Fortier MA, Little G, Bautin P, Sean M, Tétreault P. Towards repeatable and converging methods in diffusion MRI: Evidence from a longitudinal chronic pain cohort. *Aperture Neuro*. 2026;6. doi:10.52294/001c.161956

Original Research Articles

Towards repeatable and converging methods in diffusion MRI: Evidence from a longitudinal chronic pain cohort

Marc-Antoine Fortier, MSc[1,2], Graham Little, PhD[1,3], Paul Bautin, MSc[1,2], Monica Sean, PhD[1,2], Pascal Tétreault, PhD[1,2][a]

[1] Department of Anesthesiology, Université de Sherbrooke, Sherbrooke, QC, Canada, [2] Centre de Recherche du CHUS, Sherbrooke, QC, Canada, [3] Jump Ship Labs, Edmonton, AB, Canada

Keywords: diffusion MRI, tract-based spatial statistics, fractional anisotropy, white matter, longitudinal neuroimaging, reproducibility, chronic low back pain, tractometry, I2C2, image intraclass correlation coefficient

https://doi.org/10.52294/001c.161956

Aperture Neuro

Vol. 6, 2026

Studies of white matter (WM) alterations in chronic low back pain (CLBP) using diffusion MRI (dMRI) have yielded inconsistent results, highlighting a need for more reproducible methods. This study introduces a longitudinal analysis pipeline designed to identify stable and repeatable WM microstructural differences between CLBP patients and healthy controls. Diffusion MRI was acquired from 27 CLBP patients and 25 control participants at three separate visits with two-month intervals. A customized Tract-Based Spatial Statistics (TBSS) analysis was performed on fractional anisotropy (FA) maps that were averaged across visits for each participant. Resulting clusters showing group differences were subsequently filtered based on cluster size (>10 voxels) and a high repeatability threshold (image intra-class correlation coefficient [I2C2] > 0.75) to ensure findings were stable across all three imaging sessions. Results were compared against standard single-visit analyses and an alternative tractometry analysis. The repeatability analysis identified 11 clusters with stable and significant FA differences. Seven clusters, located primarily in the occipital, parietal, and frontal lobes, showed higher FA in controls. Four clusters, located in the frontal and temporal lobes, showed higher FA in the CLBP group. In contrast, single-visit analyses identified a much larger number of clusters (21 to 36), the majority of which were not spatially consistent across time and did not overlap with the final repeatable clusters. The repeatable TBSS findings demonstrated a strong spatial correspondence with group differences found using a tractometry analysis. By incorporating a longitudinal design and explicit repeatability filtering, our method effectively reduces spurious findings common in single-visit dMRI studies. This approach successfully isolated reliable WM regions with altered dMRI metrics in CLBP, demonstrating its value in improving the robustness of neuroimaging research in chronic pain.

INTRODUCTION

The association of chronic pain (CP) to alterations in human brain structure and function has become a well-established fact.[1] In-vivo magnetic resonance imaging studies have demonstrated CP related alterations to brain function and structure but have primarily focused on human grey matter.[2,3] These findings have helped clinicians appreciate CP as a complex disease involving the central nervous system (CNS) and propose mechanisms of pain chronicization from observed brain alterations.[4] Neuroimaging also shows promise as a biomarker, as it could identify functional and structural brain patterns compatible with a certain prognosis or treatment response in CP[5-8] and therefore optimize clinical decisions at an individual level.[9,10] Some examples of neuroimaging biomarker applications already appear in the literature, such as predicting chronicization of postsurgical pain,[11] predicting response to a pharmacological pain treatment[12] and using brain imaging marker for precision medicine application in low back pain.[13]

[a] **Corresponding Author:**
Pascal Tétreault, PhD (Pascal.Tetreault@USherbrooke.ca)
Department of Anesthesiology, Faculty of Medicine and Health Sciences, Université de Sherbrooke
Sherbrooke, QC, Canada

As understanding of CP has evolved, CP has been theorized to be more broadly considered a network disorder, warranting the study of the human brain's white matter (WM) which encompasses the microstructural fibers that transfer information between disparate regions of the brain. Diffusion magnetic resonance imaging (dMRI) is a technique that measures the anisotropic movement of water molecules in-vivo and is sensitive to subtle changes to WM microstructure.[14] Historically in CP literature, the most common approach of inferring microstructural alterations in human WM with dMRI data, is to fit a three-dimensional tensor and report parametric values extracted from the tensor such as fractional anisotropy (FA). Fractional anisotropy is a metric that reflects the variation of diffusion along the tensor's primary axis relative to the perpendicular axes, and is related to the underlying microstructural organization/composition of tissue within a voxel. To date, studies using dMRI to examine human WM as it relates to chronic pain remain scarce and have produced inconclusive results.[15] Results differ considerably between primary chronic pain conditions (e.g., chronic primary headache pain, chronic primary visceral pain, chronic primary musculoskeletal pain) with a multitude of dMRI acquisition and analysis methods implemented, leading to difficulty in comparing and interpreting differences between studies.

Even in studies where specific pain conditions were targeted (e.g., chronic low back pain (CLBP) and subacute back pain (SBP)), various analysis methodology was used and notable inconsistencies have been reported in the location and the degree of WM FA group differences with a more consistent finding of lower FA in CLBP patients relative to controls. Three studies have shown reduced baseline FA in CLBP compared to controls with varying sample sizes; 102 CLBP patients, 50 controls[16]; 24 CLBP patients, 22 controls[17]; 31 CLBP patients, 31 controls.[6] Two studies have reported post-therapeutic effects measurable by dMRI in CLBP patients with FA increasing in WM regions after therapy.[16,18] One study of patients with subacute back pain successfully predicted those who transitioned to a chronic pain state using measurements of FA along WM tracts,[6] with a follow-up study more robustly predicting this transition using dMRI tractography derived measures of connectivity.[19] While some agreement is found among reports of lower FA associated with CLBP, regions implicated in CLBP vary substantially between studies: left insular WM[18]; primary somatosensory cortex WM[16]; corpus callosum, bilateral anterior thalamic radiation, right posterior thalamic radiation, right superior longitudinal fasciculus, and the left anterior corona radiata[17]; the left superior longitudinal fasciculus, the left internal and external capsules, and anterior corona radiata.[6]

The problem of reproducible results is not confined to the field of chronic pain but is one of the fundamental challenges facing the neuroimaging field at large.[20] Many studies have addressed repeatability in test-retest studies of dMRI measurements reporting generally reliable dMRI measures in healthy adult controls,[21-23] across various age ranges[24-26] and in patient cohorts[27-29] with FA showing the highest reliability among diffusion tensor measurements in most cases. Repeatability has also been studied longitudinally, demonstrating that more advanced dMRI modelling produces metrics with high levels of repeatability over a 3 month interval.[30] However, these studies only assess the repeatability of dMRI metrics within controls and patients separately. To reconcile the inconsistent findings in the CP neuroimaging literature and identify robust biomarkers that accurately represent the persistent changes of neural substrates in CP, it is important to identify regional WM differences between controls and CP patients that can be found repeatedly and at multiple clinically relevant time points. Also, while acquisition strategies such as repeated dMRI acquisition within a single session can improve signal-to-noise ratio and reduce within-session variability,[31,32] they do not capture variability arising across independent imaging sessions. In contrast, multi-visit designs enable the assessment of between-session variability, including factors such as repositioning, physiological fluctuations, and scanner-related effects.[33,34] Evaluating the reproducibility of imaging findings across separate visits therefore provides an additional level of robustness, allowing identification of spatial patterns that are stable over time rather than driven by session-specific noise. Such longitudinal consistency is particularly important for the development of reliable neuroimaging biomarkers and aligns with the broader objective of advancing toward precision medicine, where stability of individual-level measures across time is essential.

Here we present a method of experimental design and voxel-wise analysis that uses the stability of dMRI derived measurements of WM microstructure as a means to filter regional differences to only include group differences that are repeatable across three scans. The proposed repeatability analysis was applied to a pilot cohort of 25 controls and 27 CLBP patients with imaging acquired at 3 separate visits. As a comparison to the most commonly reported voxel-wise analysis method in the CP literature, we chose to use tract-based spatial statistics (TBSS), projecting adjacent FA values to a voxel-wise skeleton of a subject's WM. Regions (clusters) that were found to have statistically significant between group differences were further filtered by image intra-class correlation coefficient (I2C2) values to remove regions that had non-repeatable measurements. In total, 11 regions were observed to have FA values that were stable across visits and were different between CLBP and control groups (7 higher in controls, 4 higher in CLBP). To compare our results to other more advanced methodology, an exploratory analysis was performed showing strong correspondence between the proposed repeatability analysis and tractometry, a method capable of segmenting and analyzing individual WM tracts from dMRI. The proposed repeatability analysis technique provides a method to incorporate reproducibility metrics into the analysis design and in the context of the CLBP pilot imaging study, detects regions with repeatable CLBP related group differences in FA, reducing the chance of spurious findings.

METHODS

PARTICIPANTS

This study was approved by the ethics review board of the Centre intégré universitaire de santé et de services sociaux de l'Estrie – Centre hospitalier universitaire de Sherbrooke (CIUSSS de l'Estrie – CHUS), Sherbrooke, Canada (file number: 2021-3861). The study is also registered on the Open Science Framework ("Pilot project on brain and lower back imaging of chronic pain", https://doi.org/10.17605/OSF.IO/P2Z6Y).

