



Age specificity in fornix-to-hippocampus association

Yunghin Gazes¹ · Peipei Li¹ · Emily Sun¹ · Qolamreza Razlighi¹ · Angeliki Tsapanou¹

Published online: 5 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Both white and grey matter atrophy with age, but it is still unclear how decline in white matter relates to decline in grey matter, and how this relationship varies with age. In a group of healthy adults from 20 to 80 years old, divided into three age groups by tertiles, we cross-sectionally examined the white-to-grey matter associations in the fornix and the hippocampus, and tested if and how the fornix-to-hippocampus relationship differs across the age groups. Both structures were also tested as predictors for performance on a memory test, the Selective Reminding Task (SRT). Participants were imaged with T1-weighted magnetic resonance imaging (MRI) and diffusion-weighted imaging (DWI), from which the hippocampal volume, fractional anisotropy (FA), and mean diffusivity (MD) for the bilateral crus and body of the fornix were calculated. Our data showed that even after accounting for age, sex, and motion parameters, fornix integrity predicted hippocampal volume in the two older age groups (middle and old age) for the crus of the fornix, and only in the oldest age group for the body of the fornix. Furthermore, fornix integrity significantly predicted SRT performance, whereas hippocampal volume did not; this relationship was also observed only in the oldest age group, and absent in the two younger age groups. The age specificity of the relationships suggests that the fornix-to-hippocampus relationship only manifests once brain structures begin to atrophy in old age, and that fornix integrity is a more sensitive measure for episodic memory than is hippocampal volume.

Keywords Diffusion tensor imaging · Episodic memory · Aging · Fornix · Hippocampus · T1-weighted imaging

White (WM) and grey matter (GM) are interconnected structures, but have been treated as independent factors by most studies due to the fact that different imaging techniques are sensitive to the two different structures. The structural integrity of WM is usually quantified with diffusion weighted imaging (DWI) while that of GM is measured with T1-weighted magnetic resonance imaging (MRI). A major difference between the two techniques is that DWI reflects microstructural properties such as the diffusion of water molecules in the brain, while T1-weighted MRI reflects macrostructural properties such as thickness and volume. This difference in scale suggests that even if both grey and WM integrity decline at the same rate, the damage should be detected earlier in WM due to the greater sensitivity of DWI to microstructural damage.

A small number of neuroimaging studies have examined the association between WM and GM measures in healthy and clinical populations. While both white and grey matter measures have been incorporated into the same model as are predictors for cognitive performance (He et al. 2012; Hong et al. 2015; Stricker et al. 2013), the direct association between the two structures has rarely been explored. For example, Stricker et al. (2013) controlled for cortical thickness and found significant associations between medial temporal FA and memory, as well as between parietal FA and executive function, but did not examine the direct relationship between cortical thickness and FA. Steenwijk et al. (2015) was one of the studies that directly related WM to GM measures in 208 multiple sclerosis patients, using tractography to segment WM tracts directly connected to each of the 82 GM regions (68 cortical and 14 deep gray matter regions), then relating the FA of the WM tracts to the respective GM seed regions. In relapsing-remitting MS patients, FA was found to predict GM atrophy in both cortical and deep GM regions while FA only predicted GM atrophy in deep GM regions for secondary-progressive patients. In an aging study, Storsve et al. (2016) directly related longitudinal GM atrophy to 18 major WM

✉ Yunghin Gazes
YL2107@cumc.columbia.edu

¹ Cognitive Neuroscience Division, Department of Neurology,
Columbia University Medical Center, 630 W 168th Street, P&S Box
16, New York, NY 10032, USA

tracts in 201 healthy adults aged 23 to 87 years old and demonstrated spatial specificity in the relationship between GM and WM regions.

While these studies established a direct link between GM and WM, an important question that remains unaddressed is whether the WM-to-GM relationship differs with age. Our current study examined the interaction between this relationship and age by focusing on two directly connected structures, the hippocampus and the fornix. While a complex system of connections are formed with the hippocampus, the fornix consists of the major fiber bundles that exit the hippocampus and connect with cortical areas, and plays an important role in memory and learning functions (Duvernoy et al. 2013). Thus, the fornix to hippocampus pairing is the best candidate for understanding the relationships among GM, WM, and performance on an episodic memory task in the context of aging. We hypothesized that atrophy in the hippocampus and the fornix should be closely correlated, and that age differentially influences the relationships between these two structures and task performance.

Relationships between memory performance and the structural characteristics of the hippocampus and fornix have also been reported in a number of studies (Douet and Chang 2014; Fletcher et al. 2013), but most of these studies examined the virtues of the fornix as a biomarker for Alzheimer's disease, rather than the age-related differences in the fornix-hippocampus relationship and its interaction with age. An exception was in Mielke et al. (2012)—although the focus of the paper was on the fornix as a potential AD biomarker, it also directly correlated WM and GM, finding that the fornix FA and three additional WM integrity measures were all correlated with hippocampal volume.

Based on previous literature, we used the Selective Reminding task (SRT) (Buschke and Fuld 1974) to measure episodic memory. SRT is routinely used in clinical evaluations of episodic memory ability, for which performance has been associated with the left hippocampal volume in multiple sclerosis patients (Pardini et al. 2014) and in patients with memory complaints (Quenon et al. 2016).

