



## OPEN ACCESS

## EDITED BY

Julio Ernesto Villalon-Reina,  
University of Southern California,  
United States

## REVIEWED BY

David N. Kennedy,  
University of Massachusetts Medical School,  
United States  
Prasanna Parvathaneni,  
Flagship Biosciences, Inc., United States

## \*CORRESPONDENCE

John Kruper  
✉ jk232@uw.edu

RECEIVED 22 February 2024

ACCEPTED 15 May 2024

PUBLISHED 12 June 2024

## CITATION

Kruper J, Hagen MP, Rheault F, Crane I,  
Gilmore A, Narayan M, Motwani K, Lila E,  
Rorden C, Yeatman JD and Rokem A (2024)  
Tractometry of the Human Connectome  
Project: resources and insights.  
*Front. Neurosci.* 18:1389680.  
doi: 10.3389/fnins.2024.1389680

## COPYRIGHT

© 2024 Kruper, Hagen, Rheault, Crane,  
Gilmore, Narayan, Motwani, Lila, Rorden,  
Yeatman and Rokem. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC  
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Tractometry of the Human Connectome Project: resources and insights

John Kruper<sup>1\*</sup>, McKenzie P. Hagen<sup>1</sup>, François Rheault<sup>2</sup>,  
Isaac Crane<sup>3</sup>, Asa Gilmore<sup>1</sup>, Manjari Narayan<sup>4</sup>, Keshav Motwani<sup>5</sup>,  
Eardi Lila<sup>5</sup>, Chris Rorden<sup>6</sup>, Jason D. Yeatman<sup>4</sup> and Ariel Rokem<sup>1</sup>

<sup>1</sup>Department of Psychology, University of Washington, Seattle, WA, United States, <sup>2</sup>Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada, <sup>3</sup>Department of Psychology, University of Chicago, Chicago, IL, United States, <sup>4</sup>Graduate School of Education, Stanford University, Stanford, CA, United States, <sup>5</sup>Department of Biostatistics, University of Washington, Seattle, WA, United States, <sup>6</sup>Department of Psychology, University of South Carolina, Columbia, SC, United States

**Introduction:** The Human Connectome Project (HCP) has become a keystone dataset in human neuroscience, with a plethora of important applications in advancing brain imaging methods and an understanding of the human brain. We focused on tractometry of HCP diffusion-weighted MRI (dMRI) data.

**Methods:** We used an open-source software library (pyAFQ; <https://yeatmanlab.github.io/pyAFQ>) to perform probabilistic tractography and delineate the major white matter pathways in the HCP subjects that have a complete dMRI acquisition ( $n = 1,041$ ). We used diffusion kurtosis imaging (DKI) to model white matter microstructure in each voxel of the white matter, and extracted tract profiles of DKI-derived tissue properties along the length of the tracts. We explored the empirical properties of the data: first, we assessed the heritability of DKI tissue properties using the known genetic linkage of the large number of twin pairs sampled in HCP. Second, we tested the ability of tractometry to serve as the basis for predictive models of individual characteristics (e.g., age, crystallized/fluid intelligence, reading ability, etc.), compared to local connectome features. To facilitate the exploration of the dataset we created a new web-based visualization tool and use this tool to visualize the data in the HCP tractometry dataset. Finally, we used the HCP dataset as a test-bed for a new technological innovation: the TRX file-format for representation of dMRI-based streamlines.

**Results:** We released the processing outputs and tract profiles as a publicly available data resource through the AWS Open Data program's Open Neurodata repository. We found heritability as high as 0.9 for DKI-based metrics in some brain pathways. We also found that tractometry extracts as much useful information about individual differences as the local connectome method. We released a new web-based visualization tool for tractometry—"Tractoscope" (<https://nrdg.github.io/tractoscope>). We found that the TRX files require considerably less disk space—a crucial attribute for large datasets like HCP. In addition, TRX incorporates a specification for grouping streamlines, further simplifying tractometry analysis.

## KEYWORDS

brain, MRI, diffusion MRI, tractometry, Open Data, heritability, predictive modeling, data visualization

## 1 Introduction

The long-range connections between different brain areas that form the human macro-scale connectome are essential to the distribution and integration of information in the brain (Bassett and Sporns, 2017). Healthy brain connections are also important for mental and neurological health (Bassett and Bullmore, 2009). The Human Connectome Project (HCP) is a pioneering effort to study the structure and function of the brain macro-scale connectome. The WU-Minn-Ox consortium of the HCP pursued this effort by collecting a large dataset of 1,200 young adult twin and non-twin siblings that included extensive measurements of structural (T1-weighted and T2-weighted), functional (both with a task and without one—i.e., at “rest”) and diffusion-weighted MRI (dMRI), in addition to genotype information and behavioral testing. Some of the subjects also underwent additional electrophysiological measurements and additional MRI measurements at 7T.<sup>1</sup> Rather than relying on the state of the art of MRI measurements at the time that the project was initiated, the HCP advanced the field forward, developing a large number of novel techniques for data acquisition, data processing and analysis, and created novel ways to organize and disseminate the data. This effort has generated a dataset that even now, more than a decade after the project started, stands out in its high quality and uniformity of measurement, and in the large value that the research community has drawn from it. Thus, the HCP has become a keystone dataset in human neuroscience, with more than 1,500 papers that acknowledge using the data, as of 2021 (Elam et al., 2021). Its approach serves as a source of inspiration to a large number of HCP-style follow-up studies (Glasser et al., 2016), including studies targeting life-span development (Bookheimer et al., 2019; Howell et al., 2019), and several different projects targeting specific clinical populations (e.g., Demro et al., 2021).

Measurements of dMRI in the HCP dataset leveraged several technical innovations. These included use of specialized hardware, and particularly of a strong and fast set of gradients, with a maximal gradient strength of 100 mT/m, and effective slew rate of 91 mT/m/s. Parallel imaging techniques that use multi-slice and multi-band excitation were used to accelerate the acquisition of each volume (Setsompop et al., 2012). This enabled measurements in a large number of different directions, with multiple different non-zero b-values (distributed in three shells of  $b \approx 1,000\text{s/mm}^2$ ,  $b \approx 2,000\text{s/mm}^2$ ,  $b \approx 3,000\text{s/mm}^2$ ), and with a high spatial resolution of  $1.25 \times 1.25 \times 1.25\text{mm}^3$ . In addition to these advanced acquisition techniques, HCP developed novel processing methods to address artifacts due to motion and eddy currents, and to address geometric distortions due to susceptibility. Thus, the HCP produced data that far exceeds, in terms of spatial and angular resolution, what is possible in most clinical settings, even a decade later. Therefore, these dMRI data provide unique views of the human white matter connectome.

Tractometry analysis of dMRI data focuses on the physical properties of major white matter pathways. It uses computational tractography and anatomical constraints to delineate the locations of known anatomical tracts in dMRI data, and extracts brain white matter tissue properties along the length of each tract (Yeatman et al., 2012). Tractometry provides important information about brain tissue properties and individual differences, but for large and important datasets, such as the HCP, applying cutting-edge tractometry methods requires specialized expertise, and is also very computationally demanding. The present work enables the study of brain connections in the HCP dataset by providing tractometry results in 1,041 subjects in HCP that have completed a full set of dMRI measurements and by building a set of insights and resources based on this data. In each subject in the dataset, 24 major white matter pathways were identified using the pyAFQ software (<https://yeatmanlab.github.io/pyAFQ>) (Table 1). We used probabilistic tractography to delineate the tracts and diffusion kurtosis imaging (DKI) (Jensen et al., 2005) as implemented in the open-source software DIPY (<https://dipy.org>) (Garyfallidis et al., 2014; Henriques et al., 2021) to describe white matter tissue properties along their lengths. DKI was used because it extends diffusion tensor imaging (DTI) (Basser et al., 1994), providing a more complete assessment of diffusion by measuring the deviation of the diffusion patterns from a Gaussian distribution. In addition, in previous work, we have also shown that DKI describes the HCP dMRI data more accurately and more reliably than DTI (Henriques et al., 2021). Here, we also used an extension of DKI that models biophysical white matter tissue properties (Fieremans et al., 2011) to provide additional information about the axonal white matter fraction along the length of the major white matter pathways. The results of this processing are all provided openly through the AWS Open Data program in the Open Neurodata repository (Vogelstein et al., 2018), and we provide an example of how to access this data.

We used this open dataset as a platform to examine several different aspects of the data. First, we characterized the overall distribution of tissue properties along the length of the white matter pathways that we delineated. We also used the presence of a large number of monozygotic and dizygotic twins in the sample to characterize the heritability of DKI tissue properties along the length of the tracts. Finally, we compared the predictive ability of tract profiles to other diffusion processing methods. Tract profiles of tissue properties can be used to compare different subject groups or in order to understand individual differences (Jones et al., 2005; Colby et al., 2012; Yeatman et al., 2012; Dayan et al., 2016; Richie-Halford et al., 2021). However, high-dimensional data with limited observations can challenge the accuracy of out-of-sample predictions, providing motivation to understand if there is any loss of predictive information with the dimensionality reduction provided by tract profiles. In a previous study (Rasero et al., 2021), brain-behavior correlations were assessed using the local connectome (LC) method (Yeh et al., 2016), which calculates a q-space normalized map of the density of spins between neighboring locations along tracts. The resulting feature sets from each method differ in their dimensionality—tract profiles for every standard tract results in several thousand features, while LC results in hundreds of thousands of features. In the present study, we compared the information provided by LC to the much more concise information provided in tractometry tract profiles. Open access to a standard HCP tractometry dataset will facilitate future

<sup>1</sup> This consortium was based on a collaboration between groups at Washington University, the University of Minnesota, and Oxford University; for brevity, we will refer to this consortium as “HCP” henceforth, acknowledging that another important consortium, the MGH-UCLA consortium, pursued a different and also important approach (McNab et al., 2013).

TABLE 1 Abbreviations used for the tracts saved in both the TRK and TRX format.

Tract abbreviation	Formal tract name
ATR_L	Left anterior thalamic
ATR_R	Right anterior thalamic
CST_L	Left corticospinal
CST_R	Right corticospinal
CGC_L	Left cingulum cingulate
CGC_R	Right cingulum cingulate
FP	Forceps major
FA	Forceps minor
IFO_L	Left inferior fronto-occipital
IFO_R	Right inferior fronto-occipital
ILF_L	Left inferior longitudinal
ILF_R	Right inferior longitudinal
SLF_L	Left superior longitudinal
SLF_R	Right superior longitudinal
UNC_L	Left uncinate
UNC_R	Right uncinate
ARC_L	Left arcuate
ARC_R	Right arcuate
Orbital	Orbital corpus callosum
AntFrontal	Anterior frontal callosum
SupFrontal	Superior frontal callosum
Motor	Motor corpus callosum
SupParietal	Superior parietal corpus callosum
PostParietal	Posterior parietal corpus callosum
Occipital	Occipital corpus callosum
Temporal	Temporal Corpus Callosum

research aimed at comparing additional methods for analysis of brain behavior correlations.

