SUPPLEMENTAL INFORMATION

S1 Supplemental Methods

S1.1 Specimen Imaging Details

Table S1 provides demographic data and imaging parameters for each specimen. The protocol for handling and imaging tissues evolved over the course of the project. Prior to 2011, specimens ($n = 9$) were imaged using coils with a 70 mm inner diameter, without submersion in a MRI-neutral fluid. From 2011 onward ($n = 23$), specimens were imaged using a custom-made coil with a 35 mm inner diameter and a long (80 mm) z-field. These 23 specimens were scanned in a cylindrical container filled with an MRI-neutral industrial lubricant. Specimens were cut down to the size necessary to fit in the coil and/or container. During cutting, the hippocampal formation was kept intact, while also attempting to preserve as much of the extrahippocampal MTL cortex as possible. All MRI scans were obtained using a multi-slice spin echo sequence. Sequence parameters varied slightly between specimens, with TR ranging between 4 s and 5 s and TE ranging between 21 and 23 ms. Scans had resolution of $200 \times 200 \times 200\,\mu\text{m}^3$, with the exception of one specimen that was scanned at $160 \times 160 \times 160\,\mu\text{m}^3$ resolution.

Following acquisition and reconstruction, images were corrected for bias field non-uniformity using the N4ITK algorithm [24] and normalized to a common intensity range by clipping the intensities below the 0.1 and above the 99.9 percentile, and scaling the intensity range to $[0, 1000]$. An error in scanner gradient calibration was discovered towards the end of the study and a 3D printed phantom similar to the one described in [20] was used to derive linear scaling factors between the scanner coordinate frame and the physical coordinate frame. These linear factors (6% in x, 3% in y, 11% in z) were applied to the MRI scans as part of postprocessing.

S1.2 SRLM and Hippocampus Segmentation

Segmentations of the SRLM and the hippocampus were used to guide groupwise registration. These segmentations were initialized using automatic and semi-automatic approaches and completed using extensive manual editing in ITK-SNAP software [31]. In the 20 NDRI cases, the hippocampus was labeled manually to provide "rough" quality masks required by our earlier approach to atlas-building [29]. These masks were refined (i.e., adjusted to better match image features) using leave-one-out automatic segmentation using a multi-atlas label fusion segmentation method [27], followed by manual editing in ITK-SNAP. To segment the SRLM in the 20 NDRI cases, a Hessian-based filter [13]was used to detect thin sheet-like structures inside of the hippocampus mask, followed by random forest classification of SRLM vs. non-SRLM voxels, then followed by extensive manual correction.

In the 12 cases from CNDR, an initial outline of the hippocampus was generated by interpolating outlines traced manually every 13 slices and used as input to guide multi-atlas label fusion segmentation [27], with NDRI images and segmentations serving as atlases. The resulting segmentations of the whole hippocampus and SRLM were manually corrected. Examples of hippocampus and SRLM segmentations from this multi-step procedure are in Figure S1.

S1.3 MRI Atlas Generation

S1.3.1 Atlas Generation Overview

The algorithm for constructing an atlas from ex vivo MRI scans consists of three stages, which are summarized briefly below.

1. **Shape matching via minimum spanning tree (MST).** Approximate correspondences between the SRLM and hippocampus segmentations of all specimens are computed by (1) constructing an MST on a fully connected graph in which nodes represent specimens and edges are weighted by anatomical similarity between specimens; and (2) registering each specimen to the specimen located at the root of the MST using a sequence of diffeomorphic registrations along the path between that specimen and the root specimen in the MST. The MST-based approach reduces the amount of anatomical difference

Table S1: Subject demographic data and MRI acquisition parameters. Group column gives the subject's clinical diagnostic group (AD:Alzheimer's disease; NC: Non-dementia control; OD: Other Dementia). See text for full details. Specimen NDRI12-R (marked with *) was excluded after the initial building of the atlas because of the registration failures in the atlas building procedures caused by the unusually large tail of the caudate nucleus. This specimen is not considered in any of the morphometric analyses in the paper and is excluded from Main Table 1.

Sample ID	Age (years)	Sex	Group	Fixation (days)	MRI Resolution (μm $\times$ μm $\times$ μm)	Coil	TR (ms)	TE (ms)	Scan Time (h)	Stitched	Histology
NDRI01-R	89	F	OD	108	160x160x160	Insight 70mm	5000	23	13.3		Y
NDRI02-R	76	M	NC	256	200x200x200	M2M 70mm	4000	21	15.1	Y	Y
NDRI03-L	89	F	NC	130	200x200x200	Insight 70mm	5000	21	16.1		
NDRI03-R				142	200x200x200	Insight 70mm	5000	23	13.3	Y	
NDRI04-L	82	M	NC	145	200x200x200	M2M 70mm	4000	21	15.1	Y	
NDRI05-L	61	M	AD	66	200x200x200	Insight 70mm	5000	23	16.7		
NDRI06-L	72	M	AD	580	200x200x200	M2M 35mm	4000	21	15.1	Y	Y
NDRI06-R				517	200x200x200	M2M 35mm	4000	21	15.1	Y	
NDRI07-L	75	M	NC	292	200x200x200	M2M 35mm	4000	21	15.1	Y	
NDRI08-R	75	M	NC	523	200x200x200	M2M 35mm	4000	21	7.6		Y
NDRI09-L	78	F	NC	284	200x200x200	M2M 35mm	4000	21	15.1	Y	
NDRI09-R				217	200x200x200	M2M 35mm	4000	21	16	Y	Y
NDRI10-L	78	F	NC	521	200x200x200	M2M 35mm	4000	21	15.1	Y	
NDRI10-R				132	200x200x200	M2M 70mm	4000	21	15.1	Y	
NDRI11-L	70	F	NC	142	200x200x200	M2M 70mm	4000	21	13.6	Y	
NDRI12-L	81	F	NC	527	200x200x200	M2M 35mm	4000	21	15.1	Y	Y
NDRI12-R*				138	200x200x200	M2M 70mm	4000	21	15.1	Y	
NDRI13-L	86	F	AD	255	200x200x200	M2M 35mm	4000	21	15.1	Y	
NDRI13-R				538	200x200x200	M2M 35mm	4000	21	15.1	Y	
NDRI14-L	90	F	AD	442	200x200x200	M2M 35mm	4000	21	15.1	Y	Y
CNDR01-R	76	F	OD	147	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR02-R	60	M	AD	82	200x200x200	M2M 35mm	4000	21	13.5		Y
CNDR03-R	75	M	OD	31	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR04-R	61	F	OD	43	200x200x200	M2M 35mm	4000	21	15.1	Y	
CNDR05-L	74	F	AD	24	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR06-R	54	M	AD	47	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR07-R	80	F	NC	147	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR08-R	67	F	OD	34	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR09-R	83	M	OD	43	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR10-L	73	M	OD	22	200x200x200	M2M 35mm	4000	22	15.1		Y
CNDR11-L	65	M	OD	27	200x200x200	M2M 35mm	4000	22	15.9	Y	
CNDR12-L	76	M	OD	35	200x200x200	M2M 35mm	4000	22	15.9	Y	

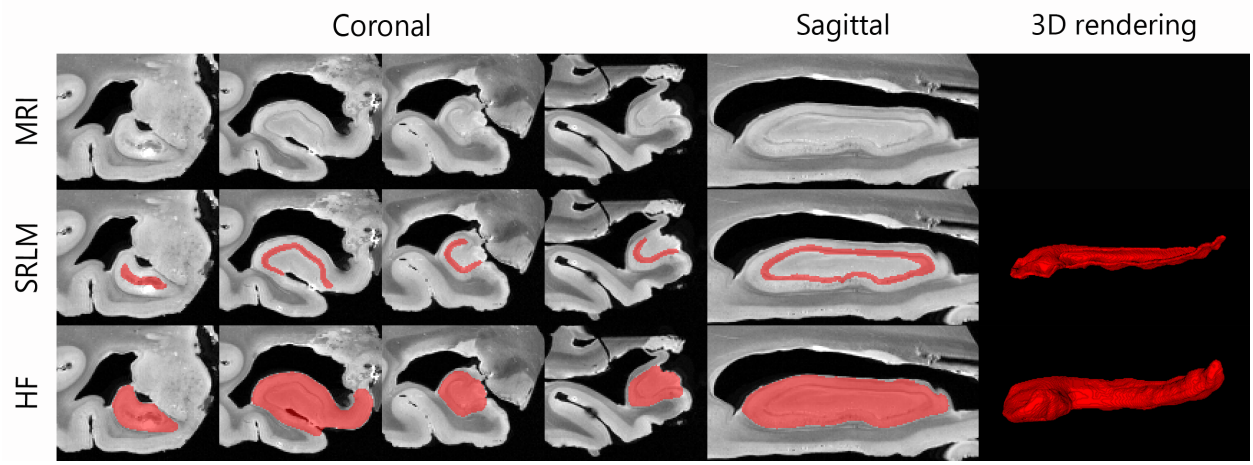


Figure S1: Example segmentation of SRLM and hippocampal formation (HF) in an ex vivo MRI scan. These segmentations were used to guide groupwise registration for atlas building.

between specimens involved in pairwise registrations. All registrations in this stage are applied to SRLM/hippocampus label images.

2. **Shape averaging via geodesic shooting (GS).** Shape correspondences computed in the MST stage are used to extract a set of landmarks on the SRLM and hippocampus surfaces in the MST root specimen and corresponding sets of landmarks in each of the other specimens. Then a computational anatomy algorithm by Vaillant et al. [25] is used to compute the mean SRLM/hippocampus shape and a set of diffeomorphic transformations between the mean shape and each of the input shapes that minimize a total deformation energy. Diffeomorphic transformations between the mean shape and input shapes are computed using geodesic shooting algorithms [3, 28]. These transformations are computed between landmark sets but can be interpolated to produce diffeomorphic transformations between images.

3. **Groupwise intensity-based registration (INT).** In this stage, specimen MRI scans are matched to a population mean image on the basis of intensity similarity. This stage uses the unbiased population atlas construction approach, which alternates between intensity averaging to obtain the mean image and intensity-based diffeomorphic registration to the mean [15, 5, 14]. This algorithm is applied to the MRI specimens after they have been transformed into the shape average space by transformations computed in the GS stage; in other words, MST and GS stages serve to provide a high-quality initialization for the groupwise intensity-based registration algorithm in the INT stage. The final output of this three-stage algorithm consists of an average or "template" image of the hippocampal formation, and a set of diffeomorphic transformations between this template and each of the input MRI scans.

Each of the three stages plays a critical role in atlas construction. Without the first stage, groupwise registration fails to match up corresponding locations between specimens, due to the complexity of the image and varying context. Without the second stage, the atlas has the shape of one of the specimens (root specimen), thus failing to capture the anatomical properties of the population. Without the third stage, the atlas is based purely on SRLM/hippocampus segmentations and does not take into account intensity features. In the Supplemental Section S2.1 we demonstrate the contribution of each stage of the atlas generation algorithm to the final atlas.

