NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Heritability and reliability of automatically segmented human hippocampal formation subregions

Christopher D. Whelan [a], Derrek P. Hibar [a], Laura S. van Velzen [b], Anthony S. Zannas [c,d], Tania Carrillo-Roa [c], Katie McMahon [e], Gautam Prasad [a], Sinéad Kelly [a], Joshua Faskowitz [a], Greig deZubiracay [f], Juan E. Iglesias [g], Theo G.M. van Erp [h], Thomas Frodl [i], Nicholas G. Martin [j], Margaret J. Wright [k], Neda Jahanshad [a], Lianne Schmaal [b], Philipp G. Sämann [c], Paul M. Thompson [a,*], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] *Imaging Genetics Center, University of Southern California, Marina del Rey, CA, USA*
[b] *Department of Psychiatry and Neuroscience Campus Amsterdam, VU University Medical Center and GGZ inGeest, Amsterdam, The Netherlands*
[c] *Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany*
[d] *Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA*
[e] *Centre for Advanced Imaging, University of Queensland, Brisbane, Australia*
[f] *Faculty of Health, Queensland University of Technology, Brisbane, Australia*
[g] *Basque Center on Cognition, Brain and Language, Donostia, Gipuzkoa, Spain*
[h] *Department of Psychiatry and Human Behavior, University of California, Irvine, USA*
[i] *Department of Psychiatry, Otto-von Guericke-University of Magdeburg, Germany*
[j] *QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia*
[k] *Queensland Brain Institute, University of Queensland, Brisbane, Australia*

ARTICLE INFO

Article history:
Received 29 July 2015
Accepted 23 December 2015
Available online xxxx

ABSTRACT

The human hippocampal formation can be divided into a set of cytoarchitecturally and functionally distinct subregions, involved in different aspects of memory formation. Neuroanatomical disruptions within these subregions are associated with several debilitating brain disorders including Alzheimer's disease, major depression, schizophrenia, and bipolar disorder. Multi-center brain imaging consortia, such as the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium, are interested in studying disease effects on these subregions, and in the genetic factors that affect them. For large-scale studies, automated extraction and subsequent genomic association studies of these hippocampal subregion measures may provide additional insight. Here, we evaluated the test–retest reliability and transplatform reliability (1.5 T *versus* 3 T) of the subregion segmentation module in the *FreeSurfer* software package using three independent cohorts of healthy adults, one young (Queensland Twins Imaging Study, N = 39), another elderly (Alzheimer's Disease Neuroimaging Initiative, ADNI-2, N = 163) and another mixed cohort of healthy and depressed participants (Max Planck Institute, MPIP, N = 598). We also investigated agreement between the most recent version of this algorithm (v6.0) and an older version (v5.3), again using the ADNI-2 and MPIP cohorts in addition to a sample from the Netherlands Study for Depression and Anxiety (NESDA) (N = 221). Finally, we estimated the heritability (h^2) of the segmented subregion volumes using the full sample of young, healthy QTIM twins (N = 728). Test–retest reliability was high for all twelve subregions in the 3 T ADNI-2 sample (intraclass correlation coefficient (ICC) = 0.70–0.97) and moderate-to-high in the 4 T QTIM sample (ICC = 0.5–0.89). Transplatform reliability was strong for eleven of the twelve subregions (ICC = 0.66–0.96); however, the hippocampal fissure was not consistently reconstructed across 1.5 T and 3 T field strengths (ICC = 0.47–0.57). Between-version agreement was moderate for the hippocampal tail, subiculum and presubiculum (ICC = 0.78–0.84; Dice Similarity Coefficient (DSC) = 0.55–0.70), and poor for all other subregions (ICC = 0.34–0.81; DSC = 0.28–0.51). All hippocampal subregion volumes were highly heritable (h^2 = 0.67–0.91). Our findings indicate that eleven of the twelve human hippocampal subregions segmented using *FreeSurfer* version 6.0 may serve as reliable and informative quantitative phenotypes for future multi-site imaging genetics initiatives such as those of the ENIGMA consortium.

* Corresponding author at: Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, 4676 Admiralty Way, Marina del Rey, CA 90292, USA.
E-mail address: pthomp@usc.edu (P.M. Thompson).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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://dx.doi.org/10.1016/j.neuroimage.2015.12.039
1053-8119/

Introduction

The mammalian hippocampal formation is one of the most important brain regions for spatial navigation (O'Keefe, 1990), episodic memory retrieval (Burgess et al., 2002), and associative learning processes (Morris, 2006). This seahorse-shaped structure in the medial temporal lobe is divided into a set of cytoarchitectonically heterogeneous subregions (Insausti and Amaral, 2004; Winterburn et al., 2013; Pipitone et al., 2014), each associated with distinct aspects of memory formation, among other functions. For example, the dentate gyrus (DG) and sectors 3 and 4 of the *cornu ammonis* (CA) are involved in declarative memory acquisition (Coras et al., 2014), whereas the subiculum and CA1 are associated with disambiguation during working memory processes (Newmark et al., 2013). The CA2 subregion, long assumed to be a simple transition point between CA3 and CA1, has recently been implicated in animal models of social memory (Hitti and Siegelbaum, 2014) and episodic time encoding (Navratilova and Battaglia, 2015). The subiculum, a subregion that exerts control over the hippocampal output, has been associated with spatial memory functions, but its ventral part may play an additional regulatory role in inhibition of the HPA axis (O'Mara, 2006).

Neuroanatomical abnormalities within these hippocampal subregions are associated with a broad range of neurological and psychiatric disorders, from ischaemic stroke, encephalitis, temporal lobe epilepsy, transient global amnesia and multiple sclerosis (Bartsch, 2012; Das et al., 2011) to bipolar disorder (BPD), major depressive disorder (MDD) and posttraumatic stress disorder (PTSD) (Sala, 2008). Some of these malformations develop as a result of head trauma, intracranial infection or other environmental influences, but genetic factors also play a fundamental role (Thompson et al., 2008; van Erp et al., 2004). Recent advances in genome-wide association (GWA) meta-analysis and large-scale collaborative brain imaging (*e.g.* Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA), the Early Growth Genetics (EGG) consortium, and the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE)) have helped identify several common genetic variants associated with structural variation in the hippocampus (Hibar et al., 2015; Stein et al., 2012; Lim et al., 2012) as well as other brain regions including the putamen, caudate nucleus (Hibar et al., 2015), intracranial volume (Ikram et al., 2012; Stein et al., 2012) and head circumference (Taal et al., 2012).

International consortia like ENIGMA are now turning their attention to specific investigations of genetic and phenotypic variation in healthy individuals as well as those diagnosed with schizophrenia, BPD, MDD, PTSD, epilepsy and many other brain illnesses (Thompson et al., 2014). Among subcortical structures assessed, the hippocampus has consistently shown the greatest effect sizes for differences between patients and controls, in both schizophrenia (van Erp et al., 2015) and major depression, particularly recurrent depression (Schmaal et al., 2015). Impaired hippocampal integrity may in turn impair treatment response, making it pivotal to detect such morphologically defined subgroups (Frodl et al., 2008; Sämann et al., 2013).

Focusing on fine-grained phenotypic variation within small subregions of the hippocampus may improve our power to localize genetic and disease-related effects on the brain as a whole. As part of its next major project, the ENIGMA consortium aims to delineate specific sub-regions of the hippocampus as quantitative phenotypes for genome-wide association and cross-sectional case:control meta-analyses. Before these new ENIGMA initiatives can begin, we first need to evaluate a non-invasive, reliable and relatively accessible technique for reconstructing the human hippocampal subfields *in vivo*. In turn, for future genetic mapping efforts, we must validate these automatically reconstructed hippocampal sub-regions as quantitative *endophenotypes* — heritable, robust brain markers that may be closer to the molecular basis of disease than diagnostic assessments in the clinic (Braskie and Ringman, 2011; Glahn et al., 2007; Gottesman and Gould, 2003; Hasler and Northoff, 2011).

Several manual segmentation techniques have been developed to reconstruct hippocampal and parahippocampal subregions from T1-weighted MRI scans acquired at 3 to 7 T field strengths (La Joie et al., 2010; Van Leemput et al., 2009; Mueller et al., 2007; Wisse et al., 2012; Adler et al., 2014). Although these methods typically segment the hippocampal subregions at remarkably fine-scaled resolution, a critical bottleneck for collaborative imaging initiatives such as ENIGMA is the need to manually label the subregion boundaries, which is laborious, time-consuming and susceptible to intra- and inter-observer variability (Van Leemput et al., 2009). Several automated protocols have been developed to address this issue, combining rules on image intensity and geometry to delineate the boundaries between hippocampal and parahippocampal subregions (Van Leemput et al.; Yushkevich et al., 2009, 2010). One often-used automated technique is provided as part of FreeSurfer, a freely available suite of neuroimaging structural analysis tools (Fischl, 2012).

Initial versions of the FreeSurfer algorithm (versions 5.1, 5.2 and 5.3) produce subregion segmentations that are largely inconsistent with brain anatomy (de Flores et al., 2015; Pluta et al., 2012; Wisse et al., 2014). An updated version of the algorithm, to be released as part of FreeSurfer version 6.0, uses a new statistical atlas constructed from ultra-high resolution *ex vivo* MRI (Iglesias et al., 2015). This revised algorithm produces subregion volume estimates that more closely match volumes derived from histological investigations (Iglesias et al., 2015). However, consensus is still lacking on the most appropriate subregion delineation protocol to use (Yushkevich et al., 2015). Here, using four independent samples, we set out to validate version 6.0 of the automated FreeSurfer algorithm from three complementary perspectives: First, we evaluated the algorithm's 'test–retest' reliability; *i.e.* its ability to extract comparable subregion measures across multiple time points in two independent cohorts with different image acquisition parameters and age characteristics (our two samples differ in mean age by approximately 50 years). Second, we examined the algorithm's 'trans-platform' reliability — defined as its ability to reproduce similar subregion measures across different MRI scanner platforms and field strengths (for example, 3 T *versus* 1.5 T). Third, we investigated overall agreement between this new algorithm, which we will refer to as 'FS6.0', and the older algorithm, version 5.3, which we will refer to as 'FS5.3'. The degree of quantitative deviation between volumes extracted using FS5.3 and volumes extracted using FS6.0 may help users of the former evaluate the necessity of re-processing their data with the latter.

Validation of a reliable, automated subregion segmentation tool may allow ENIGMA and other imaging consortia to study hippocampal subregions as fine-grained quantitative phenotypes in large-scale genome-wide association meta-analyses. However, to be considered a promising target for genetic mapping, the subregional volume estimates must show evidence of heritability (h^2). Quantitative genetic analysis of automatically segmented, T1-weighted brain images from paired twin samples has frequently been employed to estimate the heritability of global volumetric measures. Prior estimates show that total hippocampal volume is highly heritable in both healthy adults (h^2 = 0.66–0.71) (den Braber et al., 2013; Erp and Saleh, 2004; Wright et al., 2002) and children (h^2 = 0.64–0.72) (Swagerman and Brouwer, 2014). However, structural variance within the whole hippocampus may be less heritable in elderly adults (h^2 = 0.4–0.65) (DeStefano et al., 2009; Mather et al., 2015; Sullivan et al., 2001), possibly due to environmental stressors (Hedges and Woon, 2010), alterations in testosterone levels (Panizzon et al., 2012) or other endogenous biological factors. Similarly, total hippocampal volume is only moderately heritable in schizophrenia (h^2 = 0.36–0.73) (Kaymaz and Os, 2009; Roalf et al., 2015). Thus, while the heritability of total hippocampal volume is well established across many populations, the heritability of structural variations in individual subregions has yet to be delineated. Therefore, in the second part of this study, we set out to disentangle the relative contributions of additive genetic variance and

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

environmental influences on hippocampal subregion volume in two independent cohorts of healthy adults, and by this to assess the eligibility of such hippocampal subregion volumes as *endophenotypes* for future large-scale collaborative genetic association studies in ENIGMA.

Methods

Participants and imaging protocols

Four collections of MRI scans were analyzed in this study.

ADNI-2

Subjects. For our test–retest and between-version reliability analyses, we analyzed publicly available data from 163 healthy control subjects from the second phase of the Alzheimer's Disease Neuroimaging Initiative, ADNI-2 (81 women, 82 men, age mean ± SD = 73.58 ± 6.21 years) (http://adni.loni.usc.edu/). 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 (MCI) and early Alzheimer's disease (AD). Further details of the ADNI project are given in Jack et al. (2010) and at http://www.adni-info.org.

Imaging. T1-weighted MR images were acquired using a 3 T General Electric (GE) Medical Systems scanner with the following parameters: 3-dimensional MP-RAGE, 8-channel head coil, voxel size 1.2 × 1.2 × 1.2 mm, time to repeat (TR) = 400 ms, time to echo (TE) = 2.85 ms, flip angle = 11°, field of view (FOV) = 26 cm, resolution = 256 × 256 mm. A baseline and follow-up scan was acquired for all healthy controls, with an average inter-scan interval of 3.3 months. Family trios or siblings were not scanned as part of the ADNI-2 protocol, so this dataset was not included in our heritability analyses.

