Brain-wide associations between white matter and age highlight the role of fornix microstructure in brain age

Max Korbmacher[1,2,3,+], Ann Marie de Lange[2,4,5], Dennis van der Meer[2,6], Dani Beck[2,7,8], Eli Eikefjord[1,3], Arvid Lundervold[1,3,9,10], Ole A. Andreassen[2,11], Lars T. Westlye[2,8,11], Ivan I. Maximov[1,2,+]

Author affiliations
[1] Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway
[2] NORMENT Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo and Oslo University Hospital, Oslo, Norway
[3] Mohn Medical Imaging and Visualisation Center (MMIV), Bergen, Norway
[4] Department of Psychiatry, University of Oxford, Oxford, UK
[5] LREN, Centre for Research in Neurosciences - Department of Clinical Neurosciences, CHUV and University of Lausanne, Lausanne, Switzerland
[6] Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands
[7] Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway
[8] Department of Psychology, University of Oslo, Oslo, Norway
[9] Department of Radiology, Haukeland University Hospital, Bergen, Norway
[10] Department of Biomedicine, University of Bergen, Bergen, Norway
[11] KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway
+ corresponding author

Corresponding author contact
Max Korbmacher, M.Sc. & Ivan I. Maximov, PhD
Email: max.korbmacher@hvl.no, ivan.maximov@hvl.no
Postal address: Inndalsveien 28, 5063 Bergen, Norway

Abstract

Identifying white matter (WM) microstructure parameters that reflect the underlying biology of the brain will advance our understanding of ageing and brain health. In this extensive comparison of brain age predictions and age-associations of WM features from different diffusion approaches, we analysed UK Biobank diffusion Magnetic Resonance Imaging (dMRI) data across midlife and older age (N = 35,749, 44.6 to 82.8 years of age). Conventional and advanced dMRI approaches were consistent in predicting brain age; with their WM-features similarly related to and predicted by age. However, brain age was estimated best when combining approaches, showing different aspects of WM to contribute to brain age. Fornix was found as the central region for brain age predictions across diffusion approaches. We encourage the application of multiple dMRI approaches for detailed insights into WM, and the further investigation of fornix as a potential biomarker of brain age and ageing.

Introduction

Neuroscientific research over the past decades has increased our understanding of the brain mechanisms associated with tissue maturation and ageing effects[1]. A particularly fruitful source of data is magnetic resonance imaging (MRI), revealing information about structural and functional brain architecture in vivo[2]. For many MRI modalities, such as diffusion-weighted MRI (dMRI) or T_1-weighted MRI, a variety of quantitative measures can be estimated, and linked to behaviour, cognitive and health scores[3,4]. However, selection and interpretation of such parameters are difficult, largely due to intra-subject variability in ageing, for example influenced by covariates from the genetic to environmental level[4]. Hence, the use of large-scale MRI databases, such as UK Biobank (UKB)[5] or the Human Connectome Project[6], becomes inevitable, as it allows detecting and localising important brain patterns and supporting their generalisability[7]. Simultaneously, large-scale data provides sufficient power for the application of advanced multivariate statistical models, and machine learning (ML) techniques.

Brain age prediction is an example of such a technique, helping translate large amounts of complex multidimensional data into practically interpretable outputs. Brain age prediction involves training a ML model to determine trajectories of brain ageing from a

series of brain MRI features. Once the model is trained, it can predict the age of brains not included in the training data. The disparity between chronological age and predicted age, the so-called brain age gap (BAG), can be used as an indicator for neurological, neuropsychiatric and neurodegenerative disorders[10,11]. For example, BAG has been associated with stroke history, diabetes, smoking, alcohol intake, several cognitive measures[12,13], mortality risk, different brain and psychiatric disorders[14,15], cardiovascular risk factors[19], stroke risk[16], and loneliness[17]. However, besides Alzheimer's disease or schizophrenia, the evidence is mixed for the relationship of BAG and different health outcomes and a smaller BAG is not necessarily indicative of good health[4]. Moreover, recent longitudinal evidence shows early-life factors and genetics to have stronger effects on brain maturation than T_1-weighted grey matter (GM) BAG[18]. However, BAG is a promising heritable indicator of general health status[10,13,19,20].

BAG and age trajectories offer paths towards a better understanding of the ageing brain. There are various detectable age-related brain changes, such as GM and WM atrophy[8], WM de-differentiation[9], and functional connectivity changes[4] which have hence informed the choice of brain-age modelling-parameters[12,16,19,25,27-29,30]. In that context, many ML approaches have been used to make robust and clinically relevant brain age predictions from different MRI modalities[10,21–23]; yet, particularly the eXtreme Gradient Boosting[24] regressor model, using a decision tree approach, being increasingly used for brain age predictions from large-scale data due to its precision and speed [10,25,26]. Especially dMRI and structural MRI have been shown useful for brain age predictions[12,16,19,25,27-29,30]. However, further systematic, sufficiently powered assessments of dMRI-derived brain age and how diffusion metrics map onto age are needed.

DMRI-derived measures consist of unique parameters allowing both to reveal WM changes at micrometer scale and to provide the basis for a prediction of macroscopic outcomes, such as age. Conventionally, WM brain architecture is described using diffusion tensor imaging (DTI)[31]. However, recent advances offer more biophysically meaningful approaches[32], and sensible foundation for cross-validation and better comparability[25]. DTI-derived measures, namely fractional anisotropy (FA), and axial (AD), mean (MD), and radial (RD) diffusivity have all been shown to be highly age sensitive[9,25,33]. However, the DTI approach is limited by the Gaussian diffusion assumption and is unable to take into account entangled WM microstructure features[25]. In the present work, we consider 1) the Bayesian rotationally invariant approach (BRIA)[34], 2) diffusion kurtosis imaging (DKI)[35]; 3) kurtosis derived supplement, known as white matter tract integrity (WMTI)[36]; 4) spherical mean technique (SMT)[37], and 5) multi-compartment spherical mean technique (mcSMT)[38] in addition to DTI. Only a few studies have compared dMRI models directly as original brain age predictors[25,39,40]. Yet, brain age and age curve assessments of DTI, BRIA, DKI, WMTI, SMT, mcSMT (**ST10**) in a representative sample still need establishing, as well as most influential WM regions for brain ageing. Our assessments focus on the process of ageing (from midlife to late adulthood), starting by associating BAG across diffusion approaches and compare brain-age-chronological-age-correlations to assess prediction consistency. Fornix was identified as most contributing feature in these predictions exploring feature-contributions, and was the strongest correlate of age, with fornix features highly correlated across approaches. Finally, we created fornix and whole-brain-age curves expecting curvilinear relationships reflecting brain-tissue-composition at different ageing stages[25,33,52].

Results

Brain age predictions

Table 1 presents a comparison between different diffusion approaches in predicting brain age for each diffusion approach. The strongest correlation between chronological and predicted age was found in the multimodal approach including dMRI data from all six diffusion approaches, Pearson's r=0.805, 95% CI [0.800, 0.808], p<.001, and the smallest correlation in the multimodal approach including only mean scores Pearson's r =

0.627, 95% CI [0.627, 0.639], $p<.001$, respectively (corrected and non-corrected correlations are presented in **Table 1**). The strongest correlation between uncorrected age predictions and chronological age was observed for WMTI Pearson's $r=0.765$, 95% CI [0.761, 0.770], $p<.001$, and the smallest for mcSMT Pearson's $r=0.721$, 95% CI [0.716, 0.726], $p<.001$.

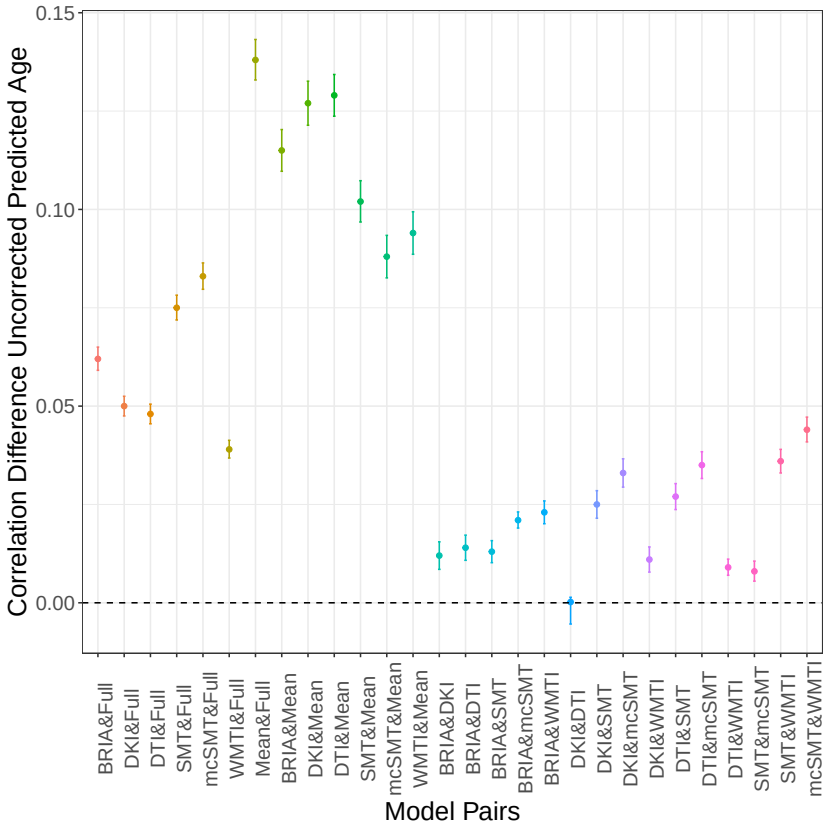
Table 1: Performance of Brain Age Prediction Models

Approach§	Number of MRI features	R^2 (SD)	RMSE (SD)	MAE (SD)	Prediction-Age Correlation* [95% CI]	Corrected Prediction-Age Correlation* [95% CI]
BRIA	690	0.550 (0.012)	5.007 (0.057)	4.002 (0.042)	0.742 [0.737, 0.747]	0.892+ [0.889, 0.894]
DKI	207	0.576 (0.015)	4.958 (0.077)	3.975 (0.068)	0.754 [0.755, 0.764]	0.903 [0.901, 0.905]
DTI	276	0.571 (0.014)	4.983 (0.072)	3.984 (0.062)	0.756 [0.751, 0.761]	0.900 [0.897, 0.902]
SMT	276	0.531 (0.010)	5.214 (0.053)	4.183 (0.036)	0.729 [0.724, 0.734]	0.899 [0.897, 0.901]
mcSMT	276	0.519 (0.011)	5.175 (0.045)	4.153 (0.036)	0.721 [0.716, 0.726]	0.892+ [0.889, 0.894]
WMTI	207	0.585 (0.012)	4.903 (0.065)	3.928 (0.050)	0.765 [0.761, 0.770]	0.902 [0.900, 0.904]
Mean multimodal	28	0.393 (0.012)	5.932 (0.051)	4.812 (0.046)	0.627 [0.621, 0.634]	0.905 [0.903, 0.907]
Full multimodal	1932	0.645 (0.011)	4.534 (0.041)	3.624 (0.037)	0.804 [0.800, 0.808]	0.907 [0.905, 0.909]

Table logic: R^2, RMSE, MAE are displayed in the format Mean (Standard Deviation), Pearson's correlations are displayed in the format Correlation Score 95% Confidence Interval [Lower Bound, Upper Bound].
Mean multimodal refers to diffusion metrics averaged over the skeleton for all six diffusion approaches. Full multimodal refers to all diffusion data from the six diffusion approaches, i.e. mean multimodal data in addition to metrics averaged over the JHU atlas regions. R^2 = variance explained, RMSE = root mean squared error, MAE = mean absolute error.
§ For an overview of the metrics contained in each of the diffusion approaches see **ST10**.
+ Details on the smallest correlation: BRIA Corrected Prediction-Age Correlation r = .89173, mcSMT Corrected Prediction-Age Correlation r = .89176
* All correlation were significant at p < .001.

Hotelling's[41] *t*-tests were used to compare correlations between uncorrected predicted age and chronological age across diffusion models and Zou's[42] method to estimate the confidence intervals around the correlation differences (**Fig.1** and **ST3**; **SF8** and **ST2** for corrected prediction correlation comparisons). These differences were not significantly different from each other for model pairs DKI and DTI ($p\approx1$). All other correlations were different from each other, Pearson's $rs_{diff}\leq0.15$, $p<.001$, with the biggest difference observed between mean and full multimodal scores' correlations (**ST2** for exact values). To identify the most influencing WM regions, we computed the permutation feature importance for each model's features (models: **Table 1**), ranked by contribution to the variance explained (**Table 2**). For feature rankings by contribution to model prediction accuracy using gain scores[43] see **ST15**. Across diffusion approaches, diffusion values estimated on the fornix had the most valuable contribution to variance explained (**Table 2**) and prediction accuracy (**ST15**). Model which had fornix features removed had lower model fit and brain-age-chronological-age correlations were smaller than for models containing fornix ($rs<-0.003$, $ps<.001$; **ST16**).

Fig.1: Differences between Pearson's correlations of chronological and uncorrected predicted ages across diffusion approaches with 95% confidence interval



Differences between Pearson's correlation coefficients of chronological and uncorrected predicted age by diffusion approach. See SF8 for correlational differences between approaches for corrected brain age predictions.

Table 2: Top five diffusion metrics ranked by their contribution to variance explained (R^2) in age

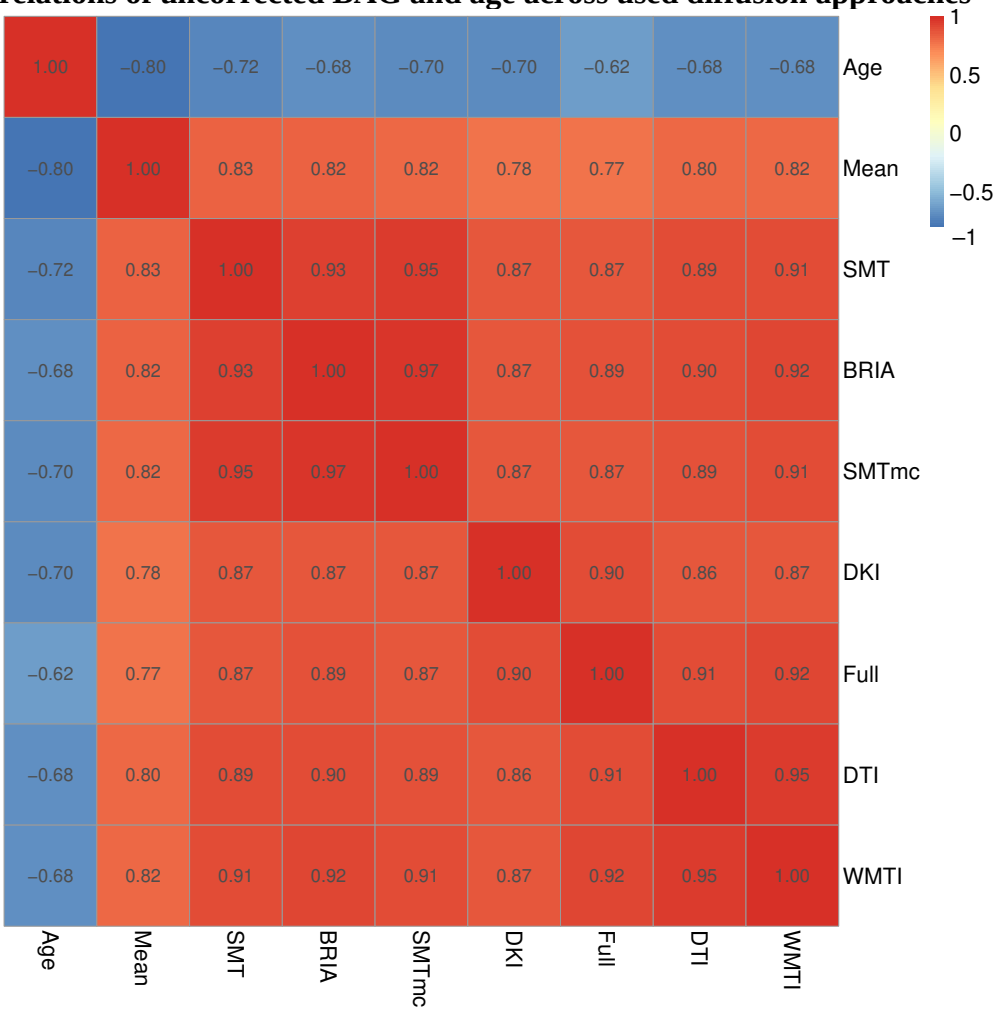
BRIA	DKI	DTI	SMT	mcSMT	WMTI	Multimodal
Micro FA fornix 0.1954±0.0027	AK right anterior limb of internal capsule 0.0984±0.0014	MD fornix 0.0712±0.0013	MD fornix 0.0795±0.0018	Extratrans fornix 0.0498±0.0013	AWF fornix 0.1699±0.0023	Micro FA fornix 0.0914±0.0011
Vextra forceps minor 0.0278±0.0007	RK fornix 0.0884±0.0016	FA forceps minor 0.0533±0.0011	FA right superior longitudinal fasciculus 0.0267±0.0007	Intra forceps minor 0.0444±0.0009	radEAD fornix to right striaterminalis 0.0283±0.0007	AK anterior limb of internal capsule 0.0055±0.0011
Vextra body of the corpus callosum 0.0261±0.0007	MK left external capsule 0.0259±0.0006	RD fornix to right Striaterminalis 0.0462±0.0009	Longitudinal fornix 0.0251±0.0006	Intra fornix 0.0289±0.0009	AWF Forceps minor 0.0194±0.0005	FA forceps minor 0.0219±0.0006
Micro FA fornix to right Striaterminalis 0.0203±0.0006	MK right superior longitudinal fasciculus 0.0214±0.0006	FA right superior cerebellar peduncle 0.0221±0.0006	Trans fornix to right striaterminalis 0.0204±0.0006	Extratrans fornix to right Striaterminalis 0.0201±0.0006	axEAD forceps minor 0.0193±0.0007	RD right fornix stria terminalis 0.0214±0.0006
Vintra right superior cerebellar peduncle 0.0194±0.0006	RK forceps minor 0.0208±0.0005	FA body of the corpus callosum 0.0218±0.0006	FA fornix 0.0192±0.0006	Extratrans right external capsule 0.0163±0.0007	axEAD left posterior limb of internal capsule 0.0173±0.0006	AK Genu corpus callosum 0.0095±0.0003

Note: Variance explained (R^2) by a single feature refers here to the part of the total variance explained by the respective model presented in Table 1. Multimodal refers to an approach using the diffusion metrics from all diffusion approaches. Cells containing Fornix are marked in green.

Brain age gap across diffusion approaches and age

In order to compare uncorrected BAG (BAG_u) calculations across the used diffusion approaches, BAG_u was correlated from different diffusion approaches and with age. Correlations between the six diffusion approaches ranged between r=0.857 and r=0.966 (**Fig.2; SF1** for corrected BAG correlations). Overall, BAG_u scores from the different approaches were strongest related to WMTI BAG_c (range: r = 0.873 to 0.952), and weakest to mean multimodal BAG_u (range: r=0.779 to r=0.828), and could be observed in one cluster containing DKI, DTI, WMTI and multimodal BAG_u and a second cluster containign BRIA, SMT, and SMTmc. However, DKI, BAG_u was more strongly correlated with full multimodal BAG_c than with other well-performing approaches DTI (Pearson's r_{diff}=0.03, p<.001) and WMTI (r_{diff}=0.03, p<.001). Vice versa, DTI BAG_c correlated strongest with WMTI BAG_c (r=0.905, p<.001).

Fig.2: Correlations of uncorrected BAG and age across used diffusion approaches

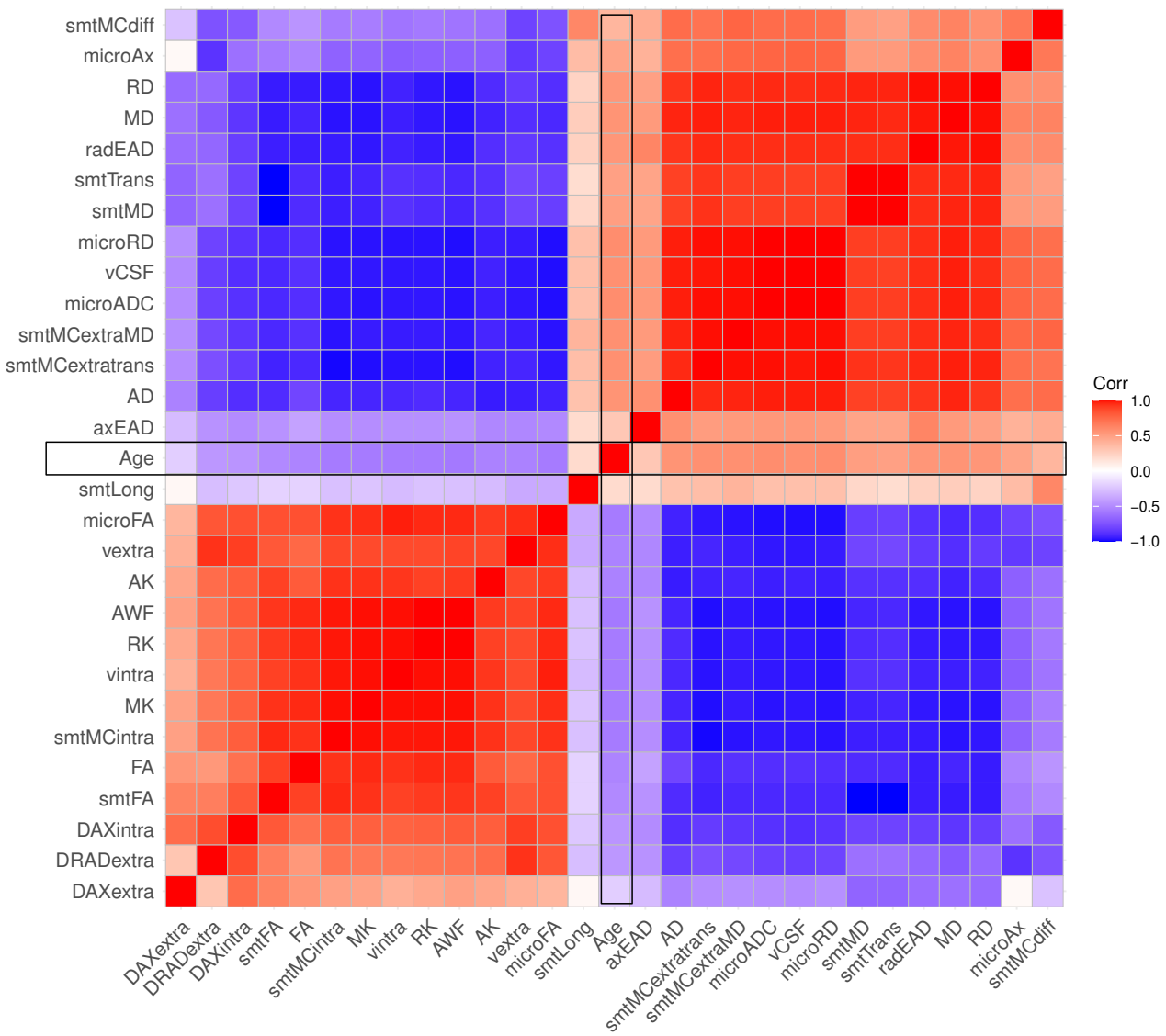


Age-BAG correlations, approximating 0, were not significant at $p_{Holm} \geq .05$. All other correlations were significant at $p_{Holm} < .001$. For the corrected BAG correlations across models see SF1.

