios that answer different neuroscience questions.

We also derived the hierarchical relation between white matter bundles. While many studies have been conducted to cluster white matter tracts^{40–45}, these clustering methods did not consider the connective pattern with the cortical regions. The clustering in this study showed that SLF II and SLF III are closely related to AF, whereas SLF I is closely related to the cingulum. These results may appear questionable and astonishing at first glance, but there are supporting references: Catani et al.²⁰ showed SLF III as the anterior segment of the AF, which did not include SLF I. Wang et al.⁴⁶ suggested that the SLF I could be viewed as part of the cingulum system. From the clinical perspective, especially in the surgical intervention of brain tumors, the

neurosurgical consensus is that the eloquent area correlated with post-surgical functional deficits includes regions innervated by AF, SLF II, and SLF III^{47–49}. These areas did not include SLF I because SLF I did not show significant language function⁵⁰. Furthermore, SLF I was delineated by anterograde tract-tracer technique in rhesus macaques⁵¹, where detailed mapping was illustrated by Schmahmann and Pandya's work⁵². In their work, among 15 cases enhancing the SLF I (cases 1, 2, 3, 4, 6, 7, 9, 17, 22, 26, 27, 28, 29, 31, and 33), 12 of them (except cases 26, 27, and 28) also enhanced the cingulum bundle as labeled by the authors. In comparison, only two cases (cases 4 and 31) enhanced SLF II, and three (cases 6, 7, and 33) enhanced SLF III. This hinted at a closer relationship between the SLF I and cingulum than with the SLF II or SLF III. Nonetheless, it is noteworthy that the clustering in this study was based on the tract-to-region connectome entries, and by no means could this be used to confirm a new naming convention or provide a new neuroanatomical definition as each bundle has well-defined

anatomical locations, as shown in Fig. 2. More functional or lesion-based studies are needed to support or refute these clustering results.

There are limitations to this study. The tract-to-region connectome did not include cerebellar, commissural, brainstem, or connections between subcortical structures. Excluding cerebellar and brainstem pathways are due to different slices coverage near the brainstem and insufficient spatial resolution to generate reliable fiber tracking results. The commissural pathways are excluded because of the limited ability of the fiber tracking methods to resolve crossing-kissing patterns when the corpus callosum crosses the hemispheres⁵³. Moreover, for bundles mapped in this study, the tract-to-region relation was determined by a simple overlap in the binary mask. This setting was used to compensate for tractography's "gyral bias" that failed to map connections at the "gyral bank." However, there could be spurious connections because tractography could not confirm innervation. For example, the connection probabilities between IFOF and insula regions (e.g., Pol1 and 52) were just due to IFOF passing by, and further histology validation is needed. On top of these limitations, the included tracts were also subject to errors such as trajectory deviation, premature termination, and incorrect routing, as discussed in a recent review⁵⁴. Although multiple strategies have been leveraged to reduce false connections, we cannot rule out possible errors in the current form of the tract-to-region connectome.

Furthermore, the existing graph-theoretical analysis⁵ may not be readily applicable to the tract-to-region connectome because the tract-to-region concept is conceptually different from the conventional region-to-region one. While the region-to-region connectome implies an undirected graph, the tract-to-region matrix appears to be a subset of an undirected graph called a bipartite graph. The bipartite graph comprises two disjoint sets of nodes, one for tracts and one for regions. The nodes representing tracts could not be equally exchanged with nodes representing regions. Since most network measures view all nodes equally in the computation, applying these network analyses to the tract-to-region connectome could lead to questionable results. Further theoretical development is thus required to translate network measures to the tract-to-region connectome.

This study mainly focuses on the concept of the tract-to-region connectome, and the variations due to different tracking recognition tools were not investigated. The tract-to-region relation could be derived using different tools or atlases, but additional customizations may be needed to address unique technical concerns when deriving the tract-to-region relation. Specifically, tract segmentation tools often used cortical parcellations^{8,55} or end regions of tracts¹⁰ to recognize a tract. The tracts defined by cortical parcellation would show connections according to the supplied cortical parcellations, leading to circular results in the tract-to-region connectome. A solution is to use white matter trajectories to recognize tracts, but trajectory-based methods may not effectively utilize all existing atlases^{9,56,57}. Most white matter atlases were voxel-based volumes that provided only masks of tract coverage and did not have trajectory coordinates needed by trajectory-based recognition. Few atlases provide trajectory coordinates for individual subjects, but a group average would be needed to minimize the individual differences. This averaging step is critical for classifiers that are sensitive to noisy data.