Following the long-standing tradition of the HCP, our development of HCP tractometry results provides a platform for developing and advancing new technologies. We used HCP tractometry as a platform to test TRX, a recently-proposed community-based file format that incorporates the benefits of several previously-developed file formats for tractography, and that advances several new innovative features (Rheault et al., 2022). In the present work, we used HCP tractometry to test the computational efficiency of TRX and its potential to conserve storage space, while retaining important information about tract profile features. Finally, interactive web-based visualization tools for exploring large datasets lower the barrier for fruitful interaction with these datasets, and serve as a point of entry for researchers who are considering how to use the data (Keshavan and Poline, 2019). In previous work, we developed AFQ-Browser (<https://yeatmanlab.github.io/AFQ-Browser>), an application that enables exploration of tractometry datasets (Yeatman et al., 2018), but the previously presented tool was limited in terms of its ability to explore the anatomical structure of each individual subject in the dataset. The recent development of the NiiVue software library enables much more facile visualization of anatomical data (Hanayik

et al., 2024), including both volumetric and tractography datasets and their combination. Here, we present Tractoscope (<https://nrdg.github.io/tractoscope>), as the next generation of web-based tools for sharing and exploring tractometry results.

## 2 Methods

### 2.1 Data

Diffusion MRI data was collected by the Human Connectome Project (HCP), as previously described in detail (Sotiropoulos et al., 2013). Briefly, data was acquired on a 3T Siemens Skyra MRI system equipped with a 32-channel coil that was modified to accommodate gradients with  $G_{max} = 100 \text{ mT/m}$  (ultimately, acquisition was conducted with a  $G_{max} = 97.4 \text{ mT/m}$  after optimization for gradient duty cycle). Multislice echo planar imaging with multiband excitation was acquired with a TR of 5.5 s and TE of 89 ms. Three diffusion-weighted shells were acquired:  $b \approx 1,000 \text{ s/mm}^2$ ,  $b \approx 2,000 \text{ s/mm}^2$ ,  $b \approx 3,000 \text{ s/mm}^2$  and the same TR/TE was used in each. In each shell, 90 non-colinear directions were selected, to optimize coverage within and across shells (Caruyer et al., 2013), resulting in the acquisition of 190 data points in each shell, corresponding to measurements in inverse phase encoding direction (LR and LR directions) and five non-diffusion weighted acquisitions. The spatial resolution of the data was  $1.25 \times 1.25 \times 1.25 \text{ mm}^3$ .

We used data provided by HCP that had already been processed using the HCP minimal preprocessing pipelines, as previously described (Glasser et al., 2013). Briefly, intensity normalization was performed across the six acquisition series based on the non diffusion-weighted images ( $b_0$ ). These  $b_0$  images were also used to estimate and correct EPI distortions using the FSL “topup” tool (Andersson et al., 2003). The FSL “eddy” tool was used to correct artifacts due to eddy currents and motion (Andersson and Sotiropoulos, 2016). Gradient spatial non-linearities were computed (Bammer et al., 2003). A spatial transform was calculated between the average  $b_0$  image and the T1-weighted data using FreeSurfer’s “BBRegister” algorithm (Greve and Fischl, 2009). The eddy-corrected data were transformed according to both the gradient nonlinearity correction and T1w registration into 1.25 mm structural volume space in a single step.

We analyzed data from 1,041 subjects from the HCP who had complete measurements of dMRI (i.e., where these measurements passed the HCP quality control process, and also included all 270 diffusion MRI volumes). Among these subjects, the average age was  $28.7 \text{ years} \pm 3.7 \text{ years}$  (standard deviation); 479/562 were male/female.

### 2.2 Tractometry analysis

We applied the pyAFQ pipeline to perform advanced tractometry analysis (Kruper et al., 2021). We used data provided by HCP that had already been pre-processed (Glasser et al., 2013; Sotiropoulos et al., 2013). Using pyAFQ, we fit constrained spherical deconvolution (CSD) and used it as the fiber orientation distribution function for probabilistic tractography implemented

in DIPY (Tournier et al., 2008; Garyfallidis et al., 2014). We used symmetric normalization (SyN) (Avants et al., 2008) diffeomorphic non-linear registration to register subjects to the Montreal Neurological Institute (MNI) template (Fonov et al., 2011). We calculated the non-linear registration because the linear registration to the T1w volume that was already applied in preprocessing does not take into account more subtle local differences in brain anatomy that need to be taken into account in defining the trajectory of major white matter pathways. Twenty-four different white matter tracts were defined in template space based on a combination of inclusion and exclusion regions of interest (ROI). Sixteen are from the original AFQ templates (Wakana et al., 2007; Yeatman et al., 2012), and eight are callosal tracts (Dougherty et al., 2007). The tracts are enumerated in Table 1. The ROIs are primarily planar “inclusion” ROIs, where streamlines transecting the ROIs are assigned to be part of the bundle. However, some of the ROIs are “endpoint” ROIs, where streamlines must either start or end in the ROI, and some are “exclusion” ROIs, where streamlines cannot transect the ROI, to be assigned. The ROIs for each tract were transformed into the individual subject anatomical coordinates using the inverse of the transformation defined by SyN from the subject to the template space. Streamlines were selected from the whole-brain tractography based on whether they passed through inclusion ROIs and did not pass through exclusion ROIs for each tract. After initial selection was conducted, individual streamlines may additionally have been excluded based on whether they were extreme outliers. Streamlines were considered outliers if their Mahalanobis distance to other streamlines is greater than three standard deviations or if their length was more than five standard deviations from the mean length. This outlier exclusion was conducted over five rounds, similar to the original AFQ procedure (Yeatman et al., 2012). The diffusion kurtosis imaging (DKI) model was fit using the DIPY implementation to create the following maps of microstructural tissue properties: fractional anisotropy (FA), mean diffusivity (MD), and mean kurtosis (MK) (Henriques et al., 2021), as well as axonal water fraction (AWF) from the White Matter Tract Integrity (WMTI) model (Fieremans et al., 2011). In each tract, every streamline was resampled to 100 nodes, and tract profiles were generated by sampling the FA, MD, MK, AWF maps using these positions. The contributions of each streamline to the tract profile at each position was inversely weighted by the distance of that node from the median of the streamline positions for that node (Yeatman et al., 2012).

### 2.3 Evaluating heritability of tract profiles

The collection of data from both monozygotic (MZ) and dizygotic (DZ) twins in the HCP dataset enables an assessment of the genetic linkage, or heritability, of traits measured in the data with Haseman-Elston regression (Haseman and Elston, 1972). In this method, identity by descent in each twin pair is regressed against the square of the difference between twins in the tissue property tract profiles at every position along each tract (Equation 1):

$$(Y_{ijk1} - Y_{ijk2})^2 = \alpha + \beta\pi_i, \quad (1)$$

where  $i$  is an index of the twin pair,  $Y_{ij1} - Y_{ij2}$  is the difference between the two members of this twin pair in the tissue property value at position  $j$  (1-100) along tract  $k$  (1-24; Table 1). The genetic linkage  $\pi_i$  is assessed through the degree of identity by descent (i.e.,  $\pi_i = 1.0$  for MZ and  $\pi_i = 0.5$  for DZ twins). Heritability of the tissue properties for position/tract  $jk$ ,  $h_{jk}^2$  is then estimated as (Equation 2):

$$h_{jk}^2 = -\beta/(2\sigma_{jk}^2), \quad (2)$$

where  $\sigma_{jk}^2$  is the variance of the squared difference  $(Y_{ijk1} - Y_{ijk2})^2$  across  $i$ .

### 2.4 Evaluating brain-behavior correlations in tractometry data

We used tractometry-generated tract profiles for every tract as input features to a regularized predictive model to investigate the brain-behavior correlations of tractometry and a variety of cognitive and non-cognitive phenotypes. Each phenotype was predicted individually using a LASSO regularized linear model where the input features were the 100 node-level FA, MD, MK and AWF measurements from each of 24 tracts. LASSO regularized linear models remove unimportant features by shrinking the model weights of coefficients to zero (Tibshirani, 1996). In addition to the LASSO regularized models, the inherent grouping of tract profiles into tracts and tissue properties provides an opportunity to use models that exploit such groupings, such as Sparse Group LASSO (SGL) (Simon et al., 2013; Richie-Halford et al., 2021). In addition to the shrinking of individual features, SGL shrinks entire groups toward zero, eliminating both uninformative features and groups. As a comparison, we also created LASSO models using a different tissue property description, the local connectome (Yeh et al., 2016). This approach calculates a q-space normalized map of the density of spins between neighboring locations along tracts, producing a much larger number of features (128,894 features for each subject in LC, compared to 9,600 tract profile features). These features were also used as input features to a LASSO regularized model. A nested 5-fold cross-validation procedure was used to determine the level of regularization that was used, for fitting and for evaluation. To evaluate the reliability of our models, each model was ran 100 times, using different splits for cross validation (CV). Because the dataset contains familial relationships, cross-validation was done with respect to family, such that individuals within the same family were always assigned to the same fold. Models were evaluated on their predictive ability using the out-of-sample coefficient of determination  $R^2$  and on reliability using 95% confidence intervals of their model weights across the different CV splits.