Our current study further contributed to research in brain structure by demonstrating age-related differences in the fornix-to-hippocampal association, as well as by providing support for the greater value of using diffusivity measures relative to using hippocampal volume measures in predicting episodic memory performance in older adults.

Method

Participants

327 participants aged 20 to 80 years old were included in this study. Demographic details are listed in Table 1, which shows

Table 1 Demographic information

| Group | n | Age, M(SD) | Education, M(SD) | Sex, %(n) F |
|---------------|-----|-------------|------------------|-------------|
| Young: <39 | 108 | 28.8 (4.71) | 16.0 (2.28) | 67% (72) |
| Middle: 39–63 | 109 | 52.5 (7.47) | 16.1 (2.26) | 48% (52) |
| Old: ≥ 64 | 110 | 69.6 (4.09) | 16.7 (2.60) | 49% (54) |

the three age groups used in group-specific analyses when probing age interactions. Groups were divided by tertiles (the 33rd and 67th percentiles divided the sample into three approximately equal subsets). Participants were recruited using established market mailing procedures to standardize the recruitment procedures of young and old adults. Participants who responded to the mailings were screened over telephone to ensure fulfillment of basic inclusion criteria (right handed, English speaking, no psychiatric or neurological disorders that could potentially affect cognition, normal or corrected-to-normal vision, and no current use of central-nervous-system-targeting medications). Individuals that passed the telephone screening were then screened in-person, and participants aged 60 years or older were required to score at least 135 on the Mattis Dementia Rating Scale for inclusion in the study. Informed consent, as approved by the Internal Review Board of the College of Physicians and Surgeons of Columbia University, was obtained prior to study participation, and after the nature and risks of the study were explained. Participants were paid for their participation in the study.

Selective reminding task

The total number of items recalled in the Selective Reminding Task (SRT) (Buschke and Fuld 1974) was used to assess memory performance. Participants in this task were first read a list of 12 words, then immediately asked to recall as many as they could. For the five following trials, they were reminded of the words that they did not report, then asked once again to recall all of the words on the list. Thus, the maximum number of total recalled words is 72 (12 items × 6 trials).

MRI acquisition

MRI images were acquired in a 3.0 T Philips Achieva magnet using a standard quadrature head coil. A T1-weighted scout image was acquired to determine subject position. 165 contiguous 1 mm axial T1-weighted images of the whole brain were acquired for each subject with an MPRAGE sequence using the following parameters: TR = 6.5 ms, TE = 3 ms, flip angle = 8°, 256 × 256 acquisition matrix, and 240 mm field of view. Two DWI images were acquired in 56 directions using the following parameters: b = 800 s/mm², TE = 69 ms, TR = 11,032 ms, flip angle = 90°, in-plane resolution 112 × 112 voxels, acquisition time = 12 min 56 s, slice thickness =

2 mm (no gap), and 75 slices. All T1 scans were reviewed by a neuroradiologist to check for potentially clinically significant findings, such as abnormal neural structure. No clinically significant finding was identified in the subjects included in our current study.

Structural T1 processing

Each subject's structural T1 scans were reconstructed using FreeSurfer v5.1 (<http://surfer.nmr.mgh.harvard.edu/>). The accuracies of FreeSurfer's subcortical segmentation and cortical parcellation (Fischl et al. 2002; Fischl et al. 2004) have been reported to be comparable to manual labeling. Each subject's WM and GM boundaries, as well as GM and cerebral spinal fluid boundaries, were visually inspected slice by slice, manual control points were added in the case of any visible discrepancy, and reconstruction was repeated until satisfactory results were reached within every subject. Hippocampal volume for both hemispheres was calculated through FreeSurfer's automated subcortical segmentation (Fischl et al. 2002). The subcortical structure borders were plotted by TKmedit visualization tools and compared against the actual brain regions. In case of discrepancies, the parcellations were corrected manually.

Diffusion tensor image processing

The two DWI images were concatenated along with their corresponding b-vectors, b-values, and processed through FreeSurfer's TRActs Constrained by UnderLying Anatomy (TRACULA) processing stream (Yendiki et al. 2011) which comprised of preprocessing through the FMRIB Software Library (FSL) FDT toolbox (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT>), including eddy current and motion correction, and diffusion tensor modeling to produce the fractional anisotropy (FA) and the mean diffusivity (MD) images. Masks from the Johns Hopkins University WM labels atlas for the left and right crura and the body of the fornix were used to extract the MD and FA for each of the three structures. Transformations of the masks from Montreal Neurological Institute (MNI) space to each participant's native space were calculated as part of the TRACULA pipeline, which consisted of an array of registration tools including tkregister2, FreeSurfer's boundary-based registration (Greve and Fischl 2009), and FSL's FLIRT (Jenkinson et al. 2002; Jenkinson and Smith 2001). For each participant, three mean MD and three FA values were produced and entered into the regression models described in the Statistical Analysis section below. Movement parameters were produced, but only the mean translation and rotation values were used in this study. Only FA and MD were examined in our study and did not include axial and radial diffusivity values because factor analysis showed that the four DTI measures are so highly correlated

that they constitute at most two unique factors (unpublished result). Therefore, we only examined two measures that dominated the two factors to minimize the chance of type I and type II errors in hypothesis testing: (1) If we didn't correct for multiple comparisons and thus overestimating the effects, then we would risk producing false positives, and (2) if we corrected for multiple comparisons, which assumes the measures are independent of each other, then we would risk producing false negatives. Therefore, we examined two measures that are more representative of the two major sources of variability in the four DTI measures and corrected for multiple comparisons as described in the Statistical analysis section below.

