PLOS ONE

Exploring the relationship between white matter integrity, cocaine use and GAD polymorphisms using Bayesian Model Averaging

Tmader Alballa[1,4]*, Edward L. Boone[1,3], Liangsuo Ma[2,3], Andrew Snyder[7], F. Gerard Moeller[3,5,6,7]

1 Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, Virginia, United States of America, **2** Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, United States of America, **3** Institute of Drug and Alcohol Studies, Virginia Commonwealth University, Richmond, Virginia, United States of America, **4** Mathematical Sciences Department, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, **5** Department of Pharmacology and Toxicology, Virginia Commonwealth University, Richmond, Virginia, United States of America, **6** Department of Neurology, Virginia Commonwealth University, Richmond, Virginia, United States of America, **7** C. Kenneth and Dianne Wright, Center for Clinical Translational Research, Virginia Commonwealth University, Richmond, Virginia, United States of America

* tmader.alballa@gmail.com

Abstract

Past investigations utilizing diffusion tensor imaging (DTI) have demonstrated that cocaine use disorder (CUD) yields white matter changes, primarily in the corpus callosum. By applying Bayesian model averaging using multiple linear regression in DTI, we demonstrate there may exist relationships between the impaired white matter and glutamic acid decarboxylase (*GAD*) polymorphisms. This work explored the two-way and three-way interactions between *GAD*1^a (SNP: rs1978340) and *GAD*1^b (SNP: rs769390) polymorphisms and years of cocaine use (*YCU*). *GAD*1^a was associated with more frontal white matter changes on its own but *GAD*1^b was associated with more midbrain and cerebellar changes as well as a greater increase in white matter changes in the context of chronic cocaine use. The three-way interaction *GAD*1^a|*GAD*1^b|*YCU* appeared to be roughly an average of the polymorphism two-way interactions *GAD*1^a|*YCU* and *GAD*1^b|*YCU*. The three-way interaction demonstrated multiple regions including corpus callosum which featured fewer significant voxel changes, perhaps suggesting a small protective effect of having both polymorphisms on corpus callosum and cerebellar peduncle.

Introduction

Cocaine use disorder (CUD) patients exhibit altered white matter integrity versus non-drug using controls [1–7], with the most consistent findings in the corpus callosum. The altered white matter integrity is particularly evident in CUD patients with greater severity of cocaine

use [4]. Moreover, the changes in the white matter have also been associated with the length of cocaine use [8]. Diffusion tensor imaging (DTI) is a non-invasive magnetic resonance imaging (MRI) technology suitable for characterizing microstructural changes in human neuropathology; DTI measures restricted diffusion of water due to cellular membranes and organelles that comprise the structural integrity of brain [9]. Fractional anisotropy (FA), the most commonly used DTI measure, describes the macroscopic axonal organization in brain with lower FA generally suggesting reduced white matter integrity [1, 10]. Consistent with these observations, preclinical studies have shown that lower FA is observed in rats exposed to a chronic regimen of cocaine [11, 12], supporting the hypothesis that exposure to cocaine may underlie the lower FA seen in human CUD patients. However, not all CUD patients show reduced FA compared to healthy controls (e.g., [13]), and few studies to date have examined factors that could enhance or reduce the impact of cocaine on white matter integrity.

However, a previous study [14] tested twenty-one genetic variants across seventeen candidate genes that may be related to cocaine addiction, where Bayesian Model Averaging with linear regression was used to explore the effects of the cocaine use on these genetic variants. Their DTI findings were related to two Single Nucleotide Polymorphisms (SNPs) $GAD1^a$ and $GAD1^b$. That previous study [14] finds that comparing the impact of $GAD1^a$ to the other 20 genetic variants on diffusion in the white matter of the brain was comprehensive with approximately 4 times more in the number of significant voxels compared to the second most effective genetic variant. However, that study only looked at single genes without looking at any interactions between SNP's and years of cocaine use in the model. The purpose of this study is to further explore how the $GAD1^a$ and $GAD1^b$ SNPs in human brain cells interact with human sex, years of cocaine use, and white matter changes as measured by FA in DTI using a Bayesian Model Averaging [15] whit linear regression model [16] by using exactly the same data-set that have been used in the previous work [14]. BMA is one of the most effective strategies that deals with model uncertainty, which takes into account all possible models. The novel contribution of this study is the demonstration of a series of two-way $GAD1^a|YCU$ / $GAD1^b|YCU$ and three-way $GAD1^a|GAD1^b|YCU$ models which enable us to assess the individual contributions of each covariate in determining white matter changes in CUD.

Study population

Informed consent was obtained from each patients and healthy controls before being included in this study. Treatment-seeking CUD patients and healthy controls were recruited via newspaper advertisements and were initially screened by a brief telephone interview. Following the phone screen, eligible individuals attended an in-person intake assessment session, during which screening for psychiatric disorders using the Structured Clinical Interview for DSM-IV [17], and a medical history and physical examination were conducted. Information about each individual's (drug and non-drug user) demographic and drug use history was also collected at the intake interview. For all individuals, the Addiction Severity Index [18] was obtained to document lifetime drug and alcohol use. Immediately prior to MRI scanning, a urine sample was obtained from each individual to screen for Δ^9-tetrohydrocannabinol, opiates, cocaine, amphetamines, benzodiazepines, and pregnancy (for females). Each individual was also screened for recent alcohol use by measuring breath alcohol concentration. Individual inclusion criteria were:

1. 18 to 55 years old.

2. free of alcohol at the time of MRI scanning.

3. CUD patients met criteria for current cocaine dependence as determined by Structured Clinical Interview for DSM-IV (SCID) [17].

4. healthy controls had no current or lifetime history of any DSM-IV substance use or psychiatric disorder.

Exclusion criteria were:

1. met current or past DSM-IV Axis I disorder other than substance abuse or dependence.

2. taking medication or having disorders that could affect the central nervous system.

3. claustrophobia during MRI simulator sessions.

4. having any definite or suspected clinically-significant brain abnormalities on the Fluid-Attenuated Inversion Recovery (FLAIR) MRI scans.

5. urine drug screen showing positive for Δ^9-tetrohydrocannabinol, opiates, cocaine, amphetamines, benzodiazepines (for controls only).