Control and CLBP participants were recruited and asked to visit the research center of the CHUS for cerebral and lumbar MRI scans for a total of three visits with a two-month interval. Saliva samples and questionnaires were also collected as a part of the larger study. Apart from their usual treatment, no therapeutic intervention was modified or administered to participants between visits. Participants were permitted to continue or initiate usual care during the study period (excluding treatments specified in the exclusion criteria) and were required to report any changes to their treatment regimen. All CLBP participants underwent pain intensity screening using the Brief Pain Inventory (BPI) short form[35] to quantify their symptoms.

Exclusion and inclusion criteria for CLBP and HC with full details described in previous studies on behavior and grey matter density.[36,37] In short, CLBP participants were included if they had low back pain for greater than 4 months, an average pain intensity of greater than or equal to 3/10 on the BPI within the 24-hour period before the initial visit and pain primarily localized to the low back. Control participants were excluded if they had a history of chronic pain, had pain at the time of testing or an outstanding painful episode within 3 months of enrollment.

In total, 27 CLBP (mean age 43 ± 16 years, 13 females) and 25 control participants (mean age 42 ± 14 years, 15 females) were recruited for this study, and had adequate imaging data for at least one of the 3 visits. Amongst these cohorts, 23 CLBP participants and 24 control participants completed all three visits or had adequate diffusion and anatomical imaging data acquired for the subsequent analysis. Study participants were selected to ensure no discrepancy in demographic, clinical or neuropsychological characteristics except for the pain inventory, as summarized in Table 1.

MRI ACQUISITION

Imaging data was acquired on a Philips Ingenia 3.0T MRI located at the CHUS. For each subject and visit, whole brain T1-weighted structural imaging and diffusion weighted imaging (DWI) were acquired. Structural T1-weighted imaging (1.0 x 1.0 x 1.0 mm^3) was acquired with TE = 3.5 ms, TR = 7.9 ms, flip angle = 8°, and a scan time of 4:19 min. DWI was acquired with a single-shot EPI spin-echo sequence (Multiband factor = 2, SENSE reduction factor = 1.9, Partial Fourier (Halfscan) factor = 0.6949152), TR = 4800 ms, TE = 92 ms, flip angle = 90°, 66 slices were acquired at a 2 mm thickness with no gap and 2.0 × 2.0 mm^2 in-plane. In total, 108 diffusion volumes acquired using 7 b = 0 s/mm^2, 8 b = 300 s/mm^2, 32 b = 1000 s/mm^2, 60 b = 2000 s/mm^2, and 1 = 0 s/mm^2, acquired in the reversed phase encode direction in a scan time of 9:19 min. All T1-weighted and diffusion-weighted images were visually inspected for motion artifacts, including ringing and blurring, and no scans were excluded on this basis. During the diffusion scan, no difference in motion was reported between CLBP and HC (Figure S1).

PREPROCESSING

Structural T1 and diffusion weighted images for each subject and visit were preprocessed using the Tractoflow pipeline.[38] In short, the Tractoflow pipeline performs image denoising,[39] magnetic field susceptibility correction,[40] eddy current correction,[41] non-linear registration from T1 imaging space to diffusion imaging space using ANTs,[42] brain extraction and tissue segmentation on T1 images,[43] image intensity normalization,[44] spatial resolution resampling, outputting DTI parametric (e.g., fractional anisotropy (FA)) and fiber orientation distribution function (fODF) maps. T1-to-dMRI co-registration did not induce any significant group variations and stayed consistent across visits as shown in supplementary analysis (Figure S2).

TRACT-BASED SPATIAL STATISTICS

To leverage the longitudinal design of the dataset, a customized repeatability analysis pipeline was used to extract regions (clusters) that had a measurable group difference at all three visits. The custom pipeline is visualized in Figure 1. Using the preprocessed FA maps output from the previous steps, Tract-Based Spatial Statistics (TBSS),[45] an automated voxel-wise diffusion analysis tool (FMRIB software suite), was performed using default parameters by first generating a WM skeleton, at a spatial resolution of 1 mm x 1 mm x 1 mm, for the entire cohort (all subjects and scans included) and then projecting FA values to this skeleton for each subject and scan.

REPEATABILITY ANALYSIS

To improve the sensitivity to detect group differences, skeletonized FA maps were first averaged across visits for each subject generating a single FA map per subject. A two-sample t-test was then performed (FSL randomise) between groups using the subject FA maps averaged across visits outputting a single parametric t-statistic map along the skeleton. The parametric map was thresholded at a t-statistic value of $|2.5|$ ($p < 0.01$) and remaining clusters were excluded if they were less than 10 voxels in size (i.e,. a volume of 10 mm^3, on 1mm^3 WM skeleton). Finally, to remove clusters that had unreliable FA values across visits, the image intra-class correlation (I2C2)[46] was calculated for each remaining cluster using only the subjects that had adequate imaging acquired at all 3 visits. Clusters with I2C2 values less than 0.75 (good to excellent reliability,[46,47]) were re-

Table 1. Characteristics of the study participants.

	CLBP	HC	p-value
Baseline (BL)	27	25	-
2 months (2M)	24	25	-
4 months (4M)	23	25	-
Intracranial volume (ICV) (cm^3)			
BL	1530.31 ± 155.30	1532.76 ± 154.37	p=0.955
2M	1537.03 ± 163.59	1530.78 ± 152.40	p=0.890
4M	1541.83 ± 153.97	1535.64 ± 153.77	p=0.890
Women	13	15	p=0.419
Men	14	10	
Age (average ± std)	43 ± 16	42 ± 14	p=0.700
Brief Pain Inventory - severity (0 – 40)			
BL	17 ± 4	NA	NA
2M	18 ± 6	NA	
4M	15 ± 7	NA	
Pain Detect - now (0 – 10)			
BL	4.4 ± 1.6	NA	NA
2M	4.5 ± 1.6	NA	
4M	3.7 ± 2.0	NA	
STAI - state (20 – 80)			
BL	33 ± 8	26 ± 6	p=8.5E-04
2M	33 ± 9	26 ± 6	p=0.0046
4M	32 ± 10	26 ± 8	p=0.023
STAI - trait (20 – 80)			
BL	35 ± 10	29 ± 9	p=0.035
2M	35 ± 12	29 ± 9	p=0.056
4M	35 ± 12	29 ± 10	p=0.079
Pain duration			
4 months - 5 months	0	NA	NA
4 months - 5 months	0	NA	
6 months - 12 months	5	NA	
1 - 4 years	8	NA	
5 years and more	14	NA	

moved leaving clusters with a measurable group difference and had repeatable values across visits. All code to perform this portion of the analysis has been made openly available (https://github.com/Tetreault-Pain-Imaging-Lab/tpil_i2c2_filter/releases/tag/v1.0.0-beta).

COMPARISON TO SINGLE VISITS

As a comparison to standard cross-sectional designs, the same group difference analysis, parameter map thresholding (t-statistics greater than |2.5|) and cluster size thresholding (clusters containing greater than 10 voxels) were performed separately for each visit. Resulting clusters were then separated into those where controls had greater FA compared to the CLBP group (t-statistics greater than 2.5) and those where CLBP had greater FA compared to the control group (t-statistics less than 2.5). Clusters retained from the three analyses performed for each visit separately were compared to those retained from the proposed repeatability analysis as follows. First, the total number of clusters retained after statistical/voxel size thresholding at each visit were totaled. Then the clusters retained at each visit were compared to those retained after the proposed repeatability analysis to identify the number of clusters that spatially overlap in both analyses. For example, for the visit 1 analysis, both the total number of clusters retained and the total number of clusters that overlap the proposed repeatability analysis are calculated.

COMPARISON TO ALTERNATIVE METHODOLOGY: TRACTOMETRY ANALYSIS

To compare results from the proposed repeatability analysis with alternative methodology, a supplementary automated tractometry analysis was performed. Whole brain tractograms were generated using the previously used

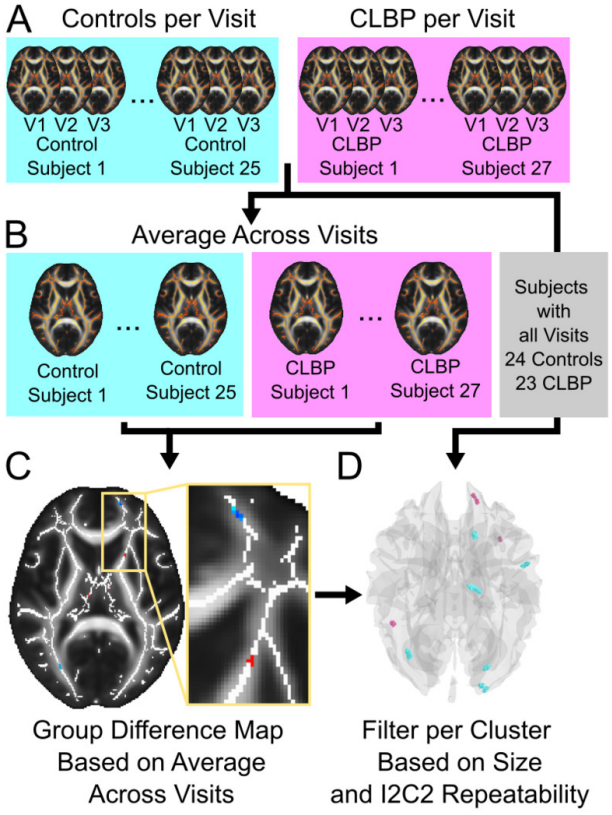


Figure 1. Workflow of analysis used to generate clusters with regional differences between controls and chronic low back pain (CLBP) subjects that are repeatedly observed across visits. A) 25 control subjects and 27 CLBP subjects were processed to generate skeletonized FA maps per subject for each visit (V1, V2, V3). B) Skeletonized FA maps were averaged across visits to attain a single map per subject. C) A group difference map was generated based on voxel-wise t-tests and thresholded to include voxels with a t-statistic value greater than |2.5| ($p < 0.01$). D) Clusters were filtered to include only clusters that were greater than 10 voxels in size and had an I2C2 value greater than 0.75. I2C2 values for each cluster were calculated using the subset of subjects that had imaging data available for all three visits.