Statistical analysis

All analyses were conducted in SPSS (IBM; v23) and $p < .05$, corrected for multiple comparisons as described in each subsection below.

Fornix integrity predicting hippocampal volume Multiple linear regression was used to test the relationship of the mean FA and MD of the fornix with hippocampal volume, in which hippocampal volume was the dependent variable and either FA or MD was the independent variable. Covariates included age, intracranial volume (ICV), sex, and the means of translational and rotational motion during the DWI scan. For the models involving the crura of the fornix, each crus was used to predict the ipsilateral hippocampal volume (i.e. the left crus predicting the left hippocampal volume). For the model with the body of the fornix as a predictor, the mean over the left and right hippocampal volumes was the dependent variable. Since two DTI measures, FA and MD, were examined along with the three segments of the fornix (the left and the right crura and the body), the p value after correction for multiple comparisons was .0083 (.05/6).

Structural measures predicting SRT performance Multiple linear regression was performed with SRT total items recalled as the dependent variable, and the hippocampal volume, either MD or FA, age, education, and motion parameters as independent variables. As there is obviously no laterality associated with SRT performance, the crus and hippocampal measures were averaged across the two hemispheres. Thus, the mean MD or FA for the crus of the fornix and the mean hippocampal volume for each participant were entered into the models. The p value after correction for multiple comparisons was .0125 (.05/4: two DTI measures with two segments of fornix).

Age-dependent differences To examine whether the fornix-hippocampus and structure-SRT performance relationships differ across age groups, an age interaction term was added

to the two above models. For the prediction of hippocampal volume, an age by DTI interaction was added to the model. For the prediction of SRT performance, an age by DTI measure and an age by hippocampal volume interaction were added to the model. For models that showed significant interaction with age, the total sample was divided into three age groups as described in the Participants section. The same model was tested in each group without interaction terms to examine the interactions in post hoc tests, which minimizes the number of multiple comparisons by conducting post-hoc tests of significant interactions with age.

Results

Two groups of models were tested: The first predicted hippocampal volume, and the second predicted SRT performance. Models predicting hippocampal volume differed by the DTI measure (FA or MD), fornix segments (crus or body), and laterality (left or right hemisphere) used; while models predicting SRT performance combined the left and right hemispheres, thus differing only by the DTI measures and fornix segments used as predictors.

Age effect on fornix DTI measures

All segments of the fornix showed lower FA and higher MD with older age, after accounting for sex and two motion parameters. Table 2 shows the statistics associated with Age and Sex effects.

Fornix predicting hippocampal volume

When employing age, fornix integrity, two motion parameters, ICV, sex, and age by fornix interaction as independent variables, age by fornix interaction was significant for the MD measure for the crus of the fornix, and both MD and FA measures for the body of the fornix. However, only the age x MD of the left Crus and age x FA of the body of the fornix exceeded the p value corrected for multiple comparisons. The statistics for these models can be found in Table 3. For significant Age x Fornix interactions, the interactions were further examined post-hoc in each of the age tertiles. In the post-hoc models, the crus and body of the fornix did not predict hippocampal volume in the young group, but did significantly predict it for the old group. The mean FA for the body of the fornix in the middle group also predicted hippocampal volume. For all of these significant associations, greater intact WM integrity (higher FA or lower MD) was associated with greater hippocampal volume. Figure 1 demonstrates the associations of the left crus of the fornix with left hippocampal volume across the three age groups.

Fornix and hippocampus predicting SRT performance

Using age, two motion parameters, education, fornix integrity, hippocampal volume, and interactions with age as independent variables, the age by fornix interaction exceeded the p value of .05 only for the mean MD of the crus of the fornix, but did not survive correction for multiple comparisons, and the age by hippocampal volume interaction was significant after correction for multiple comparisons for FA of the crus and both MD and FA for the body of the fornix. Examining the three age groups separately showed that the MD of the crus of the fornix significantly predicted SRT performance only in the old age group, while hippocampal volume was not a significant predictor of SRT performance in the old and middle age groups (see Table 4). Figure 2 shows the association between FA for the crus of the fornix and SRT performance in which greater fornix integrity (higher FA) was associated with better SRT performance.

Unexpectedly, hippocampal volume in the young group was negatively associated with SRT performance, as larger volume was associated with worse performance. However, when considering the general trend across all subjects, hippocampal volume was positively associated with SRT performance, indicating that it may have been a chance occurrence that this unexpected relationship was manifested in the young group. It would be interesting to try replicating this trend in a different sample and to then identify the source of the variability if replication were successful. Figure 3 shows the association between hippocampal volume and SRT total recalled, with the black line showing the linear trend for all of the groups combined.