The following sections define the mathematical notation used to describe registration and provide details of the MST, GS and INT stages.

S1.3.2 Registration Notation

Atlas construction involves multiple image registration steps, for which we adopt the following conventions. We let indices i, j refer to specimens. We use Υ_j to denote the MRI for specimen j, and $\mathbf{S}_j$ to denote the two-component hippocampus/SRLM segmentation, with its component images denoted S_j^H and S_j^S. In the general discussion of registration below, we use I_j to denote some k-component image associated with specimen j, i.e., $I_j : \Omega_j \in \mathbb{R}^3 \rightarrow \mathbb{R}^k$, where Ω_j is the image domain. For example, I_j may denote the three-component image formed by combinng Υ_j and $\mathbf{S}_j$.

The expression "to register specimen j to specimen i" refers to performing registration with I_i as the "fixed" image and I_j as the "moving" image, yielding a transformation $T_{i\rightarrow j} : \Omega_i \rightarrow \mathbb{R}^3$ that assigns a new coordinate $T_{i\rightarrow j}(x)$ to each point x in the fixed image domain Ω_i. We refer to the image $I_j \circ T_{i\rightarrow j}$ as the "resliced" moving image, where $\circ$ denotes function composition. The resliced moving image is defined in the fixed image domain Ω_i, and at each point $x \in \Omega_i$ has the intensity $I_j(T_{i\rightarrow j}(x))$. Registration minimizes some metric of dissimilarity between the fixed image I_i and the resliced moving image $I_j \circ T_{i\rightarrow j}$. In most general terms, registration is expressed as

$$T^*_{i\rightarrow j} \quad = \quad \arg\min_{T_{i\rightarrow j}} \mu(I_i, I_j \circ T_{i\rightarrow j}) + \lambda\rho(T_{i\rightarrow j}) \,,$$

where μ is the image dissimilarity metric, ρ is an optional regularization term, and λ is a scalar parameter.

Registration in this paper is performed using an efficient implementation of the greedy diffeomorphic algorithm described by [15] provided in an open-source software "*Greedy*"[1]. This software package is mod-

[1] `https://github.com/pyushkevich/greedy`

eled after the Symmetric Normalization (SyN) in the ANTS software package [4], but is optimized for computational speed, foregoing the symmetric registration model and implementing highly optimized image resampling and metric computation. Greedy implements common image similarity metrics including patch normalized cross-correlation (NCC) and sum of squared difference (SSD) metrics. Diffeomorphic registration is computed iteratively using the following update equation [15, 4]:

$$\begin{aligned} \psi^{\gamma} &= \mathrm{Id}+\epsilon^{\gamma} \cdot \left[G_{\sigma_\mathrm{s}} * D_{\phi^T_{i\to j}} \mu(I_i, I_j \circ \phi^{\gamma}_{i\to j}) \right] \\ \phi^{\gamma+1}_{i\to j} &= G_{\sigma_\mathrm{t}} * (\phi^{\gamma}_{i\to j} \circ \psi^{\gamma}) \\ \phi^{0}_{i\to j} &= Id \end{aligned}$$

where γ is the iteration number, $D_\phi \mu$ is the gradient of the metric with respect to ϕ, ϵ^T is the step size, $G_\sigma * \phi$ denotes convolution of ϕ with an isotropic Gaussian kernel with standard deviation σ, and Id is the identity transformation. For sufficiently small ϵ^T and sufficiently large σ_s, $\psi^T$is smooth and has positive Jacobian determinant for all $x \in \Omega_i$, hence is diffeomorphic. Since diffeomorphisms form a group under composition, $\phi^{T+1}_{i\to j}$ is diffeomorphic also. As in [4], parameters σ_s and σ_t serve to regularize the registration; larger values of σ_s result in smoother segmentations, and larger values of σ_t result in less total deformation. Registration is typically performed in a multi-resolution scheme, starting with images subsampled by factors of 2^k, then refined on images subsampled by factors of 2^{k-1}, etc., and at the end refined at full resolution.

In what follows, we use the following notation to describe diffeomorphic registration of specimen j to specimen i:

$$\phi_{i\to j} = \mathcal{R}_{\mathrm{diff}}(I_i \to I_j; \mu, \mathbf{w}, \sigma_\mathrm{s}, \sigma_\mathrm{t}, N) \ , \tag{1}$$

where the first part of the expression ($\bullet \to \bullet$) gives the fixed and moving images matched by the registration and the second part of the expression gives the parameters of the registration, which include the choice of metric μ (SSD or NCC, the latter with patch size, e.g., NCC[5x5x5]), the k-component vector $\mathbf{w}$ of weights assigned to the different image components when computing the dissimilarity metric, the regularization parameters σ_s and σ_t, and the number of iterations of greedy registration at each level of the multi-resolution pyramid (e.g., $N = \{100, 40, 0\}$ means 100 resolutions at 4x subsampling; 40 at 2x subsampling; and none at full resolution).

Greedy can also perform affine (and rigid) registration using gradient descent optimization of the dissimilarity metric. This is denoted

$$A_{i\to j} = \mathcal{R}_{\mathrm{aff}}(I_i \to I_j; \mu, w, A_0) \ , \tag{2}$$

where A_0 is the initial affine (or rigid) transformation between the images, and the other parameters have the same meaning as in (1). Lastly, Greedy can be used to match the moments of inertia (MOI) between two images. This produces a rigid transformation that matches the centers of mass of the two images and matches the eigenvectors of their second moment tensors. The latter is ambiguous with respect to negation of each eigenvector. To resolve this ambiguity, the image dissimilarity metric is evaluated for all possible matchings of the eigenvectors and the one with the minimal metric is taken. When matching specimens from different hemishpheres, a flip is applied to the moving image as part of the MOI matching. MOI matching is denoted:

$$A_{i\to j} = \mathcal{R}_{\mathrm{moi}}(I_i \to I_j; \mu, w) \ .$$

S1.3.3 Stage 1: Shape Matching via Minimum Spanning Tree (MST)

As the first step, we construct an undirected complete graph $\mathcal{G}$, in which each specimen corresponds to a node, and each pair of specimens corresponds to a directed edge. Each edge (i, j) is assigned a weight η_{ij} that describes the degree of anatomical similarity between specimens i and j. To obtain the weights, we first perform rough, i.e., *highly regularized,* deformable registration between the hippocampus/SRLM segmentations of the two specimens, and then measure residual difference between the co-registered segmentations.

The registration is performed as follows:

$$\begin{aligned} A^0_{i\to j} &= \mathcal{R}_{\text{moi}}(\mathbf{S}_i \to \mathbf{S}_j; \mu = SSD, \mathbf{w} = \{1,5\})\,, \\ A^{\text{rough}}_{i\to j} &= \mathcal{R}_{\text{aff}}(\mathbf{S}_i \to \mathbf{S}_j; \mu = \text{SSD}, \mathbf{w} = \{1,5\}, A_0 = A^0_{i\to j})\,, \\ \phi^{\text{rough}}_{i\to j} &= \mathcal{R}_{\text{diff}}(\mathbf{S}_i \to \mathbf{S}_j \circ A^{\text{rough}}_{i\to j}; \mu = \text{SSD}, \mathbf{w} = \{1,5\}, \\ &\qquad \sigma_{\text{s}} = 2\,\text{mm}, \sigma = 0.1\,\text{mm}, N = \{50,50,20,0\}). \end{aligned}$$

The weight factors $\mathbf{w} = \{1,5\}$ (i.e., weight 1 for the hippocampus and 5 for the SRLM) were chosen *a priori* based on the observation that the ratio of SRLM to hippocampus volume in the specimens was approximately $1:5$. By setting $\sigma_{\text{s}} = 2\,\text{mm}$, i.e., 10 voxels, we obtain a highly regularized deformation with a very smooth deformation field. This deformation matches the overall shape of the specimens but significant residual differences in shape remain. The weight η_{ij} is then assigned as

$$\eta_{ij} = 1 - \frac{1}{2}\text{GDSC}(\mathbf{S}_{\text{i}}, \mathbf{S}_{\text{j}} \circ \text{A}^{\text{rough}}_{\text{i}\to\text{j}} \circ \phi^{\text{rough}}_{\text{i}\to\text{j}}) - \frac{1}{2}\text{GDSC}(\mathbf{S}_{\text{i}} \circ \text{A}^{\text{rough}}_{\text{j}\to\text{i}} \circ \phi^{\text{rough}}_{\text{j}\to\text{i}}, \mathbf{S}_{\text{j}})\,,$$

where GDSC is the generalized Dice similarity coefficient [8], a measure of total overlap between multi-label segmentations. GDSC lies in the range $[0,1]$, with 1 being perfect overlap and 0 being no overlap. Smaller values of η_{ij} indicate greater anatomical similarity between specimens i and j.

We then use Prim's algorithm [22] to compute the minimum spanning tree (MST) in the graph $\mathcal{G}$. The MST is a tree that includes all nodes of $\mathcal{G}$ and has minimum total edge weight. We then define the root node of the MST as follows:

$$R = \underset{r\in[1\ldots N]}{\arg\min} \sum_{q=1}^{N} d_{\text{mst}}(r,q)\,,$$

where d_{mst} is the edge length of the path in the MST between two nodes.

Lastly, we perform a sequence of deformable registrations between each specimen j and the root specimen R, following the path between j and R in the MST. An important property of this registration sequence is that it involves smallest possible deformations between pairs of specimens, where smallest refers to having the least shape difference. Indeed, if edge (a,b) is an edge in a path from j to R in the MST, then by the cycle property of the MST, any other path in $\mathcal{G}$ that connects a to b contains an edge with weight greater than η_{ab}.

The sequence of registrations is performed as follows. Let $\{i_0,\ldots,i_p\}$ be the sequence of specimens on the path from j to R, with $i_0 = j$ and $i_p = R$. Then we perform a sequence of registrations for $m = 1,\ldots,p$:

$$\begin{aligned} A^{\text{mst}}_{i_m\to i_{m-1}} &= \mathcal{R}_{\text{aff}}(\mathbf{S}_{i_m} \to \mathbf{S}_j \circ T^{\text{mst}}_{i_{m-1}\to j}; \mu = \text{SSD}, \mathbf{w} = \{1,5\}, A_0 = A^{\text{rough}}_{i_m\to i_{m-1}})\,, \\ \phi^{\text{mst}}_{i_m\to i_{m-1}} &= \mathcal{R}_{\text{diff}}(\{\mathbf{S}_{i_m}, \Upsilon_{i_m}\} \to \{\mathbf{S}_j, \Upsilon_j\} \circ T^{\text{mst}}_{i_{m-1}\to j} \circ A^{\text{mst}}_{i_m\to i_{m-1}}, \\ &\qquad \mu = \text{SSD}, \mathbf{w} = \{1,5,\epsilon_\Upsilon\}, \\ &\qquad \sigma_{\text{s}} = 0.6\text{mm}, \sigma_{\text{t}} = 0.1\text{mm}, N = \{100,50,40,20\}). \\ T^{\text{mst}}_{i_m\to j} &= T^{\text{mst}}_{i_{m-1}\to j} \circ A^{\text{mst}}_{i_m\to i_{m-1}} \circ \phi^{\text{mst}}_{i_m\to i_{m-1}}\,. \end{aligned}$$

In this sequence of registrations, $T^{\text{mst}}_{i_m\to j}$ represents the combined transformation between specimen i_m and specimen j (with $T^{\text{mst}}_{i_0\to j} = \text{Id}$). Note that the deformable registration above includes the MRI with a very small weight ϵ_Υ. This is done in order to slightly improve the registration outside of the hippocampus mask. We set the weight empirically to $6\cdot 10^{-8}$ (the MRI scans are normalized to the range $[0,1000]$, whereas the segmentation images have range $[0,1]$). The registrations are also performed with a smaller smoothing parameter $\sigma_{\text{s}} = 0.6\text{mm}$, which results in less regularization.