### 2.5 TRX and TRK comparison

By default, pyAFQ generates outputs using the popular TrackVis file format (TRK) (Wang et al., 2007). However, this format does have limitations for our application. First of all, the format can not represent multiple tracts in a single file, requiring

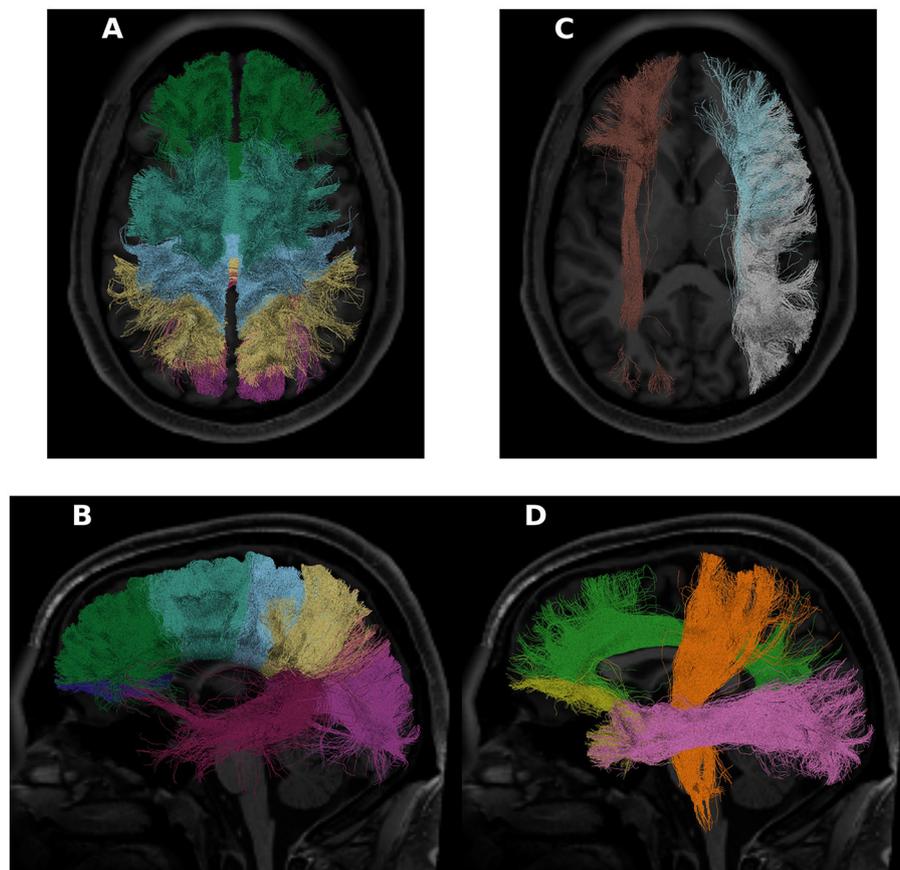


FIGURE 1

Some of the tracts recognized in a randomly chosen HCP subject (subject ID: 550439). On the left, in (A, B), we see the 8 callosal tracts visualized. In (C), we see the left inferior frontal occipital fasciculus in brown, and the right arcuate and superior longitudinal fasciculus in blue and white, respectively. In (D), the corticospinal tract is shown in orange, the cingulum is shown in green, the uncinate is shown in yellow, and the inferior longitudinal fasciculus is shown in pink. For this panel, all shown tracts are from the left hemisphere. In all panels, the subject T1 is used as the background.

many files to represent all tracts. Second, TRK files are large and slow to read, both of which impact online data visualization and analyses. Therefore, to test the new TRX format and compare it to TRK performance, the full and segmented tractograms generated during processing by pyAFQ were converted from TRK format to TRX format (Rheault et al., 2022). The data for both formats have been made available on the Open NeuroData AWS bucket. The TRX format allows users to set the data type of tractogram coordinates/vertices, and we chose to save the tractograms as half floats. We also used TRX's built-in zip compression option. We recalculated tract profiles from the TRK and TRX files while profiling for time and memory usage, in order to compare their performance.

## 2.6 Tractoscope

We developed a web-based application to visualize individual subject data from the HCP. The application was built using the Vue JavaScript framework and the NiiVue package (Hanayik et al., 2024). The application connects directly to the AWS bucket and uses the REST API provided by AWS buckets to query for the presence of expected files and to render the files into the browser

window. The application leverages the Pinia datastore library (<https://pinia.vuejs.org/>) to encapsulate and manage the large amounts of data that the application needs to operate. The source code is managed on an open-source GitHub repository (<https://github.com/nrdg/tractoscope>) and the application is deployed using npm running on the netlify continuous delivery platform to the GitHub Pages web service.

## 3 Results

### 3.1 Openly available pyAFQ HCP derivatives

All of the derivatives generated by pyAFQ to perform each of the steps in processing have been made available through the AWS Open Data programs' Open Neurodata bucket (Vogelstein et al., 2018). The results of tract recognition on a single randomly selected subject (subject ID: 550439) is shown in Figure 1. The average tract profiles from all subjects for all tracts and tissue properties are shown in Figures 2, 3.

The results can be accessed using the Amazon Web Services Command Line Interface (AWS CLI;

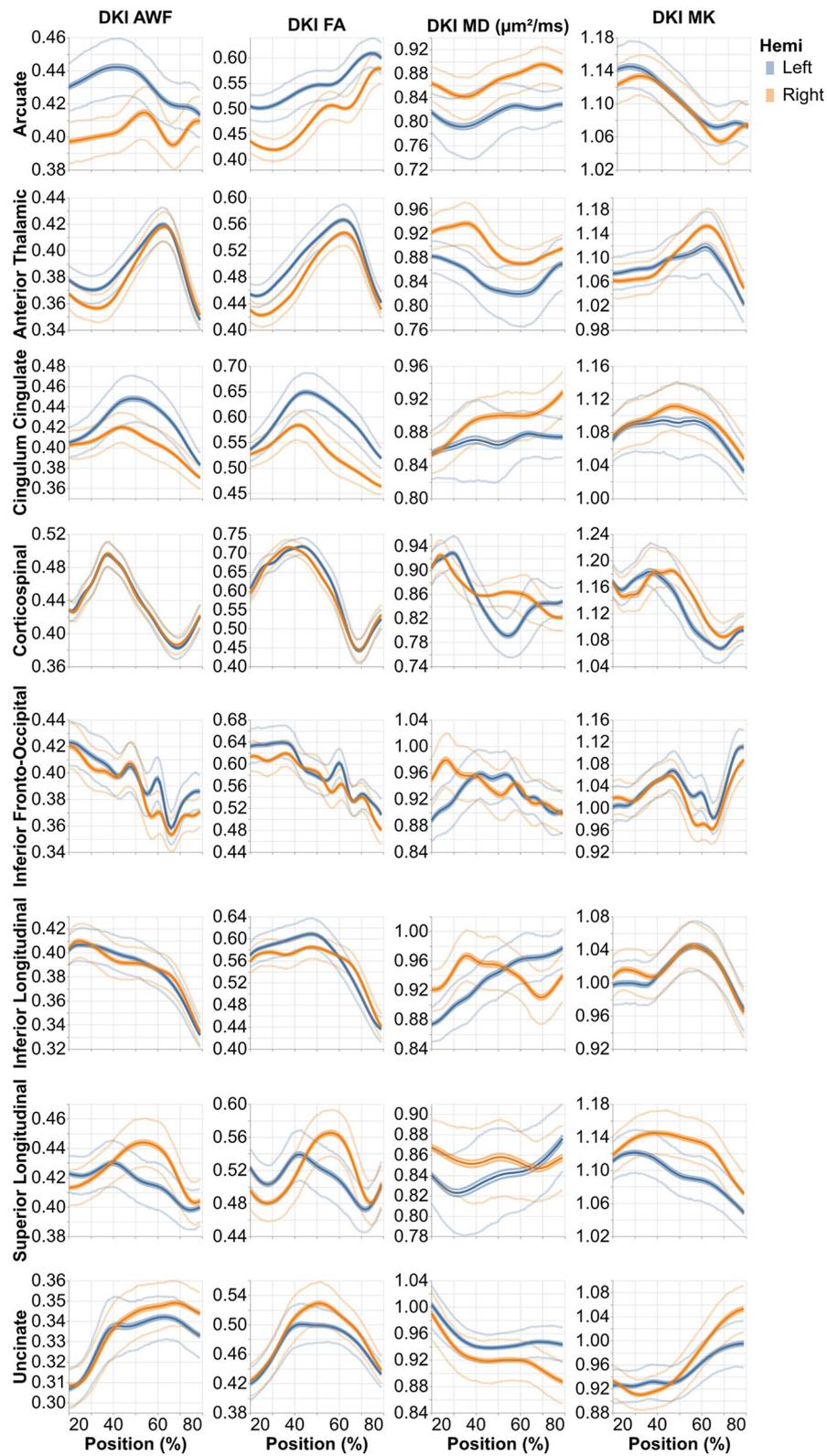
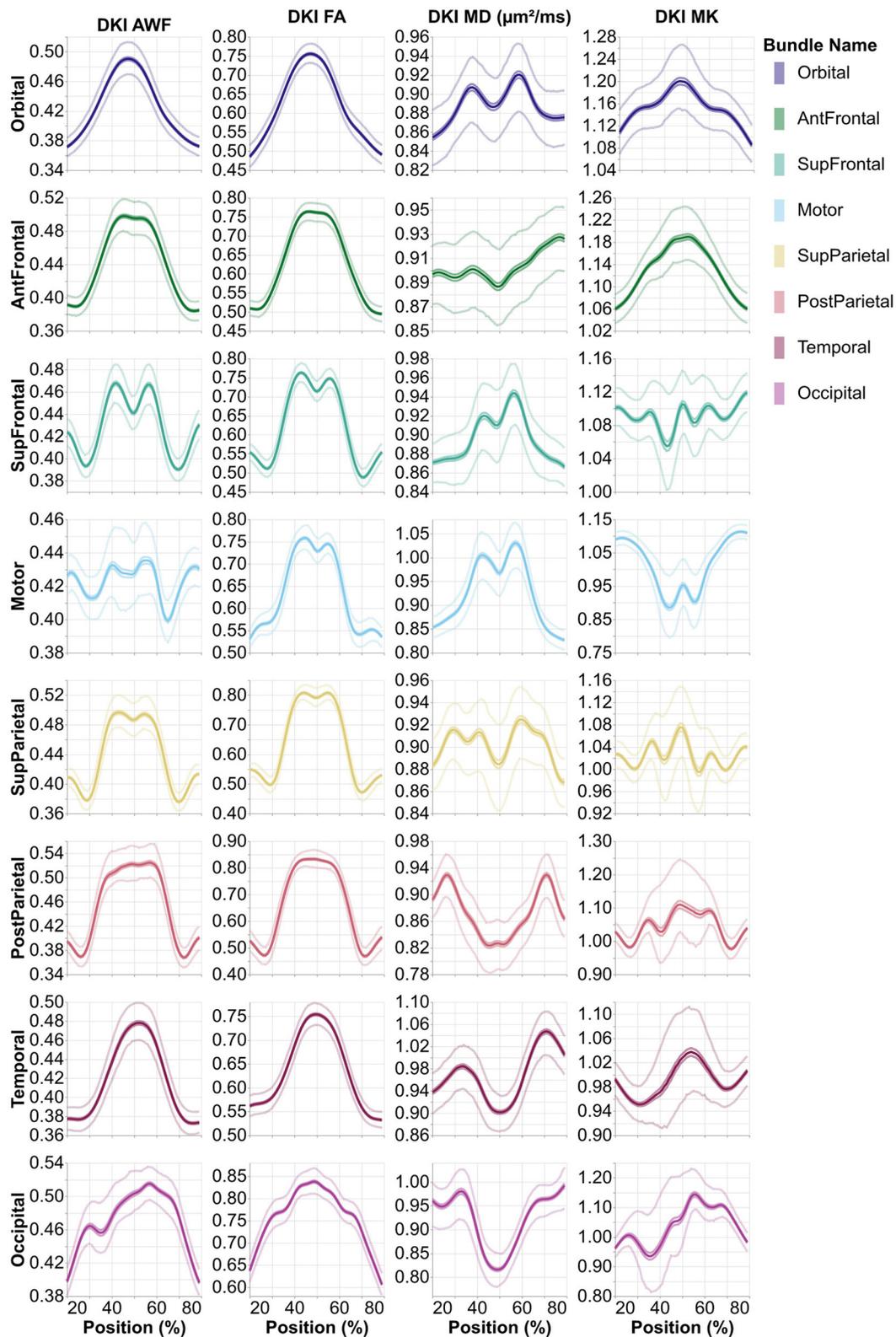


