A Large-Scale Investigation of White Matter Microstructural Associations with Reading Ability

Steven L. Meisler †
Program in Speech and Hearing Bioscience and Technology
Harvard University
43 Vassar Street, Bldg. 46, Room 4033
Cambridge, MA 02139
USA
smeisler@g.harvard.edu

† corresponding author

John D.E. Gabrieli
Department of Brain and Cognitive Sciences and
McGovern Institute for Brain Research
Massachusetts Institute of Technology
43 Vassar Street, Bldg. 46, Room 4033
Cambridge, MA 02139
USA
gabrieli@mit.edu

August 26, 2021

Abstract

Reading involves the functioning of a widely distributed brain network, and white matter tracts are responsible for sending information between constituent network nodes. Several studies have analyzed fiber bundle microstructural properties to shed insights into the neural basis of reading abilities and disabilities. Findings have been inconsistent, potentially due to small sample sizes and varying methodology. To address this, we analyzed a large data set of 690 children ages 5-18 using state-of-the-art neuroimaging acquisitions and processing techniques. We searched for associations between fractional anisotropy (FA) and single-word and nonword reading skills in both typical and poor readers across multiple tracts previously thought to contribute to reading. Consistent with prior studies, FA increased with age across all tracts. There were significant correlations between better reading skills and higher FA in several fiber bundles among poor readers, but fewer and occasionally negative associations among proficient readers. The left superior longitudinal fasciculus and arcuate fasciculus unexpectedly exhibited higher FA within those diagnosed with reading disabilities compared to those without a diagnosis, while their right-hemisphere homologs exhibited higher FA in those with higher reading scores, regardless of diagnostic status. These results suggest a different relation between white matter pathways in children with typical reading ability versus those with reading disabilities.

Keywords Reading · Dyslexia · DTI · Tractometry

Highlights

- Fractional anisotropy (FA) positively correlated with age across all examined tracts.
- We looked for group differences in FA and correlations between FA and reading skills.
- Those with reading disabilities had higher FA in left-hemisphere reading tracts.
- Several tracts' FA positively correlated with reading skills among poor readers.
- Few tracts' FA correlated with reading skill among proficient readers.

1 Introduction

The development of reading skills is vital for progress in education and communication. Reading disability is the most prevalent learning disability and is characterized by difficulty with word reading accuracy and/or fluency [Roongpraiwan et al., 2002, Lyon et al., 2003]. Multiple studies have reported structural and functional brain differences in children and adults with reading disability [Eckert, 2004, Maisog et al., 2008, Pugh et al., 2000, Cattinelli et al., 2013, Richlan et al., 2013, Jednorog et al., 2015, Linkersdörfer et al., 2012], including differences in white matter pathways as measured by diffusion tensor imaging (DTI) [Vandermosten et al., 2012b]. Studies of this nature often report the relationship between reading scores and tract fractional anisotropy (FA), which quantifies how restricted water movement is in a given area [Hagmann et al., 2006, Basser and Pierpaoli, 1996] and can be interpreted as an index of white matter coherence [Beaulieu, 2002]. Such associations may be important to identify because these fiber bundles of myelinated axons connect the gray matter regions that support reading, and especially the left-hemisphere reading network comprised of the angular gyrus, precuneus, middle temporal gyrus, superior temporal gyrus, fusiform gyrus (including the visual word form area), and inferior frontal gyrus (including Broca's area) [Cattinelli et al., 2013, Wandell and Yeatman, 2013, Murphy et al., 2019].

Studies of white matter microstructural properties as they relate to reading abilities in children and adults have yielded interesting, albeit sometimes inconsistent, results. The most commonly implicated tract in DTI studies of reading skills is the superior longitudinal fasciculus (SLF), particularly in the left hemisphere, which connects frontal and temporoparietal brain regions [Wang et al., 2016]. Multiple studies have found that higher FA in bilateral SLF is associated with better reading outcomes, whether that manifests from group comparisons between dyslexic and typically

reading individuals [Richards et al., 2008, Steinbrink et al., 2008, Carter et al., 2009, Marino et al., 2014] or correlations with reading test scores on a continuous scale [Steinbrink et al., 2008, Carter et al., 2009, Hoeft et al., 2011, Feldman et al., 2012, Lebel et al., 2013, Zhang et al., 2014, Horowitz-Kraus et al., 2015, Borchers et al., 2019]. However, other studies that have investigated the SLF, whether from a whole-brain or targeted approach, have failed to replicate this [Odegard et al., 2009, Welcome and Joanisse, 2014, Arrington et al., 2017], and a few studies have even reported negative FA-reading associations in the SLF [Carter et al., 2009, Frye et al., 2011]. The temporal sub-component of the SLF, the arcuate fasciculus (AF), has also been linked to reading performance [Rauschecker et al., 2009], and has often been analyzed separately from other SLF components due to its unique contributions to the language network [Catani et al., 2005]. Similarly, FA reductions in dyslexia have been reported in the left AF [Marino et al., 2014, Klingberg et al., 2000, Deutsch et al., 2005, Vandermosten et al., 2012a, Christodoulou et al., 2017, Su et al., 2018], and its FA has been positively associated with reading skills [Borchers et al., 2019, Klingberg et al., 2000, Deutsch et al., 2005, Christodoulou et al., 2017, Yeatman et al., 2012a, Horowitz-Kraus et al., 2014]. However, bilateral AF regions of higher FA in dyslexia have been identified [Žarić et al., 2018], and separate studies have reported negative FA-reading associations in the left AF [Christodoulou et al., 2017, Yeatman et al., 2012a, Huber et al., 2018]

Two fiber bundles that run under the SLF, the inferior longitudinal fasciculus (ILF) and inferior fronto-occipital fasciculus (IFO), serve to connect occipital and temporal-occipital areas to anterior temporal and frontal regions, respectively [Herbet et al., 2018, Martino et al., 2010], and have been identified as candidate reading tracts [Vandermosten et al., 2012a, Yeatman et al., 2013]. The left ILF has exhibited increased FA in typically developing readers compared to dyslexic readers [Steinbrink et al., 2008, Marino et al., 2014, Su et al., 2018], and bilateral ILF FA has been positively related to reading performance [Steinbrink et al., 2008, Feldman et al., 2012, Lebel et al., 2013, Zhang et al., 2014, Horowitz-Kraus et al., 2015, Odegard et al., 2009, Yeatman et al., 2012a, Horowitz-Kraus et al., 2014]. However, a few studies have found negative associations between FA and reading scores in the left ILF [Yeatman et al., 2012a, Huber et al., 2018], and a study has found increased left ILF FA in dyslexic individuals compared to their neurotypical counterparts [Banfi et al., 2019]. While studies have only identified positive correlations between FA and reading in bilateral IFO [Steinbrink et al., 2008, Feldman et al., 2012, Lebel et al., 2013, Zhang et al., 2014, Odegard et al., 2009, Welcome and Joanisse, 2014, Arrington et al., 2017], only a single study has found a reduction in left IFO FA in dyslexia [Steinbrink et al., 2008]. It is worth noting that other investigations have also yielded null results in these tracts [Borchers et al., 2019, Frye et al., 2011, Klingberg et al., 2000].

The uncinate fasciculus (UF) is thought to contribute to the ventral orthographic pathways of reading, connecting temporal and orbitofrontal regions [Catani et al., 2002, Schlaggar and McCandliss, 2007]. Despite reports of positive associations between FA and reading skills in bilateral UF [Feldman et al., 2012, Zhang et al., 2014, Odegard et al., 2009, Welcome and Joanisse, 2014, Arrington et al., 2017], the only significant group difference in FA that has been reported favored higher FA in dyslexia [Arrington et al., 2017]. The same study also reported a negative correlation between FA and one of their reading measures in the right UF [Arrington et al., 2017]. The splenium of the corpus callosum is also thought to contribute to reading, as it subserves interhemispheric communication between visual cortices [Putnam et al., 2010]. Opposing results for both group [Marino et al., 2014, Frye et al., 2008] and continuous analyses [Feldman et al., 2012, Lebel et al., 2013, Zhang et al., 2014, Odegard et al., 2009, Huber et al., 2018, Frye et al., 2008] have been reported.

One theory of the etiology of dyslexia is the cerebellar hypothesis, which implicates cerebellar dysfunction in deficits of procedural learning and reading fluency [Nicolson and Fawcett, 2007, Nicolson et al., 2001, Stoodley and Stein, 2011]. To this end, the contributions of the superior (SCP), inferior (ICP), and middle (MCP) cerebellar peduncles to reading have been investigated. The SCP contain efferent fibers that connect the deep cerebellum to inferior prefrontal regions involved in reading [Bruckert et al., 2020]. The ICP and MCP contains primarily afferent fibers which connect brainstem with the cerebellum. Such connections may facilitate automation of articulatory and oculomotor control [Bruckert et al., 2020]. Bilateral SCP FA has exhibited negative relations to reading skill [Bruckert et al., 2020, Travis et al., 2015], while the left ICP has shown both positive [Borchers et al., 2019] and negative [Travis et al., 2015] associations. In the MCP, separate studies have paradoxically reported a higher FA in dyslexia compared to typically developing readers [Richards et al., 2008] as well as a positive association between FA and reading skills [Lebel et al., 2013, Travis et al., 2015].

Despite the general trend of higher FA relating to better reading and modest agreement in reading tract outcomes, a recent meta-analysis showed no evidence for systemic FA disruptions in dyslexia [Moreau et al., 2018]. Small cohort sizes, inhomogeneous acquisition parameters, and diversity in processing and analytical methods may underlie the inconsistencies in past results [Ramus et al., 2018, Moreau et al., 2018, Schilling et al., 2021]. To address this, we leveraged the large database from the Healthy Brain Network [Alexander et al., 2017] to investigate white matter microstructural correlates of individual differences in single-word and single-nonword aptitude in children with diverse reading abilities. We additionally looked for tract-specific differences in FA between typical and poor readers. We implemented two criteria independently for delineating between reading proficiency groups, one based on diagnoses

by clinicians of specific learning disorder with impairment in reading (similar to dyslexia), and the other based on a standardized reading measure.

2 Methods

2.1 Participants

We downloaded data from 1221 participants across the first 8 data releases of the Healthy Brain Network project [Alexander et al., 2017]. Participants were all scanned at Rutgers University. All data were accessed in accordance with a data use agreement provided by the project team. The Healthy Brain Network project was approved by the Chesapeake Institutional Review Board (https://www.chesapeakeirb.com/). The research team obtained written informed consent from participants age 18 or older. For younger participants, written informed consent was collected from their legal guardians, while written assent was obtained from the participant. Full inclusion and exclusion criteria are described in the project's publication [Alexander et al., 2017]. Of note, each participant was fluent in English, had an IQ over 66, and did not have any physical or mental disorder precluding them from completing the full battery of scanning and behavioral examinations. Several behavioral and cognitive evaluations were collected as part of the project. Relevant to this study, participants completed the Test of Word Reading Efficacy 2nd edition (TOWRE) [Torgesen et al., 1999] and the Edinburgh Handedness Inventory (EHI) [Oldfield, 1971].

The TOWRE consists of two subtests, Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE). For Sight Word Efficiency, each participant is shown a list of words and asked to read the words aloud as quickly as possible. Raw scores are based on the number of words read correctly within the 45-second time limit and then converted to a standard score (population mean = 100, SD = 15). For Phonemic Decoding Efficiency, each participant is shown a list of pseudowords and asked to read the pseudowords aloud as quickly as possible. Raw scores are based on the number of pseudowords read correctly within the 45-second time limit and then converted to a standard score (population mean = 100, SD = 15). The composite TOWRE score is the mean of the two standardized scores.

After quality control (QC), there were 690 participants aged 5-18 years old. Two sets of comparisons were made in this study. The first comparison was between children with diagnosed reading disabilities (RD, n = 106) and typically reading children (TR, n = 584). Participants were included in the RD group if they were diagnosed with "Specific Learning Disorder with Impairment in Reading." These diagnoses came from a panel of clinicians associated with the Healthy Brain Network and were made following the 5th edition of the Diagnostic and Statistical Manual for Mental

Disorders [Edition et al., 2013], although the specific criteria were not made available. All other participants were included in the TR group. The second comparison used a TOWRE cutoff of 85 to delineate poor (Low-TOWRE, n = 165) and proficient (High-TOWRE, n = 525) readers (Figure 1). This threshold is 1 standard deviation below the population mean and is conventionally used for diagnosing reading disabilities [Heath et al., 2006, Pugh et al., 2014, Johnston et al., 2016].

We compared the ages, sex distribution, and handedness between the groups (Table 1). Both the TR and High-TOWRE groups were older than their poorer reading counterparts (for both comparisons, $p < 0.05$, Welch's t-test). There were overall more males than females in the cohort, and the proportions of sexes did not significantly differ between groups (between RD and TR: $p > 0.05$; between Low-TOWRE and High-TOWRE: $p > 0.9$; χ^2 tests). Handedness did not differ between groups (for both comparisons, $p > 0.4$, Welch's t-test).

Phenotype	**All**	**TR**	**RD**	**High-TOWRE**	**Low-TOWRE**
n	690	584	106	525	165
Sex [M / F]	450 / 240	389 / 195	61 / 45	343 / 182	107 / 58
Age [years]	10.71 (3.08)	10.81 (3.17)*	10.17 (2.48)*	10.87 (3.17)†	10.20 (2.73)†
Handedness [EHI]	61.38 (51.01)	60.67 (50.43)	65.32 (53.92)	60.47 (49.89)	63.01 (54.39)

Table 1: Phenotype information in the different reading proficiency groups. Values are listed as mean (standard deviation). For group comparisons of age, * denotes $p < 0.05$, and † denotes $p < 0.01$ (Welch's t-tests).

2.2 Neuroimaging Acquisition

Participants were scanned using a Siemens 3T Tim Trio scanner while wearing a standard Siemens 32-channel head coil. A high resolution T1-weighted (T1w) sequence was collected with the following parameters: TR = 2500 ms, TE = 3.15 ms, Flip Angle = 8°, 0.8 mm isotropic voxel resolution. A diffusion kurtosis imaging scan was administered with the following parameters: TR = 3320 ms, TE = 100.2 ms, Flip Angle = 90°, 1.8 mm isotropic voxel resolution, 1 b = 0 image, 64 noncollinear directions collected at b = 1000 s/mm^2 and b = 2000 s/mm^2. A pair of PEpolar fieldmaps were collected before the diffusion scan to quantify magnetic field inhomogeneity. Detailed scanner protocols are published on the Healthy Brain Network project website (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/File/mri/HBN_RU_Protocol.pdf).

2.3 Neuroimaging Preprocessing

Results included in this manuscript come from preprocessing performed using *QSIPrep* 0.13.0RC1 [Cieslak et al., 2021, RID:SCR_016216] (`https://qsiprep.readthedocs.io/en/latest/`) which is based on *Nipype* 1.6.0 [Gorgolewski et al., 2011, 2018, RRID:SCR_002502]. Many internal operations of *QSIPrep* use *Nilearn*

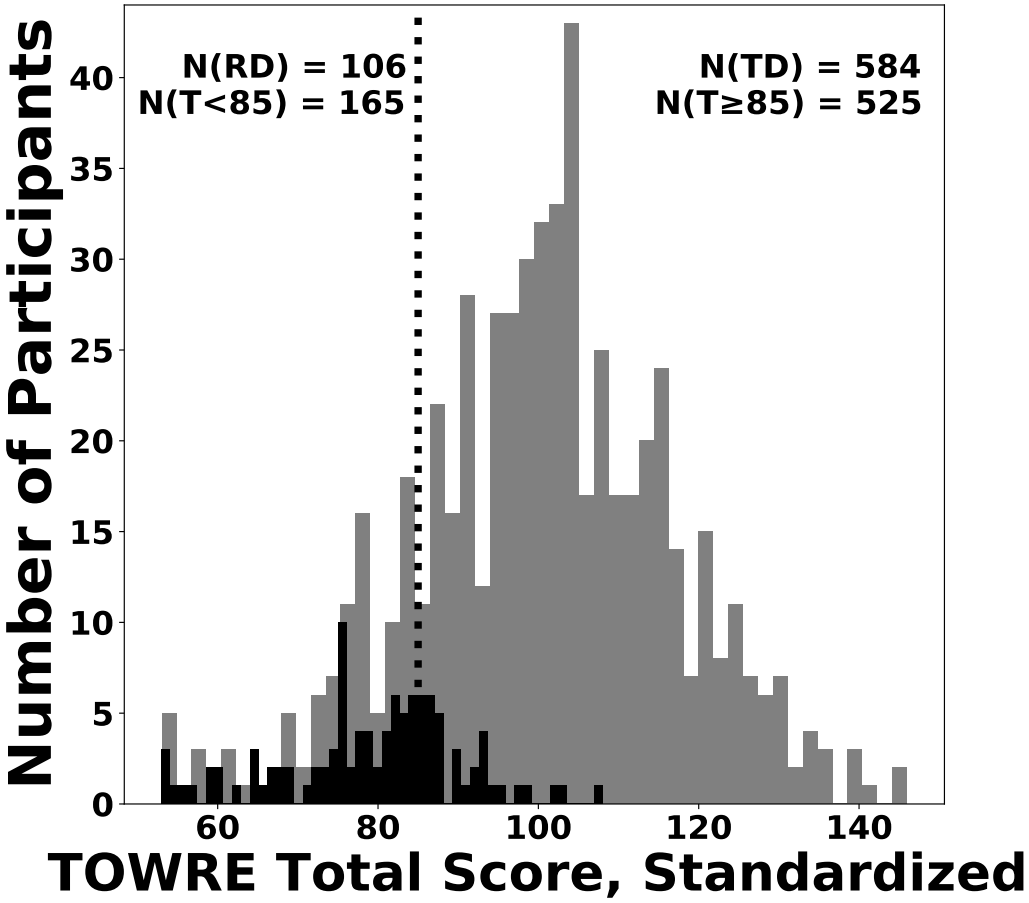

Figure 1: Distribution of standardized TOWRE composite scores across all participants. Gray bars denote TR participants, and black bars denote RD participants. The black dotted line marks a TOWRE score of 85 which separates the Low-TOWRE and High-TOWRE groups.

0.7.0 [Abraham et al., 2014, RRID:SCR_001362] and *Dipy* 1.3.0 [Garyfallidis et al., 2014, RRID:SCR_000029]. For more details of the pipelines, see the section corresponding to workflows in *QSIPrep*'s documentation (`https://qsiprep.readthedocs.io/en/latest/preprocessing.html#preprocessing-pipeline-details`). The text in the following two sections was provided by *QSIPrep* under a CC0 license with the *explicit instruction* that it be included in a manuscript for the sake of transparency and reproducibility. We made minor changes for readability.

2.3.1 Anatomical Preprocessing

The T1w image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` [Tustison et al., 2010], distributed with *ANTs* 2.3.3 [Avants et al., 2008, RRID:SCR_004757], and used as the T1w-reference

throughout the workflows. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from *ANTs*), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` [Zhang et al., 2001, *FSL* 5.0.9, RRID:SCR_002823].

2.3.2 Diffusion MRI Preprocessing

MP-PCA denoising as implemented in *MRtrix3*'s `dwidenoise` [Veraart et al., 2016] was applied with a 5-voxel window. After MP-PCA, Gibbs unringing was performed using *MRtrix3*'s `mrdegibbs` [Kellner et al., 2016]. Following unringing, B1 field inhomogeneity was corrected using `dwibiascorrect` from *MRtrix3* with the N4 algorithm [Tustison et al., 2010]. After B1 bias correction, the mean intensity of the diffusion-weighted imaging (DWI) series was adjusted so all the mean intensity of the b = 0 images matched across each separate DWI scanning sequence.