QTIM

Subjects. To estimate heritability and include an independent replication cohort for our test–retest reliability analysis, we analyzed MR images from healthy Caucasian young adults, collected as part of the Queensland Twins Imaging (QTIM) study. QTIM is a joint effort by researchers at QIMR Berghofer, The University of Queensland and the University of Southern California to study brain structure and function using T1-weighted MRI, high angular resolution diffusion imaging (HARDI) and functional MRI in a large population of young adult twins of European ancestry. Full details of the QTIM cohort are found in Zubicaray et al. (2008).

The heritability analysis included 728 individuals (132 monozygotic (MZ) sibling pairs and 232 dizygotic (DZ) sibling pairs; 465 women and 263 men with an age mean ± SD of 22.65 ± 2.73 years). The test–retest reliability analysis included a subset of the twins; 20 women, 19 men; mean age in years (±SD) = 24.03 (±2.04), who were scanned twice, with an average interval of 3 months between scanning sessions.

Imaging. 3-Dimensional T1-weighted images were acquired on a 4 T Bruker Medspec scanner using an inversion recovery rapid gradient echo protocol. Key acquisition parameters were: TI = 700 ms, TR = 1500 ms, TE = 3.35 ms, voxel size 0.94 × 0.98 × 0.98 mm, flip angle = 8°, slice thickness = 0.9 mm, 256 × 256 acquisition matrix.

Max Planck Institute of Psychiatry (MPIP)

Subjects. As part of the (i) between-version agreement and (ii) transplatform reliability analyses, high resolution T1-weighted anatomical images collected at the MPIP of Psychiatry (MPIP), Munich, Germany, from 222 healthy participants and 367 patients major depressive disorder (MDD) (334 women, 255 men, mean age ± SD = 48.4 ± 13.5, age range: 18 to 87), were included, in addition to 20 healthy controls who were scanned on a 1.5 T and 3 T platform

Imaging. The between-version comparison sample (total N = 589) was acquired on a 1.5 T General Electric clinical scanner (T1-weighted SPGR 3D volume, TR 10030 ms; TE 3.4 ms; 124 sagittal slices; matrix 256 × 256; FOV 23.0 × 23.0 cm^2; voxel size 0.8975 × 0.8975 × 1.2–1.4] mm^3; flip angle = 90°; birdcage resonator) with N = 186 of the total sample scanned after a coil upgrade (Signa Excite, sagittal T1-weighted spin echo sequence, TR 9.7 s, TE 2.1 ms). For the trans-platform sample, one image was acquired on 3 T scanner (General Electric MR750, 3D BRAVO, TR 6.1 s; TE minimum; TI 450 ms, 124 sagittal slices; matrix 256 × 256; FOV 25.6 × 25.6 cm^2; voxel size 1 × 1 × 1 mm^3; flip angle = 12°) and a second image after immediate repositioning in the 1.5 T scanner (General Electric MR450, 3D FSPGR, TR 7.9 s; TE minimum, TI 450 ms, 188 sagittal slices; matrix 320 × 256; FOV 24 × 24 cm^2; voxel size 0.9375 × 0.9375 × 1 mm^3; flip angle = 12°).

Netherlands Study of Depression and Anxiety (NESDA)

Subjects. To further assess the agreement between FreeSurfer versions, we analyzed data from 64 healthy controls and 157 patients with a diagnosis of MDD or comorbid anxiety disorder, collected as part of the Netherlands Study for Depression and Anxiety (NESDA) (145 women, 76 men, mean age ± SD = 38.14 ± 10.33 years, age range: 18 to 57).

Imaging. Imaging data were acquired using Philips 3 T magnetic resonance imaging systems (Best, The Netherlands) located at the Leiden University Medical Center, Amsterdam Medical Center, and University Medical Center Groningen. For each subject, anatomical images were obtained using a sagittal 3-dimensional gradient-echo T1-weighted sequence (repetition time, 9 ms, echo time, 3.5 ms; matrix, 256 × 256; voxel size, 1 × 1 × 1 mm; 170 slices; duration, 4.5 min).

Full participant demographics for the ADNI-2, QTIM, MPIP and NESDA samples are detailed in Table 1.

Image processing

T1-weighted images were processed using FreeSurfer (FS) version 5.3.0 using the software package's default, automated reconstruction protocol described by Anders M. Dale, Bruce Fischl and colleagues ('recon-all'—see Dale et al., 1999; Fischl et al., 1999). Briefly, each T1-weighted image was subjected to an automated segmentation process involving: (i) conversion from three-dimensional *nifti* format, (ii) affine registration into Talairach space, (iii) normalization for variable intensities caused by inhomogeneities in the radiofrequency field, (iv) 'skull-stripping', *i.e.* extraction of the skull and extra-meningeal tissues from each image, (v) segregation into left and right hemispheres using 'cutting planes', (vi) removal of the brain stem and cerebellum, (vii) correction for topology defects, (viii) definition of the gray/white matter and gray/cerebrospinal fluid boundaries using surface deformation (Fischl et al., 2004a) and (ix) parcellation of the subcortical region into distinct brain tissues, including the hippocampus, amygdala, thalamus, caudate nucleus, putamen, pallidum and accumbens (Fischl et al., 2002, 2004a, 2004b). Using FreeSurfer's native visualization toolbox, *tkmedit*, we visually inspected each image for over- or under-estimation of the gray/white matter boundaries and to identify brain areas erroneously excluded during skull stripping.

Hippocampal subregion segmentation

After successful reconstruction of the whole hippocampus and its neighboring subcortical regions, we used a revised version of the automated subregion parcellation protocol previously described by

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

Table 1
Participant demographics.

Cohort	N	Field strength	Mean age (years + SD)	Age range (years)	Female/male
ADNI-2	163	3 T	73.6	56.3–89.1	81/82
QTIM (test–retest)	39	4 T	24.03 (3.49)	20.72–27.31	20/19
QTIM (full)[a]	728	4 T	22.65 (2.73)	18.1–29.73	465/263
NESDA	221	3 T	38.14 (10.33)	18–57	145/76
MPIP	589	1.5 T	48.4 (13.5)	18–87	334/255

'SD' = standard deviation, MPIP = Max Planck Institute of Psychiatry, NESDA = Netherlands Study of Depression and Anxiety, ADNI-2 = Alzheimer's Disease NeuroImaging Initiative, QTIM = Queensland Twins Imaging Study.

[a] The QTIM cohort included 132 monozygotic twin pairs and 232 dizygotic twin pairs.

Van Leemput and colleagues (Van Leemput et al., 2008; Van Leemput et al., 2009) to segment specific subregions of the hippocampal formation in the QTIM, ADNI-2, NESDA and MPIP datasets. This revised module is compatible with FreeSurfer v5.3 (FS5.3) and will be freely distributed with FreeSurfer v6.0 (FS6.0) (Iglesias et al., 2015). Prior versions of the algorithm (FS5.1 to FS5.3) combined a single probabilistic atlas with high-resolution, T1-weighted *in-vivo* manual segmentations to predict the locations of eight hippocampal subregions. The new version (FS6.0) predicts the location of twelve hippocampal subregions, using a refined probabilistic atlas built upon a combination of manual delineations of the hippocampal formation from 15 ultra-high resolution, *ex-vivo* MRI scans and manual annotations of the surrounding subcortical structures (*e.g.*, amygdala, cortex) from an independent dataset of 39 *in-vivo*, T1-weighted, 1 mm resolution MRI scans (Iglesias et al., 2015). This revised algorithm features the following enhancements: (i) first-hand knowledge of histological staining of the hippocampus by a neuroanatomist; (ii) a cytoarchitectural atlas of the hippocampal formation (Rosene and Hoesen, 1987); and (iii) high-resolution, *ex-vivo* brain MRI scans (120 μm^3), which show definitive borders between the subregions and greater consistency with manual segmentation methods (Yushkevich et al., 2015). Previous versions of the FreeSurfer algorithm reconstructed eight subregions per hemisphere, including the CA1, CA2/3, fimbria, subiculum, presubiculum, CA4/DG, hippocampal tail and hippocampal fissure. The new algorithm provides more anatomically sensitive reconstructions of these eight subregions as well as four new subregions: the parasubiculum, the molecular layer, granule cells in the molecular layer of the DG (GC-ML-DG) and the hippocampal-amygdala transitional area (HATA).

Test–retest reliability analysis

Using FS6.0, we extracted volume estimates for the whole hippocampus and its twelve subregions from (i) the ADNI-2 and (ii) the QTIM cohorts. All QTIM and ADNI-2 images, including both test and re-test scans, were processed in parallel. After successful subregion segmentation, we used a custom-designed Matlab code to visually inspect each segmentation (see Fig. 1). Subregion volume estimates were exported to SPSS (for reliability analysis) and reformatted into phenotype covariance matrices (for heritability analysis described below).

Volume measures were imported into SPSS (IBM Corp., Version 21.0) and subjected to a series of two-way reliability analyses, using Cronbach's alpha (α) (Cronbach, 1951) as a measure of internal consistency. Cronbach's alpha is calculated as follows:

$$\propto = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$$

where N is the number of subregion volume estimates, *c-bar* is the average inter-subject covariance among these estimates and *v-bar* is the average variance. The resulting α, interpreted as the *intraclass correlation coefficient* (ICC), provides an estimate of how consistently the FreeSurfer v6.0 parcellation protocol reconstructs hippocampal subregions from baseline to follow-up scan. ICC ranges from 0 (indicating high variability between baseline and follow-up volume estimates) to 1 (denoting high reproducibility between baseline and follow-up estimates).

Between-version reliability analysis

We compared subregional hippocampal volumes estimates extracted using FS5.3 and FS6.0 from three independently acquired cohorts: (i) baseline scans of the ADNI-2 cohort (N = 163), (ii) the NESDA cohort (N = 221), and (iii) the MPIP cohort (N = 589). Volume measures for each subregion were bilaterally 'averaged' across the left and right hemispheres.

Volume measurements from FS6.0 are given in mm^3, whereas volume measurements in FS5.3 are returned on the basis of 0.5 mm isotropic. Therefore, the latter set of volume estimates was divided by a factor of 8 in order to transform them to mm^3 measurements.

Volume estimates for the eight sub-regions extracted using FS5.3 were imported into SPSS alongside eight of the twelve possible subregions extracted using FS6.0. Volume estimates for the parasubiculum, molecular layer, GC_ML_DG and HATA (extracted using FS6.0) had no direct corresponding subregions in FS5.3 and were not included in this between-version analysis. We conducted eight sets of two-way mixed reliability analyses, using the same statistical model applied for our prior test–retest comparison (Cronbach's alpha). This produced a series of ICC values measuring the agreement between the old (FS5.3) and new (FS6.0) versions of the FreeSurfer subregion segmentation algorithm.

As a second measure of reproducibility and spatial overlap between FS5.3 and FS6.0, we employed a custom-designed Matlab code to extract a series of Dice similarity coefficients (DSC) for each hippocampal subregion. The DSC, first proposed by Dice (1945), provides a validation metric for evaluating reproducibility and has previously been used to assess spatial overlap between automated MRI reconstructions (Zou et al., 2006). DSC values range from 0 (indicating no spatial overlap between two sets of binary segmentations) to 1 (full overlap between binary segmentations).

DSCs were calculated by dividing the sum of volumes segmented using FS5.3 and volumes segmented using FS6.0 by twice the volume of the intersection between these segmentations; *i.e.*

$$DSC(A, B) = 2(A \cap B)/(A + B)$$

where *A* is the first hippocampal subregion (reconstructed using FS5.3), *B* is the second hippocampal subregion volume (reconstructed using FS6.0) and $\cap$ is the intersected space between the two subregions.

Trans-platform reliability analysis

20 pairs of T1-weighted images were acquired on a 1.5 T and a 3 T scanner system to investigate the stability of both FS5.3 and FS6.0 across platforms. The repositioning between the end of the first acquisition and the start of the second acquisition was performed as fast as possible, usually taking 2–3 min. Both subregional segmentation tools (FS5.3 and FS6.0) were employed on the 2 × 20 images. Subregional volume estimates were imported into SPSS (to extract ICC values) and Matlab (to estimate DSC scores) respectively. All ICC analyses were conducted using the same statistical models previously described for the test–retest analysis.

Heritability of hippocampal subregion volumes

Heritability, defined here as the fraction of the phenotypic variability attributable to genetic variation, was calculated for each hippocampal subregion volume using a variance components model, as implemented in version 7.2.5 of the Sequential Oligogenic Linkage Analysis Routines (SOLAR) software package (http://www.nitrc.org/projects/se_linux) (Almasy and Blangero, 1998). Methods to estimate heritability in SOLAR are detailed elsewhere (Kochunov et al., 2010; Winkler et al., 2010).