6. positive pregnancy test result (for females).

Based on the inclusion and exclusion criteria, 39 CUD patients and 18 healthy controls were included in the final analysis. Demographic, drug use, and genotype characteristics of the two groups are shown in Table 1.

MRI data acquisition

MRI data were acquired on a Philips 3.0 T Intera system with a six channel receive head coil (Philips Medical Systems, Best, Netherlands). Whole brain diffusion-weighted images (DWI) were acquired in the transverse plane using a single shot diffusion sensitized spin echo echo-planar imaging (EPI) sequence, with the following parameters: b-factor = 1000 s/mm^2, repetition time = 6100 ms, echo time = 84 ms, 44 contiguous axial slices, field-of-view = 240 mm × 240 mm, 112 × 112 acquisition matrix, 256 × 256 reconstructed matrix, 0.9375 mm × 0.9375 mm reconstructed in-plane resolution, slice thickness = 3 mm, and zero interslice gap. The diffusion tensor encoding scheme was based on the uniformly distributed and balanced rotationally invariant Icosa21 (21 gradient directions) tensor-encoding set [19]. A SENSE acceleration factor of 2 was used for the DWI acqustion. The diffusion-encoded volumes were acquired with fat suppression. The DTI acquisition time was approximately 7 min. FLAIR scans and T2-weighted spin-echo scans were acquired and were read by a board-certified radiologist to rule-out any incidental brain pathology.

Table 1. Range and average for demographics of study sample in all groups.

Group	Sex	N	Age	YCU	Average of years	$GAD1^a$	$GAD1^b$
Cocaine use disorder	M	29	(23.1–52.5)	(0.25–30)	14.112	0.655	0.862
	F	10	(22.7–54.8)	(2–23)	13.00	0.5	0.8
	All	39	(22.7–54.8)	(0.25–30)	13.8269	0.615	0.846
Control	M	11	(23.2–48.8)	NA	NA	0.417	0.667
	F	7	(21–42.1)	NA	NA	0.428	1
	All	18	(21–48.8)	NA	NA	0.421	0.789

N: sample size, Age and Years of use Cocaine: range in years, YCU: Years of cocaine use.

DTI data processing

The DTI images were processed using the FMRIB Software Library www.fmrib.ox.ac.uk/fsl, version5.0.9 [20]. For each data set, the DWI images were corrected for eddy current distortions and head motion [21] after converting the Philips DICOM files into NIfTI format using dcm2nii as implemented in MRIcron http://www.mccauslandcenter.sc.edu/mricro/mricron/. Next, non-brain tissue was removed from the images using FSL's Brain Extraction Tool [22]. After these preprocessing steps, the FMRIB's Diffusion Toolbox [23] was used to fit the data to extract the DTI parameters for each voxel. The DTI parameters included fractional anisotropy (FA), mean diffusivity (MD), the three eigenvalues ($L1$, $L2$, $L3$) and the three eigenvectors ($V1$, $V2$, $V3$). Among these DTI measures, FA has been most commonly used as a general measure of white matter microstucture [24].

White matter registration was carried out using Tract-Based Spatial Statistics (TBSS) [25] as implemented in the FSL software. The FA images from all individuals were aligned to the standard MNI (Montreal Neurological Institute) space using the FSL's nonlinear registration with the `FMRIB58_FA` template image. Next, the mean FA image was created and thinned to create a mean FA skeleton representing the centers of all tracts common to all individuals, using a threshold of FA = 0.20. Each individual's aligned FA data that were local maximums within a short-radius plane normal to each point on the skeleton which were then projected onto that point of the skeleton.

Genetic data acquisition

DNA was extracted from the blood samples drawn from each subject using standard methods [26]. The candidate genetic variants $GAD1^a$ (SNP: rs1978340) and $GAD1^b$ (SNP: rs769390) were determined as described in previous studies [14, 27].

Method

Model formulation

This study was approved by the local university Committee for the Protection of Human Participants (CPHS) and was performed in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). This is a secondary data analysis, the data is the same data from a previous work [14]. This work is performed on a data-set of 57 individuals of which 39 were cocaine dependent and 18 non-drug using controls. In this study, the covariates that are considered in the model are $GAD1^a$, $GAD1^b$, human sex, and years of cocaine use, and their interactions. That leaves us with eleven covariates ($k = 11$). The following set of covariates were considered in the regression model:

$$\{ Sex, YCU, GAD1^a, GAD1^b,$$

$$Sex|YCU, Sex|GAD1^a, Sex|GAD1^b,$$

$$YCU|GAD1^a, YCU|GAD1^b, GAD1^a|GAD1^b,$$

$$YCU|GAD1^a|GAD1^b \}$$

Let the response variable $y_i(v)$ be the FA values at specific voxel $v = 1, \ldots, V$ for $i = 1, \ldots, n$ where n is the number of brain images (patients), and let X be the regression $n \times k$ matrix with demographic and genetic variables, knowing that covariates do not vary between voxels. So, the general linear regression model for the Vth voxel is given by $y(v) = X\beta(v) + \epsilon(v)$, where $\epsilon \sim N(0, \sigma^2(v))$, and $\beta(v)$ is the vector of regression coefficients.

Bayesian model averaging

The model space is $\mathbf{M} = \{M_j: j = 1, \ldots, 2^k\}$, where $2^k = 2^{11} = 2048$ is the number of all possible model combinations, and any M_j model has a subset of demographic and genetic variables. Since the covariates are the same among all voxels, the model space does not vary through all voxels. Let M_j be any model in $\mathbf{M}$ with parameters $\theta_j(v) = (\beta_j(v), \sigma_j^2(v))$ at voxel v. Where $\beta_j(v)$ is the vector of regression coefficients and $\sigma_j^2(v)$ is the variance for the jth model. The prior probability for any j^{th} model in the model space $\mathbf{M}$ indicate as $p(M_j)$. Since we do not have any prior knowledge and we do not prefer any model on others, so we assume $p(M_j) = \frac{1}{|\mathbf{M}|} = \frac{1}{2048}$ as uniform distribution among the model space with size $|\mathbf{M}| = 2048$. To obtain a posterior probability for model M_j at voxel v under data $D(v) = (Y(v), X)$, we use Bayes theorem:

