

Document downloaded from:

<http://hdl.handle.net/10251/156854>

This paper must be cited as:

Xie, L.; Wisse, LEM.; Pluta, J.; De Flores, R.; Piskin, V.; Manjón Herrera, JV.; Wang, H.... (2019). Automated segmentation of medial temporal lobe subregions on in vivo T1-weighted MRI in early stages of Alzheimer's disease. *Human Brain Mapping*. 40(12):3431-3451. <https://doi.org/10.1002/hbm.24607>



The final publication is available at

<https://doi.org/10.1002/hbm.24607>

Copyright John Wiley & Sons

#### Additional Information

This is the peer reviewed version of the following article: Xie, L, Wisse, LEM, Pluta, J, et al. Automated segmentation of medial temporal lobe subregions on in vivo T1-weighted MRI in early stages of Alzheimer's disease. *Hum Brain Mapp*. 2019; 40: 3431-3451, which has been published in final form at <https://doi.org/10.1002/hbm.24607>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

# **Automated segmentation of medial temporal lobe subregions on in vivo T1-MRI in early stages of Alzheimer's disease**

Long Xie<sup>1,4</sup>, Laura Wisse<sup>1,4,+</sup>, John Pluta<sup>1,4</sup>, Robin de Flores<sup>2,3</sup>, Jose V. Manjón<sup>6</sup>,  
Hongzhi Wang<sup>7</sup>, Sandhitsu Das<sup>1,3,4</sup>, Song-Lin Ding<sup>5</sup>, David Wolk<sup>2,3</sup>, Paul  
Yushkevich<sup>1,4</sup>, for the Alzheimer's Disease Neuroimaging Initiative\*

<sup>1</sup> Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, Philadelphia, USA

<sup>2</sup> PennMemory Center, University of Pennsylvania, Philadelphia, USA

<sup>3</sup> Department of Neurology, University of Pennsylvania, Philadelphia, USA

<sup>4</sup> Department of Radiology, University of Pennsylvania, Philadelphia, USA

<sup>5</sup> Allen Institute for Brain Science, Seattle, USA

<sup>6</sup> Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universidad Politécnica de Valencia, Valencia, Spain

<sup>7</sup> IBM Almaden Research Center, San Jose, USA

+ L.X and L.W. contributed equally to this work

\* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

# 1. Introduction

The medial temporal lobe (MTL) is the target of several neurodegenerative pathologies, most notably of neurofibrillary tangle (NFT) pathology which is thought to first affect the transentorhinal cortex, before it spreads to the entorhinal cortex and cornu ammonis region (CA) 1 of the hippocampus (Braak and Braak, 1995, 1991). As NFT pathology is closely related to neuron and synapse loss (Bobinski et al., 1997; Braak and Braak, 1991; Fukutani et al., 1995), certain medial temporal lobe *subregions* may therefore show early and selective atrophy and serve as imaging biomarker in the early stages of Alzheimer's disease. In fact, a recent in vivo MRI study showed selective atrophy in Brodmann area 35 (BA35), a region that approximates the transentorhinal region, in individuals with preclinical Alzheimer's disease (AD) compared to controls (Wolk et al., 2017). These subregions are also of interest because they are thought to subserve different cognitive functions, such as recollection and familiarity (Wolk and Dickerson, 2011; Yonelinas et al., 2007), and are part of two dissociable MTL networks, where the anterior hippocampus, entorhinal cortex (ERC) and perirhinal cortex (PRC) are part of the anterior MTL network and the posterior hippocampus and parahippocampal cortex (PHC) are part of the posterior MTL (Ranganath and Ritchey, 2012). These networks are also thought to be affected in the early stages of AD (Wolk et al., 2017).

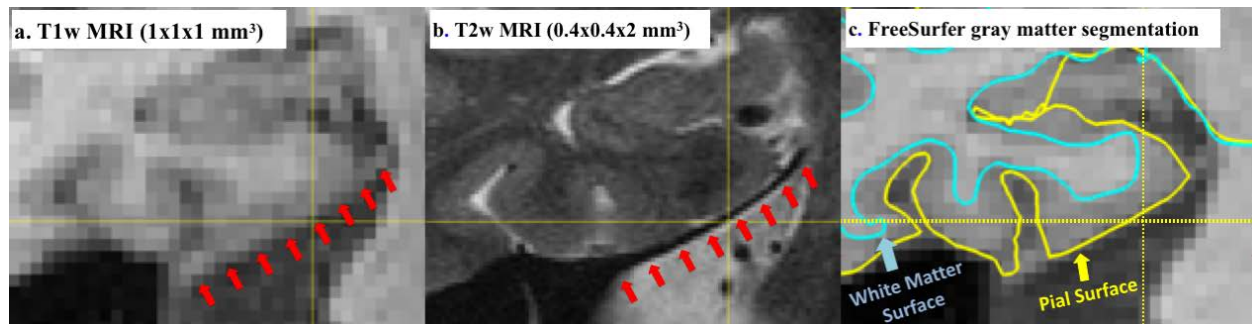
Fine-grained measurement of subregions of the MTL has therefore received increasing attention in the recent years, with many studies utilizing high resolution T2-weighted (T2w) MRI images, often with  $\sim 0.4 \times 0.4$  mm<sup>2</sup> in-plane resolution (Yushkevich et al., 2015a). The advantage of these images is that they allow for improved visualization of MTL structures, for example by the visualization of the stratum radiatum lacunosum moleculare (SRLM) which is an important border between certain subfields of the hippocampus, but also by the clear visualization of the dura mater, which is part of the meninges. The advantage of the clear visualization of the dura in these T2-weighted MRI images is that it allows for accurate segmentation of important adjacent MTL subregions, in contrast to T1-weighted (T1w) MRI images in which the dura has similar intensity as gray matter (Xie et al., 2016). Even though there are advantages of these T2w MRI images over T1w MRI images, there are large datasets of T1w MRI scans available and analyzing these datasets would allow for more power to test

hypotheses of interest. Additionally, T1-weighted images often have higher resolution in the through-plane direction which helps in better resolving the folding and branching of sulci, important for the segmentation of these MTL cortical regions.

There are methods available for the parcellation of MTL subregions for T1w MRI. Several manual approaches exist (Kivisaari et al., 2013; Nikolai V. Malykhin, 2008) which can be used to obtain granular measures of the MTL cortex on T1w MRI. Moreover, an advantage of these manual approaches is that they often take anatomical variability of the collateral sulcus into account, which greatly affects the location of the borders between MTL cortices (Ding and Van Hoesen, 2010). However, manual segmentation does not seem feasible for larger datasets like the Alzheimer's disease neuroimaging initiative (ADNI), which includes hundreds of MRI scans. There are several automated methods available, like FreeSurfer (Fischl, 2012). However, these methods generally provide coarser labels of the MTL and often provide no account of how the method dealt with anatomical variability of the collateral sulcus, which greatly affects the location of the borders between MTL cortices. It should be noted that FreeSurfer does provide a method to subdivide the hippocampus into different subfields (Iglesias et al., 2015). However, we have previously argued that standard resolution T1w MRI scans do not provide sufficient resolution for the visualization of the inner structure of the hippocampus and the parcellation of the hippocampal subfields (de Flores et al., 2014; Laura E. M. Wisse, Geert Jan Biessels, 2014). Another issue for T1w MRI scans, as mentioned above, relates to the visualization of the dura mater, part of the meninges. In the MTL, a large proportion of the ERC and parts of the PRC are located directly adjacent to the dura and as a result often appear merged with parts of the dura in T1w MRI (red arrows in Figure 1). To the best of our knowledge, none of the automatic analysis pipelines for MTL cortices using T1w MRI have addressed this confound, and the dura is often segmented as part of the gray matter by the state-of-the-art image processing algorithms (Figure 1c). This likely leads to an error in the quantification of ERC and PRC, which potentially confounds the findings of research studies. Speculating, if there is no space between the dura and cortex, as is often the case in healthy individuals, the dura may be mistakenly included in the cortex segmentation, whereas if there is space between the dura and cortex, for example in

patient groups with severe atrophy, the dura may be correctly excluded from the cortex label. This would lead to a systematic bias in the estimation of group differences.

Figure 1. The dura (red arrows) has similar intensity as gray matter in T1w MRI (a) but can be easily separated from the cortex in T2w MRI (b). It is often segmented as part of the cortex by state-of-the-art algorithms (c), e.g. FreeSurfer.



To address the issues relating to the dura and the coarseness of the MTL regions in previous automated methods, we developed a novel pipeline for the segmentation of ERC, PRC and dura on T1w images in our prior work (Xie et al., 2016), using an established multi-atlas segmentation framework (Yushkevich et al., 2015b) together with a super-resolution technique (Manjón et al., 2010). The atlas for this T1 pipeline was created using the T2 atlas set that was previously published by Yushkevich et al. (2015b), which includes measures of the ERC and subdivisions of the PRC, that is BA35 and BA36, based on a segmentation protocol that takes anatomical variability of the collateral sulcus into account, and was developed in consultation with a neuroanatomist (SLD). The segmentations of this atlas set were transformed into the T1 space after co-registration with the T2w MRI of the same subject and edited. Additionally, these segmentations were extended with a dura label informed by the T2w MRI. Evaluation of this pipeline and comparison to other methods, including FreeSurfer, indicated that a large portion of the dura received the correct label in our pipeline but not in other methods. That is, a large portion of the dura was included in gray matter in these other methods. Cross-validation experiments showed promising segmentation accuracy [Dice similarity coefficient or DSC (Dice, 1945) > 0.671] for cortical regions. Moreover, the clinical utility of the pipeline to other methods was evaluated by

examining the statistical power in discriminating controls from amnesic mild cognitive impairment (aMCI) patients, and indicated, qualitatively, that the largest area under the curve (AUC) for our pipeline was for BA35.

In the current paper we have extended this work in several ways. We have extended our previous label set to include the parahippocampal cortex (PHC) and the hippocampus, including a subdivision of anterior and posterior hippocampus and provide thickness values in addition to volumes for the MTL cortices. We have improved the registration between the T1w and T2w MRI scans allowing for a closer alignment which required less editing of the transformed segmentations in the T1-space. We completed experiments to evaluate the performance of our pipeline. We performed cross-validation experiments of the new atlas set against the manual segmentation and compared our pipeline with FreeSurfer version 6.0 (Fischl, 2012) to evaluate how the different methods label dura in T1w MRI. And we evaluated the performance of our pipeline in scans from Alzheimer's disease neuroimaging initiative (ADNI) phases GO and 2 by comparing MTL subregional volumes and thickness in amyloid negative controls with individuals with preclinical AD, prodromal AD and AD dementia. Finally, the atlas and software developed in this paper are made publicly available at <https://sites.google.com/view/ashs-dox/home>. In addition, we have also provided an easy-to-use cloud-based service of the proposed pipeline. Detailed tutorial on the cloud based serviced are available <https://sites.google.com/view/ashs-dox/cloud-ashs/overview> and briefly summarized in Supplementary Material A.

## **2. Methods and materials**

### **2.1 Participants**

#### **2.1.1 PMC atlas set**

The atlas set used in this study consists of 15 cognitively normal older controls (NC) and 14 aMCI patients. These participants were recruited from the Penn Memory Center / Alzheimer's Disease Center (PMC/ADC) at the University of Pennsylvania. Diagnosis of aMCI was made following the criteria established by Peterson and others (Peterson, 2004; Petersen et al., 2009; Winblad et al., 2004). Informed consent was provided by all

subjects. This study was approved by the Institutional Review Board of the University of Pennsylvania. This is the same atlas set that was used by Yushkevich et al. (2015b) and Xie et al (2017) to develop the atlas set using both T1w MRI and high-resolution T2w MRI. To avoid confusion, the atlas set developed in this study will be referred to as the **PMC-T1 atlas** and the one used in Yushkevich et al. (2015) and Xie et al (2017) will be referred to as the **PMC-T2 atlas**. Demographic and the mini-mental state examination (MMSE) data for the aMCI and NC groups are shown in Table 1.

Table 1 Characteristics of the PMC dataset

	Normal Control	aMCI
<b>N</b>	15	14
<b>Age (yrs)</b>	66.3 (9.5)	71.9 (6.2)
<b>Gender (M/F)</b>	7 / 8	6 / 8
<b>Education (yrs)</b>	15.6 (2.6)	16.9 (2.8)
<b>MMSE</b>	29.5 (1.0)	26.9 (1.7) ***

Note: All statistics are in comparison to cognitive normal control subjects. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . Independent two-sample t-tests (age, education, MMSE) and contingency  $\chi^2$  test (gender) were used. Standard deviation in parentheses. Abbreviations: MMSE = mini-mental state examination.

### 2.1.2 Dataset from the Alzheimer’s disease neuroimaging initiative (ADNI)

Part of the data used in the preparation of this article were obtained from the ADNI database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer’s disease. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

Cognitively normal controls and amyloid-beta ( $A\beta$ ) positive patients that have T1w MRI scan available from the ADNI GO and ADNI 2 were included in this study. The amyloid

status of each participant is determined by thresholding the global Florbetapir SUVR (derived from Florbeta PET in a standard way, with signal in the whole cerebellum as the reference region) using a threshold of SUVR  $\geq 1.11$  (Landau et al., 2012). In total, 663 participants were included and grouped into A $\beta$  negative (A $\beta$ -) controls, preclinical AD (A $\beta$  positive controls), early prodromal AD [A $\beta$  positive early mild cognitive impairment (MCI)], late prodromal AD (A $\beta$  positive LMCI) and dementia patients (A $\beta$  positive AD). Table 2 summarizes the characteristics of these subjects.

Table 2. Characteristics of the ADNI dataset

	A $\beta$ - Control	Preclinical AD	Early Prodromal AD	Late Prodromal AD	Dementia
<b>N</b>	190	95	142	109	127
<b>Age (yrs)</b>	72.3 (6.0)	74.8 (5.9) ***	73.6 (6.9)	72.3 (6.8)	74.4 (8.2) *
<b>Gender (M/F)</b>	100 / 90	31 / 64 **	81 / 61	57 / 52	68 / 59
<b>Education (yrs)</b>	16.9 (2.4)	16.1 (2.7) *	15.6 (2.8) ***	16.6 (2.6)	15.6 (2.7) ***
<b>MMSE</b>	29.0 (1.3)	29.0 (1.1)	28.0 (1.7) ***	27.2 (1.9) ***	23.0 (2.8) ***

Note: All statistics are in comparison to amyloid- $\beta$  negative (A $\beta$ -) control subjects. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . Independent two-sample t-tests (age, education, MMSE) and contingency  $\chi^2$  test (gender) were used. Standard deviation in parentheses. Abbreviations: AD = Alzheimer's disease; MMSE = mini-mental state examination.

## 2.2 Neuroimaging data acquisition

### 2.2.1 Imaging protocol atlas set

The MRI scans of the atlas set were acquired on a 3T Siemens Trio MRI scanner (Erlangen, Germany) at the University of Pennsylvania using an 8-channel array coil. The imaging protocols include (1) a whole brain T1w (Magnetization Prepared Rapid Acquisition Gradient Echo, MPRAGE) MRI scan; (2) a T2w (Turbo Spin Echo, TSE) MRI scan with partial brain coverage and oblique coronal slice positioned orthogonally to the main axis of the hippocampus (De Vita et al., 2003; Thomas et al., 2004). The parameters of the T2w MRI are: TR/TE=5310/68 ms, 18.3 ms echo spacing, 15 echo train length, 150° flip angle, 0% phase oversampling, 0.4x0.4 mm<sup>2</sup> in-plane resolution,



2.0 mm slice thickness with 0.6 mm gap, 30 interleaved slices, 7:12 min acquisition time. For the T1w MRI, they are: TR/TE/ TI=1600/3.87/950 ms, 15° flip angle, 1.0×1.0×1.0 mm<sup>3</sup> isotropic resolution, 5:13 min acquisition time.

### **2.2.2 ADNI Imaging protocol**

The protocols that used to acquire the T1w MRI scans from ADNI are variable because of the multi-center nature of the ADNI study. The T1w ADNI MRI protocol was previously described in Jack et al. (2008) and Leow et al. (2006). For Florbetapir PET, images were acquired in a 20 minutes PET brain scan session (4 frames of 5 minutes duration). 10 mCi tracer was injected followed by a 50 minutes uptake phase before imaging. We used the pre-processed images (“AV45 Coreg, Avg, Std Img and Vox Siz, Uniform Resolution”) available from the ADNI database.