Briefly, SOLAR implements a maximum likelihood variance decomposition method, expanding on prior algorithms developed by Amos

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

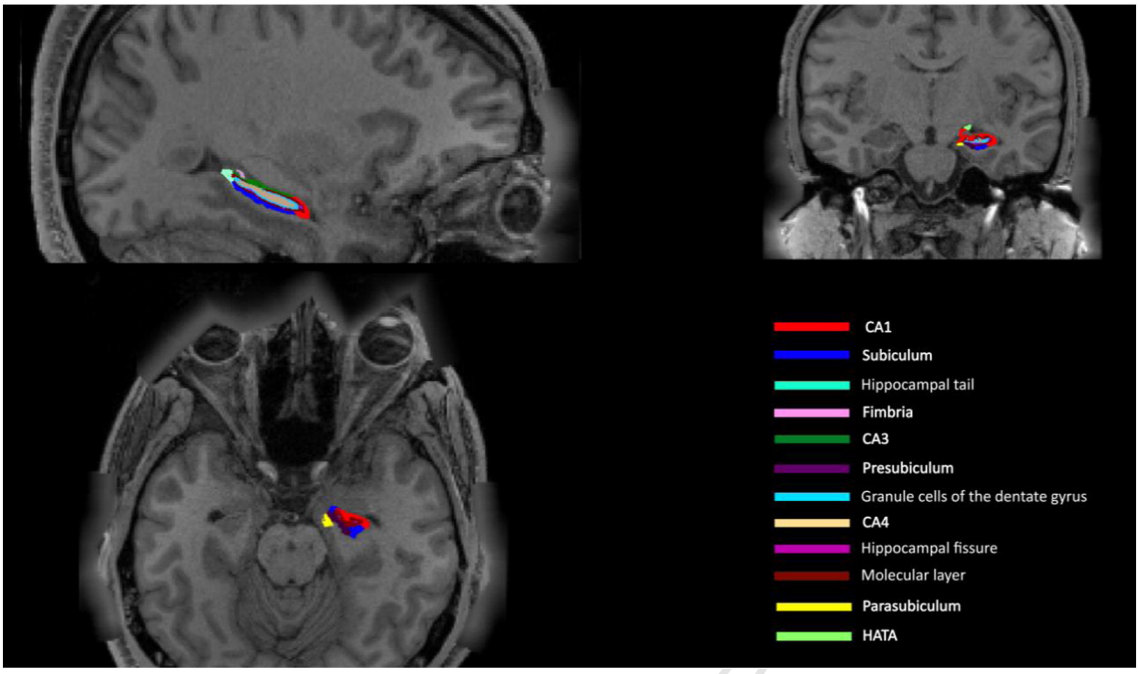

Fig. 1. Color-coded illustration of 11 hippocampal subfields in sagittal (top left), axial (bottom left) and coronal (top right) views. Subfield volumes for each participant were overlaid on their whole-brain T1-weighted image ('nu.mgz') and visually inspected for over- or under-estimation of the hippocampal subfields. In the above rendering, a representative subject from the QTIM cohort was de-identified by blurring around the edges of the skull and face. The image was generated using FreeSurfer's high-resolution visualization tool, *FreeView* (https://surfer.nmr.mgh.harvard.edu/fswiki/FreeviewGuide/).

(1994). The algorithm decomposes phenotypic variance (σ^2_P) into a genetic (σ^2_g) and a residual component (σ^2_e) — the latter represents variation not accounted for by the genetic component (*i.e.*, random environmental variation and/or experimental error). Mean volumes for the whole hippocampus and twelve of its subregions were extracted from all twin pairs in the QTIM sample (N = 132 MZ pairs and N = 232 dizygotic pairs) and reformatted into a phenotype covariance matrix. Each covariate matrix was adjusted to include sex, age, and age * sex interactions as covariates. The covariance matrix, Ω, for each pedigree of individuals was then integrated into the following expression:

$$\Omega = 2\Phi\sigma_g{}^2 + I\sigma_e{}^2$$

where Ω represents covariance between one relative and another, Φ is the pair-wise kinship coefficient representing the relationship between these relatives (0.5 for full siblings), σ^2_g represents the additive genetic component of phenotypic variance, I is the identity matrix and σ^2_e is residual non-genetic variation (*i.e.*, individual-specific environmental variance).

Heritability (h^2) was computed from this model by comparing the observed covariance matrix for phenotypic variance (σ^2_p) with the observed covariance matrix for additive genetic effects (σ^2_g), *i.e.*,

$$h^2 = \sigma^2{}_g/\sigma^2{}_P.$$

Here, h^2 is a value between 0 and 1 representing total additive genetic heritability, ranging from 0 (no genetic contributions) to 1 (all phenotypic variance reflects a genetic effect). Significance of heritability was estimated by computing a model in which $\sigma^2{}_g$ was constrained to zero, computing a second model in which $\sigma^2{}_g$ was estimated, and computing twice the difference between the first and second models' log-likelihoods. For our analysis, we employed a polygenic model that calculated the effects of specific variables (additive genetic variation, and covariates including age, sex and sex * age interactions) in explaining each subregion's volumetric variance within the QTIM population. Three main test statistics were then recorded for each subregion volume: its h^2 estimate, the significance (p-value) of this heritability estimate and its standard error. All test statistics were compared to an adjusted alpha level of $p \leq 3.84 \times 10^{-3}$ to reduce the probability of type 1 errors arising from multiple measurements (N = 13).

Results

Test–retest reliability

Test–retest reliability estimates from ADNI-2, a cohort of 163 healthy, elderly adults scanned three months apart at 3 T, revealed good reliability for all automatically segmented subregion volumes. Larger hippocampal regions (mean volume > 90 mm^3) showed highest ICC values from baseline to follow-up session. These regions included the whole hippocampus (ICC ≥ 0.94), CA1 subregion (ICC ≥ 0.91), CA3 subregion (ICC ≥ 0.88), CA4 subregion (ICC ≥ 0.9), molecular layer (ICC ≥ 0.93), subiculum (ICC ≥ 0.91), presubiculum (ICC ≥ 0.9), granule cells (ICC ≥ 0.91), hippocampal tail (ICC ≥ 0.93), hippocampal fissure (ICC ≥ 0.88) and fimbria (ICC ≥ 0.89). Automated segmentation was also stable for smaller subregions, including the HATA (ICC ≥ 0.78) and parasubiculum (ICC ≥ 0.75) (see Table 2).

Similarly, in the smaller QTIM sub-sample, consisting of 39 young, healthy adults scanned on average three months apart at 4 T, we found strong test–retest reliability for large subregions (mean volume > 90 mm^3). These subregions included the CA1 (ICC ≥ 0.86), CA3 (ICC ≥ 0.78), CA4 (ICC ≥ 0.75), molecular layer (ICC ≥ 0.86), subiculum (ICC ≥ 0.8), granule cells (ICC ≥ 0.78), hippocampal tail (ICC ≥ 0.72), hippocampal fissure (ICC ≥ 0.7) and fimbria (ICC ≥ 0.8), as well as the whole hippocampus (ICC ≥ 0.85). Test–retest reliability of the presubiculum varied considerably from the left (ICC = 0.89) to the right hemisphere (ICC = 0.65). Volume estimates were moderately reproduced for the parasubiculum (ICC ≥ 0.68) and the HATA subregion (ICC ≥ 0.5).

Between-version agreement

In the MPIP cohort (N = 589, 3 T) we found strong agreement between versions 5.3 and 6.0 of the FreeSurfer segmentation algorithm

for the subiculum (0.857). We observed moderate agreement between the following subregions: (i) the hippocampal tail (ICC = 0.778), (ii) the fimbria (ICC = 0.78), (iii) the hippocampal fissure (ICC = 0.78) and (iv) the presubiculum (ICC = 0.797). Agreement between the three major sectors of the *cornu ammonis* (CA1, CA2_3 and CA4) varied considerably; for example, the CA1 (extracted using FS6.0) showed strong agreement with the CA4/Dentate (extracted using FS5.3; ICC = 0.872) and CA2_3 (extracted using FS5.3; ICC = 0.817) but only moderately correlated with its direct counterpart, CA1 (extracted using FS5.3; ICC = 0.645). Similarly, the CA4 subregion extracted using FS6.0 only moderately correlated with the combined CA4-DG from FS5.3 (ICC = 0.66), whereas the CA3 extracted using FS6.0 correlated poorly with its closest counterpart in FS5.3, the CA2_3 (ICC = 0.383) (see Table 3).

The second set of ICCs, examining between-version agreement using volume estimates from the ADNI-2 cohort (N = 163, 3 T), revealed strong agreement between versions 5.3 and 6.0 for (i) the hippocampal tail (ICC = 0.839), (ii) the fimbria (ICC = 0.805), (iii) the presubiculum (ICC = 0.825) and (iv) the subiculum (ICC = 0.833). Between-version agreement was moderate for the hippocampal fissure (ICC = 0.628) and the CA4 (ICC = 0.633). The CA1 subregion (segmented using FS6.0) showed greater correspondence with FS5.3 reconstructions of the CA4_DG (ICC = 0.872) and CA2_3 (ICC = 0.817) than its direct anatomical counterpart, the CA1 (ICC = 0.645). Similarly, the CA3 (showed poor correlation between FS5.3 and FS6.0 (ICC = 0.344), although correlations were higher between the CA3 (extracted using FS6.0) and other subregions from FS5.3, including the CA1 (ICC = 0.523) and CA4_DG (ICC = 0.567) (see Table 4).

The third set of ICCs examined between-version agreement using values extracted from the NESDA cohort (N = 221, 3 T). This analysis revealed strong agreement between FS5.3 and FS6.0 for the subiculum (ICC = 0.815) and moderate agreement for the following subregions: (i) hippocampal tail (ICC = 0.778), (ii) fimbria (ICC = 0.758) and (iii) presubiculum (ICC = 0.783). CA1 volumes extracted using FS6.0 correlated moderately with CA1 volumes extracted using FS5.3 (ICC = 0.698), but correlated more highly with CA4_DG volumes extracted using FS5.3 (ICC = 0.856). Similarly, CA4 volumes extracted using FS6.0 correlated moderately with CA4_DG volumes from FS5.3 (ICC = 0.592), but correlated more highly with CA1 volumes from FS5.3 (ICC = 0.729). Further, the CA3 subregion extracting using FS6.0 correlated poorly with the CA2_3 subregion extracted using FS5.3 (ICC = 0.334), but correlated moderately with the CA1 (0.679) and CA4_DG (0.545). Between-version agreement was poor for the hippocampal fissure (ICC = 0.321) (see Table 5).

A complementary analysis of spatial overlap and reproducibility (as measured by the Dice Similarity Coefficient, DSC) revealed high spatial overlap across the ADNI-2, MPIP and NESDA cohorts for the whole hippocampus (DSC = 0.82–0.85). Between-version agreement was moderate for the hippocampal tail across the three cohorts (DSC = 0.67–0.70). Between-version agreement was poor-to-moderate for the CA4_DG (DSC = 0.49–0.51), fimbria (DSC = 0.45–0.53), presubiculum (DSC = 0.57–0.62) and subiculum (DSC = 0.55–0.58). Between-version agreement was poor for the CA1 (DSC = 0.39–0.4) and the CA2_3 (DSC = 0.28–0.30; see Table 6).

Trans-platform reliability

We conducted two sets of intraclass correlations, testing reliability across two MRI scanner platforms – 1.5 T and 3 T – using (i) FS5.3 and (ii) FS6.0, respectively. The subregion segmentation algorithm provided as part of FS5.3 produced stable volume estimates across scanning platforms for the following regions: (i) the whole hippocampus (ICC = 0.855), (ii) the CA2_3 (ICC = 0.856), (iii) the CA4/dentate (ICC = 0.892), (iv) the presubiculum (ICC = 0.818), (v) the subiculum (ICC = 0.866), (vi) the hippocampal tail (ICC = 0.875), (vii) the CA1 (ICC = 0.725) and (iix) the fimbria (ICC = 0.720). Volume estimates were not reliably reproduced across scanner platforms for the hippocampal fissure (ICC = 0.465) (see Table 7).

The subregion segmentation algorithm provided as part of FS6.0 produced high ICC estimates for the following regions: (i) the whole hippocampus (ICC = 0.942), (ii) the subiculum (ICC = 0.858), (iii) the CA1 (ICC = 0.915), (iv) the presubiculum (ICC = 0.853), (v) the

Table 2
Test–retest intra-class coefficients, dice similarity coefficients and mean volumes for the ADNI-2 and QTIM samples.