Tractoflow pipeline[38] with default analysis parameters. Whole brain tractograms were then segmented into separate bundles using BundleSeg[48] with a whole brain WM tract atlas (Rheault, 2023), outputting 51 WM tracts per subject. Six clusters (4 control FA > CLBP FA in the occipital lobe and 2 CLBP FA > control FA located in the superior frontal lobe) were selected for further investigation from the TBSS repeatability analysis because they were located in adjacent regions. Tracts overlapping TBSS clusters were identified by computing the spatial overlap between the thresholded TBSS cluster masks and binary voxel masks derived from a population-average WM tract atlas. Each atlas tract was converted to a binary voxel mask using DIPY's density mapping, and the voxel-wise intersection with each cluster was calculated. Tracts were considered overlapping if any streamline voxels intersected the cluster mask and were retained for tractometry analysis. To further localize where along each tract the group differences occurred, overlapping bundles were subdivided into 50 equidistant segments using centroid-based label mapping (scil_bundle_label_map) and average FA was calculated for each segment.[28] To test for segments with different FA values between controls and CLBP, two-sample t-tests ($p < 0.05$, uncorrected) were performed per segment per visit. For comparison to the repeatability analysis, tract segments that overlapped clusters output from the proposed analysis were identified.

RESULTS

GROUP AVERAGE REGIONAL DIFFERENCES IN FRACTIONAL ANISOTROPY

The repeatability analysis pipeline identified 11 WM clusters with stable group differences in fractional anisotropy (FA) across all three visits. The 11 clusters are visualized in Figure 2, along with average FA values extracted for each cluster. Seven of these clusters showed significantly higher FA in the control group compared to the CLBP group, while four clusters showed higher FA in the CLBP group. The distribution of these FA values for each participant, averaged across the three visits, is plotted in Figure 2B (Controls > CLBP) and Figure 2C (CLBP > Controls), illustrating a clear separation between the groups.

Among the 7 clusters showing higher FA in the relative to the CLBP group, 4 were located in the occipital lobe (left hemisphere: clusters 4,5,6; right hemisphere; cluster 7), 2 were located in the left parietal lobe and 1 was located in the left frontal lobe. Among the 4 clusters showing higher FA in the CLBP group relative to controls, 3 were located in the left frontal lobe (clusters A, B, C) and cluster D located in the posterior portion of the right temporal lobe.

Detailed statistics for each of the 11 clusters are presented in Table 2. The clusters ranged in size from 10 mm^3 to 48 mm^3. Effect sizes were moderate to large, with Cohen's D values ranging from 0.818 to 1.043 for clusters where controls had higher FA, and from -0.818 to -0.982 for clusters where CLBP patients had higher FA. All group differences were statistically significant ($p < 0.005$, uncorrected for multiple comparisons). All 11 clusters demonstrated good to excellent repeatability, with Image Intraclass Correlation Coefficient (I2C2) values for the entire cohort ranging from 0.7685 to 0.8821 (good to excellent[46, 47]). We also performed exploratory analyses searching for correlation between the identified clusters and some of the pain and behavioural measures collected. We observed a stable relationship between cluster 1 (located in the left Fronto-Pontine Tract) and two pain measures (severity component of the Brief Pain Inventory and reported pain at time of sampling from the PainDetect) as well as the State and Trait Anxiety Inventory (for both the State and Trait scales) (Figure S3).

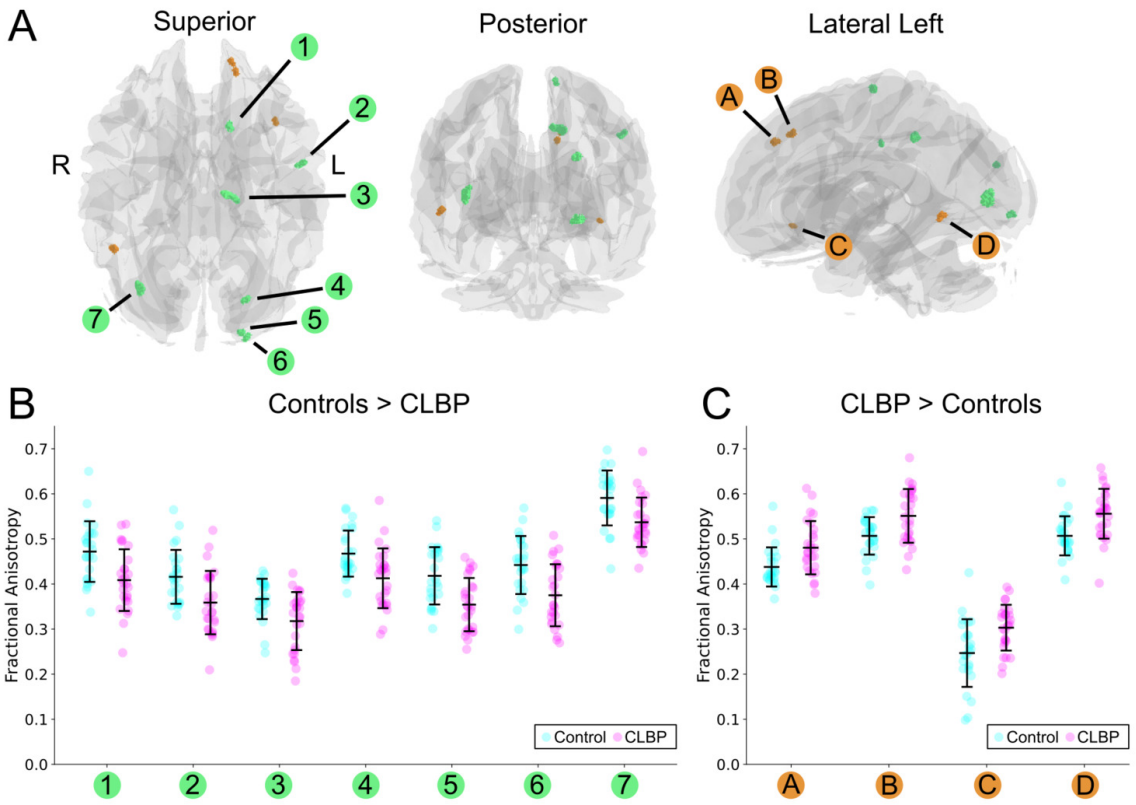


Figure 2. A) Visualization of clusters with different fractional anisotropy values between groups (t-statistic greater than |2.5|), cluster size greater than 10 voxels and an I2C2 value greater than 0.75. Seven regions where fractional anisotropy was higher in controls relative to the chronic low back pain group are shown (green) along with 4 regions where fractional anisotropy was higher in the chronic low back pain group compared to controls (orange). Average fractional anisotropy values along with standard deviations (black) are shown for individual controls (cyan) and CLBP (pink) subjects averaged across visit for B) regions with higher fractional anisotropy in controls and C) regions with higher fractional anisotropy in the chronic low back pain group.

Table 2. Group differences and repeatability of fractional anisotropy averaged for each cluster.

	Cluster Size (mm^3)	Cluster MNI Coordinates (x, y, z)	Control FA Mean ± Std (N=24)	CLBP FA Mean ± Std (N=27)	Cohen's D[a]	P-Value[b]	I2C2
Controls > Chronic Low Back Pain							
Cluster 1	15	(-11, -17, 65)	0.47 ± 0.07	0.41 ± 0.07	0.933	0.001	0.8821
Cluster 2	15	(-44, -21, 38)	0.42 ± 0.06	0.36 ± 0.07	0.873	0.003	0.8473
Cluster 3	34	(-12, -39, 40)	0.37 ± 0.04	0.32 ± 0.06	0.880	0.002	0.8178
Cluster 4	13	(-18, -82, 26)	0.47 ± 0.05	0.41 ± 0.07	0.919	0.002	0.8298
Cluster 5	12	(-17, -89, -1)	0.42 ± 0.06	0.35 ± 0.06	1.043	0.000	0.8189
Cluster 6	11	(-20, -91, 0)	0.44 ± 0.06	0.38 ± 0.07	1.006	0.001	0.7685
Cluster 7	48	(30, -71, 9)	0.59 ± 0.06	0.54 ± 0.05	0.933	0.002	0.8534
Chronic Low Back Pain > Controls							
Cluster A	16	(-13, 34, 38)	0.44 ± 0.04	0.48 ± 0.06	-0.818	0.004	0.8066
Cluster B	16	(-15, 26, 42)	0.51 ± 0.04	0.55 ± 0.06	-0.857	0.003	0.8128
Cluster C	10	(-37, 27, -8)	0.25 ± 0.08	0.3 ± 0.05	-0.885	0.003	0.8288
Cluster D	16	(43, -48, 2)	0.51 ± 0.04	0.56 ± 0.06	-0.982	0.001	0.7816