## **2.3 Manual segmentation of the MTL substructures in T1w MRI**

All edits and segmentations were performed in ITK-SNAP (Yushkevich et al., 2006).

### **2.3.1 Segmentation of parahippocampal cortices and dura**

Manual segmentations of the parahippocampal cortices from the PMC-T2 atlas set from Yushkevich et al. (2015) and Xie et al. (2017) were propagated to the space of the aligned T1w MRI, followed by manual edits and addition of the dura label. Details are described below. Figure 2 shows examples that illustrates the workflow.

Alignment between T2w MRI and the T1w MRI of the same subjects were performed following the steps below:

- (1) Rigidly aligned T1w MRI to T2w MRI using the SyN ANTs (<http://stnava.github.io/ANTs/>) with mutual information as the similarity metric.
- (2) Up-sampled T1w MRI to 0.5x0.5x1.0 mm<sup>3</sup> by applying a patch-based super-resolution (SR) technique (Manjón et al., 2010) for the purpose of bringing the

resolution of the T1w MRI closer to that of the T2w MRI. Also, the SR upsampling increases the contrast between the dura and gray matter in T1w MRI so that the boundary between them can be better visualized.

- (3) Resampled T2w MRI and the corresponding manual segmentation to  $0.4 \times 0.4 \times 1.3$  mm<sup>3</sup> using linear and nearest neighbor interpolation respectively. The purpose of this step is to make the voxel size of the T2w MRI and SR T1w MRI similar.
- (4) The upsampled T2w MRI was cropped based on its manual segmentation with a margin of 10 voxels in all directions. This is done separately for left and right hemispheres.
- (5) For each hemisphere, affine registration was performed between the SR T1w MRI and the cropped upsampled T2w MRI to further align the two modalities, initialized with the rigid transformation between the whole brain T1w MRI and T2w MRI.
- (6) Transformed and resampled SR T1w MRI to the cropped up-sampled T2w MRI space (referred to as registered SR T1w MRI), in which manual segmentations of the parahippocampal cortices and the hippocampus were performed.

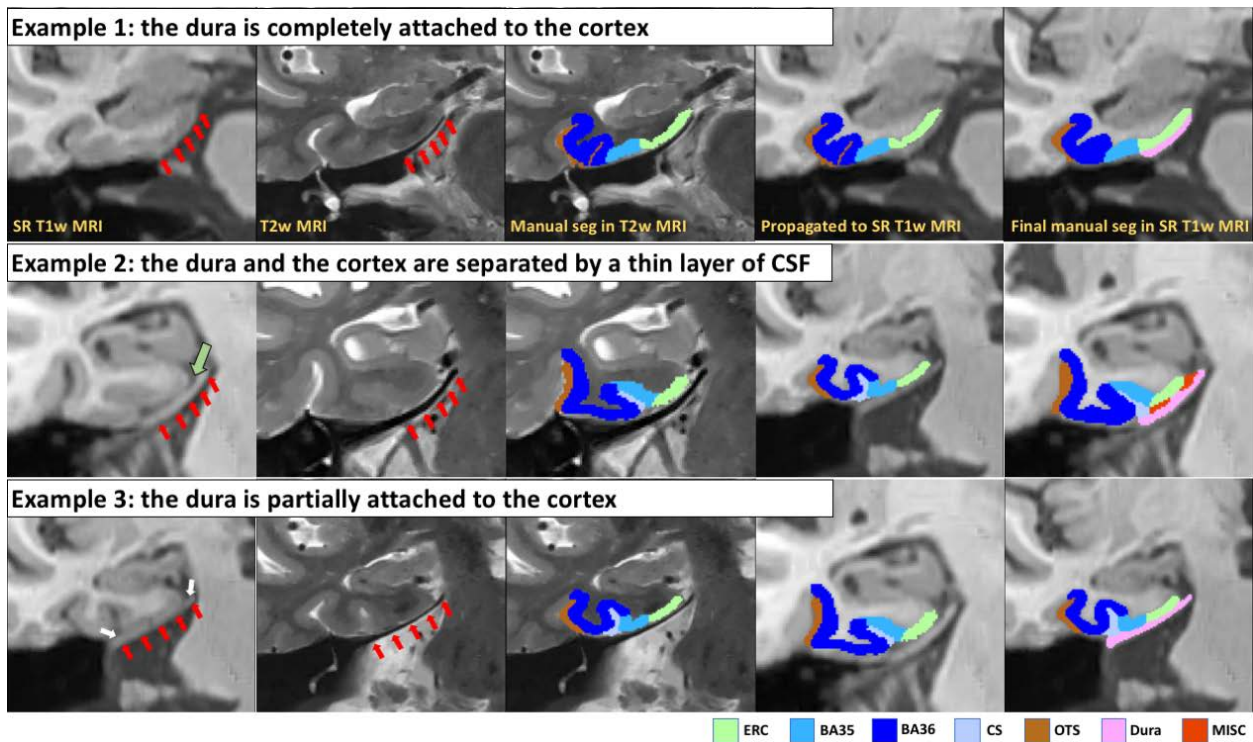
After local registration of the T2w and T1w MRI described above, the MTL region of both modalities are well aligned as shown in the first two columns in Figure 2. Labels of the parahippocampal cortices, including cortical labels (ERC, BA35, BA36, PHC) and sulcus labels [collateral sulcus (CS) and occipitotemporal sulcus (OTS)], were copied over the registered SR T1w MRI (the third and the fourth columns in Figure 2). Because of small errors in registration due to highly anisotropic voxel size of T2w MRI, affine inter-modality registration and the up-sampling of both modalities, the labels were checked and manually edited to correctly match the border with the white matter, cerebrospinal fluid (CSF) and dura largely following the procedure described in Xie et al. (2016); for these edits both the T1w and T2w MRI were used (the fifth column in Figure 2). Note that only the outer borders with surrounding regions were adjusted, not the borders between the different parahippocampal cortices. Only the last slice of the ERC was adjusted, as a transition slice, extending half the length of one slice anterior (note that these two slices translate to one slice on the T2w MRI). This is similar to the

procedure in (Berron et al., 2017). Because of small registration errors and perhaps slight differences in the visualization of the hippocampus on T1w and T2w MRI, the anterior and posterior borders of the parahippocampal cortices did not follow the protocol as described in (Yushkevich et al., 2015b) for some subjects. To reach consistency between subjects but to minimize changes to the original segmentations, an optimal anterior and posterior border was decided upon based on the full atlas set. The ERC, BA35 and BA36 extend one 1.3 mm slice anterior to the first slice of the hippocampus (was one 2.6 mm slice in the original protocol), ERC extends two 1.3 mm slices posterior to the most posterior slice of the uncus (same as in original protocol) and BA35/BA36 extends four 1.3 mm slices (same as in original protocol). The most anterior slice of the PHC is one slice posterior to the end of BA35 and BA36 (same as in original protocol) and the most posterior slice is fourth most posterior 1.3 mm slice of the hippocampus (was second most posterior 2.6 mm slice in original protocol). All subjects were visually checked and the segmentations were adjusted to match these boundary rules. Any given label needed to be extended at most two slices, where borders were matched to adjacent slices. In none of the cases the anatomy changes dramatically from one slice to the next, making these adjustments feasible.

Importantly, along the full length of the parahippocampal cortices, a label for the dura mater was assigned to the voxels inferior to the corrected parahippocampal cortices labels that have gray appearance in the registered SR T1w MRI and dark appearance in the resampled T2w MRI. Of note, the segmentation of the dura was informed by the registered T2w MRI, from which the boundary between dura and the cortex can be identified. This is especially crucial for situation when dura is completely attached to the cortex and cannot be visualized in T1w MRI (example 1 in Figure 2). In some cases, a thin layer of cerebrospinal fluid (CSF) between the dura and gray matter is visible in SR T1w MRI (green arrow in example 2 in Figure 2), i.e. a layer of voxels that have much darker intensity between the dura and gray matter in SR T1w MRI, which helps guide the dura segmentation. The CSF voxels were assigned a miscellaneous label. Moreover, in some cases, this layer of CSF is not visible, however the dura is not completely attached to the cortex either (example 3 in Figure 2). The portion of the dura

near the brain stem and inferior to the CS that is not adjacent to the cortex (white arrows in the Figure 2) provides clues for automatic and manual segmentation of the dura. The anterior and posterior extents of the dura are limited to the slices with parahippocampal cortices labels (ERC, BA35, BA36 and PHC).

Figure 2. Examples showing the procedure of manual segmenting medial temporal lobe (MTL) cortices in T1w MRI using manual labels in the space of the T2w MRI. Red and white arrows indicate the dura mater. The green arrow points to a thin layer of CSF between dura and the cortex that exists in some subjects. The white arrows show places that the dura is not completely attached to the cortex, which are important landmarks for manual and automatic segmentation.



### 2.3.2 Segmentation of the hippocampus

The European Alzheimer's Disease Consortium and Alzheimer's Disease Neuroimaging Initiative (ADNI) Harmonized Protocol (HarP) (Boccardi et al., 2015a; Frisoni et al., 2015) is a well-validated harmonized protocol for hippocampus segmentation in T1w MRI. To be consistent with the HarP protocol, we chose a subset (15 controls and 15

MCI patients) of the public available HarP training set from ADNI described in Boccardi et al. (2015b) as the training set to automatically segment the registered SR T1w MRI (obtained in Section 2.3.1) of the 29 cases in the PMC-T1 atlas set using the Automated Segmentation of Hippocampal Subfields (ASHS) package/software (see Section 2.4 for a brief description). These automatic segmentations were used to initialize the manual segmentation of the hippocampus. The characteristics of the HarP training set and the detail of the automatic segmentation pipeline is described in Supplementary material B.

All segmentations were visually checked and edited where necessary in three planes, following the HARP protocol. Two adjustments were made to the HARP protocol. First, the medial border of the hippocampus was extended to be continuous with the parahippocampal cortices generated in Section 2.3.1. The medial border therefore followed the protocol from Wisse et al. (2012). In the most posterior slices, the hippocampal medial border was located at the most medial point of the cortex not including the calcarine sulcus. The hippocampus did not always reach the parahippocampal gyrus in these most posterior slices. Second, this medial border was executed until the slice where the hippocampus was embedded in the splenium, to improve the transition to the decreasing size of the hippocampus on consecutive slices. Moreover, the hippocampus was split in an anterior and posterior region, where the border was defined by the most posterior slice of the uncus, which was included in the anterior hippocampus (Malykhin et al., 2007).

## **2.4 Automatic segmentation using ASHS**

Automatic segmentation of hippocampal subfields software/package (ASHS, <https://sites.google.com/view/ashs-dox/home>) was used to construct the atlas and perform automatic segmentation of a new subject.

### **2.4.1 Construction of ASHS-T1 atlas using the ASHS training pipeline**

The original T1w MRI, whole-brain SR T1w MRI together with the bilateral manual segmentations in the space of the SR T1w MRI are fed into the ASHS training pipeline

to generate an atlas (**ASHS-T1** atlas). The ASHS training pipeline is described in detail in Yushkevich et al. (2015b) and summarized briefly in the following steps: (1) an unbiased whole brain population template is built using the T1w MRI of all the subjects; (2) the region of interest (ROI) of each hemisphere is identified by averaging the corresponding manual segmentations that are warped to the space of the template; (3) each SR T1w MRI and the corresponding segmentation is warped to the space of the template and cropped around the ROI; (4) pairwise registrations between all the subjects are performed between the warped and cropped scans; (5) label fusion is performed for each atlas in its native space using the rest of the atlases as candidates; (6) an adaboost classifier is trained to learn the systematic error between the automatic segmentation and the manual segmentations. The ASHS-T1 atlas is publicly available at [https://www.nitrc.org/frs/shownotes.php?release\\_id=3851](https://www.nitrc.org/frs/shownotes.php?release_id=3851).

#### **2.4.2 Application of ASHS-T1 atlas to new images**

Once the ASHS-T1 atlas is trained, we can use the ASHS segmentation pipeline to automatically segment the T1w MRI scan of a new subject. Different from the pipeline described in Yushkevich et al. (2015b), the proposed pipeline only take the T1w MRI scan as input and does not require the T2w MRI scan. In brief, it involves the following steps:

- (1) The T1w MRI of the target subject is first up-sampled to  $0.5 \times 0.5 \times 1 \text{ mm}^3$  using the SR technique (Manjón et al., 2010).
- (2) The ROI around the left and right MTL are identified in the target SR T1w image by registering to a whole-brain template generated in the training pipeline.
- (3) For each target ROI, the corresponding ROIs in the atlas set are registered to it using ANTs with normalized cross-correlation metric (Avants et al., 2008).
- (4) Atlas labels are then warped to the target ROI and combined using the joint label fusion algorithm (Wang and Yushkevich, 2013).
- (5) The process is repeated in a bootstrapping fashion, where the initial segmentation of the target structures is used to initialize affine alignment between the atlas and target ROIs. This bootstrapping results in fewer failed atlas-to-target registrations and better

overall segmentation accuracy. The automatic segmentation generated from this step is referred to as the “**Heur**” output.

(6) Two adaboost classifiers, which were trained on shape features (the output referred to as the “**NoGray**”) or shape and gray-scale intensity features (“**UseGray**”) to correct for systematic errors generated in the multi-atlas label fusion step, are applied to further improve the automatic segmentation. Using the classifier outputs, i.e. NoGray and UseGray, is only recommended if the target T1w MRI scan is acquired with similar protocol as the atlas set.

Final automatic segmentations are generated in the target SR T1w MRI space. For the atlas set of 29 subjects, the automatic segmentation in the space of the SR T1w MRI was generated in a leave-one-out manner using the remaining 28 subjects as atlases. The segmentation accuracy of the “UseGray” output is reported in Table 2 and those of all the three outputs (“Heur”, “NoGray” and “UseGray”) are also computed and reported in Supplementary Table 1.

When segmenting the baseline scans of the ADNI cohort, the whole 29-subject ASHS-T1 atlas set was used. To account for the difference in imaging protocol, the segmentation output from the bootstrapping step was used [the “Heur” result generated in step (5)]. The segmentation accuracy in terms of DSC between automatic and manual segmentations of “Heur” is comparable but slightly lower (1.5% maximum DSC) than “UseGray” shown in Supplementary Table 1. Volumetric and thickness (see Section 2.5.3) measurements of bilateral anterior/posterior hippocampus, ERC, BA35, BA36 and PHC were extracted for each subject.

### **2.4.3 Quality control**

The quality of all the automatic segmentations generated by ASHS-T1 were visually checked. The pipeline successfully labels the baseline T1w MRI scans of all the 663 ADNI subjects while small errors happened in small portion of the subjects. In detail, we observed under-segmentation in the lateral border of the hippocampus in 28 (7 A $\beta$ -control, 2 preclinical AD, 8 early prodromal AD, 7 late prodromal AD and 4 dementia)

out of 663 subjects. In 16 [1 A $\beta$ - control, 1 early prodromal AD, 6 late prodromal AD (1 overlap) and 8 dementia (1 overlap)] out of 663 subjects, oversegmentation of parahippocampal cortices were identified. This is unavoidable partially due to the lack of contrast between cortex and dura. Examples of the common segmentation errors are shown Supplementary Figure 1.

## **2.5 Additional Image Processing**

### **2.5.1 Intracranial Volume (ICV)**

Intracranial space was segmented from the T1w MRI of each ADNI subject using an in-house ICV segmentation software using ASHS with a training set of 27 T1w MRI scans (15 controls and 12 aMCI) and the corresponding manual intracranial space segmentations. The unique aspect of the training set is that the manual labels were generated with the guidance of the coregistered computer tomographic (CT) scans of the same subjects. Since the boundary between skull and the soft tissue is clear in CT scans, we were able to obtain an accurate manual segmentation of the intracranial space. Supplementary C describes the detail of ICV automatic segmentation pipeline.

### **2.5.2 Cross validation experiment in the atlas set in the space of the T2w MRI (ASHS-T2)**

To compare the segmentation accuracy of the parahippocampal cortices of the proposed pipeline that only utilizes T1w MRI to that using both T1w and T2w MRI (Yushkevich et al. HBM, 2015), leave-one-out cross validation was also performed using the PMC-T2 atlas (comparisons were performed between the automatic and manual segmentations in the space of the T2w MRI). The same experiment has been done in Yushkevich et al. (2015b). However, since we have updated the ASHS software [ASHS version 2.0.0 rather than 1.0.0 ([https://www.nitrc.org/frs/?group\\_id=370](https://www.nitrc.org/frs/?group_id=370))] and the atlas manual segmentation [the PHC and OTS labels were added as described in Xie et al. (2017)], the results are slightly different from that in Yushkevich et al. (2015b). Note



that we did not perform this analysis for the hippocampus, as the segmentation protocol for the T1w and T2w hippocampus were different.