Region	Hemi	QTIM 4.0 T Bruker Medscape, N = 39							
		Mean volume (mm^3)	ICC	CI upper	CI lower	Mean volume (mm^3)	ICC	CI upper	CI lower
Whole hippocampus	Left	3494.56	.94	.91	.95	3162.66	.88	.79	.94
	Right	3565.37	.97	.95	.98	3250.55	.85	.73	.92
CA1	Left	653.08	.91	.88	.93	574.11	.86	.75	.92
	Right	676.32	.94	.91	.95	602.92	.89	.80	.94
Molecular layer	Left	572.70	.93	.90	.95	515.64	.86	.75	.92
	Right	593.24	.96	.94	.97	528.06	.88	.78	.93
Hippocampal tail	Left	510.04	.93	.91	.95	496.55	.83	.69	.90
	Right	511.88	.93	.91	.95	523.04	.72	.53	.84
Subiculum	Left	407.45	.91	.88	.93	388.12	.80	.65	.89
	Right	411.34	.94	.92	.96	388.72	.86	.74	.91
Granule cells in the molecular layer of the DG (GC-ML-DG)	Left	315.69	.91	.88	.93	277.84	.78	.62	.88
	Right	326.69	.94	.92	.96	288.63	.78	.62	.88
Presubiculum	Left	297.3	.9	.86	.92	291.6	.89	.80	.94
	Right	291.18	.92	.89	.94	276.73	.65	.42	.80
CA4	Left	271.82	.9	.87	.93	241.75	.75	.58	.86
	Right	283.69	.92	.89	.94	251.44	.77	.60	.87
CA3	Left	227.13	.88	.85	.91	203.38	.78	.62	.88
	Right	239.38	.93	.91	.95	220.03	.82	.68	.90
Hippocampal fissure	Left	159.59	.88	.84	.91	157.05	.80	.64	.88
	Right	162.1	.9	.86	.92	164.92	.70	.51	.84
Fimbria	Left	99.83	.9	.88	.93	56.77	.80	.65	.89
	Right	96.64	.89	.86	.92	52.67	.86	.75	.93
Hippocampal-amygdaloid transition area (HATA)	Left	77.19	.79	.73	.85	56.16	.50	.23	.70
	Right	72.48	.78	.71	.83	56.69	.64	.41	.79
Parasubiculum	Left	62.23	.81	.75	.86	60.69	.74	.55	.85
	Right	62.24	.75	.67	.80	61.06	.68	.46	.81

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

Table 3
Intra-class correlation coefficients for between-version agreement (MPIP cohort, N = 589, 3 T).

Region (bilateral)	Version 5.3 →							
Version 6.0 ↓	Tail	CA1	CA2_3	CA4_DG	Fimbria	Fissure	Presubiculum	Subiculum
Tail	0.778	–	–	–	–	–	–	–
CA1	–	0.645	0.817	0.872	–	–	–	–
CA3	–	0.607	0.383	0.594	–	–	–	–
CA4	–	0.673	0.405	0.661	–	–	–	–
Fimbria	–	–	–	–	0.780	–	–	–
Fissure	–	–	–	–	–	0.716	–	–
Presubiculum	–	–	–	–	–	–	0.797	–
Subiculum	–	–	–	–	–	–	–	0.857

molecular layer (ICC = 0.932), (vi) the granule cells of the dentate gyrus (ICC = 0.932), (vii) the hippocampal tail (ICC = 0.863), (iix) the CA3 (ICC = 0.827), (ix) the HATA (ICC = 0.801), (x) the CA4 (ICC = 0.792) and (xi) the fimbria (ICC = 0.721). Volume estimates were moderately correlated between scanning platforms for the parasubiculum (ICC = 0.659) and the hippocampal fissure (ICC = 0.575) (see Table 7).

Heritability of hippocampal subregion volumes

Fig. 2 shows the proportion of structural variance attributable to genetic factors for the whole hippocampus and its subregions in the QTIM sample. All regions exhibited high heritability, between 0.56 and 0.88. The highest heritability estimates ($h^2 \geq 0.7$) were observed for large regions with mean volumes of 220 mm^3 or greater (*i.e.*, the whole hippocampus, molecular layer, CA1, CA3, CA4, hippocampal tail, granule cell layer, subiculum and presubiculum). Smaller subregions (mean volume: 60–165 mm^3) showed moderate-to-high heritability ($0.55 < h^2 < 0.7$) (see Fig. 2). Table 8 shows the heritability estimates alongside their significance values and standard errors. Using a combination of FreeSurfer subregion labels and TrackVis (http://trackvis.org/), we constructed a three-dimensional visualization of each heritability estimate, this shows how large, posterior subregions (*i.e.*, the hippocampal tail) were most heritable, whereas smaller, anteromedial subregions (parasubiculum, presubiculum and fimbria) were less influenced by genetic factors (see Fig. 3).

Discussion

Here we evaluated a series of automatically segmented volumetric measures from the hippocampus and twelve of its major subregions as reliable, heritable quantitative phenotypes for future large-scale imaging genetics studies. We had four main findings. First, the most recent version of a widely employed FreeSurfer segmentation protocol (FS6.0) showed good test–retest reliability, both at 3 T and 4 T in healthy young and older adults. Spatial overlap between segmentations produced at baseline and follow-up time points was moderate-to-high for all subregions, with the exception of the hippocampal fissure. Second, segmentations produced using FreeSurfer v6.0 showed strong reproducibility from 1.5 T to 3 T field strengths. Third, subregional volume estimates varied between prior and revised versions of the FreeSurfer algorithm, with some subregions (*e.g.* the hippocampal tail) remaining stable, and others (*e.g.* the *cornu ammonis*) diverging notably from one version to the next. Fourth, genetic factors significantly affected the volume of the human hippocampus and its twelve major subregions in a sample of healthy, adult twins. Multi-site genetic analysis may therefore be feasible for automatically extracted subregion measures, building on prior studies that detected common variants associated with overall hippocampal volume (Stein et al., 2012; Hibar et al., 2015).

Table 4
Intra-class correlation coefficients for between-version agreement (ADNI-2 cohort, N = 163, 3 T).

Region (bilateral)	Version 5.3 →							
Version 6.0 ↓	Tail	CA1	CA2_3	CA4_DG	Fimbria	Fissure	Presubiculum	Subiculum
Tail	0.839	–	–	–	–	–	–	–
CA1	–	0.661	0.774	0.901	–	–	–	–
CA3	–	0.523	0.344	0.567	–	–	–	–
CA4	–	0.598	0.372	0.633	–	–	–	–
Fimbria	–	–	–	–	0.805	–	–	–
Fissure	–	–	–	–	–	0.628	–	–
Presubiculum	–	–	–	–	–	–	0.825	–
Subiculum	–	–	–	–	–	–	–	0.833

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

Table 5
Intra-class correlation coefficients for between-version agreement (NESDA cohort, N = 221, 3 T).

Region (bilateral)	Version 5.3 →							
Version 6.0 ↓	Tail	CA1	CA2_3	CA4_DG	Fimbria	Fissure	Presubiculum	Subiculum
Tail	0.778	–	–	–	–	–	–	–
CA1	–	0.698	0.694	0.856	–	–	–	–
CA3	–	0.679	0.334	0.545	–	–	–	–
CA4	–	0.729	0.343	0.592	–	–	–	–
Fimbria	–	–	–	–	0.758	–	–	–
Fissure	–	–	–	–	–	0.321	–	–
Presubiculum	–	–	–	–	–	–	0.783	–
Subiculum	–	–	–	–	–	–	–	0.815

FreeSurfer v6.0: Reliable test–retest segmentations of eleven hippocampal subregions

Automated parcellation algorithms are essential neuroimaging tools, as they facilitate the harmonized, time-efficient and precise reconstruction of brain regions across multiple sites. The automated subcortical segmentation protocol included in the *FreeSurfer* software package has been employed in several important imaging collaborations, leading to the discovery of genetic polymorphisms associated with subcortical and intracranial volumes (Hibar et al., 2015; Ikram et al., 2012; Stein et al., 2012) and the identification of robust subcortical alterations in large populations of people with schizophrenia (Van Erp et al., 2015) and major depressive disorder (Schmaal et al., 2015). FreeSurfer has been validated as a reliable method to reconstruct and measure larger brain regions (Jovicich et al., 2006; Wonderlick et al., 2009), but early versions of its hippocampal subregion segmentation module were criticized by some as anatomically inaccurate, overly reliant on low-resolution images and not yet validated against manual tracing techniques (de Flores et al., 2015; Pluta et al., 2012; Wisse et al., 2014). Here, we found that a revised version of the FreeSurfer subregion segmentation tool, due to be released with FreeSurfer v6.0, produces reliable segmentations for eleven of the twelve hippocampal subregions at 3 T and 4 T field strengths. The most reliably reconstructed subregions included the hippocampal tail, CA1, CA4, presubiculum and subiculum. These subregions showed excellent test–retest reliability in two independent tests (ICC and DSC analysis) and in two unrelated cohorts (ADNI and QTIM).

Other subregions, including the dentate gyrus, CA3, fimbria, HATA and parasubiculum, showed strong test–retest reproducibility at 3 T field strength, but a wider range of test–retest reproducibility at 4 T field strength. This discrepancy may be explained, in part, by the smaller sample size of the 4 T cohort (QTIM; N = 39) compared to the 3 T cohort (ADNI-2; N = 163). ICC estimates extracted from the 4 T cohort were associated with larger confidence intervals (CIs), many of which overlapped with CIs from the 3 T cohort (see Table 2). Voxel size differences between ADNI-2 (1.2 × 1.2 × 1.2 mm) and QTIM (0.94 × 0.98 × 0.98 mm) may have also contributed towards these discrepancies: FreeSurfer resamples MR images to 1 mm isotropic voxel size during its automated reconstruction process and this interpolation procedure may produce variable resolutions in datasets that are 'down-sampled' (*i.e.* ADNI-2) compared to those that are 'up-sampled' (*i.e.* QTIM).

Of the twelve subregions we investigated, only one – the hippocampal fissure – produced unreliable volume estimates between baseline and follow-up acquisitions. The hippocampal fissure is a vestigial sulcus located between the molecular layer of the hippocampus and the dentate gyrus. Several neuroanatomical and methodological variables may contribute to the inconsistent segmentation of this subregion. Its relatively small size and complex cytoarchitectural morphometry may make the subregion more susceptible to partial volume effects caused by changes in the subject's head positioning, variable tissue contrast profiles or even small, undetected changes in the MR signal (Morey et al., 2010). The relatively arbitrary boundary between the fissure and extrahippocampal cerebrospinal fluid (CSF) (Iglesias et al., 2015) may have also contributed towards its poor reproducibility.

Prior appraisals of the FS5.3 segmentation algorithm noted its inconsistent delineations of the hippocampal head and tail (Yushkevich et al., 2010). This new algorithm – FS6.0 – which relies upon a refined atlas built upon high-resolution *ex vivo* MRI data (Iglesias et al., 2015), appears to reconstruct the hippocampal tail and parts of the hippocampal head (CA1, CA2/3) with a high degree of spatial overlap and test–retest reproducibility. Segmentations of the dentate gyrus have also been criticized in FS5.3, as they appear to mismatch with known anatomical boundaries (Wisse et al., 2012), In FS6.0, the dentate is reconstructed as three individual subregions, namely; the hilar region (CA4), the granule cells (GC-DG) and, partially, the molecular layer. Our study showed stable test–retest reliability in all three subregions.

Prior evaluations of the FS5.3 algorithm also noted that the CA1 is the smallest of the three *cornu ammonis* segmentations (CA1, CA2 & CA3), despite *post-mortem* studies contradictorily indicating that the CA1 is the largest and the CA2&3 are the smallest subfields (Wisse et al., 2014). This neuroanatomical inconsistency may yield misleading clinical interpretations: For example, FreeSurfer-based investigations of the human hippocampal subregions have associated neurological

Table 6
DICE coefficients for between-version spatial overlap in the ADNI-2, NESDA and MPIP cohorts.

	Tail	CA1	CA2_3	CA4_DG	Fimbria	Fissure	Presubiculum	Subiculum	Whole
ADNI-2	0.68	0.40	0.30	0.50	0.45	0.30	0.60	0.56	0.85
NESDA	0.67	0.39	0.28	0.51	0.53	0.33	0.57	0.55	0.82
MPIP	0.70	0.40	0.30	0.49	0.51	0.32	0.62	0.58	0.83

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

Table 7
Trans-platform reliability across 1.5 T and 3 T field strengths, using estimates extracted from using FreeSurfer v5.3 and v6.0 (MPIP cohort, N = 10, 3 T).

Region (bilateral)	ICC (FS 5.3)	ICC (FS 6.0)
Whole hippocampus	0.855	0.960
CA1	0.725	0.915
CA2_3	0.856	0.871
CA4_DG	0.892	0.792
Fimbria	0.720	0.721
Fissure	0.465	0.575
Presubiculum	0.818	0.853
Subiculum	0.866	0.858
Tail	0.875	0.863
Parasubiculum	–	0.659
GC-ML-DG	–	0.828
Molecular_layer_HP	–	0.932
HATA	–	0.801

Median cross-platform reliability ICC across values = 0.855 (FreeSurfer 5.3), 0.853 (FreeSurfer 6.0).