[a]Cohen's D calculated as control FA mean minus CLBP FA mean divided by pooled CLBP and HC standard deviation (i.e. negative value indicates CLBP > control mean)
[b]Group differences assessed with a two sample t-test at $p < 0.05$, uncorrected for multiple comparisons

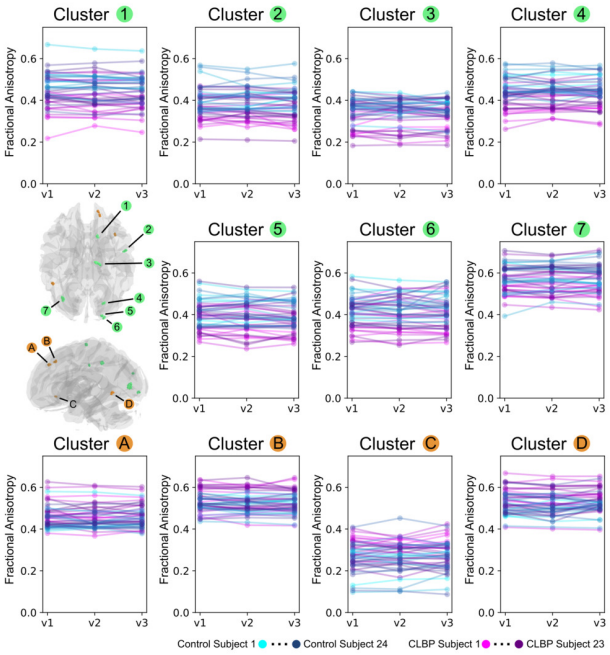


Figure 3. Mean fractional anisotropy calculated for regions remaining after the proposed repeatability analysis. The 7 regions where controls had higher FA relative to CLBP are shown in green and regions where CLBP had higher FA relative controls are shown in orange. Average fractional anisotropy is displayed for each visit for the 24 control subjects (separate colours cyan to navy) and 23 chronic low back pain subjects (separate colours pink to indigo) that had imaging data collected at all three visits. Values and group differences were stable across the three visits for all regions.

STABILITY OF FINDINGS ACROSS VISITS

To confirm the stability of these findings, the group differences within the 11 identified clusters were examined at each visit independently. The mean FA values are shown for all 11 clusters for each subject and visit in Figure 3. The mean FA values for individual subjects remained consistent throughout the three visits, with minimal intra-subject variance for both control and CLBP participants.

Table 3 provides a quantitative analysis of the stability of fractional anisotropy group differences, reporting the mean FA, Cohen's D, and p-values for each cluster at each of the three visits. As intended by the design of the repeatability analysis, the direction of the group difference (e.g., Controls > CLBP) and the statistical significance were highly consistent for all 11 clusters across each time point. For example, Cluster 7 (FA Controls > FA CLBP) had p-values of 0.002, 0.003, and 0.004 (uncorrected for multiple comparisons) for visits 1, 2, and 3, respectively along with high I2C2 (Table 2. I2C2: 0.8534) which was used for thresholding clusters. This demonstrates that the repeatability pipeline is correctly identifying regions with consistent group differences at each visit along with the previously stated minimal intra-subject variance.

REDUCED CHANCE FINDINGS RELATIVE TO A SINGLE-VISIT DESIGN

A key advantage of the proposed repeatability analysis is its ability to filter out spurious or non-repeatable findings that may arise in standard cross-sectional studies. As illustrated in Figure 4, analyzing each visit separately identified a large number of significant clusters, the majority of which were not consistent across visits. Figure 4A shows an example of a repeatable cluster, where the significant regions from each visit substantially overlap with the final cluster identified by the repeatability analysis. In contrast, Figure 4B provides an example of unreliable findings, where clusters were identified at single visits but showed no spatial agreement across time.

The bar charts in Figures 4C, 4D, and 4E quantify this effect. At Visit 1, a standard analysis yielded 36 clusters (28 for Controls > CLBP, 8 for CLBP > Controls), but only 9 of these (6/7 and 3/4) overlapped with the 11 final repeatable clusters. Similarly, at Visit 2, 21 clusters were found, with only 5 overlapping the final results. At Visit 3, 28 clusters were found, with only 5 overlapping. These results strongly suggest that most clusters identified in a single-visit analysis are chance findings that are not stable over time, highlighting the value of the proposed repeatability-based filtering approach. Further examinations revealed that clusters identified in the proposed repeatability analysis but absent in individual-visit analyses did not exceed the minimum voxel size threshold when assessed at single visits.

ANATOMICAL LOCALIZATION OF REPEATABLE FINDINGS

To place the identified clusters into anatomical context, a select subset of clusters was overlaid on major WM tracts, as shown in Figure 5. The 4 clusters residing in the occipital lobe with higher FA in controls were situated near known sensory and visual pathways. Specifically, clusters 4, 5, and 6 are adjacent to the left optic radiation and the occipital portion of the corpus callosum, while cluster 7 is near the right optic radiation. The two selected clusters with higher FA in the CLBP group were located near frontal tracts. As shown in Figure 5B, clusters A and B are adjacent to the left fronto-pontine tract, the frontal portion of the corpus callosum, and the left frontal aslant tract. It is important to note, however, that the proximity of these clusters to regions of crossing fibers impedes their definitive localization to a single WM tract.

COMPARISON OF PROPOSED METHOD WITH ALTERNATIVE TRACTOMETRY METHODOLOGY

Six WM tracts overlapped the six selected clusters from the primary analysis (four where control FA > CLBP FA and two where CLBP FA > control FA). Tract profiles for each of these six tracts are visualized per group averaged for each visit in Figure 6. Of the six tracts investigated, only the Left Optic Radiation had segments that showed statistically significant differences in FA between the control and CLBP groups across all three visits ($p<0.05$, uncor-

Table 3. Group average fractional anisotropy reported separately per visit.

		Control Mean FA ± Std	CLBP Mean FA ± Std	Cohen's D[a]	P-Value[b]
Visit 1 (Control N=24, CLBP N=27)					
Control > CLBP	Cluster 1	0.48 ± 0.07	0.41 ± 0.07	1.022	> 0.001
	Cluster 2	0.42 ± 0.07	0.36 ± 0.07	0.830	0.004
	Cluster 3	0.37 ± 0.05	0.32 ± 0.07	0.908	0.002
	Cluster 4	0.47 ± 0.05	0.41 ± 0.07	0.962	0.001
	Cluster 5	0.43 ± 0.07	0.36 ± 0.06	1.076	> 0.001
	Cluster 6	0.45 ± 0.06	0.38 ± 0.06	1.120	> 0.001
	Cluster 7	0.59 ± 0.07	0.54 ± 0.05	0.914	0.002
CLBP > Control	Cluster A	0.44 ± 0.04	0.48 ± 0.06	-0.768	0.008
	Cluster B	0.51 ± 0.04	0.56 ± 0.06	-0.841	0.004
	Cluster C	0.24 ± 0.08	0.3 ± 0.05	-0.914	0.003
	Cluster D	0.51 ± 0.05	0.56 ± 0.05	-1.028	> 0.001
Visit 2 (Control N=25, CLBP N=24)					
Control > CLBP	Cluster 1	0.47 ± 0.07	0.41 ± 0.07	0.869	0.004
	Cluster 2	0.42 ± 0.06	0.35 ± 0.06	1.183	> 0.001
	Cluster 3	0.37 ± 0.05	0.32 ± 0.07	0.807	0.008
	Cluster 4	0.47 ± 0.05	0.4 ± 0.05	1.174	> 0.001
	Cluster 5	0.42 ± 0.06	0.36 ± 0.07	0.968	0.001
	Cluster 6	0.44 ± 0.06	0.38 ± 0.08	0.778	0.01
	Cluster 7	0.59 ± 0.06	0.54 ± 0.06	0.905	0.003
CLBP > Control	Cluster A	0.44 ± 0.04	0.47 ± 0.06	-0.733	0.015
	Cluster B	0.51 ± 0.04	0.55 ± 0.06	-0.788	0.009
	Cluster C	0.25 ± 0.08	0.3 ± 0.05	-0.706	0.017
	Cluster D	0.5 ± 0.05	0.55 ± 0.06	-0.803	0.008
Visit 3 (Control N=25, CLBP N=23)					
Control > CLBP	Cluster 1	0.47 ± 0.07	0.41 ± 0.07	0.921	0.003
	Cluster 2	0.41 ± 0.06	0.34 ± 0.07	1.182	> 0.001
	Cluster 3	0.36 ± 0.05	0.31 ± 0.07	0.876	0.005
	Cluster 4	0.47 ± 0.05	0.4 ± 0.06	1.103	> 0.001
	Cluster 5	0.41 ± 0.06	0.35 ± 0.06	0.981	0.001
	Cluster 6	0.45 ± 0.07	0.37 ± 0.07	1.043	> 0.001
	Cluster 7	0.59 ± 0.06	0.54 ± 0.06	0.869	0.004
CLBP > Control	Cluster A	0.44 ± 0.05	0.48 ± 0.06	-0.724	0.017
	Cluster B	0.51 ± 0.04	0.54 ± 0.06	-0.727	0.017
	Cluster C	0.25 ± 0.07	0.3 ± 0.06	-0.799	0.008
	Cluster D	0.51 ± 0.04	0.55 ± 0.06	-0.841	0.006