### **2.5.3 Thickness measures of the parahippocampal cortices extracted from the ASHS-T1 automatic segmentation**

For parahippocampal cortices, thickness measures may be more appropriate compared to the volume ones because they are less sensitive to uncertainty in boundary estimation between cortical regions. A multi-template thickness analysis pipeline (Xie et al., 2017, 2014) was applied to the parahippocampal cortices labels (ERC, BA35, BA36 and PHC) to extract thickness. Since large anatomical variability, i.e. different branching and folding patterns of the cortex, exists at the parahippocampal cortices, traditional single-template-based approaches may not generate accurate thickness measures. The thickness pipeline takes anatomical variability into account by fitting corresponding variant-specific template to the target segmentation, which has been shown to generate more accurate thickness measurement (Xie et al., 2017, 2014).

### **2.5.4 Volume and thickness measures of hippocampus, ERC and PRC using FreeSurfer**

In order to compare the volume and thickness measurements extracted from the proposed pipeline to that from an established paradigm for T1w MRI, FreeSurfer version 6.0 (Fischl, 2012) was applied to the T1w MRI scans of both the 29 subjects in the atlas set and the ADNI dataset. Volume measurements of the hippocampus, ERC and PRC were extracted from the “aseg.stats”, “lh.BA\_exvivo.thresh.stats” and “rh.BA\_exvivo.thresh.stats” files.

## **2.6 Statistical analysis**

All statistical analyses in this paper are two-tailed with significance levels of  $p = 0.05$  unless stated otherwise. Bilateral measurements of each subregion were averaged.

### **2.6.1 Analysis of demographic and MMSE data**

To test the differences of demographic and MMSE between diagnosis groups, i.e. aMCI-NC of the PMC atlas set and each patient-control pair of the ADNI dataset, independent two-sample t-tests and contingency  $\chi^2$  test were performed for continuous (age, education, MMSE) and categorical (gender) variables respectively.

### **2.6.2 Evaluate consistency between automatic and manual segmentations**

To evaluate the automated segmentations generated by ASHS-T1 and ASHS-T2 average DSC (Dice, 1945) between the automatic segmentations and the corresponding manual segmentations among the PMC atlas sets were computed. In addition, we also computed the intra-class correlation (ICC) between volume measurements of the MTL subregions extracted from the automatic segmentations in the PMC-T1 atlas set and those obtained using the edited manual segmentations in T1w MRI space. To compare the ICC for the ASHS-T1 pipeline with that of the ASHS-T2 pipeline, similar analysis was performed for the parahippocampal cortices labels (ERC, BA35, BA36 and PHC) for the PMC-T2 atlas set as well. ICC is computed using the ICC (2,1) method in Shrout & Fleiss (1979).

### **2.6.3 Group analysis between patients and A $\beta$ - controls in ADNI**

To evaluate the clinical utility, the four patient groups were compared to A $\beta$ - controls separately. For each volume measure, a general linear model (GLM) with group membership as the factor of interest, age and ICV as covariates, was fitted to obtain the t-statistics for the NC-patient contrast. Bonferroni corrected significance level ( $p < 0.05/10$ ) is used to determine significant effects. For thickness measures, similar analysis was performed but only age was used as covariate and the Bonferroni correction significant level was set to  $p < 0.05/6$ .

## **3. Evaluation experiments and results**

We first evaluated the consistency of the automatic segmentation of ASHS-T1 with the manual ones in the space of the T1w MRI and compare the performance of ASHS-T1 with that of ASHS-T2 (Section 3.1). Then, we investigated the extent to which the established analysis methods for T1w MRI, i.e. FreeSurfer, mislabel the dura mater and the cortex (Section 3.2). Lastly, to demonstrate clinical utility of the proposed pipeline, we compared the volume and thickness measures extracted using the proposed pipeline between patients and controls using a large dataset from the ADNI and compared this with FreeSurfer (Section 3.3).

### **3.1 Evaluate consistency with manual segmentation**

Primary validation of consistency with manual segmentation was performed on the set of 29 subjects from the PMC atlas, for whom T1w MRI, T2w MRI, both automatic and manual segmentations of the SR T1w MRI and T2w MRI are available.

The DSC results are summarized in Table 2. High DSCs of anterior (0.92), posterior (0.90) and whole (0.93) hippocampus segmentation were observed. The good accuracy in segmenting dura (0.75) and the MTL cortex labels (ERC: 0.76, BA35: 0.71, BA36: 0.79, PHC: 0.80) indicates that the proposed pipeline can reliably segment the dura and the subregions of the MTL cortex. The slightly lower DSC of BA35 is not surprising given that it is a small structure and high anatomical variability exist in this region. No significant differences in segmentation accuracy, tested by two-sample t-test, were found between aMCI and NC for all labels. In addition, the DSCs of the MTL cortex labels of ASHS-T1 are comparable to that of the ASHS-T2, which has been validated to be comparable to inter-rater reliability of manual raters (Yushkevich et al., 2015b). The DSC of the proposed pipeline in segmenting ERC (DSC = 0.76) is slightly lower than that in T2w MRI (DSC = 0.79), which could be due to the limited ability to resolve gray matter boundaries because of the lower resolution and the confound of dura in T1w MRI.

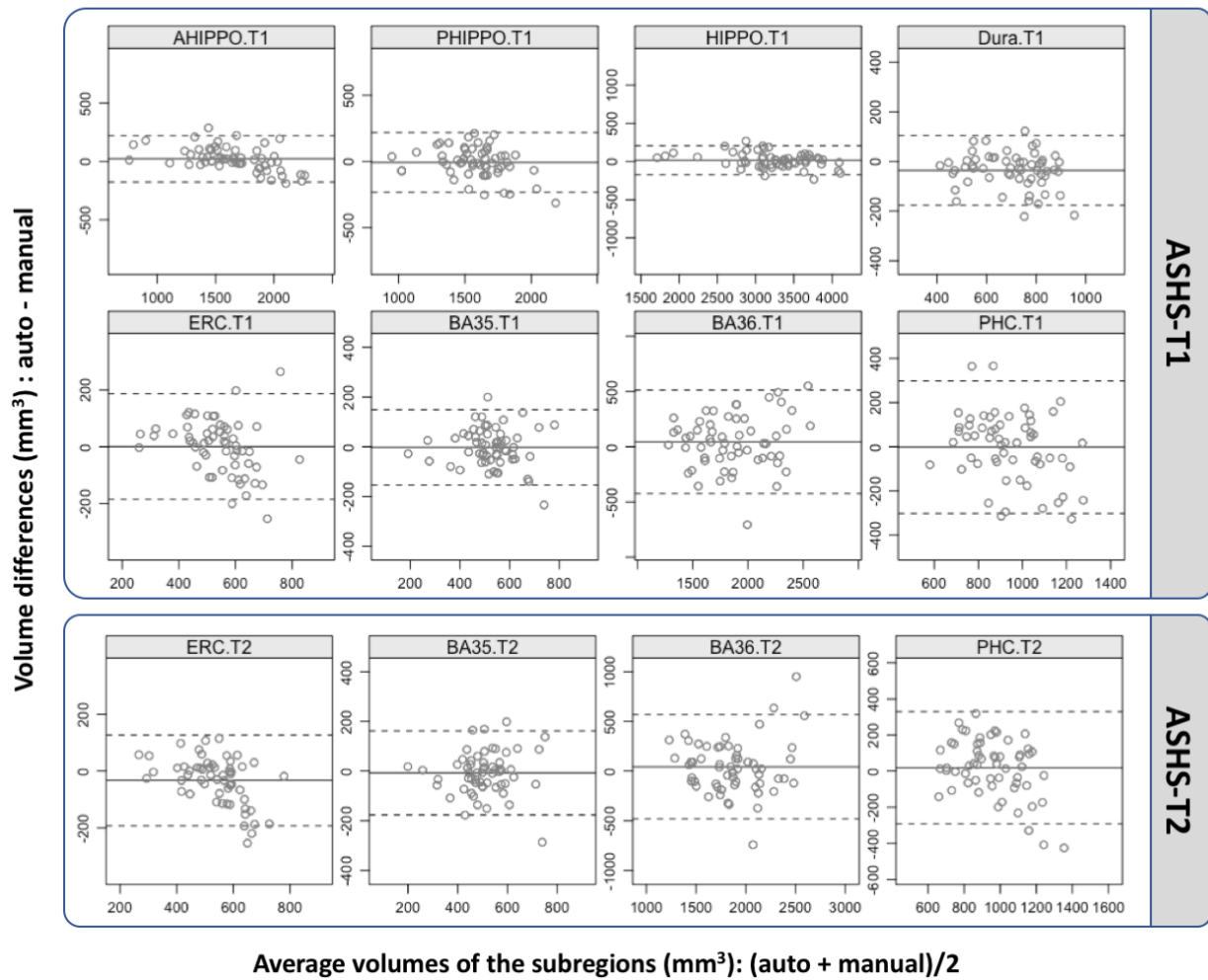
**Table 2.** Segmentation accuracy [measured by Dice similarity coefficient (DSC)] and intraclass correlation (ICC) relative to manual segmentations using leave-one-out cross validation. The atlas consists of 14 amnesic mild cognitive impairment patients (aMCI) and 15 cognitively normal controls (NC). DSCs of each substructure in both hemispheres are averaged. Mean and Standard deviation (parentheses) are reported in the table.

Substructure	T1w MRI		T2w MRI	
	ICC	DSC	ICC	DSC
	All (N=29)	All (N=29)	All (N=29)	All (n = 29)
Anterior Hippocampus	0.95	0.92 (0.02)	\	\
Posterior Hippocampus	0.89	0.90 (0.02)	\	\
<i>Whole Hippocampus</i> <sup>a</sup>	0.98	0.93 (0.01)	\	\
Entorhinal Cortex (ERC)	0.69	0.76 (0.03)	0.71	0.79 (0.03)
Brodmann Area 35 (BA35)	0.77	0.71 (0.06)	0.71	0.71 (0.06)
Brodmann Area 36 (BA36)	0.76	0.79 (0.03)	0.72	0.79 (0.04)
Parahippocampal Cortex (PHC)	0.64	0.80 (0.03)	0.64	0.79 (0.04)
Dura Mater	0.85	0.75 (0.05)	\	\

Note: <sup>a</sup>: A DSCs for the compound labels (in italics) are measured using the merged label of corresponding sub-labels (Whole hippocampus: anterior and posterior hippocampus in T1 weighted MRI).

From the ICC results, as reported in Table 2, ASHS-T1 demonstrates high consistency in segmenting anterior/posterior hippocampus (0.95 and 0.89), BA35 (0.77) and BA36 (0.76). The ICC for ERC (0.69) and PHC (0.64) were slightly lower. The ICC values did not show notable differences between the ASHS-T1 and T2 pipeline. According to the Bland-Altman plots, shown in Figure 3, there exists a small bias in ERC segmentation, i.e. the pipeline tends to undersegment larger ERC volumes and oversegment smaller ones. This would likely lead to a slight underestimation of group differences for ERC. No bias is observed for the other subregions.

**Figure 3.** Comparison of subregion volume measured by ASHS-T1 and manual segmentation in the space of T1w MRI (top panel) MRI using Bland-Altman plots. To compare with ASHS-T2 (bottom panel), the comparisons between subregion volume generated in the space of T2w MRI and the corresponding manual segmentation in the PMC-T2 atlas set are also shown (.T2). HIPPO.T1 is a compound label by merging AHIPPO.T1 and PHIPPO.T1.



### 3.2 Dura mislabeling as cortex

In this section, we performed experiments to test the two hypotheses that were introduced in Section 1, i.e. (1) the MTL cortex is commonly over-segmented by

FreeSurfer because of the mislabeling of the adjacent dura mater; (2) the degree of dura mislabeling as cortex may be different between patients and controls.

To test the first hypothesis, among subjects in the PMC-T1 atlas set, we first resampled the FreeSurfer whole brain segmentations to the space of the SR T1w MRI and then computed the average percentage of voxels labeled as dura in the manual segmentations that were mislabeled as gray matter or other by the proposed pipeline and FreeSurfer. The results, shown in Table 3, support the notion that large proportion of dura (62.4%) is segmented as gray matter by FreeSurfer. We note that FreeSurfer does not have a specific label for the dura and therefore has to label the dura voxels as something else; including them in the gray matter introduces error to cortical thickness computations. On the other hand, the majority (71.9%) of dura voxels are correctly labeled by the proposed pipeline and only 6.5% of them are labeled as gray matter and the amount of dura mislabeling as cortex is not significantly different between aMCI and NC ( $6.8 \pm 3.1\%$  vs.  $6.2 \pm 4.2\%$ ,  $p > 0.1$ , revealed by two-sample t-test).

**Table 3.** Comparisons of different analysis methods in labeling the dura mater in the PMC atlas set.

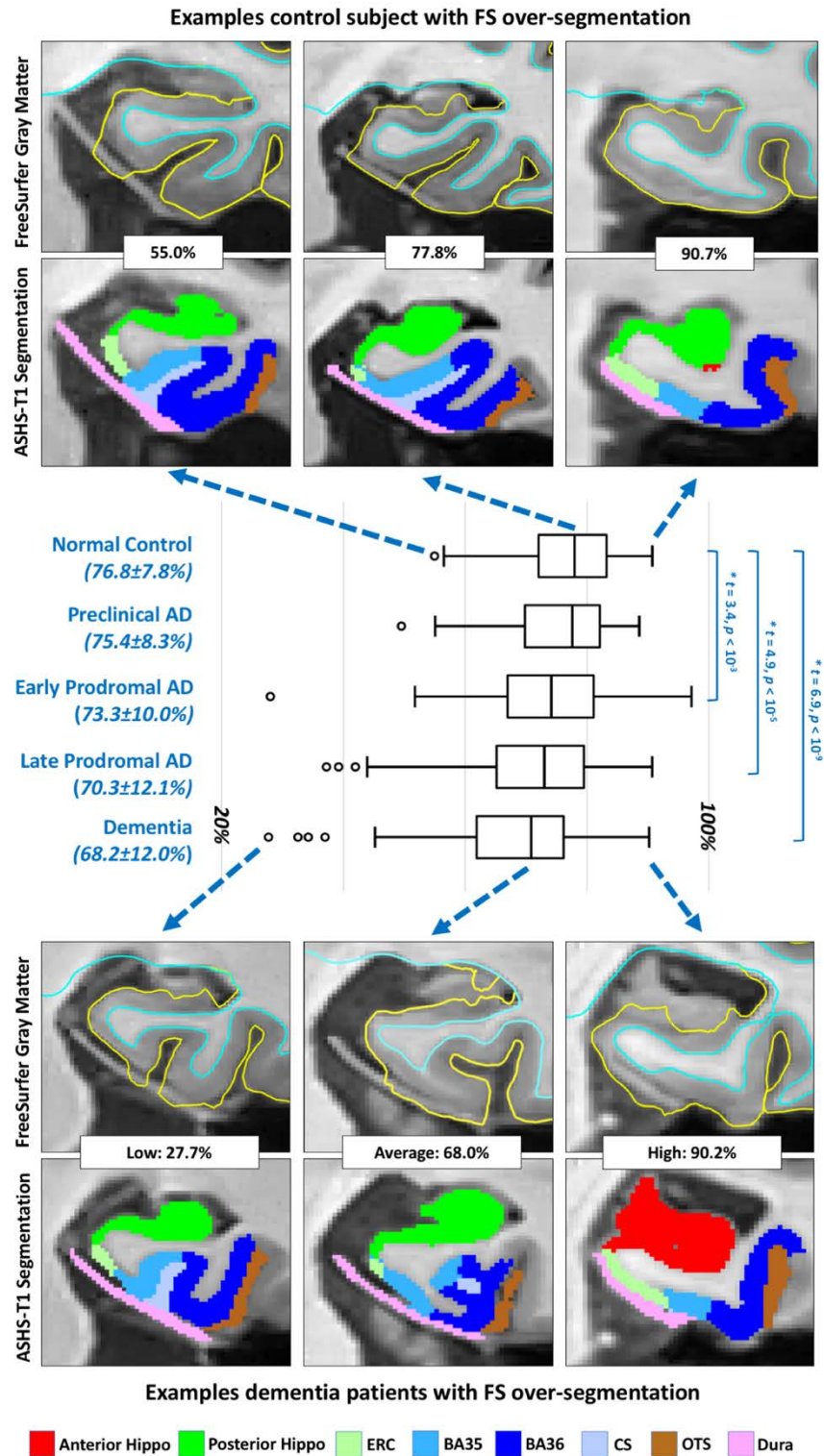
Method	% of dura voxels in manual segmentation labeled as		
	Dura	Gray matter	other
ASHS-T1	$71.9 \pm 6.4$	$6.5 \pm 3.7$	$21.6 \pm 5.9$
FreeSurfer 6.0	N/A	$62.4 \pm 10.5$	$37.6 \pm 10.5$

The second hypothesis can be tested using the ADNI dataset with controls and patients at different stages of AD. Since manual segmentation of the parahippocampal cortices and dura is not available in the ADNI dataset, the degree of dura mislabeling as gray matter by FreeSurfer is computed using the automatic ASHS-T1 segmentation, i.e. the average percentage of voxels labeled as dura by the ASHS-T1 that are labeled as gray matter. We believe this is a suitable measure because of the following evidence: (1) In the PMC-T1 atlas set, we computed the degree of dura mislabeling as gray matter by FreeSurfer relative to the dura label in the automatic segmentations generated by

ASHS-T1 and relative to the dura label from the manual segmentations. These measurements were highly correlated (Pearson correlation  $r = 0.946$ ,  $p = 9.3 \times 10^{-15}$ ), shown in Supplementary Figure 2. (2) In the PMC-T1 atlas set, no significant differences were observed between aMCI and controls in segmentation accuracy of dura (DSC reported in Table 2, 0.74 vs. 0.76) or for mislabeling of dura as cortex (6.8% vs. 6.2%) using the automatic dura segmentations generated by ASHS-T1. Therefore, it is unlikely that it will introduce bias between patients and controls. (3) All the segmentations of the ADNI subjects generated by the T1 pipeline used in this analysis were visually checked and only segmentations that have high-quality MTL cortex segmentation were used in this analysis and thus the bias induced by segmentation errors is limited.