Table 8
Heritability estimates for hippocampal subfield volumes, calculated using FreeSurfer v6.0 (QTIM cohort, N = 728, 4 T).

Region	QTIM		
	h^2	Std. error	p-Value
Hippocampal fissure	0.56	0.06	1.90×10^{-14}
Parasubiculum	0.57	0.05	6.16×10^{-17}
Fimbria	0.64	0.05	3.06×10^{-19}
HATA	0.67	0.04	2.76×10^{-24}
CA3	0.75	0.03	4.23×10^{-33}
Subiculum	0.76	0.03	5.02×10^{-32}
CA4	0.79	0.03	1.27×10^{-38}
Presubiculum	0.72	0.04	6.80×10^{-30}
CA1	0.84	0.02	2.54×10^{-47}
Granule cells of DG	0.82	0.03	5.66×10^{-41}
Molecular layer of DG	0.85	0.02	2.56×10^{-49}
Whole hippocampus	0.88	0.01	1.19×10^{-54}
Hippocampal tail	0.84	0.02	3.28×10^{-44}

conditions such as MCI or Alzheimer's disease with atrophy of the CA2&3 (Hanseeuw et al., 2011; Lim et al., 2012), whereas anatomical studies have reported the most profound atrophy in the CA1 (Simic et al., 1997; Rossler et al., 2002). Our findings suggest that this anatomical inconsistency appears to be resolved in FS6.0; the CA1 is now the largest and most reliably reconstructed of the three subfields (see Table 2). Future *in-vivo* investigations of the human hippocampal subregions should therefore prioritize the use of the revised algorithm, FS6.0, as our results show that FS6.0 reliably reproduces eleven major hippocampal subregions across two independent cohorts (QTIM and ADNI-2), despite differences in age, scanning interval and image acquisition method. Clinical findings reported using the algorithm's predecessor, FS5.3, should be interpreted with caution.

Between-version agreement and trans-platform reliability: Implications for imaging consortia

International consortia like ENIGMA typically involve large-scale implementation of harmonized segmentation protocols across diverse networks of research laboratories. Many of these laboratories may have already processed their T1-weighted images through older versions (v5.1–5.3) of the FreeSurfer subregion segmentation tool, raising questions about the need to process their data through a new version of the algorithm. Here, we found strong agreement between older (v5.3) and newer (v6.0) versions of the tool for the hippocampal tail, presubiculum and subiculum. However, versions 5.3 and 6.0 produced variable volume estimates for the *cornu ammonis*, fimbria, and hippocampal fissure. These discrepancies were expected, due to the algorithm's revised definitions of subregional borders (Iglesias et al., 2015). FS6.0 also produced four new subregions with no directly corresponding structures in FS5.3 (the parasubiculum, molecular layer, granule cells of the dentate and HATA). Furthermore, version 6.0 produced slightly more consistent estimates across lower (1.5 T) and higher (3 T) MRI scanner field strengths. Overall, these findings suggest that the latest version of the FreeSurfer subregion segmentation algorithm is a more reliable, versatile and anatomically accurate tool than its predecessors (Iglesias et al., 2015). International consortia such as ENIGMA may benefit by encouraging all participating sites to

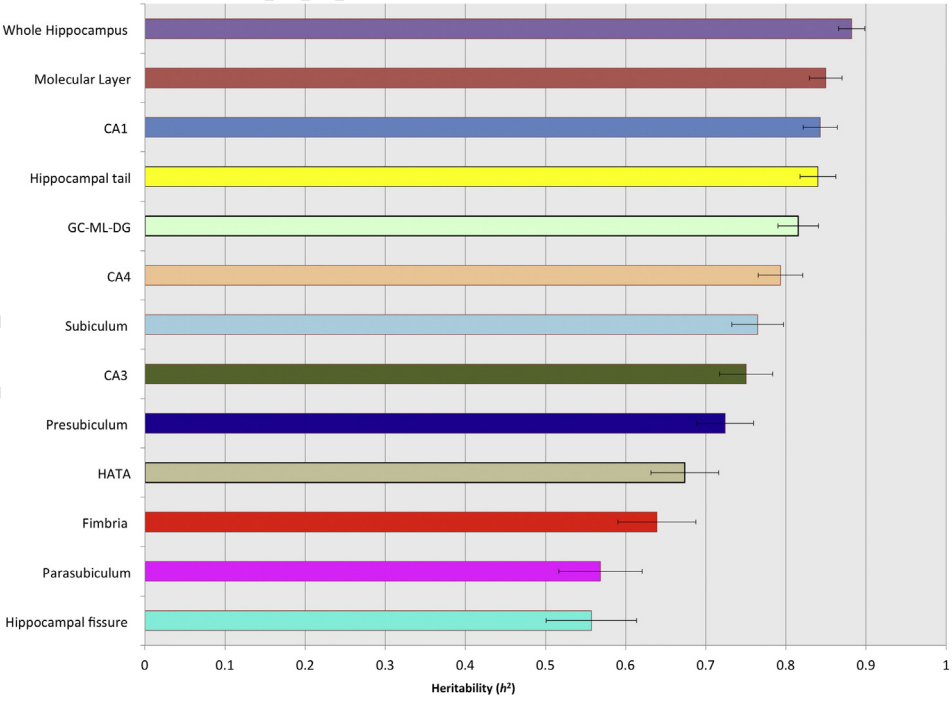

Fig. 2. Heritability of the whole hippocampus and its respective subfields in the QTIM cohort (N = 728).

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

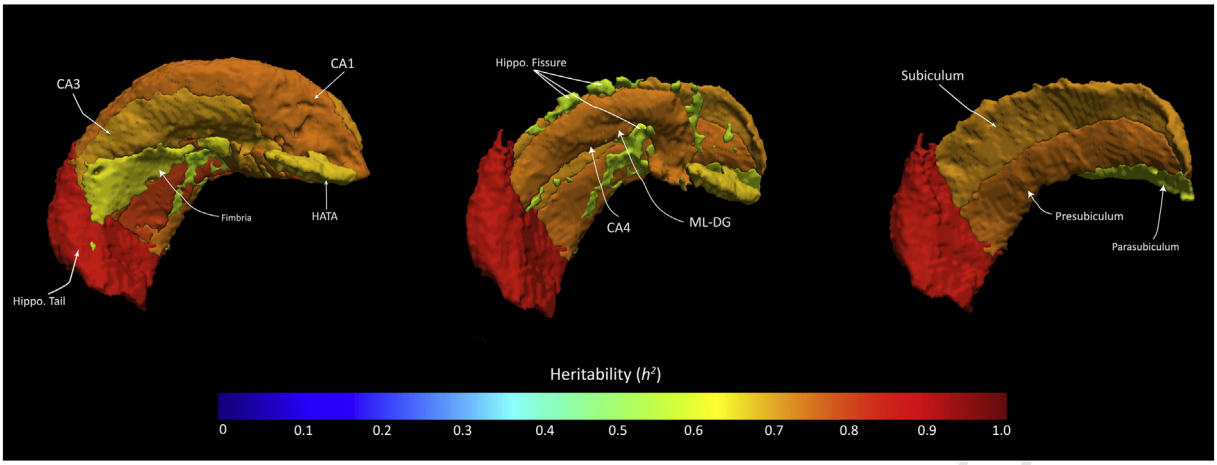

Fig. 3. Three-dimensional visualization of narrow-sense heritability within twelve subfields of the human hippocampal formation, using the average heritability estimates calculated from the QTIM cohort. Heritability is represented as a heat map, with the most heritable subregions depicted in red (see: the hippocampal tail) and moderately heritable subfields colored in green/yellow (see: the hippocampal fissure and parasubiculum). The first image (on the left) is a full reconstruction of the hippocampal formation, showing the most lateral subfields including the CA1, CA3, hippocampal tail ('*hippo. tail*'), fimbria and hippocampal-amygdaloid transition area ('*HATA*'). The middle image removes some lateral substructures, including the fimbria and CA3, in order to display mid-lying subfields including the hippocampal fissure ('*hippo. fissure*'), molecular layer and granule cells of the DG ('*ML-DG*') and CA4. The third image (on the right) further removes these subfields in order to display three remaining medial sub-regions, including the subiculum, presubiculum and parasubiculum. This rendering represents bilateral h^2 estimates, although only the left hippocampus is shown here. Image generated using TrackVis (http://trackvis.org/).

process their imaging data with the revised segmentation tool (FS6.0). The combination of volume estimates acquired using previous (FS5.3) and revised (FS6.0) algorithms is not recommended.

Validating the human hippocampal subfields as quantitative phenotypes for genetic mapping

In the second part of this manuscript, we used SOLAR to calculate the heritability of all twelve automatically segmented hippocampal subregions. The greatest genetic effects were observed in larger subregions, particularly within the granule cells of the DG, molecular layer and the hippocampal tail ($h^2 = 0.74$–0.91). Smaller subregions such as the hippocampal fissure and parasubiculum produced strong but lower heritability estimates ($h^2 = 0.56$–0.57). This pattern of heritability has previously been reported across the wider collection of subcortical structures, with larger regions (such as the thalamus) showing higher heritability than smaller regions (such as the amygdala) (see Hibar et al., 2015). These heritability fluctuations may be explained by the reduced measurement errors associated with larger segmentations. However, biological factors may also play a role. For example, the *cornu ammonis* is among the earliest brain regions to develop prenatally (Taupin, 2007), whereas the subiculum and CA2 are the first hippocampal subregions to mature postnatally (Jabès et al., 2011). The DG and hippocampal tail show accelerated patterns of neurogenesis after the first postnatal year (Insausti et al., 2010). In adult life, hippocampal neurons continue to proliferate from precursor cells in the DG (Kempermann et al., 2004). Given the early development of the CA subregions (Taupin, 2007) and hippocampal tail (Insausti et al., 2010) and the key memory-processing role of the DG in adulthood (Coras et al., 2014), it is likely that genetic factors significantly influence each region. Total hippocampal volume was also significantly heritable ($h^2 = 0.86$–0.88) — supporting prior estimates from healthy populations; this further shows the impact of genetic factors on the structure as a whole (den Braber et al., 2013; van Erp and Saleh, 2004; Swagerman and Brouwer, 2014; Wright et al., 2002).

Our main aim here was to identify reliable quantitative phenotypes that can be used in future collaborative genetic mapping efforts. A biomarker must satisfy several explicit criteria before it can be considered an endophenotype (Gottesman and Gould, 2003). First, it should be associated with illness in the population. Structural changes in the hippocampal subregions are implicated in a wide range of brain disorders, from Alzheimer's disease to epilepsy and schizophrenia (Bartsch, 2012; Sala, 2008). Second, a useful quantitative endophenotype must be heritable. In this study, all major subregions of the hippocampus were highly influenced by additive genetic effects, with heritability estimates ranging from $h^2 = 0.56$ to $h^2 = 0.91$. All subregions, with the exception of the hippocampal fissure (which shows inconsistent volume estimates across image acquisition time points and field strengths), could therefore be considered as reliable and robust quantitative phenotypes for future genetic mapping studies.

Limitations and future directions

In this collaborative investigation, we evaluated a revised version of the FreeSurfer subregion segmentation tool using data collected and analyzed at multiple, independent sites (ADNI-2, QTIM, MPIP and NESDA) at two different field strengths (3 T and 4 T) across large samples of healthy (QTIM, ADNI-2) and affected populations (MPIP, NESDA). We found that the revised algorithm produces heritable and reliable segmentations for eleven human hippocampal subregions, but future users should note some limitations. First, the algorithm has yet to be validated against manual segmentations. A recent quantitative comparison of 21 manual segmentation protocols, including the protocol used to generate manually annotated training data for the revised FreeSurfer algorithm, revealed significant variability among the labels used to define subregions, how boundaries were placed between labels, and the overall extent of the hippocampal formation that is labeled across protocols (Yushkevich et al., 2015). FS6.0 is already a reliable, accessible tool for automated subregion segmentation, but it continues to evolve alongside on-going efforts to harmonize hippocampal subfield protocols (The Hippocampal Subfields Group (HSG), 2014; see hippocampalsubfields.com). As such, it is inevitably subject to revisions as the field develops. Second, although the revised algorithm can segment T1- and T2-weighted images (and their combination; Iglesias et al., 2015), the results presented here are inferred from standard resolution, T1-weighted data only, which is more commonly available across large consortium efforts, such as ENIGMA. Test–retest reliability estimates were extracted using a series of 1.2 mm^3 and ~0.95 mm^3 isotropic images, respectively, possibly introducing measurement errors for smaller subregions like the fimbria (mean volume: 98.24 mm^3), HATA (mean volume: 74.84 mm^3) and parasubiculum (mean volume: 62.23 mm^3) (see Table 2). Future versions of the

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

FreeSurfer segmentation algorithm may yield more robust estimates for low resolution data (<1 mm^3) by combining smaller subfields such as the subiculum and CA2/3. Third, while we observed good reliability between subregion segmentations acquired at 1.5 T and 3 T field strengths, test–retest reproducibility estimates were not established at 1.5 T.