[a]Cohen's D calculated as control FA mean minus CLBP FA mean divided by pooled CLBP and HC standard deviation (i.e. negative value indicates CLBP > control mean)
[b]Group differences assessed with a two sample t-test at $p < 0.05$, uncorrected for multiple comparisons

rected). While other tracts displayed segments with significant group differences, these findings were less consistent across the three time points. In regions where statistically significant differences were observed for two or less visits, all three visits still had mean differences between groups larger than standard deviations (See zoomed in regions for the right optic radiation in Figure 6 panels E.1 and E.2) reflecting the small statistical power related to the small sample size but consistent trend level group differences across all visits. There was a strong spatial correspondence between the two methodologies in regions where the TBSS repeatability analysis yielded a statistically different group difference (highlighted 7 light red regions Figure 6). Among the tracts investigated, 7 regions overlapped retained TBSS repeatability analysis clusters, of which 5 had adjacent segments with statistically significant tractometry group differences. For study robustness, a supplementary tractometry analysis was conducted investigating the remaining 5 clusters that were more sparsely distributed across the brain, overlapping 9 tracts from the WM bundle template

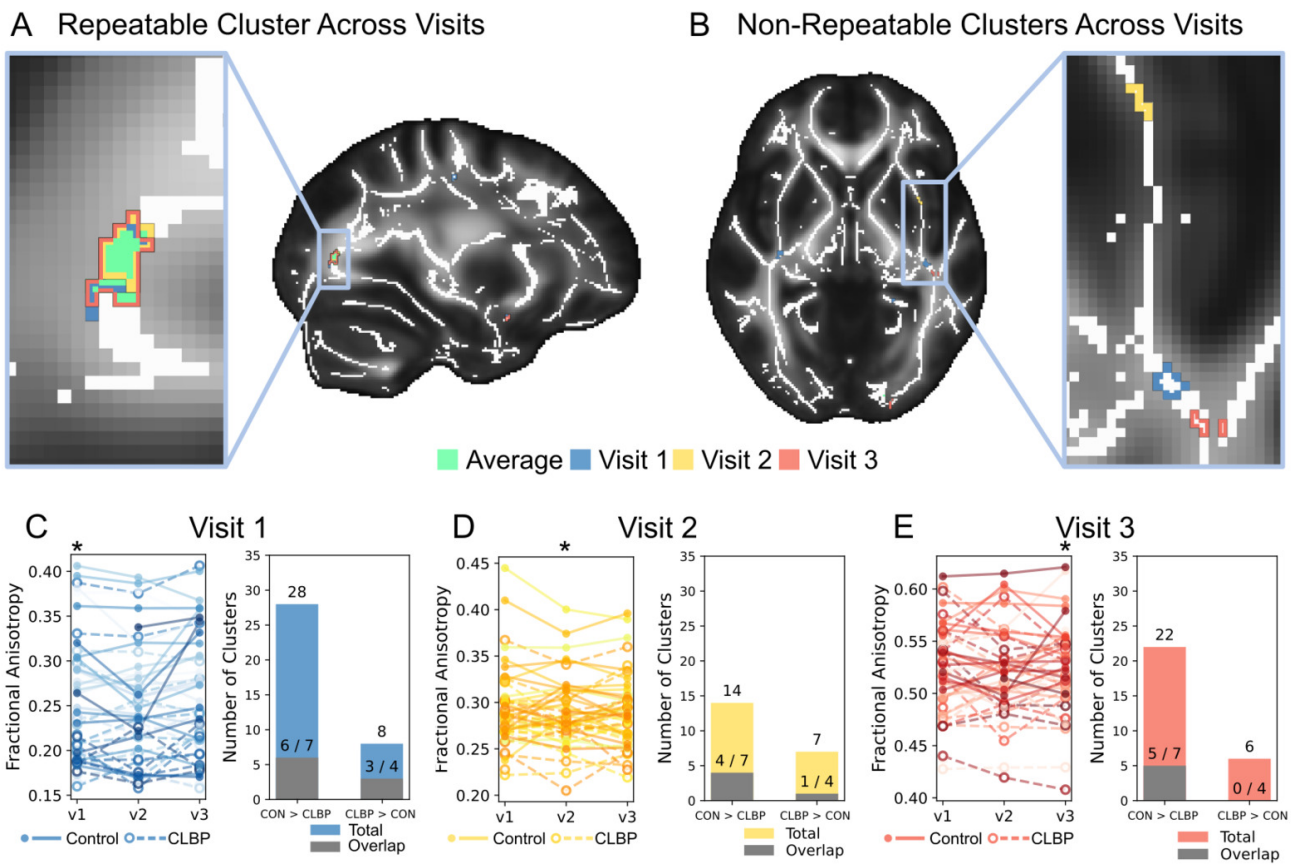


Figure 4. A) An example region where fractional anisotropy was observed to be higher in controls compared to chronic low back pain subjects at all visits. The cluster extracted with the proposed repeatability analysis is shown in green with overlapping clusters for visit 1 (blue), visit 2 (yellow), and visit 3 (red) overlaid. B) Shows an example of non-repeatable regions where fractional anisotropy differences were observed using t-statistics at single visits but there was no spatial agreement among visits. For each non-repeatable region in B, mean fractional anisotropy is plotted for each subject per visit (C left, D left, E left). Bar charts indicating the number of clusters detected across the entire brain for each visit (C right, D right, E right) are shown alongside the total number of clusters at each visit overlapping the 11 clusters output by the proposed repeatability analysis (grey, 7 control FA greater than CLBP FA, 4 CLBP FA greater than control FA). Although there are numerous clusters above threshold (voxel size greater than 10, t-statistic > |2.5|) using each visit separately, the majority of these clusters are chance findings that are not repeatable across visits.

(Figure S4). For these clusters and compared to the tractometry analysis of the first 6 clusters, the spatial correspondence between TBSS and tractometry group differences was less consistent, with only 3 tract segments having consistent TBSS and tractometry group differences among the 20 tract segments that overlapped the TBSS clusters.

DISCUSSION

CONSISTENT REGIONAL ALTERATIONS IN WHITE MATTER DTI METRICS IN CHRONIC LOW BACK PAIN

The lower FA associated with CLBP and observed in 7 regions in the current study is in agreement with previous dMRI studies reporting similarly lower FA in CLBP relative to controls.[6,16,17] Four regions were found to have higher FA in CLBP relative to controls; higher FA may also be associated with CLBP. In specific contexts, lower FA may indicate demyelination as in multiple sclerosis[49] but altered FA values and dMRI derived metrics should be interpreted in the context of specific diseases and could have many underlying microstructural explanations.[50] Thus, given the current state of the neuroimaging CP field, it is unclear the specific microstructural explanations for such alterations of FA in CLBP. Findings from one predictive study has suggested that lower FA may indicate a transition from an acute to a chronic pain condition, suggesting the alterations in WM may exist prior to the onset of chronic pain symptomatology and with these differences measurable up to at least one year later.[6] Here, the WM microstructural alterations associated with CLBP were measurable at all three visits spanning 4 months, with patients presenting pain symptoms for at least 4 months prior to their first visit, supporting the finding that microstructural changes to WM are present across a long time frame after the chronification of pain symptoms.

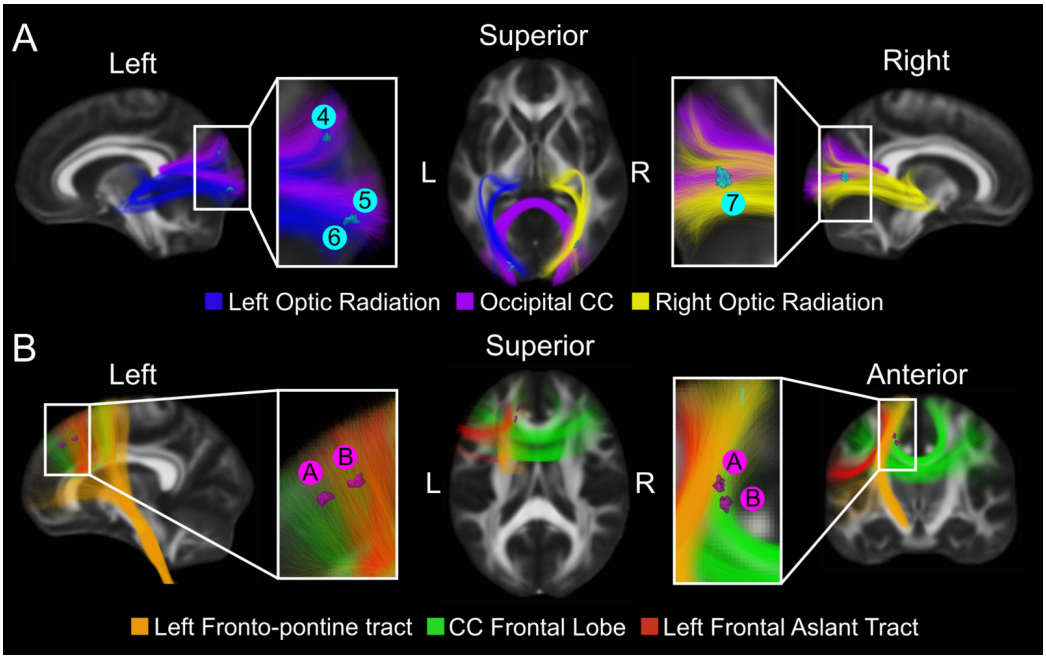


Figure 5. Visualization of select regions output from the repeatability analysis where a fractional anisotropy group difference was observed. A) 4 clusters where control fractional anisotropy was higher than chronic low back pain fractional anisotropy are displayed (cyan: 4, 5, 6, 7) and overlaid on adjacent white matter tracts (blue: left optic radiation; purple: occipital corpus callosum; yellow: right optic radiation). B) 2 clusters where chronic low back pain fractional anisotropy was greater than control fractional anisotropy are displayed (pink: A, B) and overlaid on adjacent white matter tracts (orange: left fronto-pontine tract; green: corpus callosum frontal posterior portion; red; left frontal aslant tract). Although regions overlap major white matter tracts, their proximity in fiber crossing regions impedes the exact localization of these regions to individual white matter tracts.