The final output from Stage 1 consists of a set of transformations $T^{\text{mst}}_{R\to j}$ between the root MST node and each specimen, which closely match up the hippocampus and SRLM boundaries, as shown in Figure S10. The average of the transformed MR images is shown in the first row of Figure S9.

S1.3.4 Stage 2: Shape Averaging via Geodesic Shooting (GS)

A set of L point landmarks $\mathbf{X}^R = \{X_1^R, \ldots, X_L^R\}$ is sampled from the SRLM and hippocampus surfaces of the root specimen R. These point landmarks are extracted with roughly uniform sampling using the Poisson Disk sampling algorithm [7] implemented in the VCG Library[2]. For each specimen j, the corresponding landmarks in its native space are obtained by applying the transformations from Stage 1:

$$\mathbf{X}^j = \{X_1^j, \ldots, X_L^j\} = \{T_{R\to j}^{\text{mst}}(X_1^R), \ldots, T_{R\to j}^{\text{mst}}(X_L^R)\}\,.$$

These sets of corresponding landmarks are input to the shape averaging algorithm implemented following the theoretical framework of [25]. Over iterations $m = 1, \ldots, M$ the algorithm iteratively updates the shape average $\bar{\mathbf{X}}^m$ and a set of diffeomorphic transformations between $\bar{\mathbf{X}}^m$ and $\mathbf{X}^j$. The algorithm alternates between registering $\mathbf{X}^j$ to $\bar{\mathbf{X}}^m$ for each j using the geodesic shooting algorithm [3, 28] and updating the shape average $\bar{\mathbf{X}}^{m+1}$. These two steps are summarized below.

Landmark Registration Using Geodesic Shooting This operation is performed in parallel for each specimen j. First, the landmarks $\mathbf{X}^j$ are aligned to $\bar{\mathbf{X}}^m$ using the Procrustes algorithm [11], giving a linear transform $A_j^{\text{proc},m}$ and a new set of landmarks $\hat{\mathbf{X}}^j$. We then use the geodesic shooting algorithm [3, 28] to find a spatial transformation ϕ that optimally matches the landmarks $\hat{\mathbf{X}}^j$ to $\bar{\mathbf{X}}^m$ under a regularization constraint. We adopt the notation of [3] to describe the geodesic shooting method. Landmark deformation is formulated as a dynamic system in which $q(t) = \{q_1(t), \ldots, q_L(t)\}$ describes the positions of the landmarks as a function of time ($t \in [0, 1]$) and $p(t) = \{p_1(t), \ldots, p_L(t)\}$ describes the momentum of the landmarks ($p_l(t) \in \mathbb{R}^3$). The initial positions of the landmarks are given $q(0) = \bar{\mathbf{X}}^m$, and the initial momenta of the landmarks $p(0) = \alpha_j^m$ are an unknown that the geodesic shooting algorithm optimizes for. The evolution of the system is formulated in terms of the Hamiltonian:

$$H(p, q) =< p, K(q)p >\,,$$

where $K(q)$ is a $(3L) \times (3L)$ with $L \times L$ diagonal blocks, with the (l, n)-th block equal to $G_{\sigma_{\text{gs}}}(\|q_l - q_m\|) \cdot \mathcal{I}_3$, where $G_{\sigma_{\text{gs}}}$ is a Gaussian kernel and $\mathcal{I}_3$ is an identity matrix. The system evolves according the the system of equations

$$\left\{ \begin{array}{rcl} \frac{dq}{dt} & = & \frac{\partial H}{\partial p}(q, p) \\ \frac{dp}{dt} & = & -\frac{\partial H}{\partial q}(q, p) \end{array} \right. . \tag{3}$$

$H(p, q)$ is the kinetic energy of the system and is constant over time when the system evolves according to (3). The landmark matching problem is formulated as the optimization

$$\alpha^* = \underset{\alpha \in \mathbb{R}^{3L}}{\arg\min}\, H(\bar{\mathbf{X}}^m, \alpha) + \lambda \cdot \|q(1) - \hat{\mathbf{X}}^j\|\,. \tag{4}$$

This minimization problem is discretized in time and solved using the gradient-based optimization [3]. Given the optimal solution α^*, and the corresponding landmark trajectories $q^*(t)$ and momenta $p^*(t)$ the landmark transformation can be interpolated over the entire spatial domain to yield a smooth velocity field

$$v(x, t) = \sum_{l=1}^{L} G_{\sigma_{\text{gs}}}(\|q_l^*(t) - x\|) \cdot p_l^*(t)\ , \qquad x \in \mathbb{R}^3. \tag{5}$$

From which a diffeomorphic transformation $\phi(x, t)$ can be derived by solving the differential equation [28]

$$\left\{ \begin{array}{rcl} \frac{d\phi}{dt}(x, t) & = & v(\phi(x, t), t) \\ \phi(x, 0) & = & x \end{array} \right. . \tag{6}$$

The landmark trajectories $q^*(t)$ are "embedded" in this diffeomorphic transformation, i.e. $q_l(t) = \phi(q_l(0), t)$. For specimen j and iteration m, we denote the end-point diffeomorphic transformation $\phi(x, 0)$ as $\phi_j^{\text{gs},m}$.

In summary, the geodesic shooting algorithm yields a diffeomorphic transformation $\phi_j^{\text{gs},m}$ that approximately matches landmarks $\bar{\mathbf{X}}^m$ to landmarks $\hat{\mathbf{X}}^j$, subject to a regularization term that penalizes the kinetic energy of the landmark transformation.

[2] `http://github.com/cnr-isti-vclab/vcglib/`

Shape Averaging Following [25], we update the average configuration $\bar{\mathbf{X}}^m$ by applying geodesic shooting in the direction of the average initial momenta. Specifically, given the initial momenta $\alpha_1^m, \ldots, \alpha_L^m$ obtained by optimization (4) for each specimen, we compute the average initial momentum $\bar{\alpha}_m = \frac{1}{L}\sum_{l=1}^{L} \alpha_j^m$ and apply the Hamiltonian flow (3) with $p(0) = \bar{\alpha}_m$ and $q(0) = \bar{\mathbf{X}}^m$. We then apply interpolation (5,6) to obtain a diffeomorphic transformation $\bar{\phi}_m$ corresponding to the initial momentum $\bar{\alpha}_m$.

The transformation $\bar{\phi}_m$ may include global scaling, which can cause the mean landmark configuration to grow or shrink over time. To avoid this, we perform Procrustes alignment between $\bar{\phi}_m(\bar{\mathbf{X}}^m)$ and $\bar{\mathbf{X}}^m$, which yields a transformation $\bar{A}^{\mathrm{proc},m}$. Finally, we update the landmark positions as

$$\bar{\mathbf{X}}^{m+1} = A_j^{\mathrm{proc},m}(\bar{\phi}_m(\bar{\mathbf{X}}^m)) .$$

Output of Stage 2 The steps of landmark registration and shape averaging are repeated for several iterations, until the shape average stabilizes. At each iteration, an average MRI intensity and average hippocampus/SRLM segmentation are computed by applying the transformations $T_j^{\mathrm{gs},m} = \phi_j^{\mathrm{gs},m} \circ A_j^{\mathrm{proc},m}$ to the images Υ_j and $\mathbf{S}_j$ and averaging their intensities, i.e.,

$$\begin{aligned} \bar{\Upsilon}^{\mathrm{gs},m} &= \frac{1}{N}\sum_{j=1}^{N} \Upsilon_j \circ T_j^{\mathrm{gs},m} \\ \bar{\mathbf{S}}^{\mathrm{gs},m} &= \frac{1}{N}\sum_{j=1}^{N} \mathbf{S}_j \circ T_j^{\mathrm{gs},m} \end{aligned}$$

The average images $\bar{\Upsilon}^{\mathrm{gs},M}, \bar{\mathbf{S}}^{\mathrm{gs},M}$ and the transformations $T_j^{\mathrm{gs},M}$ are the outputs of Stage 2. The average MRI intensity $\bar{\Upsilon}^{\mathrm{gs},M}$ is shown in the second row of Figure S9. A clear change in the shape of the hippocampus in the average image is observed after the application of geodesic shooting (i.e., comparing row 1 to row 2 in Figure S9).

S1.3.5 Stage 3: Groupwise Intensity-Based Registration

This stage involves computation of a unbiased population template [15, 5] using MR images $\Upsilon_1, \ldots, \Upsilon_N$ after initializing by the transformations computed in Stage 2. This is done iteratively, alternating between registration of the images to the current template, and updating the template. Let $\phi_j^{\mathrm{int},0}$ be the identity transformation. We compute the template at iteration m as

$$\begin{aligned} \bar{\Upsilon}^{\mathrm{int},m+1} &= \mathcal{S} * \sum_{j=1}^{N} \Upsilon_j \circ T_j^{\mathrm{gs},M} \circ \phi_j^{\mathrm{int},m} \\ \bar{\mathbf{S}}^{\mathrm{int},m+1} &= \sum_{j=1}^{N} \mathbf{S}_j \circ T_j^{\mathrm{gs},M} \circ \phi_j^{\mathrm{int},m} , \end{aligned}$$

where, following [6], $\mathcal{S}$ is a Laplacian unsharpening operator used to enhance the edges in the intensity template. Then, for each specimen j, the transformation to the template is updated using intensity-based registration with the normalized cross-correlation metric:

$$\begin{aligned} \phi_j^{\mathrm{int},m} = \mathcal{R}_{\mathrm{diff}}(&\bar{\Upsilon}^{\mathrm{int},m} \to \Upsilon_j \circ T_j^{\mathrm{gs},M}, \\ &\mu = \mathrm{NCC}[5 \times 5 \times 5], \\ &\sigma_{\mathrm{s}} = 0.6\mathrm{mm}, \sigma_{\mathrm{t}} = 0.2\mathrm{mm}, N = \{100, 100, 50\}). \end{aligned}$$

These steps are alternated until the template stabilizes. The final output of Stage 3 consists of the templates $\bar{\Upsilon}^{\mathrm{int},M}, \bar{\mathbf{S}}^{\mathrm{int},M}$ and the transformations $T_j^{\mathrm{int},M} = T_j^{\mathrm{gs},M} \circ \phi_j^{\mathrm{int},M}$, which map the templates into the native space of each specimen j.