$$p(M_j|D(v)) = \frac{p(D(v)|M_j)p(M_j)}{\sum_{j=1}^{2^K} p(D(v)|M_i)p(M_i)} \qquad (1)$$

where $p(D(v)|M_j)$ is the marginal probability of the data given model M_j:

$$p(D(v)|M_j) = \int L(D(v)|\theta_j, M_j)p(\theta_j|M_j)d\theta_j \qquad (2)$$

Here θ_j is the group parameters that associated with model M_j, and $p(\theta_j|M_j)$ is the prior distribution of θ_j for model M_j. For the parameters that common to all models corresponding to a given voxel such as the intercept β_{0j} and the variance σ_j^2, we used a statndard non-informative prior $p(\beta_{0j}, \sigma_j^2) = \sigma_j^{-1}$ [28]. In the special case when the prior for the regression coefficients is $\beta_j(v) \sim N(0, \sigma^2(gX_j'X_j)^{-1})$, where $g = \frac{1}{max\{n,k^2\}}$, then according to [29] there exists a closed form for $p(D(v)|M_j)$ as:

$$\begin{aligned} p(D(v)|M_j) \quad &\sim (\frac{g}{g+1})^{\frac{k_j}{2}}(\frac{1}{g+1}y(v)'[I_n - X_j(X_j'X_j)^{-1}X_j']y(v) \\ &+ \frac{g}{g+1}(y(v) - \overline{y(v)})'(y(v) - \overline{y(v)}))^{\frac{-(n-1)}{2}} \end{aligned} \qquad (3)$$

where $\overline{y(v)}$ is a vector of length n with the sample mean of $y(v)$ for each voxel, and k_j is the number of covariates in the jth model. In BMA, measuring variables importance depends on the posterior effect probability of each variable. After calculating the posterior probabilities for all models, we compute the Posterior Inclusion Probability (PIP) for every covariate as well by:

$$\pi_k(v) = p(X_k|D(v)) = \sum_{i=1}^{2^k} I_{X_k \in M_j}(M_j)p(M_j|D(v)) \qquad (4)$$

Where $p(M_j|D(v))$ is the posterior model probability for M_j, and $I_{X_j \in M_j}(M_j)$ is a function that takes only two values 0 and 1. It takes value 1 if $X_j \in M_j$ and 0 otherwise [30]. Finally, we compute the posterior distribution of coefficients $\beta_j(v)$ to determine the sign of effect at the brain locations that have high PIP by:

$$p(\beta_j(v)|D(v)) = \sum_{j=1}^{2^k} p(\beta_j(v)|D(v), M_j)p(M_j|D(v)) \qquad (5)$$

However, calculating the posterior probabilities for any model M_j is difficult to implement

since it depends on the number of models which is in our case 2048 models for each voxel, that is involved in the summation to calculate the posterior probability. The Markov chain Monte Carlo model composition procedure (MC^3) can evaluate the posterior model probabilities. According to [31] this approach randomly moves through the model space $\mathbf{M}$, assuming that we start at model M_j with a neighborhood named N_{M_j}. Where N_{M_j} contains the model M_j itself and a group of other models with either one variable more or one variable fewer than M_j. Let $M_m \in N_{M_j}$ a new model, which chosen randomly by sampling from a uniform distribution on the N_{M_j}. Suppose that the chain starts at model M_j, the chain might move to M_m with probability

$$\omega_{jm} = min\{1, \frac{p(M_m|y(v))}{p(M_j|y(v))}\} \tag{6}$$

Otherwise, the chain stays at model M_j with probability $1 - \omega_{jm}$. The procedure continues by randomly finding a new model along the existing model. In the model selection, the model M_j with the highest values of $p(M_j|y(v))$ is selected to be the best model. For more details describing the Markov chain Monte Carlo see [32].

Bayesian false discovery rate control

The aforementioned method used to find brain regions that present strong evidence for differential diffusion patterns across the specific genetic and demographic variables. There are over 600,000 voxels in this work,that number of voxels consider as high number and would cause a high false discovery rate. By apply the scheme developed by [33, 34], the False Discovery Rate (FDR) issue is resolved. Now assume that the remain voxels determine specific locations where the corresponding FA are affected by the differences in a particular genetic variant where the posterior probabilities exceed a given threshold ϕ. First need to find these posterior probabilities $\pi_k(v)$, suppose the space of those significant voxels for one specific genetic variant is given by:

$$X_\phi = \{v : \pi_k(v) \geq \phi\} \tag{7}$$

Let ϕ_δ is the corresponding threshold where δ represent the chosen level for the false discovery rate, which in our case is 10% of the highest significant voxels. To determine which probability needs to include at that space want to choose ϕ. Suppose Ψ is the space of all our posterior probabilities, and then sorting Ψ descending. Our threshold $\phi_\delta = \Psi_k(\xi)$, and that when

$$\xi = max\{k^* : \frac{1}{k^*}\sum_{k=1}^{k^*}(1 - \Psi_k) \leq \delta\} \tag{8}$$

where k^* is the number of voxels exceeding the threshold. After determining the threshold, the result would be the final set of the significant voxels with strong evidence of effect on FA.

Application

This work starts with more than 7 million voxels, but the most interesting regions are only white matter regions. By removing any black voxel and any voxel with an average of FA value lower than 0.2 in brightness which indicate that voxel has low amount of brain white matter, that leaves 601,833 voxels of interest under 2048 possible covariate combinations of models at each voxel. By using `BMS` package in open-source statistical computing software `R` [35], we calculate PIP's and the coefficients. The computation through the High-Performance Computing
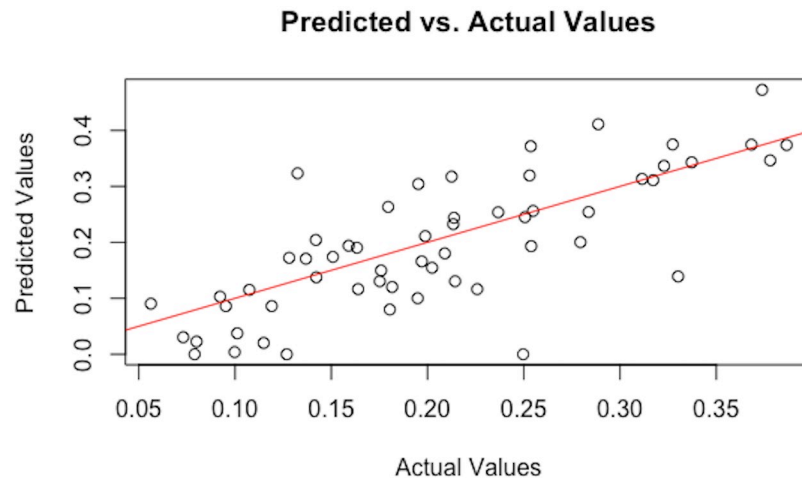
Predicted vs. Actual Values