FSL's (version 6.0.3:b862cdd5) `eddy` function was used for head motion correction and Eddy current correction [Andersson and Sotiropoulos, 2016]. The function was configured with a q-space smoothing factor of 10, a total of 5 iterations, and 1000 voxels used to estimate hyperparameters. A linear first level model and a linear second level model were used to characterize Eddy current-related spatial distortion. q-space coordinates were forcefully assigned to shells. Field offset was attempted to be separated from participant movement. Shells were aligned post-`eddy`. `eddy`'s outlier replacement was run [Andersson et al., 2016]. Data were grouped by slice, only including values from slices determined to contain at least 250 intracerebral voxels. Groups deviating by more than 4 standard deviations from the prediction had their data replaced with imputed values. Data were collected with reversed phase-encoded blips, resulting in pairs of images with distortions going in opposite directions. Here, b = 0 reference images with reversed phase encoding directions were used along with an equal number of b = 0 images extracted from the DWI scans. From these pairs the susceptibility-induced off-resonance field was estimated using a method similar to that described in [Andersson et al., 2003]. These susceptibility maps were ultimately incorporated into the Eddy current and head motion correction interpolation. Final interpolation was performed using the `jac` method.

Several confounding time-series were calculated based on the preprocessed DWI: framewise displacement (FD) using the implementation in *Nipype* (following the definitions by Power et al. [2014]). The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. Slicewise cross correlation was also calculated. The DWI time-series were resampled to ACPC, generating a preprocessed DWI run in ACPC space with 1.2mm isotropic voxels.

2.4 Tractometry and Statistical Analysis

We used *TractSeg* version 2.2 [Wasserthal et al., 2018a] (`https://github.com/MIC-DKFZ/TractSeg`), a deep-learning based white matter segmentation method, to reconstruct white matter tracts and perform tractometry with FA values. We chose this method due to its favorable balance between the accuracy of manual fiber tracking and objectivity of atlas-based methods [Genc et al., 2020]. This involved the following steps: First, we fit the diffusion tensor with a weighted least-squares algorithm [Chang et al., 2005] and produced an FA map using *TractSeg*'s `calc_FA` function, which is based off of *Dipy* [Garyfallidis et al., 2014]. Then, a DWI-to-MNI transformation was calculated with *FSL* 6.0.4's `flirt` [Jenkinson and Smith, 2001] by rigidly aligning the participant's FA map to a MNI FA template distributed with *TractSeg*. This transformation was additionally applied to the preprocessed diffusion image and brain mask. We did this to ensure orientation homogeneity across subjects. B-vectors were rotated accordingly using *TractSeg*'s `rotate_bvec` function. Multi-tissue fiber response functions were estimated using *MRtrix3*'s `dhollander` algorithm [Dhollander et al., 2016, 2019]. Fiber orientation densities (FODs) were estimated via multi-shell multi-tissue constrained spherical deconvolution (CSD) [Tournier et al., 2004, 2008, Jeurissen et al., 2014]. FODs were intensity-normalized using `mtnormalize` [Raffelt et al., 2017]. The first three principal FOD peaks were extracted and flipped along their x-axis to correct their orientation.

We then segmented 9 bilateral tracts: AF, SLF (I, II, and III), ILF, IFO, UF, SCP, and ICP. We also reconstructed the MCP and splenium of the corpus callosum (CC_7 by *TractSeg* naming convention), leading to a total of 20 fiber bundles to analyze. To make each tract, we fed the FOD peaks into *TractSeg* to create binary masks of the tract, the ROIs that define its endpoint, and a corresponding tracking output map (TOM) [Wasserthal et al., 2018b]. These were used as inputs to a pre-trained convolutional neural network to output the segmented fiber bundle [Wasserthal et al., 2019, 2020]. We generated a fixed large number (5000) of fibers per instance to account for the stochastic nature of reconstruction. This reduces inter-run variability. Using an approach based on *Dipy*'s Bundle Analytics framework [Chandio et al., 2020], FA values were calculated for 100 nodes along the length of the tract. At each node, we employed a non-parametric permutation-based statistical comparison [Nichols and Holmes, 2002] with 5000 iterations to look for correlations (Pearson's r) between node FA and standardized TOWRE composite scores, accounting for multiple comparisons given the correlative structure between adjacent nodes [Yeatman et al., 2012b]. We linearly regressed out sex (binary 0 or 1) and handedness (EHI score) before performing the regression. Additionally, we ran the same model but correlated FA against single-word (SWE) and nonword (PDE) standardized subscores separately. Finally, we looked for group differences in FA in each tract between reading proficiency groups, employing a two-sample t-test at each

node while regressing out age, sex, and handedness. Given the recency of our analysis methods, we first checked for associations between FA and age as a way of validating our method, given the well-established trend of FA increasing with age [Lebel et al., 2019]. For this model we also regressed out sex and handedness.

2.5 Data Inclusion and Quality Control

Of the original 1221 participants, 858 participants had all the necessary neuroimaging data (T1w, diffusion, and fieldmap) and were able to be run through *QSIPrep* and *TractSeg* without errors. 6 of those participants had incomplete fiber bundle segmentations. 710 of the remaining participants had all the necessary phenotypic data. An additional 17 participants were excluded for being over 18 years old. For the remaining participants, due to the sheer volume of images included in the cohort, visual inspection of each DWI image was not possible. We instead adopted an automated QC procedure described in [Yeh et al., 2019], which has been integrated into the outputs of *QSIPrep*. This involves rejecting a scan if it had different or incomplete scanning acquisitions (no participants excluded) or if over 0.1% of slices (9 slices at 72 slices/volume $\times$ 128 diffusion volumes) had significant signal dropout based on changes in slice-wise correlation (3 participants excluded). Therefore, a total of $n = 690$ participants were analyzed.

2.6 Data and Code Availability

Neuroimaging and phenotypic data can be collected following directions on the Healthy Brain Network Data Portal (`http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/index.html`) after signing a data use agreement. We cannot distribute this data publicly. All code for preprocessing neuroimaging data as well as instructions for replicating this study can be found at `https://github.com/smeisler/Meisler_ReadingFA_Associations`. With minimal modification, the preprocessing code should be able to run on most BIDS-compliant data sets using the SLURM job scheduler [Yoo et al., 2003]. Some software we used was distributed as Docker [Merkel, 2014] containers, and compiled and run with Singularity (3.6.3) [Kurtzer et al., 2017]:

- *QSIPrep* 0.13.0RC1 (`singularity build qsiprep.simg docker://pennbbl/qsiprep:0.13.0RC1`)
- *TractSeg* 2.2 (`singularity build tractseg.simg docker://brainlife/tractseg:2.2`)
- *MRtrix* 3.0.2 with *FSL* 6.0.4 and *ANTs* 2.3.4 (`singularity build mrtrix_fsl_ants.simg docker://sclove/mrtrix3.0.2:latest`)

Newer versions of these softwares have since been produced, and we encourage anyone to use the latest stable releases.

3 Results

3.1 Age-FA Correlations

To validate our methods, we checked for associations between FA and age in the entire participant cohort, expecting to see ubiquitously strong positive associations in each tract. All nodes in all 20 tracts besides the splenium, MCP, bilateral SCP, and right ILF, IFO, and UF were significantly positively correlated. In these seven fiber bundles, significant positive associations were still present, but not across the whole tract (Figure 2).

3.2 Reading-FA Associations

We correlated tract FA in each node with the participants' standardized TOWRE composite scores (Table 2). Among the Low-TOWRE group, significant positive correlations were present in all 20 tracts besides the left IFO and SLF II, which both had trends towards significance (Figures 3, S1). In the High-TOWRE group, only a single *negative* correlation between right ILF FA and TOWRE scores reached significance (Figure 4). Positive correlation prevalence was lower among the RD group compared to the Low-TOWRE group, as no nodes in the splenium, right UF, right IFO, or any cerebellar peduncle reached significance (Figure 5). In the TR group, unlike in the High-TOWRE group, a few tracts, including bilateral AF and right SLF I and II, had significant positive associations with reading aptitude, while no tract's FA negatively correlated with TOWRE scores (Figure 6). Across the entire cohort, fewer tracts' FA values were significantly positively related to reading skills compared to within either the RD or Low-TOWRE group alone (Figure S2).

We repeated the same analyses but instead correlated against the SWE and PDE TOWRE subscores (Figures S3-S12; Table 2) to probe whether there were similar or dissimilar relations between white matter and performance with words (SWE) versus pseudowords (PDE). Of note, more tracts' FA positively correlated with PDE scores across all subjects, compared to associations with SWE or TOWRE composite scores. In addition, more fiber bundles exhibited negative correlations with SWE scores among the High-TOWRE group compared to associations with other reading scores in other reading groups.

3.3 Group Differences in FA

We looked for differences in FA between the RD and TR groups and between the Low-TOWRE and High-TOWRE groups (Table 2). The right AF, IFO, and SLF I and II exhibited significantly greater FA for the High-TOWRE

group than the Low-TOWRE group (Figures 7, S13). Four tracts, the left AF and all three SLF components, showed significantly higher FA in the RD group than the TR group (Figures 8, S13).

Tract		Group Difference	Reading Metric														
			TOWRE Composite					SWE					PDE				
			TR	RD	HT	LT	All	TR	RD	HT	LT	All	TR	RD	HT	LT	All
AF	L	RD > TR	+	+		+			+				+				+
	R	HT > LT	+	+		+	+				+	+	+	+		+	+
ILF	L			+		+			+	-	+			+		+	
	R			+	-	+			+	-	+			+		+	
IFOF	L			+						-							
	R	HT > LT				+	+			-	+		+			+	+
SLF I	L	RD > TR		+		+							+				+
	R	HT > LT	+	+		+	+	+			+	+	+	+		+	+
SLF II	L	RD > TR		+			+										+
	R	HT > LT	+	+		+	+		+		+		+	+		+	+
SLF III	L	RD > TR		+		+			+		+			+		+	
	R			+		+							+				+
UF	L			+		+			+					+			+
	R					+				-	+		+			+	+
SCP	L					+					+					+	
	R					+				-			+			+	+
ICP	L					+					+		+		+	+	+
	R					+					+		+		+	+	
MCP						+					+					+	+
Splenium						+											

Table 2: A summary of results across the group difference and correlation analyses. LT and HT are used to abbreviate the Low and High-TOWRE groups, respectively. A "+" or "-" sign denotes a significant positive or negative association between FA and reading scores, respectively.

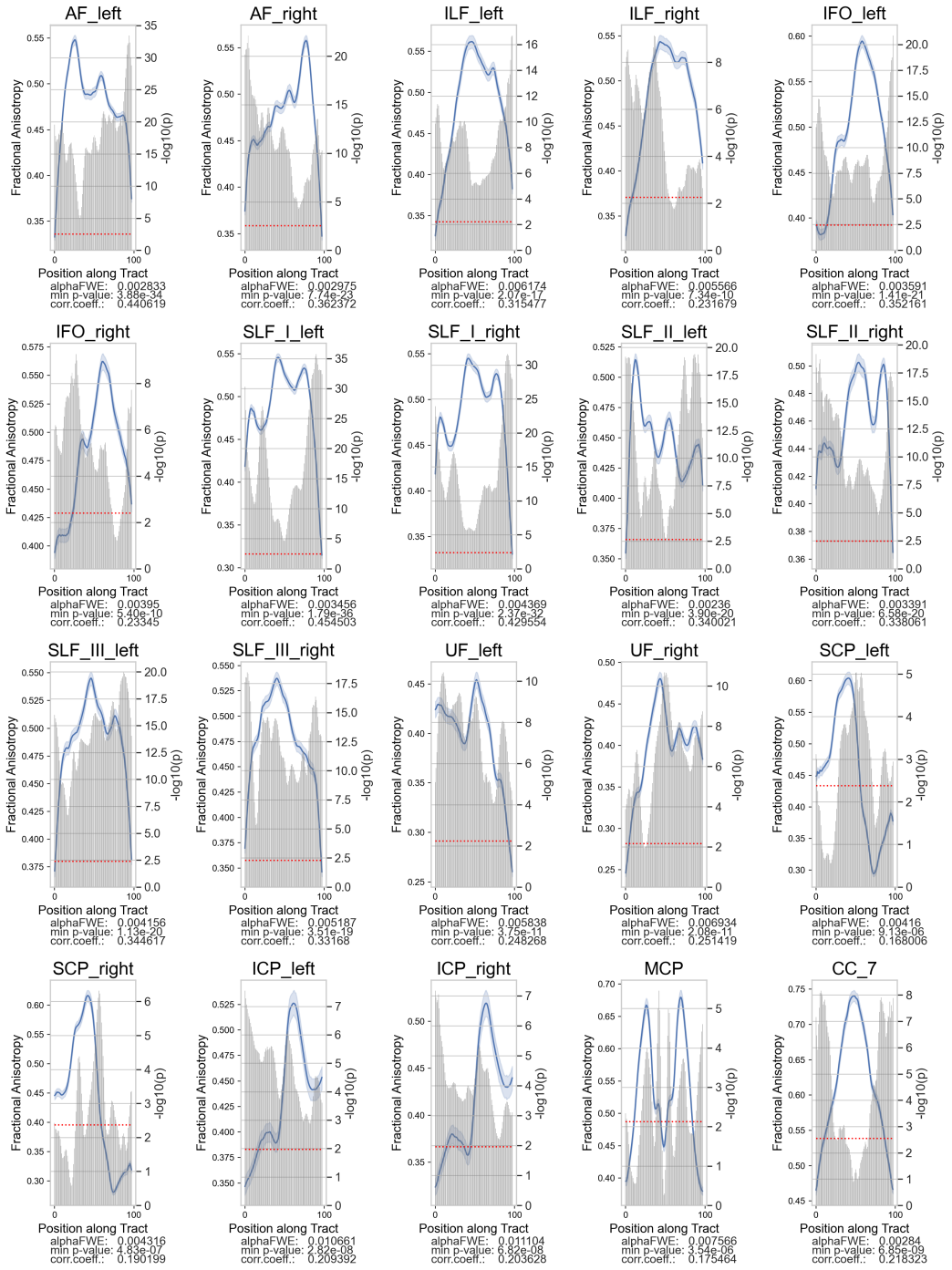

Figure 2: All 20 tracts exhibited significant FA correlations with age across the entire cohort. Each plot represents a different fiber bundle. The blue curve represents the average FA after regressing out sex and handedness. The shaded region surrounding the curve is the 95% confidence interval. The red dotted line depicts the critical FWE-adjusted α (given an uncorrected α of 0.05). The gray bars represent the p-value of the correlation at each node, measured on a -log scale. A gray bar extending above the red line ($p < \alpha$FWE) denotes a significant node. Below each plot, the critical FWE-adjusted α is listed, along with the strongest correlation achieved in the tract and its associated p-value.

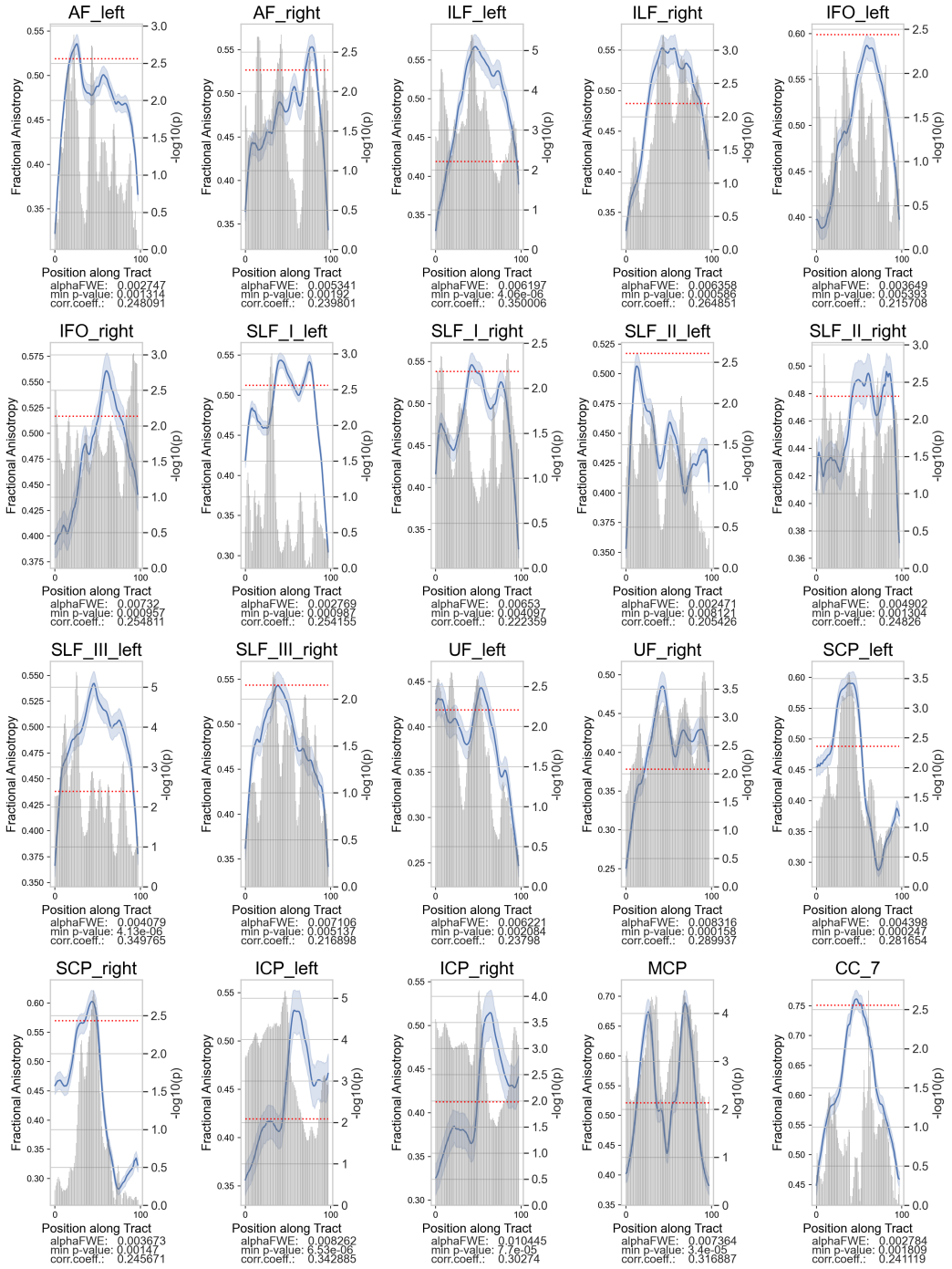

Figure 3: Several tracts exhibited significant positive FA correlations with reading ability (TOWRE composite score) among the Low-TOWRE group. Plotting convention is the same as Figure 2.