Nonetheless, studies have shown that tractography segmentation could be different due to anatomical views⁵⁸ or segmentation tools⁵⁹. The differences due to tools were direct results of different input data: region-based recognitions used cortical parcellations, while trajectory-based recognition used the topology of white matter tracts. On the other hand, the differences in anatomical views were mainly due to discrepancies in the existing categorical systems and tract nomenclature^{58,60}. Much of the recent disputes focus on detailed subcomponents and their categorical relations⁶¹. Resolving them would need new neuroanatomical evidence from tract-tracing or cadaver dissection studies. The population-based tractography and its

corresponding tract-to-region connectome are thus subject to future revisions and updates to ensure their up-to-date accuracy. Despite those discrepancies, it is noteworthy that the anatomical locations of white matter pathways are well-defined, and this study did not invent new white matter structures. The pathways mapped in this study (e.g., those shown in Figs. 2 and 3) are anatomically consistent with existing atlases from other tools and studies, and the clustering results reported were consistent across three different cortical parcellations. This consistency may support future works extending tract-to-region mapping to lifespan studies. To this end, ready-to-track data for HCP-aging, HCP-developmental, and developing HCP studies and sample processing scripts are available at <https://brain.labsolver.org> to assist further brain mapping endeavors.

Methods

Diffusion MRI acquisition

The diffusion MRI data of 1065 subjects (Fig. 1a) were acquired from the Human Connectome Project database (WashU consortium)². The age range was 22–37 years, and the average age was 28.75 years. The data were acquired using a multishell diffusion scheme with three b-values at 1000, 2000, and 3000 s/mm² with 90 sampling directions for each shell. The spatial resolution was 1.25 mm isotropic. The detailed acquisition parameters are listed in the consortium paper². The preprocessed data were used. The gradient nonlinearity was corrected for each diffusion-weighted signal at each voxel by $S' = b_0 \left(\frac{S}{b_0} \right)^{1/\|\mathbf{G}\cdot\mathbf{b}\|^2}$,⁶² where S is the raw signal, and S' is the corrected diffusion signal. b_0 is the b0 signal, and \mathbf{b} the diffusion gradient direction. \mathbf{G} is the 3-by-3 gradient nonlinearity matrix after adding an identity matrix. This per signal correction allows for keeping the original shell structures of the b-table to enable shell-based diffusion modeling.

Diffusion MRI reconstruction

The diffusion data were linearly rotated to align with the ac-pc line of the ICBM152 space and simultaneously interpolated at 1 mm using cubic spline interpolation. The b-table was also rotated accordingly. The rotated data were then reconstructed using generalized q-sampling imaging⁶³ with a diffusion sampling length ratio of 1.7. An automatic quality control routine was adopted to check the b-table orientation and ensure its accuracy⁶⁴. The reconstruction results (Fig. 1b) were further used in automated tractography.

Automated tractography

For each subject, 52 white matter bundles were mapped using the automated tractography pipeline in DSI Studio (<http://dsi-studio.labsolver.org>, developed by the author), which combined deterministic fiber tracking algorithm⁶⁵, randomized parameter saturation¹², topology-informed pruning⁶⁶, and trajectory-based tract recognition¹² (detailed in the next section) as an integrated interface. The default settings were used: the anisotropy threshold was uniformly and randomly selected from 0.5 to 0.7 Otsu threshold. The angular threshold was uniformly and randomly selected from 15 to 90 degrees. The step size was uniformly and randomly selected from 0.5 to 1.5 voxel spacing. The minimum length was 30 mm. The tracking was repeated until the ratio of streamlines to voxels reached 1.0.