Despite these limitations, the present study supports the utility of eleven automatically segmented hippocampal subregion volumes as quantitative *endophenotypes* for future imaging genetics collaborations. Progressing from macro-level investigations of large brain regions towards more fine-grained maps of specific hippocampal subregions may add more precise localization to GWAS effects. The ENIGMA consortium is now conducting related, finer-grained efforts using diffusion tensor imaging (Jahanshad et al., 2013; Kochunov et al., 2015) and shape analysis (Thompson et al., 2014). Here, we evaluated the automated reconstruction of hippocampal subregion volumes as another useful intermediate biomarker for genome-wide association. As multicenter consortium efforts continue to discover genes associated with brain measures, future quantitative genetic investigations of specific hippocampal subregions may point to a more mechanistic understanding of these genes, and how they affect cognition, behavior and neurological illness.

Conclusion

The hippocampal formation is one of the most profoundly disrupted brain regions in many neurological and psychiatric illnesses. As the present study illustrates, it is now possible to reconstruct eleven major subregions of the hippocampus using almost fully automated brain imaging methods, to a high degree of accuracy and reliability within and across populations. All eleven subregions are highly influenced by genetic factors. As the field of imaging genetics and large-scale imaging consortia continue to successfully identify genes associated with measures from the living human brain, our results may help these initiatives stratify their traits of interest and better understand the mechanisms of gene action.

Disclosures

None of the authors has any conflict of interest to disclose. We have read the Journal's position on issues involved in ethical publication and affirm that our study is consistent with the guidelines.

Uncited references

Q23

Kaymaz, 2009
Kochunov et al., 2014
Miyoshi et al., 2003
Van et al., 2008
Yamasaki et al., 2008

Acknowledgments

This work was supported in part by a Consortium grant (U54 EB020403) from the NIH contributing to the Big Data to Knowledge (BD2K) Initiative, including the NIBIB.

The study also received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 654911 (project "THALAMODEL"). Juan Eugenio Iglesias also acknowledges financial support from the Gipuzkoako Foru Aldundia (Fellows Gipuzkoa Program).

Q24 Q25

Data collection and sharing for this project was funded in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Almasy, L., Blangero, J., 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. 62, 11981211.
Amos, C.I., 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. Am. J. Hum. Genet. 54, 535–543.
Bartsch, T., 2012. The hippocampus in neurological disease. The Clinical Neurobiology of the Hippocampus: An Integrative View, p. 200.
Braskie, M.N., Ringman, J.M., 2011. Neuroimaging measures as endophenotypes in Alzheimer's disease. Int. J. Alzheimers Dis. 2011, 490140.
Burgess, N., Maguire, E., O'Keefe, J., 2002. The human hippocampus and spatial and episodic memory. Neuron 35, 625–641.
Coras, R., Pauli, E., Li, J., Schwarz, M., Rössler, K., Buchfelder, M., Hamer, H., Stefan, H., Blumcke, I., 2014. Differential influence of hippocampal subfields to memory formation: insights from patients with temporal lobe epilepsy. Brain J. Neurol. 137, 1945–1957.
Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 291–334.
Dale, A., Fischl, B., Sereno, M., 1999. Cortical surface-based analysis I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.
Das, S., Mechanic-Hamilton, D., Pluta, J., Korczykowski, M., Detre, J., Yushkevich, P., 2011. Heterogeneity of functional activation during memory encoding across hippocampal subfields in temporal lobe epilepsy. NeuroImage 58, 1121–1130.
de Flores, R., de Joie, R., Landeau, B., Perrotin, A., Mézenge, F., Sayette, V., de Eustache, F., Desgranges, B., Chételat, G., 2015. Effects of age and Alzheimer's disease on hippocampal subfields. Hum. Brain Mapp. 36, 463–474.
den Braber, A., Bohlken, M.M., Brouwer, R.M., Ent, van't, D., 2013. Heritability of subcortical brain measures: a perspective for future genome-wide association studies. NeuroImage 83, 98–102.
DeStefano, A.L., Seshadri, S., Beiser, A., 2009. Bivariate heritability of total and regional brain volumes: the Framingham Study. Alzheimer Dis Assoc Disord. 23, 218–223.
Dice, L.R., 1945. Measures of the amount of ecological association between species. Ecology 26, 297–302.
Fischl, B., 2012. FreeSurfer. NeuroImage 62, 774–781.
Fischl, B., Sereno, M., Dale, A., 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage 9, 195–207.
Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Fischl, B., Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A., 2004a. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.
Fischl, B., Salat, D., Kouwe, A., Makris, N., Ségonne, F., Quinn, B., Dale, A., 2004b. Sequence-independent segmentation of magnetic resonance images. NeuroImage 23 (Suppl. 1), S69–S84.
Frodl, T., Jäger, M., Smajstrlova, I., Born, C., Bottlender, R., Palladino, T., Reiser, M., Möller, H.-J., Meisenzahl, E.M., 2008. Effect of hippocampal and amygdala volumes on clinical outcomes in major depression: a 3-year prospective magnetic resonance imaging study. J. Psychiatry Neurosci.: JPN 33, 423.
Glahn, D.C., Thompson, P.M., Blangero, J., 2007. Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. Hum. Brain Mapp. 28, 488–501.
Gottesman, I., Gould, T., 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatr. 160, 636–645.
Hanseeuw, B.J., Van, L.K., Kavec, M., Grandin, C., Seron, X., Ivanoiu, A., 2011. Mild cognitive impairment: differential atrophy in the hippocampal subfields. AJNR Am. J. Neuroradiol. 32, 1658–1661.
Hasler, G., Northoff, G., 2011. Discovering imaging endophenotypes for major depression. Mol. Psychiatry 16, 604–619.

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

Hedges, D.W., Woon, F.L., 2010. Alcohol use and hippocampal volume deficits in adults with posttraumatic stress disorder: a meta-analysis. Biol. Psychol. 84 (2), 163–168.

Hibar, D., Stein, J., Renteria, M., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., Toro, R., Wittfeld, K., Abramovic, L., Andersson, M., Aribisala, B., Armstrong, N., Bernard, M., Bohlken, M., Boks, M., Bralten, J., Brown, A., Chakravarty, M., Chen, Q., Ching, C., Cuellar-Partida, G., Braber, A., Giddaluru, S., Goldman, A., Grimm, O., Guadalupe, T., Hass, J., Woldehawariat, G., Holmes, A., Hoogman, M., Janowitz, D., Jia, T., Kim, S., Klein, M., Kraemer, B., Lee, P., Loohuis, L., Luciano, M., Macare, C., Mather, K., Mattheisen, M., Milaneschi, Y., Nho, K., Papmeyer, M., Ramasamy, A., Risacher, S., Roiz-Santiañez, R., Rose, E., Salami, A., Sämann, P., Schmaal, L., Schork, A., Shin, J., Strike, L., Teumer, A., Donkelaar, M., Eijk, K., Walters, R., Westlye, L., Whelan, C., Winkler, A., Zwiers, M., Alhusaini, S., Athanasiu, L., Ehrlich, S., Hakobjan, M., Hartberg, C., Haukvik, U., Heister, A., Hoehn, D., Kasperaviciute, D., Liewald, D., Lopez, L., Makkinje, R., Matarin, M., Naber, M., McKay, D., Needham, M., Nugent, A., Pütz, B., Royle, N., Shen, L., Sprooten, E., Trabzuni, D., Marel, S., Hulzen, K., Walton, E., Wolf, C., Almasy, L., Ames, D., Arepalli, S., Assareh, A., Bastin, M., Brodaty, H., Bulayeva, K., Carless, M., Cichon, S., Corvin, A., Curran, J., Czisch, M., Zubicaray, G., Dillman, A., Duggirala, R., Dyer, T., Erk, S., Fedko, I., Ferrucci, L., Foroud, T., Fox, P., Fukunaga, M., Gibbs, J., Göring, H., Green, R., Guelfi, S., Hansell, N., Hartman, C., Hegenscheid, K., Heinz, A., Hernandez, D., Heslenfeld, D., Hoekstra, P., Holsboer, F., Homuth, G., Hottenga, J.-J., Ikeda, M., Jack Jr., C.R., Jenkinson, M., Johnson, R., Kanai, R., Keil, M., Kent Jr., J.W., Kochunov, P., Kwok, J., Lawrie, S., Liu, X., Longo, D., McMahon, K., Meisenzahl, E., Melle, I., Mohnke, S., Montgomery, G., Mostert, J., Mühleisen, T., Nalls, M., Nichols, T., Nilsson, L., Nöthen, M., Ohi, K., Olvera, R., Perez-Iglesias, R., Pike, G., Potkin, S., Reinvang, I., Reppermund, S., Rietschel, M., Romanczuk-Seiferth, N., Rosen, G., Rujescu, D., Schnell, K., Schofield, P., Smith, C., Steen, V., Sussmann, J., Thalamuthu, A., Toga, A., Traynor, B., Troncoso, J., Turner, J., Hernández, M., Ent, D., Brug, M., Wee, N., Tol, M.-J., Veltman, D., Wassink, T., Westman, E., Zielke, R., Zonderman, A., Ashbrook, D., Hager, R., Lu, L., McMahon, F., Morris, D., Williams, R., Brunner, H., Buckner, R., Buitelaar, J., Cahn, W., Calhoun, V., Cavalleri, G., Crespo-Facorro, B., Dale, A., Davies, G., Delanty, N., Depondt, C., Djurovic, S., Drevets, W., Espeseth, T., Gollub, R., Ho, B.-C., Hoffmann, W., Hosten, N., Kahn, R., Hellard, S., Meyer-Lindenberg, A., Müller-Myhsok, B., Nauck, M., Nyberg, L., Pandolfo, M., Penninx, B., Roffman, J., Sisodiya, S., Smoller, J., Bokhoven, H., Haren, N., Völzke, H., Walter, H., Weiner, M., Wen, W., White, T., Agartz, I., Andreassen, O., Blangero, J., Boomsma, D., Brouwer, R., Cannon, D., Cookson, M., Geus, E., Deary, I., Donohoe, G., Fernández, G., Fisher, S., Francks, C., Glahn, D., Grabe, H., Gruber, O., Hardy, J., Hashimoto, R., Pol, H., Jönsson, E., Kloszewska, I., Lovestone, S., Mattay, V., Mecocci, P., McDonald, C., McIntosh, A., Ophoff, R., Paus, T., Pausova, Z., Ryten, M., Sachdev, P., Saykin, A., Simmons, A., Singleton, A., Soininen, H., Wardlaw, J., Weale, M., Weinberger, D., Adams, H., Launer, L., Seiler, S., Schmidt, R., Chauhan, G., Satizabal, C., Becker, J., Yanek, L., Lee, S., Ebling, M., Fischl, B., Longstreth Jr., W.T., Greve, D., Schmidt, H., Nyquist, P., Vinke, L., Duijn, C., Xue, L., Mazoyer, B., Bis, J., Gudnason, V., Seshadri, S., Ikram, M., Initiative, T., Consortium, T., EPIGEN, E., IMAGEN, I., SYS, S., Martin, N., Wright, M., Schumann, G., Franke, B., Thompson, P., Medland, S., 2015. Common genetic variants influence human subcortical brain structures. Nature 520, 224–229.

Hitti, F., Siegelbaum, S., 2014. The hippocampal CA2 region is essential for social memory. Nature 508, 88–92.

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., for the Alzheimer's Disease Neuroimaging Initiative, 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.

Ikram, M., Fornage, M., Smith, A., Seshadri, S., Schmidt, R., Debette, S., Vrooman, H., Sigurdsson, S., Ropele, S., Taal, H., Mook-Kanamori, D., Coker, L., Longstreth, W., Niessen, W., DeStefano, A., Beiser, A., Zijdenbos, A., Struchalin, M., Jack, C., Rivadeneira, F., Uitterlinden, A., Knopman, D., Hartikainen, A.-L., Pennell, C., Thiering, E., Steegers, E., Hakonarson, H., Heinrich, J., Palmer, L., Jarvelin, M.-R., McCarthy, M., Grant, S., Pourcain, B., Timpson, N., Smith, G., Sovio, U., Consortium, E., Nalls, M., Au, R., Hofman, A., Gudnason, H., Lugt, A., Harris, T., Meeks, W., Vernooij, M., Buchem, M., Catellier, D., Jaddoe, V., Gudnason, V., Windham, B., Wolf, P., Duijn, C., Mosley, T., Schmidt, H., Launer, L., Breteler, M., DeCarli, C., Consortium, C., 2012. Common variants at 6q22 and 17q21 are associated with intracranial volume. Nat. Genet. 44, 539–544.

Insausti, R., Amaral, D.G., 2004. Hippocampal formation. In: Paxinos, G., Mai, J.K. (Eds.), The Human Nervous System. Elsevier, Amsterdam, pp. 871–906.

Insausti, R., Cebada-Sánchez, S., Marcos, P., 2010. Postnatal Development of the Human Hippocampal Formation, Advances in Anatomy. Springer, Embryology and Cell Biology.