Figure 4 summarizes the percentage of dura voxels segmented as gray matter by FreeSurfer in controls and the four patient groups. The amount of dura mislabeling as cortex decreases with increasing disease severity, probably due to the more distinct separation between the MTL cortex and the dura (Figure 4). The proportion of mislabeling is significantly different between controls and patients at early prodromal AD, late prodromal AD and dementia stages revealed by two-sample t-tests. Since manual segmentation of the ADNI dataset is not available, it is not feasible to evaluate the amount of dura mislabeling as cortex of the proposed method. However, since we did not see large difference of dura mislabeling as cortex between aMCI and controls in the PMC-T1 atlas set (0.6%), it seems unlikely that the observed large differences of FreeSurfer dura mislabeling between groups (3.5%, 6.5% and 8.6% between patients at early, late prodromal AD, dementia and controls respectively) are mainly due to imperfect automatic segmentation of ASHS-T1.

**Figure 4.** Percentage of dura voxels labeled as gray matter by FreeSurfer in all diagnosis groups. Examples of low, average and high over-segmentation in controls and dementia patients are shown on top and bottom respectively.



### 3.3 MTL atrophy in early stages of AD in ADNI



We compared the volume and thickness measures extracted using the proposed pipeline between patients at different stages of AD and A $\beta$ - controls in ADNI and performed a comparison with FreeSurfer. To be noted, out-of-box (without quality control) performance of the proposed pipeline is reported in this section. Excluding subjects due to poor segmentation quality for the proposed pipeline did not significantly alter the results as shown in Supplementary Table 2.

As shown in Table 4 and Table 5, no significant differences were observed in the preclinical stage. However, there was a trend level ( $F = 2.8$ ,  $p = 0.093$ , if age is not included as a covariate, i.e. independent two-sample t-test was performed, the statistical result was  $t = 2.5$ ,  $p = 0.014$ ) difference in BA35 thickness between preclinical AD patients and controls. We observed Bonferroni-corrected significant group effects at the early prodromal (MCI) stage in posterior hippocampus volume ( $F = 16.8$ ,  $p = 5.2e-5$ ), BA35 thickness ( $F = 10.4$ ,  $p = 1.4e-3$ ), ERC volume ( $F = 9.5$ ,  $p = 2.2e-3$ ), and BA35 volume ( $F = 8.2$ ,  $p = 4.5e-3$ ). Volume and thickness of all the subregions were significantly smaller in patient groups at the late prodromal stage and the differences are bigger in dementia.

Overall, the results generated by the FreeSurfer were similar to that of the proposed pipeline, i.e. (1) none of the measures from FreeSurfer showed significant differences in the preclinical AD stage; (2) Bonferroni-corrected significant effects were observed in the early prodromal AD stage [hippocampus volume ( $F = 23.0$ ,  $p = 2.0e-6$ ) and PRC volume ( $F = 9.8$ ,  $p = 1.9e-3$ )]; (3) all measurements showed significant differences in late prodromal AD and dementia stages. FreeSurfer ERC and PRC thickness were consistently about 50% thicker than the corresponding measurements (ERC and BA35) by ASHS-T1, which is probably due to the mislabeling of dura as cortex. In addition, the mislabeling of dura seems introduce instability of FreeSurfer measurements of the MTL cortex in A $\beta$ - controls and early stages of AD (preclinical and early prodromal AD). For example, FreeSurfer ERC volume decreased from A $\beta$ - controls ( $802.5 \text{ mm}^3$ ) to preclinical AD ( $768.2 \text{ mm}^3$ ) but became slightly higher in early prodromal AD ( $804.1 \text{ mm}^3$ ). Also, FreeSurfer volume and thickness measurements were more variable

(higher standard deviation of) than the corresponding measurements generated by the proposal pipeline.

Table 4. Statistical analysis results using volumetric measurements, adjusted for age and intracranial volume, in discriminating patient groups from normal controls in ADNI. Measurements that survived Bonferroni correction ( $p < 0.05/10$ ) are highlighted in bold.

Region	A $\beta$ - Control (n = 190)	Preclinical AD (n = 95)	Early Prodromal AD (n = 142)	Late Prodromal AD (n = 109)	Dementia (n = 127)
<b>ASHS-T1 Volume Measurements (mm<sup>3</sup>)</b>					
<b>Anterior Hippocampus</b>	1724.5	1711.6	1666.7	<b>1533.9</b>	<b>1440.0</b>
SD	225.6	207.0	237.9	<b>255.5</b>	<b>200.9</b>
% Diff		-0.7	-3.3	<b>-11.1</b>	<b>-16.5</b>
F stats		<2.5	4.9	<b>44.8</b>	<b>129.8</b>
p value		>0.1	0.027	<b>1.1e-10</b>	<b>2.2e-25</b>
<b>Posterior Hippocampus</b>	1646.5	1647.0	<b>1571.4</b>	<b>1405.0</b>	<b>1344.1</b>
SD	160.1	152.7	<b>172.0</b>	<b>202.7</b>	<b>167.0</b>
% Diff		0.0	<b>-4.6</b>	<b>-14.7</b>	<b>-18.4</b>
F stats		<2.5	<b>16.8</b>	<b>129.3</b>	<b>261.0</b>
p value		>0.1	<b>5.2e-5</b>	<b>4.4e-25</b>	<b>4.0e-43</b>
<b>Hippocampus</b>	3371.0	3358.6	<b>3238.2</b>	<b>2938.9</b>	<b>2784.0</b>
SD	309.9	298.7	<b>351.5</b>	<b>411.8</b>	<b>320.8</b>
% Diff		-0.4	<b>-3.9</b>	<b>-12.8</b>	<b>-17.4</b>
F stats		<2.5	<b>13.1</b>	<b>105.3</b>	<b>262.1</b>
p value		>0.1	<b>3.4e-4</b>	<b>2.5e-21</b>	<b>3.0e-43</b>
<b>ERC</b>	573.9	562.5	<b>548.0</b>	<b>499.6</b>	<b>451.4</b>
SD	73.8	69.8	<b>80.5</b>	<b>91.8</b>	<b>84.6</b>
% Diff		-5.2	<b>-6.1</b>	<b>-12.8</b>	<b>-23.7</b>
F stats		<2.5	<b>9.5</b>	<b>58.5</b>	<b>192.5</b>
p value		>0.1	<b>2.2e-3</b>	<b>2.9e-13</b>	<b>1.9e-34</b>
<b>BA35</b>	606.5	597.1	<b>579.6</b>	<b>539.8</b>	<b>481.3</b>
SD	82.2	80.3	<b>92.1</b>	<b>100.2</b>	<b>81.3</b>
% Diff		-1.6	<b>-4.4</b>	<b>-11.0</b>	<b>-20.6</b>
F stats		<2.5	<b>8.2</b>	<b>38.6</b>	<b>183.3</b>
p value		>0.1	<b>4.5e-3</b>	<b>1.8e-9</b>	<b>3.5e-33</b>
<b>BA36</b>	1881.0	1869.5	1814.6	<b>1718.6</b>	<b>1581.9</b>
SD	249.7	228.5	227.9	<b>264.3</b>	<b>230.3</b>
% Diff		-0.6	-3.5	<b>-8.6</b>	<b>-15.9</b>
F stats		<2.5	6.1	<b>27.9</b>	<b>115.4</b>
p value		>0.1	0.014	<b>2.5e-7</b>	<b>4.0e-23</b>
<b>PHC</b>	958.0	977.6	959.2	<b>909.7</b>	<b>869.4</b>
SD	117.9	129.0	138.3	<b>129.5</b>	<b>122.8</b>
% Diff		2.1	0.1	<b>-5.0</b>	<b>-9.3</b>
F stats		<2.5	<2.5	<b>10.9</b>	<b>43.5</b>
p value		>0.1	>0.1	<b>1.0e-3</b>	<b>1.8e-10</b>
<b>FreeSurfer Volume Measurements (mm<sup>3</sup>)</b>					

<b>Hippocampus</b>	3790.2	3732.6	<b>3606.2</b>	<b>3299.6</b>	<b>3146.8</b>
SD	321.9	364.1	<b>363.8</b>	<b>425.9</b>	<b>330.8</b>
% Diff		-1.5	<b>-4.9</b>	<b>-12.9</b>	<b>-17.0</b>
F stats		<2.5	<b>23.0</b>	<b>127.2</b>	<b>301.5</b>
p value		>0.1	<b>2.0e-6</b>	<b>9.2e-25</b>	<b>9.0e-48</b>
<b>ERC</b>	802.5	768.2	804.1	<b>699.5</b>	<b>662.1</b>
SD	197.1	203.1	212.2	<b>229.8</b>	<b>191.2</b>
% Diff		-4.3	0.2	<b>-12.8</b>	<b>-17.5</b>
F stats		<2.5	<2.5	<b>17.0</b>	<b>38.9</b>
p value		>0.1	>0.1	<b>4.9e-5</b>	<b>1.5e-9</b>
<b>PRC</b>	1067.5	1051.9	<b>1019.0</b>	<b>916.8</b>	<b>834.1</b>
SD	150.0	142.1	<b>147.9</b>	<b>193.4</b>	<b>161.6</b>
% Diff		-1.5	<b>-4.5</b>	<b>-14.1</b>	<b>-21.9</b>
F stats		<2.5	<b>9.8</b>	<b>56.9</b>	<b>171.6</b>
p value		>0.1	<b>1.9e-3</b>	<b>5.7e-13</b>	<b>1.5e-31</b>

Table 5. Statistical analysis results using thickness measurements, adjusted for age, in discriminating patient groups from normal controls in ADNI. Measurements that survived Bonferroni correction ( $p < 0.05/6$ ) are highlighted in bold.

Region	A $\beta$ - Control (n = 190)	Preclinical AD (n = 95)	Early Prodromal AD (n = 142)	Late Prodromal AD (n = 109)	Dementia (n = 127)
<b>ASHS-T1 Thickness Measurements (mm)</b>					
<b>ERC</b>	2.02	2.02	1.99	<b>1.93</b>	<b>1.78</b>
SD	0.16	0.17	0.18	<b>0.18</b>	<b>0.22</b>
% Diff		0.1	-1.2	<b>-4.3</b>	<b>-11.5</b>
F stats		<2.5	<2.5	<b>18.4</b>	<b>122.0</b>
p value		>0.1	>0.1	<b>2.4e-5</b>	<b>3.5e-24</b>
<b>BA35</b>	2.35	2.33	<b>2.29</b>	<b>2.20</b>	<b>2.06</b>
SD	0.16	0.18	<b>0.19</b>	<b>0.23</b>	<b>0.23</b>
% Diff		-1.1	<b>-2.6</b>	<b>-6.7</b>	<b>-12.5</b>
F stats		2.8	<b>10.4</b>	<b>49.1</b>	<b>184.4</b>
p value		0.093	<b>1.4e-3</b>	<b>1.6e-11</b>	<b>2.3e-33</b>
<b>BA36</b>	2.41	2.39	2.39	<b>2.29</b>	<b>2.21</b>
SD	0.23	0.23	0.21	<b>0.22</b>	<b>0.24</b>
% Diff		-0.7	-0.8	<b>-5.1</b>	<b>-8.5</b>
F stats		<2.5	<2.5	<b>20.3</b>	<b>57.8</b>
p value		>0.1	>0.1	<b>1.0e-5</b>	<b>3.4e-13</b>
<b>PHC</b>	2.15	2.17	2.14	<b>2.09</b>	<b>2.00</b>
SD	0.13	0.16	0.13	<b>0.15</b>	<b>0.16</b>
% Diff		0.7	-0.3	<b>-2.7</b>	<b>-7.2</b>
F stats		<2.5	<2.5	<b>13.3</b>	<b>94.2</b>
p value		>0.1	>0.1	<b>3.2e-4</b>	<b>1.2e-19</b>
<b>FreeSurfer Thickness Measurements (mm)</b>					
<b>ERC</b>	3.18	3.17	3.14	<b>2.90</b>	<b>2.74</b>
SD	0.30	0.33	0.34	<b>0.40</b>	<b>0.40</b>
% Diff		-0.2	-1.1	<b>-8.8</b>	<b>-13.6</b>
F stats		<2.5	<2.5	<b>46.3</b>	<b>122.6</b>
p value		>0.1	>0.1	<b>5.7e-11</b>	<b>2.9e-24</b>
<b>PRC</b>	3.44	3.47	3.36	<b>3.11</b>	<b>2.87</b>

SD	0.31	0.33	0.34	<b>0.41</b>	<b>0.37</b>
% Diff		1.0	-2.2	<b>-9.5</b>	<b>-16.4</b>
<i>F</i> stats		<2.5	5.2	<b>59.8</b>	<b>217.0</b>
<i>p</i> value		>0.1	0.023	<b>1.7e-13</b>	<b>1.1e-37</b>

## 4 Discussion

In this paper, we present an automatic segmentation pipeline for T1w MRI for measuring granular MTL subregions accounting for the confound of dura. Our pipeline showed accuracy of >0.76 DSC, except for BA35, and >0.76 ICC, except for ERC and PHC, for the automatic compared to the manual segmentation, which is comparable to the performance of our T2 pipeline (except for ERC for which the accuracy is slightly lower). The experiments using the PMC-T1 atlas showed that proposed pipeline can reliably separate dura from gray matter, only mislabeling 6.5% of the dura as gray matter, whereas the FreeSurfer segmentation mislabels 71.9% of dura as gray matter indeed leading to about 50% thicker cortex in ERC and PRC. In an analysis in the ADNI dataset, we showed that the degree of dura mislabeling decreases with increasing disease severity, indicating a bias where cortex in controls are oversegmented to a larger extent than in patients. This could potentially lead to an overestimation of group differences in later stages of the disease. Finally, in the ADNI dataset we demonstrated that our pipeline picks up in early MCI in the MTL, including in ERC and BA35, that match the known progression of NFT pathology, but also in the posterior hippocampus. Moreover, the volume and thickness loss become more severe and widespread with increasing disease stages.