ANATOMICAL LOCALIZATION OF WHITE MATTER ALTERATIONS IN CLBP

Repeatable decreases in FA associated with CLBP were found primarily in the occipital lobe (4 of 7 clusters) which has not been typically reported in CLBP but found in other chronic pain conditions such as in cluster headaches.[51,52] The two lower FA clusters found in the current study in the superior parietal regions were located adjacent to motor regions agreeing with lower FA found in the primary somatosensory cortex WM[16] and supports the literature suggesting alterations to primary motor networks in these conditions.[53] While no differences were found within the internal and external capsule WM regions as in other studies,[6,18] higher and lower FA associated with CLBP were found in multiple regions along tracts that traverse these regions.

Isolating voxel-wise differences to singular WM tracts was difficult because regional differences were found in areas adjacent to the cortex where multiple tracts traverse. For example, in Figure 5, the regions with lower FA in CLBP were located along the corpus callosum and left/right optic radiations, demonstrating the difficulty in interpreting these regional differences as microstructural alterations along specific tracts, as others have reported when using TBSS analysis.[6,17] Importantly, the 6 regions located in the occipital lobe (4 clusters control > CLBP) and frontal lobe (2 clusters CLBP < control) with altered FA reported from TBSS analysis in the current study agreed substantially with tract segments found with altered FA extracted from tractometry, albeit often without agreement between all 3 visits. This is an encouraging finding given the differences between how FA is measured between the two methods; TBSS projects the maximum FA value from adjacent voxels to the skeleton; tractometry averages across all voxel values overlapping a segment. A likely explanation for the small discrepancies between methods is that tractometry is an average over large WM regions, whereas TBSS is a more isolated WM measurement, and highlights the importance when interpreting results extracted from differential methodology between studies. In supplementary analysis (Figure S4) the remaining TBSS clusters were examined for consistent findings via tractometry, with these clusters being notably smaller and more sparsely distributed throughout the brain. In this case, the agreement between TBSS and tractometry were markedly less consistent than the previously analyzed regions suggesting that the likelihood of between method agreement is increased when more densely located clusters or larger regions are observed with TBSS.

PROPOSED REPEATABILITY ANALYSIS

Repeatability in neuroimaging literature is typically reported as intra-class correlation (ICC) values and are analogous to the I2C2 values (an extension of ICC for data acquired via imaging) reported in the current study. In previous works, ICC for FA values have been reported in the good-excellent range when assessing test-retest,[23,26,27]

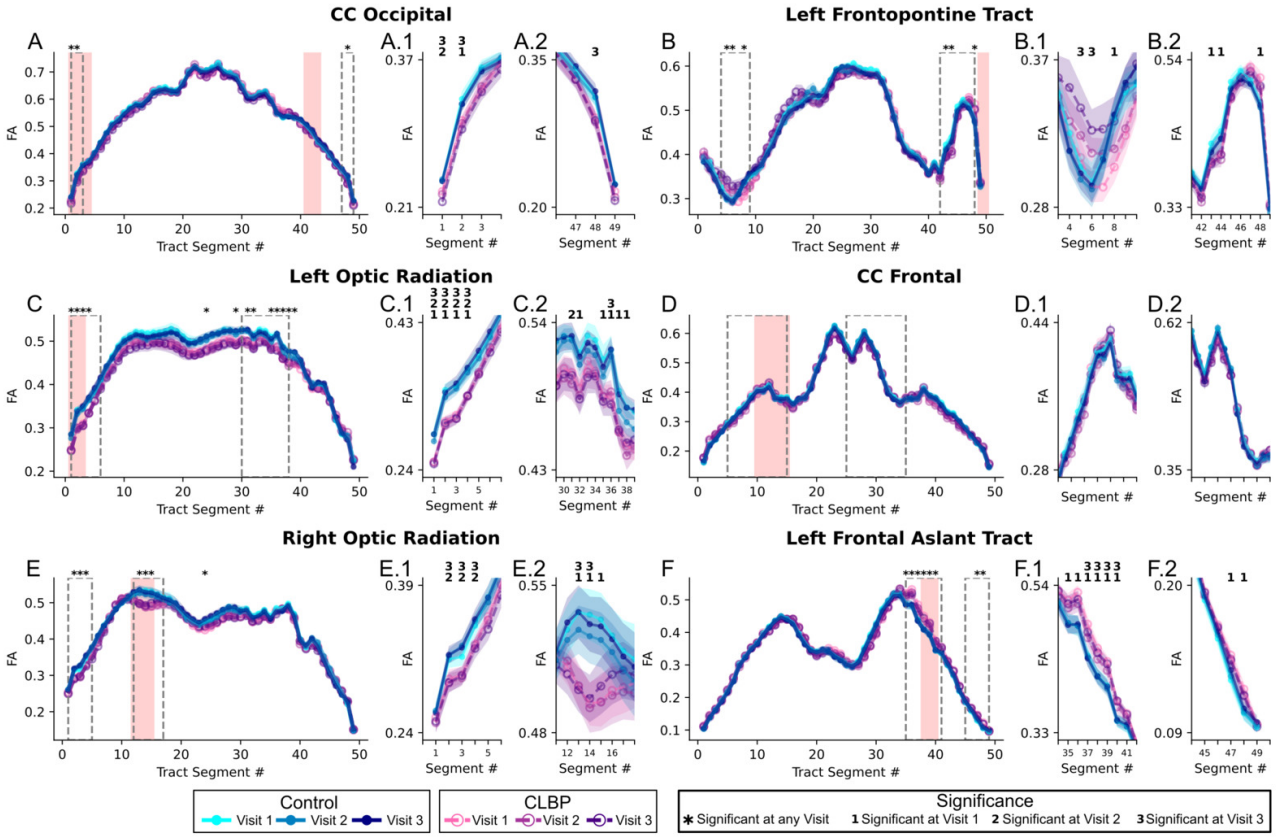


Figure 6. Tractometry profiles of fractional anisotropy for six selected white matter tracts that overlapped select regions output from the TBSS repeatability analysis (Clusters: 4, 5, 6, 7, A, B); A) Occipital Portion of the Corpus Callosum (Clusters: 4, 5, 6, 7), B) Left Frontopontine Tract (Clusters: A, B), C) Left Optic Radiation (Clusters: 5, 6), D) Frontal Portion of the Corpus Callosum (Clusters: A, B) E) Right Optic Radiation (Cluster 7), F) Left Frontal Aslant Tract (Cluster B). Average fractional anisotropy values are shown for 50 tract segments for controls (cyan to dark blue) and CLBP (pink to dark purple) with zoomed regions depicted in subplots (e.g., dotted grey boxes in A, zoomed regions in A.1, A.2). For each group and visit standard deviations are displayed with transparency in the associated colour. Segments that overlap clusters identified in the proposed TBSS repeatability analysis are highlighted (light red). Regions with a statistically significant difference (two sample t-test, $p < 0.05$, uncorrected) for at least one of three visits are identified (*), and in the zoomed plots the specific visit with a group difference is identified with a number indicating which visit showed the significant difference.

cross-scanner data with denoising having a positive effect on repeatability.[21] This current study specifically aimed to find regions that had repeatable measurements in both controls and CLBP patients, having I2C2 values in line with previous work (I2C2 values of 0.7685 - 0.8821) in the current study. This enabled the removal of regions where CLBP vs control group differences were observed at a singular visit but not at all three time points. For example, at visit 1, 26 regions surpassed cluster size and statistics thresholds with FA values lower in CLBP patients compared to controls; however, while using the proposed repeatability analysis only 7 such regions were retained. While it is notable that these repeated measures design comes with the added cost, effort and time needed to repeat the entire imaging acquisition, the technique has a substantial benefit to reduce the reporting of chance findings that can arise solely by measurement fluctuations when only single time points are considered.

LIMITATIONS AND FUTURE DIRECTIONS

Even though the results are extremely repeatable, the study is still limited by a small sample size (25 controls, 27 CLBP) making it difficult to generalize these results to the entire CLBP population at large. This is a common limitation observed in the chronic pain dMRI literature with similar sized cohorts being used previously.[6,15,17,18] The small sample size also likely impeded full inter-method agreement between the TBSS and tractometry analyses; while clear, trend-level group differences were often observed across all visits in the same regions for both methods, only a subset of these differences reached statistical significance, likely reflecting low statistical power.