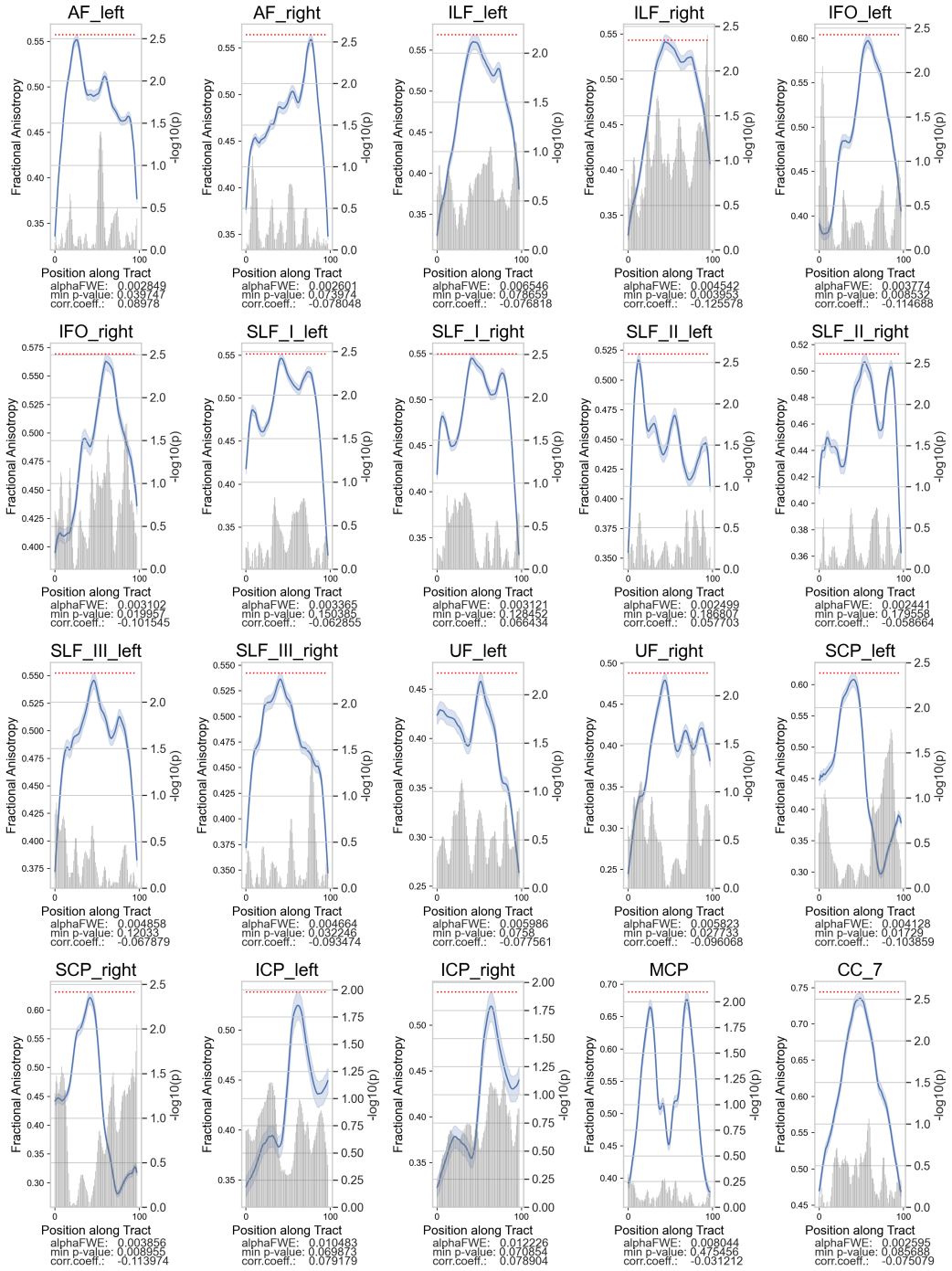

Figure 4: Only the right ILF exhibited a negative correlation with TOWRE composite scores among the High-TOWRE group. Plotting convention is the same as the previous figure.

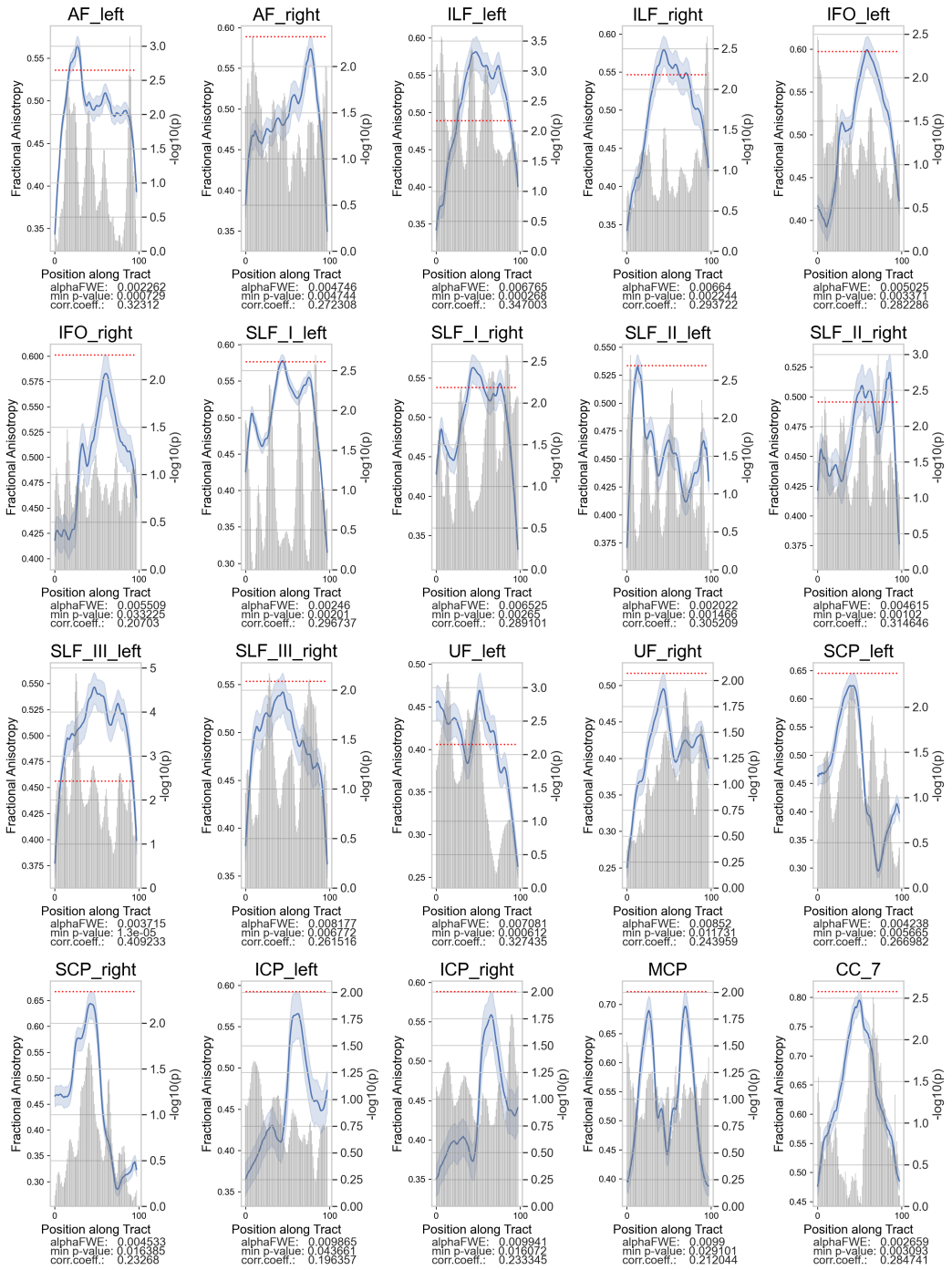

Figure 5: Several tracts exhibited significant positive FA associations with TOWRE composite scores among the RD group. Plotting convention is the same as the previous figure.

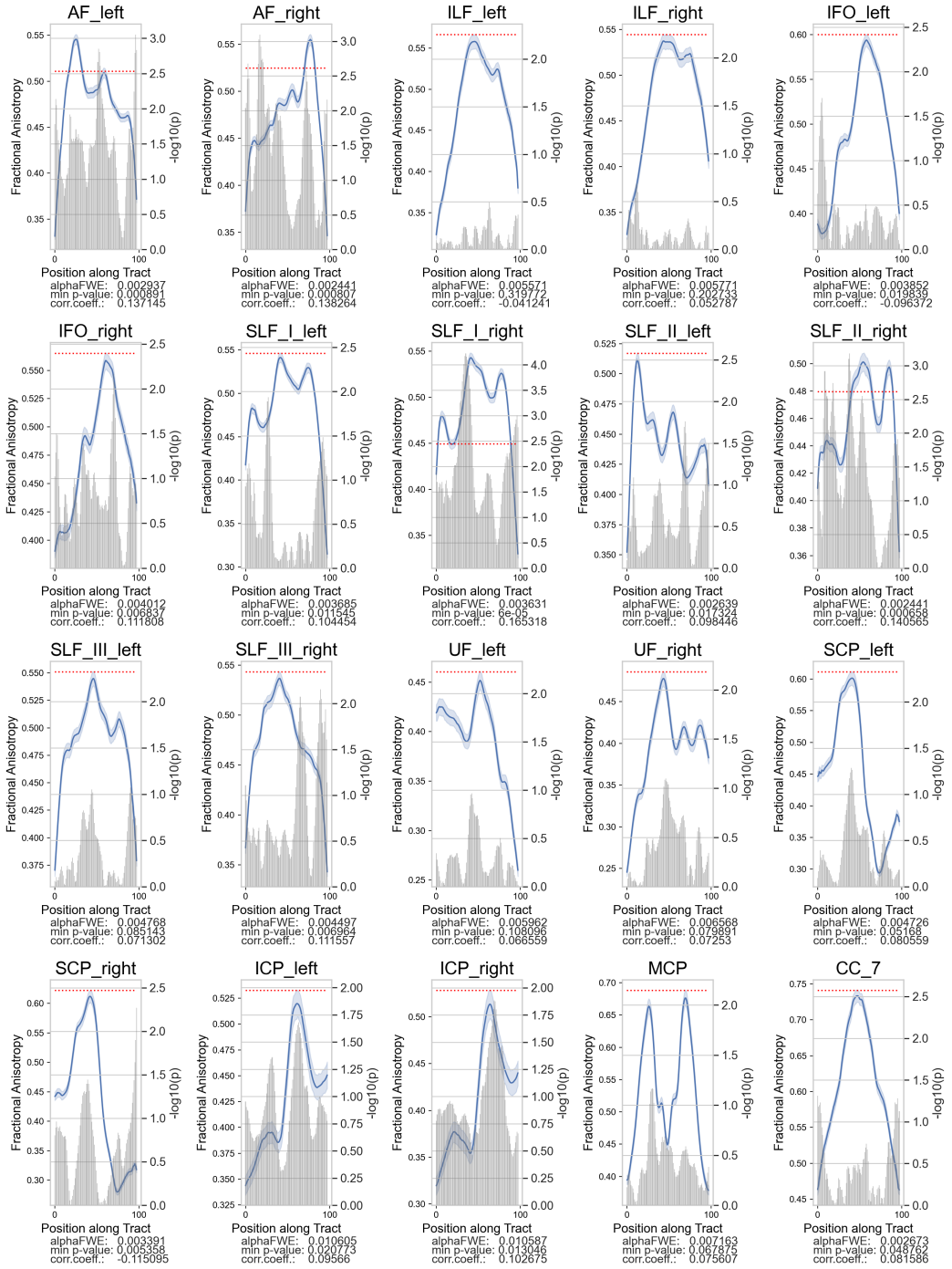

Figure 6: Only a few tracts exhibited significant positive FA associations with TOWRE composite scores among the TR group. Plotting convention is the same as the previous figure.

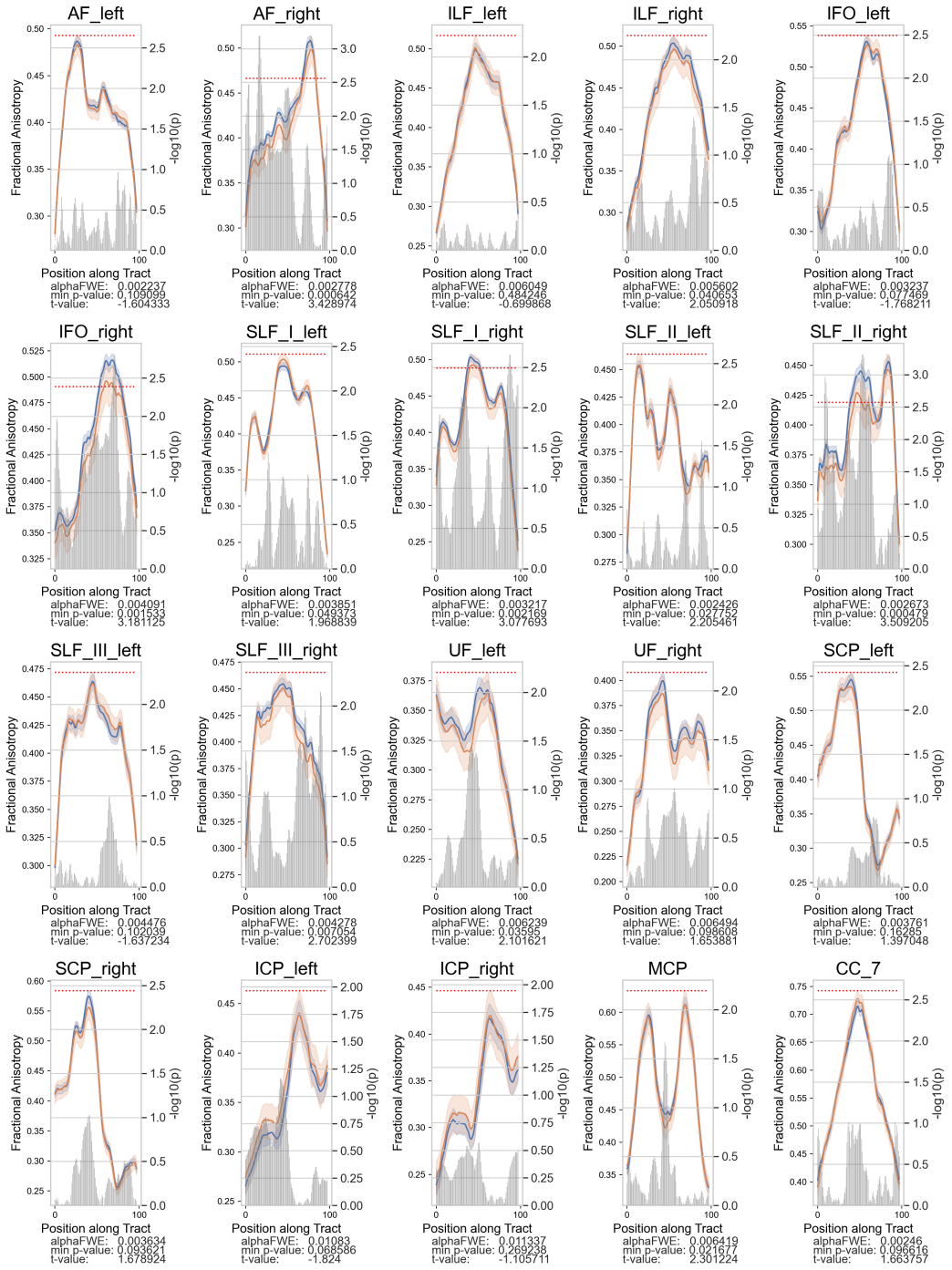

Figure 7: The FA of the right AF, SLF, and IFO was greater in the High-TOWRE group compared to the Low-TOWRE group. Each plot represents a different fiber bundle. The blue line represents the average FA in the High-TOWRE group, and the orange curve is the average FA in the Low-TOWRE group. Both lines have age, sex, and handedness regressed out. The shaded regions surrounding the lines are the 95% confidence intervals. Below each plot, the FWE-adjusted α (given an original $\alpha = 0.05$) is listed, along with the largest t-statistic (positive indicates High-TOWRE > Low-TOWRE) achieved in the tract and its associated p-value. The gray bars and dotted red lines retain their meanings from previous figures.

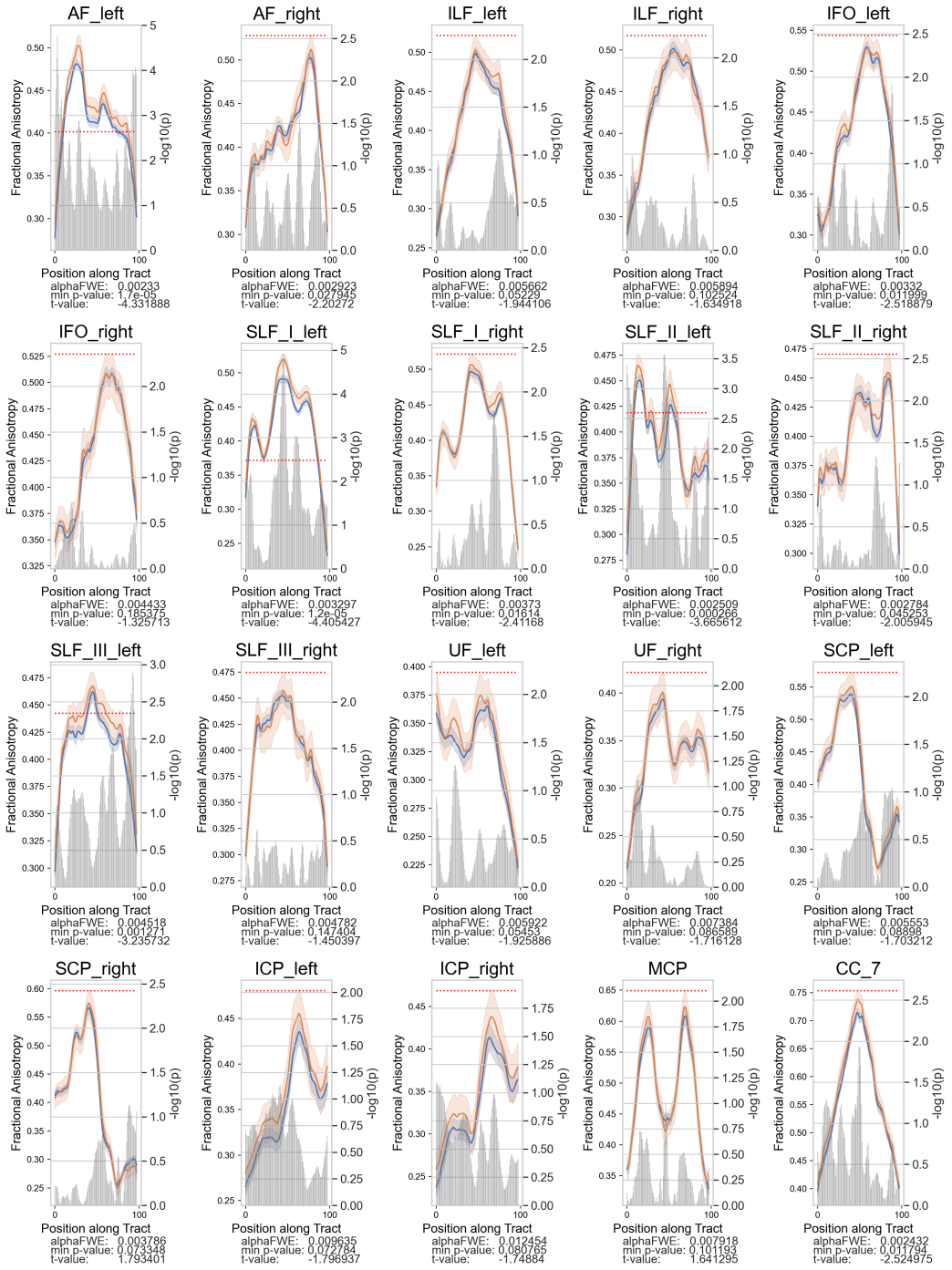

Figure 8: The FA of the left AF and SLF was greater in the RD group compared to the TR group. Each plot represents a different fiber bundle. The blue line represents the average FA in the TR group, and the orange curve is the average FA in the RD group. Both lines have age, sex, and handedness regressed out. The shaded regions surrounding the lines are the 95% confidence intervals. Below each plot, the FWE-adjusted α (given an original $\alpha = 0.05$) is listed, along with the largest t-statistic (positive indicates TR > RD) achieved in the tract and its associated p-value. The gray bars and dotted red lines retain their meanings from previous figures.