Discussion

In a group of participants ranging from young to old adulthood, our study examined if and how the integrity of two sections of the fornix relate to hippocampal volume, and how the two structures contribute to variability in the performance of a typical episodic memory task, SRT. Inclusion of a wide age range also conferred the ability to test the effect of age on the fornix-hippocampus relationship and on the structure-to-SRT performance relationship. By conducting post hoc tests on the significant age interactions, we observed that in the oldest age group, the integrity of the fornix predicted hippocampal volume even after accounting for motion during scans, and that fornix integrity predicted SRT performance even after accounting for hippocampal volume and a number of other covariates. None of these results were observed in the youngest age group, and were observed only for the fornix-SRT relationship in the middle age group.

Even though our analysis was cross-sectional, the association found between fornix integrity and hippocampal volume is consistent with a longitudinal aging study in which decline in GM thickness over an average of 3.6 years was associated

Table 2 Age and Sex effects on each of the fornix DTI measures

| IV | L Crus FA | | L Crus MD | | R Crus FA | | R Crus MD | | Body FA | | Body MD | |
|-----|-----------|-------|-----------|-------|-----------|-------|-----------|-------|---------|-------|---------|-------|
| | t | p | t | p | t | p | t | p | t | p | t | p |
| Age | -11.6 | <.001 | 11.7 | <.001 | -13.0 | <.001 | 12.1 | <.001 | -15.0 | <.001 | 14.2 | <.001 |
| Sex | -2.61 | 0.01 | 4.14 | <.001 | -2.04 | 0.042 | 2.98 | 0.003 | -3.04 | 0.003 | 2.96 | 0.003 |

IV Independent variable; L left; R right

with decrease in WM integrity for 18 major WM tracts (Storsve et al. 2016). This relationship was reported for the thickness of the end regions connected by the axonal tracts, as well as for the thickness of the surrounding cortex. However, the study only examined the entire sample, which consisted of participants aged 23 to 87 years, and did not examine the effects of interaction with age. Our study elaborates on these results by demonstrating that the GM to WM relationship may only hold for older age groups, albeit only considering the relationship specifically in the hippocampus and fornix. It would be instrumental to examine age interaction effects in other brain structures for future studies.

Our finding that the integrity of the fornix predicted hippocampal volume only in the oldest age group suggests that in the absence of any age-related damage, the variability in hippocampal volume may be more closely related to other factors such as lifestyle and genetics than it is to the integrity of the fornix. As the two connected structures decline with age, and atrophy in one structure induces atrophy in the other, their association becomes stronger. A similar relationship between fornix integrity and hippocampal volume was reported in patients with Mild Cognitive Impairment and Alzheimer's disease (Mielke et al. 2012; Zhuang et al. 2013), both of which

possess structural atrophy as a hallmark. Other conditions were also associated with a simultaneous alteration of various fornical and hippocampal aspects, although the relationship between the two was not often directly examined. For example, one study found that lower fornical FA, as well as smaller hippocampal volume, was present in early-psychosis subjects (Baumann et al. 2016), and another found that subjects with right or left temporal lobe epilepsy exhibited increased MD in both the bilateral hippocampus and fornix (Chiang et al. 2016). In all these diseases, the fornix integrity was closely correlated with measures of hippocampal integrity.

Even though we used fornix integrity to predict hippocampal volume in our models, it is important to note that the same exact significant statistics (t and p values, not the beta value) would be obtained if hippocampal volume were used to predict fornix integrity. We chose to use fornix integrity as the predictor in this study because the DTI measures for the fornix are theoretically more sensitive to structural damages than is T1-weighted imaging, as T1-weighted MRI characterizes the hippocampus at the macrostructural level, providing information about GM thickness and volume. Microstructural measures calculated from DWI enable description of WM structure at a level indiscernible at the macroscopic level. Models

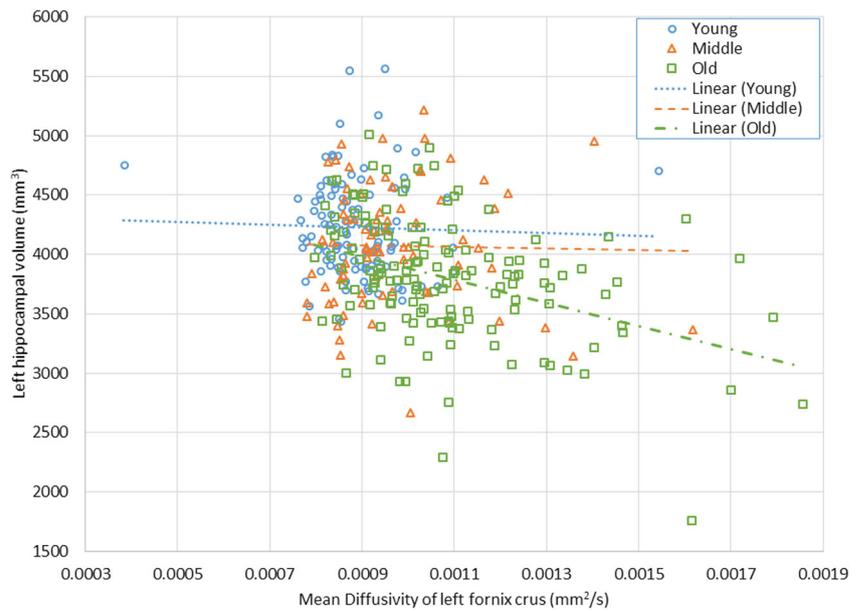
Table 3 Statistical results for the prediction of hippocampal volume