Fig 1. Actual FA values vs predicted values. This is a scatter plot of actual FA values vs predicted FA values for one significant voxel $v = (150, 80, 75)$ across 57 individuals.

took approximately 4 days. The PIP and coefficient data was restored and the PIPs were passed through the FDR algorithm to identify significant voxels for each of the covariates. This gave 6,646 significant voxels/covariate combinations. Then these voxels were then matched to the John Hopkins University (JHU) Atlas [36–38] map. There are 48 white matter regions of interest in brain based on the white matter JHU Atlas, which can help to determine the brain regions that present proof of changing of white matter among that significant voxels occur. Fig 1 is an example of one voxel where FA values vs predicted values. In general, it follows a linear line but this pattern is vary from voxel to voxel.

Results

The number of significant voxels in each region for all covariate combinations as well as total number of voxels and regions with significant white matter changes are shown in S1 Table. One-way years of cocaine use (YCU) is associated with 1046 voxels significant for white matter changes, distributed across 34 JHU Brain Atlas regions; the majority of these regions are notably frontal corpus callosum, corona radiata, and internal capsule regions. One-way $GAD1^a$ demonstrated a more prolific effect on the tractography data than one-way $GAD1^b$ with a distinctly greater number of significant voxels (212 versus 73 total significant voxels). Nearly twice as many JHU brain regions were represented for $GAD1^a$ (27) versus $GAD1^b$ (14), suggesting a wider distribution of FA changes associated with $GAD1^a$. Regions implicated in $GAD1^a$ include primarily the splenium (19), anterior limb of the internal capsule (13), bilateral superior corona radiata (24 and 12 for right and left, respectively), superior longitudinal fasciculus (19), as well as the posterior thalamic radiation (13). Most pronounced among the $GAD1^b$ clusters was the anterior limb of the internal capsule (39) with no other areas exceeding a threshold of 10 voxels.

With the two-way models, there was a substantial increase in significant voxels associated with $GAD1^b$ versus $GAD1^a$ and the number of regions with significant white matter changes was about equal. A total of 403 significant voxels for $GAD1^a|YCU$ among 33 primarily frontal and colossal JHU regions and 1553 for $GAD1^b|YCU$ among 34 mostly posterior, cerebellar, and connective tract regions. There were fourteen regions across $GAD1^a|YCU$ with more than 10 voxels, amongst which right external capsule was the largest significant region with 44
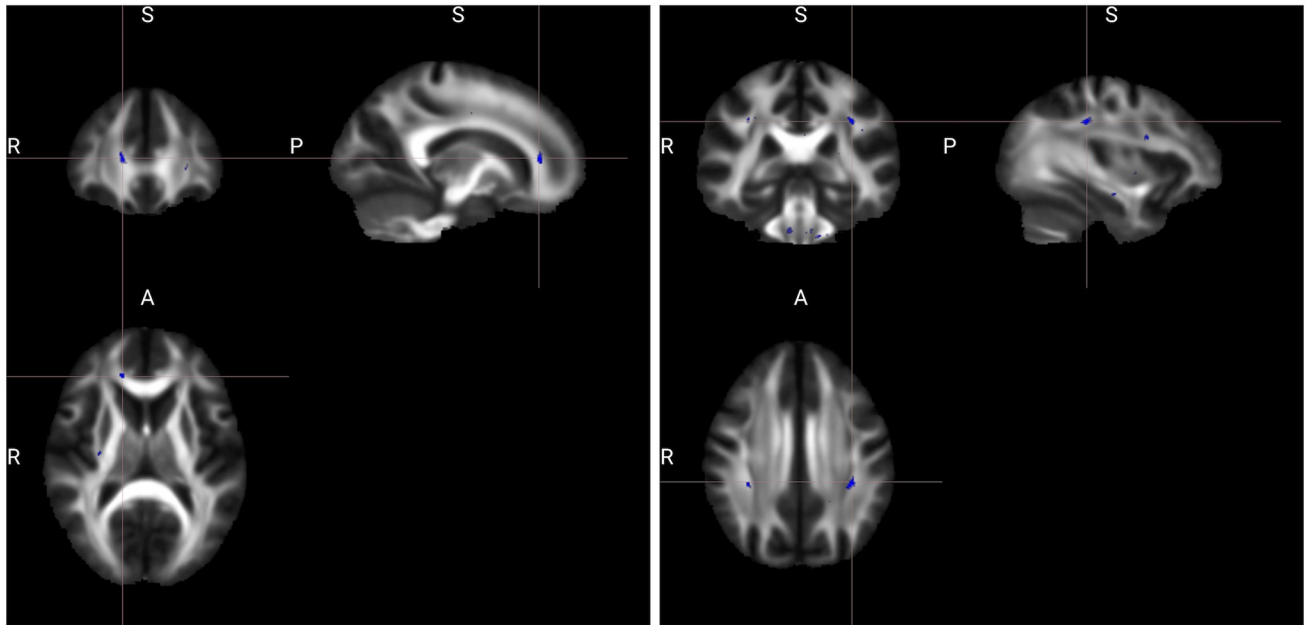
Fig 2. Significant regions in which the FA showing significant effect, among $GAD1^a|YCU$ (left) and $GAD1^b|YCU$ (right). The background is the average FA map. The viewer's left side of the axial slice is the left hemisphere of the displayed brain. The clusters shown are the "thickened" version of the skeletonized results, using the FSL tbss_fill script.

voxels. Other significant areas for $GAD1^a|YCU$ included the bilateral external capsule (44 and 33 for right and left), right anterior and right superior corona radiata (33 and 32, respectively). The body and genu of the corpus callosum (33 and 24, respectively) were also implicated. $GAD1^b|YCU$ demonstrated fifteen white matter regions with more than 10 voxels. The middle cerebellar peduncle was the largest region with 991 voxels; other cerebellar areas were implicated including right and left inferior cerebellar peduncles (49 and 62, respectively). Other major areas included in $GAD1^b|YCU$ were the right and left longitudinal fasciculus (28 and 101), right and left anterior corona radiata (26 and 13), as well as right and left medial lemniscus (44 and 51). Some genu involvement was noted as well (24), among others. In Fig 2 some significant FA findings are shown for $GAD1^a|YCU$ and $GAD1^b|YCU$.