FIGURE 2

Average profiles from the 16 standard non-callosal pyAFQ tracts for all HCP subjects. The x-axis encodes position along the given bundle, discretized into 100 positions per bundle. The thin lines that tightly hug the average profile indicate the 95% confidence interval, and they are often hard to see as they closely follow the mean, due to the large sample size. The thinner lines indicate the interquartile range. Different rows correspond to different tracts, with color showing the hemisphere. The different columns show different tissue properties, from left to right: axonal water fraction, fractional anisotropy, mean diffusivity, and mean kurtosis.



**FIGURE 3**  
 Average tract profiles from the eight absence of any commercial or financial relationships callosal pyAFQ tracts for all HCP subjects. The x-axis encodes position along the given bundle, discretized into 100 positions per bundle. The thin lines that tightly hug the average profile indicate the 95% confidence interval, and they are often hard to see as they closely follow the mean, due to the large sample size. The thinner lines indicate the interquartile range. Different rows and colors correspond to different subdivisions of the callosal tracts. The different columns show different tissue properties, from left to right: axonal water fraction, fractional anisotropy, mean diffusivity, and mean kurtosis.

<https://aws.amazon.com/cli/>) at the following S3 address: `s3://open-neurodata/rokem/hcp1200/`. The dataset is organized using principles adapted from the Brain Imaging Data Structure (BIDS), a standard for organizing and describing neuroimaging data (Gorgolewski et al., 2016), to facilitate easy access and exploration of the data, and interoperability with other datasets. Detailed examples of data access using the AWS CLI and the `boto3` Python library are provided in the [Supplementary material](#).

### 3.2 Heritability of tract profiles of tissue properties

The heritability of tract profiles varies between tissue properties, tracts, and within each tract (Figures 4, 5). Averaging across all tracts and positions along the tracts, the heritability of the different tissue properties is: FA:  $h^2 = 0.33 \pm 0.17$ , MD:  $h^2 = 0.29 \pm 0.15$ , MK:  $h^2 = 0.42 \pm 0.25$ , AWF:  $h^2 = 0.47 \pm 0.2$  (standard deviations across tracts and positions are reported). In most cases, we observe some symmetry across the midline, mirroring the laterality of tissue properties observed in Figures 2, 3, although this symmetry is less clear than with the tissue properties themselves. A notable exception to this symmetry is in the heritability of MK in the arcuate fasciculus, which is substantially lower in the left hemisphere than in the right hemisphere.

### 3.3 Accuracy and reliability of brain-phenotype models based on tract profile features

Regularized regression models were used to assess brain-phenotype correlations (Figure 6). Variance explained ( $R^2$ ) was assessed as a measure of the accuracy of the correlations, using cross-validation to mitigate the potential for overfitting within the data used for fitting. Variability of this estimate was assessed using bootstrapping. For both tractometry and LC features, accuracy across a range of phenotypes varies between almost no predictive accuracy for all models (e.g., Attention - LC:  $R^2 = 0.0064$  95% CI [0.00010, 0.030], SGL: 0.0044 [0.00011, 0.013], LASSO:  $R^2 = 0.0033$  [0.00012, 0.010]) and moderate predictive accuracy (e.g., Age - LC:  $R^2 = 0.18$  [0.10, 0.26], SGL:  $R^2 = 0.31$  [0.21, 0.42], LASSO:  $R^2 = 0.30$  [0.19, 0.39]). Though there are nominal differences between LC and tract profile predictions in some phenotypes (e.g., Age and Reading Ability), we found no significant differences in accuracy or reliability of models that used the two methods to derive features for predictive modeling.

While model accuracy did not vary significantly by model choice (Table 2, the reliability of the model weights for LASSO and SGL models did) (Table 3). Across phenotypes, LASSO tended to assign high model weights to individual nodes, with large variances across bootstraps. In contrast, SGL assigned smaller model weights to adjacent nodes within tracts, with much smaller variances in model weights across bootstraps (Figure 7). This pattern occurs across all phenotypes (Supplementary Figures S1–S4).

### 3.4 TRX provides a storage-efficient file format for tractometry data

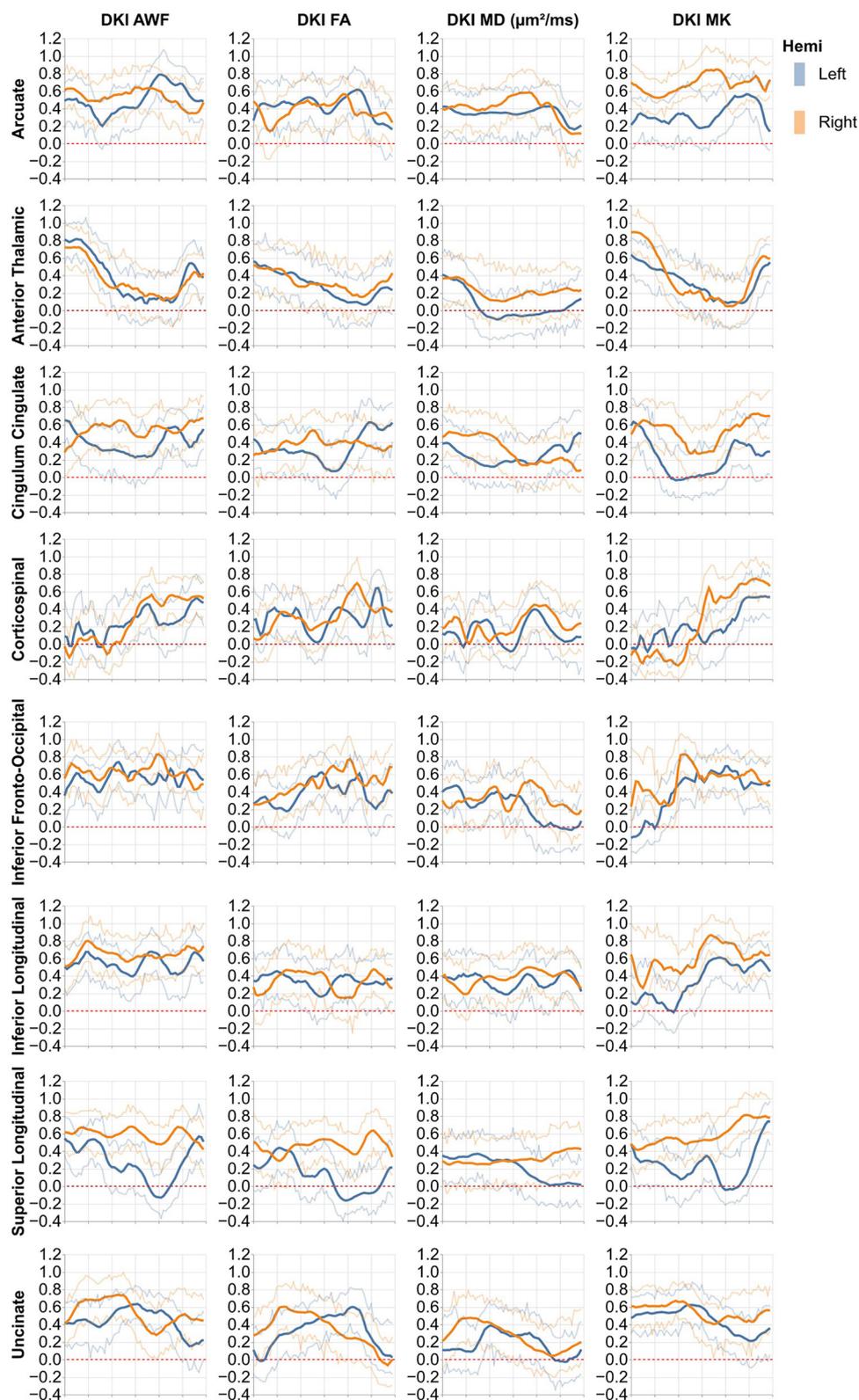
To assess the performance of the TRX file format, we calculated tract profiles from each of the tracts using the data that was stored in the TRX file format, and calculated the ratio of the elapsed time for TRX/TRK. Performance did not substantially differ between the file formats (Figure 8A), except in some cases where calculation of profiles from TRX was substantially faster than with TRK. Similarly, memory usage of TRK and TRX are very similar (Figure 8B). A similar ratio was computed for the FA along the length of the tracts (Figure 8C). Despite the decreased numerical precision, and the large substantial decrease in the file sizes on disk, which often exceed a factor of 0.5X (Figure 8D), all differences in the tract profiles were smaller than 0.1%.