| Model | | | | | Fornix | | Age | | Age x Fornix | |
|--------|----------------------|-----|--------|--------|--------------|-------|-------------|-------|--------------|-------|
| Group | IVs | DTI | DV | df | F/t | p | F/t | p | F | p |
| All | L Crus, age x L Crus | FA | L hVol | 1, 319 | 0.013 | 0.909 | 5.66 | 0.018 | 1.26 | 0.262 |
| All | R Crus, age x R Crus | FA | R hVol | 1, 319 | 0.966 | 0.326 | 9.45 | 0.002 | 3.46 | 0.064 |
| All | R Crus, age x R Crus | MD | R hVol | 1, 319 | 1.22 | 0.270 | 1.78 | 0.183 | 5.22 | 0.023 |
| All | Body, age x body | MD | hVol | 1, 319 | 0.275 | 0.601 | 1.13 | 0.289 | 5.19 | 0.023 |
| All | L Crus, age x L Crus | MD | L hVol | 1, 319 | 1.59 | 0.208 | 2.58 | 0.109 | 7.4* | 0.007 |
| Young | L Crus | MD | L hVol | 106 | -0.641 | 0.523 | -0.963 | 0.338 | – | – |
| Middle | L Crus | MD | L hVol | 107 | -1.91 | 0.059 | -1.22 | 0.226 | – | – |
| Old | L Crus | MD | L hVol | 108 | -3.87 | <.001 | -1.45 | 0.150 | – | – |
| All | Body, age x body | FA | hVol | 1, 319 | 0.56 | 0.455 | 17.8 | <.001 | 9.1* | 0.003 |
| Young | Body | FA | hVol | 106 | 1.44 | 0.152 | -0.328 | 0.744 | – | – |
| Middle | Body | FA | hVol | 107 | 2.61 | 0.011 | -1.73 | 0.087 | – | – |
| Old | Body | FA | hVol | 108 | 3.26 | 0.002 | -1.45 | 0.151 | – | – |

All models included age, sex, ICV, and motion parameters as independent variables in addition to the ones listed. Models with all age groups have F statistics listed whereas models for specific age groups have t statistics listed. Bolded = $p < .05$. * = $p < .05$ corrected for multiple correction.

IVs independent variables; FA fractional anisotropy; MD mean diffusivity; hVol hippocampal volume; L left; R right

Fig. 1 Left hippocampal volume vs. mean diffusivity of left fornix crus for the three age groups. Regression lines are shown for each age group



such as DTI enable detection of microstructural damages, and thus should detect decline at an earlier time point than can T1-weighted imaging.

With both fornix integrity and hippocampal volume in the same model as predictors of SRT performance, our observation that fornix integrity, but not hippocampal volume, significantly predicted SRT performance in older adults is also consistent with the idea that DTI is a more sensitive method of measuring atrophy. The macro- and microstructural development, as well as the decline, of underlying brain regions may result in one structure becoming the main determining factor

of cognitive processes at one age range, and another structure becoming the main factor at a later age. In old age, when neuronal atrophy is more pervasive, the most sensitive measure of brain integrity should be the most predictive factor for cognitive outcome.

In addition, neuronal atrophy with aging may also manifest more in WM than in GM. Postmortem brain studies reported that myelinated axonal fibers decreased by 45% from age 20 to age 80 (Mamer et al. 2003) and that WM volume decreased by 28% from age 20 to 90 (Pakkenberg and Gundersen 1997), while GM volume only decreased by 12.3% from age 20 to 90

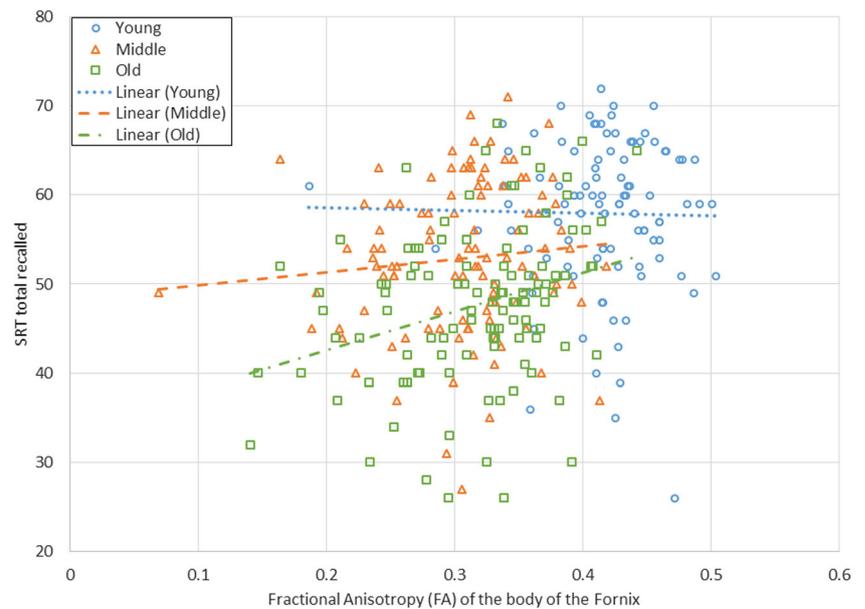
Table 4 Statistical results for the prediction of SRT performance