This pipeline has several unique aspects and strengths. First, it provides granular measures of the MTL, including subdivision of the PRC and hippocampus, for T1w MRI. It could therefore be very useful in several large large-scale studies in older populations (e.g. ADNI) in the interrogation of, for example, AD or age-related effects on the MTL, the anterior and posterior MTL networks and memory processes. Moreover, the segmentation protocol for the MTL cortices was developed in collaboration with a

neuroanatomist and takes the anatomical variability into account which can largely affect borders between cortices in contrast to most previous methods for T1w MRI. The accuracy of our automated segmentations generated by our pipeline compared to the manual segmentations is good with a DSC  $>0.76$ , except for BA35, and ICC  $>0.76$ , except for ERC and PHC. The MTL cortices are at the lower end of these ICC and DSC values and thickness rather than volume measurements might therefore be a more powerful approach. Additionally, thickness measures are also less sensitive to border placement. As far as we know, this is the first validation of automated segmentation of MTL cortices against manual segmentations on T1w MRI and the only other study performing such a validation on T2w MRI was of this same pipeline. The T1 pipeline show similar accuracy for the MTL cortices as the T2w pipeline, with a slightly lower accuracy for the ERC. With regard to hippocampus, our pipeline performs comparable to state-of-the-art methods (Xie et al., 2018).

In addition, to the best of our knowledge, this is the first automated pipeline that explicitly accounts for the confound of dura in segmenting MTL subregions in T1w MRI. Experimental results indicate that the dura can be reliably separated from the gray matter (the DSC and ICC are 0.75 and 0.85 respectively), indicating that the portions of the dura that not merged with the cortex (write arrows in Figure 2) provide sufficient features for automatic segmentation when there is low or even no contrast between the dura and the cortex. The importance of accounting for this confound was shown in the analyses of dura labeling in the FreeSurfer pipeline which indicated that not taking the dura into account can lead to 1) mislabeling of dura as gray matter causing errors in volume or thickness estimations and 2) a bias where this mislabeling is larger in controls than patients. However, it should be noted that while our pipeline explicitly accounts for the dura, it still makes a small error where small portions of the dura are counted towards the gray matter. With the limited contrast differences between dura and gray matter, this can unfortunately not be avoided. Moreover, this slight mislabeling of dura may explain the slightly lower accuracy of the ERC as it is adjacent dura for a relatively larger extent than the other MTL cortices. With the lower resolution and limited contrast in T1w MRI, it is more difficult to resolve this boundary than in T2w MRI. For

that reason, MTL cortex segmentation on high resolution T2-weighted images is still preferred. Another limitation is that the most anterior portions of the ERC and PRC are not included in the current automated pipeline, which is especially of interest for diseases such as semantic variant Primary Progressive Aphasia which show a clear anterior-to-posterior gradient of atrophy in the MTL (Chan et al., 2001; Davies et al., 2009). We will include these regions in future work. Moreover, the anterior and posterior border of ERC and PRC are directly determined by the extent of the hippocampal head. This could potentially introduce an error where ERC and PRC volume changes along with hippocampal head volume changes, that is, if the hippocampal head extends for a lower number of slices, ERC and PRC will automatically do so as well. This will affect thickness measures in a lesser extent, but will mostly affect volume measures. Indeed, in later stages of the disease the percentage volume loss in ERC and PRC is larger than the percentage thickness loss, which may reflect this bias. This is not reflected in the p-values as volume measures are also noisier. Qualitatively a stronger association of anterior hippocampal volume with ERC and BA35 volume is found than with ERC and BA35 thickness, separately for both hemispheres (Left: ERC volume:  $\rho=0.74$ , thickness:  $\rho=0.57$ ; BA35 volume:  $\rho=0.68$ , thickness:  $\rho=0.57$  – Right: ERC volume:  $\rho=0.75$ , thickness:  $\rho=0.53$ , BA35 volume:  $\rho=0.61$ , thickness:  $\rho=0.49$ ).

In light of above described strengths and limitations we would like to give some recommendations for users of the proposed T1w pipeline. Careful assessment of the MRI scans and segmentations is important, especially in light of small errors in dura labeling. Additionally, we noticed some errors in a small number of ADNI subjects at the lateral aspect of the hippocampus. Because of the composition of our atlas set the most appropriate target population is older adults and MCI patients. In this study, we applied the atlas also to images of patients with early AD dementia and careful quality assessment indicated that the atlas also performed well in this population. This matches our recent findings that atlas compositions of only controls, only MCI patients or a mixture did not significantly affect segmentation accuracy in these different groups (Xie et al., 2018). However, care is warranted when this atlas and pipeline are applied to other populations including other ages and diseases or very different imaging protocols.

When this atlas is applied to images acquired at a different platform or with a different MRI protocol, it is recommended to use the “Heur” output (step 5 in Section 2.4.2).

To assess the clinical validity and utility of our pipeline, we applied it to the ADNI dataset and compared different stages of AD with amyloid- $\beta$  negative controls on MTL subregional volume and thickness. Compared to the amyloid- $\beta$  negative controls, we observed a trend difference in BA35 thickness in amyloid- $\beta$  positive controls (preclinical AD), a significant difference in ERC volume, BA35 volume and thickness and posterior and total hippocampal volume in early MCI and in all regions in late MCI and early AD. The observed earliest effect on BA35 is consistent with the earliest accumulation of NFT pathology in this region (Braak and Braak, 1995, 1991)(REF). A recent study in a different, only partially overlapping, subset of ADNI showed a similar, but significant, decrease in BA35 thickness in preclinical AD (Wolk et al., 2017) using T2w MRI. The difference in significance may be due to more reliable segmentation of the MTL cortex because of a better contrast and separation of dura in T2w MRI as compared to T1w MRI. In light of the recently published A/T/N model, in future work it will be interesting to further select cases that are also tau-positive and investigate whether these subjects show increased neurodegeneration in BA35.

The spreading of atrophy to adjacent ERC and hippocampus in early MCI also matches the known spreading of NFT pathology (Braak and Braak, 1995, 1991) and to other studies investigating MTL atrophy patterns in the early stages (Killiany et al., 2002; Krumm et al., 2016; Olsen et al., 2017; Stoub et al., 2010; Xu et al., 2000; Yushkevich et al., 2015b). The volume loss in posterior hippocampus, rather than anterior hippocampus, was surprising, given that pathology starts in BA35, part of the PRC, which is thought to be more strongly connected to the anterior hippocampus, at least in the primate MTL (Aggleton, 2012). One might therefore speculate that the anterior hippocampus could be earlier affected than the posterior hippocampus in AD. Only a few studies investigated atrophy in the anterior and posterior hippocampus in MCI, where one study reported specific atrophy in anterior regions (Martin et al., 2010), but another did not (Greene and Killiany, 2012). Moreover, a qualitative inspection of

studies using shape analysis of the hippocampus to investigate granular effects of MCI shows inconsistent findings not clearly pointing towards an anterior-to-posterior gradient of atrophy in MCI (Chételat et al., 2008; Liana G. Apostolova, 2012; Qiu et al., 2009). Additionally, tractography studies in primates indicate that the posterior hippocampus is more strongly connected with the parahippocampal cortex which is in turn connected via the cingulum bundle with regions such as the posterior cingulate cortex and precuneus (this has also been supported by fMRI studies (Aggleton, 2012; Mufson and Pandya, 1984; Poppenk et al., 2013)) which have been indicated recently to show the earliest amyloid pathology (Palmqvist et al., 2017). This amyloid pathology, which is likely already present for years by the time subjects reach the early MCI stage, may have indirectly affected posterior hippocampal integrity. Moreover, the posterior hippocampus is part of the posterior MTL network (Ranganath and Ritchey, 2012), which has been found to already show atrophy in early MCI (Das et al., 2016). Finally, the increasing severity and widespread atrophy of the MTL in late MCI and early AD dementia again matches known spreading of NFT pathology (Braak and Braak, 1995, 1991) and other in vivo MTL work (de Flores et al., 2015; Dickerson et al., 2001; Jauhiainen et al., 2009; Stoub et al., 2010).

In general, FreeSurfer performed fairly similar in characterizing the MTL atrophy pattern in the different AD stages by picking atrophy in PRC and hippocampus in early MCI and increasing atrophy, including ERC, at later stages. The most evident difference in the early stages is a lack of significant ERC volume or thickness loss in early MCI using FreeSurfer. In fact, when looking carefully at the ERC volume measures a fluctuation can be observed where ERC volume loss is observed in preclinical AD compared to controls but then an increase is observed in early MCI, where ERC volumes again match those in the control group. This may be due to mislabeling dura as ERC which may introduce additional noise. Given that ERC atrophy is expected to be subtle at this stage, and the bias with regard to the dura mislabeling was observed at later disease stages, the inclusion of dura in the ERC label may lead only to increased measurement error. Surprisingly, even though we observed a bias in FreeSurfer of decreasing mislabeling of dura, this did not lead to larger effect sizes for group differences between



late MCI or early AD compared to controls. Perhaps this effect is counteracted by some other features of the labels, e.g. the effect size may be weakened by the larger extent of ERC and PRC which may potentially not show equal neurodegeneration along the full length. An important note is that the ERC and PRC in FreeSurfer do not represent completely the same regions as the ERC and the combined BA35 and BA36 into PRC in our pipeline. ERC and PRC in FreeSurfer actually show a 44.7% and 40.8% overlap with BA35 in our pipeline, and in fact ERC and PRC in FreeSurfer have about 37% overlap with each other. Having a granular label of BA35 rather than including it in ERC or a larger PRC label is advantageous, especially in the earliest stages of AD where NFT pathology is only thought to affect the transentorhinal cortex, which approximates our BA35 label, and a small portion of the lateral ERC. We did observe BA35 thinning in preclinical AD compared to amyloid-negative controls with our pipeline, although only at a trend level, which could potentially due to the heterogeneity in disease severity of the preclinical group.

## 5 Conclusions

In conclusion, we present a reliable automated pipeline for segmenting granular measures of the MTL for T1w MRI, explicitly accounting for the confound of dura. Moreover, we demonstrated the clinical utility by showing atrophy of early Braak regions in early MCI which becomes more severe and widespread in later stages. These findings should be replicated in other cohorts. Interesting and important future directions are establishing change in MTL regions over time, as longitudinal atrophy is more closely linked to clinical status and is important for tracking disease progression or as potential marker in clinical trials and establishing the association with Tau-PET uptake to better understand the drivers of neurodegeneration. This pipeline could be particularly useful for investigating Tau-PET tracers that show high uptake in the dura. We hope that this publicly available atlas and software including a cloud-based service (<https://sites.google.com/view/ashs-dox/home> and <https://sites.google.com/view/ashs->

[dox/cloud-ashs/overview](#)) will serve the scientific community and enable the interrogation of the role of the MTL in aging, dementia and cognition.

## Acknowledgements

This work was supported by NIH (grant numbers R01-AG056014, R01-AG040271, P30-AG010124, R01-EB017255, AG055005) and the donors of Alzheimer's Disease Research, a program of the BrightFocus Foundation (L.E.M.W.).

## References

- Aggleton, J.P., 2012. Multiple anatomical systems embedded within the primate medial temporal lobe: Implications for hippocampal function. *Neurosci. Biobehav. Rev.* 36, 1579–1596. <https://doi.org/10.1016/J.NEUBIOREV.2011.09.005>
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. <https://doi.org/10.1016/j.media.2007.06.004>
- Berron, D., Vieweg, P., Hochkeppler, A., Pluta, J.B., Ding, S.-L., Maass, A., Luther, A., Xie, L., Das, S.R., Wolk, D.A., Wolbers, T., Yushkevich, P.A., Düzel, E., Wisse, L.E.M., 2017. A protocol for manual segmentation of medial temporal lobe subregions in 7 Tesla MRI. *NeuroImage Clin.* 15. <https://doi.org/10.1016/j.nicl.2017.05.022>
- Bobinski, M., Wegiel, J., Tarnawski, M., Bobinski, M., Reisberg, B., de Leon, M.J., Miller, D.C., Wisniewski, H.M., 1997. Relationships between regional neuronal loss and neurofibrillary changes in the hippocampal formation and duration and severity of Alzheimer disease. *J. Neuropathol. Exp. Neurol.* 56, 414–20.
- Boccardi, M., Bocchetta, M., Apostolova, L.G., Barnes, J., Bartzokis, G., Corbetta, G.,

- DeCarli, C., deToledo-Morrell, L., Firbank, M., Ganzola, R., Gerritsen, L., Henneman, W., Killiany, R.J., Malykhin, N., Pasqualetti, P., Pruessner, J.C., Redolfi, A., Robitaille, N., Soininen, H., Tolomeo, D., Wang, L., Watson, C., Wolf, H., Duvernoy, H., Duchesne, S., Jack, C.R., Frisoni, G.B., 2015a. Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. *Alzheimer's Dement.* 11, 126–138.  
<https://doi.org/10.1016/J.JALZ.2014.02.009>
- Boccardi, M., Bocchetta, M., Morency, F.C., Collins, D.L., Nishikawa, M., Ganzola, R., Grothe, M.J., Wolf, D., Redolfi, A., Pievani, M., Antelmi, L., Fellgiebel, A., Matsuda, H., Teipel, S., Duchesne, S., Jack, C.R., Frisoni, G.B., EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation and for the Alzheimer's Disease Neuroimaging Initiative, 2015b. Training labels for hippocampal segmentation based on the EADC-ADNI harmonized hippocampal protocol. *Alzheimer's Dement.* 11, 175–83.  
<https://doi.org/10.1016/j.jalz.2014.12.002>
- Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* 16, 271–278.
- Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–59.
- Chan, D., Fox, N.C., Scahill, R.I., Crum, W.R., Whitwell, J.L., Leschziner, G., Rossor, A.M., Stevens, J.M., Cipelotti, L., Rossor, M.N., 2001. Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann. Neurol.* 49, 433–442.  
<https://doi.org/10.1002/ana.92>
- Chételat, G., Fouquet, M., Kalpouzos, G., Denghien, I., De la Sayette, V., Viader, F., Mézenge, F., Landeau, B., Baron, J.C., Eustache, F., Desgranges, B., 2008. Three-dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. *Neuropsychologia* 46, 1721–1731.  
<https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2007.11.037>
- Das, S.R., Mancuso, L., Olson, I.R., Arnold, S.E., Wolk, D.A., 2016. Short-Term Memory Depends on Dissociable Medial Temporal Lobe Regions in Amnesic Mild

- Cognitive Impairment. *Cereb. Cortex* 26, 2006–2017.  
<https://doi.org/10.1093/cercor/bhv022>
- Davies, R.R., Halliday, G.M., Xuereb, J.H., Kril, J.J., Hodges, J.R., 2009. The neural basis of semantic memory: Evidence from semantic dementia. *Neurobiol. Aging* 30, 2043–2052. <https://doi.org/10.1016/J.NEUROBIOLAGING.2008.02.005>
- de Flores, R., La Joie, R., Chételat, G., 2015. Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. *Neuroscience* 309, 29–50.  
<https://doi.org/10.1016/j.neuroscience.2015.08.033>
- de Flores, R., La Joie, R., Landeau, B., Perrotin, A., Mézenge, F., de La Sayette, V., Eustache, F., Desgranges, B., Chételat, G., 2014. Effects of age and Alzheimer's disease on hippocampal subfields: Comparison between manual and freesurfer volumetry. *Hum. Brain Mapp.* 00. <https://doi.org/10.1002/hbm.22640>
- De Vita, E., Thomas, D.L., Roberts, S., Parkes, H.G., Turner, R., Kinchesh, P., Shmueli, K., Yousry, T.A., Ordidge, R.J., 2003. High resolution MRI of the brain at 4.7 Tesla using fast spin echo imaging. *Br. J. Radiol.* 76, 631–7.  
<https://doi.org/10.1259/bjr/69317841>
- Dice, L.R., 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology* 26, 297–302. <https://doi.org/10.2307/1932409>
- Dickerson, B.C., Goncharova, I., Sullivan, M.P., Forchetti, C., Wilson, R.S., Bennett, D.A., Beckett, L.A., deToledo-Morrell, L., 2001. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiol. Aging* 22, 747–754. [https://doi.org/10.1016/S0197-4580\(01\)00271-8](https://doi.org/10.1016/S0197-4580(01)00271-8)
- Ding, S.-L., Van Hoesen, G.W., 2010. Borders, extent, and topography of human perirhinal cortex as revealed using multiple modern neuroanatomical and pathological markers. *Hum. Brain Mapp.* 31, 1359–79.  
<https://doi.org/10.1002/hbm.20940>
- Fischl, B., 2012. FreeSurfer. *Neuroimage* 62, 774–81.  
<https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Frisoni, G.B., Jack, C.R., Bocchetta, M., Bauer, C., Frederiksen, K.S., Liu, Y., Preboske, G., Swihart, T., Blair, M., Cavado, E., Grothe, M.J., Lanfredi, M., Martinez, O., Nishikawa, M., Portegies, M., Stoub, T., Ward, C., Apostolova, L.G., Ganzola, R.,