4 Discussion

This study is the first large-scale investigation of tractometry-derived white matter microstructural associations with single-word reading abilities across a diverse child participant cohort. The size of our participant cohort is large relative to other DTI studies, particularly those relating to reading aptitude. We used high-quality publicly available data and state-of-the-art analytical methods, showcasing the rapid advances in diffusion acquisition and processing techniques. Because clinical diagnoses of reading disabilities are somewhat subjective and reading thresholds in cutoff-based approaches are set somewhat arbitrarily, we examined both of these two methods of delineating poor and typical readers separately. In general, more associations between higher FA and better reading skills were present in poorer readers, regardless of how they were defined. There were very few correlations between FA and reading skills in the more proficient reading groups, and among them were associations between lower FA and higher reading skills. The reading disability (RD) group exhibited higher FA than the typical reading (TR) group in left-hemisphere tracts often associated with reading, including the left arcuate fasciculus (AF) and the left superior longitudinal fasciculus (SLF). Conversely, when children were divided into groups based on reading skills alone (TOWRE scores), children with better reading skills (High-TOWRE), relative to children with worse reading skills (Low-TOWRE), exhibited higher FA in multiple right-hemisphere tracts.

We found significant correlations between higher FA and age in all tracts. Our study is the largest investigation of linear white matter development patterns to date (for a review, see Lebel et al. [2019]; for a similarly-sized group-based analysis, see Chiang et al. [2011]). Our findings are consistent with prior reports of linear FA-age relationships during child development [Barnea-Goraly et al., 2005, Bonekamp et al., 2007, Muetzel et al., 2008, Mukherjee et al., 2002, Schmithorst et al., 2002], although more recent studies suggest an exponential trajectory [Lebel et al., 2008, Tamnes et al., 2010, Taki et al., 2013, Simmonds et al., 2014] which we did not evaluate in this study. The consistency of our developmental findings of widespread growth of FA with age with prior studies also confirms the quality of the DTI acquisition and the validity of our analytic method.

This is the first study of reading skills that has used *TractSeg*, which has been shown to outperform several popular white matter segmentation methods [Wasserthal et al., 2018a]. Newer tract-based approaches, as employed here, represent a paradigm shift from several earlier DTI studies. Many previous papers of FA-reading relationships, particularly before the publication of Vandermosten and colleagues' review of DTI applications to reading [Vandermosten et al., 2012b], performed whole-brain voxel-based analyses (VBA). VBA sensitivity suffers from stricter multiple-comparison correction across the entire brain. In addition, VBA methods tend to be less precise due to their being run on a

group-averaged image because the shape of long-range fiber bundles varies among people [Wassermann et al., 2011, Yeatman et al., 2011]. Spatial smoothing and affine transformation to MNI before group analysis in VBA, as opposed to the rigid alignment employed in this study to ensure orientation homogeneity, may obfuscate unique properties of a participant's anatomy [Christensen et al., 1997]. Significant findings from VBA are not always assigned to tracts. In fact, a single voxel may contain multiple fiber bundles, so it is possible early studies may have incorrectly ascribed significant voxels to fiber bundles due to not also considering the primary diffusion directionality in these areas. Tract-Based Spatial Statistics (TBSS, [Smith et al., 2006]), a popular method that improves upon traditional VBA by restraining analyses to a skeletonized white matter voxel map, also suffers in tract localization [Tsang et al., 2010]. Methodological delves into TBSS have revealed bias in the FA map used for spatial normalization, sensitivity to pathologies and noise that affect brain anatomy, incorrect voxel-to-tract assignments [Bach et al., 2014], and instability dependent on tensor-fitting methods [Maximov et al., 2015]. VBA tends to depend on FA to determine tissue class (e.g., white vs gray matter). FA is prone to underestimation and noise in voxels that contain both gray and white matter or crossing fibers [Oouchi et al., 2007], which would especially confound segmentation of the corpus callosum. In contrast, tracking with CSD methods first, as employed here, resolves crossing fibers better [Tournier et al., 2008]. VBA can still be valuable as a data exploratory tool, but tract-based methods likely provide for superior sensitivity and reduced ambiguity of fiber bundle localization (for a review, see Ramus et al. [2018]). These benefits have already been observed in multiple clinical populations [Wasserthal et al., 2020, Kamagata et al., 2013, Kuchling et al., 2018].

With these limitations in mind, it may not be surprising that tract-based methods have yielded occasionally opposing findings to voxel-based methods. For example, in the left AF, observations of *lower* FA being associated with higher reading scores came from tract-based methods [Yeatman et al., 2012a, Christodoulou et al., 2017, Huber et al., 2018], as did findings of *higher* FA in dyslexia compared to controls [Žarić et al., 2018]. This could reflect increased anatomical specificity and/or mitigated partial volume effects. However, tractography applications to reading are still in their infancy, and different tractography measures may provide different insights. For example, CSD methods are better at resolving fiber bundles that pure DTI methods have difficulty with, such as the right AF [Yeatman et al., 2011, Catani et al., 2007, Zhao et al., 2016]. The field may best be served by harmonizing diffusion acquisition and analysis protocols [Cieslak et al., 2021, Schilling et al., 2021, Ramus et al., 2018], particularly scanning with multiple shells at high b-values (at least b = 1300) to optimally use CSD.

There is debate surrounding whether dyslexia can properly be described as a neurodevelopmental difference. Population reading skills tend to follow a normal distribution, even when including those with developmental dyslexia (Figure 1).

One may argue that dyslexia is not a brain difference, but rather a way to group those on the lower end of the reading skills distribution (or bell curve). From this perspective, both reading differences and associated brain differences should lie along a continuum [Protopapas and Parrila, 2018, 2019]. Findings that would most strongly align with this view would be if reading-FA correlations were positively correlated across the *entire* participant cohort. However, that was rarely the case (Table 2), with the most obvious difference being the near-ubiquitous positive reading-FA associations in the Low-TOWRE groups (Figure 3) and non-significant correlations in the High-TOWRE group (Figure 4). While we cannot draw any causal conclusions, our results suggest a different relation between white matter structural properties and reading skills in children with lower reading ability versus those with typical reading skills.

This stark difference in FA-reading associations between proficiency groups may also be the result of extensive reading experience (or training), in and out of school, and additionally intervention experiences for children with reading difficulty. Training in a given skill, such as reading, can lead to an initial increase in FA in putative white matter regions, followed by a decline when reaching proficiency [Scholz et al., 2009]. Such trends have been observed in mathematical abilities in dyslexia [Koerte et al., 2016], musicianship [Schmithorst and Wilke, 2002], and dance [Hänggi et al., 2010]. This could explain why we observed near-ubiquitous positive associations between FA and reading abilities in the Low-TOWRE group (who are not proficient), but not in the High-TOWRE group (who are proficient).

There has also been some evidence that poor readers utilize right-hemisphere pathways to compensate for limitations of the left-hemisphere pathways that are typically most associated with single-word reading. Several studies have reported enhanced right-hemisphere brain activation in individuals with dyslexia while reading, which has been interpreted as a compensatory mechanism [Hoeft et al., 2011, Milne et al., 2002, Temple et al., 2003, Waldie et al., 2013]. In the Low-TOWRE group, every right-sided tract we analyzed contained significant positive FA correlations with better reading skill. Right-hemisphere pathway recruitment during reading training in low proficiency readers may reflect this continuing training towards single-word reading proficiency.

The present findings are surprising in several ways, especially given the exceptional statistical power of the present sample. First, it is noteworthy that there were so few associations between white-matter microstructure and single-word reading proficiency among the 584 typically reading children. The use of age-standardized reading scores controlled for simple age-related gains in reading, but there was little relation between reading ability and white-matter pathways. Second, the finding that children classified as having a reading disability had greater FA in left-hemisphere pathways associated with reading (left AF and SLF) differs from a frequent focus in dyslexia research on reduced FA in left-hemisphere tracts connecting the major occipito-temporal, temporal-parietal, and frontal nodes of the reading network.

Reduced FA in these tracts has been observed in both children and adults with dyslexia [Klingberg et al., 2000, Richards et al., 2008, Steinbrink et al., 2008, Carter et al., 2009, Christodoulou et al., 2017, Su et al., 2018, Žarić et al., 2018], children with reduced pre-reading language skills [Saygin et al., 2013], and even infants from families with a history of dyslexia [Langer et al., 2017]. In contrast, multiple other studies have either failed to find this difference or had the opposite findings of increased FA in the left-hemisphere tracts [Arrington et al., 2017, Žarić et al., 2018]. Indeed, a meta-analysis concluded that there were no reliable findings relating either reading skill as a continuous ability or comparing groups with versus without dyslexia to variation in white-matter microstructure [Moreau et al., 2018].

The results of this study should be interpreted in the context of some limitations. First, the composition of the participant cohort is unique in that most participants in the Healthy Brain Network have at least one psychological, neurodevelopmental, or learning disorder [Alexander et al., 2017]. Incentives for participating in the study included a cash reward as well as a free behavioral examination, and advertisements were targeted towards parents who may have been concerned about their child's psychological state. The proportion of Low-TOWRE and RD participants in the cohort were approximately 24% and 15% respectively. Both of these figures exceed estimates of dyslexia prevalences which tend to range from 5-10% [Roongpraiwan et al., 2002]. Secondly, we used FA as our tract metric of interest since it has largely been interpreted as a measure of tract integrity. That is, a higher FA may relate to more myelination, which in turn may increase synaptic efficiency [Lebel and Deoni, 2018]. While myelination certainly modulates FA, other factors such as axonal diameter, density, and coherence also influence it [Shemesh, 2018, Friedrich et al., 2020, Beaulieu, 2009]. The dueling mechanisms underlying FA may confound results, as both axonal pruning and increased myelination may relate to positive brain development but have opposing effects on FA [Yeatman et al., 2012a]. Analysis of T1 relaxation time in tracts using quantitative MRI may better quantify the degree of myelination per se [Lutti et al., 2014, Schurr et al., 2018]. The tract-based approach employed here, in which FA is sampled at several points across the length of the tract, allows researchers to report group differences that may be localized to certain portions of a tract, even if the mean FA does not differ between groups. However, it is unclear if these localized differences are a result of true structural differences in a given tract, or more global structural differences such as inhomogeneities in the location of a crossing fiber [Ramus et al., 2018]. While our results lend themselves to theories of the etiology of dyslexia, we cannot interpret our findings to make causal statements of white matter microstructural properties' connection to reading disabilities. Longitudinal interventional studies should continue to be performed to better understand this relationship. Finally, multivariate pattern analysis could shed insight into the relative contributions of different tract microstructural properties to classifying reading aptitude, similar to what was done in the study by Cui et al. [2016].

In several ways, this study of the relation between white-matter microstructure and single-word reading ability had features that ought to have yielded conclusive findings, but the study raises some open questions. The number of participants was exceptionally large, the TOWRE is a widely used and well-established measure of single-word reading ability, and the analyses took advantage of progress in DTI analytic methods. Children could be categorized by clinical criteria for reading disability or by TOWRE performance, and analyses could be conducted between groups or continuously within groups. There were a number of findings that aligned with some but not other previous findings. Only future research can resolve the apparent contradictions among findings, but one possibility is that there is greater diversity among poor readers than has been heretofore imagined, perhaps combined with developmental variation from infancy to beginning readers to more advanced readers.

CRediT Authorship Contributions

Steven L. Meisler: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Data Curation, Writing - original draft, Writing - review & editing. **John D.E. Gabrieli:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgements

We thank the Healthy Brain Network team for their diligence in collecting the data and generosity in sharing it. We thank all of the participants and their families for volunteering their time to participate in the study. Finally, we thank Amanda O'Brien for proofreading the manuscript.

Funding

This work was funded by the National Institutes of Health (grant numbers 5T32DC000038-28 and 5T32DC000038-29) and the Halis Family Foundation.

Conflicts of Interest

Declarations of interest: none

References

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8:14, 2014.

Lindsay M Alexander, Jasmine Escalera, Lei Ai, Charissa Andreotti, Karina Febre, Alexander Mangone, Natan Vega-Potler, Nicolas Langer, Alexis Alexander, Meagan Kovacs, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific data*, 4(1):1–26, 2017.

Jesper LR Andersson and Stamatios N Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging. *Neuroimage*, 125:1063–1078, 2016.

Jesper LR Andersson, Stefan Skare, and John Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888, 2003.

Jesper LR Andersson, Mark S Graham, Enikő Zsoldos, and Stamatios N Sotiropoulos. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion mr images. *Neuroimage*, 141:556–572, 2016.

C Nikki Arrington, Paulina A Kulesz, Jenifer Juranek, Paul T Cirino, and Jack M Fletcher. White matter microstructure integrity in relation to reading proficiency. *Brain and language*, 174:103–111, 2017.

Brian B. Avants, Charles L. Epstein, Murray Grossman, and James C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008. ISSN 1361-8415. doi: 10.1016/j.media.2007.06.004. URL http://www.sciencedirect.com/science/article/pii/S1361841507000606.

Michael Bach, Frederik B Laun, Alexander Leemans, Chantal MW Tax, Geert J Biessels, Bram Stieltjes, and Klaus H Maier-Hein. Methodological considerations on tract-based spatial statistics (tbss). *Neuroimage*, 100:358–369, 2014.

Chiara Banfi, Karl Koschutnig, Kristina Moll, Gerd Schulte-Körne, Andreas Fink, and Karin Landerl. White matter alterations and tract lateralization in children with dyslexia and isolated spelling deficits. *Human brain mapping*, 40 (3):765–776, 2019.

Naama Barnea-Goraly, Vinod Menon, Mark Eckert, Leanne Tamm, Roland Bammer, Asya Karchemskiy, Christopher C Dant, and Allan L Reiss. White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cerebral cortex*, 15(12):1848–1854, 2005.

Peter J Basser and Carlo Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *Journal of magnetic resonance*, 213(2):560–570, 1996.

Christian Beaulieu. The basis of anisotropic water diffusion in the nervous system–a technical review. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, 15(7-8):435–455, 2002.

Christian Beaulieu. *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press, 2009.

David Bonekamp, Lidia M Nagae, Mahaveer Degaonkar, Melissa Matson, Wael MA Abdalla, Peter B Barker, Susumu Mori, and Alena Horská. Diffusion tensor imaging in children and adolescents: reproducibility, hemispheric, and age-related differences. *Neuroimage*, 34(2):733–742, 2007.

Lauren R Borchers, Lisa Bruckert, Cory K Dodson, Katherine E Travis, Virginia A Marchman, Michal Ben-Shachar, and Heidi M Feldman. Microstructural properties of white matter pathways in relation to subsequent reading abilities in children: a longitudinal analysis. *Brain Structure and Function*, 224(2):891–905, 2019.

Lisa Bruckert, Katherine E Travis, Aviv A Mezer, Michal Ben-Shachar, and Heidi M Feldman. Associations of reading efficiency with white matter properties of the cerebellar peduncles in children. *The Cerebellum*, 19(6):771–777, 2020.

John C Carter, Diane C Lanham, Laurie E Cutting, Amy M Clements-Stephens, Xuejing Chen, Muhamed Hadzipasic, Joon Kim, Martha B Denckla, and Walter E Kaufmann. A dual dti approach to analyzing white matter in children with dyslexia. *Psychiatry Research: Neuroimaging*, 172(3):215–219, 2009.

Marco Catani, Robert J Howard, Sinisa Pajevic, and Derek K Jones. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage*, 17(1):77–94, 2002.

Marco Catani, Derek K Jones, and Dominic H Ffytche. Perisylvian language networks of the human brain. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 57(1):8–16, 2005.

Marco Catani, Matthew PG Allin, Masud Husain, Luca Pugliese, Marsel M Mesulam, Robin M Murray, and Derek K Jones. Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences*, 104(43):17163–17168, 2007.

Isabella Cattinelli, N Alberto Borghese, Marcello Gallucci, and Eraldo Paulesu. Reading the reading brain: a new meta-analysis of functional imaging data on reading. *Journal of neurolinguistics*, 26(1):214–238, 2013.

Bramsh Qamar Chandio, Shannon Leigh Risacher, Franco Pestilli, Daniel Bullock, Fang-Cheng Yeh, Serge Koudoro, Ariel Rokem, Jaroslaw Harezlak, and Eleftherios Garyfallidis. Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations. *Scientific reports*, 10(1):1–18, 2020.

Lin-Ching Chang, Derek K Jones, and Carlo Pierpaoli. Restore: robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 53(5):1088–1095, 2005.

Ming-Chang Chiang, Katie L McMahon, Greig I de Zubicaray, Nicholas G Martin, Ian Hickie, Arthur W Toga, Margaret J Wright, and Paul M Thompson. Genetics of white matter development: a dti study of 705 twins and their siblings aged 12 to 29. *Neuroimage*, 54(3):2308–2317, 2011.

Gary E Christensen, Sarang C Joshi, and Michael I Miller. Volumetric transformation of brain anatomy. *IEEE transactions on medical imaging*, 16(6):864–877, 1997.

Joanna A Christodoulou, Jack Murtagh, Abigail Cyr, Tyler K Perrachione, Patricia Chang, Kelly Halverson, Pamela Hook, Anastasia Yendiki, Satrajit Ghosh, and John DE Gabrieli. Relation of white-matter microstructure to reading ability and disability in beginning readers. *Neuropsychology*, 31(5):508, 2017.

Matthew Cieslak, Philip A Cook, Xiaosong He, Fang-Cheng Yeh, Thijs Dhollander, Azeez Adebimpe, Geoffrey K Aguirre, Danielle S Bassett, Richard F Betzel, Josiane Bourque, et al. Qsiprep: an integrative platform for preprocessing and reconstructing diffusion mri data. *Nature Methods*, pages 1–4, 2021.

Zaixu Cui, Zhichao Xia, Mengmeng Su, Hua Shu, and Gaolang Gong. Disrupted white matter connectivity underlying developmental dyslexia: a machine learning approach. *Human brain mapping*, 37(4):1443–1458, 2016.

Gayle K Deutsch, Robert F Dougherty, Roland Bammer, Wai Ting Siok, John DE Gabrieli, and Brian Wandell. Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex*, 41(3):354–363, 2005.

Thijs Dhollander, David Raffelt, and Alan Connelly. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion mr data without a co-registered t1 image. In *ISMRM Workshop on Breaking the Barriers of Diffusion MRI*, volume 5, page 5, 2016.

Thijs Dhollander, Remika Mito, David Raffelt, and Alan Connelly. Improved white matter response function estimation for 3-tissue constrained spherical deconvolution. In *Proc. Intl. Soc. Mag. Reson. Med*, page 555, 2019.

Mark Eckert. Neuroanatomical markers for dyslexia: a review of dyslexia structural imaging studies. *The neuroscientist*, 10(4):362–371, 2004.

Fifth Edition et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, 21, 2013.

Heidi M Feldman, Eliana S Lee, Jason D Yeatman, and Kristen W Yeom. Language and reading skills in school-aged children and adolescents born preterm are associated with white matter properties on diffusion tensor imaging. *Neuropsychologia*, 50(14):3348–3362, 2012.

Patrick Friedrich, Christoph Fraenz, Caroline Schlüter, Sebastian Ocklenburg, Burkhard Mädler, Onur Güntürkün, and Erhan Genç. The relationship between axon density, myelination, and fractional anisotropy in the human corpus callosum. *Cerebral Cortex*, 30(4):2042–2056, 2020.