For the three-way model ($GAD1^a|GAD1^b|YCU$), there were a total of 842 significant voxels distributed across 32 regions. Eight white matter regions had more than 10 voxels in middle cerebellar peduncle (585 voxels), right anterior limb of internal capsule (24 voxels), and left retrolenticular part of internal capsule (14 voxels). Other cerebellar involvement included the right and left inferior cerebellar peduncles (38 and 51). Right and left medial lemniscus (20 and 21) as well as left superior longitudinal fasciculus (18 voxels) also were included. Perhaps most notable, the three-way model featured 14 regions with a lower number of voxels than either of the two-way models for $GAD1^a$ or $GAD1^b$; these regions included most of the pontine crossing tract, all corpus callosum areas, right-sided corticospinal tract, bilateral anterior corona radiata, right-sided posterior corona radiata, left-sided sagital stratum, bilateral external capsule, left-sided cingulum, left-sided stria terminalis, and right-sided superior longitudinal fasciculus. There were also 6 regions in which the three-way model had more significant voxels than either of the two-way models. These areas included fornix, right-sided anterior limb of internal capsule, left-sided posterior limb of the internal capsule, left-sided retrolenticular part of the internal capsule, right-sided cingulum, and left-sided uncinate fasciculus.
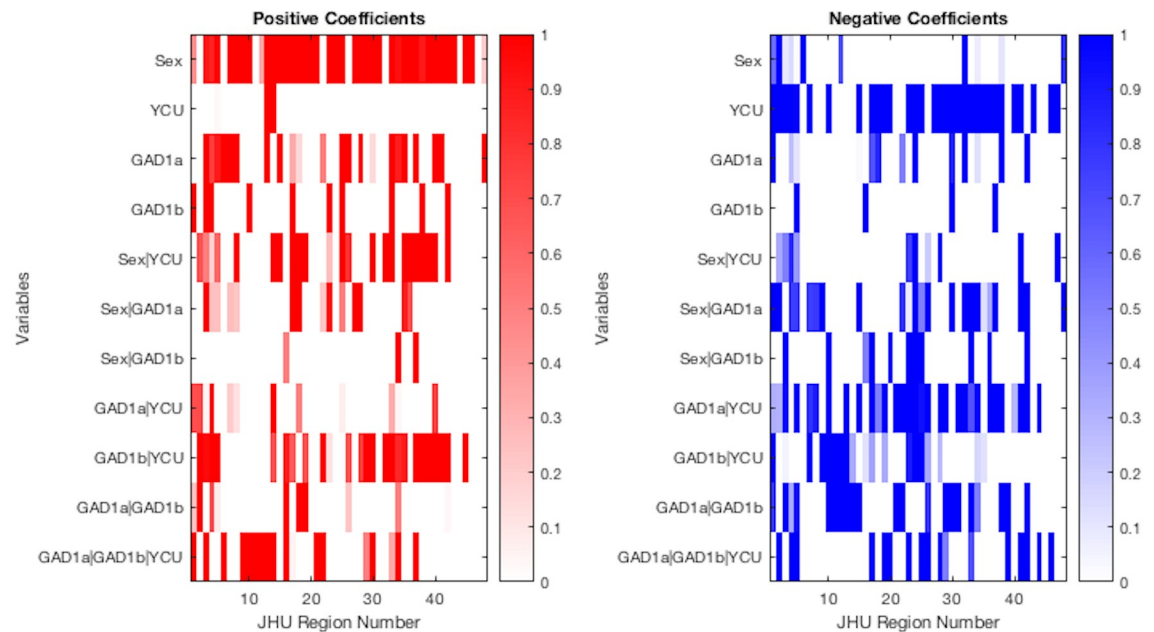
Fig 3. Proportion of positive coefficients (left) for significant voxels by variables across JHU Atlas regions, and negative (right).
The color key represents the extent of FA alteration. White color indicates the absence of significant voxels for the specific combination of brain regions by JHU Atlas among genetic variants, Sex, and YCU.

https://doi.org/10.1371/journal.pone.0254776.g003

Among the model coefficients, male sex was associated with fewer white matter changes and YCU was associated with more across essentially the entire brain as shown in Fig 3. As a one-way variable, $GAD1^a$ was associated with fewer white matter changes primarily in the corpus callosum, cingulum, and superior corona radiata, demonstrated as a relative prevalence of positive coefficients. $GAD1^b$ was associated with overall slightly fewer white matter changes in 10 regions throughout the brain and more white matter changes in 4 areas distributed throughout the brain.

The two-way interaction of Sex with other covariates appeared to reduce the number of regions with positive coefficients while increasing the number with negative. For example, whereas one-way $GAD1^a$ on its own demonstrated more positive coefficients in the corpus callosum and frontal areas among others (23), many of these coefficients were moderated and reduced for $GAD1^a|Sex$ (11 total). For the negative coefficients, $GAD1^a$ on its own featured relatively few regions with changes (12) but $GAD1^a|Sex$ featured relatively more (23).

Perhaps most notably, $GAD1^a|YCU$ demonstrated primarily negative coefficients (31) in frontal areas and connective fiber tracts including internal capsule, corona radiata, thalamic radiations, and longitudinal fasciculus; positive coefficients (14) were relatively more sparse but present primarily for cerebellum and corpus callosum. $GAD1^b|YCU$ demonstrated primarily positive coefficients (25) in corpus callosum, other thalamic radiations, and also longitudinal fasciculus; negative coefficients (18) were shown primarily for cerebellar peduncle as well as anterior and superior corona radiata. The three-way interaction $GAD1^a|GAD1^b|YCU$ had positive coefficients (17) primarily for cerebellar peduncle and negative coefficients (17) for corpus callosum and internal capsule.