S1.4 Histology-MRI Alignment and Cytoarchitectural Annotation of Hippocampal Subfields

S1.4.1 Histological Imaging

Nine specimens, each from a different individual, underwent histological processing (Table S1). Blocking, sectioning, staining and imaging followed the protocol derscribed in [2]. In brief, specimens were cut into $\sim$ 1cm thick blocks (46 in total). Blocks were scanned on the Varian 9.4 tesla scanner at $200 \times 200 \times 200\,\mu\text{m}^3$ resolution using the same multi-slice spin echo sequence as the intact specimens, but with fewer averages for 1-2 hour acquisition per block. After MRI, blocks were embedded in paraffin, sectioned with 5μm thickness at approximately 200μm intervals, stained using the Klüver-Barrera [16] method, and optically scanned at $0.5 \times 0.5\,\mu\text{m}^2$ resolution. Further details of the histological imaging procedure are provided in [2], and a visual summary is given in Figure 2 of that paper.

S1.4.2 Reconstruction

For each block, the scanned histology images are reconstructed in 3D and aligned to the MRI of the block. Following the graph-theoretical framework described in [2], each histology slide undergoes linear and deformable registration to the neighboring slides and to the matched cross-section of the MRI scan. The pipeline in the current paper makes several deviations from the approach in [2], as described below.

The framework in [2] is fully automatic and consists of three stages: (1) histology slides are aligned linearly to neighboring slices, with a graph-based approach used to skip over slices with missing content or poor image quality; (2) the resulting 3D histology stack is aligned to the MRI using affine registration in 3D; (3) histology slices are iteratively aligned and deformed to the corresponding MRI slice and to the neighboring histology slices, providing a reconstruction that is both continuous across histology and matched up with the MRI. In the present work, steps 1 and 2 are omitted, and instead, an interactive software tool *HistoloZee*[3] is used to perform the initial histology 3D reconstruction and alignment to the MRI. HistoloZee is a multi-featured tool for histology reconstruction. It allows concurrent visualization of multiple histology slices and MRI at full resolution and varying levels of zoom; rotation, translation and scaling of histology slices in-plane; 3D rotation, scaling and translation of the MRI volume; and annotation of structures on histology slides [1, Appendix A]. Although using HistoloZee to reconstruct histology stacks and align them to MRI required considerable manual effort, interactive processing allowed us to account for presence of gaps between sections, large differences in pose between adjacent slides, and better coped with slides with poor staining. The alignment from HistoloZee was then used to initialize iterative histology-MRI alignment, analogous to step 3 from [2].

Differently from [2], when registering histology images to MRI slices, we first remap the intensity of the histology images to simulate MRI-like appearance, which allows us to use the NCC metric for deformable registration, rather than the mutual information metric in [2]. Histology images have three color channels (red, green and blue, RGB). We use non-linear least squares regression to predict MRI appearance given histology RGB components. Let $\mathfrak{h}_1, \mathfrak{h}_2, \mathfrak{h}_3$ be the random variables representing the RGB intensities of a histology pixel and let Υ be the MRI intensity of the corresponding pixel. The we fit a fifth order polynomial model

$$\Upsilon = P_5(\mathfrak{h}_1, \mathfrak{h}_2, \mathfrak{h}_3; \mathcal{C}) \; + N(0, \sigma),$$

where $\mathcal{C}$ are the coefficients of the polynomial, to a set of 5000 pixels randomly sampled from a histology slide and the corresponding set of pixels in the slice of the block MRI aligned to the histology slide in HistoloZee. The fitted model is then applied to all pixels in the histology image, and the resulting simulated MRI intensities are clipped to the range of the MRI slice. Figure S2 gives an example of a histology slice, matched MRI slice, and the simulated MRI obtained by fitting this model.

Finally, as in [2], the block MRI is registered to the intact specimen MRI using highly constrained deformable registration after rigid-body initialization with manually placed point landmarks, providing a coordinate mapping between the intact specimen MRI and the reconstructed histology. The alignment between histology and MRI in the space of block MRI (where all histology slices are parallel, and each histology slice undergoes only an in-plane deformable registration) is illustrated in Figure S3. The alignment

[3] `http://www.nitrc.org/projects/historecon`

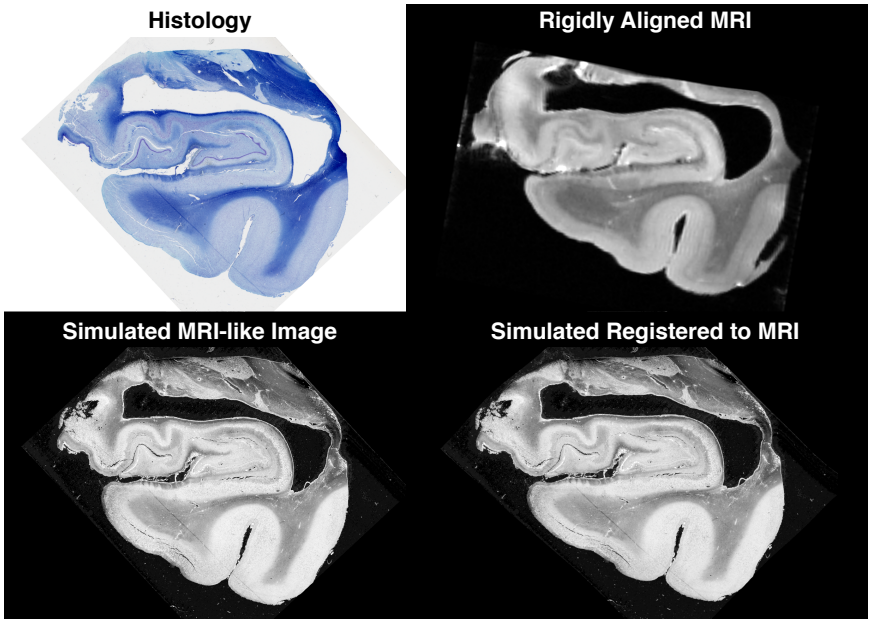


Figure S2: Remapping of histology red, green and blue channels into a single-component "simulated" MRI image. Intensity remapping makes it possible to use the normalized cross-correlation metric for deformable registration between MRI slices and histology.

between the histology and MRI in the space of whole-specimen MRI, where histology slices are no longer parallel to each other (due to block MRI to whole specimen MRI transformations) is shown in Figure S4.

S1.4.3 Histology Annotation

Hippocampal subfields were identified on the basis of cytoarchitectural features in histology images. Subfields were labeled in HistoloZee software following the anatomical rules described by Ding and Van Hoesen [10]. First, author JP annotated subfield boundaries on histology scans using line drawings and text labels in Adobe Photoshop software. Annotations were drawn on histology slices downsampled to approximately $20 \times 20\,\mu\text{m}^2$ resolution, but full-resolution histology scans loaded in HistoloZee were used to determine the actual locations of cytoarchitectural boundaries (e.g., the transition between CA1 and CA2 marked by increase in the density of pyramidal cell bodies and change in their shape from more ovoid to more triangular [12, 10]). These annotated histology images were then reviewed by the first author of [10] and edited as necessary. An example of an annotated histology slide is shown in Supplementary Figure S5. The annotations denote boundaries between hippocampal subfields CA1, CA2, CA3, DG, and SUB. However, SRLM is not separed from the CA and DG layers in the histological annotation. Instead, the atlas relies on MRI-based labeling of the SRLM as described in Section S1.2.

S1.4.4 Subfield Segmentation in MRI Space

Following the automated histology alignment and registration to the whole-hippocampus MRI, the cytoarchitecturally-derived subfield boundary annotations were mapped into the 3D MRI space and used to generate a contiguous 3D subfield segmentation. The approach in this paper is a departure from our previous approach in [2], where subfields were first fully segmented in each histology slice, and these segmentations were reconstructed in 3D MRI space and then manually edited for 3D continuity. The new procedure forgoes the time-consuming segmentations in histology slices, relying instead on the boundary annotations. The approach of segmenting hippocampal subfields in MRI space on the basis of annotations derived from histology has several advantages over the alternative approach of segmenting subfields in histology images and warping these 2D segmentations into the 3D MRI space. These include greater throughput, robustness to

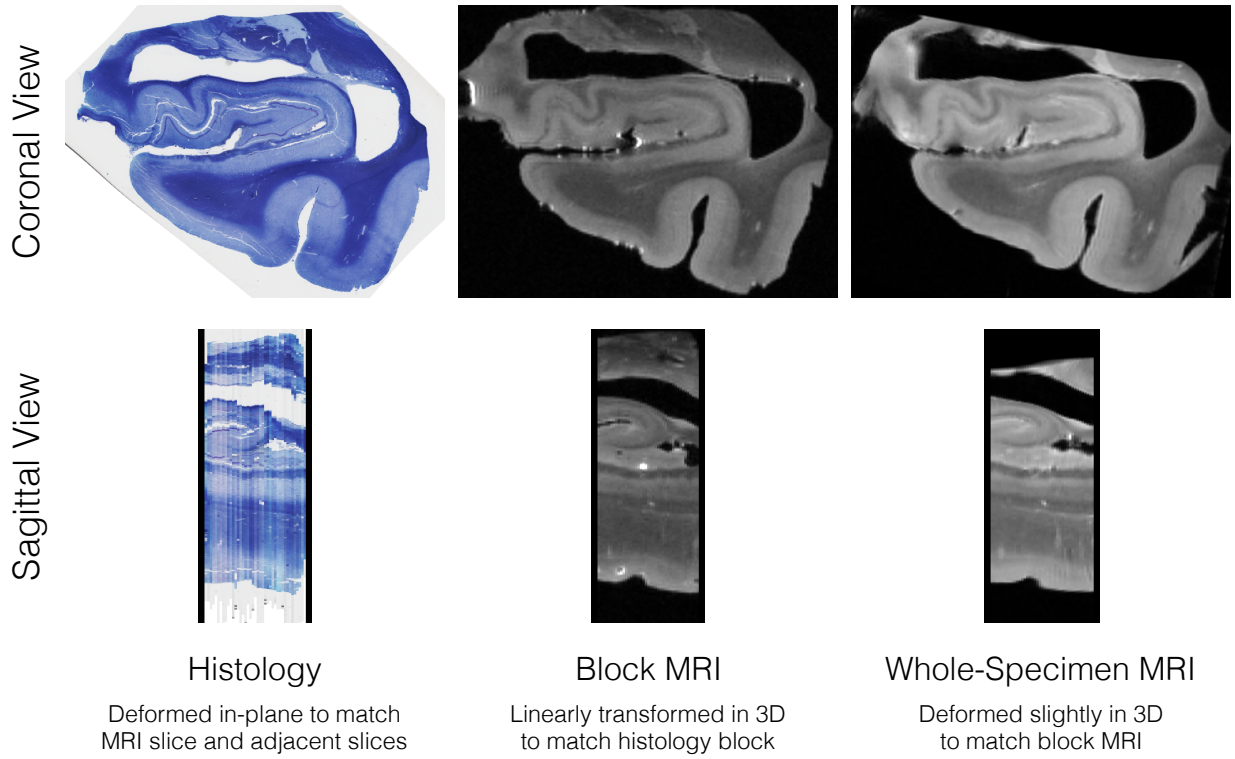


Figure S3: Histology reconstruction and alignment to block and whole-specimen MRI. The reconstructions are shown in the space of the block MRI, where histology slices are parallel and transformed in-plane. The sagittal view of the histology illustrates the relative continuity of the 3D histology reconstruction. The whole-specimen MRI, deformed under strong regularization constraints to match the block MRI, is also shown.