Richard E Frye, Khader Hasan, Lian Xue, David Strickland, Benjamin Malmberg, Jacqueline Liederman, and Andrew Papanicolaou. Splenium microstructure is related to two dimensions of reading skill. *Neuroreport*, 19(16):1627, 2008.

Richard E Frye, Jacqueline Liederman, Khader M Hasan, Alexis Lincoln, Benjamin Malmberg, John McLean III, and Andrew Papanicolaou. Diffusion tensor quantification of the relations between microstructural and macrostructural indices of white matter and reading. *Human brain mapping*, 32(8):1220–1235, 2011.

Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

Sila Genc, Chantal MW Tax, Erika P Raven, Maxime Chamberland, Greg D Parker, and Derek K Jones. Impact of b-value on estimates of apparent fibre density. *Human brain mapping*, 41(10):2583–2595, 2020.

Krzysztof Gorgolewski, Christopher D. Burns, Cindee Madison, Dav Clark, Yarosalv O. Halchenko, Michael L. Waskom, and Satrajit Ghosh. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 5:13, 2011. doi: 10.3389/fninf.2011.00013.

Krzysztof J. Gorgolewski, Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, Hans Johnson, Christopher Burns, Alexandre Manhães-Savio, Carlo Hamalainen, Benjamin Yvernault, Taylor Salo, Kesshi Jordan, Mathias Goncalves, Michael Waskom, Daniel Clark, Jason Wong, Fred Loney, Marc Modat, Blake E Dewey, Cindee Madison, Matteo Visconti di Oleggio Castello, Michael G. Clark, Michael Dayan, Dav Clark, Anisha Keshavan, Basile Pinsard, Alexandre Gramfort, Shoshana Berleant, Dylan M. Nielson, Salma Bougacha, Gael Varoquaux, Ben Cipollini, Ross Markello, Ariel Rokem, Brendan Moloney, Yaroslav O.

Halchenko, Demian Wassermann, Michael Hanke, Christian Horea, Jakub Kaczmarzyk, Gilles de Hollander, Elizabeth DuPre, Ashley Gillman, David Mordom, Colin Buchanan, Rosalia Tungaraza, Wolfgang M. Pauli, Shariq Iqbal, Sharad Sikka, Matteo Mancini, Yannick Schwartz, Ian B. Malone, Mathieu Dubois, Caroline Frohlich, David Welch, Jessica Forbes, James Kent, Aimi Watanabe, Chad Cumba, Julia M. Huntenburg, Erik Kastman, B. Nolan Nichols, Arman Eshaghi, Daniel Ginsburg, Alexander Schaefer, Benjamin Acland, Steven Giavasis, Jens Kleesiek, Drew Erickson, René Küttner, Christian Haselgrove, Carlos Correa, Ali Ghayoor, Franz Liem, Jarrod Millman, Daniel Haehn, Jeff Lai, Dale Zhou, Ross Blair, Tristan Glatard, Mandy Renfro, Siqi Liu, Ari E. Kahn, Fernando Pérez-García, William Triplett, Leonie Lampe, Jörg Stadler, Xiang-Zhen Kong, Michael Hallquist, Andrey Chetverikov, John Salvatore, Anne Park, Russell Poldrack, R. Cameron Craddock, Souheil Inati, Oliver Hinds, Gavin Cooper, L. Nathan Perkins, Ana Marina, Aaron Mattfeld, Maxime Noel, Lukas Snoek, K Matsubara, Brian Cheung, Simon Rothmei, Sebastian Urchs, Joke Durnez, Fred Mertz, Daniel Geisler, Andrew Floren, Stephan Gerhard, Paul Sharp, Miguel Molina-Romero, Alejandro Weinstein, William Broderick, Victor Saase, Sami Kristian Andberg, Robbert Harms, Kai Schlamp, Jaime Arias, Dimitri Papadopoulos Orfanos, Claire Tarbert, Arielle Tambini, Alejandro De La Vega, Thomas Nickson, Matthew Brett, Marcel Falkiewicz, Kornelius Podranski, Janosch Linkersdörfer, Guillaume Flandin, Eduard Ort, Dmitry Shachnev, Daniel McNamee, Andrew Davison, Jan Varada, Isaac Schwabacher, John Pellman, Martin Perez-Guevara, Ranjeet Khanuja, Nicolas Pannetier, Conor McDermottroe, and Satrajit Ghosh. Nipype. *Software*, 2018. doi: 10.5281/zenodo.596855.

Patric Hagmann, Lisa Jonasson, Philippe Maeder, Jean-Philippe Thiran, Van J Wedeen, and Reto Meuli. Understanding diffusion mr imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. *Radiographics*, 26(suppl_1):S205–S223, 2006.

Jürgen Hänggi, Susan Koeneke, Ladina Bezzola, and Lutz Jäncke. Structural neuroplasticity in the sensorimotor network of professional female ballet dancers. *Human brain mapping*, 31(8):1196–1206, 2010.

Steve M Heath, Dorothy VM Bishop, John H Hogben, and Neil W Roach. Psychophysical indices of perceptual functioning in dyslexia: A psychometric analysis. *Cognitive neuropsychology*, 23(6):905–929, 2006.

Guillaume Herbet, Ilyess Zemmoura, and Hugues Duffau. Functional anatomy of the inferior longitudinal fasciculus: from historical reports to current hypotheses. *Frontiers in neuroanatomy*, 12:77, 2018.

Fumiko Hoeft, Bruce D McCandliss, Jessica M Black, Alexander Gantman, Nahal Zakerani, Charles Hulme, Heikki Lyytinen, Susan Whitfield-Gabrieli, Gary H Glover, Allan L Reiss, et al. Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences*, 108(1):361–366, 2011.

Tzipi Horowitz-Kraus, Jennifer J Vannest, Darren Kadis, Nicole Cicchino, Yingying Y Wang, and Scott K Holland. Reading acceleration training changes brain circuitry in children with reading difficulties. *Brain and Behavior*, 4(6): 886–902, 2014.

Tzipi Horowitz-Kraus, Molly Grainger, Mark DiFrancesco, Jennifer Vannest, and Scott K Holland. Right is not always wrong: Dti and fmri evidence for the reliance of reading comprehension on language-comprehension networks in the right hemisphere. *Brain imaging and behavior*, 9(1):19–31, 2015.

Elizabeth Huber, Patrick M Donnelly, Ariel Rokem, and Jason D Yeatman. Rapid and widespread white matter plasticity during an intensive reading intervention. *Nature communications*, 9(1):1–13, 2018.

Katarzyna Jednorog, Artur Marchewka, Irene Altarelli, Ana Karla Monzalvo Lopez, Muna van Ermingen-Marbach, Marion Grande, Anna Grabowska, Stefan Heim, and Franck Ramus. How reliable are gray matter disruptions in specific reading disability across multiple countries and languages? insights from a large-scale voxel-based morphometry study. *Human brain mapping*, 36(5):1741–1754, 2015.

Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, 2001. ISSN 1361-8415. doi: 10.1016/S1361-8415(01)00036-6. URL http://www.sciencedirect.com/science/article/pii/S1361841501000366.

Ben Jeurissen, Jacques-Donald Tournier, Thijs Dhollander, Alan Connelly, and Jan Sijbers. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *NeuroImage*, 103:411–426, 2014.

Richard Johnston, Nicola J Pitchford, Neil W Roach, and Timothy Ledgeway. Why is the processing of global motion impaired in adults with developmental dyslexia? *Brain and Cognition*, 108:20–31, 2016.

Koji Kamagata, Yumiko Motoi, Hiroyuki Tomiyama, Osamu Abe, Kenji Ito, Keigo Shimoji, Michimasa Suzuki, Masaaki Hori, Atsushi Nakanishi, Tamotsu Sano, et al. Relationship between cognitive impairment and white-matter alteration in parkinson's disease with dementia: tract-based spatial statistics and tract-specific analysis. *European radiology*, 23(7):1946–1955, 2013.

Elias Kellner, Bibek Dhital, Valerij G Kiselev, and Marco Reisert. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic resonance in medicine*, 76(5):1574–1581, 2016.

Torkel Klingberg, Maj Hedehus, Elise Temple, Talya Salz, John DE Gabrieli, Michael E Moseley, and Russell A Poldrack. Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron*, 25(2):493–500, 2000.

Inga K Koerte, Anna Willems, Marc Muehlmann, Kristina Moll, Sonia Cornell, Silvia Pixner, Denise Steffinger, Daniel Keeser, Florian Heinen, Marek Kubicki, et al. Mathematical abilities in dyslexic children: a diffusion tensor imaging study. *Brain imaging and behavior*, 10(3):781–791, 2016.

Joseph Kuchling, Yael Backner, Frederike C Oertel, Noa Raz, Judith Bellmann-Strobl, Klemens Ruprecht, Friedemann Paul, Netta Levin, Alexander U Brandt, and Michael Scheel. Comparison of probabilistic tractography and tract-based spatial statistics for assessing optic radiation damage in patients with autoimmune inflammatory disorders of the central nervous system. *NeuroImage: Clinical*, 19:538–550, 2018.

Gregory M Kurtzer, Vanessa Sochat, and Michael W Bauer. Singularity: Scientific containers for mobility of compute. *PloS one*, 12(5):e0177459, 2017.

Nicolas Langer, Barbara Peysakhovich, Jennifer Zuk, Marie Drottar, Danielle D Sliva, Sara Smith, Bryce LC Becker, P Ellen Grant, and Nadine Gaab. White matter alterations in infants at risk for developmental dyslexia. *Cerebral Cortex*, 27(2):1027–1036, 2017.

Catherine Lebel and Sean Deoni. The development of brain white matter microstructure. *Neuroimage*, 182:207–218, 2018.

Catherine Lebel, Lindsay Walker, Alexander Leemans, Linda Phillips, and Christian Beaulieu. Microstructural maturation of the human brain from childhood to adulthood. *Neuroimage*, 40(3):1044–1055, 2008.

Catherine Lebel, Bennett Shaywitz, John Holahan, Sally Shaywitz, Karen Marchione, and Christian Beaulieu. Diffusion tensor imaging correlates of reading ability in dysfluent and non-impaired readers. *Brain and language*, 125(2): 215–222, 2013.

Catherine Lebel, Sarah Treit, and Christian Beaulieu. A review of diffusion mri of typical white matter development from early childhood to young adulthood. *NMR in Biomedicine*, 32(4):e3778, 2019.

Janosch Linkersdörfer, Jan Lonnemann, Sven Lindberg, Marcus Hasselhorn, and Christian J. Fiebach. Grey matter alterations co-localize with functional abnormalities in developmental dyslexia: An ale meta-analysis. *PLOS ONE*, 7 (8):1–10, 08 2012. doi: 10.1371/journal.pone.0043122. URL https://doi.org/10.1371/journal.pone.0043122.

Antoine Lutti, Frederic Dick, Martin I Sereno, and Nikolaus Weiskopf. Using high-resolution quantitative mapping of r1 as an index of cortical myelination. *Neuroimage*, 93:176–188, 2014.

G Reid Lyon, Sally E Shaywitz, and Bennett A Shaywitz. A definition of dyslexia. *Annals of dyslexia*, 53(1):1–14, 2003.

José M Maisog, Erin R Einbinder, D Lynn Flowers, Peter E Turkeltaub, and Guinevere F Eden. A meta-analysis of functional neuroimaging studies of dyslexia. *Annals of the New York Academy of Sciences*, 1145(1):237–259, 2008.

Cecilia Marino, Paola Scifo, Pasquale A Della Rosa, Sara Mascheretti, Andrea Facoetti, Maria L Lorusso, Roberto Giorda, Monica Consonni, Andrea Falini, Massimo Molteni, et al. The dcdc2/intron 2 deletion and white matter disorganization: focus on developmental dyslexia. *Cortex*, 57:227–243, 2014.

Juan Martino, Christian Brogna, Santiago G Robles, Francesco Vergani, and Hugues Duffau. Anatomic dissection of the inferior fronto-occipital fasciculus revisited in the lights of brain stimulation data. *cortex*, 46(5):691–699, 2010.

Ivan I Maximov, Heike Thönneßen, Kerstin Konrad, Laura Amort, Irene Neuner, and N Jon Shah. Statistical instability of tbss analysis based on dti fitting algorithm. *Journal of Neuroimaging*, 25(6):883–891, 2015.

Dirk Merkel. Docker: lightweight linux containers for consistent development and deployment. *Linux journal*, 2014 (239):2, 2014.

R Duncan Milne, Ari Syngeniotis, Graeme Jackson, and Michael C Corballis. Mixed lateralization of phonological assembly in developmental dyslexia. *Neurocase*, 8(3):209–209, 2002.

David Moreau, Josephine E Stonyer, Nicole S McKay, and Karen E Waldie. No evidence for systematic white matter correlates of dyslexia: an activation likelihood estimation meta-analysis. *Brain research*, 1683:36–47, 2018.

Ryan L Muetzel, Paul F Collins, Bryon A Mueller, Ann M Schissel, Kelvin O Lim, and Monica Luciana. The development of corpus callosum microstructure and associations with bimanual task performance in healthy adolescents. *Neuroimage*, 39(4):1918–1925, 2008.

Pratik Mukherjee, Jeffrey H Miller, Joshua S Shimony, Joseph V Philip, Deepika Nehra, Abraham Z Snyder, Thomas E Conturo, Jeffrey J Neil, and Robert C McKinstry. Diffusion-tensor mr imaging of gray and white matter development during normal human brain maturation. *American Journal of Neuroradiology*, 23(9):1445–1456, 2002.

Kimberly A Murphy, Jigar R Jogia, and Joel B Talcott. On the neural basis of word reading: A meta-analysis of fmri evidence using activation likelihood estimation. *Journal of Neurolinguistics*, 49:71–83, 2019.

Thomas E Nichols and Andrew P Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

Roderick I Nicolson and Angela J Fawcett. Procedural learning difficulties: reuniting the developmental disorders? *TRENDS in Neurosciences*, 30(4):135–141, 2007.

Roderick I Nicolson, Angela J Fawcett, and Paul Dean. Developmental dyslexia: the cerebellar deficit hypothesis. *Trends in neurosciences*, 24(9):508–511, 2001.

Timothy N Odegard, Emily A Farris, Jeremiah Ring, Roderick McColl, and Jeffrey Black. Brain connectivity in non-reading impaired children and children diagnosed with developmental dyslexia. *Neuropsychologia*, 47(8-9): 1972–1977, 2009.

Richard C Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1): 97–113, 1971.

Hiroyuki Oouchi, Kei Yamada, Koji Sakai, Osamu Kizu, Takao Kubota, Hirotoshi Ito, and Tsunehiko Nishimura. Diffusion anisotropy measurement of brain white matter is affected by voxel size: underestimation occurs in areas with crossing fibers. *American Journal of Neuroradiology*, 28(6):1102–1106, 2007.

Jonathan D. Power, Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage*, 84 (Supplement C):320–341, 2014. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2013.08.048. URL http://www.sciencedirect.com/science/article/pii/S1053811913009117.

Athanassios Protopapas and Rauno Parrila. Is dyslexia a brain disorder? *Brain sciences*, 8(4):61, 2018.

Athanassios Protopapas and Rauno Parrila. Dyslexia: Still not a neurodevelopmental disorder. *Brain sciences*, 9(1):9, 2019.

Kenneth R Pugh, W Einar Mencl, Annette R Jenner, Leonard Katz, Stephen J Frost, Jun Ren Lee, Sally E Shaywitz, and Bennett A Shaywitz. Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Mental retardation and developmental disabilities research reviews*, 6(3):207–213, 2000.

Kenneth R Pugh, Stephen J Frost, Douglas L Rothman, Fumiko Hoeft, Stephanie N Del Tufo, Graeme F Mason, Peter J Molfese, W Einar Mencl, Elena L Grigorenko, Nicole Landi, et al. Glutamate and choline levels predict individual differences in reading ability in emergent readers. *Journal of Neuroscience*, 34(11):4082–4089, 2014.

Mary Colvin Putnam, Megan S Steven, Karl W Doron, Adam C Riggall, and Michael S Gazzaniga. Cortical projection topography of the human splenium: hemispheric asymmetry and individual differences. *Journal of cognitive neuroscience*, 22(8):1662–1669, 2010.

David Raffelt, Thijs Dhollander, J-Donald Tournier, Rami Tabbara, Robert E Smith, Eric Pierre, and Alan Connelly. Bias field correction and intensity normalisation for quantitative analysis of apparent fibre density. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 25, page 3541, 2017.

Franck Ramus, Irene Altarelli, Katarzyna Jednoróg, Jingjing Zhao, and Lou Scotto Di Covella. Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience & Biobehavioral Reviews*, 84:434–452, 2018.

Andreas M Rauschecker, Gayle K Deutsch, Michal Ben-Shachar, Armin Schwartzman, Lee M Perry, and Robert F Dougherty. Reading impairment in a patient with missing arcuate fasciculus. *Neuropsychologia*, 47(1):180–194, 2009.

Todd L Richards, Jim Stevenson, J Crouch, L Clark Johnson, Kenneth R Maravilla, Pat Stock, Robert D Abbott, and Virginia W Berninger. Tract-based spatial statistics of diffusion tensor imaging in adults with dyslexia. *American Journal of Neuroradiology*, 29(6):1134–1139, 2008.

Fabio Richlan, Martin Kronbichler, and Heinz Wimmer. Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Human brain mapping*, 34(11):3055–3065, 2013.

Rawiwan Roongpraiwan, Nichara Ruangdaraganon, Pongsak Visudhiphan, and Kanitta Santikul. Prevalence and clinical characteristics of dyslexia in primary school students. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet*, 85:S1097–103, 2002.

Zeynep M Saygin, Elizabeth S Norton, David E Osher, Sara D Beach, Abigail B Cyr, Ola Ozernov-Palchik, Anastasia Yendiki, Bruce Fischl, Nadine Gaab, and John DE Gabrieli. Tracking the roots of reading ability: white matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. *Journal of Neuroscience*, 33(33):13251–13258, 2013.

Kurt G Schilling, François Rheault, Laurent Petit, Colin B Hansen, Vishwesh Nath, Fang-Cheng Yeh, Gabriel Girard, Muhamed Barakovic, Jonathan Rafael-Patino, Thomas Yu, et al. Tractography dissection variability: what happens when 42 groups dissect 14 white matter bundles on the same dataset? *bioRxiv*, pages 2020–10, 2021.

Bradley L Schlaggar and Bruce D McCandliss. Development of neural systems for reading. *Annu. Rev. Neurosci.*, 30: 475–503, 2007.

Vincent J Schmithorst and Marko Wilke. Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience letters*, 321(1-2):57–60, 2002.

Vincent J Schmithorst, Marko Wilke, Bernard J Dardzinski, and Scott K Holland. Correlation of white matter diffusivity and anisotropy with age during childhood and adolescence: a cross-sectional diffusion-tensor mr imaging study. *Radiology*, 222(1):212–218, 2002.

Jan Scholz, Miriam C Klein, Timothy EJ Behrens, and Heidi Johansen-Berg. Training induces changes in white-matter architecture. *Nature neuroscience*, 12(11):1370–1371, 2009.

Roey Schurr, Yiran Duan, Anthony M Norcia, Shumpei Ogawa, Jason D Yeatman, and Aviv A Mezer. Tractography optimization using quantitative t1 mapping in the human optic radiation. *Neuroimage*, 181:645–658, 2018.

Noam Shemesh. Axon diameters and myelin content modulate microscopic fractional anisotropy at short diffusion times in fixed rat spinal cord. *Frontiers in Physics*, 6:49, 2018.