Associations between diffusion metrics and age

A correlational analysis was used to demonstrate associations among Fornix diffusion metrics and age (**Fig.3**, including QC outliers: **SF4**). Association strengths ranged from to r=-0.997 (smtTrans and smtMCintra) to r=0.999 (smtTrans and smtMD). Correlations between fornix metrics and age ranged from r=-0.558 (smtMCintra) to r=0.570 (microRD).

Fig.3: Correlation matrix for fornix diffusion metrics and chronological age



All correlations were significant at Holm-corrected $p_{Holm} < .05$.

For region-wide associations between age and diffusion metrics, all diffusion metrics were correlated with age and displayed for p_{Holm}<0.001 (**Fig.4**). Among these correlations, Pearson's r values > 0.5 were name-labelled showing various correlations between diffusion metrics in the fornix and age.

However, when controlling for covariates, only relatively small proportions of the variance in single local and global diffusion metrics could be predicted from the whole model with small contributions of age to the models (**SF11**).

Fig.4: Correlations between diffusion metrics and age

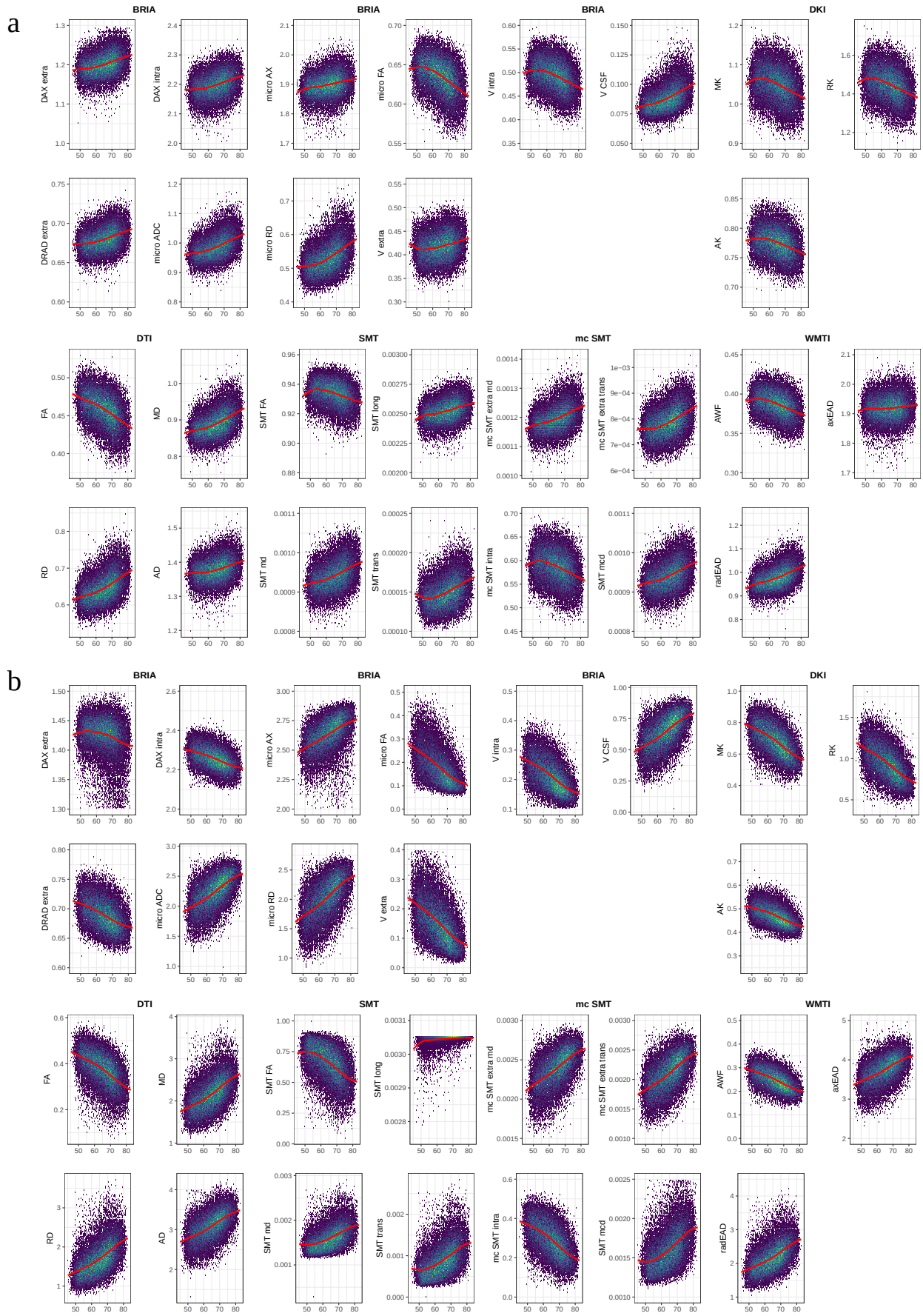
Note: Each point indicates one correlation between a diffusion metric and chronological age. Names of diffusion metrics are displayed when correlations between the metric and age reached a Pearson correlation of $|r|>0.5$. Holm correction[44] was used for FDR-correction, and all displayed values were significant at $p < .001$.
For the distribution of the correlations see **SF12**.

Age Trajectories of Diffusion Features

In **Fig.5** we present absolute diffusion metrics for the whole brain (**Fig.5a**) and fornix (**Fig.5b**) across ages for the examined six diffusion approaches (overview of metrics: **ST10**). Age-metric relationships for fornix were approximating linearity closer than more curvilinear global age-curves.

Several fornix-age relationships for BRIA extra-axonal and intra-axonal radial and axonal diffusivity opposed whole-brain-age relationships.

Fig.5: Whole-brain and fornix diffusion metrics across age



Note: The presented plots represent diffusion metrics for each of the six diffusion models from the full sample N = 35,749 for a) whole-brain diffusion metrics, b) fornix diffusion metrics. Brighter colours indicate higher density and red lines are fitted lines to the relationship between age and diffusion metric.

Whole-brain (**Fig.6**) and fornix (**SF9**) diffusion metrics M were predicted from age, sex and scanner site to create age curves (**Fig.5A-B**) which can be compared to raw *Z*-score-normalisation curves (**Fig.5C-D)**:

$$M = \beta_0 + \beta_1 Age + \beta_2 Age^2 + \beta_3 \times Site*Sex + \beta_4 Sex*Age + \beta_5 Sex + \beta_6 Site \quad (1)$$

A general trend was observed of most features crossing the mean at the same age, around 65 (**Fig.6**, **SF9**). Model fit metrics R^2_{adj} and Standard Error (SE) for the models accounting for age, sex and scanner site (Equation 1) when predicting diffusion metrics were calculated (**Fig.5E**). Highest SE, R^2_{adj} and variability across metrics was observed when predicting BRIA metrics (R^2_{adj} = .21), as well as lowest $R^2_{adj}\approx 0$ in BRIA Vextra, respectively. While DTI metrics could also be predicted well from the model, lowest variability in R^2_{adj} was found in WMTI and DKI. For fornix metrics, SE and R^2_{adj} was generally higher across diffusion approaches (**SF9**).

To test age-sensitivity of the mean features, likelihood ratio tests were conducted comparing models derived from Equation 1 against models derived from the same formula with age removed: (Equation 2).

$$M = \beta_0 + \beta_1 Site*Sex + \beta_2 Sex + \beta_3 Site \quad (2)$$

All models showed significant age dependence, with DTI RD (χ^2= 9,640.26, p_{Holm}<.001), BRIA microRD (χ^2=9,496.19, p_{Holm}<.001), and DTI FA (χ^2= 8,803.13, p_{Holm}<.001) being the most age-sensitive metrics, and WMTI axEAD (χ^2=6.66, p_{Holm}=.084), mcSMT diffusion coefficient (χ^2=238.47, p_{Holm}<.001), and WMTI radEAD(χ^2=418.26, p_{Holm}<.001) the least age-sensitive metrics (**ST11**).

In a set of additional analyses, we examined age-sensitivity of fornix features and whether the relationship between whole-brain as well as fornix diffusion metrics and age are better described as linear or non-linear. Fornix diffusion metrics were age sensitive (**SF9**) but model fit did not differ between linear and non-linear models for whole-brain (**ST12**) or fornix metrics (**ST9**).

Finally, to observe BAG-WM associations, principal components of regional and whole-brain WM metrics for each of the eight models (**Table 1**) were only weakly correlated with uncorrected BAG_u, and similarly related to corrected BAG_c, $age_{chronological}$ and $age_{predicted}$ (**SF10**). Furthermore, when predicting the most important WM components (**SF10**, **ST14**) or single regional or whole-brain metrics (**SF11**) from BAG_c and BAG_u and covariates, models predicted relatively small proportions of variance, with small contributions of BAG to the model (**SF10-11**).

Fig.6: Raw and predicted whole-brain WM diffusion metrics by chronological age

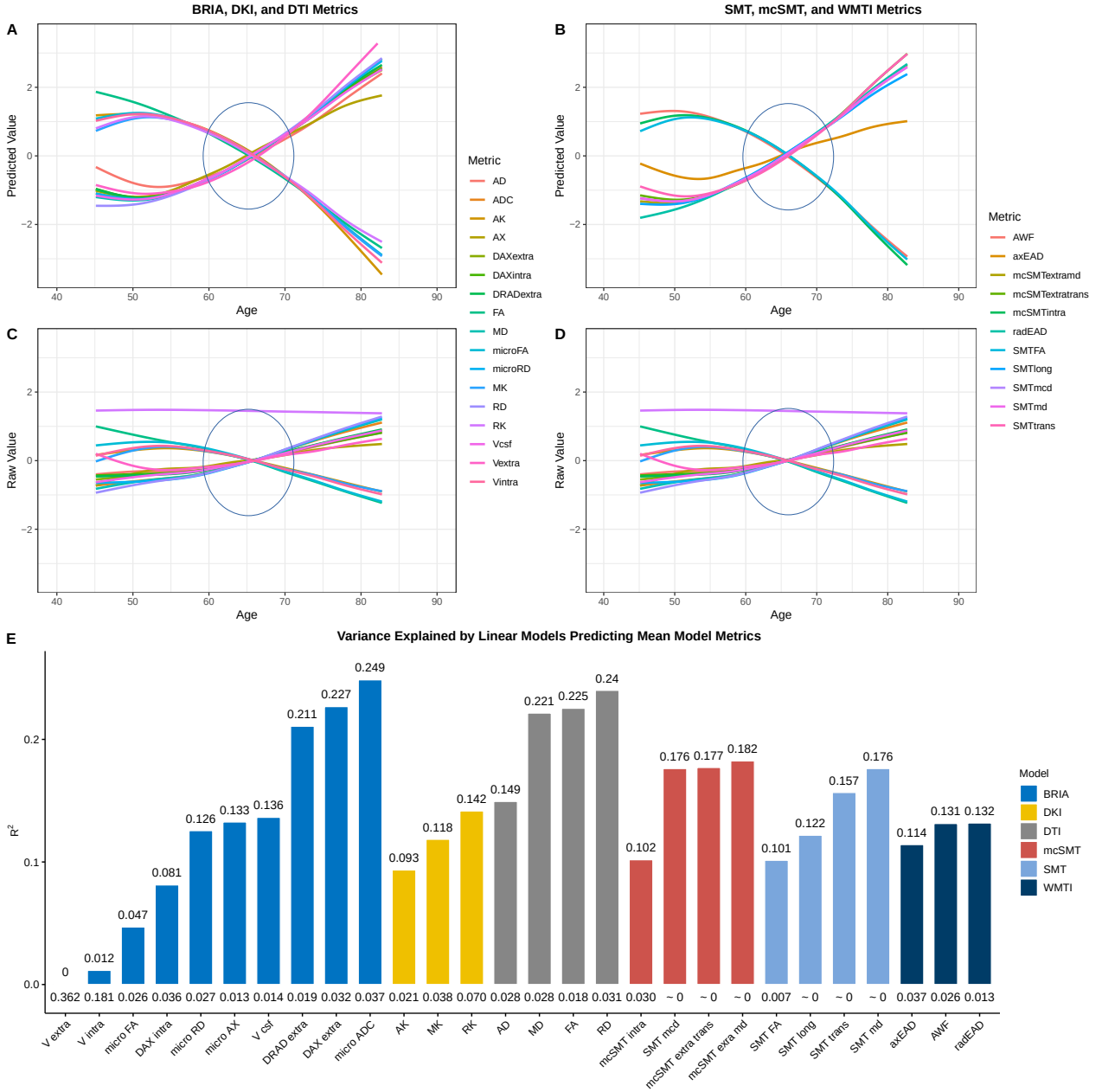


Fig.6A-D shows age curves for each standardised (z-score) diffusion metric's mean skeleton value (y-axis) plotted as a function of age (x-axis). Shaded areas represent 95% CI. Curves fitted to raw values (Fig.6 C-D) serve as a comparison to the lm-derived predicted values from Equation 1 (Fig.A-B). Fig.6E indicates the model fit for the linear models from Fig.6A-B, showing R^2_{adj} values on top and Standard Error (SE) on the bottom of the bars which each represent a Fornix skeleton value for one of the seven models. Lines crossing at age 65 are marked with ovals. Model summaries of all 28 mean models can be found in ST5. The same visualisation of fornix diffusion values can be found in SF9.

Discussion

We revealed that both conventional DTI and advanced diffusion approaches (WMTI, DKI, BRIA, SMT, mcSMT) perform consistently on brain age predictions, as indicated previously[25]. As a novel finding, our results show strong contributions of fornix microstructures explaining variance in age and reducing error for dMRI/WM-based brain age predictions, and model fit for brain age prediction models without fornix is reduced. Additionally, Mass-WM-age-correlations reveal strongest correlations between fornix microstructure and age. This suggest that the fornix is a key WM region of cross-sectional brain age, with fornix and whole-brain dMRI metrics' age trajectories following similar patterns such as steepening slopes at later ages.

On the other hand, there are multiple challenges related to fornix as a driver of brain age estimates, particularly multicollinearity, which might bias estimates of the importance of fornix (gain and permutation feature importance) for brain age predictions, and second, data processing artefacts. UKB offers diffusion data acquired with the most typical two-shell-diffusion protocol. Nevertheless, the standard diffusion model[66] based on differentiation of intra- and extra-axonal water pools could not be solved using this measurement strategy[66]. As a result, the derived diffusion metrics have both numerical uncertainties and the variability introduced from non-biological parameters[66]. Quantitative metrics derived from the different diffusion approaches allow to investigate such non-biological variability and to grade the subject variability in terms of used covariances. Yet, the aforementioned technical limitation might play a decisive role in a clinical context[50,66].

Besides obstacles resulting from modelling assumptions, our sample is cross-sectional in design and limited to adults older than forty, which, in turn, influences predictions and model evaluation metrics. Metrics such as r and R^2 are expected to be lower than in samples with wider age-ranges[45]. Additionally, the UKB imaging sub-sample shows better health than the non-imaging UKB subjects[67]. Another open question is the exact interpretation of BAG and its relationship with WM metrics, which was found to be small for principal WM components (**SF10**) and single diffusion metrics (**SF11**). Although previous research shows no relationship between the rate of change in longitudinal regional and global T_1-weighted-feature-retrieved BAG[18], further investigation of longitudinal as well as voxel-wise WM-derived BAG provide additional avenues to increase the interpretability of BAG.

We found the different diffusion metrics to be highly correlated (fornix, **Fig.3**), and show similar age trajectories (**SF9A-B**), which provokes the question of whether some of the metrics are redundant. The identification of redundant metrics and the combination of metrics across diffusion approaches is a matter of future research comparing diffusion approaches by probing them in practical settings such as in clinical samples[70].

Only few studies[56,57] address the fornix across ages. A possible reason is fornix' artefact-susceptibility induced from its proximity to the cerebrospinal-fluid, while being a small tubular region. Recent processing pipelines such as TBSS minimise such artefacts[80]. Yet, the influence of cerebrospinal-fluid artefacts in small tubular structures like the fornix remains unclear[68]. Fornix is a relatively small anatomical structure, and, for example, fornix BRIA cerebrospinal-fluid fraction is higher (vCSF>0.5) than global measures (vCSF>0.075), suggesting a presence of strong partial volume effect. In order to overcome such distorting effects, voxel-wise techniques are recommended, demanding the development of novel approaches incorporating techniques such as deep learning showing better performance than traditional ML, especially on large population samples[69].

Consistency across diffusion approaches

Overall, the results of brain age predictions are similar across diffusion approaches, with WMTI, DTI and DKI predicting age better than SMT, mcSMT and BRIA considering model fit and prediction-outcome correlations (**Table 1**). This finding could be explained in terms of diffusion approaches; i.e., the attempt to introduce more biophysically accurate parameters into the model might simultaneously reduce the general sensitivity of the used approaches to the tissue changes. Integrative approaches such as DTI or DKI are able to localise brain changes, however, without providing information about the underlying mechanisms. Our study support a previous study with a smaller but more age-differentiated sample (n=702) of DTI and WMTI being superior to mcSMT at brain age predictions in terms of model performance[25]. When examining additional diffusion models on a larger sample, we find DKI metrics to have higher predictive power than in Beck and colleagues[25]. Simultaneously, differences between diffusion approaches, and both variance explained and prediction error (RMSE, MAE) were smaller in this study. These differences are likely due to the narrower age range in our study[45], whereas our

significantly larger sample emphasises the reliability of our findings.

While brain age predictions from single diffusion approaches were grossly similar, predictions from combined approaches were best (**Table 1**). Correlations between predicted and chronological age were consistent across diffusion approaches, as differences between correlations were small (**Fig.1**, **SF8**). This shows that addressing a wider range of WM characteristics improves predictive models compared to models with single diffusion approach metrics (e.g., only DTI), which would be intuitive when considering BAG as a general indicator of health[10,13,19,20].Vice versa, reducing spatial specificity by averaging diffusion metrics across all WM reduced prediction accuracy. Conventionally used DTI on its own is limited in its ability to present biophysically meaningful measures of the underlying microstructure. As a result, the advanced modelling is recalled including intra- and extra-axonal spaces and tissue peculiarities being influenced by individual differences in myelin and fibre architecture (crossing/bending fibres, and axonal characteristics)[25]. Hence, adding additional information to DTI better allow to infer the underlying neurobiology of tissue, for example, expressed in differential WM-age-dependences (**Fig.5-7**) or brain age predictions (**Table 1**)[25].

We observed that BAG exhibits strong correlations across all diffusion approaches (**Fig.2,SF1**). Congruently with the correlational differences (**Fig.1, SF8**), BAG based on averaged skeleton values was least correlated to all other diffusion approaches (**Fig.2**), indicating inferiority of global compared to region-wide approaches. BAG obtained from WMTI, DTI and DKI were closest related to BAG from the multimodal approach (which predicted age best), both for age-bias corrected and uncorrected BAG (**Fig.2, SF1**). This is in agreement with the observed age-prediction model performance (**Table 1**). BAG correlations were observed in three clusters: 1) WMTI and DTI, 2) mcSMT, SMT, BRIA, and 3) DKI, indicative of similar measurements within these clusters (**Fig.2, SF1**). To a certain extent, these clusters reflect similarities in the underlying mathematics of the clustering diffusion approaches. For example, mcSMT and SMT are closely related models[37], whereas DKI's non-Gaussianity might reveal another quality of age-sensitive WM microstructures not captured by the other approaches[46]. Additionally, the cluster differences indicate that the observed diffusion approaches measure different age(ing)-sensitive characteristics, supporting the argument for a combination of diffusion approaches when assessing the ageing brain.

Age trajectories and fornix as a brain age feature

Based on the presented findings on fornix, we further investigate details of fornix, keeping discussed limitations to the generalizability of the findings in mind. Diffusion metrics describing fornix microstructure were consistently related to each other and age across all diffusion approaches in two clusters. Values were positively correlated within each cluster and negatively between clusters (see **Fig.3**). In the first cluster, different approaches' FA, kurtosis metrics (MK, RK, AK), water fractions (vintra and vextra from BRIA and AWF from WMTI), and BRIA intra-axonal and extra-axonal radial and axial diffusivity were positively correlated. The second cluster, which was negatively related to the first cluster but positive to age, contained metrics of mean, axial and radial diffusivity, and cerebrospinal-fluid fraction of the different diffusion approaches, which were positively related to each other. Interestingly, both clusters consisted of unit-less values, for example water fractions, and diffusivities, which might have the same meaning as extra-axonal axial diffusivities from different diffusion approaches, for example BRIA vs SMTmc. Such consistencies of similar metrics across diffusion approaches were more apparent for the fornix when QC-identified outliers were removed (compare **Fig.3** and **SF4**), which supports the reliability of our findings of fornix-age-dependencies. Furthermore, fornix metrics were most strongly related to age across diffusion approaches (**Fig.4, SF11**), supporting the importance of fornix in reducing error of brain age predictions (**Table 2**). Not surprisingly, all fornix features were age-sensitive (**ST4**), and more age sensitive than whole-brain metrics (**ST11**). Whole-brain trajectories are in

agreement with previous results, showing-age sensitivity of various mean diffusion metrics[25], and the same directionality of age trajectories of metrics for DTI[9,33], mcSMT, DKI, WMTI[25].