### 3.5 A browser-based application for exploring the HCP tractometry results

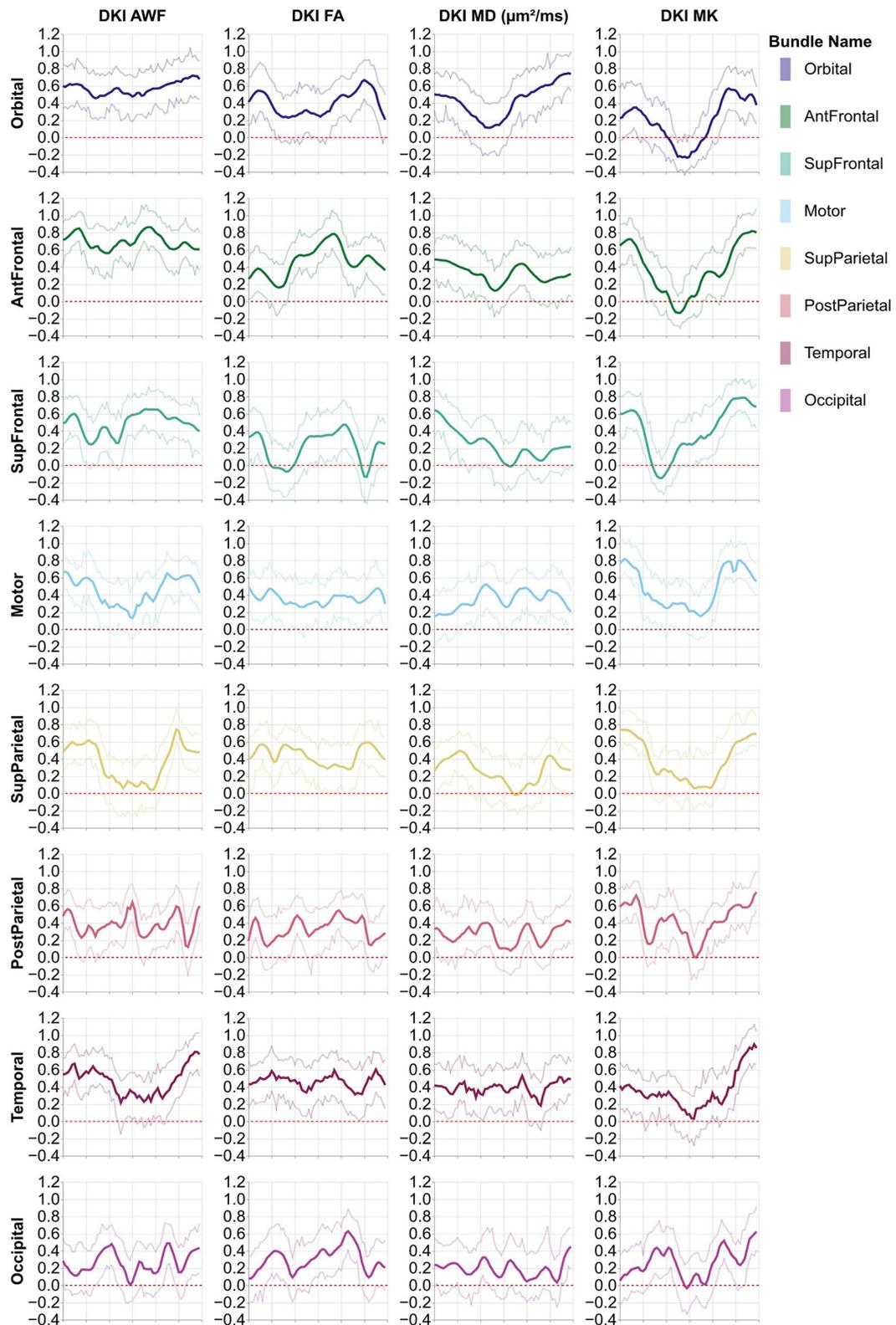
Evaluating tractometry results and viewing them without downloading any data is possible using the Tractoscope web app. Tractoscope was implemented to work with both TRK and TRX file formats, allowing users to easily explore and visualize tractography files in the HCP dataset, as well as other datasets that comply with a similar BIDS-inspired data layout. The tool is available publicly (<https://github.com/nrdg/tractoscope>). Any pyAFQ-compliant dataset hosted on AWS S3 buckets can be connected to the existing application with minimal configuration changes, by adding an entry to a `datasets.json` file. Once the AWS S3 bucket is configured to be publicly available and has HTTPS enabled, Tractoscope will be able to connect to it and visualize the dataset. The application currently enables visualizations of both the HCP dataset described here (Figure 9), as well as another dataset: the HBN-POD2 dataset, previously described in [Richie-Halford et al. \(2022\)](#).

## 4 Discussion

The open availability of datasets like HCP promotes collaborative studies and enhances methodological approaches. This tractometry analysis of HCP diffusion MRI data using pyAFQ and its visualization through Tractoscope exemplifies the practical benefits of accessible data. This approach facilitates a broad range of research possibilities, where different groups can use the tissue properties we share to get a more detailed understanding of white matter pathways, which are crucial for studies on neurological disorders, brain development, and cognitive functions. Some of the potential uses of the resources that we have created include: (i) as a normative sample, to be compared to various patient populations, (ii) integration with the other data that was collected by HCP in the same subjects (e.g., functional MRI measurements), (iii) further exploration of the relationships between white matter tissue properties and other phenotypic measurements, and (iv) as an educational resource for learning about the structure of human brain white matter.



**FIGURE 4**  
 Heritability profiles from the 16 standard non-callosal pyAFQ tracts for all HCP subjects. The x-axis encodes position along the bundle. Thin lines indicate 95% confidence interval. Different rows correspond to different tracts, with color showing the hemisphere. The different columns show different tissue properties, from left to right: axonal water fraction, fractional anisotropy, mean diffusivity, mean kurtosis.



**FIGURE 5**  
Heritability profiles from the eight callosal pyAFQ tracts for all HCP subjects. The x-axis encodes position along the bundle. Thin lines show the 95% confidence interval. Different rows and colors correspond to different subdivisions of the callosal tracts. The different columns show different tissue properties, from left to right: axonal water fraction, fractional anisotropy, mean diffusivity, mean kurtosis.

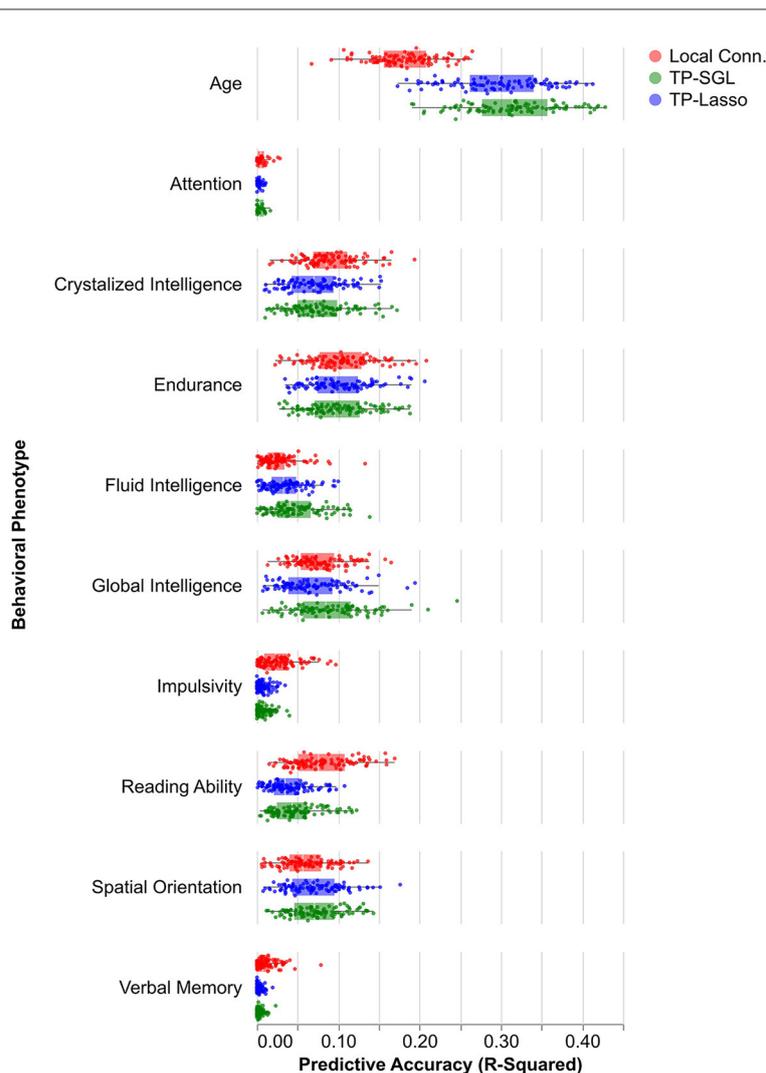


FIGURE 6

Predictive model performance by phenotype. Box and whisker plots show the distribution model accuracies by model type and input feature. Boxes show the middle 50% of accuracy values (quantified by using out of sample  $R^2$ ), and each point is one model run.

The granular approach of tractometry potentially enables a more nuanced understanding of white matter variation. Additionally, by focusing on known tracts, the results of tractometry have been shown to be reliable across scans and robust to choice of model (Kruper et al., 2021). To improve interoperability between this dataset and others, we used the BIDS standard as inspiration for organizing and describing the data (Gorgolewski et al., 2016). BIDS is structured to improve the accessibility, organization, and ease of sharing complex brain imaging datasets. It employs a consistent naming scheme and directory structure, making it easier for researchers to store, analyze, and share their data with others in the field.

Analysis methods focus on various aspects of dMRI data. For example, many analysis approaches focus on generating connectivity matrices, or graphs. Connectivity results from the HCP dataset have already been published (Kiar et al., 2018). We provide a complement here, using tractometry, which allows for the evaluation of diffusion characteristics along the lengths

of known tracts. Similar, tractometry-based analysis results for a subset of HCP subjects have been published as a part of larger data releases containing subjects from multiple datasets (Avesani et al., 2019; Lerma-Usabiaga et al., 2020; Hayashi et al., 2023). Here, we provide tractometry results for all subjects in HCP that have a complete dMRI acquisition. We also provide an initial characterization of population-level tract profiles in Figure 2. This characterization replicates previously known properties of human brain tract profiles. For example, there is a substantial lateralization of tissue properties in the arcuate fasciculus compared to other tracts, which is known feature of this tract (Bain et al., 2019).

#### 4.1 The heritability of tract profiles

Brain structure and function has a substantial genetic component. Heritability assesses the amount of variance within

TABLE 2 Variances of SGL and LASSO model weights for each phenotype across tracts.