Trajectory-based tract recognition

The tract recognition used the nearest neighbor method to classify tracts. Since the nearest neighbor classifier is sensitive to noisy data and prone to overfitting, training data preparation was critical for best performance. Most population-based atlases have substantial individual variations and thus would require additional averaging to minimize individual differences. Therefore, in this study, the recognition used a population-averaged tractography atlas²⁸, which was aggregated from the young adult population and vetted by a team of

neuroanatomists. An updated version of the atlas in the ICBM152 nonlinear asymmetry space (publicly available at <https://brain.labsolver.org>) was used in this study.

For each subject, the tractography atlas was nonlinearly warped to subject space using the diffeomorphic mapping derived between the subject's anisotropy image and the ICBM152-space anisotropy image. The Hausdorff distance⁶⁷ was computed between each subject and the atlas tract. The shortest distance then determined the label of subject tracts. Some tracts might substantially deviate from all atlas trajectories, and thus a maximum allowed Hausdorff distance (termed tolerance distance) of 16 was used to remove them. The minimum Hausdorff distance was increased to 18 and 20 mm if no bundle was found after topology-informed pruning⁶⁶. In this study, 4 out of the 52 bundles did not reach a 100% yield rate for all 1065 subjects: left CBTs (983/1065), right CBTs (1054/1065), right AF (1046/1065), right occipital corticopontine tract (1064/1065). The missing of the corticobulbar and corticopontine tracts in some subjects could be due to the limitation of the fiber tracking algorithm to capture substantial turning of the pathways. The right AF in some subjects was entirely labeled as SLF III due to no connection to the superior temporal lobe.

The steps mentioned above, including subject-space fiber tracking, parameter saturation, randomized parameters, topology-informed pruning, anisotropy-based warping, Hausdorff distance computation, were integrated as the automated tractography function in DSI Studio. The source code is also available on GitHub repository at <https://github.com/frankyeh/DSI-Studio>.

The computation was conducted at the Pittsburgh Supercomputing Center provided through the Extreme Science and Engineering Discovery Environment (XSEDE) resource⁶⁸. The tractography result for each subject and each white matter tract are shared on <http://brain.labsolver.org>.

Brain parcellations and tract-to-region connectome

The white matter bundles of each subject were then exported to the ICBM152 2009 nonlinear space to facilitate integration across the entire subject group. As shown in Fig. 1d, for each of the 1065 subjects, the trajectories of white matter bundles in the ICBM152 space were examined with the Brodmann, Kleist, and HCP-MMP parcellations to derive the population-based tract-to-region connectome. A random parcellation was also used as a comparison.

We used the ICBM152-space version of newly reconstructed Brodmann and Kleist atlases¹³. On the other hand, the ICBM152-space version of HCP-MMP was obtained from <https://neurovault.org/collections/1549/> (asymmetric, improved reconstruction) and further inspected for each cortical region to manually remove the cross-sulci leakage using DSI Studio. The revised version of the HCP-MMP was shared with the DSI Studio package and is publicly available at <http://dsi-studio.labsolver.org>. The random parcellation was derived from level 33 of Craddock's fine-grained random parcellations¹⁸ by assigning the labels to a gray matter mask in the ICBM152 space using the shortest distance.

A binary tract-to-region connection matrix was obtained for each subject by calculating the intersection between the voxel-wise mapping of white matter bundles and cortical regions (Fig. 1e). The binary matrices of 1065 subjects were then aggregated to compute the population probability of the tract-to-region connection, and one matrix was generated for each parcellation, respectively. The tract-to-region connectome can be downloaded from <http://brain.labsolver.org>.

Hierarchical clustering

We used row vectors of the tract-to-region matrices as the feature vectors to derive the hierarchical relation between cortical regions (Fig. 1f). The similarity matrices between each region pair were quantified by their correlation. Since the correlation could be nonlinear, we

used the nonparametric Spearman's rank correlation to consider possible nonlinearity relations in population probability. The hierarchical clustering was conducted using weighted average distance⁶⁹ provided by the "linkage" function in MATLAB to avoid the high variability drawback of simple single linkage clustering. For each cortical parcellation (Brodmann, Kleist, HCP-MMP), a dendrogram was generated to reveal the hierarchical relation of the cortical regions. The hierarchical clustering for white matter bundles was conducted using the column vector of the HCP-MMP connectome matrices (Fig. 1g). The clustering routine also used weighted average distance to generate the dendrogram to reveal the hierarchical relation of white matter bundles.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The tract-to-region matrices, similarity matrices, clustering code, and scripts to generate identical figures are provided in the "Source Data.zip". The population-based tractography and tract-to-region connectome are publicly available at <http://brain.labsolver.org>. Source Data are provided with this paper.