| Model | | | | | Fornix | | Age | | Age x Fornix | | hVol | | Age x hVol | |
|--------|------------------------------------|-----|-----|--------|-------------|-------|--------------|-------|--------------|-------|--------------|-------|--------------|-------|
| Group | IVs | DTI | DV | df | F/t | p | F/t | p | F | p | F/t | p | F | p |
| All | Crus, hVol, age x Crus, age x hVol | MD | SRT | 1, 308 | 1.74 | 0.188 | 1.97 | 0.162 | 4.35 | 0.038 | 8.39 | 0.004 | 6.1 | 0.014 |
| All | Crus, hVol, age x Crus, age x hVol | FA | SRT | 1, 308 | 0.629 | 0.429 | 21.5 | <.001 | 2.49 | 0.115 | 9.41 | 0.002 | 7.91* | 0.005 |
| Young | Crus, hVol | FA | SRT | 99 | -0.341 | 0.734 | -1.71 | 0.091 | - | - | -2.95 | 0.004 | - | - |
| Middle | Crus, hVol | FA | SRT | 104 | 0.705 | 0.482 | -2.15 | 0.034 | - | - | 0.472 | 0.638 | - | - |
| Old | Crus, hVol | FA | SRT | 108 | 2.49 | 0.014 | -2.05 | 0.043 | - | - | 0.259 | 0.796 | - | - |
| All | Body, hVol, age x body, age x hVol | MD | SRT | 1, 308 | <.001 | 0.997 | 9.2 | 0.003 | 0.337 | 0.562 | 11.1 | 0.001 | 9.86* | 0.002 |
| Young | Body, hVol | MD | SRT | 99 | -0.088 | 0.93 | -1.67 | 0.099 | - | - | -3.06 | 0.003 | - | - |
| Middle | Body, hVol | MD | SRT | 104 | -0.73 | 0.467 | -2.11 | 0.037 | - | - | 0.36 | 0.717 | - | - |
| Old | Body, hVol | MD | SRT | 108 | -1.33 | 0.188 | -2.28 | 0.024 | - | - | 0.423 | 0.673 | - | - |
| All | body, hVol, age x body, age x hVol | FA | SRT | 1, 308 | 0.559 | 0.455 | 20.9 | <.001 | 1.64 | 0.201 | 10.3 | 0.001 | 8.72* | 0.003 |
| Young | Body, hVol | FA | SRT | 99 | -0.412 | 0.682 | -1.6 | 0.114 | - | - | -3.03 | 0.003 | - | - |
| Middle | Body, hVol | FA | SRT | 104 | 0.238 | 0.813 | -2.09 | 0.039 | - | - | 0.44 | 0.663 | - | - |
| Old | Body, hVol | FA | SRT | 108 | 1.57 | 0.119 | -2.28 | 0.024 | - | - | 0.256 | 0.798 | - | - |

All models included age, edu, and motion parameters as independent variables in addition to the ones listed. Models with all age groups have F statistics listed whereas models for specific age groups have t statistics listed. Bolded = $p < .05$. * = $p < .05$ corrected for multiple correction.

IVs independent variables; FA fractional anisotropy; MD mean diffusivity; hVol hippocampal volume; L left; R right

Fig. 2 SRT total recalled vs. average of the mean fractional anisotropy for bilateral fornix crura for the three age groups. Regression lines are shown for each age group

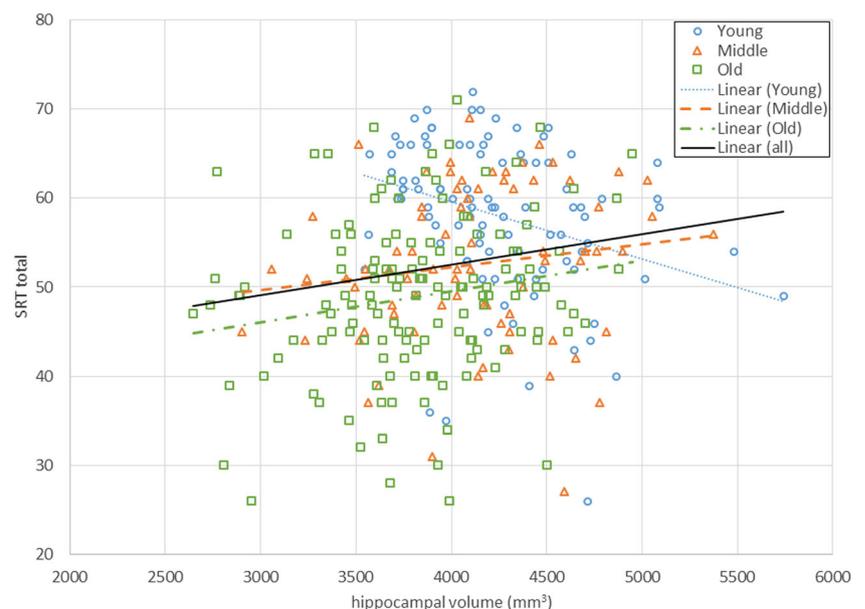


(Pakkenberg and Gundersen 1997). Given that complete loss of axonal tracts would result in neuronal death, the much smaller decline in GM volume suggests that WM decline should be in the form of axonal arbor reduction (Adalbert and Coleman 2013). However, decline in axonal arbor reduction should result in a decrease in the number of dendritic spines, and thus be observed as changes in thickness and volume in the connected GM structures.