We displayed that fornix microstructure measures have differential behaviours across diffusion approaches (**Figs.5-6**). Focussing on absolute diffusion values (**Fig.5**), it can be observed that diffusion measures which are correlated (**Figs.2-3**) exhibit similar age dependences. Additionally, slopes of fornix compared to whole-brain diffusion metrics were generally steeper and closer approximating linearity, indicating stronger changes, such as quicker WM degeneration in the fornix compared to the whole-brain average (see **Fig.5**). Particularly BRIA metrics show visually detectable differences between the fornix and the whole brain (**Fig.5**, DAXextra, DAXintra, DRADextra, Vextra); as opposed to global developments, fornix intra and extra-axonal diffusion decreased, indicating fornix shrinkage with increasing age. Periventricular shrinkage is linked to enlarging ventricles[47], which has been related to ageing and neurodegenerative disorder progression[48]. This effect was observed by a positive relationship between age and cerebrospinal fluid (CSF) fraction in BRIA. Another metric which revealed larger differences in the fornix than for the whole-brain average was intra-axonal water fractions, which can be treated as a proxy for the axonal density, decreased with increasing age (see **Fig.5**, BRIA:Vintra; SMTmc:intra; WMTI:AWF) while the CSF fraction (BRIA) increases. Such WM microstructure changes are not only directly linked to different neurobiological features but can be markers of clinical outcomes, such as dementia[49,50].

A selection of metrics is comparable across diffusion approaches taking DTI as reference point, showing similar age trajectories. DTI metrics AD, RD, and MD tend to increase over the lifespan and FA tends to decrease across brain regions (**Figs.6-7**)[25,33,51,52] as well as in fornix (**Fig.5b**, **SF9**), implying processes such as de-myelination, changes in axonal and general WM integrity. Such DTI age-dependences are reflected by according BRIA, SMT, and WMTI metrics, whereas DKI shows opposite age-relationships, as presented previously[25]. Deterioration effects, measured by the age-dependency of axonal water fractions, were generally stronger in fornix compared to whole-brain metrics (**Fig.5**). Interestingly, opposed to global metrics, radial diffusivity measures from DKI and BRIA (DRADextra) decreased in fornix (**Fig.5**), suggesting higher fornix than global plasticity, potentially being an antecedent of age-related hippocampal changes[55].

Additional, unique information about age dynamics was presented by standardised scores accounting for age, sex and scanner site and standardised uncorrected scores across ages (**Fig.6**, **SF9**). After standardisation and accounting for covariates, most fornix metrics follow a tightly resembling near-linear trend either increasing or decreasing by age (**SF9A-B**), as opposed to whole-brain metrics which follow a rather curvilinear line, as previously shown[25,33,52]. Diffusion metrics' variance explained across models indicates fornix metrics to be more sensitive to a combination of covariates age, sex, and scanner site than whole-brain metrics (**Fig.6, SF9**). In the fornix, only BRIA extra-axonal axial diffusivity (DAX extra) and the SMT longitudinal diffusion coefficient (SMT long) showed non-linear trajectories, however, both measures are weakly correlated to other diffusion parameters (**Fig.2**). Yet, when comparing model metrics such as variance explained of linear and non-linear models predicting fornix and whole-brain diffusion metrics from age, sex and scanner site and their interactions, there were no apparent differences between models (**ST9, ST12**). This implies that contrary to previous research observing the entire lifespan presenting curvilinear DTI age trajectories[25,33], or trends towards curvilinearity (with yet better linear fit for selected regions)[52], we found that fornix and whole-brain age trajectories from age 40 can be described as linear when accounting for covariates sex, age, and scanner site. While the crossing of the x-axis at age 65 (**Fig.6**, **SF9**) is a reflection of the sample's age distribution (**Fig.7**), in addition to the shapes of the different age-trajectories, it reveals that the different diffusion approaches are similarly age-sensitive or measure similar underlying ageing-related changes. For whole-brain metrics, changes become exacerbated from 65

onwards (**Fig.7**), with reasons potentially laying in an accelerated neurodegeneration also reflected in the exponentially increasing risk to develop neurodegenerative disorders from age 65 onwards[53]. For example, in the USA, 3% of 65-74 year olds, 17% of the 75-84 year olds, and 32% of those ag 85+ developed Alzheimer's dementia[54]. Subclinical or preclinical states are, however, not captured by these approximations, and WM changes usually precede clinical detections, making WM monitoring a promising tool for early detection.

Beyond WM, fornix changes seem to play an important role for GM changes, particularly in the hippocampus: for example, fornix glia damages lead to hippocampal GM atrophy[55]. This might be reflected by dis-connectivity of fornix with other brain regions as described by decreasing extra axonal space coefficients (**Fig.5b**), and following changes in fornix function. Potentially, the consequences of age-related fornix changes thereby affect functionality of a selection of brain regions, such as the hippocampus. While several studies have presented ageing-related fornix microstructure changes in humans[56,57] and monkeys[58] in small samples, only one large-scale study revealed findings connected to the fornix, namely strongest default mode network GM volume covariation with fornix WM microstructure[59]. This suggests that fornix, a key connector of the limbic system with the cortex, might also be critical for default mode network functioning. Moreover, memory and episodic recall have been related to fornix[60]. Hence, fornix changes might play an important role in known ageing-dependent temporal lobe changes, and specifically hippocampal changes for ageing-related pathological developments[61–64]. Previous studies presented age-related fornix DTI metric changes[55–57] which potentially appear prior to hippocampal volume changes[55,56], and are related to declining episodic memory performance[55]. Hence, fornix changes potentially serve to predict future pathological development, suggesting WM changes in the fornix as a potential ageing biomarker and therapeutic target. This supports previous findings showing network re-activations, metabolic and GM changes after fornix deep-brain-stimulation antagonising the progression of neurodegenerative disorders[65].

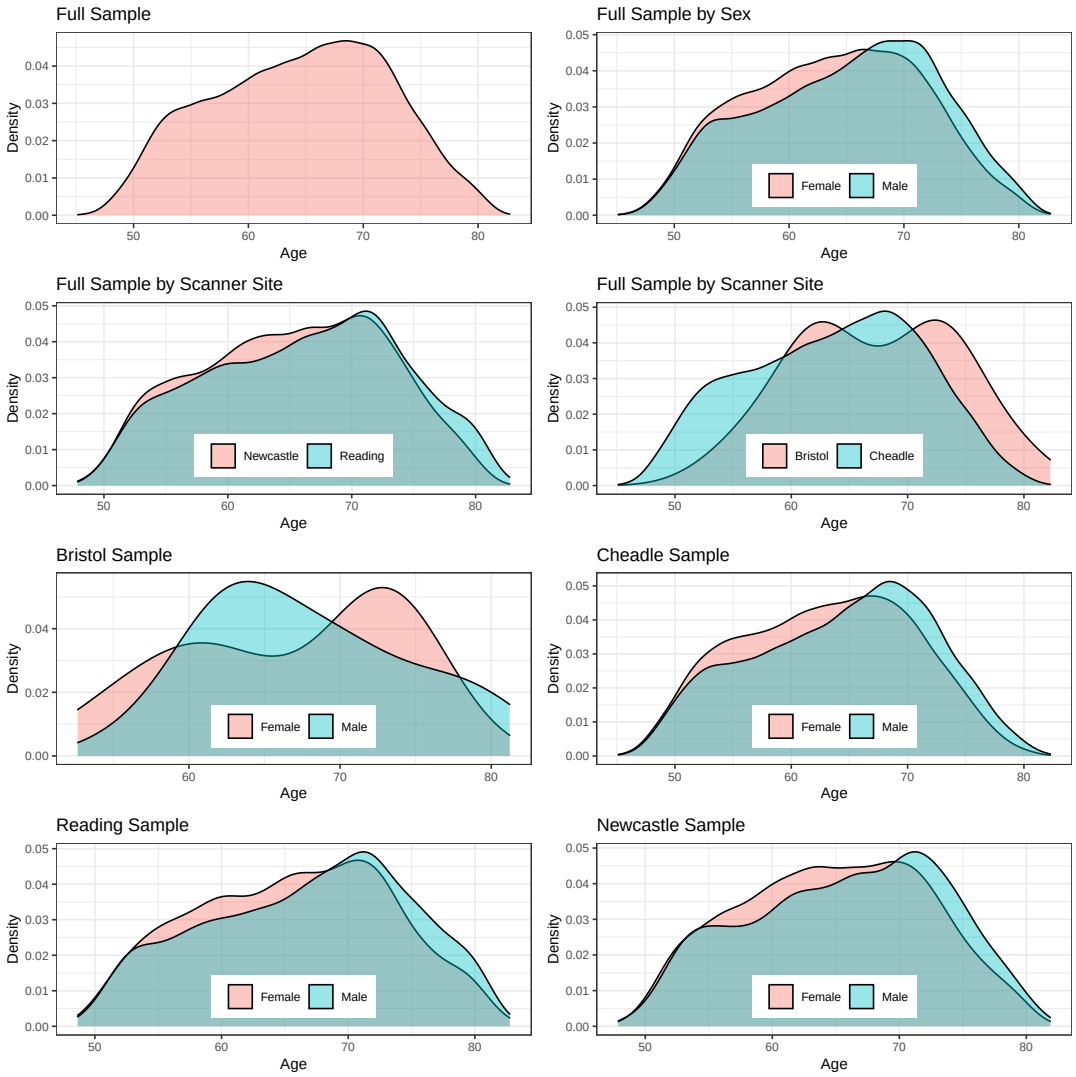
The current study gives for the first time a detailed account on region-wise-to-global WM-age relationships for multiple diffusion approaches in a representative sample, and highlights fornix as an important structure for age predictions across diffusion approaches. Brain age was estimated best when combining approaches, showing different aspects of WM to contribute to brain age with fornix being the central region for these predictions.

Methods

Sample characteristics

The original UKB[5] diffusion MRI data consisted of N = 42,208 participants. After exclusions, based on later withdrawn consent and an ICD-10 diagnosis from categories F, G, I, and stroke (excluded: N = 3,521), and data sets not meeting quality control standards (N = 2,938) using the YTTRIUM method[39], we obtained a final sample consisting of 35,749 healthy adults (age range 44.57 to 82.75, M_{age} = 64.46, SD_{age} = 7.62, Md_{age} = 64.97; 52.96% females, 47.04% males). Participants were recruited and scanned at four different sites: 57.62% in Cheadle, 26.30% in Newcastle, 15.96% in Reading, and 0.12% in Bristol (**Fig.7**).

Fig.7: Density plots for the sample's age by sex and scanner site



MRI acquisition, diffusion pipeline and TBSS analysis

UKB MRI data acquisition procedures are described elsewhere[5].

Diffusion data preprocessing was conducted as described in Maximov et al.[71], using an optimised pipeline which includes corrections for noise[72], Gibbs ringing[73], susceptibility-induced and motion distortions, and eddy current artefacts[74]. Isotropic Gaussian smoothing was carried out with the FSL[75] function *fslmaths* with a Gaussian kernel of 1 mm^3. After that DTI, DKI, and WMTI metrics were estimated using Matlab 2017b[76]. Employing the multi-shell data, DKI and WMTI metrics were estimated using Matlab code (https://github.com/NYU-DiffusionMRI/DESIGNER)[36]. SMT, and mcSMT metrics were estimated using original code (https://github.com/ekaden/smt)[37], as well as Bayesian estimates / BRIA were estimated by the original Matlab code (https://bitbucket.org/reisert/baydiff/src/master/)[34].

In total, we obtained 28 metrics from six diffusion approaches (DTI, DKI, WMTI, SMT, mcSMT, BRIA)[25,38,71,77–79]. In order to normalise all metrics, we used tract-based spatial statistics (TBSS)[80], as part of FSL[81]. In brief, initially all BET-extracted[82] FA images were aligned to MNI space using non-linear transformation (FNIRT)[75]. Afterwards, the mean FA image and related mean FA skeleton were derived. Each diffusion scalar map was projected onto the mean FA skeleton using the TBSS procedure. In order to provide a quantitative description of diffusion metrics we evaluated averaged values over the skeleton and two white matter atlases, namely the JHU atlas[83] and the JHU tractographic atlas[84]. Finally, we obtained 20 WM tracts and 48 regions of interest (ROIs) based on a probabilistic white matter atlas (JHU) (Hua et al., 2008) for each of the 28 metrics, including

the mean skeleton values. Altogether, 1932 features per individual were derived (28 metrics * (48 ROIs + 1 skeleton mean + 20 tracts); see number of dMRI features in **Table 1**)).

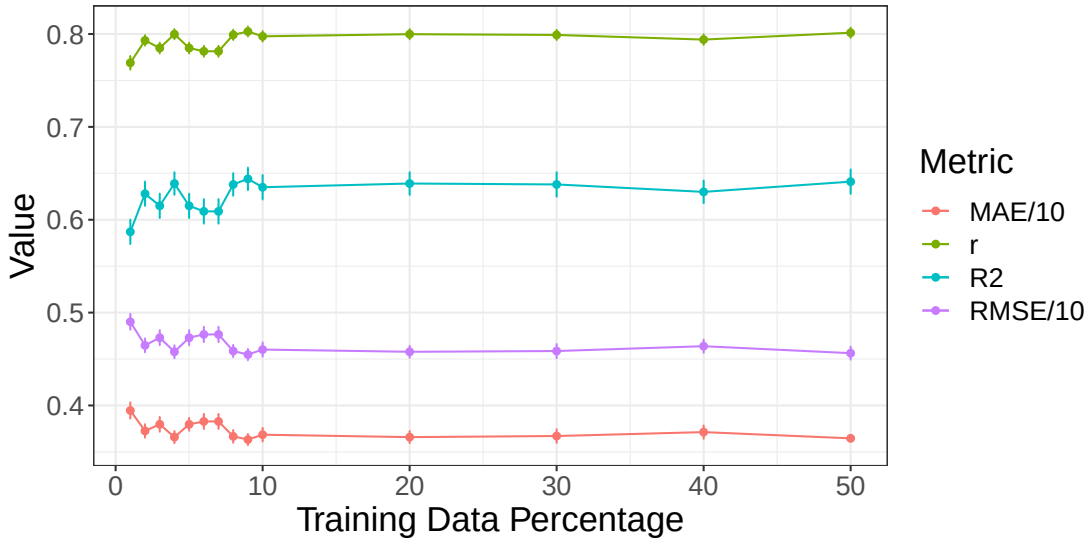
Statistical Analyses

All statistical analyses were carried out using Python, version 3.7.1 and R, version 3.6.0 (www.r-project.org/). *p*-values were adjusted for multiple comparison using Holm correction[44].

Brain Age Predictions

First, brain age predictions were performed using XGBoost[24] in Python. To evaluate how much training data was needed for accurate brain age predictions from all 1940 brain features, we divided the full dataset (N=35,749) into two equal parts: one pre-test set and one pre-training set for independent parameter-tuning. From the pre-training set, data was randomly sampled for sub-samples consisting of 358, 715, 1,073, 1,430, 1,788, 2,145, 2,503, 2,860, 3,218, 3,575, 7,150, 10,725, 14,300, or 17,875 participants, corresponding to 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40% and 50% of the total subjects, respectively (**Fig.8**). Models were trained on these sub-samples and then tested on the remaining half, i.e., the 50% pre-test sample, using 10-fold cross validation showing model performance to not further improve past the 10% training data mark, informing our training-test-split (**Fig.8, ST1**).

Fig.8: Model performance for different train-test splits



Model metrics R^2, RMSE, MAE and their standard deviations, as well as the Pearson's correlations between predicted and chronological age and its 95% confidence interval are displayed for different training data percentages of the total data (x-axis). For visualisation purposes, RMSE and MAE were divided by 10. For exact values see Suppl. Table ST1.

Second, in order to compare the different diffusion approaches, based on the previous steps, the training-test split was fixed at previously used 10% training data (N = 3,575) and 90% test data (N = 32,174) which indicated a best fit at a learning rate = 0.05, max layers/depth = 3 and number of trees = 750. These tuned parameters were used for 10-fold cross-validations brain age predictions on the test data of all six individual models, one multimodal model combining all metrics from all diffusion models, and one multimodal model using only mean values from all diffusion models (**Table 1**).

Third, BAG was calculated as the difference between chronological age Ω and predicted age P:

$$BAG_{uncorrected} = P - \Omega \quad (3)$$

As a supplement, age-bias-corrected predicted age was calculated from the intercept and slope of age predictions as previously described[26,85]:

$$P = \alpha \times \Omega + \beta \quad (4)$$
$$BAG_{corrected} = (P + [\Omega - (\alpha \times \Omega + \beta)]) - \Omega \quad (5)$$

P represents predicted age modelled from chronological age Ω, with intercept β and slope α. This age-bias correction allowed for a bias-corrected BAG estimate.

Data Availability
All raw data are available from the UKB[5] (www.ukbiobank.ac.uk). Synthetic datasets with the synthpop[89] R package based on the original data for all six diffusion approaches (resulting in six datasets) to run the code are openly available at the Open Science Framework: (https://osf.io/nv8ea/). Synthetic datasets are simulated datasets closely mimicking the statistical characteristics of the original data while protecting data privacy and anonymity.

Code Availability
Code needed to run brain age predictions in Python, and for all analyses and visualisations in R is available at the Open Science Framework: (https://osf.io/nv8ea/).

Acknowledgements
This research was funded by the Research Council of Norway (#223273) This study has been conducted using UKB data under Application 27412. UKB has received ethics approval from the National Health Service National Research Ethics Service (ref 11/NW/0382). The work was performed on the Service for Sensitive Data (TSD) platform, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo IT-Department (USIT). Computations were performed using resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway. Finally, we want to thank all UKB participants and facilitators who made this research possible.

Author contributions
M.K.: Study design, Software, Formal analysis, Visualisations, Project administration, Writing – original draft, Writing – review & editing
A.M.d.L.: Software, Writing – review & editing
D.v.d.M.: Software, Writing – review & editing
A.L.: Writing – review & editing, Funding acquisition
E.E.: Writing – review & editing
D.B.: Writing – review & editing
O.A.A.: Writing – review & editing, Funding acquisition
L.W.: Writing – review & editing, Funding acquisition
I.I.M. supervision, Study design, Data pre-processing and quality control, Writing – review & editing, Funding acquisition

Conflicts of Interesting
The authors have no conflicts of interest to disclaim.

References

1. Grady, C. L. The cognitive neuroscience of ageing. *Nat. Rev. Neurosci.* **13**, 491–505 (2012).
2. Symms, M., Jäger, H. R., Schmierer, K. & Yousry, T. A. A review of structural magnetic resonance neuroimaging. *J. Neurol. Neurosurg. Psychiatry* **75**, 1235–1244 (2004).
3. Grady, C. The cognitive neuroscience of ageing. *Nat. Rev. Neurosci. 2012 137* **13**, 491–505 (2012).
4. Wrigglesworth, J. *et al.* Factors associated with brain ageing - a systematic review. *BMC Neurol. 2021 211* **21**, 1–23 (2021).
5. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* **12**, e1001779 (2015).

6. Van Essen, D. C. *et al.* The Human Connectome Project: a data acquisition perspective. *NeuroImage* **62**, 2222–2231 (2012).
7. Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nat. 2022 6037902* **603**, 654–660 (2022).
8. Lawrence, K. E. *et al.* Age and sex effects on advanced white matter microstructure measures in 15,628 older adults: A UK biobank study. *Brain Imaging Behav.* **15**, 2813–2823 (2021).
9. Cox, S. R. *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat. Commun. 2016 71* **7**, 1–13 (2016).
10. Kaufmann, T. *et al.* Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617 (2019).
11. Franke, K. & Gaser, C. Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained? *Front. Neurol.* **10**, (2019).
12. Cole, J. H. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. *Neurobiol. Aging* **92**, 34–42 (2020).
13. Leonardsen, E. H. *et al.* Deep neural networks learn general and clinically relevant representations of the ageing brain. *medRxiv* (2021) doi:10.1101/2021.10.29.21265645.
14. Cole, J. H. & Franke, K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences* vol. 40 681–690 Preprint at https://doi.org/10.1016/j.tins.2017.10.001 (2017).
15. Rokicki, J. *et al.* Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Hum. Brain Mapp.* **42**, 1714–1726 (2021).
16. de Lange, A. M. G. *et al.* Multimodal brain-age prediction and cardiovascular risk: The Whitehall II MRI sub-study. *NeuroImage* **222**, 117292 (2020).
17. de Lange, A. M. G. *et al.* Prominent health problems, socioeconomic deprivation, and higher brain age in lonely and isolated individuals: A population-based study. *Behav. Brain Res.* **414**, (2021).
18. Vidal-Pineiro, D. *et al.* Individual variations in ’brain age’ relate to early-life factors more than to longitudinal brain change. *elifesciences.org* (2021) doi:10.7554/eLife.
19. Beck, D. *et al.* Cardiometabolic risk factors associated with brain age and accelerate brain ageing. *Hum. Brain Mapp.* **43**, 700–720 (2022).