Daniel J Simmonds, Michael N Hallquist, Miya Asato, and Beatriz Luna. Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (dti) study. *Neuroimage*, 92:356–368, 2014.

Stephen M Smith, Mark Jenkinson, Heidi Johansen-Berg, Daniel Rueckert, Thomas E Nichols, Clare E Mackay, Kate E Watkins, Olga Ciccarelli, M Zaheer Cader, Paul M Matthews, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006.

Claudia Steinbrink, Katrin Vogt, A Kastrup, H-P Müller, FD Juengling, J Kassubek, and A Riecker. The contribution of white and gray matter differences to developmental dyslexia: insights from dti and vbm at 3.0 t. *Neuropsychologia*, 46(13):3170–3178, 2008.

Catherine J Stoodley and John F Stein. The cerebellum and dyslexia. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 2011.

Mengmeng Su, Jingjing Zhao, Michel Thiebaut de Schotten, Wei Zhou, Gaolang Gong, Franck Ramus, and Hua Shu. Alterations in white matter pathways underlying phonological and morphological processing in chinese developmental dyslexia. *Developmental cognitive neuroscience*, 31:11–19, 2018.

Yasuyuki Taki, Benjamin Thyreau, Hiroshi Hashizume, Yuko Sassa, Hikaru Takeuchi, Kai Wu, Yuka Kotozaki, Rui Nouchi, Michiko Asano, Kohei Asano, et al. Linear and curvilinear correlations of brain white matter volume, fractional anisotropy, and mean diffusivity with age using voxel-based and region-of-interest analyses in 246 healthy children. *Human brain mapping*, 34(8):1842–1856, 2013.

Christian K Tamnes, Ylva Østby, Anders M Fjell, Lars T Westlye, Paulina Due-Tønnessen, and Kristine B Walhovd. Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure. *Cerebral cortex*, 20(3):534–548, 2010.

Elise Temple, Gayle K Deutsch, Russell A Poldrack, Steven L Miller, Paula Tallal, Michael M Merzenich, and John DE Gabrieli. Neural deficits in children with dyslexia ameliorated by behavioral remediation: evidence from functional mri. *Proceedings of the National Academy of Sciences*, 100(5):2860–2865, 2003.

Joseph K Torgesen, Carol Alexander Rashotte, and Richard K Wagner. *TOWRE: Test of word reading efficiency*. Psychological Corporation Toronto, Ontario, 1999.

J-Donald Tournier, Fernando Calamante, David G Gadian, and Alan Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *Neuroimage*, 23(3):1176–1185, 2004.

J-Donald Tournier, Chun-Hung Yeh, Fernando Calamante, Kuan-Hung Cho, Alan Connelly, and Ching-Po Lin. Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage*, 42(2):617–625, 2008.

Katherine E Travis, Yael Leitner, Heidi M Feldman, and Michal Ben-Shachar. Cerebellar white matter pathways are associated with reading skills in children and adolescents. *Human brain mapping*, 36(4):1536–1553, 2015.

JM Tsang, RF Dougherty, and BA Wandell. Tract alignment errors decrease detection power in group analyses of diffusion data with tbss. *Society for Neuroscience. San Diego, CA*, 2010.

Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010. ISSN 0278-0062. doi: 10.1109/TMI.2010.2046908.

Maaike Vandermosten, Bart Boets, Hanne Poelmans, Stefan Sunaert, Jan Wouters, and Pol Ghesquiere. A tractography study in dyslexia: neuroanatomic correlates of orthographic, phonological and speech processing. *Brain*, 135(3): 935–948, 2012a.

Maaike Vandermosten, Bart Boets, Jan Wouters, and Pol Ghesquière. A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neuroscience & Biobehavioral Reviews*, 36(6):1532–1552, 2012b.

Jelle Veraart, Dmitry S Novikov, Daan Christiaens, Benjamin Ades-Aron, Jan Sijbers, and Els Fieremans. Denoising of diffusion mri using random matrix theory. *NeuroImage*, 142:394–406, 2016.

Karen E Waldie, Charlotte E Haigh, Gjurgjica Badzakova-Trajkov, Jude Buckley, and Ian J Kirk. Reading the wrong way with the right hemisphere. *Brain Sciences*, 3(3):1060–1075, 2013.

Brian A Wandell and Jason D Yeatman. Biological development of reading circuits. *Current Opinion in Neurobiology*, 23(2):261–268, 2013.

Xuhui Wang, Sudhir Pathak, Lucia Stefaneanu, Fang-Cheng Yeh, Shiting Li, and Juan C Fernandez-Miranda. Subcomponents and connectivity of the superior longitudinal fasciculus in the human brain. *Brain Structure and Function*, 221(4):2075–2092, 2016.

Demian Wassermann, Yogesh Rathi, Sylvain Bouix, Marek Kubicki, Ron Kikinis, Martha Shenton, and Carl-Fredrik Westin. White matter bundle registration and population analysis based on gaussian processes. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 320–332. Springer, 2011.

Jakob Wasserthal, Peter Neher, and Klaus H Maier-Hein. Tractseg-fast and accurate white matter tract segmentation. *NeuroImage*, 183:239–253, 2018a.

Jakob Wasserthal, Peter F Neher, and Klaus H Maier-Hein. Tract orientation mapping for bundle-specific tractography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 36–44. Springer, 2018b.

Jakob Wasserthal, Peter F Neher, Dusan Hirjak, and Klaus H Maier-Hein. Combined tract segmentation and orientation mapping for bundle-specific tractography. *Medical image analysis*, 58:101559, 2019.

Jakob Wasserthal, Klaus H Maier-Hein, Peter F Neher, Georg Northoff, Katharina M Kubera, Stefan Fritze, Anais Harneit, Lena S Geiger, Heike Tost, Robert C Wolf, et al. Multiparametric mapping of white matter microstructure in catatonia. *Neuropsychopharmacology*, 45(10):1750–1757, 2020.

Suzanne E Welcome and Marc F Joanisse. Individual differences in white matter anatomy predict dissociable components of reading skill in adults. *Neuroimage*, 96:261–275, 2014.

Jason D Yeatman, Robert F Dougherty, Elena Rykhlevskaia, Anthony J Sherbondy, Gayle K Deutsch, Brian A Wandell, and Michal Ben-Shachar. Anatomical properties of the arcuate fasciculus predict phonological and reading skills in children. *Journal of cognitive neuroscience*, 23(11):3304–3317, 2011.

Jason D Yeatman, Robert F Dougherty, Michal Ben-Shachar, and Brian A Wandell. Development of white matter and reading skills. *Proceedings of the National Academy of Sciences*, 109(44):E3045–E3053, 2012a.

Jason D Yeatman, Robert F Dougherty, Nathaniel J Myall, Brian A Wandell, and Heidi M Feldman. Tract profiles of white matter properties: automating fiber-tract quantification. *PloS one*, 7(11):e49790, 2012b.

Jason D Yeatman, Andreas M Rauschecker, and Brian A Wandell. Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. *Brain and language*, 125(2):146–155, 2013.

Fang-Cheng Yeh, Islam M Zaydan, Valerie R Suski, David Lacomis, R Mark Richardson, Joseph C Maroon, and Jessica Barrios-Martinez. Differential tractography as a track-based biomarker for neuronal injury. *Neuroimage*, 202:116131, 2019.

Andy B Yoo, Morris A Jette, and Mark Grondona. Slurm: Simple linux utility for resource management. In *Workshop on Job Scheduling Strategies for Parallel Processing*, pages 44–60. Springer, 2003.

Gojko Žarić, Inge Timmers, Patty Gerretsen, Gorka Fraga González, Jurgen Tijms, Maurits W van der Molen, Leo Blomert, and Milene Bonte. Atypical white matter connectivity in dyslexic readers of a fairly transparent orthography. *Frontiers in psychology*, 9:1147, 2018.

Mingxia Zhang, Chuansheng Chen, Gui Xue, Zhong-lin Lu, Leilei Mei, Hongli Xue, Miao Wei, Qinghua He, Jin Li, and Qi Dong. Language-general and-specific white matter microstructural bases for reading. *NeuroImage*, 98: 435–441, 2014.

Yongyue. Zhang, Michael. Brady, and Stephen Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1): 45–57, 2001. ISSN 0278-0062. doi: 10.1109/42.906424.

Jingjing Zhao, Michel Thiebaut de Schotten, Irene Altarelli, Jessica Dubois, and Franck Ramus. Altered hemispheric lateralization of white matter pathways in developmental dyslexia: Evidence from spherical deconvolution tractography. *Cortex*, 76:51–62, 2016.

Supplementary Materials

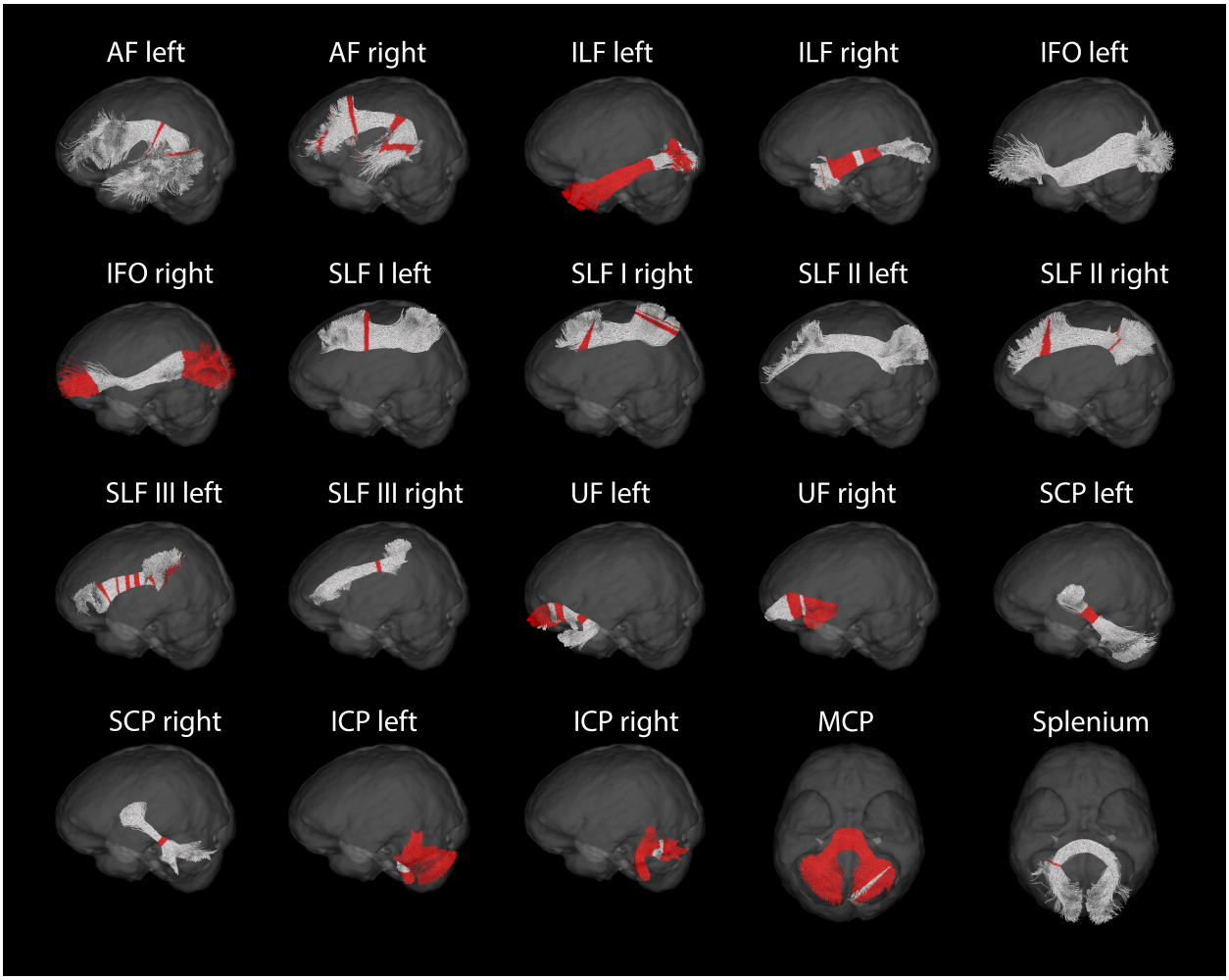

Figure S1: Anatomical depictions of white matter regions whose FA correlated with TOWRE composite scores in the Low-TOWRE group. Red coloring denotes regions of significant positive correlation. The left IFO and SLF III had no significantly correlated regions.

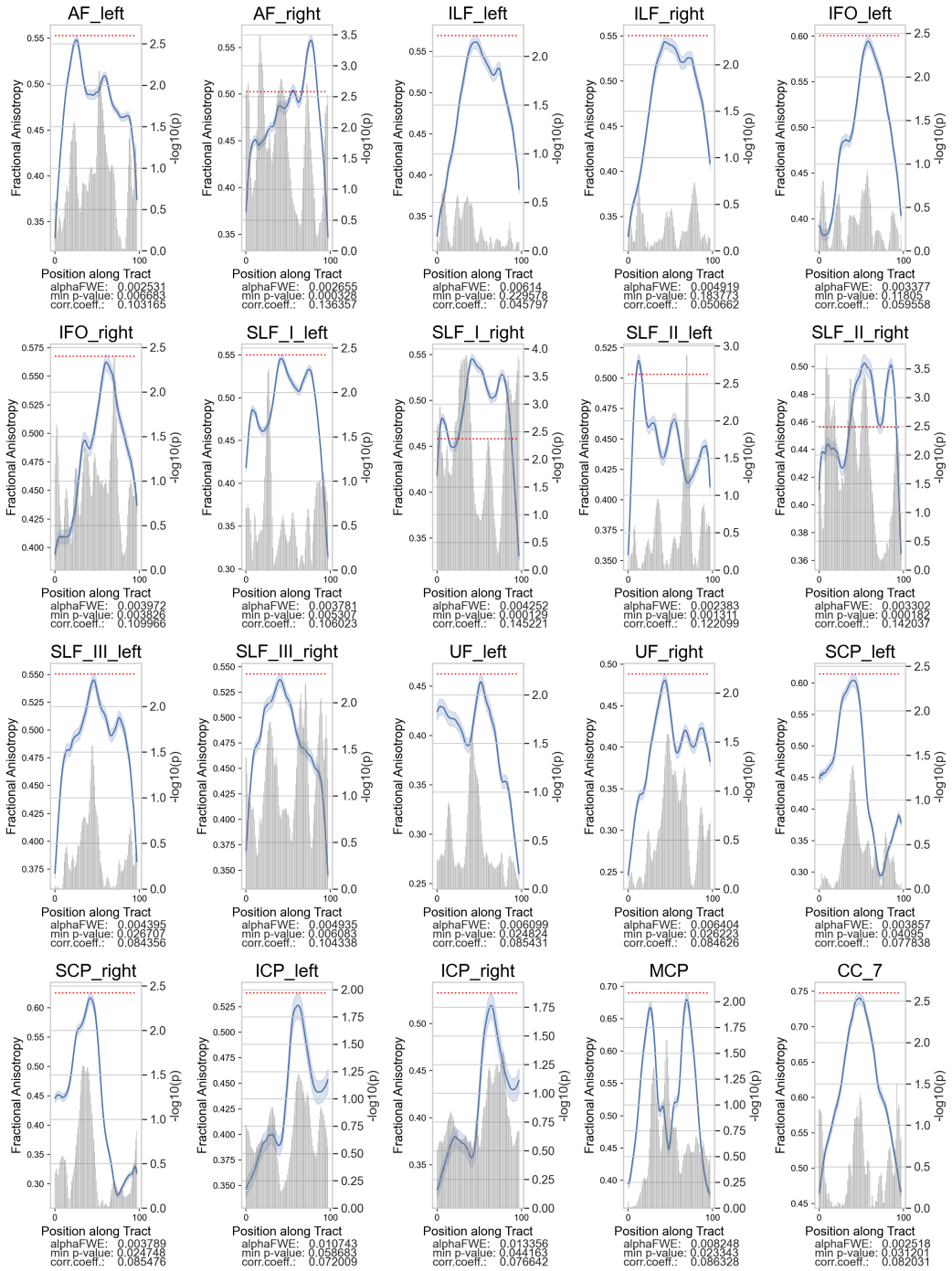

Figure S2: Few tracts exhibited significant FA correlations with TOWRE composite scores across the entire participant cohort. Each plot represents a different fiber bundle. The blue curve represents the average FA after regressing out sex and handedness. The shaded region surrounding the curve is the 95% confidence interval. The red dotted line depicts the critical FWE-adjusted α (given an uncorrected α of 0.05). The gray bars represent the p-value of the correlation at each node, measured on a -log scale. A gray bar extending above the red line ($p < \alpha$FWE) denotes a significant node. Below each plot, the critical FWE-adjusted α is listed, along with the strongest correlation achieved in the tract and its associated p-value.

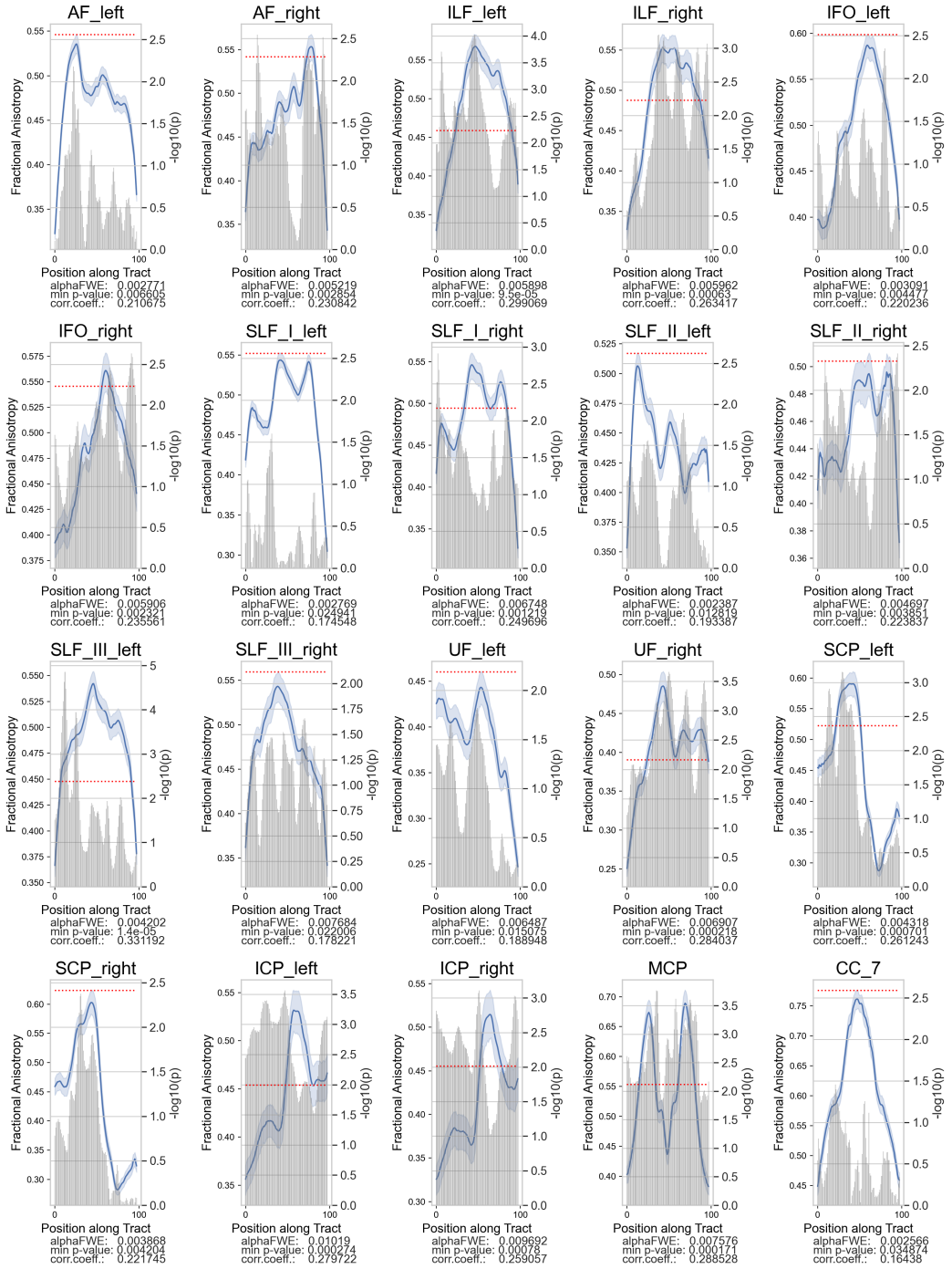

Figure S3: Several fiber bundles exhibited significant FA correlations with TOWRE SWE scores among the Low-TOWRE group. Plotting convention is the same as in the previous figure.