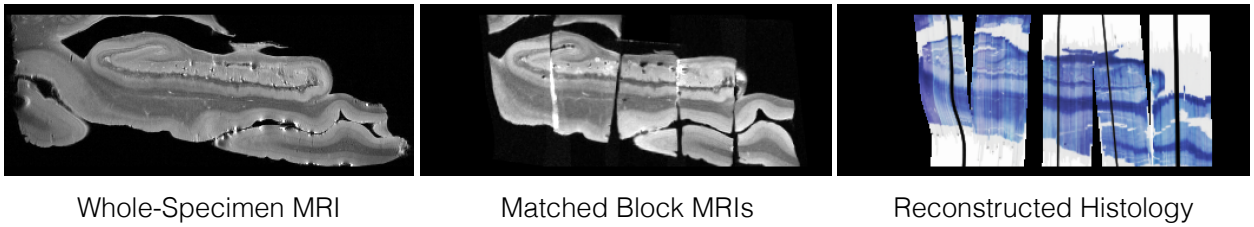


Figure S4: Histology reconstruction in the space of the whole-specimen MRI. The native-space whole-specimen MRI, block MRIs resampled into whole MRI space, and reconstructed histology blocks are shown in the sagittal section. Black spaces in the histology reconstruction correspond to gaps between blocks and missing spans of slides within blocks, as well as some residual tissue deformation not recovered by the block MRI to whole-specimen MRI registration.

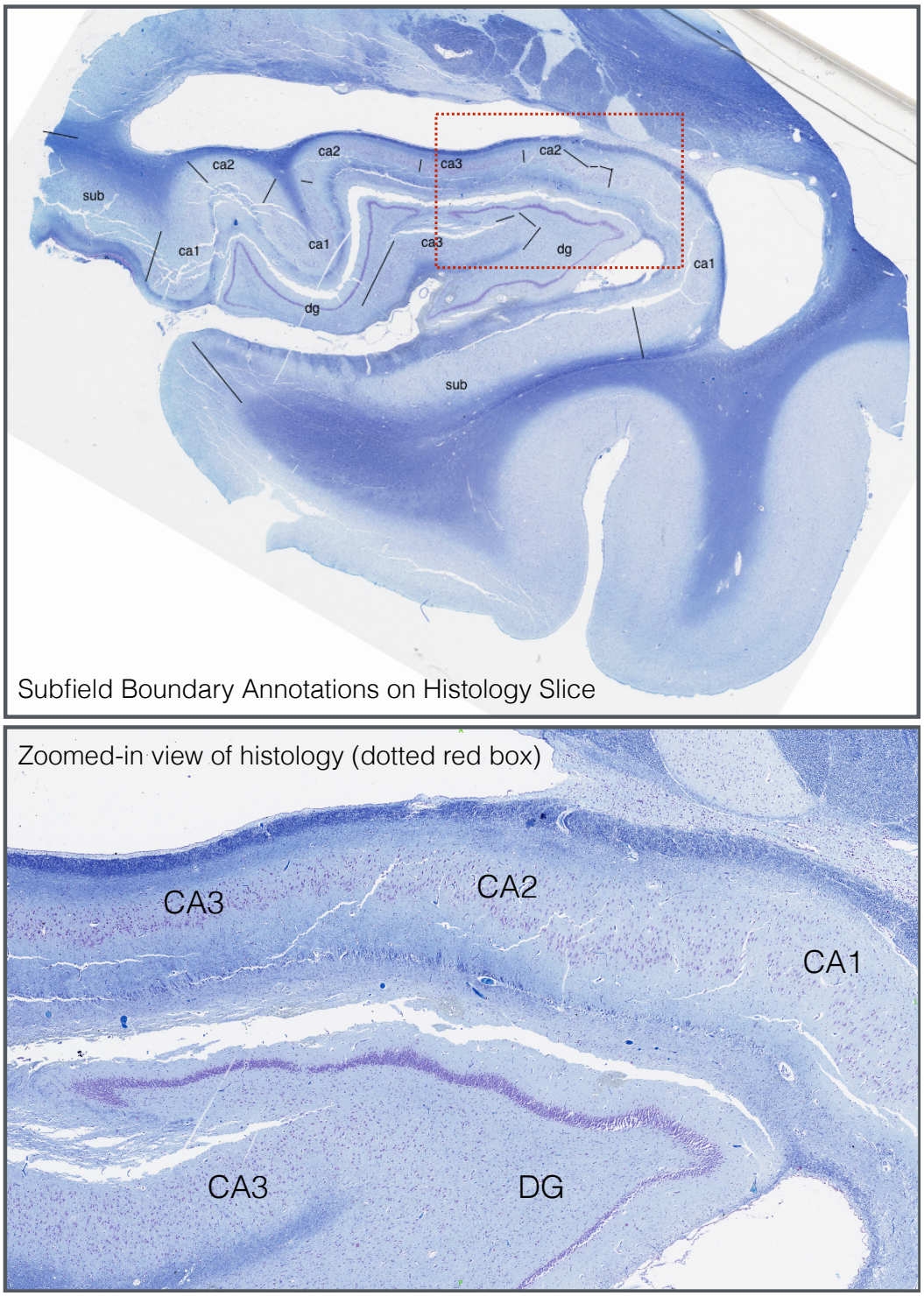


Figure S5: Histology annotation example. **Top:** a histology slide in the hippocampal head, with line annotations drawn by author JP and reviewed by author SLD. **Bottom:** a zoomed-in view of the region outlined by a red dashed box in full-resolution histology data. Full resolution histology was used to determine the location of annotated boundaries. Additional annotations are shown at a larger scale in Figures S16-S20.

errors in histology-to-MRI registration, and better continuity of the segmentation in 3D space. Moreover, segmenting hippocampal subfields directly in MRI space has the advantage of mitigating potential distorting effects of tears or other slight in-plane misalignments between the histology sections and MRI scans. In the presence of a registration mismatch between MRI and histology, the manual tracer can visually map subfield boundaries from the histology space to MRI space, instead of "blindly" relying on the registration as in the earlier approach. Note that the SRLM label is already traced in the MRI space (Section S1.2).

In the new subfield segmentation procedure, the co-registered MRI and histology with overlaid boundary annotations are displayed side-by-side in ITK-SNAP [31, 30], and subfield outlines are manually traced in MRI space. This is done separately for each block. In "block space", the histology slices are deformed in-plane to match the corresponding slices of the MRI. Thus it is still possible to see full detail of histology and the boundary annotations. Yet, since the MRI has 3D continuity in block space, it is also possible to trace subfields in a way that is consistent three-dimensionally. Author LW traced the boundaries of CA1, CA2, CA3, dentate gyrus and subiculum subfields in coronal MRI slices in each of the 46 blocks. Histology annotations were used to define CA1/CA2, CA2/CA3, CA3/DG, CA1/SUB boundaries, which cannot be clearly observed on MRI. MRI gray/white contrast was used to define inner and outer boundaries of the subfields. In certain head slices with very complex geometry of subfield boundaries, histological boundary annotations were not enough to infer the full segmentation, and segmentation performed directly in histology space (as in [2]) was consulted in order to label subfields in MRI space. Additionally, in 2 of the 46 blocks, a slight misalignment of histology and MRI was detected along the histology slicing direction, and matching histology slices were manually selected by author LW, instead of relying on the automatic histology/MRI reconstruction. Figure S6 illustrates this subfield segmentation procedure in a single slice.

To reduce segmentation time, subfield boundaries were traced by LW in block MRI space on every third coronal slice (every 0.6 mm) for the blocks containing hippocampal head and tail; and every fifth slice (every 1.0 mm) for the hippocampal body blocks. These sparse segmentations were then interpolated in the space of the whole-hippocampus MRI (Figure S7). Gaps between blocks and missing blocks were not interpolated. After interpolation, author LW reviewed the reconstructed segmentations in ITK-SNAP. In regions where 3D discontinuity between boundaries was observed, author LW consulted the block-space histology segmentations and histology annotations, and if warranted, edited the whole-MRI space subfield segmentation locally to improve continuity while still adhering to the cytoarchitectural evidence of subfield boundaries (white circle in Figure S7).

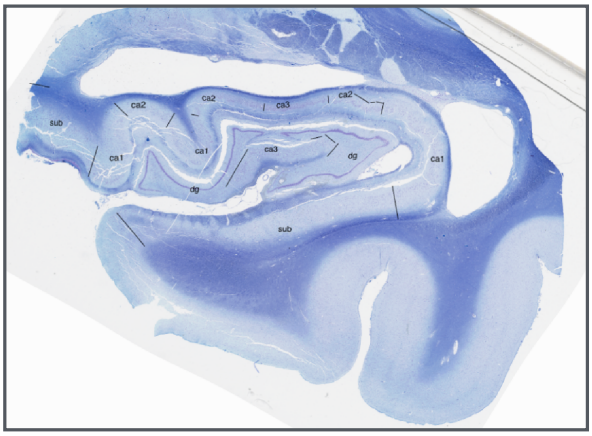
S1.4.5 Computation of Consensus Segmentation in MRI Atlas Space

A consensus segmentation of the hippocampal subfields in the MRI atlas was computed by majority voting among the warped subfield segmentations of the nine specimens with histology (shown in Figure S12), with slight regularization by a Markov random field prior. Let us consider the discrete representation of the MRI atlas space with voxels labeled $1, \ldots, N_{\text{vox}}$. For a voxel x in MRI atlas space, let $V[x,k]$ be the total number of "votes" for the label k among the nine warped segmentations, i.e., $V[x,k] = \sum_j S_j^k[x]$, where S_j^k is the warped binary segmentation image for label $k \in \{1, \ldots, K\}$ in specimen j. Since segmentations derived from histology have missing data (gaps between blocks), the total number of votes at a voxel, $\hat{V}[x] = \sum_{k=1}^{K} V[x,k]$, may be different at different voxels (K denotes the number of histology labels). The consensus multi-label segmentation of the atlas is obtained by solving the optimization problem

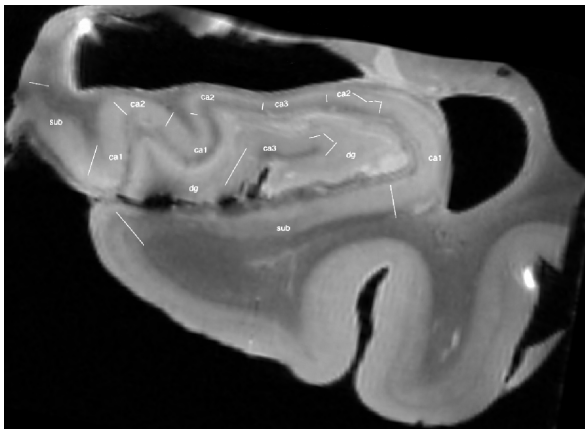
$$S^* = \underset{S \in \{1,\ldots,K\}^{N_{\text{vox}}}}{\arg\min} \sum_{x=1}^{N_{\text{vox}}} \left(\hat{V}[x] - V[x, S[x]] \right) + \beta \sum_{\{x,y\} \in \mathcal{E}} \delta(S[x], S[y]) \,, \tag{7}$$

where β is a scalar weight, $\mathcal{E}$ is the set of all face-adjacent voxels, and δ is the Kronecker delta function. This optimization problem corresponds to finding the mode of a Gibbs probability density, and can be solved using a polynomial-time graph cut algorithm [17]. The first (data) term in (7) penalizes the number of votes at a voxel x that disagree with the consensus label $S[x]$, while the second term penalizes spatial discontinuity in labels, and acts as a regularizer.