20. Cole, J. H. *et al.* Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* **163**, 115–124 (2017).
21. Dosenbach, N. U. F. *et al.* Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361 (2010).
22. Franke, K., Ziegler, G., Klöppel, S. & Gaser, C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage* **50**, 883–892 (2010).
23. Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C. & Mechelli, A. Machine learning for brain age prediction: Introduction to methods and clinical applications. *eBioMedicine* **72**, 103600 (2021).
24. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **13-17-August-2016**, 785–794 (2016).
25. Beck, D. *et al.* White matter microstructure across the adult lifespan: A mixed longitudinal and cross-sectional study using advanced diffusion models and brain-age prediction. *NeuroImage* **224**, 117441 (2021).
26. de Lange, A. M. G. *et al.* Population-based neuroimaging reveals traces of childbirth in the maternal brain. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22341–22346 (2019).
27. Salih, A. *et al.* Brain age estimation at tract group level and its association with daily life measures, cardiac risk factors and genetic variants. *Sci. Rep. 2021 111* **11**, 1–14 (2021).
28. de Lange, A. M. G. *et al.* Multimodal brain-age prediction and cardiovascular risk: The Whitehall II MRI sub-study. *NeuroImage* **222**, 117292 (2020).
29. Chen, C. Le *et al.* Generalization of diffusion magnetic resonance imaging–based brain age prediction model through transfer learning. *NeuroImage* **217**, 116831 (2020).
30. Richard, G. *et al.* Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ* **2018**, (2018).
31. Basser, P. J., Mattiello, J. & LeBihan, D. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* **66**, 259–267 (1994).
32. Novikov, D. S., Fieremans, E., Jespersen, S. N. & Kiselev, V. G. Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation. *NMR Biomed.* **32**, e3998 (2019).

33. Westlye, L. T. *et al.* Life-Span Changes of the Human Brain White Matter: Diffusion Tensor Imaging (DTI) and Volumetry. *Cereb. Cortex* **20**, 2055–2068 (2010).
34. Reisert, M., Kellner, E., Dhital, B., Hennig, J. & Kiselev, V. G. Disentangling micro from mesostructure by diffusion MRI: A Bayesian approach. *NeuroImage* **147**, 964–975 (2017).
35. Jensen, J. H., Helpern, J. A., Ramani, A., Lu, H. & Kaczynski, K. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn. Reson. Med.* **53**, 1432–1440 (2005).
36. Fieremans, E., Jensen, J. H. & Helpern, J. A. White matter characterization with diffusional kurtosis imaging. *NeuroImage* **58**, 177–188 (2011).
37. Kaden, E., Kruggel, F. & Alexander, D. C. Quantitative mapping of the per-axon diffusion coefficients in brain white matter. *Magn. Reson. Med.* **75**, 1752–1763 (2016).
38. Kaden, E., Kelm, N. D., Carson, R. P., Does, M. D. & Alexander, D. C. Multi-compartment microscopic diffusion imaging. *NeuroImage* **139**, 346–359 (2016).
39. Maximov, I. I. *et al.* Fast qualitY conTrol meThod foR derIved diffUsion Metrics (YTTRIUM) in big data analysis: U.K. Biobank 18,608 example. *Hum. Brain Mapp.* **42**, 3141–3155 (2021).
40. Raghavan, S. *et al.* Diffusion models reveal white matter microstructural changes with ageing, pathology and cognition. *Brain Commun.* **3**, (2021).
41. Hotelling, H. Relations Between Two Sets Of Variates. *Biometrika* **28**, 321–377 (1936).
42. Zou, G. Y. Toward Using Confidence Intervals to Compare Correlations. *Psychol. Methods* **12**, 399–413 (2007).
43. XBGoost Developers. XGBoost Documentation - Introduction to Boosted Trees. https://xgboost.readthedocs.io/en/latest/tutorials/model.html (2021).
44. Sture Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
45. de Lange, A. M. G. *et al.* Mind the gap: Performance metric evaluation in brain-age prediction. *Hum. Brain Mapp.* **43**, 3113–3129 (2022).
46. De Santis, S., Gabrielli, A., Palombo, M., Maraviglia, B. & Capuani, S. Non-Gaussian diffusion imaging: a brief practical review. *Magn. Reson. Imaging* **29**, 1410–1416 (2011).
47. Kwon, Y. H., Jang, S. H. & Yeo, S. S. Age-related changes of lateral ventricular width and periventricular white matter in the human brain: a diffusion tensor imaging study. *Neural Regen. Res.* **9**, 986 (2014).

48. Pinaya, W. H. L. *et al.* Using normative modelling to detect disease progression in mild cognitive impairment and Alzheimer's disease in a cross-sectional multi-cohort study. *Sci. Rep.* **11**, 15746 (2021).
49. Meeter, L. H., Kaat, L. D., Rohrer, J. D. & Van Swieten, J. C. Imaging and fluid biomarkers in frontotemporal dementia. *Nat. Rev. Neurol.* **13**, 406–419 (2017).
50. Thomas, A. G., Koumellis, P. & Dineen, R. A. The fornix in health and disease: An imaging review. *Radiographics* **31**, 1107–1121 (2011).
51. Cox, S. R. *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat. Commun.* **7**, (2016).
52. Davis, S. W. *et al.* Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *NeuroImage* **46**, 530 (2009).
53. Nichols, E. *et al.* Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105–e125 (2022).
54. 2020 Alzheimer's disease facts and figures. *Alzheimers Dement.* **16**, 391–460 (2020).
55. Metzler-Baddeley, C. *et al.* Fornix white matter glia damage causes hippocampal gray matter damage during age-dependent limbic decline. *Sci. Rep. 2019 91* **9**, 1–14 (2019).
56. Chen, D. Q., Strauss, I., Hayes, D. J., Davis, K. D. & Hodaie, M. Age-related changes in diffusion tensor imaging metrics of fornix subregions in healthy humans. *Stereotact. Funct. Neurosurg.* **93**, 151–159 (2015).
57. Christiansen, K. *et al.* The status of the precommissural and postcommissural fornix in normal ageing and mild cognitive impairment: An MRI tractography study. *NeuroImage* **130**, 35–47 (2016).
58. Peters, A., Sethares, C. & Moss, M. B. How the primate fornix is affected by age. *J. Comp. Neurol.* **518**, 3962–3980 (2010).
59. Kernbach, J. M. *et al.* Subspecialization within default mode nodes characterized in 10,000 UK Biobank participants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12295–12300 (2018).
60. Senova, S., Fomenko, A., Gondard, E. & Lozano, A. M. Anatomy and function of the fornix in the context of its potential as a therapeutic target. *J. Neurol. Neurosurg. Psychiatry* **91**, 547–559 (2020).

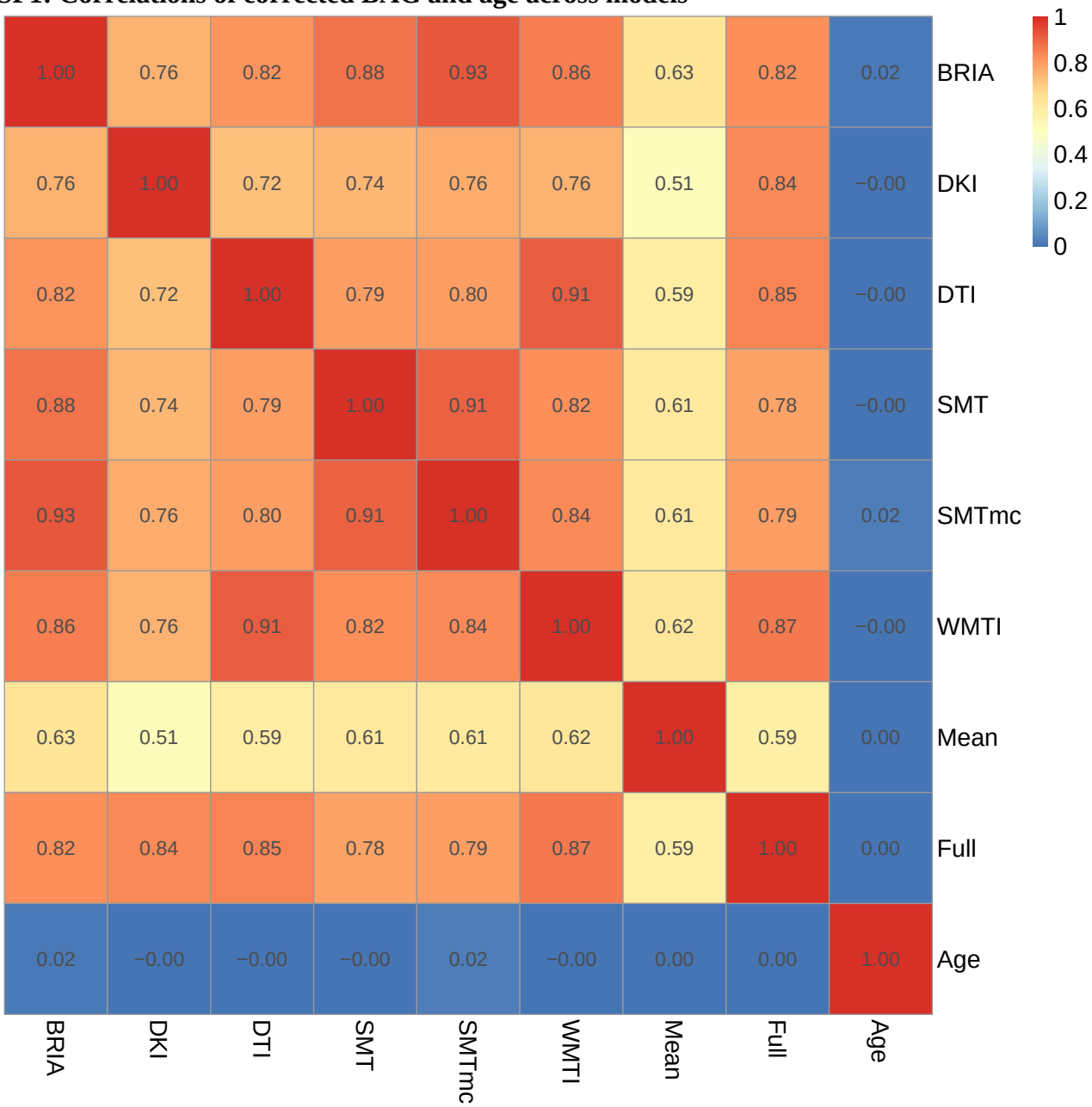
61. Hedden, T. & Gabrieli, J. D. E. Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci. 2004 52* **5**, 87–96 (2004).
62. Burke, S. N. & Barnes, C. A. Neural plasticity in the ageing brain. *Nat. Rev. Neurosci. 2006 71* **7**, 30–40 (2006).
63. Pluvinage, J. V. & Wyss-Coray, T. Systemic factors as mediators of brain homeostasis, ageing and neurodegeneration. *Nat. Rev. Neurosci.* **21**, 93–102 (2020).
64. Cabeza, R. *et al.* Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nat. Rev. Neurosci.* doi:10.1038/s41583-018-0068-2.
65. Jakobs, M., Lee, D. J. & Lozano, A. M. Modifying the progression of Alzheimer's and Parkinson's disease with deep brain stimulation. *Neuropharmacology* **171**, 107860 (2020).
66. Novikov, D. S., Kiselev, V. G. & Jespersen, S. N. On modeling. *Magn. Reson. Med.* **79**, 3172–3193 (2018).
67. Lyall, D. M. *et al.* Quantifying bias in psychological and physical health in the UK Biobank imaging sub-sample. *Brain Commun.* **4**, (2022).
68. Bach, M. *et al.* Methodological considerations on tract-based spatial statistics (TBSS). *NeuroImage* **100**, 358–369 (2014).
69. Popescu, S. G., Glocker, B., Sharp, D. J. & Cole, J. H. Local Brain-Age: A U-Net Model. *Front. Aging Neurosci.* **13**, 838 (2021).
70. Kantarci, K. Fractional anisotropy of the fornix and hippocampal atrophy in Alzheimer's disease. *Front. Aging Neurosci.* **6**, 316 (2014).
71. Maximov, I. I., Alnæs, D. & Westlye, L. T. Towards an optimised processing pipeline for diffusion magnetic resonance imaging data: Effects of artefact corrections on diffusion metrics and their age associations in UK Biobank. *Hum. Brain Mapp.* **40**, 4146–4162 (2019).
72. Veraart, J., … E. F.-M. resonance in & 2016, undefined. Diffusion MRI noise mapping using random matrix theory. *Wiley Online Libr.* **76**, 1582–1593 (2016).
73. Kellner, E., Dhital, B., … V. K.-M. resonance in & 2016, undefined. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Wiley Online Libr.* **76**, 1574–1581 (2016).
74. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* **125**, 1063–1078 (2016).

75. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
76. Mathworks. Matlab version 9.3.0.713579 (R2017b). Preprint at (2017).
77. Benitez, A., Jensen, J. H., Falangola, M. F., Nietert, P. J. & Helpern, J. A. Modeling white matter tract integrity in aging with diffusional kurtosis imaging. *Neurobiol. Aging* **70**, 265–275 (2018).
78. Hope, T. R. *et al.* Diffusion tensor and restriction spectrum imaging reflect different aspects of neurodegeneration in Parkinson's disease. *PLOS ONE* **14**, e0217922 (2019).
79. Pines, A. R. *et al.* Leveraging multi-shell diffusion for studies of brain development in youth and young adulthood. *Dev. Cogn. Neurosci.* **43**, 100788 (2020).
80. Smith, S. M. *et al.* Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage* **31**, 1487–1505 (2006).
81. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219 (2004).
82. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
83. Mori, S., Wakana, S., Zijl, P. Van & Nagae-Poetscher, L. MRI atlas of human white matter. (2005).
84. Hua, K. *et al.* Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage* **39**, 336–347 (2008).
85. de Lange, A. M. G. & Cole, J. H. Commentary: Correction procedures in brain-age prediction. *NeuroImage Clin.* **26**, (2020).
86. Hotelling, H. The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters. *https://doi.org/10.1214/aoms/1177731867* **11**, 271–283 (1940).
87. Schuirmann, D. L. On hypothesis-testing to determine if the mean of a normal-distribution is contained in a known interval. *Biometrics* **37**, 617 (1981).
88. Kirkwood, T. B. L. & Westlake, W. J. Bioequivalence Testing -- A Need to Rethink. *Biometrics* **37**, 589 (1981).
89. Nowok, B., Raab, G. M. & Dibben, C. synthpop: Bespoke Creation of Synthetic Data in R. *J. Stat. Softw.* **74**, 1–26 (2016).

Supplement

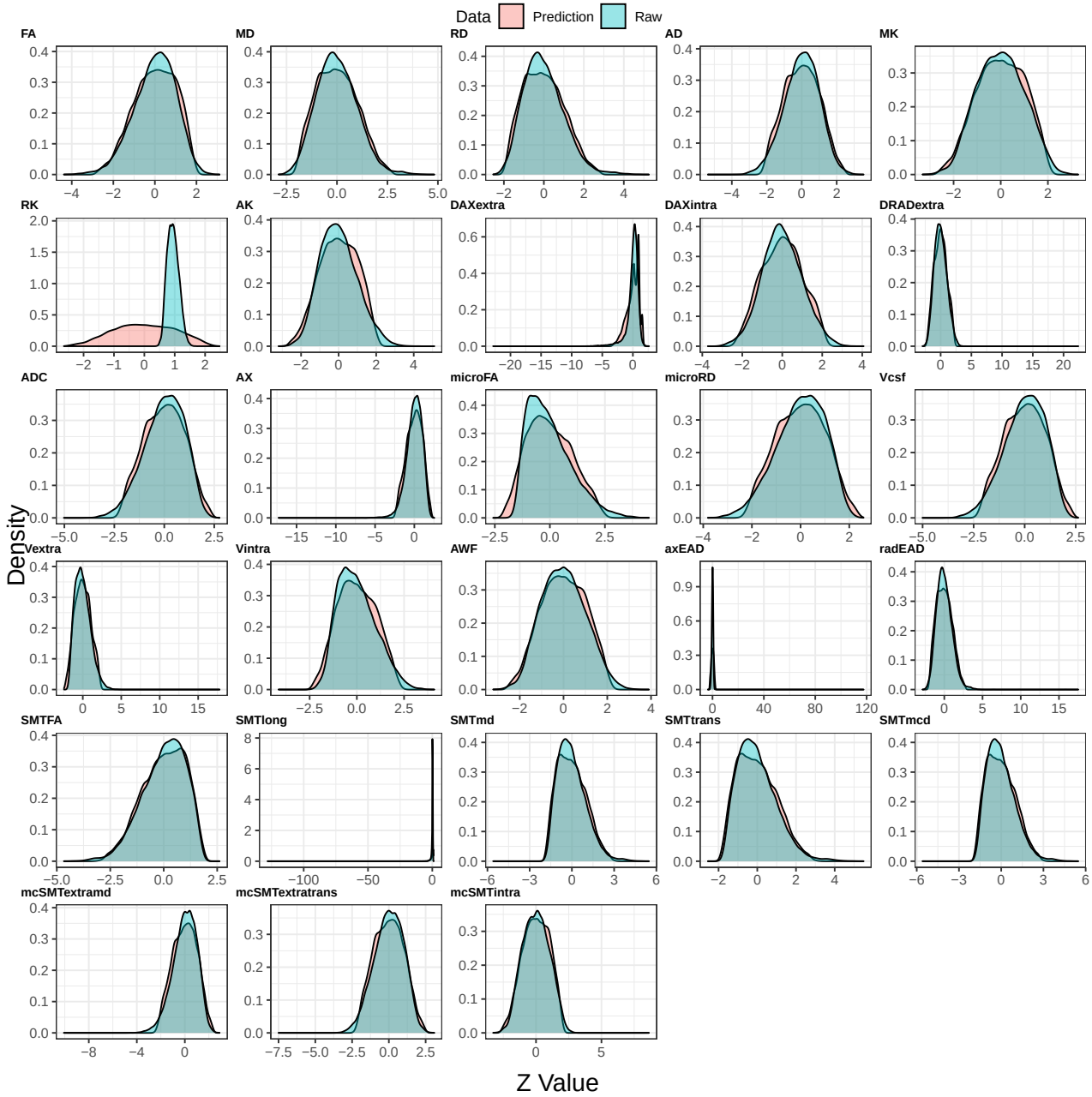
Supplementary Figures

SF1: Correlations of corrected BAG and age across models



Mean = multimodal model including only mean metrics; Full = full multimodal model including all diffusion indices. All correlations were significant at $p_{Holm} < .001$.

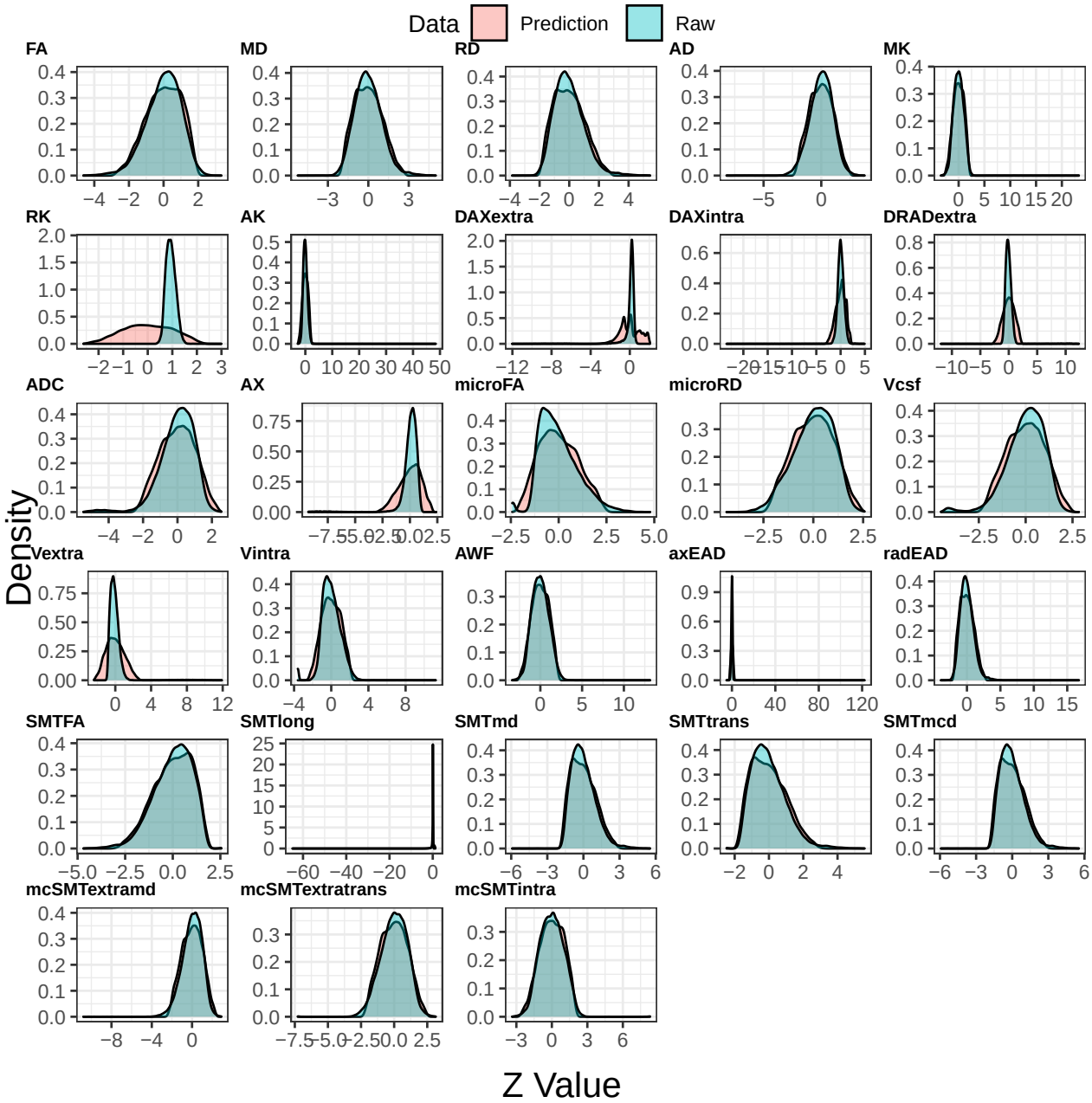
SF2: Comparison of predicted and raw fornix Z-scored diffusion metrics' density



Density plots for each Z-scored (standardised) raw and predicted values for each fornix metric from the six observed diffusion models. Predictions were made from the linear model described in Equation 1.
Find the same density plot for data including QC outliers in SF3.