Wolf, D., Barkhof, F., Bartzokis, G., DeCarli, C., Csernansky, J.G., deToledo-Morrell, L., Geerlings, M.I., Kaye, J., Killiany, R.J., Lehericy, S., Matsuda, H., O'Brien, J., Silbert, L.C., Scheltens, P., Soininen, H., Teipel, S., Waldemar, G., Fellgiebel, A., Barnes, J., Firbank, M., Gerritsen, L., Henneman, W., Malykhin, N., Pruessner, J.C., Wang, L., Watson, C., Wolf, H., deLeon, M., Pantel, J., Ferrari, C., Bosco, P., Pasqualetti, P., Duchesne, S., Duvernoy, H., Boccardi, M., Albert, M.S., Bennet, D., Camicioli, R., Collins, D.L., Dubois, B., Hampel, H., denHeijer, T., Hock, C., Jagust, W., Launer, L., Maller, J.J., Mueller, S., Sachdev, P., Simmons, A., Thompson, P.M., Visser, P.-J., Wahlund, L.-O., Weiner, M.W., Winblad, B., 2015. The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: Evidence of validity. *Alzheimer's Dement.* 11, 111–125. <https://doi.org/10.1016/J.JALZ.2014.05.1756>

Fukutani, Y., Kobayashi, K., Nakamura, I., Watanabe, K., Isaki, K., Cairns, N.J., 1995. Neurons, intracellular and extracellular neurofibrillary tangles in subdivisions of the hippocampal cortex in normal ageing and Alzheimer's disease. *Neurosci. Lett.* 200, 57–60. [https://doi.org/10.1016/0304-3940\(95\)12083-G](https://doi.org/10.1016/0304-3940(95)12083-G)

Greene, S.J., Killiany, R.J., 2012. Hippocampal Subregions are Differentially Affected in the Progression to Alzheimer's Disease. *Anat. Rec. Adv. Integr. Anat. Evol. Biol.* 295, 132–140. <https://doi.org/10.1002/ar.21493>

Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., Fischl, B., Van Leemput, K., 2015. A computational atlas of the hippocampal formation using ex vivo , ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *Neuroimage* 115, 117–137. <https://doi.org/10.1016/j.neuroimage.2015.04.042>

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., Weiner, M.W., Study, for the A., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI

- methods. *J. Magn. Reson. Imaging* 27, 685–91. <https://doi.org/10.1002/jmri.21049>
- Jauhiainen, A.M., Pihlajamäki, M., Tervo, S., Niskanen, E., Tanila, H., Hänninen, T., Vanninen, R.L., Soininen, H., 2009. Discriminating accuracy of medial temporal lobe volumetry and fMRI in mild cognitive impairment. *Hippocampus* 19, 166–175. <https://doi.org/10.1002/hipo.20494>
- Killiany, R.J., Hyman, B.T., Gomez-Isla, T., Moss, M.B., Kikinis, R., Jolesz, F., Tanzi, R., Jones, K., Albert, M.S., 2002. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology* 58, 1188–96.
- Kivisaari, S., Probst, A., Taylor, K., 2013. The perirhinal, entorhinal, and parahippocampal cortices and hippocampus: an overview of functional anatomy and protocol for their segmentation in MR images. *fMRI* 239–267.
- Krumm, S., Kivisaari, S.L., Probst, A., Monsch, A.U., Reinhardt, J., Ulmer, S., Stippich, C., Kressig, R.W., Taylor, K.I., 2016. Cortical thinning of parahippocampal subregions in very early Alzheimer’s disease. *Neurobiol. Aging* 38, 188–196. <https://doi.org/10.1016/J.NEUROBIOLAGING.2015.11.001>
- Landau, S.M., Mintun, M.A., Joshi, A.D., Koeppe, R.A., Petersen, R.C., Aisen, P.S., Weiner, M.W., Jagust, W.J., 2012. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann. Neurol.* 72, 578–586. <https://doi.org/10.1002/ana.23650>
- Laura E. M.Wisse, Geert Jan Biessels, M.I.G., 2014. A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer 39, 127–34. <https://doi.org/10.1503/jpn.130070>
- Leow, A.D., Klunder, A.D., Jack, C.R., Toga, A.W., Dale, A.M., Bernstein, M.A., Britson, P.J., Gunter, J.L., Ward, C.P., Whitwell, J.L., Borowski, B.J., Fleisher, A.S., Fox, N.C., Harvey, D., Kornak, J., Schuff, N., Studholme, C., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2006. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage* 31, 627–640. <https://doi.org/10.1016/j.neuroimage.2005.12.013>
- Liana G. Apostolova, A.E.G.S.B.K.S.H.Y.C.A.W.T.P.M.T., 2012. Hippocampal Atrophy and Ventricular Enlargement in Normal Aging, Mild Cognitive Impairment (mci), and Alzheimer Disease. *Alzheimer Dis. &* 26, 17–27.

<https://doi.org/10.1097/wad.0b013e3182163b62>

- Malykhin, N. V., Bouchard, T.P., Ogilvie, C.J., Coupland, N.J., Seres, P., Camicioli, R., 2007. Three-dimensional volumetric analysis and reconstruction of amygdala and hippocampal head, body and tail. *Psychiatry Res. Neuroimaging* 155, 155–165. <https://doi.org/10.1016/J.PSCYCHRESNS.2006.11.011>
- Manjón, J. V, Coupé, P., Buades, A., Collins, L.D., Robles, M., 2010. MRI superresolution using self-similarity and image priors. *Int. J. Biomed. Imaging* 2010, 425891. <https://doi.org/10.1155/2010/425891>
- Martin, S.B., Smith, C.D., Collins, H.R., Schmitt, F.A., Gold, B.T., 2010. Evidence that volume of anterior medial temporal lobe is reduced in seniors destined for mild cognitive impairment. *Neurobiol. Aging* 31, 1099–1106. <https://doi.org/10.1016/J.NEUROBIOLAGING.2008.08.010>
- Mufson, E.J., Pandya, D.N., 1984. Some observations on the course and composition of the cingulum bundle in the rhesus monkey. *J. Comp. Neurol.* 225, 31–43. <https://doi.org/10.1002/cne.902250105>
- Nikolai V. Malykhin, T.P.B.R.C.N.J.C., 2008. Aging hippocampus and amygdala. *Neuroreport* 19, 543–547. <https://doi.org/10.1097/wnr.0b013e3282f8b18c>
- Olsen, R.K., Yeung, L.-K., Noly-Gandon, A., D'Angelo, M.C., Kacollja, A., Smith, V.M., Ryan, J.D., Barense, M.D., 2017. Human anterolateral entorhinal cortex volumes are associated with cognitive decline in aging prior to clinical diagnosis. *Neurobiol. Aging* 57, 195–205. <https://doi.org/10.1016/J.NEUROBIOLAGING.2017.04.025>
- Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., Blennow, K., Landau, S., Jagust, W., Hansson, O., 2017. Earliest accumulation of  $\beta$ -amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat. Commun.* 8, 1214. <https://doi.org/10.1038/s41467-017-01150-x>
- Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–94. <https://doi.org/10.1111/j.1365-2796.2004.01388.x>
- Petersen, R.C., Roberts, R.O., Knopman, D.S., Boeve, B.F., Geda, Y.E., Ivnik, R.J., Smith, G.E., Jack, C.R., 2009. Mild cognitive impairment: ten years later. *Arch. Neurol.* 66, 1447–55. <https://doi.org/10.1001/archneurol.2009.266>
- Poppenk, J., Evensmoen, H.R., Moscovitch, M., Nadel, L., 2013. Long-axis

specialization of the human hippocampus. *Trends Cogn. Sci.* 17, 230–240.  
<https://doi.org/10.1016/j.tics.2013.03.005>

- Qiu, A., Fennema-Notestine, C., Dale, A.M., Miller, M.I., Alzheimer's Disease Neuroimaging Initiative, A.D.N., 2009. Regional shape abnormalities in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 45, 656–61.
- Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. *Nat. Rev. Neurosci.* 13, 713–26. <https://doi.org/10.1038/nrn3338>
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–8.
- Stoub, T.R., Rogalski, E.J., Leurgans, S., Bennett, D.A., deToledo-Morrell, L., 2010. Rate of entorhinal and hippocampal atrophy in incipient and mild AD: Relation to memory function. *Neurobiol. Aging* 31, 1089–1098.  
<https://doi.org/10.1016/J.NEUROBIOLAGING.2008.08.003>
- Thomas, D.L., De Vita, E., Roberts, S., Turner, R., Yousry, T.A., Ordidge, R.J., 2004. High-resolution fast spin echo imaging of the human brain at 4.7 T: implementation and sequence characteristics. *Magn. Reson. Med.* 51, 1254–64.  
<https://doi.org/10.1002/mrm.20106>
- Wang, H., Yushkevich, P. a, 2013. Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation. *Front. Neuroinform.* 7, 27.  
<https://doi.org/10.3389/fninf.2013.00027>
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., Petersen, R.C., 2004. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J. Intern. Med.* 256, 240–6.  
<https://doi.org/10.1111/j.1365-2796.2004.01380.x>
- Wisse, L.E.M., Gerritsen, L., Zwanenburg, J.J.M., Kuijf, H.J., Luijten, P.R., Biessels, G.J., Geerlings, M.I., 2012. Subfields of the hippocampal formation at 7T MRI: In vivo volumetric assessment. *Neuroimage* 61, 1043–1049.  
<https://doi.org/10.1016/j.neuroimage.2012.03.023>



- Wolk, D.A., Das, S.R., Mueller, S.G., Weiner, M.W., Yushkevich, P.A., Alzheimer's Disease Neuroimaging Initiative, 2017. Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's disease. *Neurobiol. Aging* 49, 204–213. <https://doi.org/10.1016/j.neurobiolaging.2016.09.011>
- Wolk, D.A., Dickerson, B.C., 2011. Fractionating verbal episodic memory in Alzheimer's disease. *Neuroimage* 54, 1530–1539. <https://doi.org/10.1016/J.NEUROIMAGE.2010.09.005>
- Xie, L., Pluta, J., Wang, H., Das, S.R., Mancuso, L., Kliot, D., Avants, B.B., Ding, S.-L., Wolk, D.A., Yushkevich, P.A., 2014. Automatic clustering and thickness measurement of anatomical variants of the human perirhinal cortex, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. pp. 81–88.
- Xie, L., Pluta, J.B., Das, S.R., Wisse, L.E.M., Wang, H., Mancuso, L., Kliot, D., Avants, B.B., Ding, S.-L., Manjón, J. V., Wolk, D.A., Yushkevich, P.A., 2017. Multi-template analysis of human perirhinal cortex in brain MRI: Explicitly accounting for anatomical variability. *Neuroimage* 144, 183–202. <https://doi.org/10.1016/j.neuroimage.2016.09.070>
- Xie, L., Shinohara, R.T., Ittyerah, R., Kuijff, H.J., Pluta, J.B., Blom, K., Kooistra, M., Reijmer, Y.D., Koek, H.L., Zwanenburg, J.J.M., Wang, H., Luijten, P.R., Geerlings, M.I., Das, S.R., Biessels, G.J., Wolk, D.A., Yushkevich, P.A., Wisse, L.E.M., 2018. Automated Multi-Atlas Segmentation of Hippocampal and Extrahippocampal Subregions in Alzheimer's Disease at 3T and 7T: What Atlas Composition Works Best? *J. Alzheimer's Dis.* 63, 217–225. <https://doi.org/10.3233/JAD-170932>
- Xie, L., Wisse, L.E.M., Das, S.R., Wang, H., Wolk, D.A., Manjón, J. V., Yushkevich, P.A., 2016. Accounting for the Confound of Meninges in Segmenting Entorhinal and Perirhinal Cortices in T1-Weighted MRI, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*. pp. 564–571. [https://doi.org/10.1007/978-3-319-46723-8\\_65](https://doi.org/10.1007/978-3-319-46723-8_65)
- Xu, Y., Jack, C.R., O'Brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., Petersen, R.C., 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. *Neurology* 54, 1760–7.
- Yonelinas, A.P., Widaman, K., Mungas, D., Reed, B., Weiner, M.W., Chui, H.C., 2007.

Memory in the aging brain: Doubly dissociating the contribution of the hippocampus and entorhinal cortex. *Hippocampus* 17, 1134–1140.

<https://doi.org/10.1002/hipo.20341>

Yushkevich, P.A., Amaral, R.S.C., Augustinack, J.C., Bender, A.R., Bernstein, J.D., Boccardi, M., Bocchetta, M., Burggren, A.C., Carr, V.A., Chakravarty, M.M., Chételat, G., Daugherty, A.M., Davachi, L., Ding, S.-L., Ekstrom, A., Geerlings, M.I., Hassan, A., Huang, Y., Iglesias, J.E., La Joie, R., Kerchner, G.A., LaRocque, K.F., Libby, L.A., Malykhin, N., Mueller, S.G., Olsen, R.K., Palombo, D.J., Parekh, M.B., Pluta, J.B., Preston, A.R., Pruessner, J.C., Ranganath, C., Raz, N., Schlichting, M.L., Schoemaker, D., Singh, S., Stark, C.E.L., Suthana, N., Tompariy, A., Turowski, M.M., Van Leemput, K., Wagner, A.D., Wang, L., Winterburn, J.L., Wisse, L.E.M., Yassa, M.A., Zeineh, M.M., 2015a. Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards a harmonized segmentation protocol. *Neuroimage* 111, 526–541. <https://doi.org/10.1016/j.neuroimage.2015.01.004>

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31, 1116–28. <https://doi.org/10.1016/j.neuroimage.2006.01.015>

Yushkevich, P.A., Pluta, J.B., Wang, H., Xie, L., Ding, S., Gertje, E.C., Mancuso, L., Klot, D., Das, S.R., Wolk, D.A., 2015b. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum. Brain Mapp.* 36, 258–287. <https://doi.org/10.1002/hbm.22627>

# Supplementary Material

**Supplementary Table 1.** Segmentation accuracy of the three outputs of ASHS-T1 relative to manual segmentations, measured by Dice similarity coefficient (DSC) using leave-one-out cross validation. DSCs of each substructure in both hemispheres are averaged. Mean and Standard deviation (parentheses) are reported in the table.

Substructure	Heur	NoGray	UseGray
Anterior Hippocampus	0.91 (0.02)	0.91 (0.02)	0.92 (0.02)
Posterior Hippocampus	0.89 (0.02)	0.89 (0.02)	0.90 (0.02)
<i>Whole Hippocampus<sup>a</sup></i>	0.92 (0.01)	0.92 (0.02)	0.93 (0.01)
Entorhinal Cortex (ERC)	0.76 (0.03)	0.76 (0.04)	0.76 (0.03)
Brodmann Area 35 (BA35)	0.70 (0.06)	0.71 (0.07)	0.71 (0.06)
Brodmann Area 36 (BA36)	0.78 (0.04)	0.79 (0.05)	0.79 (0.03)
Parahippocampal Cortex (PHC)	0.79 (0.03)	0.80 (0.04)	0.80 (0.03)
Dura Mater	0.74 (0.05)	0.75 (0.06)	0.75 (0.05)

<sup>a</sup> DSCs for the compound labels (in italics) are measured using the merged label of corresponding sub-labels (Whole hippocampus: anterior and posterior hippocampus).