Code availability

The analysis tool DSI Studio⁷⁰ and its source code are available at <https://github.com/frankyeh/DSI-Studio>.

References

1. Akil, H., Martone, M. E. & Van Essen, D. C. Challenges and opportunities in mining neuroscience data. *Science* **331**, 708–712 (2011).
2. Glasser, M. F. et al. The Human Connectome Project's neuroimaging approach. *Nat. Neurosci.* **19**, 1175–1187 (2016).
3. Sporns, O., Tononi, G. & Kotter, R. The human connectome: a structural description of the human brain. *PLoS Comput Biol.* **1**, e42 (2005).
4. Biswal, B. B. et al. Toward discovery science of human brain function. *Proc. Natl Acad. Sci. USA* **107**, 4734–4739 (2010).
5. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
6. O'Donnell, L. J. et al. Automated white matter fiber tract identification in patients with brain tumors. *Neuroimage Clin.* **13**, 138–153 (2017).
7. Rheault, F. et al. Bundle-specific tractography with incorporated anatomical and orientational priors. *Neuroimage* **186**, 382–398 (2019).
8. Zollei, L., Jaimes, C., Saliba, E., Grant, P. E. & Yendiki, A. TRActs constrained by UnderLying INfant anatomy (TRACULInA): an automated probabilistic tractography tool with anatomical priors for use in the newborn brain. *Neuroimage* **199**, 1–17 (2019).
9. Zhang, F. et al. An anatomically curated fiber clustering white matter atlas for consistent white matter tract parcellation across the lifespan. *Neuroimage* **179**, 429–447 (2018).
10. Wasserthal, J., Neher, P. & Maier-Hein, K. H. TractSeg – fast and accurate white matter tract segmentation. *Neuroimage* **183**, 239–253 (2018).
11. Maffei, C. et al. Using diffusion MRI data acquired with ultra-high gradient strength to improve tractography in routine-quality data. *Neuroimage* **245**, 118706 (2021).
12. Yeh, F. C. Shape analysis of the human association pathways. *Neuroimage* **223**, 117329 (2020).
13. Pijnenburg, R. et al. Myelo- and cytoarchitectonic microstructural and functional human cortical atlases reconstructed in common MRI space. *Neuroimage* **239**, 118274 (2021).

14. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
15. Catani, M. & de Schotten, M. T. *Atlas of Human Brain Connections* (Oxford University Press, 2012).
16. Thiebaut de Schotten, M. et al. Atlas location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *Neuroimage* **54**, 49–59 (2011).
17. Hansen, C. B. et al. Pandora: 4-D white matter bundle population-based atlases derived from diffusion MRI fiber tractography. *Neuroinformatics* **19**, 447–460 (2021).
18. Craddock, R. C., James, G. A., Holtzheimer, P. E. III, Hu, X. P. & Mayberg, H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* **33**, 1914–1928 (2012).
19. Bernal, B. & Ardila, A. The role of the arcuate fasciculus in conduction aphasia. *Brain* **132**, 2309–2316 (2009).
20. Catani, M., Jones, D. K. & ffytche, D. H. Perisylvian language networks of the human brain. *Ann. Neurol.* **57**, 8–16 (2005).
21. Glasser, M. F. & Rilling, J. K. DTI tractography of the human brain's language pathways. *Cereb. Cortex* **18**, 2471–2482 (2008).
22. Rilling, J. K. et al. The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* **11**, 426–428 (2008).
23. Saur, D. et al. Ventral and dorsal pathways for language. *Proc. Natl Acad. Sci. USA* **105**, 18035–18040 (2008).
24. Fridriksson, J. et al. Revealing the dual streams of speech processing. *Proc. Natl Acad. Sci. USA* **113**, 15108–15113 (2016).
25. Milner, A. D. & Goodale, M. A. Two visual systems re-viewed. *Neuropsychologia* **46**, 774–785 (2008).
26. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11800–11806 (2000).
27. Mishkin, M., Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417 (1983).
28. Yeh, F. C. et al. Population-averaged atlas of the macroscale human structural connectome and its network topology. *Neuroimage* **178**, 57–68 (2018).
29. Ghulam-Jelani, Z. et al. Redundancy circuits of the commissural pathways in human and rhesus macaque brains. *Hum. Brain Mapp.* **42**, 2250–2261 (2021).
30. Gajardo-Vidal, A. et al. Damage to Broca's area does not contribute to long-term speech production outcome after stroke. *Brain* **144**, 817–832 (2021).
31. Ardila, A., Bernal, B. & Rosselli, M. Why Broca's area damage does not result in classical Broca's aphasia. *Front. Hum. Neurosci.* **10**, 249 (2016).
32. Fridriksson, J., Guo, D., Fillmore, P., Holland, A. & Rorden, C. Damage to the anterior arcuate fasciculus predicts non-fluent speech production in aphasia. *Brain* **136**, 3451–3460 (2013).
33. Salvalaggio, A., De Filippo De Grazia, M., Zorzi, M., Thiebaut de Schotten, M. & Corbetta, M. Post-stroke deficit prediction from lesion and indirect structural and functional disconnection. *Brain* **143**, 2173–2188 (2020).
34. Thiebaut de Schotten, M., Foulon, C. & Nachev, P. Brain disconnections link structural connectivity with function and behaviour. *Nat. Commun.* **11**, 5094 (2020).
35. Griffis, J. C., Metcalf, N. V., Corbetta, M. & Shulman, G. L. Lesion Quantification Toolkit: a MATLAB software tool for estimating grey matter damage and white matter disconnections in patients with focal brain lesions. *Neuroimage Clin.* **30**, 102639 (2021).
36. Greene, C. et al. Finding maximally disconnected subnetworks with shortest path tractography. *Neuroimage Clin.* **23**, 101903 (2019).
37. Sohn, Y., Choi, M. K., Ahn, Y. Y., Lee, J. & Jeong, J. Topological cluster analysis reveals the systemic organization of the *Caenorhabditis elegans* connectome. *PLoS Comput. Biol.* **7**, e1001139 (2011).
38. Akiki, T. J. & Abdallah, C. G. Determining the hierarchical architecture of the human brain using subject-level clustering of functional networks. *Sci. Rep.* **9**, 19290 (2019).
39. Raut, R. V., Snyder, A. Z. & Raichle, M. E. Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proc. Natl Acad. Sci. USA* **117**, 20890–20897 (2020).
40. Siless, V., Chang, K., Fischl, B. & Yendiki, A. AnatomicCuts: hierarchical clustering of tractography streamlines based on anatomical similarity. *Neuroimage* **166**, 32–45 (2018).
41. Garyfallidis, E. et al. Recognition of white matter bundles using local and global streamline-based registration and clustering. *Neuroimage* **170**, 283–295 (2018).
42. Guevara, P. et al. Robust clustering of massive tractography datasets. *Neuroimage* **54**, 1975–1993 (2011).
43. Jin, Y. et al. Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics. *Neuroimage* **100**, 75–90 (2014).
44. Maddah, M., Mewes, A. U., Haker, S., Grimson, W. E. & Warfield, S. K. Automated atlas-based clustering of white matter fiber tracts from DTMRI. *Med. Image Comput. Comput. Assist. Inter.* **8**, 188–195 (2005).
45. Wang, Q., Yap, P. T., Wu, G. & Shen, D. Application of neuroanatomical features to tractography clustering. *Hum. Brain Mapp.* **34**, 2089–2102 (2013).
46. Wang, X. et al. Subcomponents and connectivity of the superior longitudinal fasciculus in the human brain. *Brain Struct. Funct.* **221**, 2075–2092 (2016).
47. Maldonado, I. L. et al. Surgery for gliomas involving the left inferior parietal lobule: new insights into the functional anatomy provided by stimulation mapping in awake patients. *J. Neurosurg.* **115**, 770–779 (2011).
48. Sanai, N., Martino, J. & Berger, M. S. Morbidity profile following aggressive resection of parietal lobe gliomas. *J. Neurosurg.* **116**, 1182–1186 (2012).
49. Southwell, D. G. et al. Language outcomes after resection of dominant inferior parietal lobule gliomas. *J. Neurosurg.* **127**, 781–789 (2017).
50. Chang, E. F., Raygor, K. P. & Berger, M. S. Contemporary model of language organization: an overview for neurosurgeons. *J. Neurosurg.* **122**, 250–261 (2015).
51. Petrides, M. & Pandya, D. N. Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* **228**, 105–116 (1984).
52. Schmahmann, J. D. & Pandya, D. N. *Fiber Pathways of the Brain* (Oxford University Press, 2009).
53. Maier-Hein, K. H. et al. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* **8**, 1349 (2017).
54. Yeh, F. C., Irimia, A., Bastos, D. C. A. & Golby, A. J. Tractography methods and findings in brain tumors and traumatic brain injury. *Neuroimage* **245**, 118651 (2021).
55. Warrington, S. et al. XTRACT – standardised protocols for automated tractography in the human and macaque brain. *Neuroimage* **217**, 116923 (2020).
56. Oishi, K., Faria, A. V., Zijl, P. C. M. v. & Mori, S. *MRI Atlas of Human White Matter* 2nd edn (Elsevier, 2010).
57. Catani, M. & Thiebaut de Schotten, M. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* **44**, 1105–1132 (2008).
58. Schilling, K. G. et al. Tractography dissection variability: what happens when 42 groups dissect 14 white matter bundles on the same dataset. *Neuroimage* **243**, 118502 (2021).