Supplementing the density plots, two one-sided tests for equivalence testing (TOST)[87,88] were used to test whether mean differences between the model's predictions (**SF9A-B**) and the raw scores (**SF9C-D**) are equal to zero with the assumptions that observed *Z*-score differences smaller |0.5| are equal to 0. Following this assumption, differences were equal to zero for all metrics, except the DKI metric RK: $M_{diff} = 0.943$, 95% CI [0.935, 0.951], $p \approx 1$.

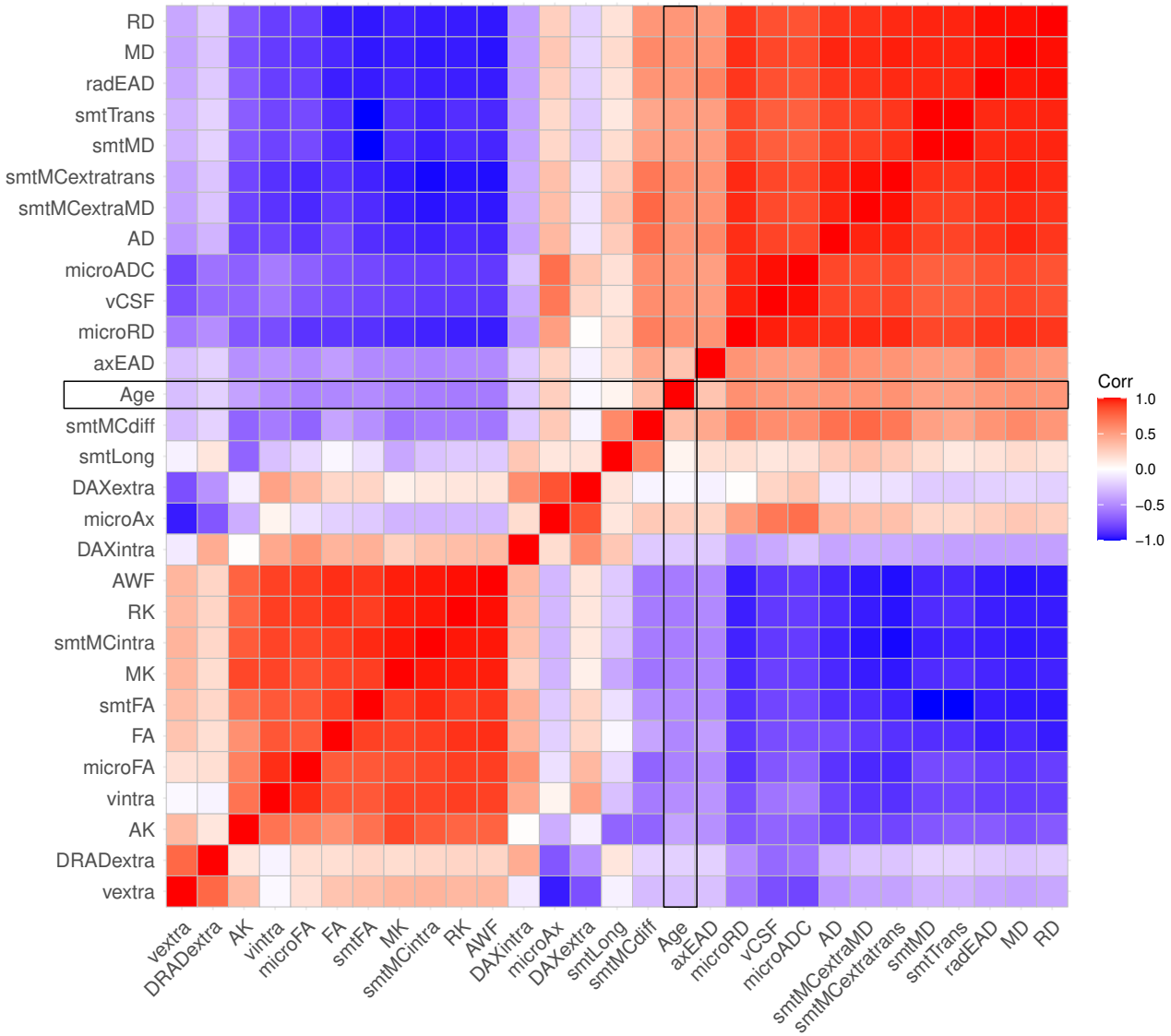
SF3: Comparison of predicted and raw Fornix Z-scored diffusion metrics' density including QC outliers



Density plots for each Z-scored (standardised) raw and predicted values for each fornix metric from the six observed diffusion models on data *containing QC outliers*. Predictions were made from the linear model described in Equation 1.

Outliers were defined by the YTTRIUM method[39] including outlier removal based on density-based spatial clusterisation (k-means)The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

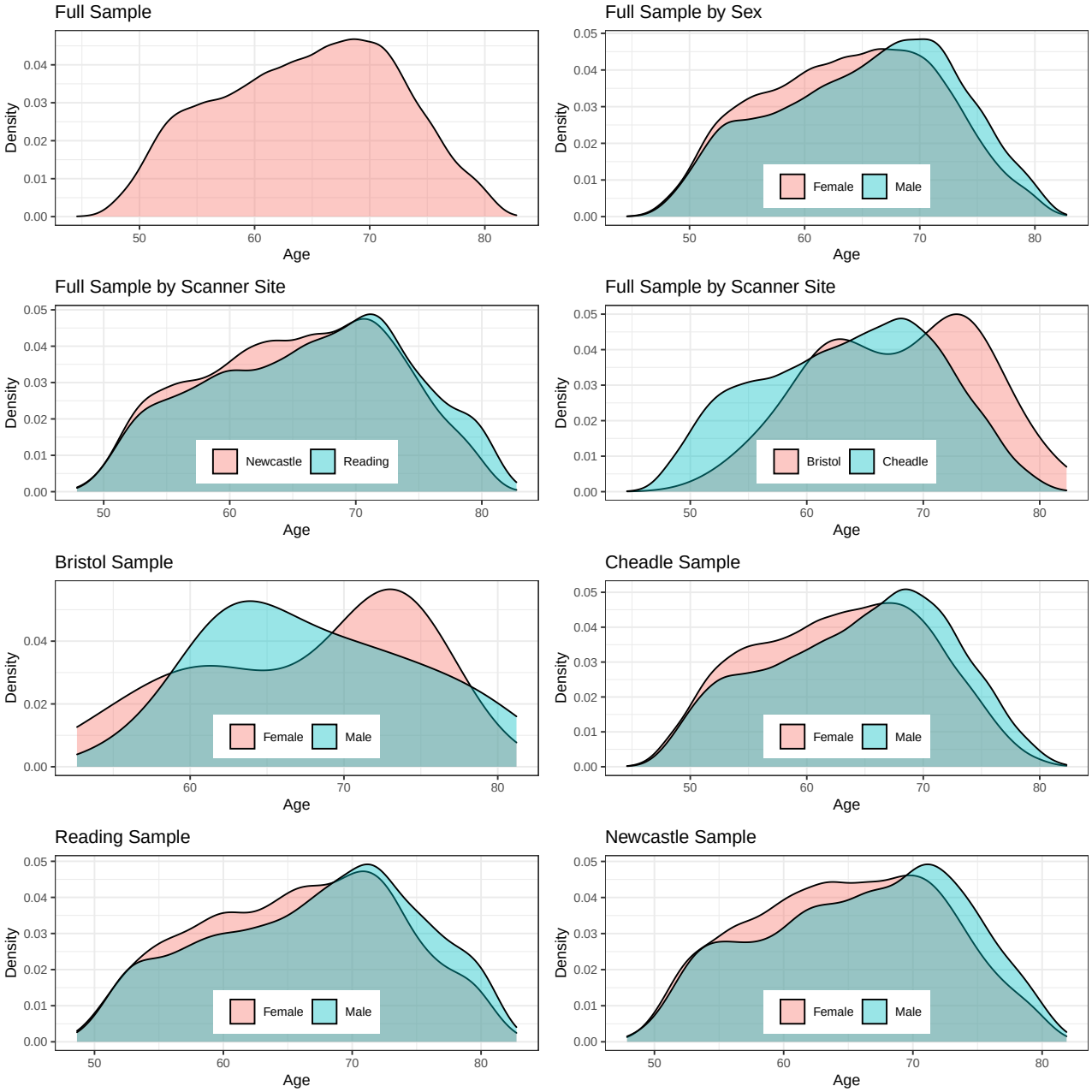
SF4: Correlations between Fornix diffusion metrics and chronological age for data including QC outliers



All correlations were significant at FWE-corrected $p_{Holm} < .05$.

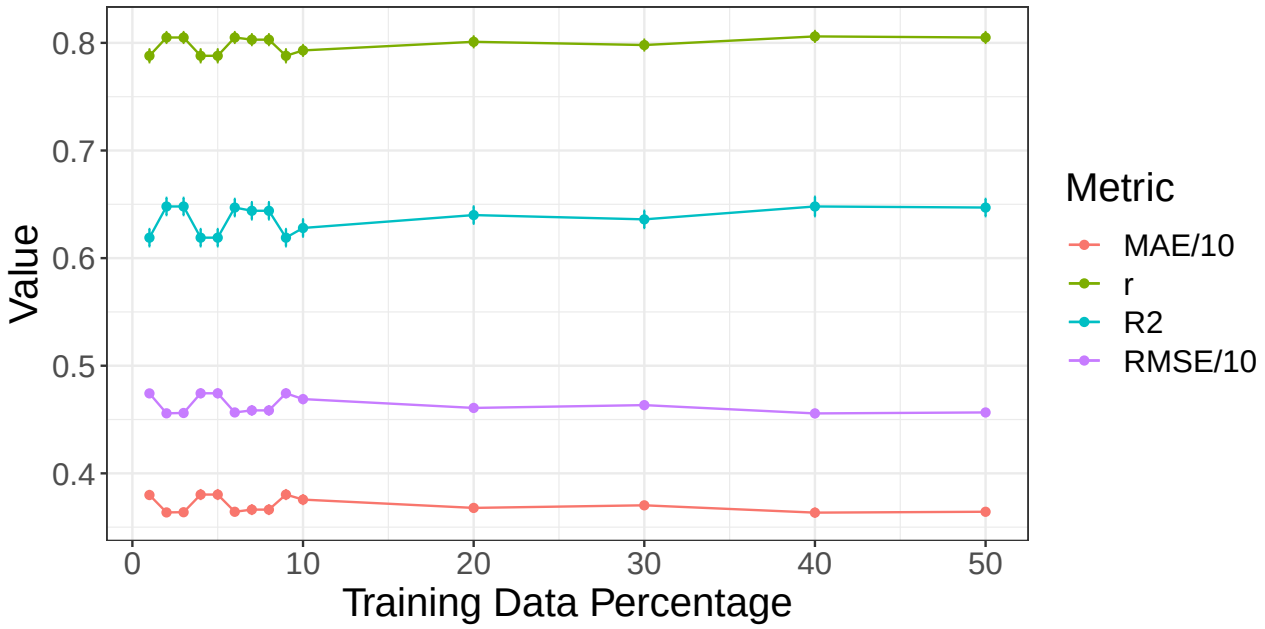
Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

SF5: Density plots for the sample's age by sex and scanner site for data including QC outliers



Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

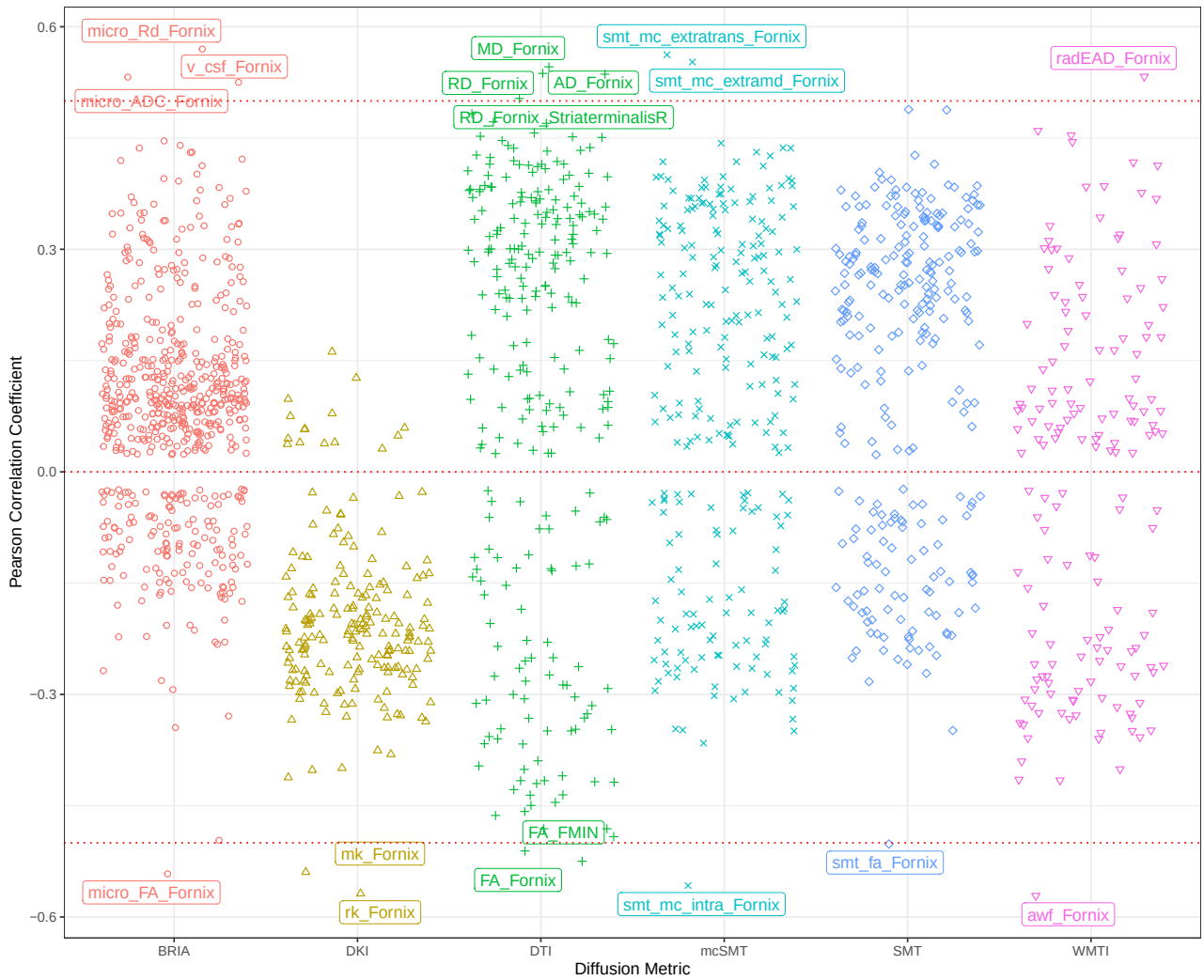
SF6: Model performance for different train-test splits for data *including QC outliers*



Model metrics R^2, RMSE, MAE and their standard deviations, as well as the Pearson's correlations between predicted and chronological age and its 95% confidence interval are displayed for different training data percentages of the total data (x-axis). For visualisation purposes, RMSE and MAE were divided by 10. For exact values see Suppl. Table ST8.

Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

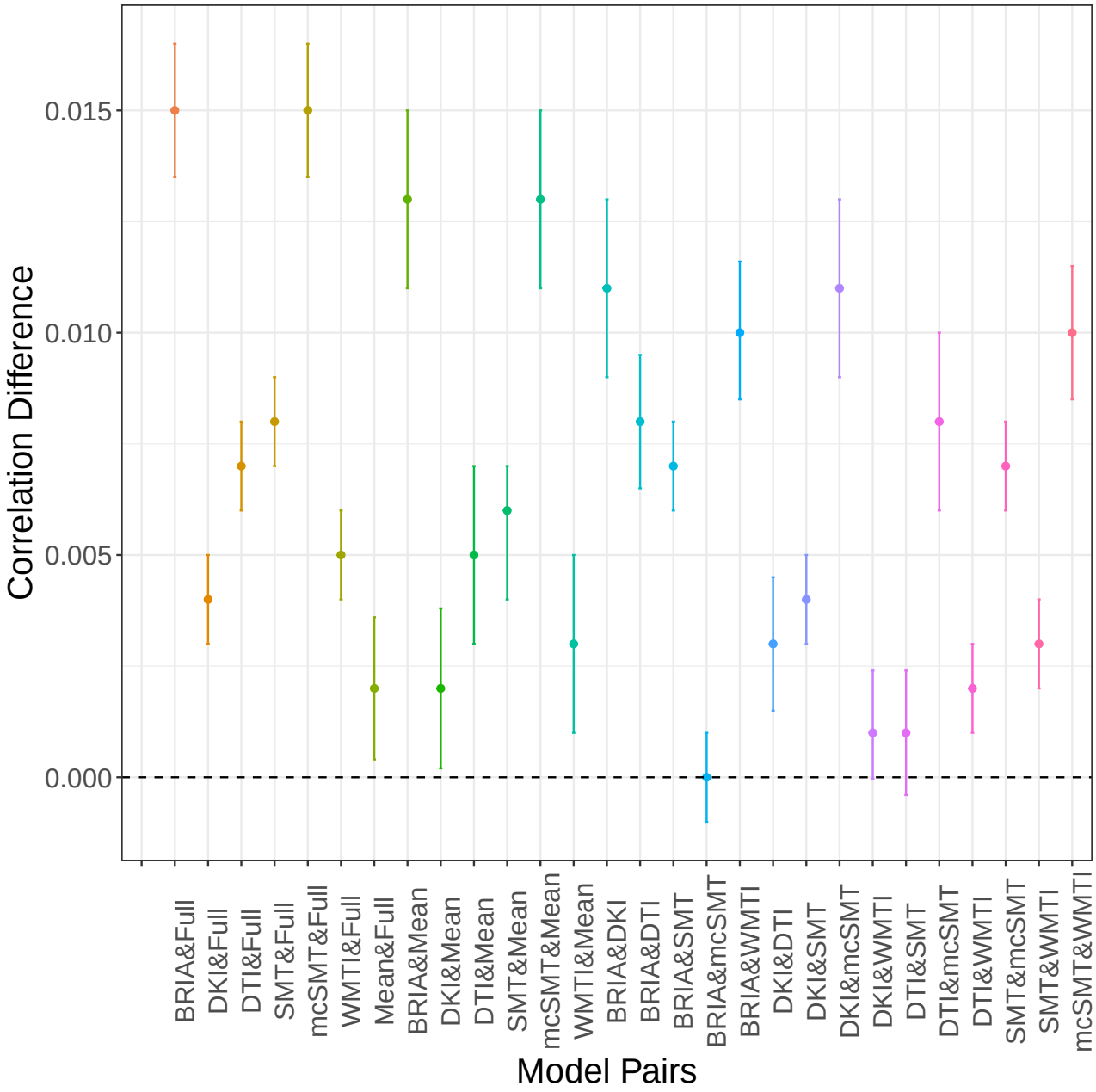
SF7: Correlations between diffusion metrics and chronological age for data including QC outliers



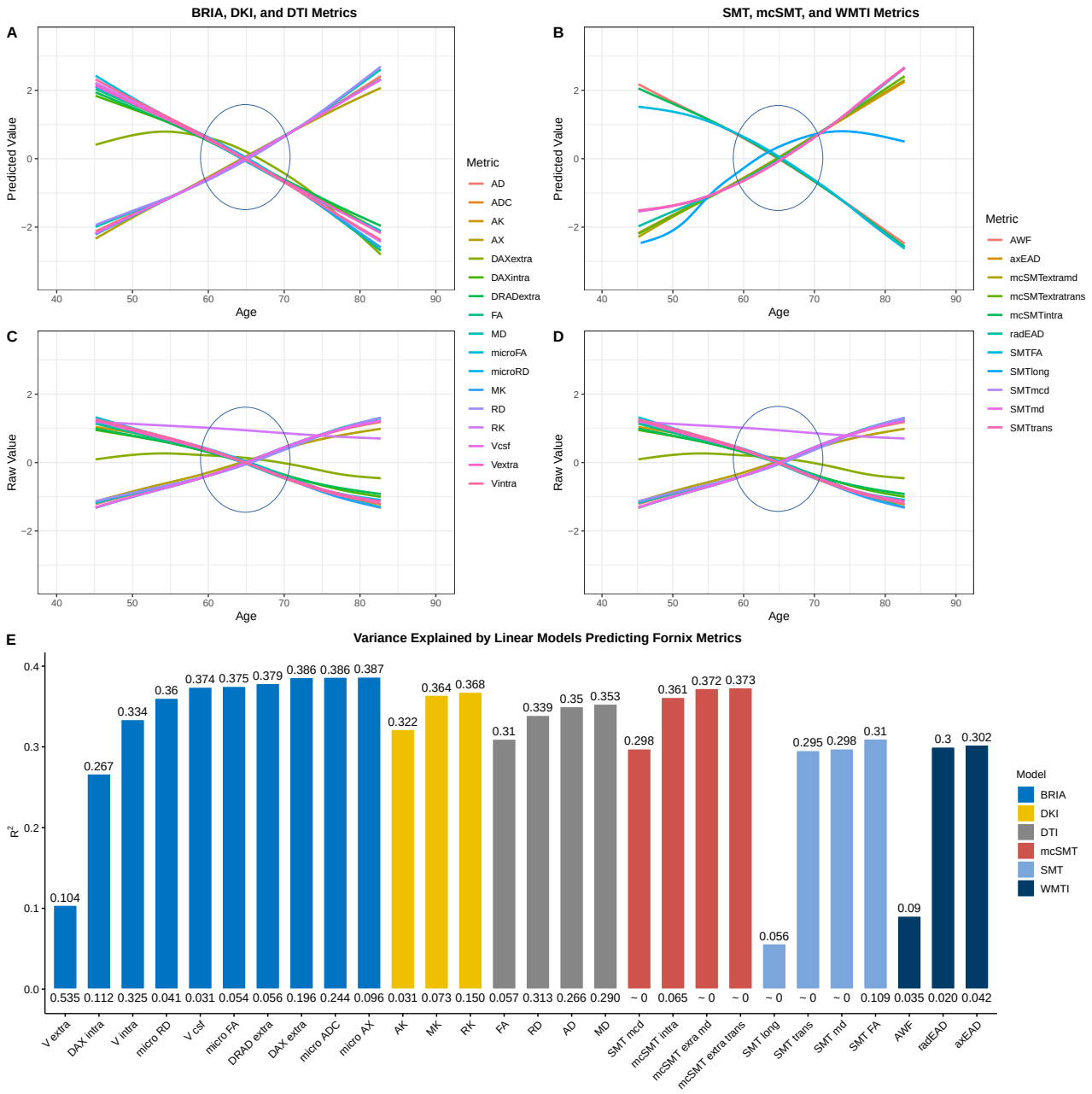
Note: Each point indicates one correlation between a diffusion feature and chronological age. Names of diffusion features are displayed when correlations between the feature and age reached a Pearson correlation of $|r|>0.5$. Holm correction was used for FDR-correction, and all displayed values were significant at $p < .001$.
Results for the analysis run on data *not* including QC outliers (N = 35,749) can be found in Fig.4.

Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

SF8: Differences between correlations of chronological and *corrected* predicted age across diffusion approaches with 95% confidence interval



SF9: Raw and predicted fornix diffusion metrics by chronological age

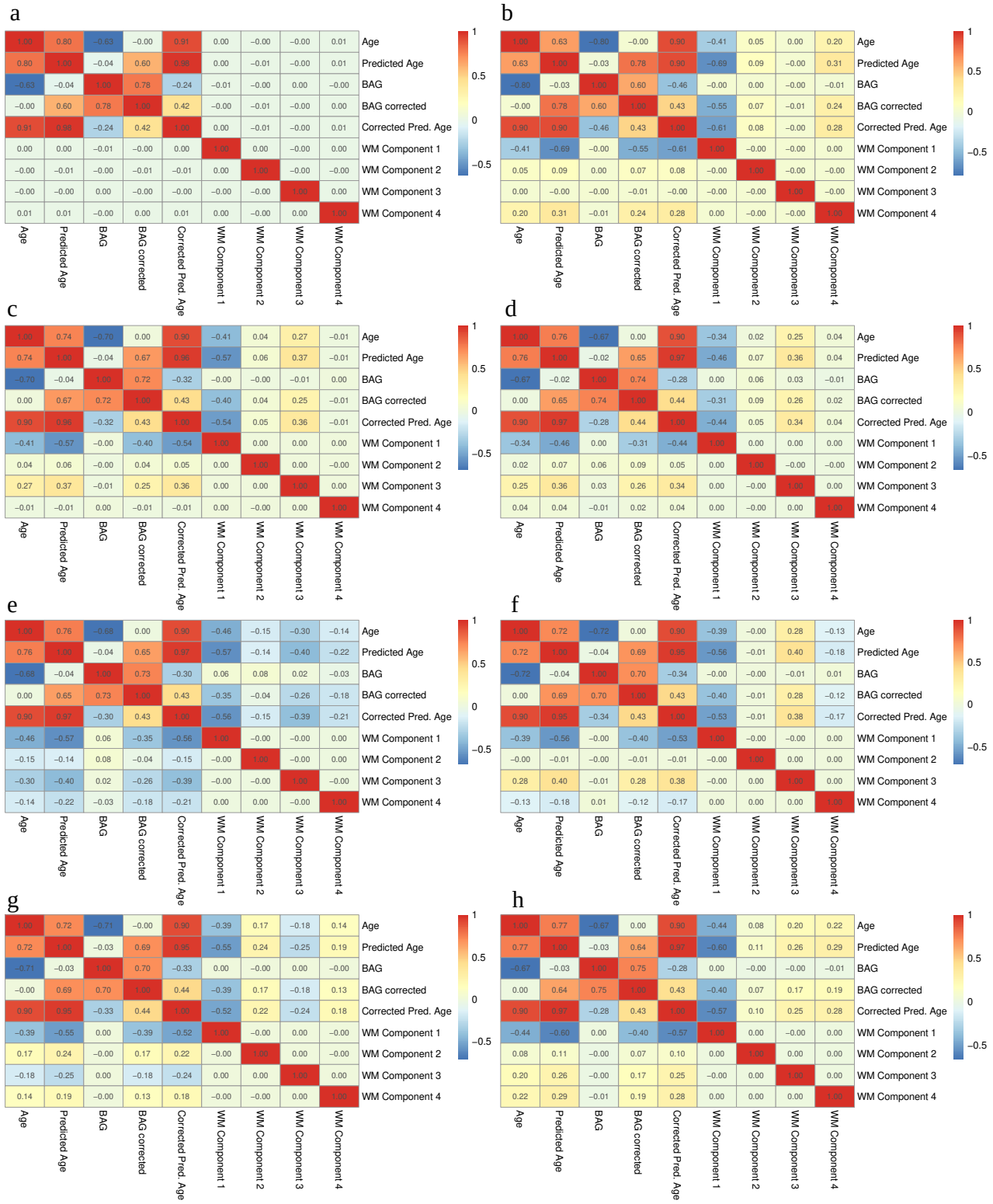


SF9A-D shows age curves for each standardised (z-score) fornix diffusion skeleton value (y-axis) plotted as a function of age (x-axis). Shaded areas represent 95% CI. Curves fitted to raw values (SF9C-D) serve as a comparison to the lm-derived predicted values from Equation 1 (Fig.A-B). SF9E indicates the model fit for the linear models from SF9A-B, showing R^2_{adj} values on top and Standard Error (SE) on the bottom of the bars which each represent a Fornix skeleton value for one of the seven models. Lines crossing at age 65 are marked with circles. Model summaries of all 28 Fornix models can be found in ST5. The same visualisation of diffusion values averaged across the brain can be found in Fig.6.

Model fit metrics R^2_{adj} and Standard Error (SE) for the models accounting for age, sex and scanner site (Equation 1) when predicting fornix metrics were calculated (**SF9E;** see **Fig.6** for whole brain metrics). Highest R^2_{adj} and variability across metrics were observed when predicting BRIA fornix features, lowest R^2_{adj} when predicting SMT fornix metrics. DKI, DTI and mcSMT fornix diffusion metric predictions were most consistent, with BRIA, mcSMT and SMT having one outlier each, Vextra, SMTlong, and AWF, respectively, being less sensitive to age, sex and scanner site. Highest SE could be observed in the BRIA model and the lowest SE in SMT.

To test age-sensitivity of the fornix features, likelihood ratio tests were conduced comparing models derived from Equation 1 against models derived from the same formula with age removed (Equation 2). All models showed significant age dependence, with BRIA microRD (χ^2= 14,480.54, $p_{Holm} < .001$), microADC (χ^2= 14,384.87, $p_{Holm} < .001$) and SMT vCSF (χ^2= 14,311.47, $p_{Holm} < .001$) being the most age-sensitive metrics, and mcSMT smtLong (χ^2= 1,554.49, $p_{Holm} < .001$), BRIA DAXextra (χ^2= 1,824.54, $p_{Holm} < .001$) and axEAD (χ^2= 3,024.74, $p_{Holm} < .001$) the least age-sensitive metrics (**ST4**).

SF10. Pearson's *r* for age, brain age and WM principal components' relationships



All correlations with Pearson's *r* > .01 were significant at p < .001
Read row-wise from top-left to right with matrixes indicating a) full multimodal data, b) mean/whole brain average data, c) BRIA, d) DKI, e) DTI, f) SMT, g) SMT mc, h) WMTI

The first five principal components of the respective number of WM metrics for each of the eight principal components analyses were related to age, predicted (brain) age, corrected predicted age, uncorrected and corrected BAG (see **ST13** for overview of variance explained by principal components). Notably, BAG was not or only weakly related to WM components, and relationships of age, predicted age, corrected predicted age and corrected BAG with WM components followed the same pattern of direction and strength of associations, suggesting age-dependencies of these measures.

When predicting the first 4 components retrieved from the respective models (as done for brain age predictions), using BAG, age, sex, site, as well as age-sex and sex-site interactions as predictors (as specified in Equation 1), different sized proportions of the variance in the components could be explained with corrected and uncorrected BAG models not differing in variance predicted and beta values. Average data BAG models explained most variance in its first component R^2 = .505, with b_{BAG} = -0.673, followed by WMTI R^2 = .372, with b_{BAG} = -0.847, and the DTI BAG model R^2 = .358, with b_{BAG} = -1.082. The second component was best predicted by a DTI BAG model R^2 = .152, with b_{BAG} = -0.059. The third component was best predicted by the DKI BAG model R^2 = .256, b_{BAG} = 0.170, followed by the DTI BAG model R^2 = .250, b_{BAG} = -0.210; and the SMT BAG model R^2 = .247, b_{BAG} = 0.291. Finally, the last component was best predicted by the full BAG model, R^2 = .128, b_{BAG} = 0.0002. For an overview of all BAG models' performance see **ST14**. For a more nuanced follow-up analysis of global and regional individual diffusion metric predictions see **SF11**.

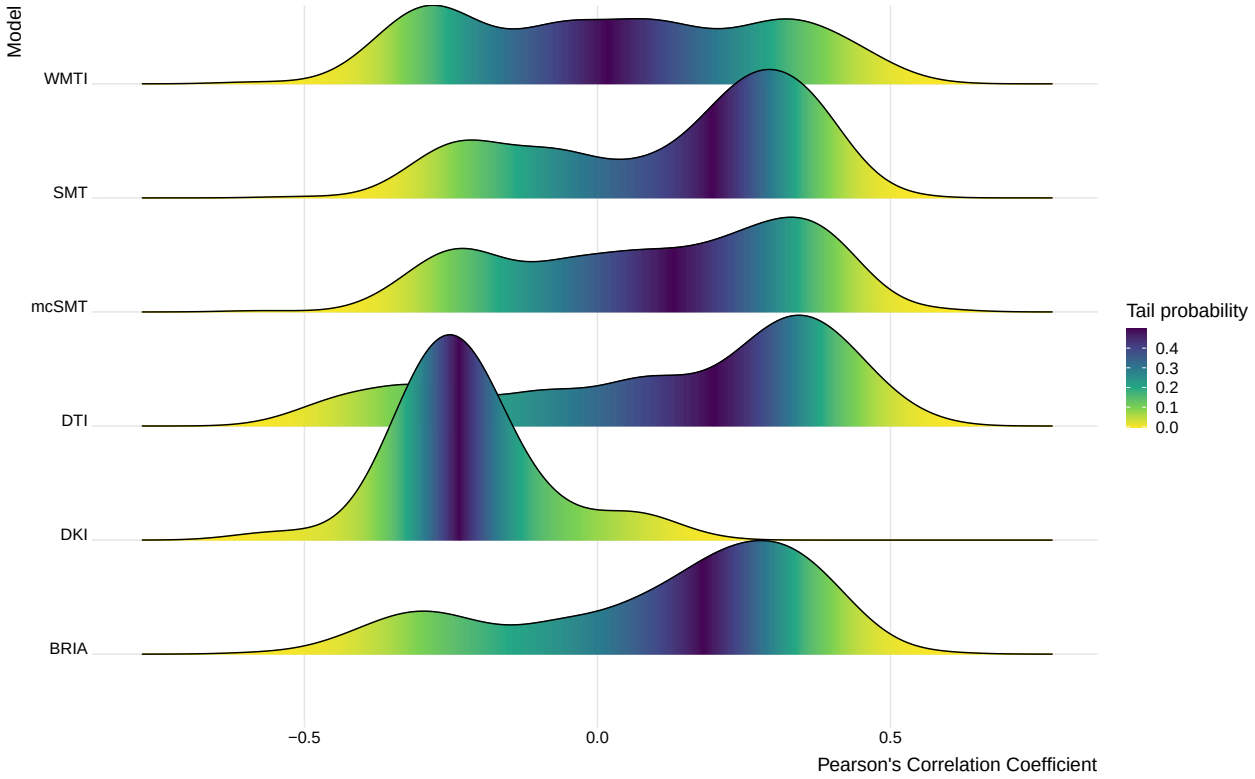
SF11. Predictions of individual global and regional diffusion metrics

Panels indicate used models: a) full multimodal model including all approaches global and local fatures, b) mean multimodal modal, including only global metrics of all diffusion approaches, c) BRIA, d) DKI, e) DTI, f) SMT, g) mcSMT, h) WMTI.

We predicted the individual 1940 regional and global WM diffusion metrics from BAG, site, sex, age, as well as sex-age and sex-site interaction terms. While there were no differences in explaining variance between corrected and uncorrected BAG, models coefficients differed (see **SF11**).

Variance explained across statistically significant models (at Bonferroni-corrected $p < 0.05/1940$) ranged from adjusted $R^2_{min} = .001$ to $R^2_{max} = .387$ ($R^2_{mean} = .108$, SD = 0.062), and beta values for BAG ranged from $b_{BAG} > -0.001$ to $b_{BAG} < 0.001$, with most variance explained in metrics Fornix v csf ($R_{adj}^2 = .387$, $b_{BAG} > -0.001$, $b_{age} = 0.009$), Fornix micro RD ($R_{adj}^2 = .386$, $b_{BAG} > -0.001$, $b_{age} = 0.009$), and Fornix micro ADC ($R_{adj}^2 = .386$, $b_{BAG} > -0.001$, $b_{age} = 0.019$).

SF12. Density plots feature-age correlation across diffusion approaches with tail probabilities



This figure is a supplement to Figure 5, showing the distributions of the correlations between age and each models' diffusion metrics.

ST1: Brain age predictions from different train-test splits

% of Data	Best Fitting Model	Train Results				Test Results on 50% of Data			
		R^2	RMSE	MAE	$r_{age\ x\ pred}$	R^2	RMSE	MAE	$r_{age\ x\ pred}$
1	E = 0.05 D_{max}= 3 T_{max}= 450	0.503 (0.087)	5.300 (0.599)	4.283 (0.555)	0.719 [0.665, 0.766]	0.621 (0.012)	4.693 (0.071)	3.759 (0.070)	0.788 [0.783, 0.794]
2	E = 0.05 D_{max}= 3 T_{max}= 350	0.551 (0.081)	5.107 (0.517)	4.172 (0.474)	0.750 [0.716, 0.780]	0.613 (0.011)	4.743 (0.071)	3.803 (0.073)	0.783 [0.778, 0.789]
3	E = 0.05 D_{max}= 3 T_{max}= 800	0.549 (0.078)	5.129 (0.401)	4.171 (0.393)	0.751 [0.724, 0.776]	0.635 (0.012)	4.605 (0.067)	3.683 (0.065)	0.797 [0.792, 0.802]
4	E = 0.05 D_{max}= 4 T_{max}= 200	0.561 (0.062)	5.136 (0.325)	4.090 (0.257)	0.741 [0.716, 0.763]	0.606 (0.011)	4.784 (0.068)	3.840 (0.071)	0.780 [0.774, 0.785]
5	E = 0.05 D_{max}= 3 T_{max}= 300	0.586 (0.037)	4.994 (0.217)	4.043 (0.241)	0.764 [0.744, 0.783]	0.607 (0.012)	4.779 (0.074)	3.834 (0.074)	0.780 [0.774, 0.785]
6	E = 0.05 D_{max}= 4 T_{max}= 800	0.576 (0.035)	4.962 (0.210)	3.953 (0.149)	0.763 [0.745, 0.780]	0.641 (0.012)	4.569 (0.058)	3.652 (0.057)	0.801 [0.795, 0.806]
7	E = 0.05 D_{max}= 3 T_{max}= 900	0.592 (0.042)	4.887 (0.204)	3.930 (0.150)	0.774 [0.757, 0.789]	0.637 (0.012)	4.591 (0.068)	3.669 (0.064)	0.799 [0.793, 0.804]
8	E = 0.01 D_{max}= 4 T_{max}= 950	0.598 (0.028)	4.881 (0.212)	3.920 (0.216)	0.764 [0.749, 0.779]	0.605 (0.011)	4.790 (0.072)	3.848 (0.072)	0.779 [0.773, 0.785]
9	E = 0.05 D_{max}= 4 T_{max}= 950	0.591 (0.036)	4.882 (0.279)	3.917 (0.175)	0.774 [0.760, 0.788]	0.643 (0.011)	4.554 (0.056)	3.638 (0.055)	0.802 [0.797, 0.807]
10	E = 0.05 D_{max}= 3 T_{max}= 750	0.598 (0.033)	4.886 (0.259)	3.899 (0.230)	0.777 [0.764, 0.790]	0.633 (0.012)	4.614 (0.067)	3.691 (0.064)	0.796 [0.791, 0.802]
20	E = 0.05 D_{max}= 5 T_{max}= 600	0.619 (0.025)	4.748 (0.080)	3.754 (0.124)	0.787 [0.778, 0.795]	0.638 (0.011)	4.587 (0.061)	3.665 (0.060)	0.799 [0.794, 0.804]
30	E = 0.05 D_{max}= 4 T_{max}= 800	0.633 (0.009)	4.623 (0.066)	3.693 (0.065)	0.798 [0.791, 0.804]	0.641 (0.012)	4.569 (0.058)	3.652 (0.057)	0.801 [0.795, 0.806]
40	E = 0.05 D_{max}= 5 T_{max}= 400	0.641 (0.014)	4.584 (0.088)	3.628 (0.050)	0.797 [0.791, 0.803]	0.631 (0.011)	4.628 (0.062)	3.701 (0.061)	0.795 [0.790, 0.800]
50	E = 0.05 D_{max}= 5 T_{max}= 850	0.637 (0.017)	4.576 (0.068)	3.630 (0.049)	0.805 [0.799, 0.810]	0.641 (0.012)	4.566 (0.059)	3.647 (0.059)	0.801 [0.796, 0.806]

R^2 = variance explained, RMSE = root mean square error, MAE = mean absolute error, $r_{age\ x\ pred}$ = correlation of chronological and predicted age. Numbers in round brackets indicate standard deviations. Numbers in square brackets indicate confidence intervals. E = eta (learning rate), D_{max} = maximum depth, T_{max} = maximum number of trees. The best fitting model was selected via grid search focussed on RMSE.

ST2: Differences between correlations of chronological and *corrected* predicted age across models with 95% confidence interval

	BRIA	DKI	DTI	SMT	mcSMT	WMTI	Mean
DKI	0.011 [0.009, 0.013]						
DTI	0.008 [0.007, 0.010]	0.003 [0.002, 0.005]					
SMT	0.007 [0.006, 0.008]	0.004 [0.003, 0.005]	0.001 [-0.0004, 0.0024][4]				
mcSMT	≈0 [-0.001, 0.001][1]	0.011 [0.009, 0.013]	0.008 [0.006, 0.010]	0.007 [0.006, 0.008]			
WMTI	0.010 [0.009, 0.012]	0.001 [-0.0004, 0.0024][2]	0.002 [0.001, 0.003]	0.003 [0.002, 0.004]	0.010 [0.009, 0.012]		
Mean	0.013 [0.011, 0.015]	0.002 [0.0002, 0.038][3]	0.005 [0.003, 0.007]	0.006 [0.004, 0.007]	0.013 [0.011, 0.015]	0.003 [0.001, 0.005]	
FULL	0.015 [0.014, 0.017]	0.004 [0.003, 0.005]	0.007 [0.006, 0.008]	0.008 [0.007, 0.009]	0.015 [0.014, 0.017]	0.005 [0.004, 0.006]	0.002 [0.0004, 0.0036]

Confidence Intervals are based on Zou [42]. Unmarked differences were significant at $p < .001$.
1 Hotelling's [86] $t(32171) \approx 0$, $p \approx 1$.
2 Hotelling's (1940) $t(32171) = 1.4232$, $p = .1547$.
3 Hotelling's (1940) $t(32171) = 2.2174$, $p = .0266$.
4 Hotelling's (1940) $t(32171) = 2.4176$, $p = .0156$.