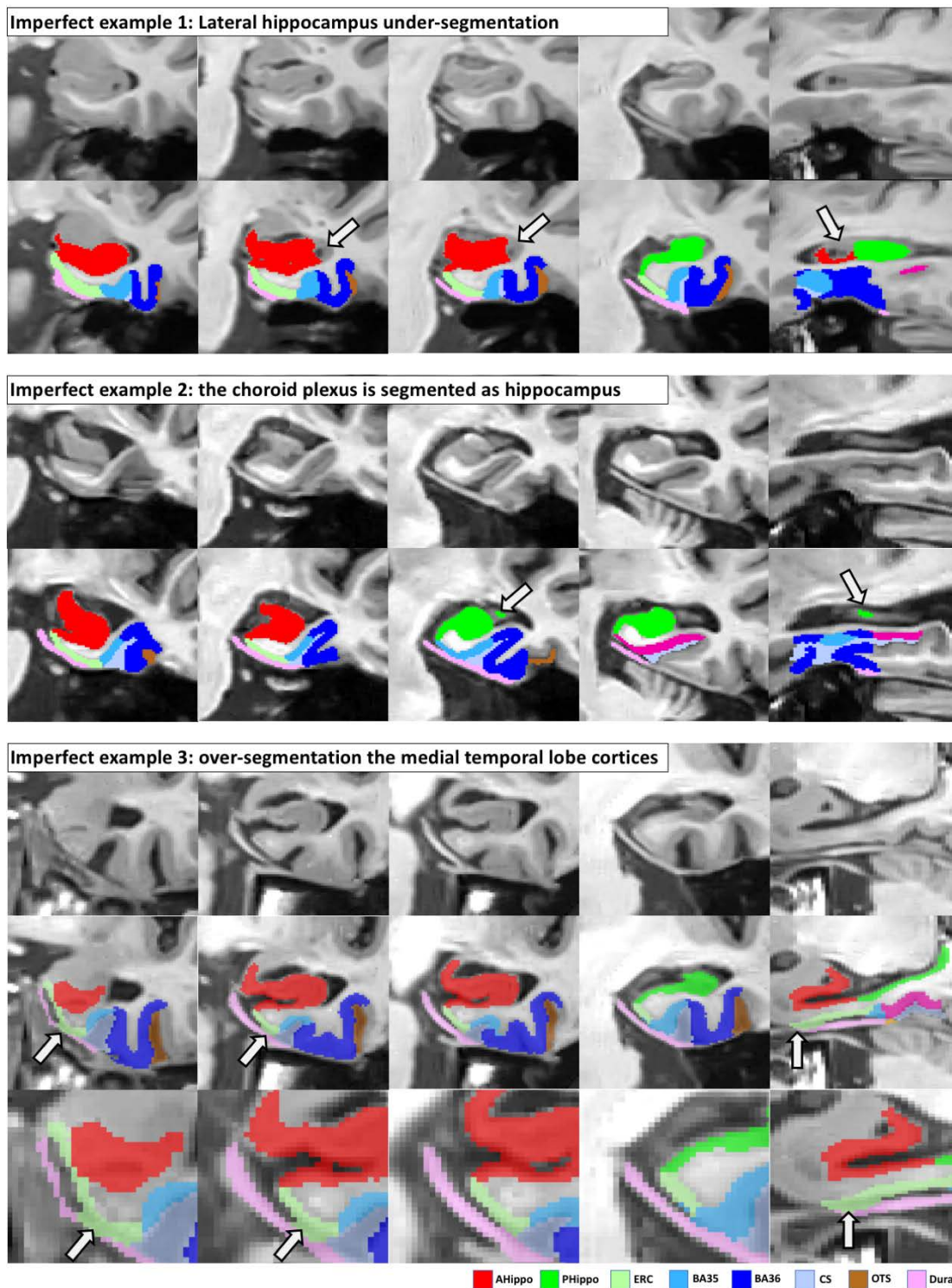
Supplementary Table 2. Statistical analysis results using volumetric measurements in discriminating patient groups from normal controls in ADNI. Quality control was performed to exclude low quality automatic segmentations. Measurements that survived

Bonferroni correction ( $p < 0.05/4$  for hippocampal measurements and  $p < 0.05/6$  for parahippocampal cortices measurements) are highlighted in bold.

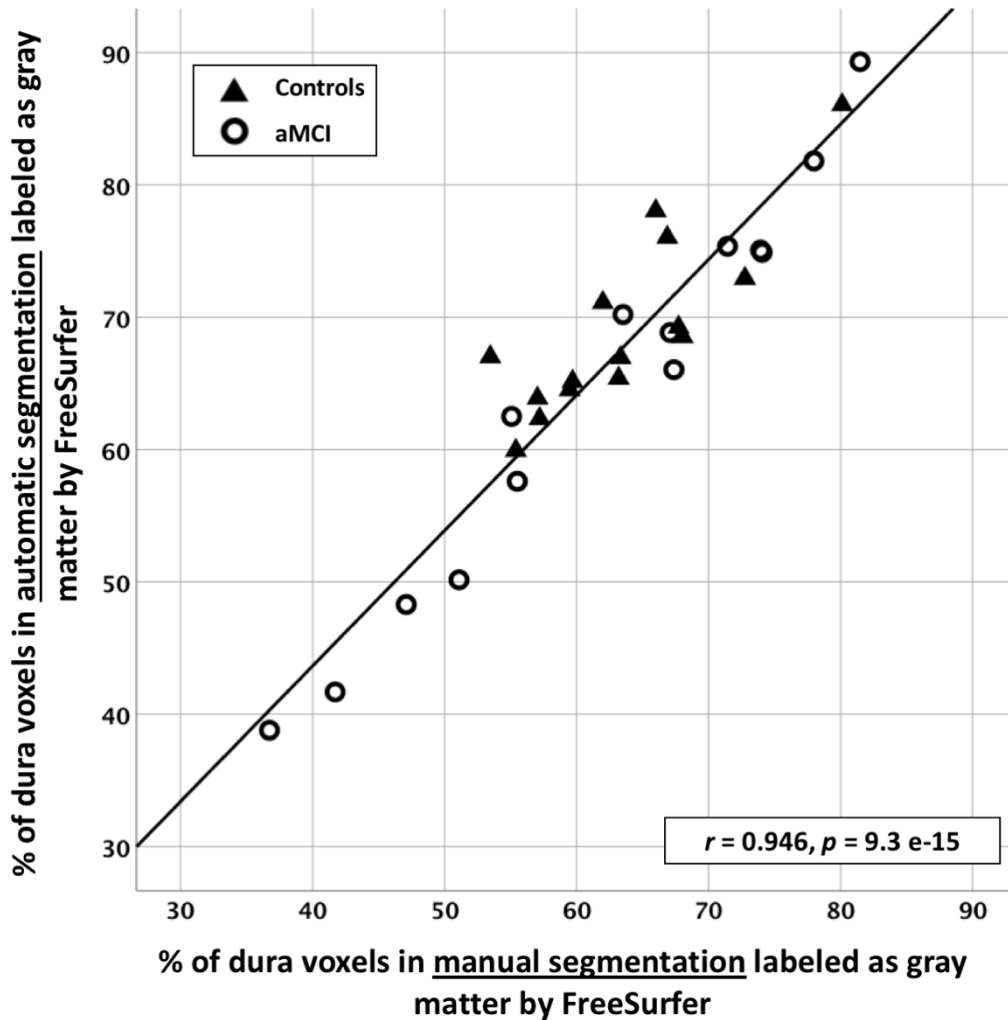
Region	A $\beta$ - Control	Preclinical AD	Early Prodromal AD	Late Prodromal AD	Dementia
<b>Hippocampal Volume Measurements (mm<sup>3</sup>), adjusted for age and ICV</b>					
<b>N</b>	183	93	134	102	123
<b>ASHS-T1 Anterior Hippocampus</b>	1721.1	1714.9	1660.6	<b>1530.2</b>	<b>1434.5</b>
SD	224.6	205.4	237.3	<b>249.6</b>	<b>199.4</b>
% Diff		-0.4	-3.5	<b>-11.1</b>	<b>-16.7</b>
F stats		<2.5	5.2	<b>43.8</b>	<b>128.7</b>
p value		>0.1	0.023	<b>1.9e-10</b>	<b>4.4e-25</b>
<b>ASHS-T1 Posterior Hippocampus</b>	1647.3	1648.8	<b>1566.7</b>	<b>1401.1</b>	<b>1345.6</b>
SD	160.9	152.9	<b>172.6</b>	<b>199.7</b>	<b>168.6</b>
% Diff		0.1	<b>-4.9</b>	<b>-14.9</b>	<b>-18.3</b>
F stats		<2.5	<b>18.3</b>	<b>128.8</b>	<b>245.8</b>
p value		>0.1	<b>2.5e-5</b>	<b>7.9e-25</b>	<b>6.0e-41</b>
<b>ASHS-T1 Hippocampus</b>	3368.4	3363.8	<b>3227.3</b>	<b>2931.2</b>	<b>2780.1</b>
SD	313.4	296.1	<b>348.8</b>	<b>402.7</b>	<b>322.1</b>
% Diff		-0.1	<b>-4.2</b>	<b>-13.0</b>	<b>-17.5</b>
F stats		<2.5	<b>14.1</b>	<b>103.6</b>	<b>248.9</b>
p value		>0.1	<b>2.1e-4</b>	<b>6.4e-21</b>	<b>2.6e-41</b>
<b>FS Hippocampus</b>	3779.9	3741.3	<b>3583.6</b>	<b>3292.0</b>	<b>3130.7</b>
SD	320.8	332.8	<b>357.0</b>	<b>415.6</b>	<b>320.4</b>
% Diff		-1.0	<b>-5.2</b>	<b>-12.9</b>	<b>-17.2</b>
F stats		<2.5	<b>25.2</b>	<b>123.1</b>	<b>304.1</b>
p value		>0.1	<b>8.9e-7</b>	<b>6.4e-21</b>	<b>1.3e-47</b>
<b>Parahippocampal Cortices Volume Measurements (mm<sup>3</sup>), adjusted for age and ICV</b>					
<b>N</b>	189	95	141	103	119
<b>ASHS-T1 ERC</b>	573.9	562.5	<b>549.7</b>	<b>502.3</b>	<b>451.6</b>
SD	74.0	69.8	<b>78.1</b>	<b>93.2</b>	<b>80.7</b>
% Diff		-2.0	<b>-4.2</b>	<b>-12.5</b>	<b>-21.3</b>
F stats		<2.5	<b>8.6</b>	<b>51.5</b>	<b>193.7</b>
p value		>0.1	<b>3.7e-3</b>	<b>6.0e-12</b>	<b>2.1e-34</b>
<b>ASHS-T1 BA35</b>	606.5	596.9	<b>581.5</b>	<b>543.1</b>	<b>481.4</b>
SD	82.0	80.2	<b>89.8</b>	<b>100.2</b>	<b>80.7</b>
% Diff		-1.6	<b>-4.1</b>	<b>-10.4</b>	<b>-20.6</b>
F stats		<2.5	<b>7.2</b>	<b>33.6</b>	<b>176.2</b>
p value		>0.1	<b>7.7e-3</b>	<b>1.7e-8</b>	<b>5.0e-32</b>
<b>ASHS-T1 BA36</b>	1878.3	1869.5	1816.6	<b>1719.4</b>	<b>1585.2</b>
SD	247.4	228.5	227.4	<b>259.4</b>	<b>229.0</b>
% Diff		-0.5	-3.3	<b>-8.5</b>	<b>-15.6</b>
F stats		<2.5	5.4	<b>26.3</b>	<b>190.2</b>
p value		>0.1	0.021	<b>2.4e-7</b>	<b>4.9e-22</b>
<b>ASHS-T1 PHC</b>	958.3	977.6	960.4	<b>910.0</b>	<b>867.1</b>
SD	118.2	129.0	138.0	<b>129.4</b>	<b>122.3</b>
% Diff		2.0	0.2	<b>-5.0</b>	<b>-9.5</b>

<i>F</i> stats		<2.5	<2.5	<b>10.5</b>	<b>44.6</b>
<i>p</i> value		>0.1	>0.1	<b>1.3e-3</b>	<b>1.1e-10</b>
<b>FS ERC</b>	802.8	768.2	805.0	<b>703.1</b>	<b>661.9</b>
SD	197.6	203.1	212.7	<b>230.7</b>	<b>194.1</b>
% Diff		-4.3	0.3	<b>-12.4</b>	<b>-17.5</b>
<i>F</i> stats		<2.5	<2.5	<b>15.2</b>	<b>37.0</b>
<i>p</i> value		>0.1	>0.1	<b>1.2e-4</b>	<b>3.6e-9</b>
<b>FS PRC</b>	1066.9	1051.9	<b>1022.2</b>	<b>922.3</b>	<b>839.8</b>
SD	150.2	142.1	<b>143.5</b>	<b>193.2</b>	<b>162.6</b>
% Diff		-1.4	<b>-4.2</b>	<b>-13.6</b>	<b>-21.3</b>
<i>F</i> stats		<2.5	<b>8.7</b>	<b>50.4</b>	<b>157.0</b>
<i>p</i> value		>0.1	<b>3.4e-3</b>	<b>9.8e-12</b>	<b>2.6e-29</b>
<b>Parahippocampal Cortices Thickness Measurements (mm), adjusted for age</b>					
<b>N</b>	189	95	141	103	119
<b>ASHS-T1 ERC</b>	2.02	2.02	2.00	<b>1.94</b>	<b>1.79</b>
SD	0.16	0.17	0.17	<b>0.18</b>	<b>0.22</b>
% Diff		0.0	-1.1	<b>-3.8</b>	<b>-11.3</b>
<i>F</i> stats		<2.5	<2.5	<b>14.3</b>	<b>116.0</b>
<i>p</i> value		>0.1	>0.1	<b>1.9e-4</b>	<b>3.9e-23</b>
<b>ASHS-T1 BA35</b>	2.35	2.33	<b>2.30</b>	<b>2.20</b>	<b>2.06</b>
SD	0.16	0.18	<b>0.17</b>	<b>0.22</b>	<b>0.23</b>
% Diff		-1.0	<b>-2.3</b>	<b>-6.4</b>	<b>-12.4</b>
<i>F</i> stats		2.8	<b>9.2</b>	<b>44.2</b>	<b>178.0</b>
<i>p</i> value		0.097	<b>2.7e-3</b>	<b>1.5e-10</b>	<b>2.7e-32</b>
<b>ASHS-T1 BA36</b>	2.41	2.39	2.39	<b>2.30</b>	<b>2.21</b>
SD	0.23	0.23	0.21	<b>0.21</b>	<b>0.24</b>
% Diff		-0.7	-0.7	<b>-4.8</b>	<b>-8.2</b>
<i>F</i> stats		<2.5	<2.5	<b>17.4</b>	<b>52.0</b>
<i>p</i> value		>0.1	>0.1	<b>4.0e-5</b>	<b>4.3e-12</b>
<b>ASHS-T1 PHC</b>	2.15	2.16	2.14	<b>2.09</b>	<b>2.00</b>
SD	0.13	0.16	0.13	<b>0.15</b>	<b>0.15</b>
% Diff		0.7	-0.2	<b>-2.7</b>	<b>-7.1</b>
<i>F</i> stats		<2.5	<2.5	<b>11.9</b>	<b>94.0</b>
<i>p</i> value		>0.1	>0.1	<b>6.4e-4</b>	<b>1.5e-19</b>
<b>FS ERC</b>	3.18	3.17	3.15	<b>2.92</b>	<b>2.75</b>
SD	0.30	0.33	0.34	<b>0.39</b>	<b>0.41</b>
% Diff		-0.2	-1.0	<b>-8.2</b>	<b>-13.5</b>
<i>F</i> stats		<2.5	<2.5	<b>40.4</b>	<b>113.7</b>
<i>p</i> value		>0.1	>0.1	<b>8.1e-10</b>	<b>8.9e-23</b>
<b>FS PRC</b>	3.44	3.47	3.37	<b>3.13</b>	<b>2.88</b>
SD	0.31	0.33	0.33	<b>0.41</b>	<b>0.37</b>
% Diff		0.9	-2.0	<b>-8.8</b>	<b>-16.1</b>
<i>F</i> stats		<2.5	4.4	<b>51.0</b>	<b>201.9</b>
<i>p</i> value		>0.1	0.036	<b>7.4e-12</b>	<b>1.6e-35</b>

Supplementary Figure 1. The three common errors of the ASHS-T1 pipeline, indicated by the white arrows. In order to better visualize the error in the third example, the transparency of the segmentation was set to 60% and zoom-in images were provided.



Supplementary Figure 2. The percentages of dura voxels labeled as gray matter by FreeSurfer computed using automatic and manual segmentation are highly correlated ( $r = 0.946$ ,  $p = 9.3e-15$ ).