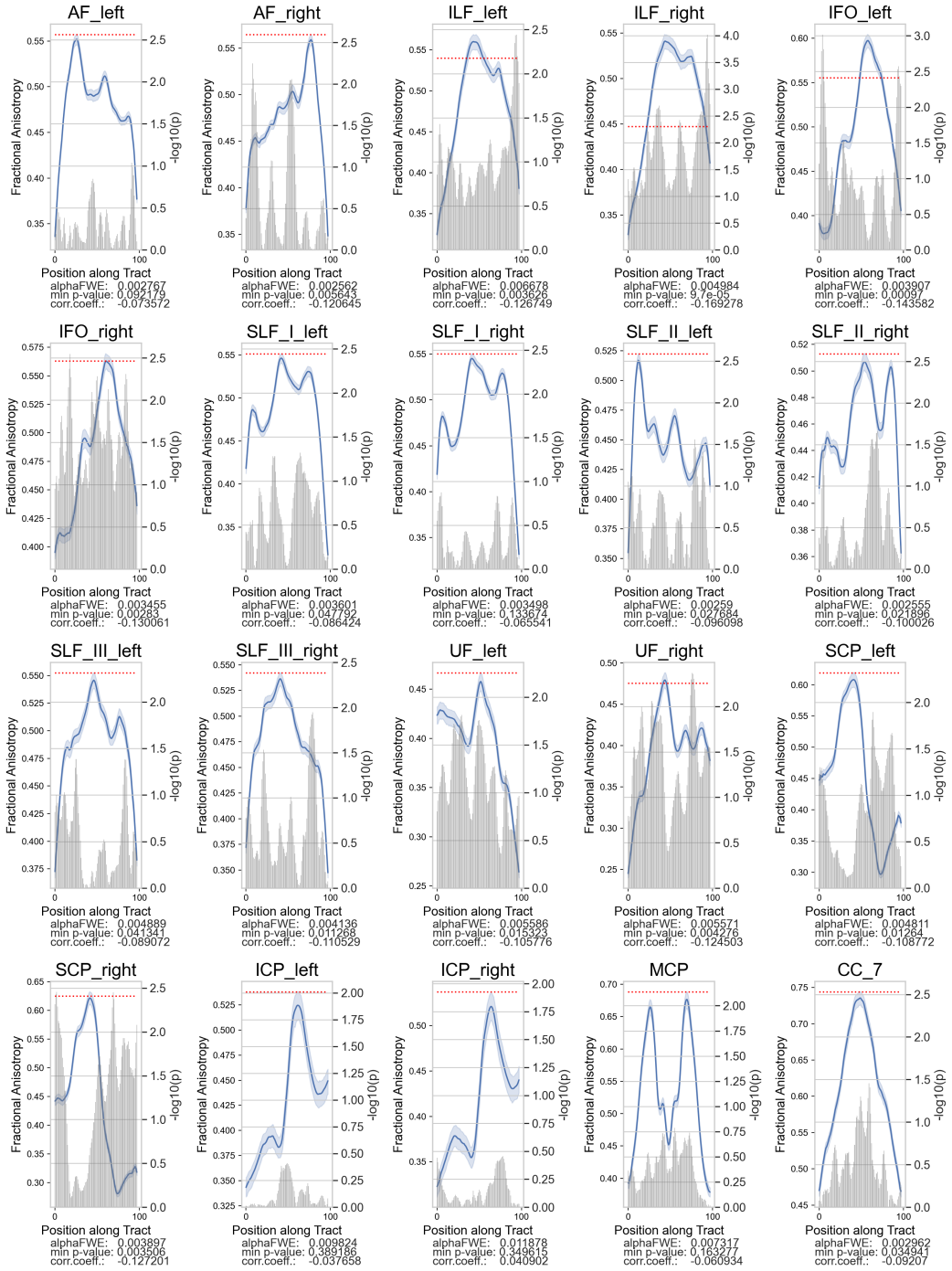

Figure S4: Multiple tracts exhibited significant negative FA correlations with TOWRE SWE scores among the High-TOWRE group. Plotting convention is the same as in the previous figure.

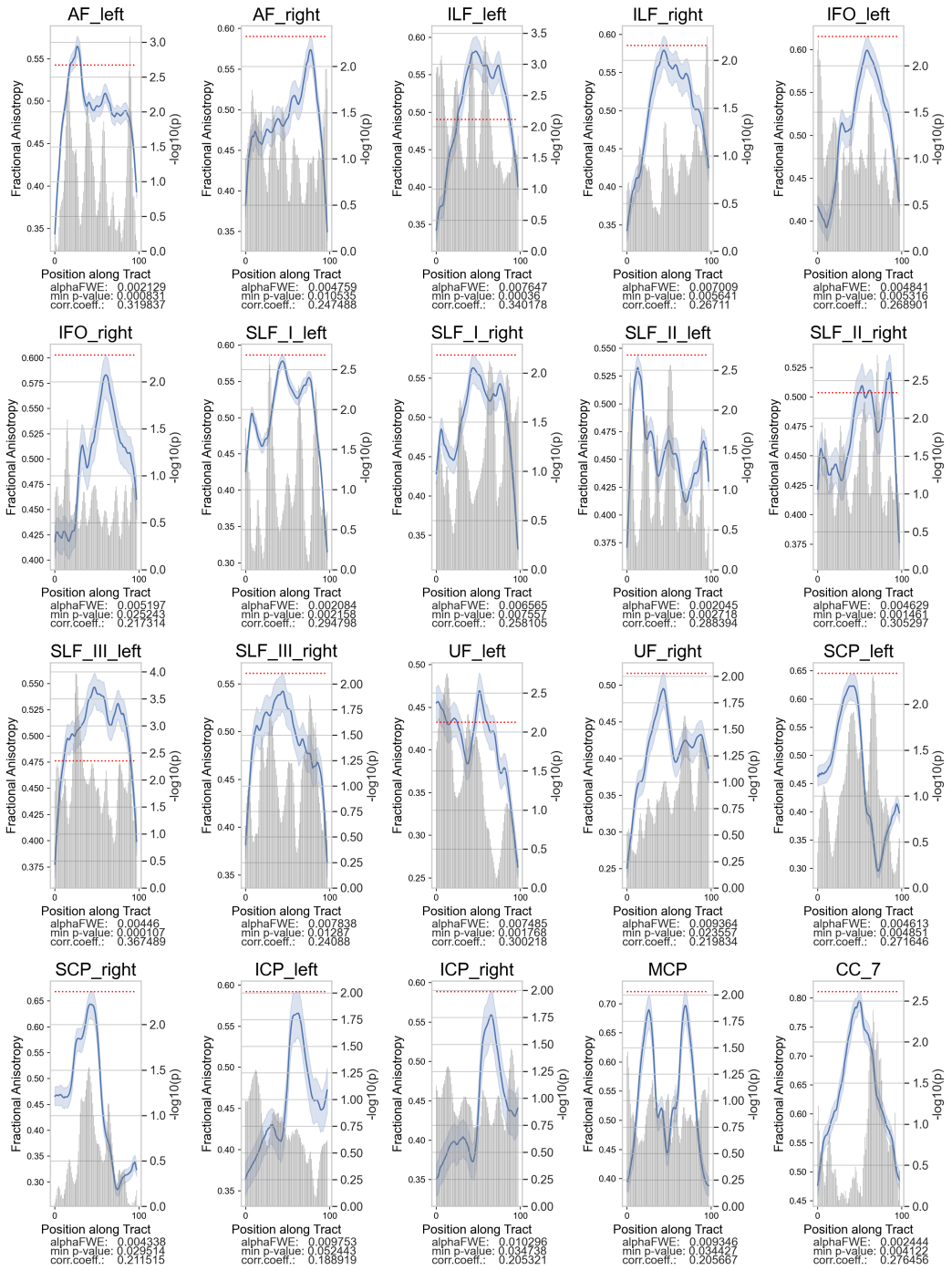

Figure S5: Several fiber bundles exhibited significant FA correlations with TOWRE SWE scores among the RD group. Plotting convention is the same as in the previous figure.

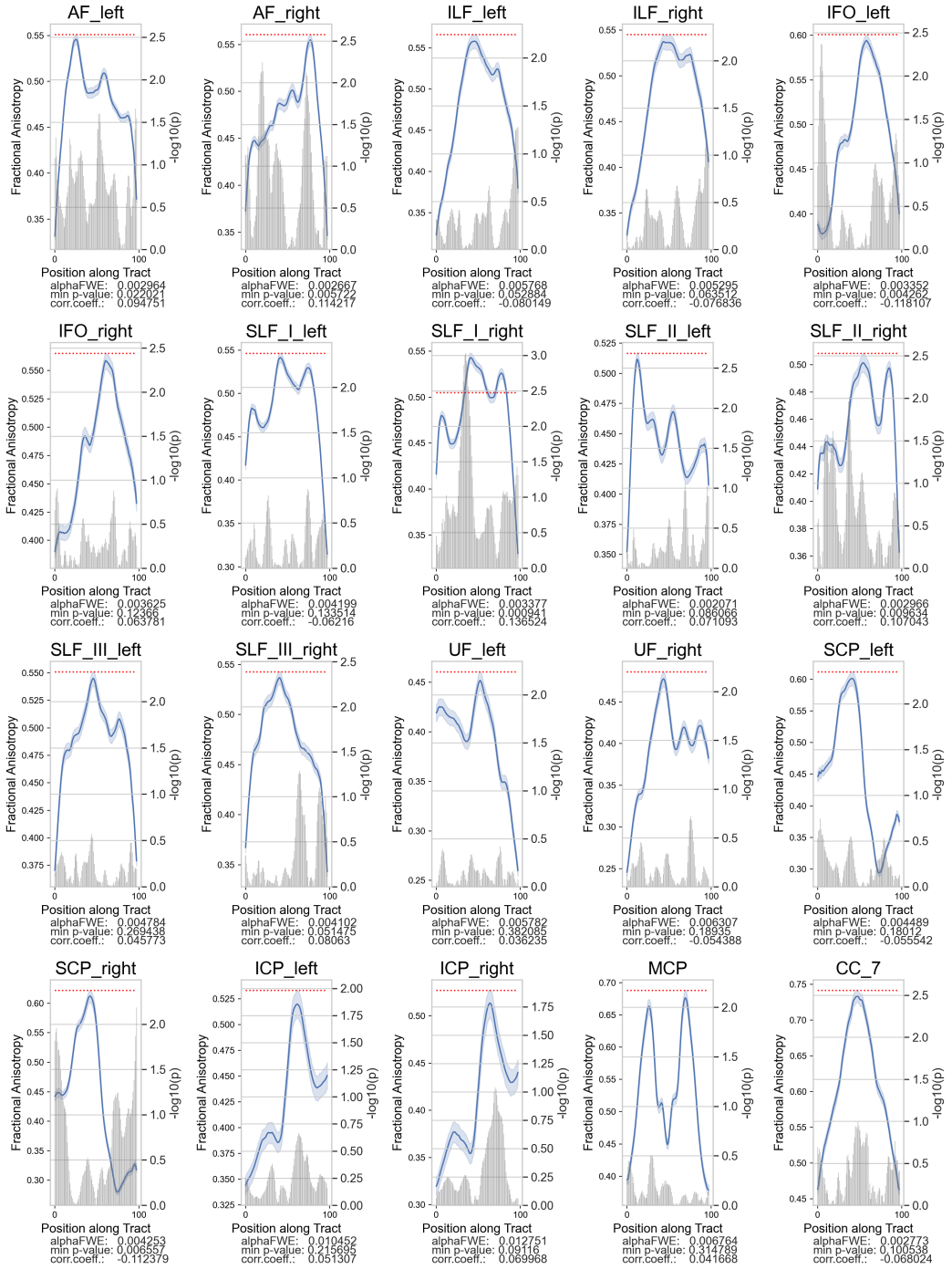

Figure S6: Only the FA of right SLF I was positively associated with TOWRE SWE scores among the TR group. Plotting convention is the same as in the previous figure.

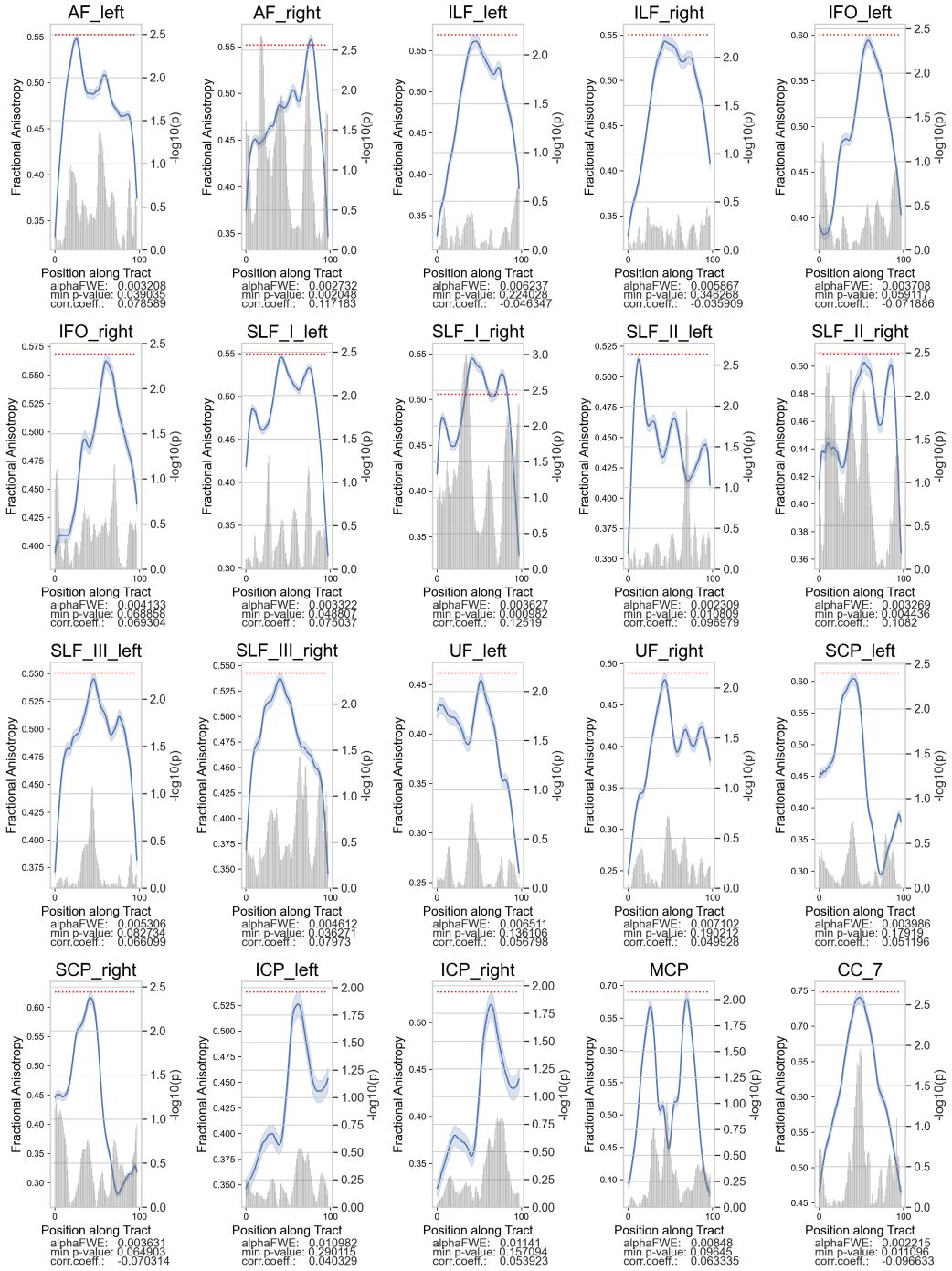

Figure S7: Only the right AF and SLF I FA values were positively associated with TOWRE SWE scores across the entire cohort. Plotting convention is the same as in the previous figure.

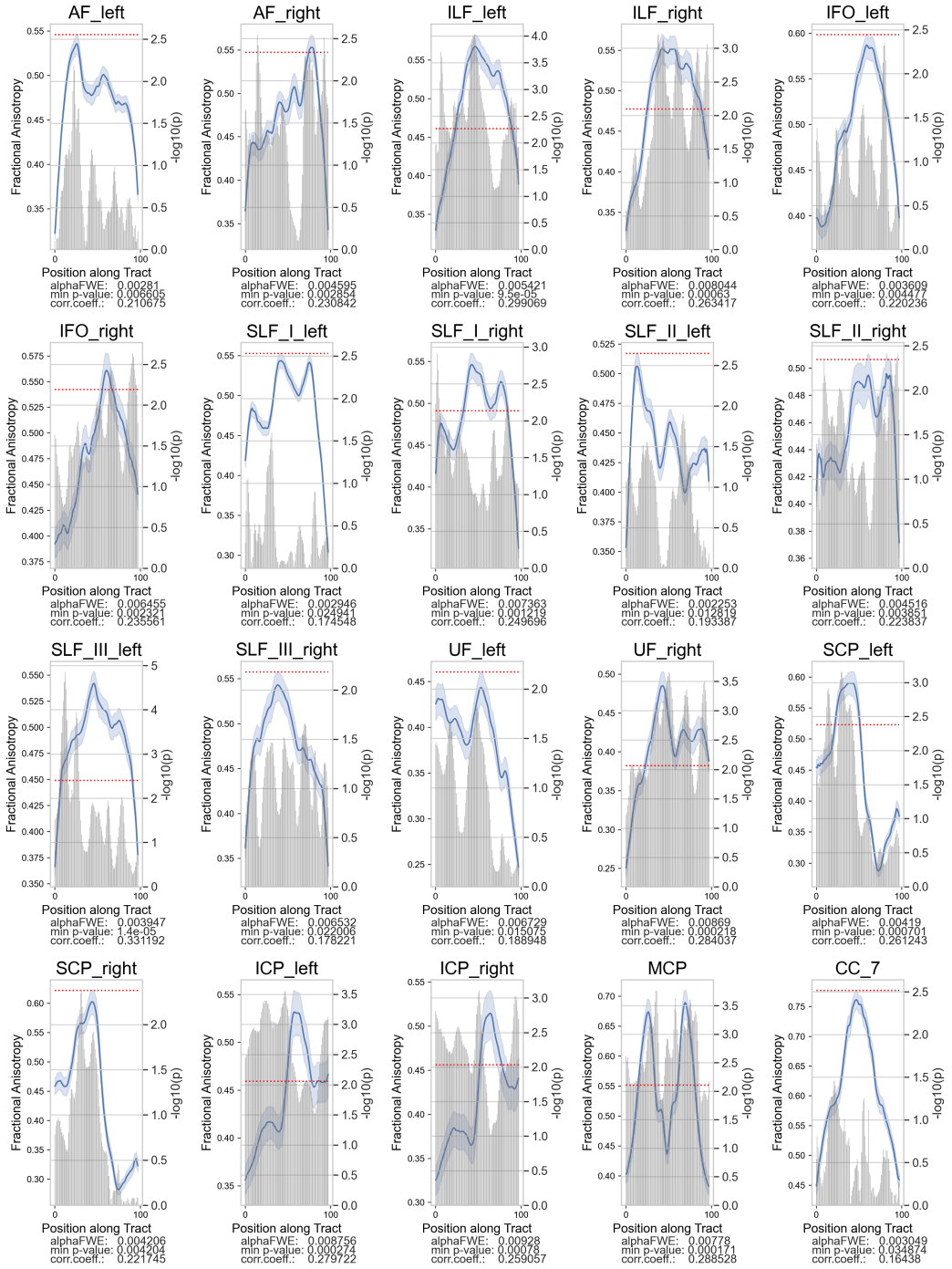

Figure S8: Several fiber bundles exhibited significant FA correlations with TOWRE PDE scores among the Low-TOWRE group. Plotting convention is the same as in the previous figure.