Whole-brain neuroimaging analyses involve testing thousands of voxels simultaneously, requiring correction for multiple comparisons to control false positives. Many of these correction frameworks (e.g., Gaussian random field theory and cluster-based inference) were originally devel-

oped and extensively validated in the context of task-based fMRI, where assumptions regarding spatial smoothness, signal distribution, and temporal modelling are relatively well characterized.[54-56] Structural MRI modalities such as diffusion imaging present different statistical properties, including non-Gaussian signal distributions,[57] modality-specific spatial covariance,[58] and dependence on preprocessing steps such as tensor fitting or skeletonization,[32] which may challenge some of these assumptions.

In addition, when sample sizes are modest, stringent voxel-wise corrections can substantially reduce statistical power, reflecting the well-described trade-off between sensitivity and control of false positives in mass-univariate neuroimaging analyses,[55] a limitation that is particularly relevant in whole-brain analyses. In the present study, each group included approximately 25 participants, but each participant was scanned at three time points across four months. In this context, applying highly conservative correction thresholds may obscure biologically meaningful effects. Therefore, in addition to conventional statistical testing, we examined spatial patterns of effects using uncorrected maps and evaluated whether these patterns were consistently observed across all three visits. While this approach does not replace formal multiple-comparisons correction, the replication of spatially consistent effects across independent time points provides complementary evidence supporting the robustness of the findings, as reproducibility across repeated measurements is known to improve the reliability of neuroimaging metrics.[59-61] Importantly, such within-cohort reproducibility is aligned with the broader objective of advancing neuroimaging toward precision medicine, where identifying stable and biologically meaningful signals at the individual or small-cohort level is challenging, but essential for the development of clinically relevant biomarkers.[10,62,63]

A further limitation, inherent to the resolution of dMRI and tensor-based analysis, is the difficulty in definitively localizing observed effects to a single WM tract, especially in regions with complex fiber crossings. Future work could address this by applying more complex modelling techniques that consider fiber crossings, such as fixel-based analysis,[64] to disentangle the contributions of individual fiber populations within a voxel and pinpoint the specific tracts driving the observed diffusion differences. We acknowledge that the repeated-measures design involving three separate time points is costly and difficult to implement. However, now that highly repeatable clusters have been identified, a future direction could be to use these locations as a priori regions of interest. Validating these specific regions in a new, larger cohort may only require one or two scans, leveraging the knowledge gained from this longitudinal design to create more efficient and targeted studies moving forward.

CONCLUSION

In this study, we introduce a longitudinal dMRI study design and analysis that prioritizes the repeatability of findings to investigate WM microstructural alterations in CLBP. Here, 11 regions were identified with differences in FA between CLBP patients and healthy controls that were consistent across 3 imaging sessions spanning four months. Lower FA in CLBP was found in 7 regions whereas higher FA was found in 4 regions. The primary contribution of this work is the demonstration that incorporating repeatability metrics directly into the analysis framework can substantially increase the reliability of results. Standard single-visit analyses revealed many significant clusters that were not stable across time, suggesting a high rate of chance findings. Our method effectively filters these spurious results by retaining only those regions that show both a significant group difference and high test-retest reliability. In future work, this cohort will be expanded to a larger, more representative sample, which will allow for more robust statistical testing and better generalization of these findings to the broader CLBP population. By leveraging longitudinal data to filter for repeatability, this methodology offers a path toward reliable and clinically meaningful biomarkers in chronic pain and other neurological conditions facing similar challenges of reproducibility.

DATA AND CODE AVAILABILITY

Data will be made accessible on openpain.org in 2026.

ETHICAL STATEMENT

This study was approved by the ethics review board of the Centre intégré universitaire de santé et de services sociaux de l'Estrie – Centre hospitalier universitaire de Sherbrooke (CIUSSS de l'Estrie – CHUS), Sherbrooke, Canada (file number: 2021-3861).

FUNDING SOURCES

The authors declare the following funding sources: Fonds de recherche du Québec (#311042; 10.13039/501100020951); Arthritis Society (#21-0000000041; 10.13039/501100000142), Réseau québécois de recherche sur la douleur (10.13039/501100017033), and Réseau de Bio-Imagerie du Quebec (10.13039/100010571) to Pascal Tétreault; Canadian Institutes of Health Research (10.13039/501100000024) to Marc-Antoine Fortier and Fonds de recherche du Québec (10.13039/501100020951) to Graham Little.

AI USE DISCLOSURE

Large language models were used to assist in the software development for this study (GitHub Autopilot, Chat GPT 4.1) with all output reviewed by an expert analyst (GL).

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

Submitted: December 18, 2025 CDT. Revised: April 02, 2026 CDT. Accepted: April 24, 2026 CDT. Published: May 28, 2026 CDT.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.

REFERENCES

1. Baliki MN, Apkarian AV. Nociception, Pain, Negative Moods, and Behavior Selection. *Neuron*. 2015;87(3):474-491. doi:10.1016/j.neuron.2015.06.005

2. Kregel J, Meeus M, Malfliet A, et al. Structural and functional brain abnormalities in chronic low back pain: A systematic review☆. *Seminars in Arthritis and Rheumatism*. 2015;45(2):229-237. doi:10.1016/j.semarthrit.2015.05.002

3. van der Miesen MM, Lindquist MA, Wager TD. Neuroimaging-based biomarkers for pain: State of the field and current directions. *PAIN Reports*. 2019;4(4):e751. doi:10.1097/PR9.0000000000000751

4. Martucci KT, Mackey SC. Neuroimaging of Pain: Human Evidence and Clinical Relevance of Central Nervous System Processes and Modulation. *Anesthesiology*. 2018;128(6):1241-1254. doi:10.1097/ALN.0000000000002137

5. Baliki MN, Petre B, Torbey S, et al. Corticostriatal functional connectivity predicts transition to chronic back pain. *Nature Neuroscience*. 2012;15(8):1117-1119. doi:10.1038/nn.3153

6. Mansour AR, Baliki MN, Huang L, et al. Brain white matter structural properties predict transition to chronic pain. *Pain*. 2013;154(10):2160-2168. doi:10.1016/j.pain.2013.06.044

7. Seminowicz DA, Wideman TH, Naso L, et al. Effective treatment of chronic low back pain in humans reverses abnormal brain anatomy and function. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 2011;31(20):7540-7550. doi:10.1523/JNEUROSCI.5280-10.2011

8. Tétreault P, Baliki MN, Baria AT, Bauer WR, Schnitzer TJ, Apkarian AV. Inferring distinct mechanisms in the absence of subjective differences: Placebo and centrally acting analgesic underlie unique brain adaptations. *Human Brain Mapping*. 2018;39(5):2210-2223. doi:10.1002/hbm.23999

9. Davis KD, Aghaeepour N, Ahn AH, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: Challenges and opportunities. *Nature Reviews Neurology*. 2020;16(7). doi:10.1038/s41582-020-0362-2

10. Mouraux A, Iannetti GD. The search for pain biomarkers in the human brain. *Brain*. 2018;141(12):3290-3307. doi:10.1093/brain/awy281

11. Sluka KA, Wager TD, Sutherland SP, et al. Predicting chronic postsurgical pain: Current evidence and a novel program to develop predictive biomarker signatures. *PAIN*. 2023;164(9):1912. doi:10.1097/j.pain.0000000000002938

12. Tétreault P, Mansour A, Vachon-Presseau E, Schnitzer TJ, Apkarian AV, Baliki MN. Brain Connectivity Predicts Placebo Response across Chronic Pain Clinical Trials. *PLOS Biology*. 2016;14(10):e1002570. doi:10.1371/journal.pbio.1002570

13. Reckziegel D, Tétreault P, Ghantous M, et al. Sex-Specific Pharmacotherapy for Back Pain: A Proof-of-Concept Randomized Trial. *Pain and Therapy*. 2021;10(2):1375-1400. doi:10.1007/s40122-021-00297-2

14. Beaulieu C. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR in Biomedicine*. 2002;15(7–8):435-455. doi:10.1002/nbm.782

15. Bautin P, Fortier MA, Sean M, et al. What has brain diffusion magnetic resonance imaging taught us about chronic primary pain: A narrative review. *PAIN*. 2025;166(2):243. doi:10.1097/j.pain.0000000000003345

16. Kim H, Mawla I, Lee J, et al. Reduced tactile acuity in chronic low back pain is linked with structural neuroplasticity in primary somatosensory cortex and is modulated by acupuncture therapy. *NeuroImage*. 2020;217:116899. doi:10.1016/j.neuroimage.2020.116899

17. Ma J, Wang X, Qiu Q, Zhan H, Wu W. Changes in Empathy in Patients With Chronic Low Back Pain: A Structural-Functional Magnetic Resonance Imaging Study. *Frontiers in Human Neuroscience*. 2020;14:326. doi:10.3389/fnhum.2020.00326

18. Čeko M, Shir Y, Ouellet JA, Ware MA, Stone LS, Seminowicz DA. Partial recovery of abnormal insula and dorsolateral prefrontal connectivity to cognitive networks in chronic low back pain after treatment. *Human Brain Mapping*. 2015;36(6):2075-2092. doi:10.1002/hbm.22757