Jabès, A., Lavenex, P., Amaral, D., Lavenex, P., 2011. Postnatal development of the hippocampal formation: a stereological study in macaque monkeys. J. Comp. Neurol. 519, 1051–1070.

Jack, C.R., Bernstein, M.A., Borowski, B.J., Gunter, J.L., Fox, N.C., Thompson, P.M., Schuff, N., Krueger, G., Killiany, R.J., Decarli, C.S., Dale, A.M., Carmichael, O.W., Tosun, D., Weiner, M.W., 2010. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. Alzheimers Dement. 6, 212–220.

Jahanshad, N., Kochunov, P., Sprooten, E., Mandl, R., Nichols, T., Almassy, L., Blangero, J., Brouwer, R., Curran, J., Zubicaray, G., Duggirala, R., Fox, P., Hong, L., Landman, B., Martin, N., McMahon, K., Medland, S., Mitchell, B., Olvera, R., Peterson, C., Starr, J., Sussmann, J., Toga, A., Wardlaw, J., Wright, M., Pol, H., Bastin, M., McIntosh, A., Deary, I., Thompson, P., Glahn, D., 2013. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA DTI working group. NeuroImage 81, 455–469.

Jovicich, J., Czanner, S., Greve, D., Haley, E., Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage 30, 436443.

Kaymaz, N., Os, V.J., 2009. Heritability of structural brain traits: an endophenotype approach to deconstruct schizophrenia. Int. Rev. Neurobiol. 89, 85–130.

Kempermann, G., Jessberger, S., Steiner, B., Kronenberg, G., 2004. Milestones of neuronal development in the adult hippocampus. Trends Neurosci. 27, 447–452.

Kochunov, P., Glahn, D.C., Fox, P.T., Lancaster, J.L., 2010. Genetics of primary cerebral gyrification: heritability of length, depth and area of primary sulci in an extended pedigree of Papio baboons. NeuroImage 53, 1126–1134.

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T., Mandl, R., Almasy, L., Booth, T., Brouwer, R., Curran, J., Zubicaray, G., Dimitrova, R., Duggirala, R., Fox, P., Hong, L., Landman, B., Lemaître, H., Lopez, L., Martin, N., McMahon, K., Mitchell, B., Olvera, R., Peterson, C., Starr, J., Sussmann, J., Toga, A., Wardlaw, J., Wright, M., Wright, S., Bastin, M., McIntosh, A., Boomsma, D., Kahn, R., Braber, A., Geus, E., Deary, I., Pol, H., Williamson, D., Blangero, J., Ent, D., Thompson, P., Glahn, D., 2014. Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and megaanalytical approaches for data pooling. NeuroImage 95, 136–150.

Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T., Wright, S., Hong, L., Patel, B., Behrens, T., Jbabdi, S., Andersson, J., Lenglet, C., Yacoub, E., Moeller, S., Auerbach, E., Ugurbil, K., Sotiropoulos, S., Brouwer, R., Landman, B., Lemaitre, H., Braber, A., Zwiers, M., Ritchie, S., Hulzen, K., Almasy, L., Curran, J., deZubicaray, G., Duggirala, R., Fox, P., Martin, N., McMahon, K., Mitchell, B., Olvera, R., Peterson, C., Starr, J., Sussmann, J., Wardlaw, J., Wright, M., Boomsma, D., Kahn, R., Geus, E., Williamson, D., Hariri, A., Ent, D., Bastin, M., McIntosh, A., Deary, I., Pol, H., Blangero, J., Thompson, P., Glahn, D., Essen, D., 2015. Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. NeuroImage 111, 300–311.

La Joie, R., Fouquet, M., Mézenge, F., Landeau, B., Villain, N., Mevel, K., Pélerin, A., Eustache, F., Desgranges, B., Chételat, G., 2010. Differential effect of age on hippocampal subfields assessed using a new high-resolution 3 T MR sequence. NeuroImage 53, 506–514.

Lim, H.K., Hong, S.C., Jung, W.S., Ahn, K.J., Won, W.Y., Hahn, C., et al., 2012. Automated segmentation of hippocampal subfields in drug-naïve patients with Alzheimer's Disease. AJNR Am. J. Neuroradiol. 34, 747–751.

Mather, K.A., Armstrong, N.J., Wen, W., Kwok, J.B., 2015. Investigating the genetics of hippocampal volume in older adults without dementia. PLoS One 10, e0116920.

Miyoshi, K., Honda, A., Baba, K., Taniguchi, M., Oono, K., Fujita, T., Kuroda, S., Katayama, T., Tohyama, M., 2003. Disrupted-In-Schizophrenia 1, a candidate gene for schizophrenia, participates in neurite outgrowth. Mol. Psychiatry 8, 685–694.

Morey, R.A., Selgrade, E.S., Wagner, H.R., Huettel, S.A., Wang, L., McCarthy, G., 2010. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Hum. Brain Mapp. 31, 1751–1762.

Morris, R., 2006. Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas. Eur. J. Neurosci. 23, 2829–2846.

Mueller, S., Stables, L., Du, A., Schuff, N., Truran, D., Cashdollar, N., Weiner, M., 2007. Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4 T. Neurobiol. Aging 28, 719–726.

Navratilova, Z., Battaglia, F.P., 2015. CA2: it's about time—and episodes. Neuron 85, 8–10.

Newmark, R., Schon, K., Ross, R., Stern, C., 2013. Contributions of the hippocampal subfields and entorhinal cortex to disambiguation during working memory. Hippocampus 23, 467–475.

O'Keefe, J., 1990. A computational theory of the hippocampal cognitive map. Prog. Brain Res. 83, 301–312.

O'Mara, S., 2006. Controlling hippocampal output: the central role of subiculum in hippocampal information processing. Behav. Brain Res. 174, 304–312.

Panizzon, M.S., Hauger, R.L., Eaves, L.J., Chen, C.H., 2012. Genetic influences on hippocampal volume differ as a function of testosterone level in middle-aged men. NeuroImage 59, 1123–1131.

Pipitone, J., Park, M.T.M., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos, A.N., Chakravarty, M.M., 2014. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. NeuroImage 101, 494–512.

Pluta, J., Yushkevich, P., Das, S., Wolk, D., 2012. In vivo analysis of hippocampal subfield atrophy in mild cognitive impairment via semi-automatic segmentation of T2-weighted MRI. J. Alzheimers Dis.: JAD 31, 85–99.

Roalf, D.R., Vandekar, S.N., Almasy, L., Ruparel, K., 2015. Heritability of subcortical and limbic brain volume and shape in multiplex–multigenerational families with schizophrenia. Biol. Psychiatry 77, 137–146.

Rosene, D., Hoesen, G., 1987. The hippocampal formation of the primate brain, a review of some comparative aspects of cytoarchitecture and connections. Cereb. Cortex 6, 345–456.

Rossler, M., Zarski, R., Bohl, J., Ohm, T.G., 2002. Stage-dependent and sector-specific neuronal loss in hippocampus during Alzheimer's disease. *Acta Neuropathol.* 103, 369-369.

Sala, M., 2008. Stress and hippocampal abnormalities in psychiatric disorders. Eur. Neuropsychopharmacol. 14, 393–405.

Sämann, P., Höhn, D., Chechko, N., Kloiber, S., Lucae, S., Ising, M., Holsboer, F., Czisch, M., 2013. Prediction of antidepressant treatment response from gray matter volume across diagnostic categories. Eur. Neuropsychopharmacol. 23, 1503–1515.

Schmaal, L., Veltman, D.J., van Erp, T.G.M., Sämann, P.G., Frodl, T., Jahanshad, N., Loeher, E., Tiemeier, H., Hofman, A., Niessen, W.J., Vernooij, M.W., Ikram, M.A., Wittfield, K., Grabe, H.J., Block, A., Hegenscheid, K., Völzke, H., Hoehn, D., Czisch, M., Lagopoulos, J., Hatton, S.N., Hickie, I.B., Goya-Maldonado, R., Krämer, B., Gruber, O., Couvy-Duchesne, B., Rentería, M.E., Strike, L.T., Mills, N.T., de Zubicaray, G.I., McMahon, K.L., Medland, S.E., Martin, N.G., Gillespie, N.A., Wright, M.J., Hall, G.B., MacQueen, G.M., Frey, E.M., Carballedo, A., van Velzen, L.S., van Tol, M.J., van der Wee, N.J., Veer, I.M., Walter, H., Schnell, K., Schramm, E., Normann, C., Schoepf, D., Konrad, C., Zurowski, B., Nickson, T., McIntosh, A.M., Papmeyer, M., Whalley, H.C., Sussman, J.E., Godlewska, B.R., Cowen, P.J., Fischer, F.H., Rose, M., Penninx, B.W.J.H., Thompson, P.M., Hibar, D.P. for the ENIGMA-Major Depressive Disorder Working Group, 2015. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. Mol. Psychiatry (in press).

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039

Simic, G., Kostovic, I., Winblad, B., Bogdanovic, N., 1997. Volume and number of neurons of the human hippocampal formation in normal aging and Alzheimer's disease. J. Comp. Neurol. 379, 482–494.

Stein, J., Medland, S., Vasquez, A., Hibar, D., Senstad, R., Winkler, A., Toro, R., Appel, K., Bartecek, R., Bergmann, Ø., Bernard, M., Brown, A., Cannon, D., Chakravarty, M., Christoforou, A., Domin, M., Grimm, O., Hollinshead, M., Holmes, A., Homuth, G., Hottenga, J.-J., Langan, C., Lopez, L., Hansell, N., Hwang, K., Kim, S., Laje, G., Lee, P., Liu, X., Loth, E., Lourdusamy, A., Mattingsdal, M., Mohnke, S., Maniega, S., Nho, K., Nugent, A., O'Brien, C., Papmeyer, M., Pütz, B., Ramasamy, A., Rasmussen, J., Rijpkema, M., Risacher, S., Roddey, J., Rose, E., Ryten, M., Shen, L., Sprooten, E., Strengman, E., Teumer, A., Trabzuni, D., Turner, J., Eijk, K., Erp, T., Tol, M.-J., Wittfeld, K., Wolf, C., Woudstra, S., Aleman, A., Alhusaini, S., Almasy, L., Binder, E., Brohawn, D., Cantor, R., Carless, M., Corvin, A., Czisch, M., Curran, J., Davies, G., Almeida, M., Delanty, N., Depondt, C., Duggirala, R., Dyer, T., Erk, S., Fagerness, J., Fox, P., Freimer, N., Gill, M., Göring, H., Hagler, D., Hoehn, D., Holsboer, F., Hoogman, M., Hosten, N., Jahanshad, N., Johnson, M., Kasperaviciute, D., Kent, J., Kochunov, P., Lancaster, J., Lawrie, S., Mandl, R., Matarin, M., Mattheisen, M., Meisenzahl, E., Melle, I., Moses, E., Mühleisen, T., Nauck, M., Nöthen, M., Olvera, R., Pandolfo, M., Pike, G., Puls, R., Reinvang, I., Rentería, M., Rietschel, M., Roffman, J., Royle, N., Rujescu, D., Savitz, J., Schnack, H., Schnell, K., Seiferth, N., Smith, C., Steen, V., Hernández, M., Heuvel, M., Wee, N., Haren, N., Veltman, J., Völzke, H., Walker, R., Westlye, L., Whelan, C., Agartz, I., Boomsma, D., Cavalleri, G., Dale, A., Djurovic, S., Drevets, W., Hagoort, P., Hall, J., Heinz, A., Jack, C., Foroud, T., Hellard, S., Macciardi, F., Montgomery, G., Poline, J., Porteous, D., Sisodiya, S., Starr, J., Sussmann, J., Toga, A., Veltman, D., Walter, H., Weiner, M., Bis, J., Ikram, M., Smith, A., Gudnason, V., Tzourio, C., Vernooij, M., Launer, L., DeCarli, C., Seshadri, S., 2012. Identification of common variants associated with human hippocampal and intracranial volumes. Nat. Genet. 44, 552–561.

Sullivan, E.V., Pfefferbaum, A., Swan, G.E., 2001. Heritability of hippocampal size in elderly twin men: equivalent influence from genes and environment. Hippocampus 11, 754–762.

Swagerman, S.C., Brouwer, R.M., 2014. Development and heritability of subcortical brain volumes at ages 9 and 12. Genes Brain Behav. 13 (8), 733–742.