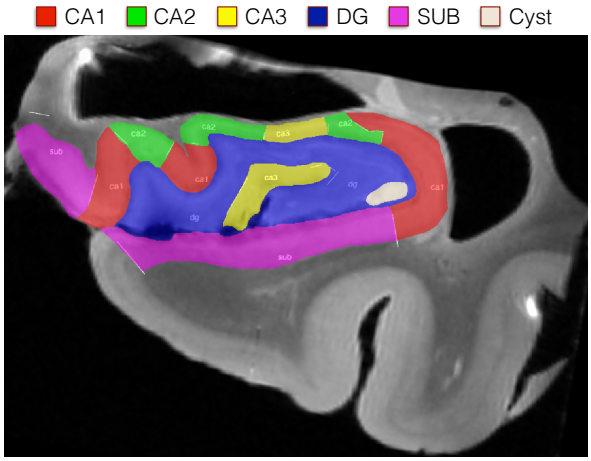
Regularization was employed because at some voxels there is little data available (due to gaps in the histology segmentations) and continuity may help determine the correct label assignment at those voxels. In determining the relative weight of the regularization term (β), we obtained consensus segmentations using



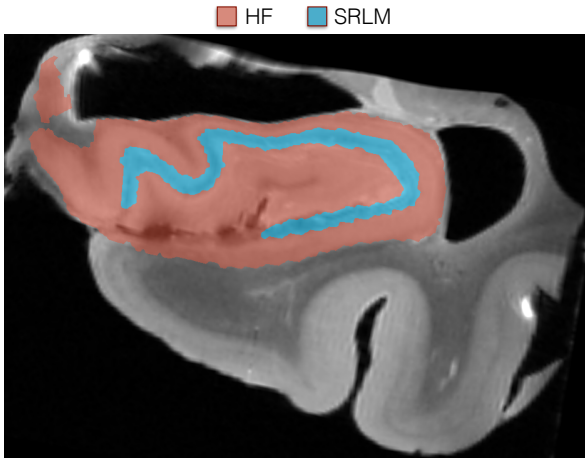
Histology with boundary annotations warped to the blocked MRI space



MRI slice with overlaid boundary annotations



Hippocampal subfield labels traced in block MRI space using the boundary annotations from histology



Whole-hippocampus and SRLM labels created directly from the MRI

Figure S6: Example of a histology annotation mapped onto an MRI slice and used to segment hippocampal subfields in block MRI space. The same histology slide is shown in Fig. S5. The white markings on the MRI slices correspond to the dark lines and text on the histology slide, and can be viewed with larger font in Fig. S5. There is a tear in the tissue adjacent to SRLM in the histology slide, which is not uncommon and is one of the reasons why we chose to perform subfield segmentation in MRI space rather than histology space.

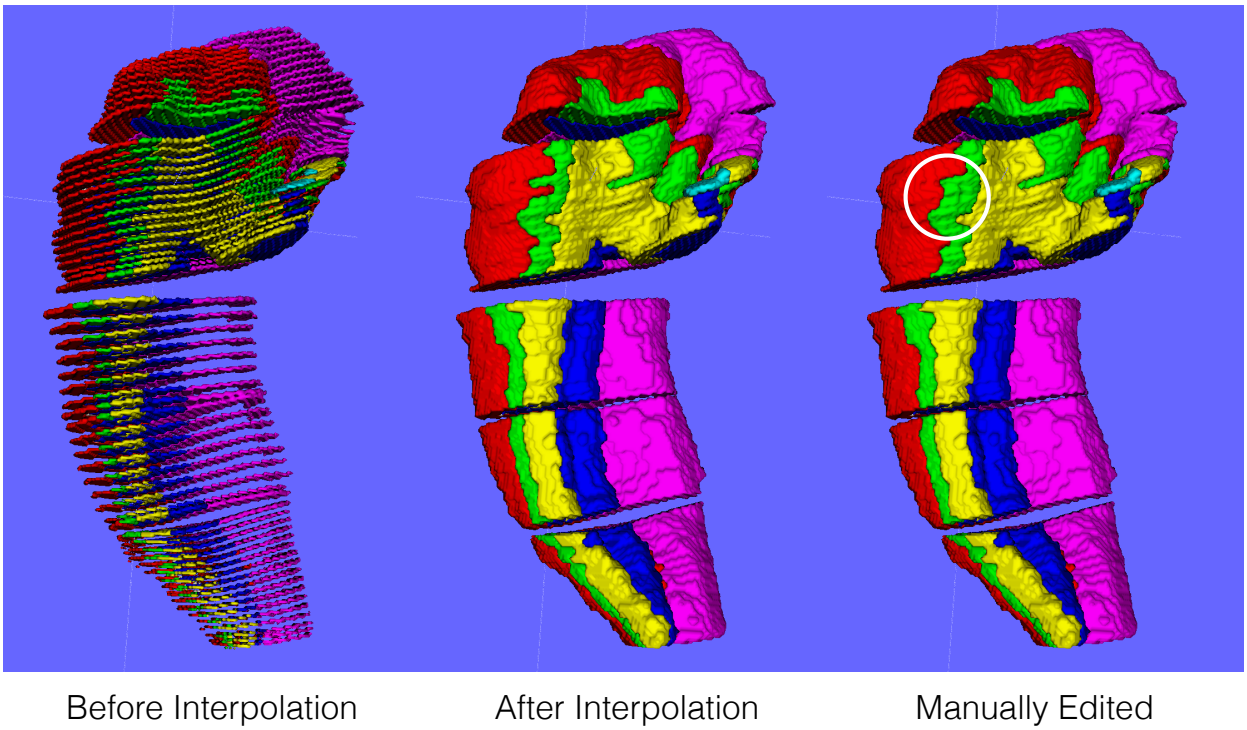


Figure S7: Mapping of histology segmentations performed in block MRI space into whole MRI space for a single specimen. On the left, the segmentations drawn on every third/fifth slice in block space are warped into whole MRI space. In the middle, these warped slice segmentations are interpolated within regions where blocks cover the whole-specimen MRI. On the right, small corrections are introduced manually to improve 3D continuity of the segmentation.

a range of β values and for each β value computed the average generalized Dice coefficient (GDSC) [8] between the consensus segmentation and all of the warped specimen segmentations. We empirically chose the largest β such that mean GDSC with regularization was no more than 0.2% of the mean GDSC without regularization (hence the degree of regularization is very *slight*). The selected value was $\beta = 0.05$.

S1.4.6 Computation of Regional Thickness of Hippocampal Layers

The thickness of three hippocampal layers (DG forming the innermost layer, SRLM forming the middle layer, and the outermost layer composed of CA1-3 and subiculum) was computed as follows. The marching cubes algorithm [18] is applied to the consensus segmentation of the MRI atlas to extract a surface mesh of each layer. The surface mesh is smoothed slightly using the Taubin algorithm [23] implemented in the VCG library [4] to remove sharp voxel edges. The transformations $T_j^{\text{int},M}$ are applied to the surface meshes to warp them into specimen native space for all 31 specimens. The Voronoi skeleton [21] is extracted from each surface mesh, and twice the distance from each mesh vertex to the skeleton is recorded as thickness for that specimen at that vertex. Thickness values are brought back into the atlas space, forming a set of 31 thickness observations at each vertex, which are analyzed as described in the Statistical Analysis Section of the main paper.

S1.4.7 Visualization of Shape Variability

Geodesic shooting [25, 19] was used to visualize shape differences between control and AD hippocampi. As in Step 2 of the MRI atlas generation approach (Section S1.3.4), geodesic shooting was applied between landmarks sampled in the MRI atlas space along the SRLM and hippocampus surfaces and corresponding landmarks in each input specimen. Correspondence between the two sets of landmarks was established by the three-step registration procedure in Section S1.3). Geodesic shooting compactly represents the diffeomorphic transformation between the template landmarks and a given specimen's landmarks as an initial momentum vector at each landmark. By (5), the initial momentum vector at each landmark can be transformed into an initial velocity vector at each landmark, and vice versa. Similar to [25], statistical shape analysis is performed similar on these initial velocity vector fields.

Shape analysis is performed as follows. First, principal components analysis (PCA) is used to derive the mean initial velocity field and a set of orthogonal principal directions, which are the eigenvectors of the covariance matrix of the initial velocity vector fields, ordered by the magnitude of their eigenvalue. The set of principal directions is culled to the P directions with largest eigenvalues that together account for 95% of total variance. Each specimen is then represented by a P-dimensional vector of loadings on these principal directions. A support vector machine (SVM) classifier with a linear kernel [26] was trained using loadings as features and diagnostic group (control or AD; other dementia cases were excluded from the SVM) as class membership. The vector $\mathbf{w}$ orthogonal to the SVM hyperplane was assumed to be the direction in the space of loadings that best discriminates between AD and control. The initial velocity vector field corresponding to $\mathbf{w}$ was computed and geodesic shooting was used to obtain a time-varying velocity field $v^{\mathbf{w}}(x,t)$ over the entire image domain. Integrating the flow equation over this field for different values of t yields family of diffeomorphic transformations of the MRI atlas corresponding to movement along direction $\mathbf{w}$. These transformations are then visualized as a movie representing shape change between control and AD groups.

S1.4.8 Measurement of hippocampal digitations and folds

Hippocampal digitations in the head refer to superior or laterosuperior protuberances of the surface of the hippocampus; note that we are not referring to protuberances of the dentate gyrus. Similarly, folds refer to the protuberances of the surface of the hippocampus in the body and tail, located often on the inferior or latero-inferior surface in the body and sometimes on the lateral surface in the tail. The head digitations were counted in the coronal plane by scrolling anterior and posterior to identify all digitations (See Figure 8a-c). The folds of the posterior hippocampus were mostly counted in the sagittal plane but also in the other planes, especially counting folds in the tail (See Figure 8d-g).The 3D rendering of the hippocampus was used to track all digitations/folds.