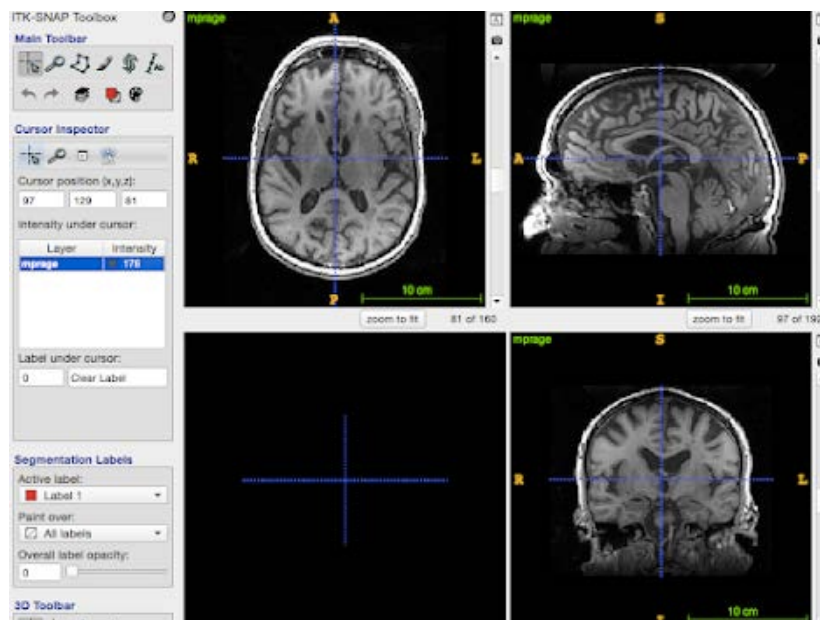
**Supplementary material A. Cloud-based ASHS-T1 service tutorial**

This is a tutorial for using the cloud-based ASHS-T1 service, made available via ITK-SNAP 3.8.0-BETA or higher version (available to download from: <http://www.itksnap.org/pmwiki/pmwiki.php?n=Downloads.SNAP3>). This service is designed to segment the medial temporal lobe (MTL) subregions, including anterior/posterior hippocampus, ERC, BA35, BA36 and PHC, from 3T whole-brain T1-weighted MRI scans (MPRAGE, ~1.0 mm<sup>3</sup> isotropic resolution) of older adults (55+ age). No skull-stripping is necessary. Below is the step-by-step instruction:

### (1) Create an ITK-SNAP workspace

To send your images to the cloud, you need to create an ITK-SNAP workspace that contains the T1-weighted images.

- Download and install ITK-SNAP 3.8-BETA or higher version
- Open ITK-SNAP
- Use "*File->Open Main Image...*" to load the **T1-weighted image** as the main image



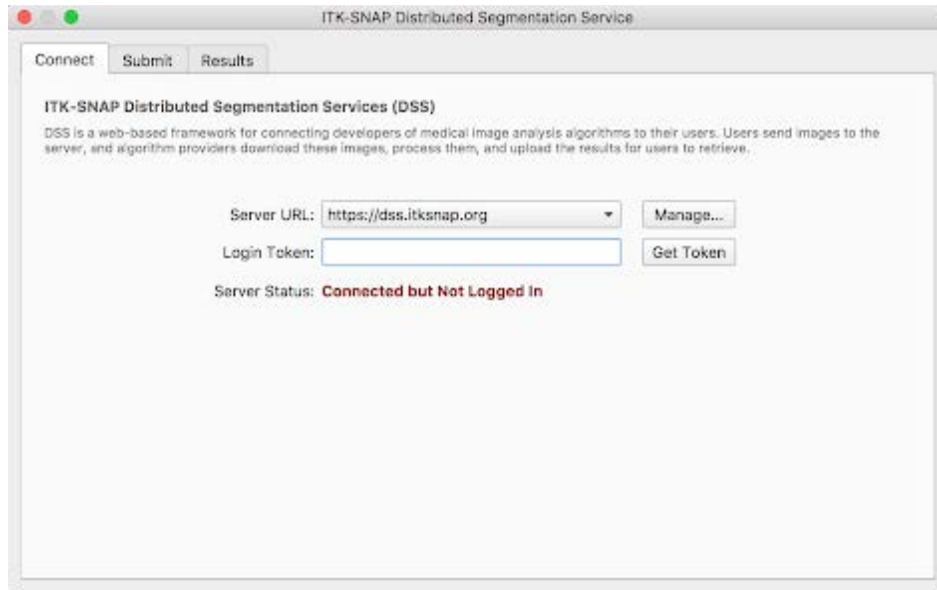
- Save the ITK-SNAP workspace to a file using "*Workspace->Save Workspace*"

### (2) Login to the Distributed Segmentation System (DSS)

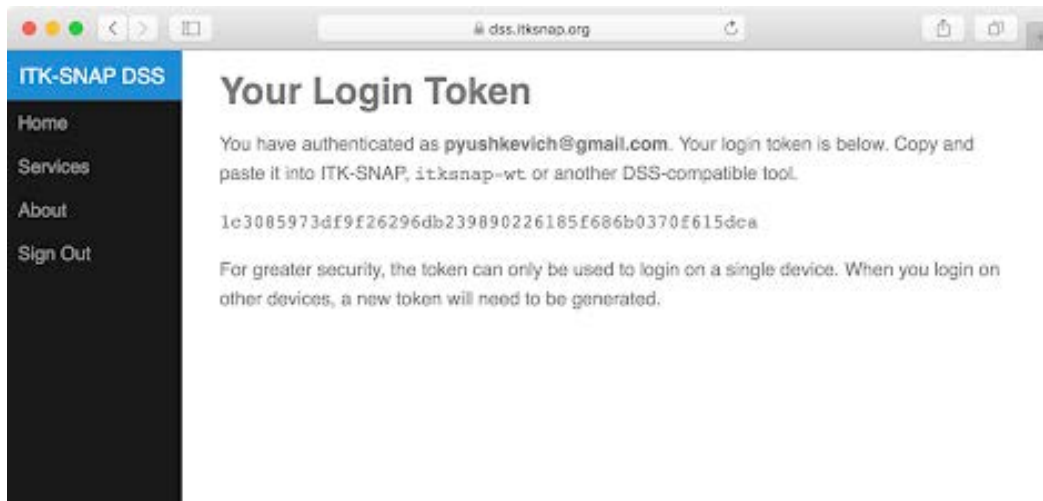
Next, you need to login to the DSS server. This requires a Google account.

- Open the DSS window using "*Tools->Distributed Segmentation Service*"





- Select the "Connect" tab if not already selected
- Press the "Get Token" button. This will open a web browser and take you to the login page for DSS, shown to the right.
  - Follow the login prompts on the website. Once you login with your Google account, you will see a 40-character **login token**. Copy the token and paste it back into ITK-SNAP.



- The ITK-SNAP DSS window should show an updated status, in green, indicating that you successfully logged into the system
  - Next time you open ITK-SNAP on the same device, your login information will be saved. You will not need to repeat the token-based login process.

### (3) Upload Your Workspace and Create a Ticket

In this step, you will send your workspace to the cloud for segmentation with ASHS.

- Select the "Submit" tab in the ITK-SNAP DSS window

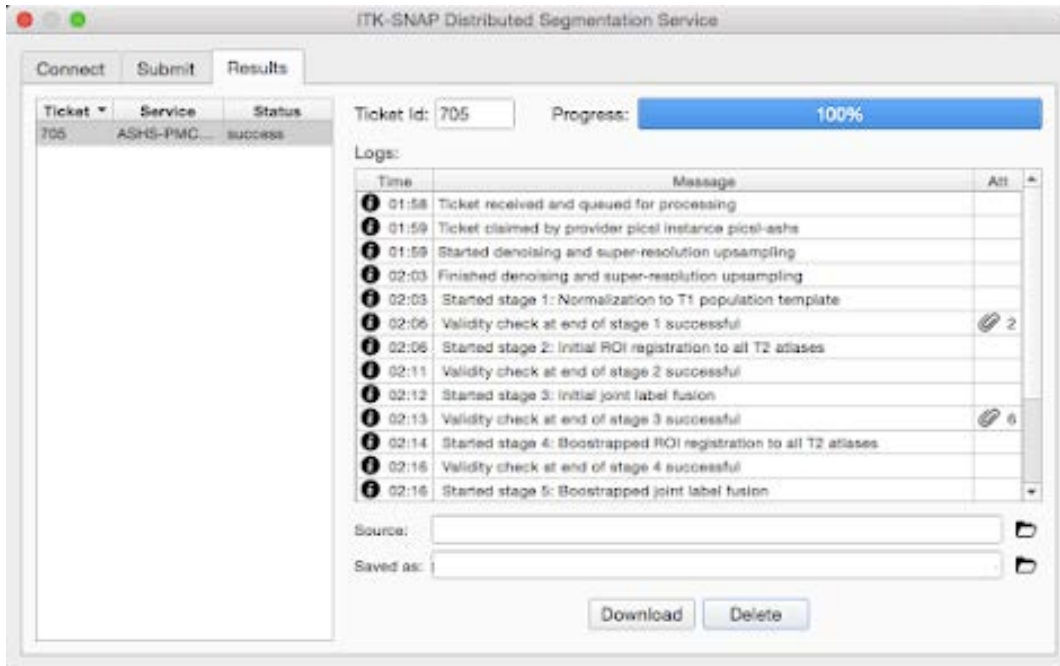
Tag	Type	Required	Target Object
T1-MRI	Main Image	<input checked="" type="checkbox"/>	mprage
AC	Point Landmark	<input type="checkbox"/>	Unassigned
PC	Point Landmark	<input type="checkbox"/>	Unassigned

- Select the service you wish to use. In this tutorial, we are using the ASHS-PMC-T1 service. The image specifications for this service are described here.
- Ignore the AC/PC landmarks in the table. These are not yet supported.
- Press "Submit". The workspace will be uploaded to the server and a new ticket will be created.

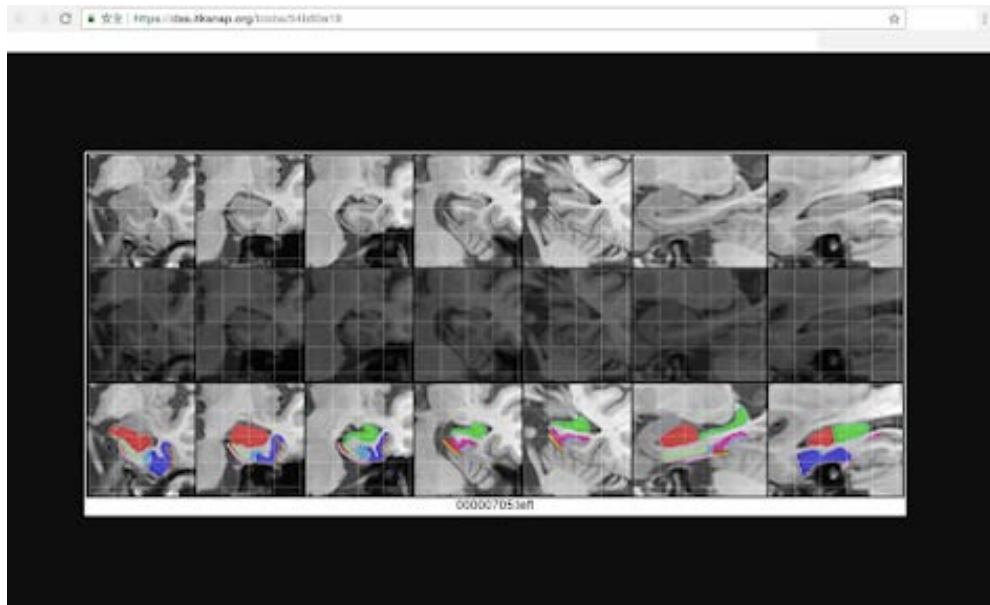
### (4) Monitoring Ticket Status

Once a ticket is created, the DSS window will switch to the "Results" tab, where you can monitor the progress of your ticket. While the ticket is being processed, you can close ITK-SNAP or submit additional tickets.

- Your ticket will be highlighted in the list of tickets on the left of the window, and the progress of your ticket will be shown on the right.



- Note the status of your ticket:
  - **ready**: the ticket has been uploaded and is in the queue to be processed
  - **claimed**: the ticket is actively being processed
  - **success**: the ticket was successfully processed and results are available
  - **failed/timed out**: an error occurred while servicing the ticket
- Once your ticket is in **claimed** state, log messages show the progress of ASHS segmentation. Some log messages have attachments in the form of QC (quality control) images. Click the attachment button to list the attachments and open them in a web browser.

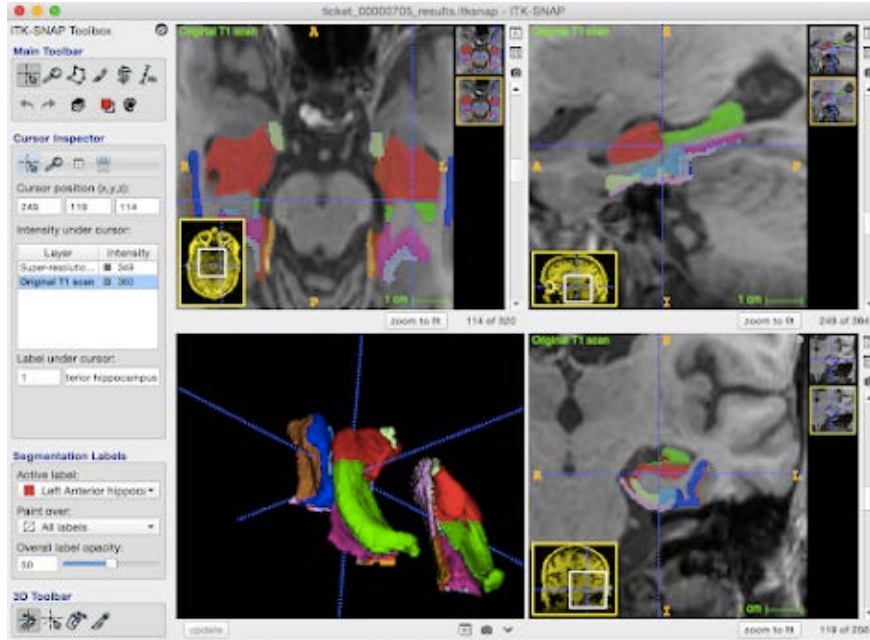


- To abort ticket processing, press the "Delete" button

#### (5) Download and Examine ASHS Results

Once a ticket completes successfully, you can download the results as an ITK-SNAP workspace.

- Select a ticket that is in **success** state and press the "*Download*" button
- A dialog shown on the right will appear.
- Choose where to save the result workspace and whether to open it after downloading.
- The segmentation will be opened as shown below.



In addition to ASHS-T1, more services are available to analyze other datasets. Please check <https://dss.itksnap.org/services> for a complete list of services.

## Supplementary material B. HARP atlas set

We selected T1-weighted MRI scans and their hippocampal manual segmentations of from the HARP training set (Boccardi et al., 2015b) from ADNI that is publicly available (<http://www.hippocampal-protocol.net/SOPs/index.php>). In total, 15 controls and 15 MCI patients were selected. The Automated Segmentation of Hippocampal Subfields (ASHS) algorithm (see Section 2.4 for a brief description) was run to build an atlas, for which the cross-validation showed a Dice similarity coefficient of 0.90 and 0.91 for left and right hippocampus respectively. Demographic information of the selected subjects are available in Supplementary Table 4.

Supplementary Table 4. Characteristics of the HARP atlas

	Control	MCI	AD
N	11	13	8

<b>Age (yrs)</b>	77.3 (8.8)	75.6 (7.6)	77.5 (6.2)
<b>Gender (M/F)</b>	7 / 4	8 / 5	6 / 2
<b>Education (yrs)</b>	16.8 (2.7)	16.2 (2.8)	15.0 (2.6)

Note: All statistics are in comparison to amyloid- $\beta$  negative (A $\beta$ -) control subjects. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . Independent two-sample t-tests (age, education) and contingency  $\chi^2$  test (gender) were used. Standard deviation in parentheses. Abbreviations: AD = Alzheimer's disease.

## Supplementary material C. ICV atlas set

Whole-brain T1-weighted MRI and CT scans of 15 controls, 12 aMCI subjects from the PMC/ADC center were used to develop a training set for automatic intracranial volume (ICV) segmentation. Intracranial volume was manually traced in the space of the T1-weighted MRI with the guidance of the coregistered CT scans of the same subjects. Since the boundary between the skull and soft tissue can be clearly visualized in CT scans, we were able to obtain an accurate manual segmentation of the intracranial space. Similarly, the Automated Segmentation of Hippocampal Subfields (ASHS) algorithm (see Section 2.4 for a brief description) was run to build an atlas, for which the cross-validation showed a Dice similarity coefficient of 0.98. Demographic information of the ICV atlas set is shown in Supplementary Table 4.

Supplementary Table 4. Characteristics of the ICV atlas

	<b>Normal Control</b>	<b>aMCI</b>
<b>N</b>	15	12
<b>Age (yrs)</b>	70.2 (7.9)	74.0 (9.2)
<b>Gender (M/F)</b>	3 / 12	9 / 3 **
<b>Education (yrs)</b>	16.9 (2.6)	16.7 (3.1)

Note: All statistics are in comparison to cognitive normal control subjects. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . Independent two-sample t-tests (age, education) and contingency  $\chi^2$  test (gender) were used. Standard deviation in parentheses.