Taal, H., Pourcain, B., Thiering, E., Das, S., Mook-Kanamori, D., Warrington, N., Kaakinen, M., Kreiner-Møller, E., Bradfield, J., Freathy, R., Geller, F., Guxens, M., Cousminer, D., Kerkhof, M., Timpson, N., Ikram, M., Beilin, L., Bønnelykke, K., Buxton, J., Charoen, P., Chawes, B., Eriksson, J., Evans, D., Hofman, A., Kemp, J., Kim, C., Klopp, N., Lahti, J., Lye, S., McMahon, G., Mentch, F., Müller-Nurasyid, M., O'Reilly, P., Prokopenko, I., Rivadeneira, F., Steegers, E., Sunyer, J., Tiesler, C., Yaghootkar, H., Consortium, C., Breteler, M., Decarli, C., Breteler, M., Debette, S., Fornage, M., Gudnason, V., Launer, L., Lugt, A., Mosley, T., Seshadri, S., Smith, A., Vernooij, M., Consortium, E., Blakemore, A., Chiavacci, R., Feenstra, B., Fernandez-Banet, J., Grant, S., Hartikainen, A.-L., Heijden, A., Iñiguez, C., Lathrop, M., McArdle, W., Mølgaard, A., Newnham, J., Palmer, L., Palotie, A., Pouta, A., Ring, S., Sovio, U., Standl, M., Uitterlinden, A., Wichmann, H.-E., Vissing, N., DeCarli, C., Duijn, C., McCarthy, M., Koppelman, G., Estivill, X., Hattersley, A., Melbye, M., Bisgaard, H., Pennell, C., Widen, E., Hakonarson, H., Smith, G., Heinrich, J., Jarvelin, M.-R., Jaddoe, V., Consortium, E., 2012. Common variants at 12q15 and 12q24 are associated with infant head circumference. Nat. Genet. 44, 532–538.

Taupin, P., 2007. The hippocampus: neurotransmission and plasticity in the nervous system. Nova Biomedical Books.

The Hippocampal Subfields Group (HSG), 2014. Towards a harmonized protocol for segmentation of hippocampal subfields and parahippocampal cortical subregions in *in vivo* MRI: a white paper. HS3.3 Meeting, Irvine, CA (http://www.hippocampalsubfields.com/whitepaper/.

Thompson, C.L., Pathak, S.D., Jeromin, A., Ng, L.L., 2008. Genomic anatomy of the hippocampus. Neuron 60, 1010–1021.

Thompson, P., Stein, J., Medland, S., Hibar, D., Vasquez, A., Rentería, M., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M., Martin, N., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N., Andreassen, O., Apostolova, L., Appel, K., Armstrong, N., Aribisala, B., Bastin, M., Bauer, M., Bearden, C., Bergmann, Ø., Binder, E., Blangero, J., Bockholt, H., Bøen, E., Bois, C., Boomsma, D., Booth, T., Bowman, I., Bralten, J., Brouwer, R., Brunner, H., Brohawn, D., Buckner, R., Buitelaar, J., Bulayeva, K., Bustillo, J., Calhoun, V., Cannon, D., Cantor, R., Carless, M., Caseras, X., Cavalleri, G., Chakravarty, M., Chang, K., Ching, C., Christoforou, A., Cichon, S., Clark, V., Conrod, P., Coppola, G., Crespo-Facorro, B., Curran, J., Czisch, M., Deary, I., Geus, E., Braber, A., Delvecchio, G., Depondt, C., Haan, L., Zubicaray, G., Dima, D., Dimitrova, R., Djurovic, S., Dong, H., Donohoe, G., Duggirala, R., Dyer, T., Ehrlich, S., Ekman, C., Elvsåshagen, T., Emsell, L., Erk, S., Espeseth, T., Fagerness, J., Fears, S., Fedko, I., Fernández, G., Fisher, S., Foroud, T., Fox, P., Francks, C., Frangou, S., Frey, E., Frodl, T., Frouin, V., Garavan, H., Giddaluru, S., Glahn, D., Godlewska, B., Goldstein, R., Gollub, R., Grabe, H., Grimm, O., Gruber, O., Guadalupe, T., Gur, R., Gur, R., Göring, H., Hagenaars, S., Hajek, T., Hall, G., Hall, J., Hardy, J., Hartman, C., Hass, J., Hatton, S., Haukvik, U., Hegenscheid, K., Heinz, A., Hickie, I., Ho, B.-C., Hoehn, D., Hoekstra, P., Hollinshead, M., Holmes, A., Homuth, G., Hoogman, M., Hong, L., Hosten, N., Hottenga, J.-J., Pol, H., Hwang, K., Jack Jr., C.R., Jenkinson, M., Johnston, C., Jönsson, E., Kahn, R., Kasperaviciute, D., Kelly, S., Kim, S., Kochunov, P., Koenders, L., Krämer, B., Kwok, J., Lagopoulos, J., Laje, G., Landen, M., Landman, B., Lauriello, J., Lawrie, S., Lee, P., Hellard, S., Lemaître, H., Leonardo, C., Li, C., Liberg, B., Liewald, D., Liu, X., Lopez, L., Loth, E., Lourdusamy, A., Luciano, M., Macciardi, F., Machielsen, M., MacQueen, G., Malt, U., Mandl, R., Manoach, D., Martinot, J.-L., Matarin, M., Mather, K., Mattheisen, M., Mattingsdal, M., Meyer-Lindenberg, A., McDonald, C., McIntosh, A., McMahon, F., McMahon, K., Meisenzahl, E., Melle, I., Milaneschi, Y., Mohnke, S., Montgomery, G., Morris, D., Moses, E., Mueller, B., Maniega, S., Mühleisen, T., Müller-Myhsok, B., Mwangi, B., Nauck, M., Nho, K., Nichols, T., Nilsson, L.-G., Nugent, A., Nyberg, L., Olvera, R., Oosterlaan, J., Ophoff, R., Pandolfo, M., Papalampropoulou-Tsiridou, M., Papmeyer, M., Paus, T., Pausova, Z., Pearlson, G., Penninx, B., Peterson, C., Pfennig, A., Phillips, M., Pike, G., Poline, J., Potkin, S., Pütz, B., Ramasamy, A., Rasmussen, J., Rietschel, M., Rijpkema, M., Risacher, S., Roffman, J., Roiz-Santiañez, R., Romanczuk-Seiferth, N., Rose, E., Royle, N., Rujescu, D., Ryten, M., Sachdev, P., Salami, A., Satterthwaite, T., Savitz, J., Saykin, A., Scanlon, C., Schmaal, L., Schnack, H., Schork, A., Schulz, S., Schür, R., Seidman, L., Shen, L., Shoemaker, J., Simmons, A., Sisodiya, S., Smith, C., Smoller, J., Soares, J., Sponheim, S., Sprooten, E., Starr, J., Steen, V., Strakowski, S., Strike, L., Sussmann, J., Sämann, P., Teumer, A., Toga, A., Tordesillas-Gutierrez, D., Trabzuni, D., Trost, S., Turner, J., Heuvel, M., Wee, N., Eijk, K., Erp, T., Haren, N., Ent, D., Tol, M.-J., Hernández, M., Veltman, D., Versace, A., Völzke, H., Walker, R., Walter, H., Wang, L., Wardlaw, J., Weale, M., Weiner, M., Wen, W., Westlye, L., Whalley, H., Whelan, C., White, T., Winkler, A., Wittfeld, K., Woldehawariat, G., Wolf, C., Zilles, D., Zwiers, M., Thalamuthu, A., Schofield, P., Freimer, N., Lawrence, N., Drevets, W., 2014. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav. 8, 153–182.

van Erp, T.G., Saleh, P.A., Huttunen, M., Lönnqvist, J., Kaprio, J., Salonen, O., Valanne, L., Poutanen, V.P., Standertskjöld-Nordenstam, C.G., Cannon, T.D., 2004. Hippocampal volumes in schizophrenic twins. Arch. Gen. Psychiatry 61, 346–353.

van Erp, T.G., Hibar, D.P., Rasmussen, J.M., Glahn, D.C., Pearlson, G.D., Andreassen, O.A., Agartz, A., Westlye, L.T., Hauvk, U.K., Dale, A.M., Melle, I., Hartberg, C.B., Gruber, O., Kraemer, B., Zilles, D., Donohoe, G., Kelly, S., McDonald, C., Morris, D.W., Cannon, D.M., Corvin, A., Machielsen, M.W.J., Koenders, L., de Haan, L., Veltman, D.J., Satterhwaite, T.D., Wolf, D.H., Gur, R.C., Gur, R.E., Potkin, S.G., Mathalon, D.H., Mueller, B.A., Preda, A., Macciardi, F., Ehrlich, S., Walton, E., Hass, J., Calhoun, V.D., Bockholt, H.J., Sponheim, S.R., Shoemaker, J.M., van Haren, N.E.M., Pol, H.E.H., Ophoff, R.A., Kahn, R.S., Roiz-Sanitiañez, R., Crespo-Facorro, B., Wang, L., Alpert, K., Jönsson, E.G., Dimitrova, R., Bois, C., Whalley, H.C., McIntosh, A.M., Lawrie, S.M., Hashimoto, R., Thompson, P.M., Turner, HA for the ENIGMA Schizophrenia Working Group, 2015. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Mol. Psychiatry (in press). doi: 10.1038/mp.2015.63.

Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L., Augustinack, J., Dickerson, B., Golland, P., Fischl, B., 2009. Automated segmentation of hippocampal subfields from ultra high resolution in vivo MRI. Hippocampus 19, 549–557.

Van, L., Van, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L.L., Augustinack, J., Dickerson, B.C., Golland, P., Fischl, B., 2008. Model-based segmentation of hippocampal subfields in ultra-high resolution in vivo MRI. Medical Image Computing and Computer-Assisted Intervention: MICCAI 11, pp. 235–243.

Winkler, A.M., Kochunov, P., Blangero, J., Almasy, L., 2010. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. NeuroImage 53, 1135–1146.

Winterburn, J.L., Pruessner, J.C., Chavez, S., Schira, M.M., Lobaugh, N.J., Voineskos, A.N., et al., 2013. A novel in vivo atlas of human hippocampal subfields using high-resolution 3 T magnetic resonance imaging. NeuroImage 74, 254–265.

Wisse, L., Gerritsen, L., Zwanenburg, J., Kuijf, H., Luijten, P., Biessels, G., Geerlings, M., 2012. Subfields of the hippocampal formation at 7 T MRI: in vivo volumetric assessment. NeuroImage 61, 1043–1049.

Wisse, L., Biessels, G., Geerlings, M., 2014. A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer. Front. Aging Neurosci. 6, 261.

Wonderlick, J., Ziegler, D., Hosseinivarnamkhasti, P., Locasio, J., Bakkhour, A., Vanderkouwe, A., Triantafyllou, C., Corkin, S., Dickerson, B., 2009. Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. NeuroImage 44, 13241333.

Wright, I.C., Sham, P., Murray, R.M., Weinberger, D.R., 2002. Genetic contributions to regional variability in human brain structure: methods and preliminary results. NeuroImage 17, 256–271.

Yamasaki, N., Maekawa, M., Kobayashi, K., Kajii, Y., Maeda, J., Soma, M., Takao, K., Tanda, K., Ohira, K., Toyama, K., Kanzaki, K., Fukunaga, K., Sudo, Y., Ichinose, H., Ikeda, M., Iwata, N., Ozaki, N., Suzuki, H., Higuchi, M., Suhara, T., Yuasa, S., Miyakawa, T., 2008. Alpha-CaMKII deficiency causes immature dentate gyrus, a novel candidate endophenotype of psychiatric disorders. Mol. Brain 1, 6.

Yushkevich, P.A., Avants, B.B., Pluta, J., Das, S., Minkoff, D., Mechanic-Hamilton, D., Glynn, S., Pickup, S., Liu, W., Gee, J.C., Grossman, M., Detre, J.A., 2009. A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 T. NeuroImage 44, 385–398.

Yushkevich, P., Wang, H., Pluta, J., Das, S., Craige, C., Avants, B., Weiner, M., Mueller, S., 2010. Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. NeuroImage 53, 1208–1224.

Yushkevich, P.A., Amaral, R., Augustinack, J., Bender, A., Bernstein, J., Boccardi, M., Bocchetta, M., Burggren, A., Carr, V., Chakravarty, M., Chetelat, G., Daugherty, A., Davachi, L., Ding, S.-L., Ekstrom, A., Geerlings, M., Hassan, A., Huang, Y., Iglesias, J., Joie, R., Kerchner, G., LaRocque, K., Libby, L., Malykhin, N., Mueller, S., Olsen, R., Palombo, D., Parekh, M., Pluta, J., Preston, A., Pruessner, J., Ranganath, C., Raz, N., Schlichting, M., Schoemaker, D., Singh, S., Stark, C., Suthana, N., Tompary, A., Turowski, M., Leemput, K., Wagner, A., Wang, L., Winterburn, J., Wisse, L., Yassa, M., Zeineh, M., (HSG), H., 2015. 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.

Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, M.C., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R., 2006. Statistical validation of image segmentation quality based on a spatial overlap index. Acad. Radiol. 11 (2), 178–189.

Zubicaray, D.G., Chiang, M.C., McMahon, K.L., 2008. Meeting the challenges of neuroimaging genetics. Brain Imaging Behav. 2 (4), 258–263.

Please cite this article as: Whelan, C.D., et al., Heritability and reliability of automatically segmented human hippocampal formation subregions, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.039