ST3: Differences between correlations of *uncorrected* predicted and chronological age across diffusion approaches with 95% confidence interval

	BRIA	DKI	DTI	SMT	mcSMT	WMTI	Mean
DKI	0.012 [0.009, 0.015]						
DTI	0.014 [0.011, 0.017]	0.0002 [-0.005, 0.001]					
SMT	0.013 [0.010, 0.016]	0.025 [0.022, 0.029]	0.027 [0.024, 0.030]				
mcSMT	0.021 [0.019, 0.023]	0.033 [0.029, 0.037]	0.035 [0.032, 0.038]	0.008 [0.006, 0.011]			
WMTI	0.023 [0.020, 0.026]	0.011 [0.008, 0.014]	0.009 [0.007, 0.011]	0.036 [0.033, 0.039]	0.044 [0.041, 0.047]		
Mean	0.115 [0.110, 0.120]	0.127 [0.121, 0.133]	0.129 [0.124, 0.134]	0.102 [0.097, 0.107]	0.088 [0.083, 0.093]	0.0940 [0.089, 0.099]	
Full	0.062 [0.059, 0.065]	0.050 [0.048, 0.053]	0.048 [0.046, 0.051]	0.075 [0.072, 0.078]	0.083 [0.080, 0.086]	0.039 [0.037, 0.041]	0.138 [0.133, 0.143]

Confidence Intervals are based on Zou (2007). Unmarked differences were significant at $p < .001$.
* Hotelling's (1940) $t(34801) = 0.9648$, $p = 0.3347$

ST4: Fornix metrics' age sensitivity: comparing diffusion metric prediction models with and without age

Model	Metric	Full[1]	Reduced[2]	χ^2	p	p_{Holm}
BRIA	vintra	63395.90	56372.64	14046.51	<.001	<.001
BRIA	vextra	53488.40	47076.64	12823.52	<.001	<.001
BRIA	vCSF	32902.83	25747.10	14311.47	<.001	<.001
BRIA	microRD	-304.68	-7544.95	14480.54	<.001	<.001
BRIA	microFA	52365.03	45215.37	14299.32	<.001	<.001
BRIA	microAx	27419.04	23070.45	8697.18	<.001	<.001
BRIA	microADC	7489.46	297.02	14384.87	<.001	<.001
BRIA	DRADextra	88839.19	84573.36	8531.64	<.001	<.001
BRIA	DAXintra	62832.56	58452.18	8760.75	<.001	<.001
BRIA	DAXextra	69448.84	68536.57	1824.54	<.001	<.001
DKI	RK	17066.24	10008.43	14115.62	<.001	<.001
DKI	AK	73031.59	67042.47	11978.25	<.001	<.001
DKI	MK	42812.92	35761.79	14102.27	<.001	<.001
DTI	FA	51988.61	46349.96	11277.31	<.001	<.001
DTI	MD	-6425.12	-12850.70	12851.17	<.001	<.001
DTI	RD	-9157.86	-15294.22	12272.70	<.001	<.001
DTI	AD	-3330.43	-9581.75	12502.64	<.001	<.001
SMT	smtFA	28610.35	23053.56	11113.58	<.001	<.001
SMT	smtLong	343020.49	342243.25	1554.49	<.001	<.001
SMT	smtMD	253964.29	248676.32	10575.95	<.001	<.001
SMT	smtTrans	239584.11	234337.78	10492.66	<.001	<.001
mcSMT	smtMCintra	46725.61	39797.67	13855.88	<.001	<.001
mcSMT	smtMCextraMD	260580.57	253725.97	13709.21	<.001	<.001
mcSMT	smtMCextratrans	249976.89	242973.81	14006.16	<.001	<.001
mcSMT	smtMCd	287624.52	284705.87	5837.30	<.001	<.001
WMTI	AWF	73001.96	65850.35	14303.22	<.001	<.001
WMTI	axEAD	-28343.53	-29855.90	3024.74	<.001	<.001
WMTI	radEAD	-10582.93	-16576.91	11987.96	<.001	<.001

1 Full = full model log likelihood
2 Reduced = reduced model log likelihood

ST5: Model summaries for all 28 Fornix models

Effect	β	Std. Error	t-value	p	Metric	β	Std. Error	t-value	p	Metric
age	0.002	0.001	3.521	0	vintra	0.002	0.001	3.521	0	vextra
age^2	0	0	-10.532	0	vintra	0	0	-10.532	0	vextra
sex	0.009	0.018	0.481	0.63	vintra	0.009	0.018	0.481	0.63	vextra
siteCheadle	0.01	0.012	0.825	0.41	vintra	0.01	0.012	0.825	0.41	vextra
siteNewcastle	0.004	0.012	0.327	0.744	vintra	0.004	0.012	0.327	0.744	vextra
siteReading	0.009	0.012	0.748	0.454	vintra	0.009	0.012	0.748	0.454	vextra
age:sex	0	0	-5.532	0	vintra	0	0	-5.532	0	vextra
sex:siteCheadle	-0.003	0.017	-0.183	0.854	vintra	-0.003	0.017	-0.183	0.854	vextra
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	vintra	-0.003	0.017	-0.168	0.867	vextra
sex:siteReading	-0.005	0.017	-0.283	0.777	vintra	-0.005	0.017	-0.283	0.777	vextra
age	0.002	0.001	3.521	0	vCSF	0.002	0.001	3.521	0	microRD
age^2	0	0	-10.532	0	vCSF	0	0	-10.532	0	microRD
sex	0.009	0.018	0.481	0.63	vCSF	0.009	0.018	0.481	0.63	microRD
siteCheadle	0.01	0.012	0.825	0.41	vCSF	0.01	0.012	0.825	0.41	microRD
siteNewcastle	0.004	0.012	0.327	0.744	vCSF	0.004	0.012	0.327	0.744	microRD
siteReading	0.009	0.012	0.748	0.454	vCSF	0.009	0.012	0.748	0.454	microRD
age:sex	0	0	-5.532	0	vCSF	0	0	-5.532	0	microRD
sex:siteCheadle	-0.003	0.017	-0.183	0.854	vCSF	-0.003	0.017	-0.183	0.854	microRD
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	vCSF	-0.003	0.017	-0.168	0.867	microRD
sex:siteReading	-0.005	0.017	-0.283	0.777	vCSF	-0.005	0.017	-0.283	0.777	microRD
age	0.002	0.001	3.521	0	microFA	0.002	0.001	3.521	0	microAx
age^2	0	0	-10.532	0	microFA	0	0	-10.532	0	microAx
sex	0.009	0.018	0.481	0.63	microFA	0.009	0.018	0.481	0.63	microAx
siteCheadle	0.01	0.012	0.825	0.41	microFA	0.01	0.012	0.825	0.41	microAx
siteNewcastle	0.004	0.012	0.327	0.744	microFA	0.004	0.012	0.327	0.744	microAx
siteReading	0.009	0.012	0.748	0.454	microFA	0.009	0.012	0.748	0.454	microAx
age:sex	0	0	-5.532	0	microFA	0	0	-5.532	0	microAx
sex:siteCheadle	-0.003	0.017	-0.183	0.854	microFA	-0.003	0.017	-0.183	0.854	microAx
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	microFA	-0.003	0.017	-0.168	0.867	microAx
sex:siteReading	-0.005	0.017	-0.283	0.777	microFA	-0.005	0.017	-0.283	0.777	microAx
age	0.002	0.001	3.521	0	microADC	0.002	0.001	3.521	0	DRADextra
age^2	0	0	-10.532	0	microADC	0	0	-10.532	0	DRADextra
sex	0.009	0.018	0.481	0.63	microADC	0.009	0.018	0.481	0.63	DRADextra
siteCheadle	0.01	0.012	0.825	0.41	microADC	0.01	0.012	0.825	0.41	DRADextra
siteNewcastle	0.004	0.012	0.327	0.744	microADC	0.004	0.012	0.327	0.744	DRADextra
siteReading	0.009	0.012	0.748	0.454	microADC	0.009	0.012	0.748	0.454	DRADextra
age:sex	0	0	-5.532	0	microADC	0	0	-5.532	0	DRADextra
sex:siteCheadle	-0.003	0.017	-0.183	0.854	microADC	-0.003	0.017	-0.183	0.854	DRADextra
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	microADC	-0.003	0.017	-0.168	0.867	DRADextra
sex:siteReading	-0.005	0.017	-0.283	0.777	microADC	-0.005	0.017	-0.283	0.777	DRADextra
age	0.002	0.001	3.521	0	DAXintra	0.002	0.001	3.521	0	DAXextra
age^2	0	0	-10.532	0	DAXintra	0	0	-10.532	0	DAXextra
sex	0.009	0.018	0.481	0.63	DAXintra	0.009	0.018	0.481	0.63	DAXextra
siteCheadle	0.01	0.012	0.825	0.41	DAXintra	0.01	0.012	0.825	0.41	DAXextra
siteNewcastle	0.004	0.012	0.327	0.744	DAXintra	0.004	0.012	0.327	0.744	DAXextra
siteReading	0.009	0.012	0.748	0.454	DAXintra	0.009	0.012	0.748	0.454	DAXextra
age:sex	0	0	-5.532	0	DAXintra	0	0	-5.532	0	DAXextra
sex:siteCheadle	-0.003	0.017	-0.183	0.854	DAXintra	-0.003	0.017	-0.183	0.854	DAXextra

sex:siteNewcastle	-0.003	0.017	-0.168	0.867	DAXintra	-0.003	0.017	-0.168	0.867	DAXextra
sex:siteReading	-0.005	0.017	-0.283	0.777	DAXintra	-0.005	0.017	-0.283	0.777	DAXextra
age	0.002	0.001	3.521	0	RK	0.002	0.001	3.521	0	AK
age^2	0	0	-10.532	0	RK	0	0	-10.532	0	AK
sex	0.009	0.018	0.481	0.63	RK	0.009	0.018	0.481	0.63	AK
siteCheadle	0.01	0.012	0.825	0.41	RK	0.01	0.012	0.825	0.41	AK
siteNewcastle	0.004	0.012	0.327	0.744	RK	0.004	0.012	0.327	0.744	AK
siteReading	0.009	0.012	0.748	0.454	RK	0.009	0.012	0.748	0.454	AK
age:sex	0	0	-5.532	0	RK	0	0	-5.532	0	AK
sex:siteCheadle	-0.003	0.017	-0.183	0.854	RK	-0.003	0.017	-0.183	0.854	AK
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	RK	-0.003	0.017	-0.168	0.867	AK
sex:siteReading	-0.005	0.017	-0.283	0.777	RK	-0.005	0.017	-0.283	0.777	AK
age	0.002	0.001	3.521	0	MK	0.002	0.001	3.521	0	FA
age^2	0	0	-10.532	0	MK	0	0	-10.532	0	FA
sex	0.009	0.018	0.481	0.63	MK	0.009	0.018	0.481	0.63	FA
siteCheadle	0.01	0.012	0.825	0.41	MK	0.01	0.012	0.825	0.41	FA
siteNewcastle	0.004	0.012	0.327	0.744	MK	0.004	0.012	0.327	0.744	FA
siteReading	0.009	0.012	0.748	0.454	MK	0.009	0.012	0.748	0.454	FA
age:sex	0	0	-5.532	0	MK	0	0	-5.532	0	FA
sex:siteCheadle	-0.003	0.017	-0.183	0.854	MK	-0.003	0.017	-0.183	0.854	FA
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	MK	-0.003	0.017	-0.168	0.867	FA
sex:siteReading	-0.005	0.017	-0.283	0.777	MK	-0.005	0.017	-0.283	0.777	FA
age	0.002	0.001	3.521	0	MD	0.002	0.001	3.521	0	RD
age^2	0	0	-10.532	0	MD	0	0	-10.532	0	RD
sex	0.009	0.018	0.481	0.63	MD	0.009	0.018	0.481	0.63	RD
siteCheadle	0.01	0.012	0.825	0.41	MD	0.01	0.012	0.825	0.41	RD
siteNewcastle	0.004	0.012	0.327	0.744	MD	0.004	0.012	0.327	0.744	RD
siteReading	0.009	0.012	0.748	0.454	MD	0.009	0.012	0.748	0.454	RD
age:sex	0	0	-5.532	0	MD	0	0	-5.532	0	RD
sex:siteCheadle	-0.003	0.017	-0.183	0.854	MD	-0.003	0.017	-0.183	0.854	RD
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	MD	-0.003	0.017	-0.168	0.867	RD
sex:siteReading	-0.005	0.017	-0.283	0.777	MD	-0.005	0.017	-0.283	0.777	RD
age	0.002	0.001	3.521	0	AD	0.002	0.001	3.521	0	smtFA
age^2	0	0	-10.532	0	AD	0	0	-10.532	0	smtFA
sex	0.009	0.018	0.481	0.63	AD	0.009	0.018	0.481	0.63	smtFA
siteCheadle	0.01	0.012	0.825	0.41	AD	0.01	0.012	0.825	0.41	smtFA
siteNewcastle	0.004	0.012	0.327	0.744	AD	0.004	0.012	0.327	0.744	smtFA
siteReading	0.009	0.012	0.748	0.454	AD	0.009	0.012	0.748	0.454	smtFA
age:sex	0	0	-5.532	0	AD	0	0	-5.532	0	smtFA
sex:siteCheadle	-0.003	0.017	-0.183	0.854	AD	-0.003	0.017	-0.183	0.854	smtFA
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	AD	-0.003	0.017	-0.168	0.867	smtFA
sex:siteReading	-0.005	0.017	-0.283	0.777	AD	-0.005	0.017	-0.283	0.777	smtFA
age	0.002	0.001	3.521	0	smtLong	0.002	0.001	3.521	0	smtMD
age^2	0	0	-10.532	0	smtLong	0	0	-10.532	0	smtMD
sex	0.009	0.018	0.481	0.63	smtLong	0.009	0.018	0.481	0.63	smtMD
siteCheadle	0.01	0.012	0.825	0.41	smtLong	0.01	0.012	0.825	0.41	smtMD
siteNewcastle	0.004	0.012	0.327	0.744	smtLong	0.004	0.012	0.327	0.744	smtMD
siteReading	0.009	0.012	0.748	0.454	smtLong	0.009	0.012	0.748	0.454	smtMD
age:sex	0	0	-5.532	0	smtLong	0	0	-5.532	0	smtMD
sex:siteCheadle	-0.003	0.017	-0.183	0.854	smtLong	-0.003	0.017	-0.183	0.854	smtMD
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	smtLong	-0.003	0.017	-0.168	0.867	smtMD
sex:siteReading	-0.005	0.017	-0.283	0.777	smtLong	-0.005	0.017	-0.283	0.777	smtMD
age	0.002	0.001	3.521	0	smtTrans	0.002	0.001	3.521	0	smtMCintra
age^2	0	0	-10.532	0	smtTrans	0	0	-10.532	0	smtMCintra
sex	0.009	0.018	0.481	0.63	smtTrans	0.009	0.018	0.481	0.63	smtMCintra
siteCheadle	0.01	0.012	0.825	0.41	smtTrans	0.01	0.012	0.825	0.41	smtMCintra
siteNewcastle	0.004	0.012	0.327	0.744	smtTrans	0.004	0.012	0.327	0.744	smtMCintra
siteReading	0.009	0.012	0.748	0.454	smtTrans	0.009	0.012	0.748	0.454	smtMCintra

age:sex	0	0	-5.532	0	smtTrans	0	0	-5.532	0	smtMCintra
sex:siteCheadle	-0.003	0.017	-0.183	0.854	smtTrans	-0.003	0.017	-0.183	0.854	smtMCintra
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	smtTrans	-0.003	0.017	-0.168	0.867	smtMCintra
sex:siteReading	-0.005	0.017	-0.283	0.777	smtTrans	-0.005	0.017	-0.283	0.777	smtMCintra
age	0.002	0.001	3.521	0	smtMCextraMD	0.002	0.001	3.521	0	smtMCextratrans
age^2	0	0	-10.532	0	smtMCextraMD	0	0	-10.532	0	smtMCextratrans
sex	0.009	0.018	0.481	0.63	smtMCextraMD	0.009	0.018	0.481	0.63	smtMCextratrans
siteCheadle	0.01	0.012	0.825	0.41	smtMCextraMD	0.01	0.012	0.825	0.41	smtMCextratrans
siteNewcastle	0.004	0.012	0.327	0.744	smtMCextraMD	0.004	0.012	0.327	0.744	smtMCextratrans
siteReading	0.009	0.012	0.748	0.454	smtMCextraMD	0.009	0.012	0.748	0.454	smtMCextratrans
age:sex	0	0	-5.532	0	smtMCextraMD	0	0	-5.532	0	smtMCextratrans
sex:siteCheadle	-0.003	0.017	-0.183	0.854	smtMCextraMD	-0.003	0.017	-0.183	0.854	smtMCextratrans
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	smtMCextraMD	-0.003	0.017	-0.168	0.867	smtMCextratrans
sex:siteReading	-0.005	0.017	-0.283	0.777	smtMCextraMD	-0.005	0.017	-0.283	0.777	smtMCextratrans
age	0.002	0.001	3.521	0	smtMCd	0.002	0.001	3.521	0	AWF
age^2	0	0	-10.532	0	smtMCd	0	0	-10.532	0	AWF
sex	0.009	0.018	0.481	0.63	smtMCd	0.009	0.018	0.481	0.63	AWF
siteCheadle	0.01	0.012	0.825	0.41	smtMCd	0.01	0.012	0.825	0.41	AWF
siteNewcastle	0.004	0.012	0.327	0.744	smtMCd	0.004	0.012	0.327	0.744	AWF
siteReading	0.009	0.012	0.748	0.454	smtMCd	0.009	0.012	0.748	0.454	AWF
age:sex	0	0	-5.532	0	smtMCd	0	0	-5.532	0	AWF
sex:siteCheadle	-0.003	0.017	-0.183	0.854	smtMCd	-0.003	0.017	-0.183	0.854	AWF
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	smtMCd	-0.003	0.017	-0.168	0.867	AWF
sex:siteReading	-0.005	0.017	-0.283	0.777	smtMCd	-0.005	0.017	-0.283	0.777	AWF
age	0.002	0.001	3.521	0	axEAD	0.002	0.001	3.521	0	radEAD
age^2	0	0	-10.532	0	axEAD	0	0	-10.532	0	radEAD
sex	0.009	0.018	0.481	0.63	axEAD	0.009	0.018	0.481	0.63	radEAD
siteCheadle	0.01	0.012	0.825	0.41	axEAD	0.01	0.012	0.825	0.41	radEAD
siteNewcastle	0.004	0.012	0.327	0.744	axEAD	0.004	0.012	0.327	0.744	radEAD
siteReading	0.009	0.012	0.748	0.454	axEAD	0.009	0.012	0.748	0.454	radEAD
age:sex	0	0	-5.532	0	axEAD	0	0	-5.532	0	radEAD
sex:siteCheadle	-0.003	0.017	-0.183	0.854	axEAD	-0.003	0.017	-0.183	0.854	radEAD
sex:siteNewcastle	-0.003	0.017	-0.168	0.867	axEAD	-0.003	0.017	-0.168	0.867	radEAD
sex:siteReading	-0.005	0.017	-0.283	0.777	axEAD	-0.005	0.017	-0.283	0.777	radEAD

ST6: Brain age prediction model performance *for data including QC outliers*

Name	MRI features	R^2	RMSE	MAE	Prediction-Age Correlation*	Corrected Prediction-Age Correlation*
BRIA	700	0.538 (0.009)	5.103 (0.044)	4.096 (0.038)	0.734 [0.729, 0.739]	0.902 [0.900, 0.904]
DKI	210	0.561 (0.009)	5.078 (0.053)	4.073 (0.047)	0.750 [0.745, 0.754]	0.876 [0.874, 0.879]
DTI	280	0.565 (0.009)	5.052 (0.039)	4.041 (0.038)	0.752 [0.748, 0.757]	0.874 [0.872, 0.877]
SMT	280	0.522 (0.009)	5.297 (0.035)	4.254 (0.031)	0.723 [0.718, 0.728]	0.870 [0.868, 0.873]
mcSMT	280	0.508 (0.008)	5.263 (0.040)	4.227 (0.034)	0.714 [0.708, 0.719]	0.901 [0.899, 0.903]
WMTI	210	0.574 (0.009)	4.999 (0.036)	4.003 (0.034)	0.758 [0.754, 0.763]	0.875 [0.873, 0.877]
Mean scores multimodal	28	0.400 (0.068)	5.945 (0.082)	4.820 (0.068)	0.633 [0.627, 0.639]	0.875 [0.873, 0.878]
Full model multimodal	1932	0.648 (0.009)	4.557 (0.077)	3.637 (0.066)	0.805 [0.801, 0.808]	0.877 [0.875, 0.880]

Model selection was based on a grid search with stopping rule when model performance did not not improve after 20 rounds. Model selection of all models was based on multimodal model training on 10% of the data, indicating best fit for learning rate = 0.05, maximum depth = 4, maximum number of trees = 750 as indicated in Fig.8 and ST1.
R^2:variance explained, RMSE: root mean squared error, MAE: mean absolute error
Note: R^2, RMSE, MAE are displayed in the format Mean (Standard Deviation), Pearson's correlations are displayed in the format Correlation Score 95% Confidence Interval [Lower Bound, Upper Bound].
*All correlation were significant at $p < .001$.

Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

ST7: Top five diffusion metrics ranked by gain in age prediction accuracy *for data including QC outliers*

BRIA	DKI	DTI	SMT	mcSMT	WMTI	Full
Micro RD ATRR (67436)	MK Fornix (99885)	MD Fornix (79767)	MD Fornix (72129)	Extratrans Fornix (55429)	AWF Fornix (69033)	AWF fornix (278977)
Micro FA Fornix (63804)	RK Fornix (28196)	RD left anterior corona radiata (51592)	MD FMIN (48161)	Intra Fornix (46212)	RadEAD anterior right corona radiata (34381)	Micro RD fornix (71175)
Micro RD right external capsule (27069)	AK anterior right limb of internal capsule (22725)	RD FMIN (25201)	MD right anterior corona radiata (44331)	Extratrans right external capsule (16611)	RadEAD IFOFR (22512)	Micro FA fornix(35049)
Micro RD Fornix right striaterminalis (17090)	AK Fornix (14401)	RD Fornix right stria terminalis (22951)	FA Fornix (19697)	ExtraMD Fornix (10267)	RadEAD FMIN (20286)	MD right tapetum (34008)
Micro FA FMIN (14335)	AK superior frontooccipital left fasciculus (7978)	MD anterior limb of internal left capsule (15589)	Long left tapetum (14596)	ExtraMD anterior left limb of internal capsule (7830)	RadEAD ATRL (16666)	RadEAD anterior right corona radiata (27907)

Table values can be read as feature name (gain). Mean refers to the multimodal model containing only mean scores and full to the full model containing all features. Cells including Fornix are marked in green.
laLC = left anterior limb of internal capsule; raLC = right anterior limb of internal capsule; lST = lSTria terminalis; rST = rSTria terminalis; lsfoF = left superior frontal occipital fasciculus; laCR = left anterior corona radiata; raCR = right anterior corona radiata; rEC = right external capsule

Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

ST8: Brain age predictions from different train-test splits *for data including QC outliers*

% of Data	Best Fitting Model	Train Results				Test Results on 50% of Data			
		R^2	RMSE	MAE	$r_{age\ x\ pred}$	R^2	RMSE	MAE	$r_{age\ x\ pred}$
1	E = 0.05 D_{max}= 4 T_{max}= 250	0.503 (0.113)	5.313 (0.448)	4.287 (0.233)	0.716 [0.664, 0.761]	0.619 (0.008)	4.743 (0.029)	3.799 (0.036)	0.788 [0.782, 0.793]
2	E = 0.05 D_{max}= 4 T_{max}= 950	0.543 (0.094)	5.162 (0.365)	4.101 (0.250)	0.744 [0.711, 0.774]	0.648 (0.008)	4.558 (0.025)	3.637 (0.025)	0.805 [0.800, 0.810]
3	E = 0.05 D_{max}= 4 T_{max}= 900	0.543 (0.052)	5.226 (0.334)	4.149 (0.297)	0.741 [0.714, 0.766]	0.648 (0.008)	4.561 (0.024)	3.639 (0.026)	0.805 [0.800, 0.810]
4	E = 0.05 D_{max}= 3 T_{max}= 550	0.572 (0.033)	5.110 (0.266)	4.043 (0.184)	0.758 [0.736, 0.779]	0.619 (0.008)	4.744 (0.036)	3.803 (0.041)	0.788 [0.782, 0.793]
5	E = 0.05 D_{max}= 3 T_{max}= 350	0.575 (0.051)	5.038 (0.312)	4.015 (0.233)	0.760 [0.740, 0.778]	0.619 (0.008)	4.744 (0.036)	3.803 (0.041)	0.788 [0.782, 0.793]
6	E = 0.05 D_{max}= 4 T_{max}= 850	0.574 (0.047)	5.144 (0.239)	4.048 (0.195)	0.758 [0.740, 0.774]	0.647 (0.008)	4.566 (0.024)	3.643 (0.026)	0.805 [0.800, 0.810]
7	E = 0.05 D_{max}= 3 T_{max}= 900	0.589 (0.032)	4.998 (0.147)	4.004 (0.186)	0.770 [0.754, 0.785]	0.644 (0.008)	4.585 (0.039)	3.663 (0.040)	0.803 [0.798, 0.808]
8	E = 0.05 D_{max}= 3 T_{max}= 550	0.595 (0.017)	4.944 (0.174)	3.953 (0.138)	0.775 [0.761, 0.789]	0.644 (0.008)	4.585 (0.039)	3.663 (0.040)	0.803 [0.798, 0.808]
9	E = 0.05 D_{max}= 3 T_{max}= 350	0.602 (0.025)	4.915 (0.195)	3.921 (0.169)	0.773 [0.759, 0.786]	0.619 (0.008)	4.744 (0.036)	3.803 (0.041)	0.788 [0.782, 0.793]
10	E = 0.05 D_{max}= 3 T_{max}= 550	0.608 (0.033)	4.869 (0.188)	3.882 (0.114)	0.780 [0.768, 0.792]	0.628 (0.008)	4.690 (0.037)	3.756 (0.041)	0.793 [0.788, 0.798]
20	E = 0.05 D_{max}= 4 T_{max}= 450	0.624 (0.013)	4.706 (0.097)	3.779 (0.092)	0.791 [0.782, 0.799]	0.640 (0.008)	4.608 (0.024)	3.679 (0.028)	0.801 [0.796, 0.806]
30	E = 0.05 D_{max}= 4 T_{max}= 900	0.641 (0.017)	4.611 (0.066)	3.683 (0.064)	0.641 [0.798, 0.811]	0.636 (0.008)	4.634 (0.023)	3.703 (0.029)	0.798 [0.793, 0.803]
40	E = 0.05 D_{max}= 5 T_{max}= 950	0.643 (0.019)	4.596 (0.063)	3.668 (0.053)	0.803 [0.797, 0.808]	0.648 (0.009)	4.557 (0.033)	3.635 (0.037)	0.806 [0.801, 0.810]
50	E = 0.05 D_{max}= 4 T_{max}= 850	0.643 (0.020)	4.584 (0.122)	3.637 (0.088)	0.806 [0.801, 0.811]	0.647 (0.008)	4.566 (0.024)	3.643 (0.026)	0.805 [0.800, 0.810]

Note: Numbers in round brackets indicate standard deviations. Numbers in square brackets indicate confidence intervals. E = eta, D_{max} = maximum depth, T_{max} = maximum number of trees. The best fitting model was determined by grid search.

Outliers were defined by the YTTRIUM method[38] including outlier removal based on density-based spatial clusterisation (k-means). The total data used here was $N_{full+outliers}$ = 38,687, including the full data N_{full} = 35,749 used for all analyses and $N_{outliers}$ = 2,938 datasets defined as outliers. This dataset does not include participants who withdrew their consent or participants with an ICD-10 diagnosis categories G or F or stroke, category I.

ST9: Comparisons of linear and generalized additive models predicting fornix diffusion metrics

Metric	LM AIC	GAM AIC	LM BIC	GAM BIC	LM R^2_{adj}	GAM R^2_{adj}
vintra	-126767.79	-126755.60	-126665.98	-126662.27	0.36	0.36
vextra	-106952.81	-106929.99	-106851.00	-106836.66	0.38	0.37
vCSF	-65781.67	-65773.53	-65679.86	-65680.20	0.39	0.39
microRD	633.36	643.11	735.18	736.43	0.39	0.39
microFA	-104706.06	-104634.30	-104604.25	-104540.97	0.38	0.38
microAx	-54814.08	-54790.59	-54712.27	-54697.26	0.27	0.27
microADC	-14954.91	-14946.59	-14853.10	-14853.26	0.39	0.39
DRADextra	-177654.37	-177648.58	-177552.56	-177555.25	0.30	0.30
DAXintra	-125641.12	-125639.53	-125539.31	-125546.20	0.30	0.30
DAXextra	-138873.67	-138808.20	-138771.86	-138714.87	0.09	0.09
RK	-34108.48	-34109.41	-34006.67	-34016.08	0.37	0.37
AK	-146039.18	-146041.08	-145937.37	-145947.75	0.32	0.32
MK	-85601.85	-85603.57	-85500.04	-85510.24	0.36	0.36
FA	-103953.22	-103924.62	-103851.41	-103831.30	0.31	0.31
MD	12874.23	12878.93	12976.04	12972.25	0.35	0.35
RD	18339.73	18359.69	18441.54	18453.02	0.34	0.34
AD	6684.86	6689.75	6786.67	6783.08	0.35	0.35
smtFA	-57196.69	-57161.31	-57094.88	-57067.98	0.31	0.31
smtLong	-686016.98	-685869.48	-685915.17	-685776.15	0.06	0.05
smtMD	-507904.58	-507869.45	-507802.77	-507776.13	0.30	0.30
smtTrans	-479144.23	-479104.60	-479042.42	-479011.28	0.30	0.29
smtMCintra	-93427.21	-93427.11	-93325.40	-93333.78	0.36	0.36
smtMCextraMD	-521137.14	-521125.59	-521035.33	-521032.26	0.37	0.37
smtMCextratrans	-499929.78	-499929.66	-499827.97	-499836.34	0.37	0.37
smtMCd	-575225.03	-574996.85	-575123.22	-574903.53	0.22	0.21
AWF	-145979.93	-145981.86	-145878.11	-145888.53	0.37	0.37
axEAD	56711.05	56715.91	56812.86	56809.24	0.10	0.10
radEAD	21189.85	21203.80	21291.67	21297.12	0.33	0.33

LM = linear model, GAM = generalized additive model, AIC = Akaike information criterion, BIC = Bayesian information criterion. The numbers are derived from the six diffusion approaches' 28 metrics following Equation 1 for linear models and all variables of the equation allowing splines for non-linear models.

ST10: Overview of diffusion metrics by diffusion approach

Diffusion Approach	Metrics
Bayesian Rotationally Invariant Approach (BRIA)	intra-axonal axial diffusivity (DAX intra)
	extra-axonal radial diffusivity (DRAD extra)
	microscopic fractional anisotropy (micro FA)
	extra-axonal axial diffusivity (DAX extra)
	intra-axonal water fraction (V intra)
	extra-axonal water fraction (V extra)
	cerebrospinal fluid fraction (vCSF)
	microscopical axial diffusivity (micro AX)
	microscopic radial diffusivity (micro RD)
	microscopical apparent diffusion coefficient (micro ADC)
Diffusion Kurtosis Imaging (DKI)	mean kurtosis (MK)
	radial kurtosis (RK)
	axial kurtosis (AK)
Diffusion Tensor Imaging (DTI)	fractional anisotropy (FA)
	axial diffusivity (AD)
	mean diffusivity (MD)
	radial diffusivity (RD)
Spherical Mean Technique (SMT)	fractional anisotropy (SMT FA)
	mean diffusivity (SMT md)
	transverse diffusion coefficient (SMT trans)
	longitudinal diffusion coefficient (SMT long)
Multi-compartment Spherical Mean Technique (mcSMT)	extra-neurite microscopic mean diffusivity (mcSMT extra md)
	extra-neurite transverse microscopic diffusivity (mcSMT extra trans)
	mc SMTdiffusion coefficient (SMT mcd)
	intra-neurite volume fraction (mcSMT intra)
White Matter Tract Integrity (WMTI)	axonal water fraction (AWF)
	radial extra-axonal diffusivity (radEAD)
	axial extra-axonal diffusivity (axEAD)

ST11. Whole-brain metrics' age sensitivity: comparing diffusion metric prediction models with and without age

Metric	Full[1]	Reduced[2]	χ^2	*p*	p_{Holm}
vintra	77831.60	75474.15	4714.90	<.001	<.001
vextra	80010.15	79172.34	1675.61	<.001	<.001
vCSF	104258.51	102045.76	4425.50	<.001	<.001
microRD	67462.88	62714.78	9496.19	<.001	<.001
microFA	91462.73	87450.56	8024.34	<.001	<.001
microAx	68450.93	67363.35	2175.17	<.001	<.001
microADC	72076.62	67934.52	8284.22	<.001	<.001
DRADextra	105039.68	102688.39	4702.58	<.001	<.001
DAXintra	66812.02	64826.61	3970.82	<.001	<.001
DAXextra	80174.07	77914.51	4519.12	<.001	<.001
RK	44127.87	41757.04	4741.66	<.001	<.001
AK	87172.98	85483.34	3379.29	<.001	<.001
MK	66245.64	64166.15	4158.99	<.001	<.001
FA	93186.73	88785.16	8803.13	<.001	<.001
MD	76490.81	72273.83	8433.95	<.001	<.001
RD	73140.15	68320.02	9640.26	<.001	<.001
AD	76516.84	74699.38	3634.93	<.001	<.001
smtFA	125878.97	124092.64	3572.65	<.001	<.001
smtLong	287398.36	285277.70	4241.31	<.001	<.001
smtMD	320273.27	317056.32	6433.91	<.001	<.001
smtTrans	342834.04	339946.36	5775.35	<.001	<.001
smtMCintra	74365.62	72478.15	3774.94	<.001	<.001
smtMCextraMD	314618.75	311214.26	6809.00	<.001	<.001
smtMCextratrans	303909.44	300508.10	6802.69	<.001	<.001
smtMCd	290508.70	290389.47	238.47	<.001	<.001
AWF	102689.44	100136.63	5105.61	<.001	<.001
axEAD	-14423.53	-14426.86	6.66	0.08	0.08
radEAD	10331.86	10122.73	418.26	<.001	<.001

1 Full = full model log likelihood
2 Reduced = reduced model log likelihood

ST12. Comparisons of linear and generalized additive models predicting whole-brain diffusion metrics

Metric	LM AIC	GAM AIC	LM BIC	GAM BIC	LM R^2_{adj}	GAM R^2_{adj}
vintra	-155639.20	-155638.58	-155537.39	-155545.25	0.13	0.13
vextra	-159996.29	-159998.05	-159894.48	-159904.72	0.05	0.05
vCSF	-208493.03	-208487.27	-208391.21	-208393.94	0.13	0.13
microRD	-134901.75	-134894.29	-134799.94	-134800.96	0.25	0.25
microFA	-182901.46	-182896.42	-182799.64	-182803.09	0.21	0.21
microAx	-136877.86	-136869.60	-136776.05	-136776.28	0.08	0.08
microADC	-144129.25	-144118.74	-144027.44	-144025.42	0.23	0.23
DRADextra	-210055.37	-210046.77	-209953.56	-209953.44	0.13	0.13
DAXintra	-133600.03	-133591.04	-133498.22	-133497.71	0.11	0.11
DAXextra	-160324.15	-160312.75	-160222.34	-160219.42	0.13	0.13
RK	-88231.74	-88223.96	-88129.93	-88130.63	0.14	0.14
AK	-174321.97	-174323.26	-174220.16	-174229.93	0.09	0.09
MK	-132467.28	-132465.03	-132365.47	-132371.70	0.12	0.12
FA	-186349.46	-186338.20	-186247.65	-186244.88	0.23	0.23
MD	-152957.62	-152950.05	-152855.81	-152856.72	0.22	0.22
RD	-146256.29	-146248.25	-146154.48	-146154.92	0.24	0.24
AD	-153009.69	-153006.22	-152907.88	-152912.90	0.15	0.15
smtFA	-251733.94	-251735.36	-251632.13	-251642.03	0.10	0.10
smtLong	-574772.72	-574753.50	-574670.90	-574660.18	0.12	0.12
smtMD	-640522.55	-640508.56	-640420.74	-640415.23	0.18	0.18
smtTrans	-685644.08	-685645.94	-685542.27	-685552.61	0.16	0.16
smtMCintra	-148707.24	-148707.45	-148605.43	-148614.13	0.10	0.10
smtMCextraMD	-629213.51	-629198.88	-629111.69	-629105.56	0.18	0.18
smtMCextratrans	-607794.88	-607787.65	-607693.07	-607694.32	0.18	0.18
smtMCd	-580993.40	-580991.12	-580891.59	-580897.79	0.01	0.01
AWF	-205354.88	-205349.62	-205253.07	-205256.29	0.14	0.14
axEAD	28871.07	28869.86	28972.88	28963.19	0.00	0.00
radEAD	-20639.72	-20639.85	-20537.91	-20546.52	0.01	0.01

LM = linear model, GAM = generalized additive model, AIC = Akaike information criterion, BIC = Bayesian information criterion. The numbers are derived from the six diffusion approaches' 28 metrics following Equation 1 for linear models and all variables of the equation allowing splines for non-linear models.

ST13. Variance explained by principal components of white matter metrics

	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6	Component 7	Component 8	Component 9	Component 10
Full Multimodal	0.3500	0.1065	0.0583	0.0393	0.0328	0.0208	0.0193	0.0172	0.0144	0.0133
Mean Multimodal	0.6474	0.2085	0.0690	0.0354	0.0199	0.0069	0.0046	0.0029	0.0020	0.0010
BRIA	0.3759	0.0963	0.0706	0.0533	0.0377	0.0291	0.0238	0.0177	0.0140	0.0128
DKI	0.4358	0.0909	0.0479	0.0309	0.0294	0.0228	0.0174	0.0153	0.0137	0.0123
DTI	0.4072	0.0816	0.0532	0.0434	0.0353	0.0249	0.0206	0.0165	0.0145	0.0138
SMT	0.3393	0.1942	0.0603	0.0336	0.0255	0.0215	0.0207	0.0173	0.0166	0.0130
SMT mc	0.3404	0.1648	0.0585	0.0450	0.0329	0.0224	0.0203	0.0165	0.0155	0.0127
WMTI	0.2711	0.1277	0.0421	0.0379	0.0284	0.0279	0.0217	0.0195	0.0178	0.0168

Eight principal component analyses (PCA) were run: six PCA addressing the different diffusion approaches, one addressing the multimodal average scores (mean multimodal) and one the multimodal model, containing all data (full multimodal). The first four components from all PCA were deemed meaningful based on the proportion of variance explained in the WM data.

ST14. Model performance and BAG beta values for multimodal and diffusion-approach specific principal component predictions from multimodal and diffusion approach-specific BAG and covariates

Predicted Component	Approach	R^2	R_{adj}^2	*T*	*p*	b_{BAG}
1	Full Multimodal	0.0141	0.0138	46.0160	<0.0001	-0.0151
1	Mean Multimodal	0.5054	0.5052	3651.2999	<0.0001	-0.6726
1	BRIA	0.3439	0.3437	1685.8645	<0.0001	-1.8093
1	DKI	0.2263	0.2261	940.7306	<0.0001	-0.8142
1	DTI	0.3575	0.3573	1789.7236	<0.0001	-1.0817
1	SMT	0.3462	0.3460	1841.6494	<0.0001	-1.0953
1	SMT mc	0.3152	0.3150	1480.6879	<0.0001	-1.0782
1	WMTI	0.3720	0.3718	1905.4699	<0.0001	-0.8474
2	Full Multimodal	0.0247	0.0244	81.5815	<0.0001	-0.0332
2	Mean Multimodal	0.0179	0.0177	65.2632	<0.0001	0.0522
2	BRIA	0.0400	0.0397	134.1182	<0.0001	0.1030
2	DKI	0.0191	0.0188	62.7254	<0.0001	0.1090
2	DTI	0.1517	0.1514	575.2169	<0.0001	-0.0587
2	SMT	0.0022	0.0019	7.6338	<0.0001	-0.0159
2	SMT mc	0.0784	0.0781	273.5101	<0.0001	0.2988
2	WMTI	0.0563	0.0560	191.8700	<0.0001	0.0876
3	Full Multimodal	0.1148	0.1145	417.0057	<0.0001	-0.0086
3	Mean Multimodal	0.0003	0.0001	1.1795	0.2991	-0.0023
3	BRIA	0.1842	0.1839	726.2248	<0.0001	0.4627
3	DKI	0.2564	0.2562	1109.0911	<0.0001	0.1705
3	DTI	0.2500	0.2498	1072.1366	<0.0001	-0.2096
3	SMT	0.2465	0.2463	1137.8183	<0.0001	0.2908
3	SMT mc	0.1844	0.1841	726.9496	<0.0001	-0.1839
3	WMTI	0.1751	0.1749	682.7749	<0.0001	0.1266
4	Full Multimodal	0.1279	0.1276	471.5012	<0.0001	0.0002
4	Mean Multimodal	0.0999	0.0997	396.8655	<0.0001	0.0672
4	BRIA	0.0880	0.0877	310.4268	<0.0001	-0.0049
4	DKI	0.0688	0.0685	237.5608	<0.0001	0.0302
4	DTI	0.0696	0.0693	240.5501	<0.0001	-0.1682
4	SMT	0.1040	0.1038	403.8253	<0.0001	-0.0603
4	SMT mc	0.1639	0.1637	630.6907	<0.0001	0.0803
4	WMTI	0.2239	0.2236	927.6834	<0.0001	0.1036

The table shows predictions of the first four components retrieved from the respective models (as done for brain age predictions, see **Table 1**), using BAG, age, sex, site, as well as age-sex and sex-site interactions as predictors (Equation 1). Both these four components as well as multimodal and approach-specific BAGs are based on data limited to the particular uni- or multi-modal approach and vary therefore in their number of metrics (**Table 1**).

ST15. Top five diffusion metrics ranked by gain in age prediction accuracy

BRIA	DKI	DTI	SMT	mcSMT	WMTI	Multimodal
Micro FA fornix (54957)	MK fornix (39662)	MD fornix (50535)	MD fornix (43563)	Intra fornix (38043)	AWF fornix (52531)	Micro FA Fornix (67749)
Micro RD right external capsule (22860)	RK fornix (26954)	RD FMIN (18386)	MD right anterior corona radiata (24675)	Extra trans Fornix (35799)	RadEAD ATRL (12328)	RD Fornix right Stria terminalis (17664)
Micro FA FMIN (10081)	AK right anterior limb of internal capsule (16340)	RD fornix right stria terminalis (15431)	MD SLFR (19451)	Extratrans right external capsule (15369)	RadEAD right anterior corona radiata	AK anterior right limb of internalcapsule (17664)
Micro FA fornix right stria terminlis (9853)	AK fornix (10516)	AD fornix (9637)	MD FMIN (13527)	Extra MD anterior left limb of internal capsule (6254)	RadEAD IFOFR (9828)	RadEAD right anterior corona radiata (17375)
Micro RD Fornix right stria terminalis (9812)	AK left superior fronto occipital fasciculus (6850)	FA fornix left stria terminalis (9283)	FA fornix (12011)	Extra trans anterior right limb of internal capsule (6126)	RadEAD right external capsule (9793)	RadEAD SLFR (15840)

Table values can be read as feature name (gain value). Gain refers to the improvement in accuracy brought by a feature to the branches it is on[43]. Multimodal refers to an approach using the diffusion metrics from all diffusion approaches. Cells containing Fornix are marked in green.
Tracts are abbreviated as follows: ATRL = anterior thalamic radiation left, FMIN = Forceps minor, IFOFR = inferior fronto-occipital fasciculus right, SLFR = superior longitudinal fasciculus right

ST16. Brain age prediction model performance e*xcluding fornix features* and uncorrected brain age – chronological age correlations comparison

Name	MRI features	R^2	RMSE	MAE	Prediction-Age Correlation	Uncorrected Brain Age Correlation Difference to All Data
BRIA	700	0.527 (0.010)	5.131 (0.042)	4.129 (0.033)	0.727 [0.722, 0.732]	-0.007* [-0.009, -0.004]
DKI	182	0.550 (0.015)	5.108 (0.070)	4.105 (0.065)	0.742 [0.737, 0.747]	-0.006* [-0.008, -0.003]
DTI	252	0.555 (0.013)	5.078 (0.066)	4.079 (0.061)	0.745 [0.745, 0.750]	-0.005* [-0.007, -0.003]
SMT	252	0.507 (0.008)	5.347 (0.042)	4.309 (0.028)	0.713 [0.707, 0.718]	-0.009* [-0.011, -0.006]
mcSMT	252	0.488 (0.011)	5.342 (0.045)	4.303 (0.036)	0.699 [0.693, 0.705]	-0.015* [-0.018, -0.012]
WMTI	182	0.566 (0.012)	5.018 (0.062)	4.031 (0.052)	0.753 [0.748, 0.757]	-0.003* [-0.006, -0.001]
Mean scores multimodal	28	0.393 (0.012)	5.932 (0.051)	4.812 (0.046)	0.627 [0.621, 0.634]	0 [-0.0001, 0.0001]
Full model multimodal	1904	0.636 (0.012)	4.591 (0.077)	3.677 (0.039)	0.798 [0.794, 0.802]	-0.006* [-0.007, -0.004]

In the above only fornix features are excluded, while QC and all other steps are kept as described in the Methods section. *Importantly*, radiations from the fornix to other tracts such as fornix to stria terminalis radiations were not excluded. Compare results from the full model in Table 1 for uncorrected prediction-age correlations which were the basis for the final column.
* p<.001

Brain age predictions from models containing fornix metrics are consistently stronger correlated with age than predictions from models not containing fornix (*rs*<-0.003,*ps*<.001).