Discussion

We applied BMA methods to DTI and genetic data collected from CUD individuals and controls, in order to investigate interaction effects of two-way and three-way variables among the

$GAD1^a$ and $GAD1^b$ SNPs as well as clinical covariate years of cocaine use (YCU) and demographic covariate Sex in the human brain. One-way $GAD1^a$ was associated with broader and more significant FA changes throughout the brain with respect to $GAD1^b$, a notable 3 times difference and comparable perhaps to the roughly 4 times difference noted for the same relationship by [14]. Significant changes for one-way $GAD1^a$ were more consistent with typical frontal and forebrain white matter changes associated with cocaine use disorder including corpus callosum, corona radiata, internal capsule, and thalamic projections; hindbrain white matter was shown in the cerebral peduncle, as well. One-way $GAD1^b$ showed FA changes in the internal capsule alone with only scant findings elsewhere in the brain.

The inclusion of the YCU covariate revealed more substantial distribution of white matter changes for the two-way $GAD1^b|YCU$ including in the corpus callosum, additional frontal (medial lemniscus, internal capsule, corona radiata, cingulate gyrus), as well as thalamic and longitudinal fasciculus FA changes. The middle cerebellar, cerebellar peduncle FA, and superior longitudinal fasciculus changes were perhaps the most impressive with 991, combined 118 (inferior and superior cerebellar peduncle), and 129 voxels, respectively. The corpus callosal and forebrain FA changes combined with the length of use variable are perhaps more characteristic of white matter changes most typically associated with cocaine use disorder [4, 5, 7]. Increased grey matter cerebellar changes have also been associated with the length of cocaine use [8]. $GAD1^a|YCU$ was also associated with corpus callosum as well as similar frontal white matter FA changes (internal capsule, corona radiata, external capsule). Though there are some notable differences between the two models, the relative similarity of the region of FA changes between $GAD1^a|YCU$ and $GAD1^b|YCU$ suggests that YCU may be the primary driver of these FA changes. For example, $GAD1^b$ on its own features only a single voxel in middle cerebellar peduncle whereas $GAD1^b|YCU$ has 991. However, there is a distinct contribution of $GAD1^b$ with this particular relationship because $GAD1^a|YCU$ demonstrates only 13 significant voxels in the same region. Though there is clearly an effect of YCU on both $GAD1^a$ and $GAD1^b$, this particularly impressive finding would demonstrate an interaction between $GAD1^b|YCU$.

Both two-way polymorphism and YCU models were associated with a relative increase in (Genu and Body) corpus callosum, anterior corona radiata, internal capsule, and longitudinal fasciculus involvement with respect to their one-way models. Many of these changes—particularly the corpus callosum involvement—were novel for the two-way models with respect to the one-way polymorphism models, suggesting again that YCU was the primary driver. This finding is consistent with prior research implicating corpus callosum changes associated with chronic cocaine use [7]. Notable differences include greater corticospinal tract, cerebral peduncle, and external capsule involvement for $GAD1^a|YCU$ as well as greater medial lemniscus, cerebellar, thalamic radiation, and longitudinal fasciculus involvement for $GAD1^b|YCU$. Many of these differences were also novel for the two-way with respect to the one-way models —again, pointing to the relative contribution of the YCU variable. Again, broader frontal brain and cerebellar FA changes have been associated with cocaine use [3, 12]. White matter changes were similarly broad for $GAD1^a|YCU$ and $GAD1^b|YCU$ which featured 33 and 34 regions with white matter changes, respectively. In addition to differential involvement of certain regions, there was a substantially greater number of much larger concentration of white matter changes in the cerebellum for $GAD1^b|YCU$ and, in general, a greater posterior/mid/hindbrain-centric distribution of white matter changes. $GAD1^a|YCU$, on the other hand, featured a larger distribution of frontal white matter changes. It is perhaps remarkable that the number of voxels with white matter changes increases by a factor of 20 for $GAD1^b$ with the addition of the YCU covariate whereas the increase is only by a factor of 2 for $GAD1^a$. However, if we exclude the middle cerebellar peduncle from the $GAD1^b|YCU$ model, the relative increase is only about a factor of 7. Based on the voxel-wise white matter changes data, it

appears that while $GAD1^a$ is associated with more white matter changes on its own in the one-way model there may be a relative protective effect associated with this polymorphism with respect to chronic cocaine use. Though there appears to be relatively fewer white matter changes on its own in the one-way, $GAD1^b$ related white matter changes increase dramatically with the addition of the YCU covariate, suggesting perhaps a relative vulnerability to chronic cocaine use in individuals with the $GAD1^b$ polymorphism, particularly in the cerebellum.

The model coefficients in Fig 3 provide additional evidence of the differential effects of each of the covariates. In particular, the effect of YCU on FA was negative across the majority of brain regions, as would be expected. Male sex also appears to demonstrate a protective effect against white matter changes with positive coefficients across most brain regions. The two-way interaction $GAD1^a|YCU$ featured negative coefficients for most white matter regions (31) with relatively fewer positive coefficients (10). $GAD1^b|YCU$ exhibited somewhat the opposite trend in the sense that, except for the cerebellar regions, there were relatively few negative coefficients (17) and more positive coefficients (25). The effect of the Sex covariates on FA was positive among approximately all coefficients, suggesting that male sex was a protective factor. The roughly equal number of positive (18) and negative coefficients (17) for $GAD1^a|GAD1^b|YCU$ suggests that there isn't a significant interaction between the polymorphisms and YCU that meaningfully increases or decreases white matter changes across the whole brain for individuals who have both polymorphisms expressed. That said, the presence of 14 regions in which the number of voxels with white matter changes were less than either two-way polymorphism model suggests that there may be a protective effect for individual regions (including corpus callosum, anterior corona radiata, among others). On the other hand, the presence of 6 regions in which the three-way model included slightly more voxels than either two-way model (including fornix and internal capsule) points to the possibility that having both polymorphisms may be associated with a deleterious effect in certain regions, as well.