[4] http://vcg.isti.cnr.it/vcglib/

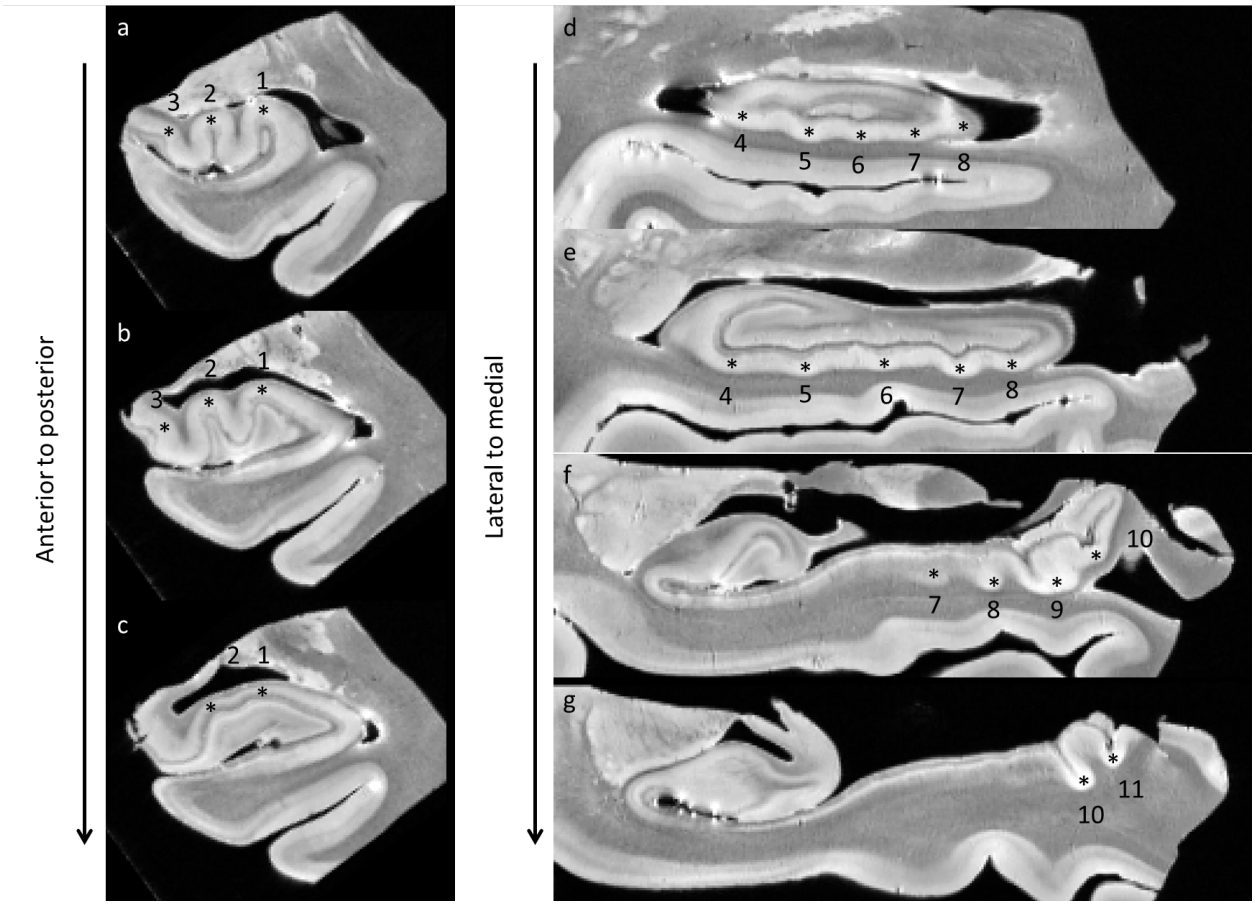


Figure S8: Measurement of hippocampal digitations and folds. The coronal plane, figure a-c, was used to count digitations in the hippocampal head. In this subject 3 head digitations can be counted. In c it can be seen that digitation 2 can no longer be identified, which demonstrates the importance of scrolling back and forth to identify all digitations. In figure d-f, the sagittal plane of the same subject is shown which was primarily used to count folds. Not all folds can be identified on each of the sagittal slices with fold 4-6 disappearing on the two more medial slices, and vice versa for folds 9-11. In this subject the most posterior folds could be identified in the sagittal plane, but this is often not the case. For this reason all three planes were inspected when counting folds.

S1.4.9 Qualitative Analysis of Hippocampal Tail Morphology

Although there is no consensus definition, the hippocampal tail is often defined based on external landmarks which were not available in the excised specimen MRIs. We therefore adopted a different working definition to be able to follow the curve of the tail in a consistent manner. We defined the tail as the posterior 1/3 of the hippocampus. To test the hypothesis that at the macroscopic level the structure of the hippocampal tail is similar to that of the hippocampal body, we defined three cut planes in each specimen MRI that followed the curve of the tail. These cut planes were defined by drawing line segments in the axial plane following rigid alignment of each specimen to the MRI atlas. Line segments were drawn by author LW using the line annotation tool in ITK-SNAP at approximately 6/9, 7/9 and 8/9 of the total hippocampus length and in the direction perpendicular to the lateral CA border. Cut planes were defined as the planes orthogonal to the axial plane that contained these line segments. Each specimen MRI was resliced along the three cut planes and the resulting 2D images were examined visually. Examples from two representative specimens are shown in Figure S14.

S2 Supplemental Results

S2.1 MRI Atlas Quality

Figure S9 shows atlases obtained at different stages and variations of the algorithm in Section S1.3. Atlases labeled "MST, MST+GS and MST+GS+INT" correspond to the three stages in Section S1.3 (minimum spanning tree, geodesic shooting, and intensity-based groupwise registration). The shape averaging effect of the geodesic shooting step is clearly visible, with the shape of the atlas markedly different between MST and MST+GS atlases. Atlas "INT only" is generated to demonstrate that the more conventional computational anatomy approach frequently used for in vivo brain morphometry [14, 15, 5] does not yield as good of an atlas as the proposed approach. The white arrows in the figure point out locations of marked difference between "MST+GS+INT" and "INT only" approaches, with "INT only" generating a less well-defined pattern in the head of the hippocampus, likely due to poor initial matching of SRLM surfaces.

The quality of the registration between individual specimen images and each of the atlases in Figure S9 is visualized in Figure S10. Overall, the warped specimen images for the MST+GS+INT strategy are more similar to each other and to the atlas, than for the INT strategy.

Figure S11 reports quantitative metrics of atlas quality for the three stages in Section S1.3 and the "INT only" approach. For each pair of specimens registered to a given atlas, we compute the Dice similarity coefficient, a measure of relative overlap [9], between their respective hippocampus and SRLM segmentations. For all pairs of specimens with histology, we compute the generalized Dice coefficient (GDSC, a multi-label overlap measure) [8] between their histologically derived multi-label subfield segmentations in atlas space. GDSC is computed across voxels that are labeled in both specimens and fall inside of the hippocampus mask in the given atlas. Lastly, we compute the average normalized cross-correlation (NCC) image dissimilarity metric (with patch size $9 \times 9 \times 9$) between each pair of specimens registered to the atlas. For each pair of specimens, the NCC is integrated over the hippocampus label in MRI atlas space.

The quantitative metrics in Figure S11 require careful interpretation. The hippocampus and SRLM segmentations are directly used by the MST and MST+GS stages of atlas generation, so it is to be expected that overlap for these segmentations would be highest for these two stages. Likewise, the NCC metric is optimized by intensity-based registration, so we expect it to be higher for "MST+GS+INT" and "INT only" approaches. The key observation is that "MST+GS+INT" not only has higher hippocampus/SRLM overlap than "INT only" but also achieves greater NCC. This indicates that the "INT only" approach is initialized farther from the optimal groupwise alignment than the "MST+GS+INT" approach. In other words, the initialization provided by the MST+GS stages helps groupwise intensity-based registration to find a better solution than initialization based on groupwise affine registration. Lastly, "MST+GS+INT" does best at aligning histology labels between specimens.

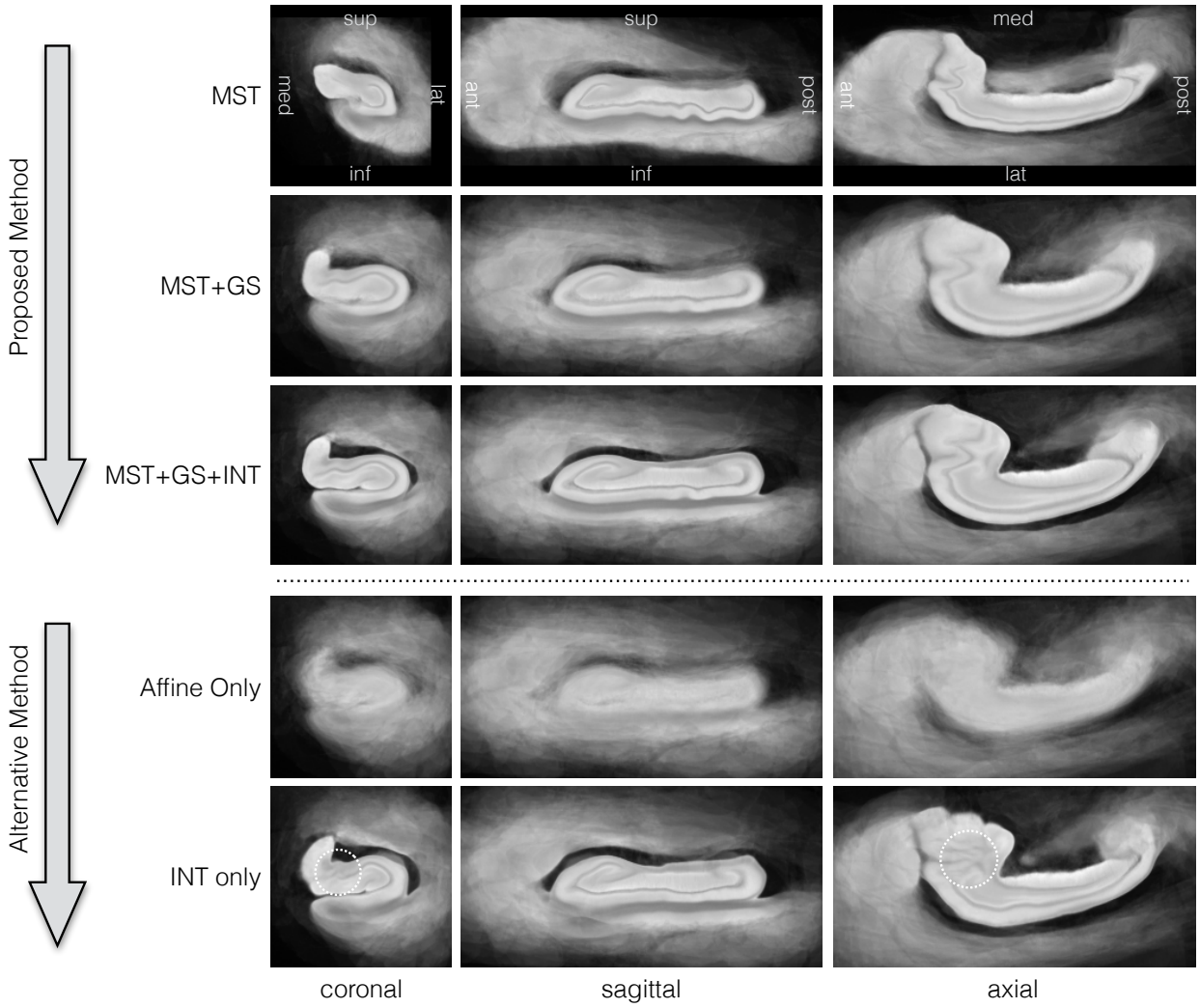


Figure S9: A comparison of MRI atlases generated by the three stages of the atlas building approach in Section S1.3 with an alternative approach. The first three rows show atlases generated at the end of the minimum spanning tree (MST), geodesic shooting (MST+GS) and intensity-based groupwise registration (MST+GS+INT). The alternative approach is to use hippocampus and SRLM labels for affine alignment (Affine only" and then to perform only intensity-based groupwise registration (INT only). Each atlas is generated by averaging the registered specimen MRI scans with equal weighting. The dashed white circles point out regions where the INT template exhibits noticable differences from the MST+GS+INT template.