	SGL	LASSO
Age	5.7e-03	1.5e-02
Crystallized intelligence	8.5e-03	3.0e-02
Fluid intelligence	7.4e-03	2.4e-02
Global intelligence	1.2e-02	3.6e-02
Impulsivity	5.4e-05	1.8e-04
Endurance	7.4e-03	2.8e-02
Verbal memory	8.1e-04	3.4e-03
Reading ability	9.1e-03	2.6e-02
Attention	3.5e-06	2.6e-05
Spatial orientation	3.4e-03	1.1e-02

a studied trait that can be explained by genetic differences. Because of their known shared genetic background, twin pairs are often studied to assess heritability. The HCP was designed with this in mind, recruiting 149 MZ and 94 DZ twin pairs (138 MZ pairs and 75 DZ pairs were included in our heritability analysis, because of missingness of DWI data in some participants). Previous research has already demonstrated that DTI-derived tissue properties are heritable at the level of tract averages both in the HCP (Kochunov et al., 2015; Gao et al., 2021), as well as in other datasets (Gustavson et al., 2019). In a few cases, heritability of DTI metrics was also assessed along the length of tracts (Lee et al., 2015). In line with these previous findings, we also found that DKI metrics can have substantial heritability up to approximately  $h^2 = 0.9$  for the DKI-specific metrics (MK and AWF) and slightly lower for metrics that are estimated in both DTI and DKI (FA and MD, which both do not exceed  $h^2 = 0.8$ ). Higher heritability seems to correspond to smaller error bars in the tract profiles, suggesting that heritability of a white matter tissue property is easier to discern when the signal is more reliably measured. The spatial variability of heritability across the length of the tracts is notable and mirrors to some extent the spatial variability of tract profiles of tissue properties. Variability in the heritability of tissue properties themselves may reflect interactions with other parts of the tissue, or different sensitivity of portions of the tracts to environmental or genetic factors.

## 4.2 Comparing tract profiles and local connectome

One of the promises of the large-scale data collection of the HCP was that the data would illuminate individual variability in a variety of behavioral measures and differences in cognitive abilities. There are a variety of different ways to assess brain-behavior correlations that are at the foundation of establishing the brain basis of individual differences. Here, we

TABLE 3 Average accuracy for each phenotype and model.

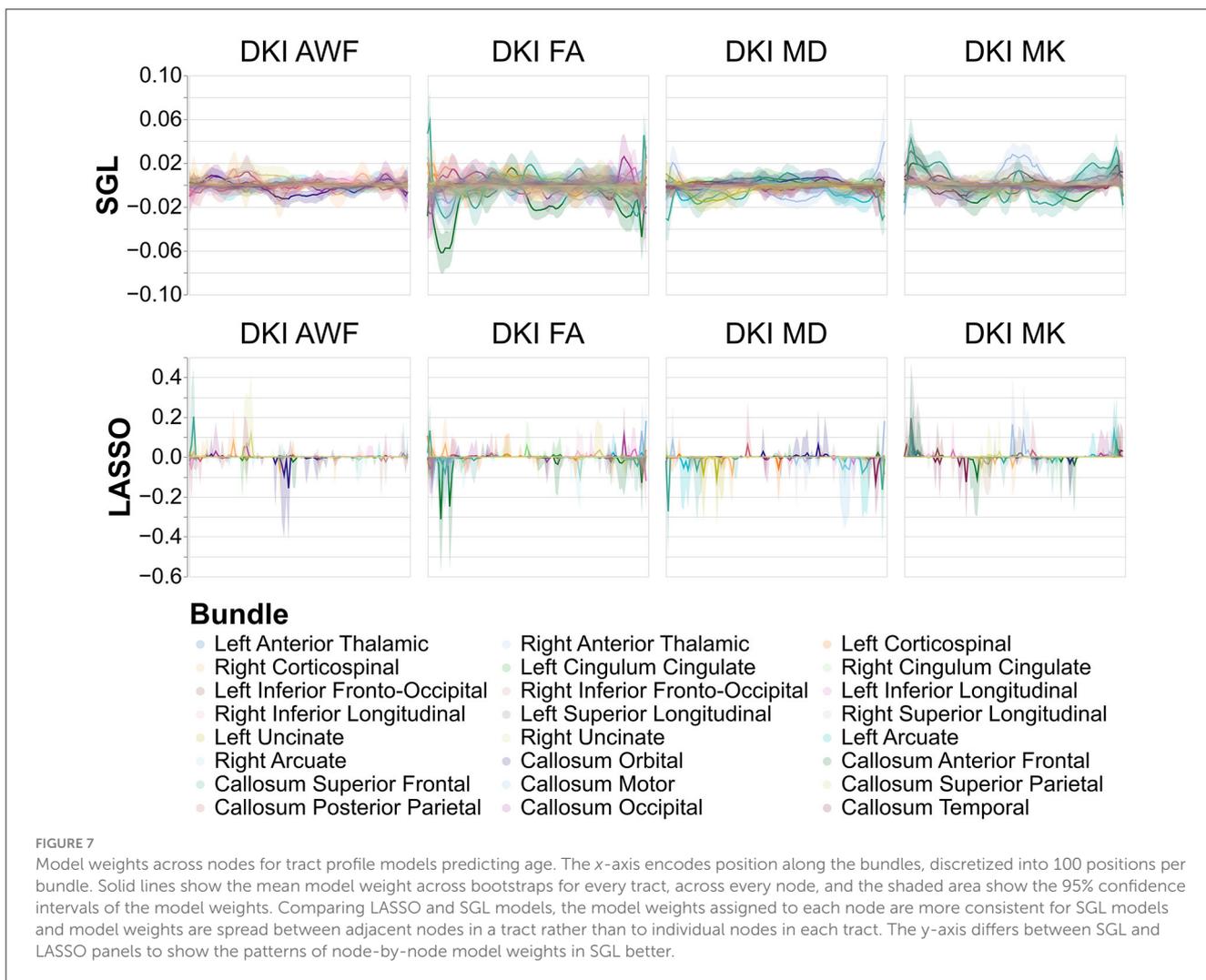
	LASSO	LC	SGL
Age	3e-01	1.8e-01	3.1e-01
Crystallized intelligence	7.2e-02	9e-02	7.6e-02
Fluid intelligence	3.5e-02	2.7e-02	4.8e-02
Global intelligence	6.8e-02	7.7e-02	8.9e-02
Impulsivity	7.1e-03	2.7e-02	8.2e-03
Endurance	1e-01	1e-01	1e-01
Verbal memory	3.3e-03	1.2e-02	3.7e-03
Reading ability	3.9e-02	8.1e-02	4.8e-02
Attention	3.3e-03	6.4e-03	4.4e-03
Spatial orientation	7e-02	6.1e-02	7.2e-02

assessed the information that is available in white matter tract profiles using regularized regression approaches. As a baseline for comparison, we used features of the white matter extracted using the local connectome (LC) approach (Yeh et al., 2016). We found that both tract profiles and local connectome had small predictive skill for most phenotypes, with nominal but insignificant differences in predictive accuracy of models using tract profiles or LC as their input features (Figure 6). In line with previous literature, we found that phenotypes varied by their ability to be predicted regardless of input features, with some phenotypes like attention, verbal memory, and impulsivity having predictive accuracies near zero (Rasero et al., 2021; Roy et al., 2024). Other phenotypes, like age, had average  $R^2$  values around 0.30 for all models. Though SGL and LASSO did not differ in terms of their average accuracy, they differ substantially in terms of the variability in their feature selection properties. SGL provides much smoother and less variable selection of features.

Taken together this set of results suggests that tractometry of the human white matter extracts much of the useful information about individual differences that is present in the LC method, but the number of features is smaller by approximately an order of magnitude. This indicates that tractometry dramatically reduces the dimensionality of dMRI data, while preserving many of the features that are relevant to individual differences, to the extent that those are reflected in brain white matter tissue properties.

## 4.3 Comparing TRK and TRX

The availability of comprehensive and accessible data resources is instrumental in driving forward research in understanding brain function in health and disease. File formats and standards for storing scientific data are an important key component of the cyberinfrastructure used to disseminate and reuse scientific results, as intended here. The TRX format is a



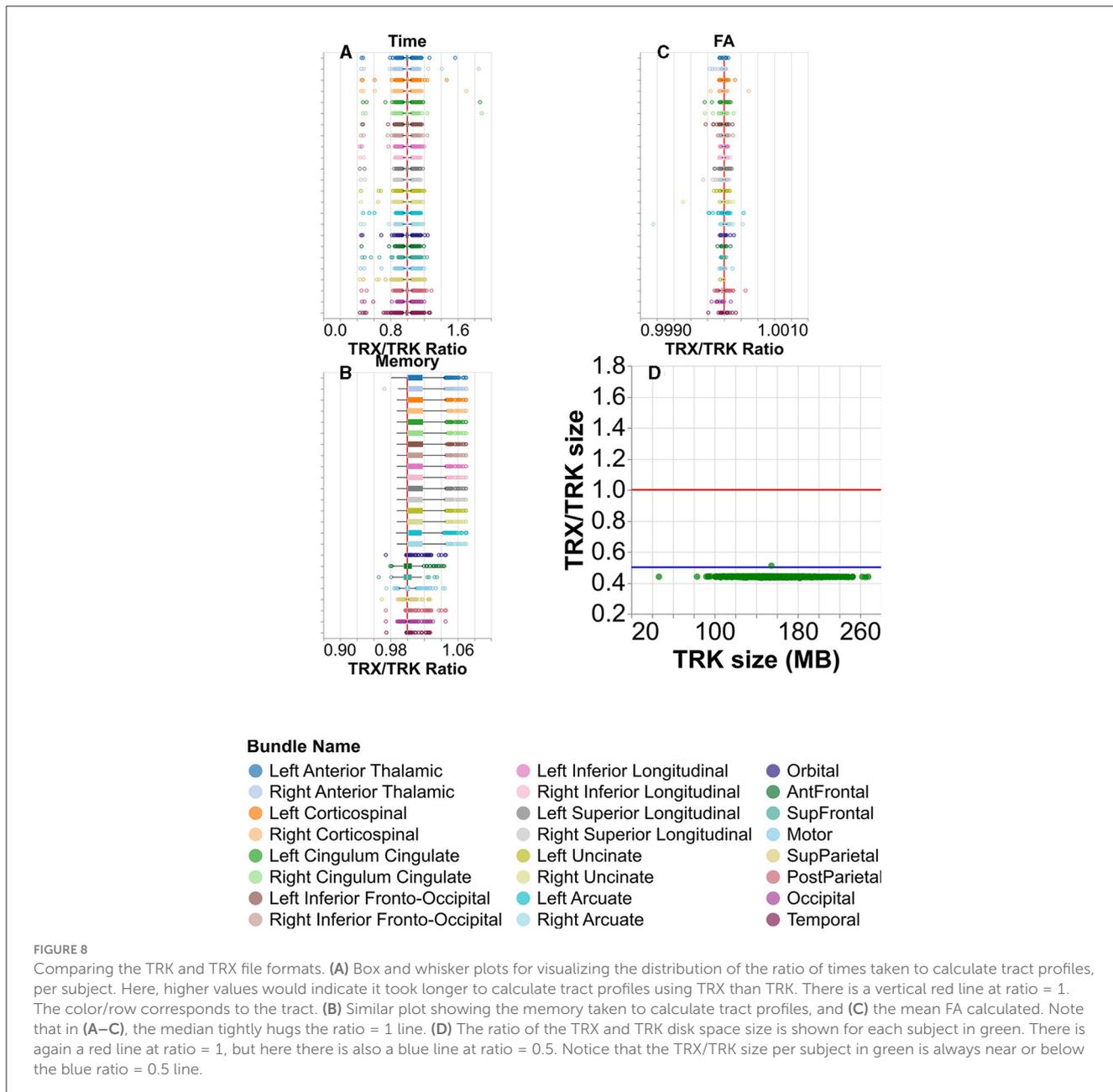
recent proposal to improve storage and access to datasets of computational tractography results (Rheault et al., 2022). The use of the TRX file format should help address the challenges of efficiently managing large neuroimaging datasets that contain such results.