To better understand the mechanism of WM degeneration however, more precise models, such as Neurite Orientation Dispersion and Density Imaging (Zhang et al. 2012), and imaging techniques, such as myelin water fraction (Billiet et al. 2015) should be used. While diffusion studies frequently

interpret the diffusion measures in terms of specific neural degenerative mechanisms such as axonal damage and myelin loss, and a number of studies have shown convincing evidence that changes in axonal damage were reflected in axial diffusivity (diffusivity in the longest axis) while myelin loss was reflected in radial diffusivity (mean diffusivity in the two shortest axes) (Concha et al. 2006; Song et al. 2003), other studies have shown that diffusivity measures from diffusion tensor models may be influenced by a myriad of other factors (Beaulieu 2002). For example, Concha et al. (2010) demonstrated that only the membrane of the axon influenced the diffusivity measures in axons of various diameters and with or without myelination, indicating that while axial and radial

Fig. 3 SRT total recalled vs. average of bilateral hippocampal volumes for the three age groups. Regression lines are shown for each age group and for all three groups together



diffusivity measures are influenced by different degenerative events, changes in these measures are not sufficient evidence for a specific degenerative event.

Our study demonstrated the age specificity of the fornix-hippocampus relationship, and illustrated that fornix integrity is a better predictor for memory performance in healthy adults than hippocampal volume. As with all cross-sectional data analyses, causal inference is limited and must be confirmed with longitudinal data to truly test the directional relationship between hippocampal volume and fornix integrity. The age specificity of the fornix-hippocampus relationship should also be tested in other WM-GM pairings to examine if it is similarly representative of WM-GM relationships in the rest of the brain. Replication of this relationship in the rest of the brain and with longitudinal changes will advance clinical diagnosis of neural degenerative diseases by placing more emphasis on diffusion weighted images as a clinical diagnostic tool.

Funding This study was funded by National Institute of Health/Aging under grant numbers K01AG051777, RF1AG038465, and R01AG026158.

Compliance with ethical standards

Ethical approval Participants in this study were treated in accordance with the ethical standards of the Columbia University Institutional Review Board. They were only tested after they had a complete understanding of the risks and benefits involved in this research and had provided written consent for participation and use of their data.

Conflict of interest All authors declare that they have no conflict of interest.