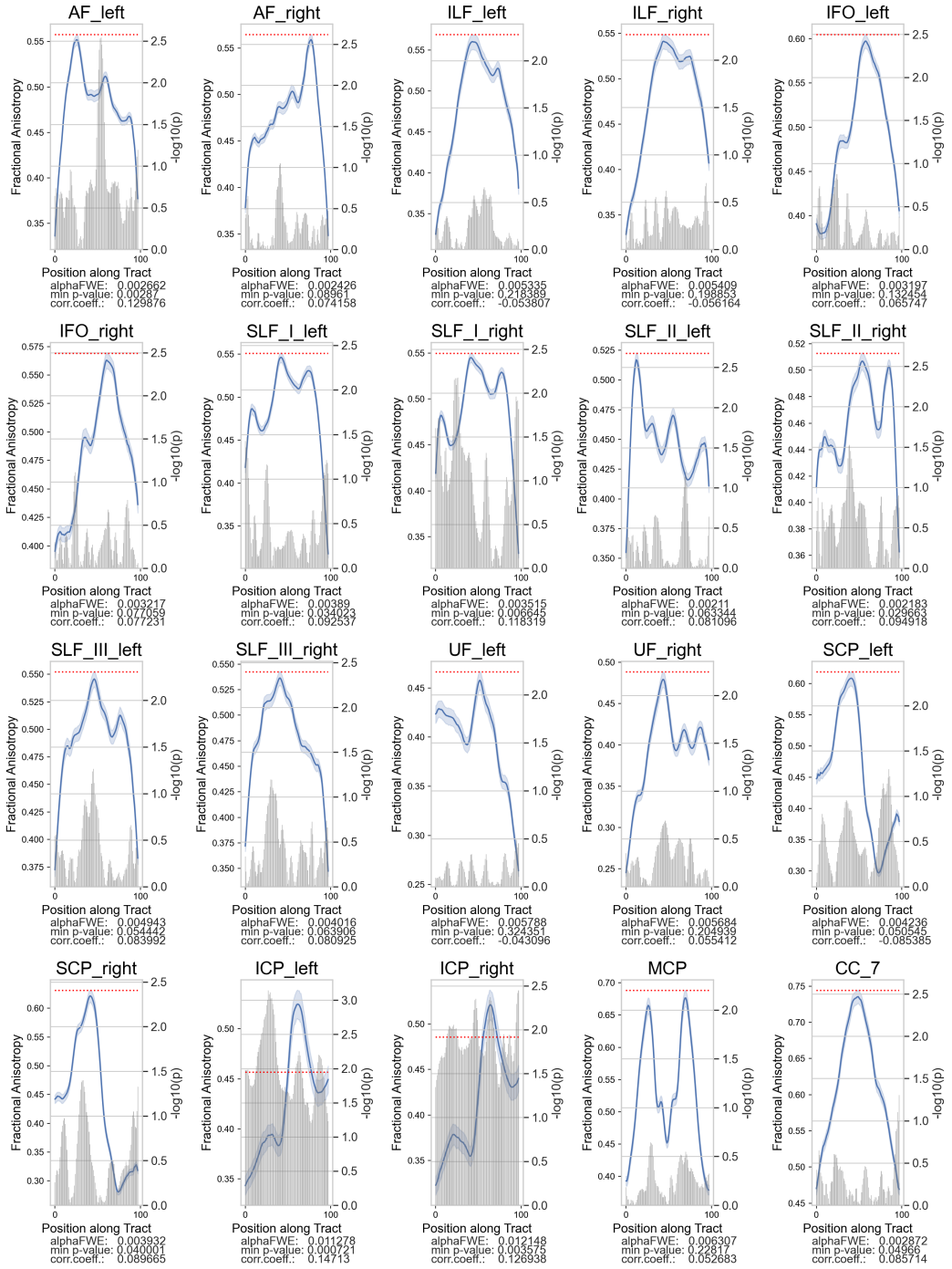

Figure S9: Only bilateral ICP displayed significant FA correlations with TOWRE PDE scores among the High-TOWRE group. Plotting convention is the same as in the previous figure.

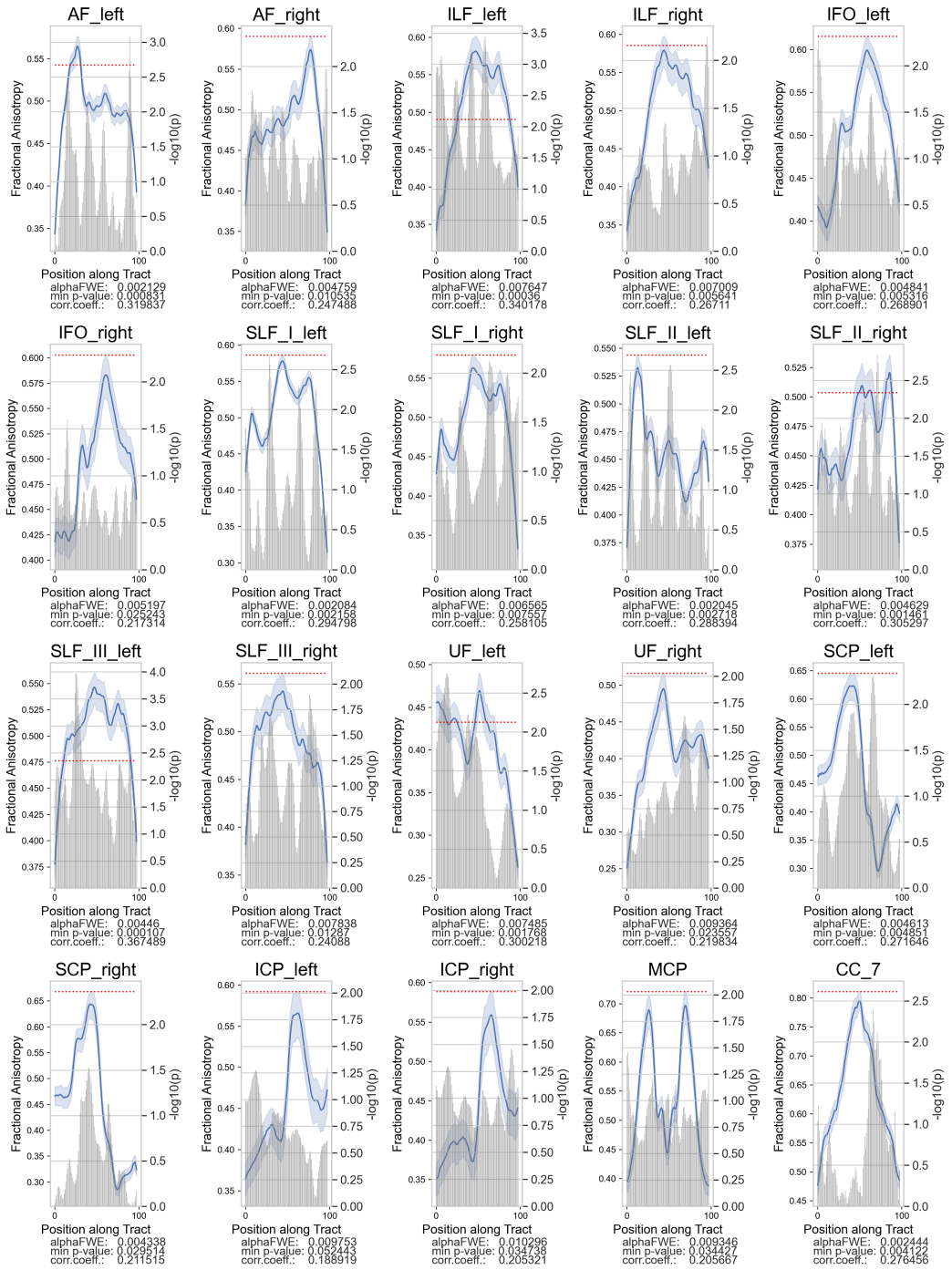

Figure S10: Several fiber bundles exhibited significant FA correlations with TOWRE PDE scores among the RD group. Plotting convention is the same as in the previous figure.

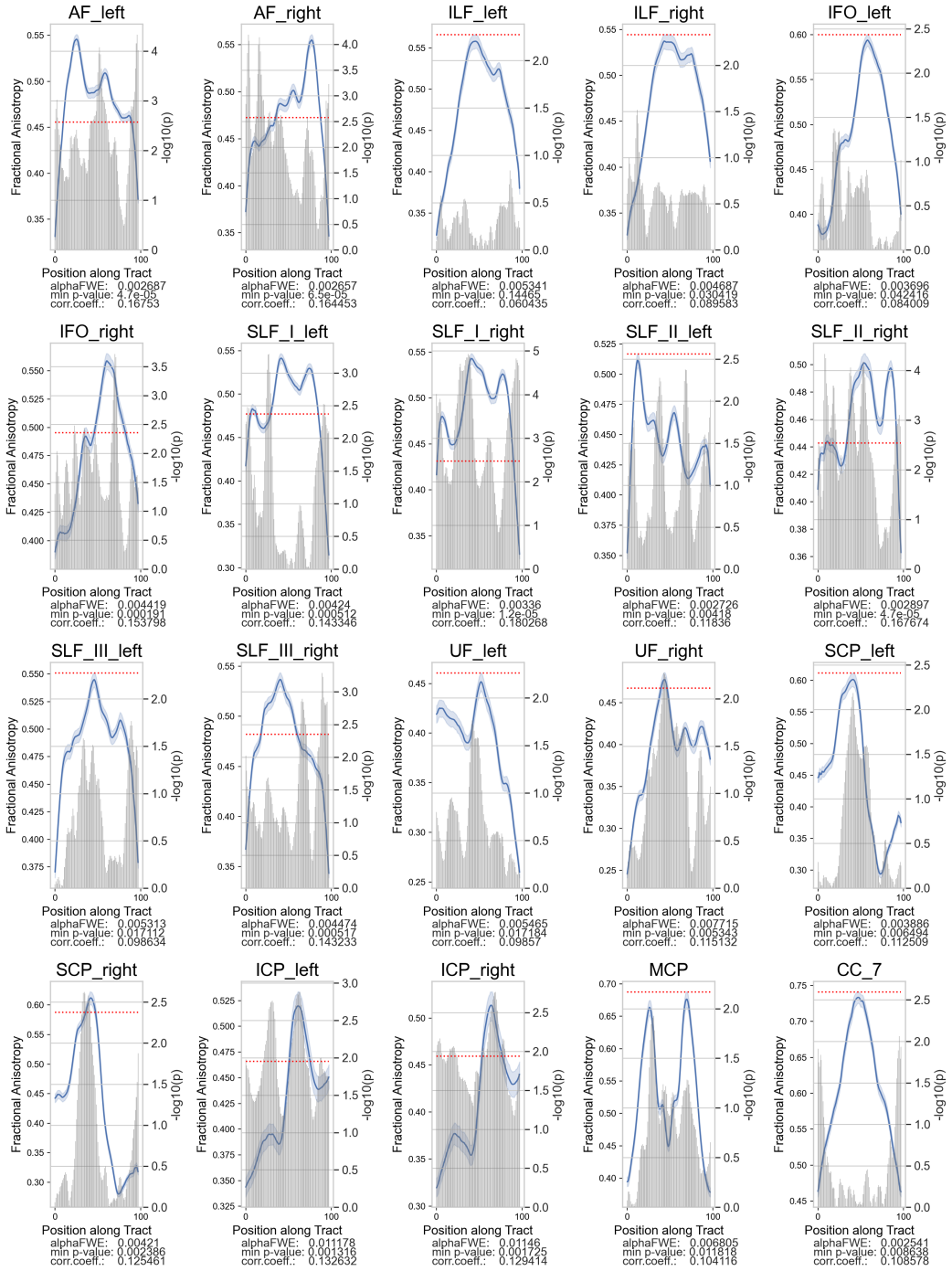

Figure S11: Several fiber bundles exhibited significant FA correlations with TOWRE PDE scores among the TR group. Plotting convention is the same as in the previous figure.

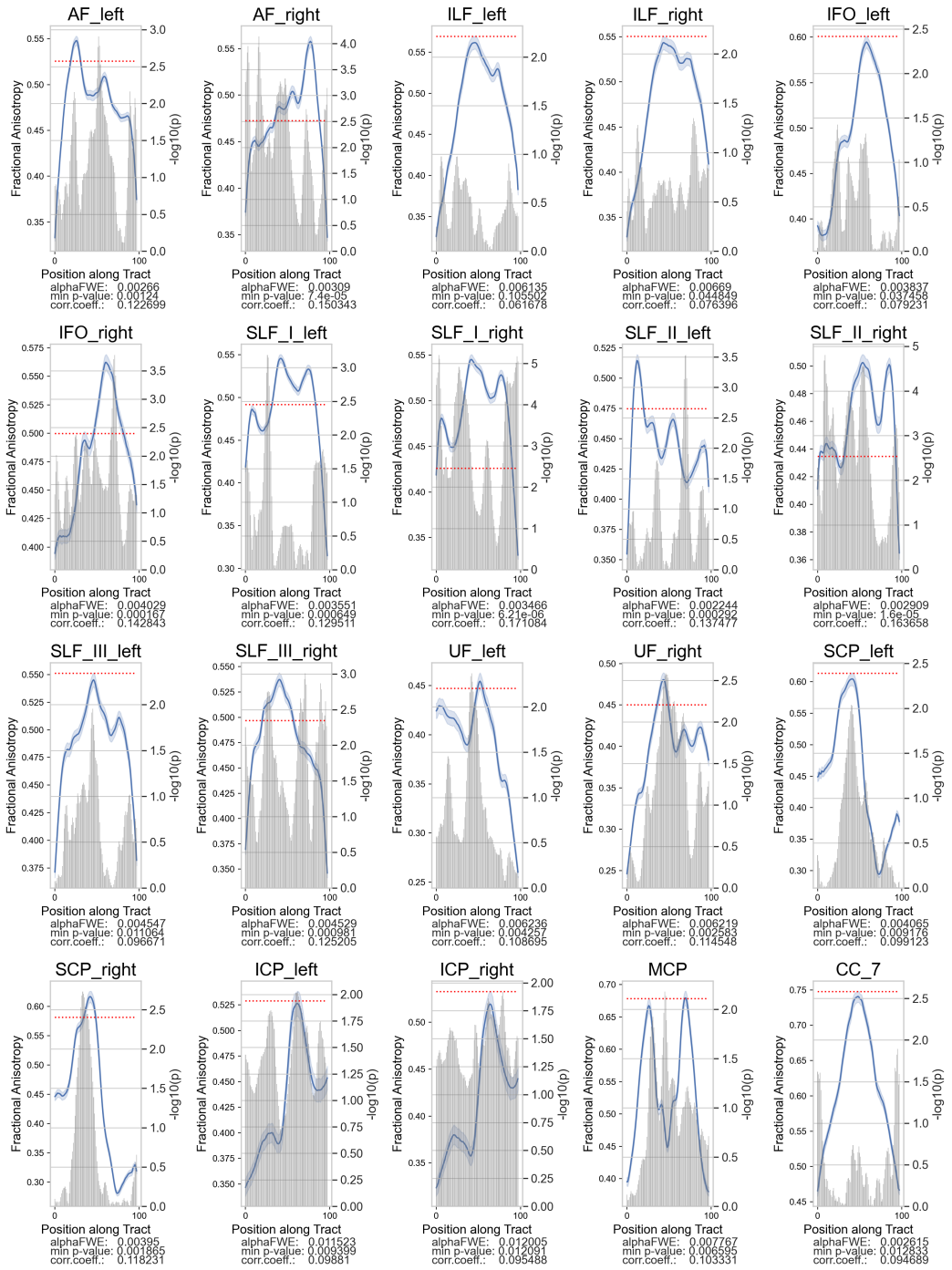

Figure S12: Several fiber bundles exhibited significant FA correlations with TOWRE PDE scores across the entire participant cohort. Plotting convention is the same as in the previous figure.

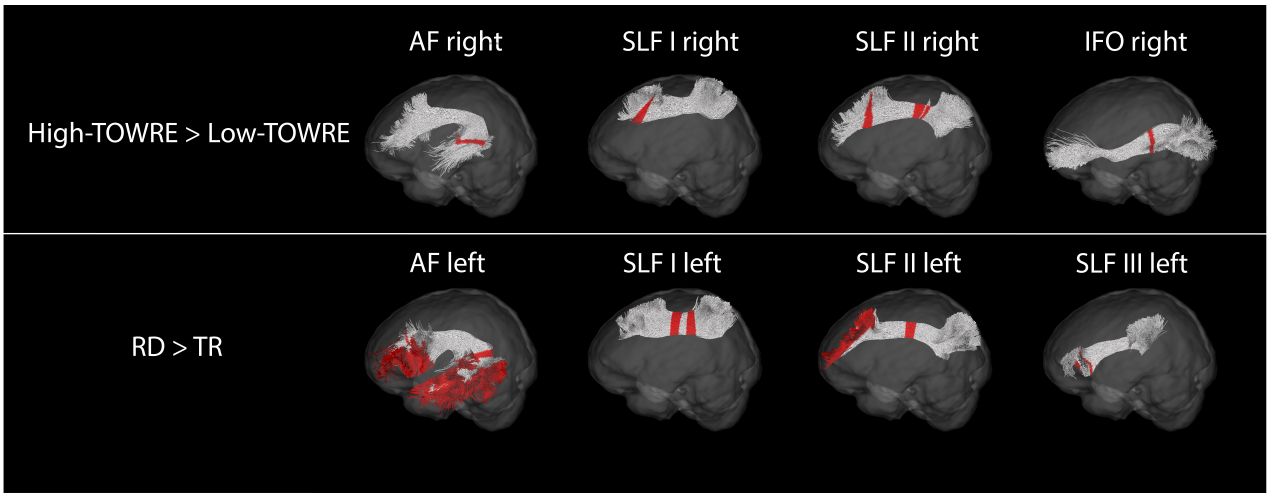

Figure S13: Anatomical depictions of white matter regions where group differences in FA existed between the Low and High-TOWRE groups and between the RD and TR groups.