For the long term goal which is to find a treatment for CUD, now we know how the polymorphisms interact with each other and with lifetime cocaine use, also we know which is more effective than others. For future work, This can provide genetic testing that allowance position to better prescribe a treatment. Also allowance the doctors really understanding what is in the patient's future given their genetic profile, which can be used at least as personalized medicine.

Conclusion

Bayesian analysis was used to analyze the relationships between genetic data, sex, years of cocaine use, and fractional anisotropy white matter changes. The underlying advantage of this modeling approach is taking into account the model uncertainty through Bayesian model averaging. This method was building on genetic data from the previous work of [14] exploring in greater detail the relative effects of years of cocaine use as well as the possession of $GAD1^a$ and $GAD1^b$ polymorphisms. The primary novel addition in this study is the added consideration of years of cocaine use and sex in a series of two- and three-way models of the interaction of the genetic variants on brain white matter in users of cocaine. Specifically, the goal of this work was to identify brain regions that present strong evidence for differential diffusion patterns across the two-way and three-way interactions of $GAD1^a|YCU$, $GAD1^b|YCU$, and $GAD1^a|GAD1^b|YCU$. Ancillary goals included evaluating the effect of sex on white matter changes; in these data, male sex appears to have a mild protective effect. $GAD1^a$ appears to be associated primarily with frontal white matter changes with only a moderate increase in voxel-wise volume with chronic cocaine use. $GAD1^b$, on the other hand, was associated with a more dramatic increase in cerebellar white matter changes as well as other primarily mid/hindbrain regions based on the voxel-wise data. The presence of 14 regions in the three-way model

which feature lower voxel counts than either of the two-way polymorphism models points to a possible protective effect of having both polymorphisms. The model coefficients indicate that $GAD1^a|YCU$ is associated with more negative white matter changes and $GAD1^b|YCU$ with more positive. The three-way model $GAD1^a|GAD1^b|YCU$ demonstrated fewer negative coefficients than either two-way model as well as a cluster of novel positive coefficients in the medial lemniscus and cerebellar peduncle not shared by the two-way models—again, pointing to a possibly protective effect in having both polymorphisms.

However, since the values of the response variable $y(v)$ are belong to the standard unit interval (0, 1), then the linear regression is not the most appropriate method to analyze the data. That because the predicted values could be out of the interval (0, 1) when using linear regression. One of the suggestions solution of this issue is to transform the response variable to be values on the real line, then analyze the transformed data using linear regression. However, this method has some weaknesses such as: the regression parameters are explicated in terms of the transformed data and it is difficult to interpret it in terms of the un-transformed data. Moreover, the distributions of proportions are not symmetric, then conclusions relying on the normality assumption could be invalid. In order to overcome this problem, a regression model for continuous variables that belong to the standard unit interval (0, 1) has been proposed by [39]. This proposed model is under an assumption that the response variable is Beta distributed, and the model is called Beta regression model. The underlying advantage to use the Beta regression model is the flexibility that achieved by the assumption of the response variable which is following Beta distribution.

Supporting information

S1 Table. Number of voxels with white matter (FA) changes by JHU Brain Atlas region for the variables and covariates.
(XLSX)

Author Contributions

Formal analysis: Tmader Alballa.

Funding acquisition: Liangsuo Ma, F. Gerard Moeller.

Investigation: Andrew Snyder.

Methodology: Tmader Alballa.

Project administration: F. Gerard Moeller.

Software: Tmader Alballa, Edward L. Boone.

Supervision: Edward L. Boone.

Visualization: Edward L. Boone.

Writing – original draft: Tmader Alballa, Liangsuo Ma.

Writing – review & editing: Edward L. Boone, Andrew Snyder.

References

1. Lim KO, Choi SJ, Pomara N, Wolkin A, Rotrosen JP. Reduced frontal white matter integrity in cocaine dependence: a controlled diffusion tensor imaging study. Biological psychiatry. 2002; 51(11):890–895. https://doi.org/10.1016/S0006-3223(01)01355-5

2. Moeller FG, Hasan KM, Steinberg JL, Kramer LA, Dougherty DM, Santos RM, et al. Reduced anterior corpus callosum white matter integrity is related to increased impulsivity and reduced discriminability in cocaine-dependent subjects: diffusion tensor imaging. Neuropsychopharmacology. 2005; 30(3):610–617. https://doi.org/10.1038/sj.npp.1300617 PMID: 15637640

3. Lim KO, Wozniak JR, Mueller BA, Franc DT, Specker SM, Rodriguez CP, et al. Brain macrostructural and microstructural abnormalities in cocaine dependence. Drug and alcohol dependence. 2008; 92(1-3):164–172. https://doi.org/10.1016/j.drugalcdep.2007.07.019 PMID: 17904770

4. Ma L, Hasan KM, Steinberg JL, Narayana PA, Lane SD, Zuniga EA, et al. Diffusion tensor imaging in cocaine dependence: regional effects of cocaine on corpus callosum and effect of cocaine administration route. Drug and alcohol dependence. 2009; 104(3):262–267. https://doi.org/10.1016/j.drugalcdep.2009.05.020 PMID: 19595517

5. Lane SD, Steinberg JL, Ma L, Hasan KM, Kramer LA, Zuniga EA, et al. Diffusion tensor imaging and decision making in cocaine dependence. PLoS one. 2010; 5(7). https://doi.org/10.1371/journal.pone.0011591 PMID: 20661285

6. Bell RP, Foxe JJ, Nierenberg J, Hoptman MJ, Garavan H. Assessing white matter integrity as a function of abstinence duration in former cocaine-dependent individuals. Drug and alcohol dependence. 2011; 114(2-3):159–168.