19. Vachon-Presseau E, Tétreault P, Petre B, et al. Corticolimbic anatomical characteristics predetermine risk for chronic pain. *Brain: A Journal of Neurology*. 2016;139(Pt 7):1958-1970. doi:10.1093/brain/aww100

20. Botvinik-Nezer R, Wager TD. Reproducibility in Neuroimaging Analysis: Challenges and Solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2023;8(8):780-788. doi:10.1016/j.bpsc.2022.12.006

21. Ades-Aron B, Coelho S, Lemberskiy G, et al. Denoising Improves Cross-Scanner and Cross-Protocol Test–Retest Reproducibility of Diffusion Tensor and Kurtosis Imaging. *Human Brain Mapping*. 2025;46(4):e70142. doi:10.1002/hbm.70142

22. Boekel W, Forstmann BU, Keuken MC. A test-retest reliability analysis of diffusion measures of white matter tracts relevant for cognitive control. *Psychophysiology*. 2017;54(1):24-33. doi:10.1111/psyp.12769

23. Duan F, Zhao T, He Y, Shu N. Test–retest reliability of diffusion measures in cerebral white matter: A multiband diffusion MRI study. *Journal of Magnetic Resonance Imaging*. 2015;42(4):1106-1116. doi:10.1002/jmri.24859

24. Luque Laguna PA, Combes AJE, Streffer J, et al. Reproducibility, reliability and variability of FA and MD in the older healthy population: A test-retest multiparametric analysis. *NeuroImage: Clinical*. 2020;26:102168. doi:10.1016/j.nicl.2020.102168

25. Merisaari H, Tuulari JJ, Karlsson L, et al. Test-retest reliability of Diffusion Tensor Imaging metrics in neonates. *NeuroImage*. 2019;197:598-607. doi:10.1016/j.neuroimage.2019.04.067

26. Rosberg A, Tuulari JJ, Kumpulainen V, et al. Test-retest reliability of diffusion tensor imaging scalars in 5-year-olds. *Human Brain Mapping*. 2022;43(16):4984-4994. doi:10.1002/hbm.26064

27. Cole JH, Farmer RE, Rees EM, et al. Test-Retest Reliability of Diffusion Tensor Imaging in Huntington's Disease. *PloS Currents*. 2014;6. doi:10.1371/currents.hd.f19ef63fff962f5cd9c0e88f4844f43b

28. Cousineau M, Jodoin PM, Garyfallidis E, et al. A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical*. 2017;16:222-233. doi:10.1016/j.nicl.2017.07.020

29. Keihaninejad S, Zhang H, Ryan NS, et al. An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's disease. *NeuroImage*. 2013;72:153-163. doi:10.1016/j.neuroimage.2013.01.044

30. Newman BT, Dhollander T, Reynier KA, Panzer MB, Druzgal TJ. Test–retest reliability and long-term stability of three-tissue constrained spherical deconvolution methods for analyzing diffusion MRI data. *Magnetic Resonance in Medicine*. 2020;84(4):2161-2173. doi:10.1002/mrm.28242

31. Farrell JAD, Landman BA, Jones CK, et al. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *Journal of Magnetic Resonance Imaging: JMRI*. 2007;26(3):756-767. doi:10.1002/jmri.21053

32. Jones DK, Cercignani M. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR in Biomedicine*. 2010;23(7):803-820. doi:10.1002/nbm.1543

33. Shahim P, Holleran L, Kim JH, et al. Test-retest reliability of high spatial resolution diffusion tensor and diffusion kurtosis imaging. *Sci Rep*. 2017;7:11141. doi:10.1038/s41598-017-11747-3

34. Zuo XN, Xing XX. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews*. 2014;45:100-118. doi:10.1016/j.neubiorev.2014.05.009

35. Poquet N, Lin C. The Brief Pain Inventory (BPI). *Journal of Physiotherapy*. 2016;62(1):52. doi:10.1016/j.jphys.2015.07.001

36. Sean M, Coulombe-Lévêque A, Nadeau W, et al. Counting your chickens before they hatch: Improvements in an untreated chronic pain population, beyond regression to the mean and the placebo effect. *Pain Reports*. 2024;9(3):e1157. doi:10.1097/PR9.0000000000001157. PMID:38689593

37. Sean M, Cote S, Coulombe-Lévêque A, et al. Distinct Grey Matter Alterations in Untreated Mild to Moderate Chronic Low Back Pain: A Longitudinal Magnetic Resonance Imaging Study. *European Journal of Pain*. 2026;30(4):e70272. doi:10.1002/ejp.70272

38. Theaud G, Houde JC, Boré A, Rheault F, Morency F, Descoteaux M. TractoFlow: A robust, efficient and reproducible diffusion MRI pipeline leveraging Nextflow & Singularity. *NeuroImage*. 2020;218:116889. doi:10.1016/j.neuroimage.2020.116889

39. Veraart J, Novikov DS, Christiaens D, Ades-Aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *NeuroImage*. 2016;142:394-406. doi:10.1016/j.neuroimage.2016.08.016

40. Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*. 2003;20(2):870-888. doi:10.1016/S1053-8119(03)00336-7

41. Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*. 2016;125:1063-1078. doi:10.1016/j.neuroimage.2015.10.019

42. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. 2008;12(1):26-41. doi:10.1016/j.media.2007.06.004

43. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*. 2001;20(1):45-57. doi:10.1109/42.906424

44. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908

45. Smith SM, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*. 2006;31(4):1487-1505. doi:10.1016/j.neuroimage.2006.02.024

46. Shou H, Eloyan A, Lee S, et al. Quantifying the reliability of image replication studies: The image intra-class correlation coefficient (I2C2). *Cognitive, Affective & Behavioral Neuroscience*. 2013;13(4):714-724. doi:10.3758/s13415-013-0196-0

47. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

48. St-Onge E, Schilling K, Rheault F. BundleSeg: A versatile, reliable and reproducible approach to white matter bundle segmentation. Published online 2023. doi:10.1007/978-3-031-47292-3_5

49. Inglese M, Bester M. Diffusion imaging in multiple sclerosis: Research and clinical implications. *NMR in Biomedicine*. 2010;23(7):865-872. doi:10.1002/nbm.1515

50. Horsfield MA, Jones DK. Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases – a review. *NMR in Biomedicine*. 2002;15(7–8):570-577. doi:10.1002/nbm.787

51. Szabó N, Kincses ZT, Párdutz Á, et al. White matter disintegration in cluster headache. *The Journal of Headache and Pain*. 2013;14(1):64. doi:10.1186/1129-2377-14-64

52. Teepker M, Menzler K, Belke M, et al. Diffusion Tensor Imaging in Episodic Cluster Headache. *Headache: The Journal of Head and Face Pain*. 2012;52(2):274-282. doi:10.1111/j.1526-4610.2011.02000.x

53. Chang WJ, O'Connell NE, Beckenkamp PR, Alhassani G, Liston MB, Schabrun SM. Altered Primary Motor Cortex Structure, Organization, and Function in Chronic Pain: A Systematic Review and Meta-Analysis. *The Journal of Pain*. 2018;19(4):341-359. doi:10.1016/j.jpain.2017.10.007

54. Nichols TE. Multiple testing corrections, nonparametric methods, and random field theory. *NeuroImage*. 2012;62(2):811-815. doi:10.1016/j.neuroimage.2012.04.014

55. Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*. 2003;12(5):419-446. doi:10.1191/0962280203sm341ra

56. Smith SM, Nichols TE. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. 2009;44(1):83-98. doi:10.1016/j.neuroimage.2008.03.061

57. Le Bihan D, Iima M. Diffusion Magnetic Resonance Imaging: What Water Tells Us about Biological Tissues. *PLoS Biology*. 2015;13(7):e1002203. doi:10.1371/journal.pbio.1002203

58. Alexander-Bloch A, Giedd JN, Bullmore E. Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience*. 2013;14(5):322-336. doi:10.1038/nrn3465

59. Bennett CM, Miller MB. How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*. 2010;1191(1):133-155. doi:10.1111/j.1749-6632.2010.05446.x

60. Danielian LE, Iwata NK, Thomasson DM, Floeter MK. Reliability of Fiber Tracking Measurements in Diffusion Tensor Imaging for Longitudinal Study. *NeuroImage*. 2010;49(2):1572-1580. doi:10.1016/j.neuroimage.2009.08.062

61. Heiervang E, Behrens TEJ, Mackay CE, Robson MD, Johansen-Berg H. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *NeuroImage*. 2006;33(3):867-877. doi:10.1016/j.neuroimage.2006.07.037

62. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single Subject Prediction of Brain Disorders in Neuroimaging: Promises and Pitfalls. *NeuroImage*. 2017;145(Pt B):137-165. doi:10.1016/j.neuroimage.2016.02.079

63. Intensive longitudinal brain imaging reveals personalized signatures of chronic pain. *Nature Neuroscience*. Published online 2026:1-2. doi:10.1038/s41593-026-02238-8

64. Dhollander T, Clemente A, Singh M, et al. Fixel-based Analysis of Diffusion MRI: Methods, Applications, Challenges and Opportunities. *NeuroImage*. 2021;241:118417. doi:10.1016/j.neuroimage.2021.118417

SUPPLEMENTARY MATERIALS

Supplementary Materials

Download: https://apertureneuro.org/article/161956-towards-repeatable-and-converging-methods-in-diffusion-mri-evidence-from-a-longitudinal-chronic-pain-cohort/attachment/343835.docx