Our study includes a performance comparison between TRK and TRX formats in profiling the tracts that we delineated in HCP. From Figures 8A, B, we see that the means are centered on the vertical red line, indicating that the time and memory required for calculation of tract profiles using TRX are comparable to those using TRK. From Figure 8C, we see that the differences in the resulting profiles are typically much smaller than 0.01%, with one outlier having a difference of approximately 0.01%. Additionally, TRX's integrated zip functionality and flexible data saving options enable more efficient use of disk space for storing tractograms, providing a potential for more than 2X improvement in storage, with almost no loss in information. Furthermore, the use of TRX's built-in grouping feature for

segmented tractograms offers a more convenient approach compared to TRK to manage results of tractometry analysis. In TRK, segmented tracts typically necessitate additional files for storing tract identification metadata, whereas TRX simplifies this process, enhancing the efficiency of data management in neuroimaging studies.

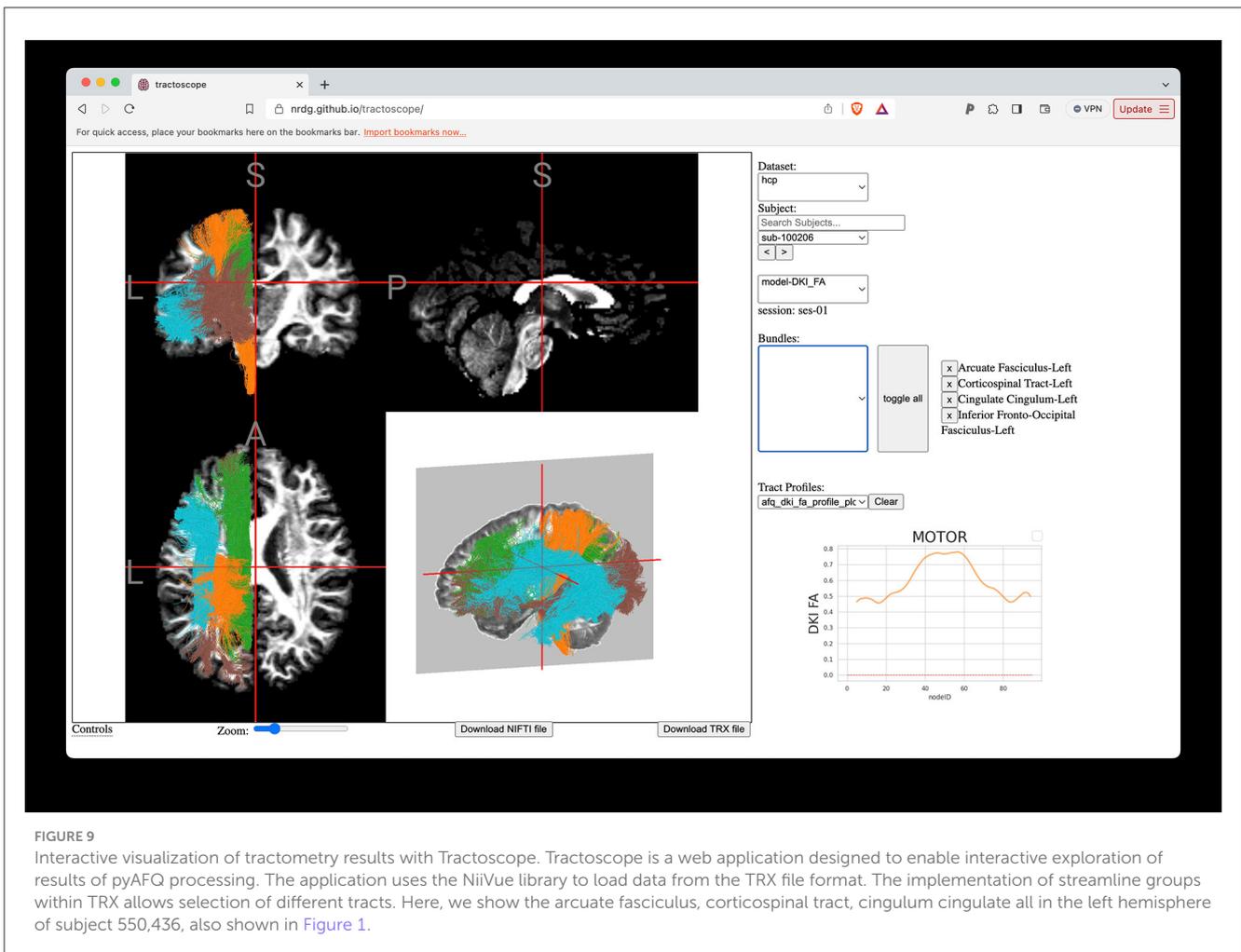
#### 4.4 Visualizing the data with Tractoscope

We developed Tractoscope, a NiiVue-based web-viewer for neuroimaging data that allows users to visualize large datasets hosted on the cloud. Tractoscope enables visualization and exploration of cloud-hosted pyAFQ-processed datasets. Tractoscope is built to work with the Amazon Web Services API, which allows it to interact dynamically with datasets that comply with the structure expected for outputs of the pyAFQ software. This significantly decreases the amount of work developers would



have to do to connect the tool to future datasets. The tool is also highly configurable, allowing developers to select which scans and tracts should be made available to the user for selection through the application graphical user interface. The tool also has the ability to display tract profiles, such as those generated by pyAFQ, so long as those are stored in the graphical output format that pyAFQ generates per default. The result is a user-friendly, configurable website that can display any and all structural and diffusion imaging for datasets in the pyAFQ output format. If available, Tractoscope uses TRX files due to their increased efficiency, but it is still compatible with datasets that use TRK files.

Tractoscope demonstrates that the development of standard ways to represent large datasets facilitates the development of a wide range of standards-compliant applications, which can be universally applied to any dataset formatted according to the standard (Pestilli et al., 2021). By doing so, we ensure compatibility and interoperability across various research tools and datasets, significantly enhancing the efficiency and scope of neuroimaging research. pyAFQ operates according to these principles, as does Tractoscope. For example, Tractoscope already also visualizes subjects from the Healthy Brain Network (Alexander et al., 2017; Richie-Halford et al., 2022), in addition to HCP tractometry.



## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://registry.opendata.aws/open-neurodata/>.

## Ethics statement

Ethical approval was not required for the current study in accordance with the local legislation and institutional requirements, as the study used publicly available deidentified human neuroimaging datasets, and written informed consent was obtained by the Human Connectome Project for data collection. Additional written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

## Author contributions

JK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing. MH: Conceptualization, Formal analysis, Funding acquisition, Investigation, Visualization, Writing—original draft, Writing—review & editing. FR: Methodology, Software, Writing—review & editing. IC: Software, Writing—original draft, Writing—review & editing. AG: Software, Writing—original draft, Writing—review & editing. MN: Formal analysis, Methodology, Software, Writing—review & editing. KM: Methodology, Writing—review & editing. EL: Methodology, Funding Acquisition, Writing—review & editing. CR: Software, Funding Acquisition, Writing—review & editing. JY: Conceptualization, Funding acquisition, Writing—review & editing. AR: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Project administration, Writing—original draft, Writing—review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Data were provided 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 MacDonnell Center for Systems Neuroscience at Washington University. This project was funded by NSF grant 1934292 (PI: Balazinska), NIH grant RF1 MH121868 (PI: Rokem), NIH grant R01EB027585 (PI: Garyfallidis), NIH grant R01HD095861 (PI: Yeatman), NIH grant RF1MH133701 (PI: Rorden), NIH grant NIH R03 EB033001 (PI: Lila), and NSF grant DMS-2210064 (PI: Lila). JK was supported through the NSF Graduate Research Fellowship DGE-2140004. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0023112 (Awardee: McKenzie Hagen). This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply

its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Cloud computing resources were provided by Azure through the University of Washington eScience Institute and by the Amazon Web Services Open Data program.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2024.1389680/full#supplementary-material>