59. Schilling, K. G. et al. Fiber tractography bundle segmentation depends on scanner effects, vendor effects, acquisition resolution, diffusion sampling scheme, diffusion sensitization, and bundle segmentation workflow. *Neuroimage* **242**, 118451 (2021).
60. Dick, A. S. & Tremblay, P. Beyond the arcuate fasciculus: consensus and controversy in the connective anatomy of language. *Brain* **135**, 3529–3550 (2012).
61. Bullock, D. N. et al. A taxonomy of the brain's white matter: twenty-one major tracts for the 21st century. *Cereb. Cortex* **32**, 1–15 (2022). <https://doi.org/10.1093/cercor/bhab500> (2022).
62. Malkyarenko, D. I. & Chenevert, T. L. Practical estimate of gradient nonlinearity for implementation of apparent diffusion coefficient bias correction. *J. Magn. Reson. Imaging* **40**, 1487–1495 (2014).
63. Yeh, F. C., Wedeen, V. J. & Tseng, W. Y. Generalized q-sampling imaging. *IEEE Trans. Med. Imaging* **29**, 1626–1635 (2010).
64. Schilling, K. G. et al. A fiber coherence index for quality control of B-table orientation in diffusion MRI scans. *Magn. Reson. Imaging* **58**, 82–89 (2019).
65. Yeh, F. C., Verstynen, T. D., Wang, Y., Fernandez-Miranda, J. C. & Tseng, W. Y. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS One* **8**, e80713 (2013).
66. Yeh, F. C. et al. Automatic removal of false connections in diffusion MRI tractography using topology-informed pruning (TIP). *Neurotherapeutics* **16**, 52–58 (2019).
67. Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863 (1993).
68. Towns, J. et al. XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* **16**, 62–74 (2014).
69. Sokal, R. R. A statistical method for evaluating systematic relationships. *Univ. Kans., Sci. Bull.* **38**, 1409–1438 (1958).
70. Yeh, F. *DSI Studio* <https://doi.org/10.5281/zenodo.4978980> (2021).

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1548562. The computation resources were allocated under TG-CIS200026. The author was supported by NIH grant R01 NS120954. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and

by the McDonnell Center for Systems Neuroscience at Washington University.

Author contributions

F.-C.Y. did the data analysis and wrote the manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information The online version contains

supplementary material available at <https://doi.org/10.1038/s41467-022-32595-4>.

Correspondence and requests for materials should be addressed to Fang-Cheng Yeh.

Peer review information *Nature Communications* thanks Klaus Maier-Hein and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022