7. Ma L, Steinberg JL, Wang Q, Schmitz JM, Boone EL, Narayana PA, et al. A preliminary longitudinal study of white matter alteration in cocaine use disorder subjects. Drug and alcohol dependence. 2017; 173:39–46. https://doi.org/10.1016/j.drugalcdep.2016.12.016 PMID: 28192722

8. Sim ME, Lyoo IK, Streeter CC, Covell J, Sarid-Segal O, Ciraulo DA, et al. Cerebellar gray matter volume correlates with duration of cocaine use in cocaine-dependent subjects. Neuropsychopharmacology. 2007; 32(10):2229–2237. https://doi.org/10.1038/sj.npp.1301346 PMID: 17299505

9. Taylor D, Bushell M. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. Physics in Medicine & Biology. 1985; 30(4):345. https://doi.org/10.1088/0031-9155/30/4/009

10. Song SK, Sun SW, Ramsbottom MJ, Chang C, Russell J, Cross AH. Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. Neuroimage. 2002; 17(3):1429–1436. https://doi.org/10.1006/nimg.2002.1267

11. Narayana PA, Ahobila-Vajjula P, Ramu J, Herrera J, Steinberg JL, Moeller FG. Diffusion tensor imaging of cocaine-treated rodents. Psychiatry Research: Neuroimaging. 2009; 171(3):242–251. https://doi.org/10.1016/j.pscychresns.2008.04.004

12. Narayana PA, Herrera JJ, Bockhorst KH, Esparza-Coss E, Xia Y, Steinberg JL, et al. Chronic cocaine administration causes extensive white matter damage in brain: diffusion tensor imaging and immunohistochemistry studies. Psychiatry Research: Neuroimaging. 2014; 221(3):220–230. https://doi.org/10.1016/j.pscychresns.2014.01.005 PMID: 24507117

13. Romero MJ, Asensio S, Palau C, Sanchez A, Romero FJ. Cocaine addiction: diffusion tensor imaging study of the inferior frontal and anterior cingulate white matter. Psychiatry Research: Neuroimaging. 2010; 181(1):57–63. https://doi.org/10.1016/j.pscychresns.2009.07.004

14. Azadeh S, Hobbs BP, Ma L, Nielsen DA, Moeller FG, Baladandayuthapani V. Integrative Bayesian analysis of neuroimaging-genetic data with application to cocaine dependence. NeuroImage. 2016; 125:813–824. https://doi.org/10.1016/j.neuroimage.2015.10.033

15. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Statistical science. 1999; p. 382–401.

16. Seber GA, Lee AJ. Linear regression analysis. vol. 329. John Wiley & Sons; 2012.

17. First MB. Structured clinical interview for DSM-IV axis I disorders. Biometrics Research Department. 1997;.

18. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al. The fifth edition of the Addiction Severity Index. Journal of substance abuse treatment. 1992; 9(3):199–213. https://doi.org/10.1016/0740-5472(92)90062-S PMID: 1334156

19. Hasan KM, Narayana PA. Computation of the fractional anisotropy and mean diffusivity maps without tensor decoding and diagonalization: theoretical analysis and validation. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2003; 50 (3):589–598. https://doi.org/10.1002/mrm.10552

20. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012; 62 (2):782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

21. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Medical image analysis. 2001; 5(2):143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

22. Smith SM. Fast robust automated brain extraction. Human brain mapping. 2002; 17(3):143–155. https://doi.org/10.1002/hbm.10062

23. Behrens TE, Johansen-Berg H, Woolrich M, Smith S, Wheeler-Kingshott C, Boulby P, et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature neuroscience. 2003; 6(7):750–757. https://doi.org/10.1038/nn1075 PMID: 12808459

24. Hasan KM, Narayana PA. Retrospective measurement of the diffusion tensor eigenvalues from diffusion anisotropy and mean diffusivity in DTI. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2006; 56(1):130–137. https://doi.org/10.1002/mrm.20935

25. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage. 2006; 31(4):1487–1505. https://doi.org/10.1016/j.neuroimage.2006.02.024 PMID: 16624579

26. Oosterhuis BE, LaForge KS, Proudnikov D, Ho A, Nielsen DA, Gianotti R, et al. Catechol-O-methyltransferase (COMT) gene variants: Possible association of the Val158Met variant with opiate addiction in hispanic women. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2008; 147(6):793–798. https://doi.org/10.1002/ajmg.b.30716 PMID: 18270997

27. Azadeh S, Hobbs BP, Ma L, Nielsen DA, Moeller FG, Baladandayuthapani V. Integrative Bayesian analysis of neuroimaging-genetic data through hierarchical dimension reduction. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE; 2016. p. 824–828.

28. Berger JO. Prior information and subjective probability. In: Statistical Decision Theory and Bayesian Analysis. Springer; 1985. p. 74–117.

29. Fernandez C, Ley E, Steel MF. Benchmark priors for Bayesian model averaging. Journal of Econometrics. 2001; 100(2):381–427. https://doi.org/10.1016/S0304-4076(00)00076-2

30. Lee AH, Boone EL. A frequentist assessment of Bayesian inclusion probabilities for screening predictors. Journal of Statistical Computation and Simulation. 2011; 81(9):1111–1119. https://doi.org/10.1080/00949651003702135

31. Madigan D, York J, Allard D. Bayesian graphical models for discrete data. International Statistical Review/Revue Internationale de Statistique. 1995; p. 215–232.

32. Brooks S. Markov chain Monte Carlo method and its application. Journal of the royal statistical society: series D (the Statistician). 1998; 47(1):69–100. https://doi.org/10.1111/1467-9884.00117

33. Morris JS, Brown PJ, Herrick RC, Baggerly KA, Coombes KR. Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. Biometrics. 2008; 64(2):479–489. https://doi.org/10.1111/j.1541-0420.2007.00895.x

34. Baladandayuthapani V, Ji Y, Talluri R, Nieto-Barajas LE, Morris JS. Bayesian random segmentation models to identify shared copy number aberrations for array CGH data. Journal of the american statistical association. 2010; 105(492):1358–1375. https://doi.org/10.1198/jasa.2010.ap09250

35. Zeugner S, Feldkircher M, et al. Bayesian model averaging employing fixed and flexible priors: The BMS package for R. Journal of Statistical Software. 2015; 68(4):1–37. https://doi.org/10.18637/jss.v068.i04

36. Mori S, Wakana S, Van Zijl PC, Nagae-Poetscher L. MRI atlas of human white matter. Elsevier; 2005.

37. Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, et al. Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage. 2007; 36(3):630–644. https://doi.org/10.1016/j.neuroimage.2007.02.049 PMID: 17481925

38. Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, et al. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. Neuroimage. 2008; 39(1):336–347. https://doi.org/10.1016/j.neuroimage.2007.07.053 PMID: 17931890

39. Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. Journal of applied statistics. 2004; 31(7):799–815. https://doi.org/10.1080/0266476042000214501