References

- Adalbert, R., & Coleman, M. P. (2013). Review: Axon pathology in age-related neurodegenerative disorders. *Neuropathology and Applied Neurobiology*, 39(2), 90–108. <https://doi.org/10.1111/j.1365-2990.2012.01308.x>.
- Baumann, P. S., Griffa, A., Fournier, M., Golay, P., Ferrari, C., Alameda, L., . . . Conus, P. (2016). Impaired fornix-hippocampus integrity is linked to peripheral glutathione peroxidase in early psychosis. *Translational Psychiatry*, 6(7), e859. <https://doi.org/10.1038/tp.2016.117>.
- Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, 15(7–8), 435–455. <https://doi.org/10.1002/nbm.782>.
- Billiet, T., Vandenberghe, M., Madler, B., Peeters, R., Dhollander, T., Zhang, H., . . . Emsell, L. (2015). Age-related microstructural differences quantified using myelin water imaging and advanced diffusion MRI. *Neurobiology of Aging*, 36(6), 2107–2121. <https://doi.org/10.1016/j.neurobiolaging.2015.02.029>.
- Buschke, H., & Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology*, 24(11), 1019–1025.
- Chiang, S., Levin, H. S., Wilde, E., & Haneef, Z. (2016). White matter structural connectivity changes correlate with epilepsy duration in temporal lobe epilepsy. *Epilepsy Research*, 120, 37–46. <https://doi.org/10.1016/j.eplepsyres.2015.12.002>.
- Concha, L., Gross, D. W., Wheatley, B. M., & Beaulieu, C. (2006). Diffusion tensor imaging of time-dependent axonal and myelin degradation after corpus callosotomy in epilepsy patients. *Neuroimage*, 32(3), 1090–1099. <https://doi.org/10.1016/j.neuroimage.2006.04.187>.
- Concha, L., Livy, D. J., Beaulieu, C., Wheatley, B. M., & Gross, D. W. (2010). In vivo diffusion tensor imaging and histopathology of the fimbria-fornix in temporal lobe epilepsy. *The Journal of Neuroscience*, 30(3), 996–1002. <https://doi.org/10.1523/JNEUROSCI.1619-09.2010>.
- Douet, V., & Chang, L. (2014). Fornix as an imaging marker for episodic memory deficits in healthy aging and in various neurological disorders. *Frontiers in Aging Neuroscience*, 6, 343. <https://doi.org/10.3389/fnagi.2014.00343>.
- Duvernoy, H., Cattin, F., Risold, P.-Y., & ebrary Inc. (2013). The human hippocampus functional anatomy, vascularization and serial sections with MRI retrieved from <http://www.columbia.edu/cgi-bin/cul/resolve?clio10415430>.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., . . . Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1), 11–22.
- Fletcher, E., Raman, M., Huebner, P., Liu, A., Mungas, D., Carmichael, O., & Decarli, C. (2013). Loss of fornix white matter volume as a predictor of cognitive impairment in cognitively normal elderly individuals. *JAMA Neurology*, 70, 1389–1395. <https://doi.org/10.1001/jamaneurol.2013.3263>.
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>.
- He, J., Wong, V. S., Fletcher, E., Maillard, P., Lee, D. Y., Iosif, A. M., . . . DeCarli, C. (2012). The contributions of MRI-based measures of gray matter, white matter hyperintensity, and white matter integrity to late-life cognition. *AJNR. American Journal of Neuroradiology*, 33(9), 1797–1803. <https://doi.org/10.3174/ajnr.A3048>.
- Hong, Z., Ng, K. K., Sim, S. K., Ngeow, M. Y., Zheng, H., Lo, J. C., . . . Zhou, J. (2015). Differential age-dependent associations of gray matter volume and white matter integrity with processing speed in healthy older adults. *Neuroimage*, 123, 42–50. <https://doi.org/10.1016/j.neuroimage.2015.08.034>.
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156.
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2), 825–841.
- Marnier, L., Nyengaard, J. R., Tang, Y., & Pakkenberg, B. (2003). Marked loss of myelinated nerve fibers in the human brain with age. *The Journal of Comparative Neurology*, 462(2), 144–152. <https://doi.org/10.1002/cne.10714>.
- Mielke, M. M., Okonkwo, O. C., Oishi, K., Mori, S., Tighe, S., Miller, M. I., . . . Lyketsos, C. G. (2012). Fornix integrity and hippocampal volume predict memory decline and progression to Alzheimer's disease. *Alzheimers Dement*, 8(2), 105–113. <https://doi.org/10.1016/j.jalz.2011.05.2416>.
- Pakkenberg, B., & Gundersen, H. J. (1997). Neocortical neuron number in humans: Effect of sex and age. *The Journal of Comparative Neurology*, 384(2), 312–320.
- Pardini, M., Bergamino, M., Bommarito, G., Bonzano, L., Luigi Mancardi, G., & Roccatagliata, L. (2014). Structural correlates of subjective and objective memory performance in multiple sclerosis. *Hippocampus*, 24(4), 436–445. <https://doi.org/10.1002/hipo.22237>.
- Quenon, L., Dricot, L., Woodard, J. L., Hanseeuw, B., Gilis, N., Lhommel, R., & Ivanoiu, A. (2016). Prediction of free and cued

- selective reminding test performance using volumetric and amyloid-based biomarkers of Alzheimer's disease. *Journal of the International Neuropsychological Society*, 22(10), 991–1004. <https://doi.org/10.1017/S1355617716000813>.
- Song, S. K., Sun, S. W., Ju, W. K., Lin, S. J., Cross, A. H., & Neufeld, A. H. (2003). Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage*, 20(3), 1714–1722.
- Steenwijk, M. D., Daams, M., Pouwels, P. J., L., J. B., Tewarie, P. K., Geurts, J. J., . . . Vrenken, H. (2015). Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. *Human Brain Mapping*, 36(5), 1796–1807. <https://doi.org/10.1002/hbm.22738>.
- Storsve, A. B., Fjell, A. M., Yendiki, A., & Walhovd, K. B. (2016). Longitudinal changes in white matter tract integrity across the adult lifespan and its relation to cortical thinning. *PLoS One*, 11(6), e0156770. <https://doi.org/10.1371/journal.pone.0156770>.
- Stricker, N. H., Salat, D. H., Foley, J. M., Zink, T. A., Kellison, I. L., McFarland, C. P., . . . Leritz, E. C. (2013). Decreased white matter integrity in neuropsychologically defined mild cognitive impairment is independent of cortical thinning. *Journal of the International Neuropsychological Society*, 19(8), 925–937. <https://doi.org/10.1017/S1355617713000660>.
- Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zollei, L., Augustinack, J., . . . Fischl, B. (2011). Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Frontiers in Neuroinformatics*, 5, 23. <https://doi.org/10.3389/fninf.2011.00023>.
- Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., & Alexander, D. C. (2012). NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4), 1000–1016. <https://doi.org/10.1016/j.neuroimage.2012.03.072>.
- Zhuang, L., Sachdev, P. S., Trollor, J. N., Reppermund, S., Kochan, N. A., Brodaty, H., & Wen, W. (2013). Microstructural white matter changes, not hippocampal atrophy, detect early amnesic mild cognitive impairment. *PLoS One*, 8(3), e58887. <https://doi.org/10.1371/journal.pone.0058887>.