## References

- Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., et al. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* 4:170181. doi: 10.1038/sdata.2017.181
- Andersson, J. L. R., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20, 870–888. doi: 10.1016/S1053-8119(03)00336-7
- Andersson, J. L. R., and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* 125, 1063–1078. doi: 10.1016/j.neuroimage.2015.10.019
- Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004
- Avesani, P., McPherson, B., Hayashi, S., Caiafa, C. F., Henschel, R., Garyfallidis, E., et al. (2019). The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Sci. Data* 6:69. doi: 10.1038/s41597-019-0073-y
- Bain, J. S., Yeatman, J. D., Schurr, R., Rokem, A., and Mezer, A. A. (2019). Evaluating arcuate fasciculus laterality measurements across dataset and tractography pipelines. *Hum. Brain Mapp.* 40, 3695–3711. doi: 10.1002/hbm.24626
- Bammer, R., Markl, M., Barnett, A., Acar, B., Alley, M. T., Pelc, N. J., et al. (2003). Analysis and generalized correction of the effect of spatial gradient field distortions in diffusion-weighted imaging. *Magn. Reson. Med.* 50, 560–569. doi: 10.1002/mrm.10545
- Basser, P. J., Mattiello, J., and Le Bihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267. doi: 10.1016/S0006-3495(94)80775-1
- Bassett, D. S., and Bullmore, E. T. (2009). Human brain networks in health and disease. *Curr. Opin. Neurol.* 22, 340–347. doi: 10.1097/WCO.0b013e32832d293dd
- Bassett, D. S., and Sporns, O. (2017). Network neuroscience. *Nat. Neurosci.* 20, 353–364. doi: 10.1038/nn.4502
- Bookheimer, S. Y., Salat, D. H., Terpstra, M., Ances, B. M., Barch, D. M., Buckner, R. L., et al. (2019). The lifespan human connectome project in aging: an overview. *Neuroimage* 185, 335–348. doi: 10.1016/j.neuroimage.2018.10.009
- Caruyer, E., Lenglet, C., Sapiro, G., and Deriche, R. (2013). Design of multishell sampling schemes with uniform coverage in diffusion MRI. *Magn. Reson. Med.* 69, 1534–1540. doi: 10.1002/mrm.24736
- Colby, J. B., Soderberg, L., Lebel, C., Dinov, I. D., Thompson, P. M., and Sowell, E. R. (2012). Along-tract statistics allow for enhanced tractography analysis. *Neuroimage* 59, 3227–3242. doi: 10.1016/j.neuroimage.2011.11.004
- Dayan, M., Monohan, E., Pandya, S., Kuceyeski, A., Nguyen, T. D., Raj, A., et al. (2016). Profilmetry: a new statistical framework for the characterization of white matter pathways, with application to multiple sclerosis. *Hum. Brain Mapp.* 37, 989–1004. doi: 10.1002/hbm.23082
- Demro, C., Mueller, B. A., Kent, J. S., Burton, P. C., Olman, C. A., Schallmo, M.-P., et al. (2021). The psychosis human connectome project: an overview. *Neuroimage* 241:118439. doi: 10.1016/j.neuroimage.2021.118439
- Dougherty, R. F., Ben-Shachar, M., Deutsch, G. K., Hernandez, A., Fox, G. R., and Wandell, B. A. (2007). Temporal-callosal pathway diffusivity predicts phonological skills in children. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8556–8561. doi: 10.1073/pnas.0608961104
- Elam, J. S., Glasser, M. F., Harms, M. P., Sotiropoulos, S. N., Andersson, J. L. R., Burgess, G. C., et al. (2021). The human connectome project: a retrospective. *Neuroimage* 244:118543. doi: 10.1016/j.neuroimage.2021.118543
- Fieremans, E., Jensen, J. H., and Helpen, J. A. (2011). White matter characterization with diffusional kurtosis imaging. *Neuroimage* 58, 177–188. doi: 10.1016/j.neuroimage.2011.06.006

- Fonov, V., Evans, A. C., Botteron, K., Almlí, C. R., McKinstry, R. C., Collins, D. L., et al. (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54, 313–327. doi: 10.1016/j.neuroimage.2010.07.033
- Gao, S., Donohue, B., Hatch, K. S., Chen, S., Ma, T., Ma, Y., et al. (2021). Comparing empirical kinship derived heritability for imaging genetics traits in the UK biobank and human connectome project. *Neuroimage* 245:118700. doi: 10.1016/j.neuroimage.2021.118700
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., et al. (2014). Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* 8:8. doi: 10.3389/fninf.2014.00008
- Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., et al. (2016). The human connectome project's neuroimaging approach. *Nat. Neurosci.* 19, 1175–1187. doi: 10.1038/nn.4361
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44
- Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. doi: 10.1016/j.neuroimage.2009.06.060
- Gustavson, D. E., Hatton, S. N., Elman, J. A., Panizzon, M. S., Franz, C. E., Hagler, D. J. Jr., et al. (2019). Predominantly global genetic influences on individual white matter tract microstructure. *Neuroimage* 184, 871–880. doi: 10.1016/j.neuroimage.2018.10.016
- Hanayik, T., Drake, C., Rorden, C., Hardcastle, N., and Androulakis, A. (2024). *niivue/niivue: 0.43.3* (0.43.3). Zenodo. Available online at: <https://zenodo.org/records/11204589>
- Haseman, J. K., and Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* 2, 3–19. doi: 10.1007/BF01066731
- Hayashi, S., Caron, B. A., Heinsfeld, A. S., Vinci-Booher, S., McPherson, B., Bullock, D. N., et al. (2023). brainlife.io: a decentralized and open source cloud platform to support neuroscience research. *arXiv*. doi: 10.1038/s41592-024-02296-5
- Henriques, R. N., Correia, M. M., Marrale, M., Huber, E., Kruper, J., Koudoro, S., et al. (2021). Diffusional kurtosis imaging in the diffusion imaging in python project. *Front. Hum. Neurosci.* 15:675433. doi: 10.3389/fnhum.2021.675433
- Howell, B. R., Styner, M. A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., et al. (2019). The UNC/UMN baby connectome project (BCP): an overview of the study design and protocol development. *Neuroimage* 185, 891–905. doi: 10.1016/j.neuroimage.2018.03.049
- Jensen, J. H., Helpert, J. A., Ramani, A., Lu, H., and Kaczynski, K. (2005). Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn. Reson. Med.* 53, 1432–1440. doi: 10.1002/mrm.20508
- Jones, D. K., Travis, A. R., Eden, G., Pierpaoli, C., and Basser, P. J. (2005). PASTA: pointwise assessment of streamline tractography attributes. *Magn. Reson. Med.* 53, 1462–1467. doi: 10.1002/mrm.20484
- Keshavan, A., and Poline, J.-B. (2019). From the wet lab to the web lab: a paradigm shift in brain imaging research. *Front. Neuroinform.* 13:3. doi: 10.3389/fninf.2019.00003
- Kiar, G., Bridgeford, E. W., Roncal, W. R. G., Chandrasekhar, V., Mhembe, D., Ryman, S., et al. (2018). A high-throughput pipeline identifies robust connectomes but troublesome variability. *bioRxiv*. doi: 10.1101/188706
- Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T. E., et al. (2015). Heritability of fractional anisotropy in human white matter: a comparison of human connectome project and ENIGMA-DTI data. *Neuroimage* 111, 300–311. doi: 10.1016/j.neuroimage.2015.02.050
- Kruper, J., Yeatman, J. D., Richie-Halford, A., Bloom, D., Grotheer, M., Caffarra, S., et al. (2021). Evaluating the reliability of human brain white matter tractometry. *Apert Neuro* 1, 1–25. doi: 10.52294/e6198273-b8e3-4b63-babb-6e6b0da10669
- Lee, S. J., Steiner, R. J., Luo, S., Neale, M. C., Styner, M., Zhu, H., et al. (2015). Quantitative tract-based white matter heritability in twin neonates. *Neuroimage* 111, 123–135. doi: 10.1016/j.neuroimage.2015.02.021
- Jerma-Usabiaga, G., Mukherjee, P., Perry, M. L., and Wandell, B. A. (2020). Data-science ready, multisite, human diffusion MRI white-matter-tract statistics. *Sci. Data* 7:422. doi: 10.1038/s41597-020-00760-3
- McNab, J. A., Edlow, B. L., Witzel, T., Huang, S. Y., Bhat, H., Heberlein, K., et al. (2013). The human connectome project and beyond: initial applications of 300 mt/m gradients. *Neuroimage* 80, 234–245. doi: 10.1016/j.neuroimage.2013.05.074
- Pestilli, F., Poldrack, R., Rokem, A., Satterthwaite, T., Feingold, F., Duff, E., et al. (2021). A community-driven development of the brain imaging data standard (BIDS) to describe macroscopic brain connections. *OSF Preprint*. doi: 10.17605/OSF.IO/U4G5P
- Rasero, J., Sentis, A. I., Yeh, F.-C., and Verstynen, T. (2021). Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLoS Comput. Biol.* 17:e1008347. doi: 10.1371/journal.pcbi.1008347
- Rheault, F., Hayot-Sasson, V., Smith, R., Rorden, C., Tournier, J.-D., Garyfallidis, E., et al. (2022). *TRX: A Community-Oriented Tractography File Format*. Glasgow: Organization for Human Brain Mapping.
- Richie-Halford, A., Cieslak, M., Ai, L., Caffarra, S., Covitz, S., Franco, A. R., et al. (2022). An analysis-ready and quality controlled resource for pediatric brain white-matter research. *Sci. Data* 9, 1–27. doi: 10.1038/s41597-022-01695-7
- Richie-Halford, A., Yeatman, J. D., Simon, N., and Rokem, A. (2021). Multidimensional analysis and detection of informative features in human brain white matter. *PLoS Comput. Biol.* 17:e1009136. doi: 10.1371/journal.pcbi.1009136
- Roy, E., Richie-Halford, A., Kruper, J., Narayan, M., Bloom, D., Nedelec, P., et al. (2024). White matter and literacy: a dynamic system in flux. *Dev. Cogn. Neurosci.* 65:101341. doi: 10.1016/j.dcn.2024.101341
- Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., and Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.* 67, 1210–1224. doi: 10.1002/mrm.23097
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse-group lasso. *J. Comput. Graph. Stat.* 22, 231–245. doi: 10.1080/10618600.2012.681250
- Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., et al. (2013). Advances in diffusion MRI acquisition and processing in the human connectome project. *Neuroimage* 80, 125–143. doi: 10.1016/j.neuroimage.2013.05.057
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Tournier, J.-D., Yeh, C.-H., Calamante, F., Cho, K.-H., Connelly, A., and Lin, C.-P. (2008). Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* 42, 617–625. doi: 10.1016/j.neuroimage.2008.05.002
- Vogelstein, J. T., Perlman, E., Falk, B., Baden, A., Gray Roncal, W., Chandrasekhar, V., et al. (2018). A community-developed open-source computational ecosystem for big neuro data. *Nat. Methods* 15, 846–847. doi: 10.1038/s41592-018-0181-1
- Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., et al. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36, 630–644. doi: 10.1016/j.neuroimage.2007.02.049
- Wang, R., Benner, T., Sorensen, A. G., and Wedeen, V. J. (2007). "Diffusion toolkit: a software package for diffusion imaging data processing and tractography," in *Proc Intl Soc Mag Reson Med, Volume 15* (Berlin).
- Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., and Feldman, H. M. (2012). Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS ONE* 7:e49790. doi: 10.1371/journal.pone.0049790
- Yeatman, J. D., Richie-Halford, A., Smith, J. K., Keshavan, A., and Rokem, A. (2018). A browser-based tool for visualization and analysis of diffusion mri data. *Nat. Commun.* 9:940. doi: 10.1038/s41467-018-03297-7
- Yeh, F.-C., Badre, D., and Verstynen, T. (2016). Connectometry: a statistical approach harnessing the analytical potential of the local connectome. *Neuroimage* 125, 162–171. doi: 10.1016/j.neuroimage.2015.10.053