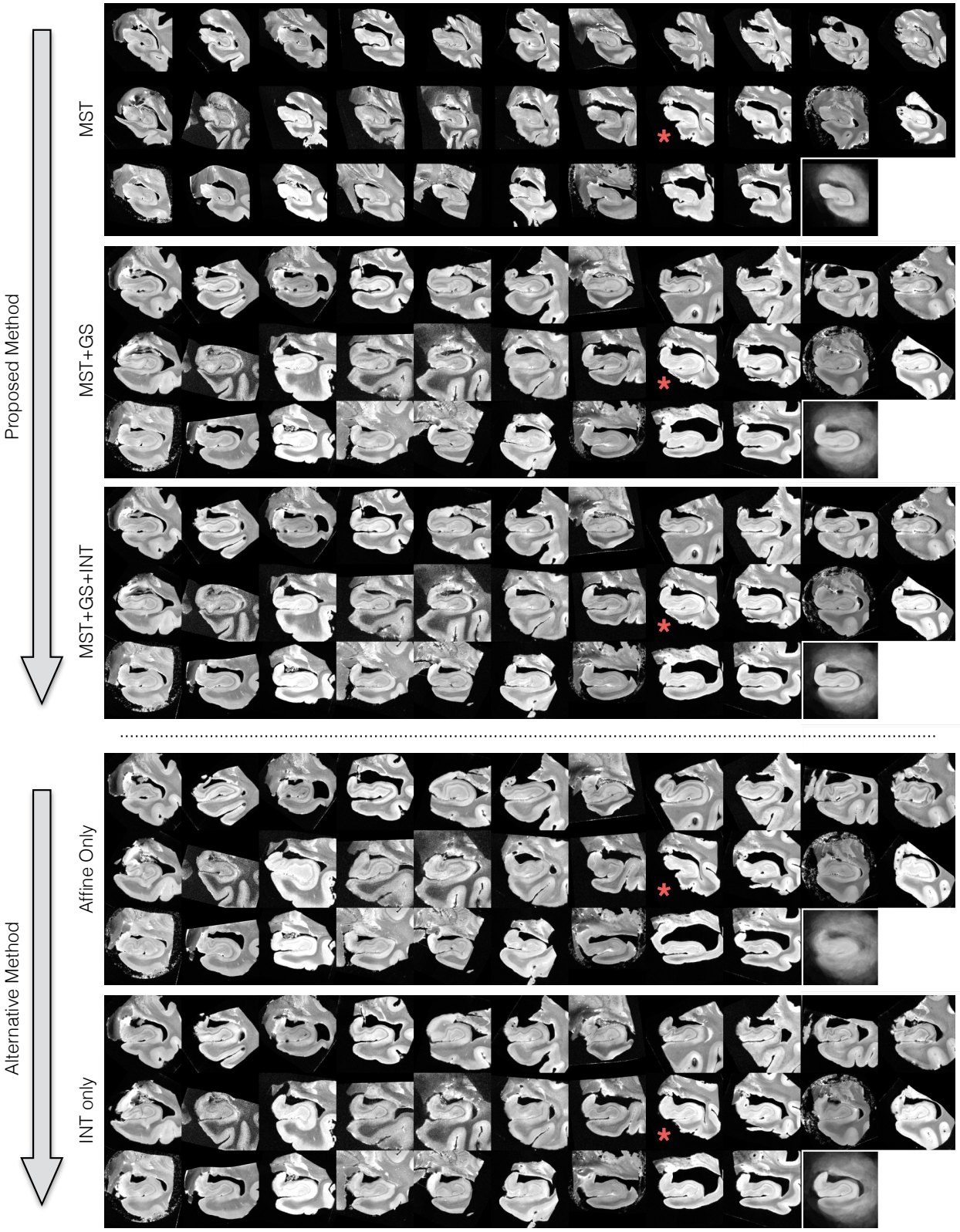


Figure S10: MRI scans of 31 specimens warped into MRI atlas space for each of the five atlases in Figure S9. The more similar the warped images, the better the atlas quality. The specimen that was the root of the MST is marked by the red asterisk.

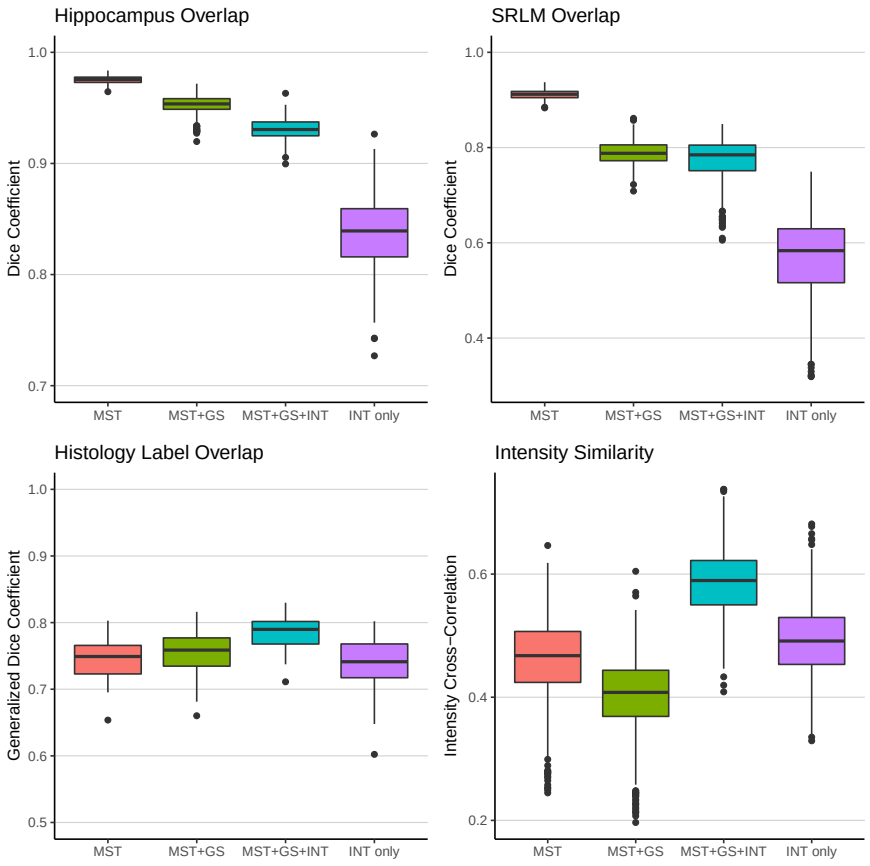


Figure S11: Quantitative metrics of groupwise registration quality after each stage of atlas-building in Section S1.3 and for the alternative "INT only" approach. See text for the description.

S3 Additional Supplemental Figures

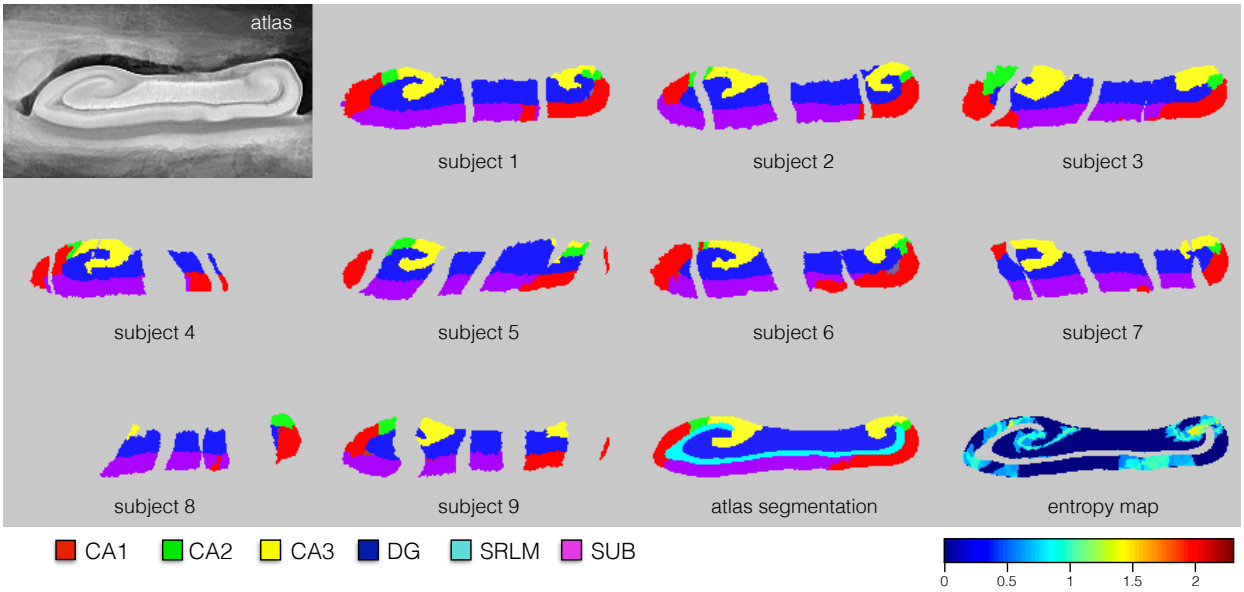


Figure S12: **Consensus Segmentation Labeling of MRI Atlas.** Individual histology-based subfield segmentations mapped into the MRI atlas space and the consensus segmentation. The consensus segmentation includes the SRLM label obtained from MRI-based segmentations of SRLM in the 31 specimens. The last panel shows the entropy map for the consensus segmentation; low values of entropy indicate that the nine constituent segmentations are consistent with each other at a voxel, while higher values of entropy at a voxel indicate greater disagreement between the nine segmentations.

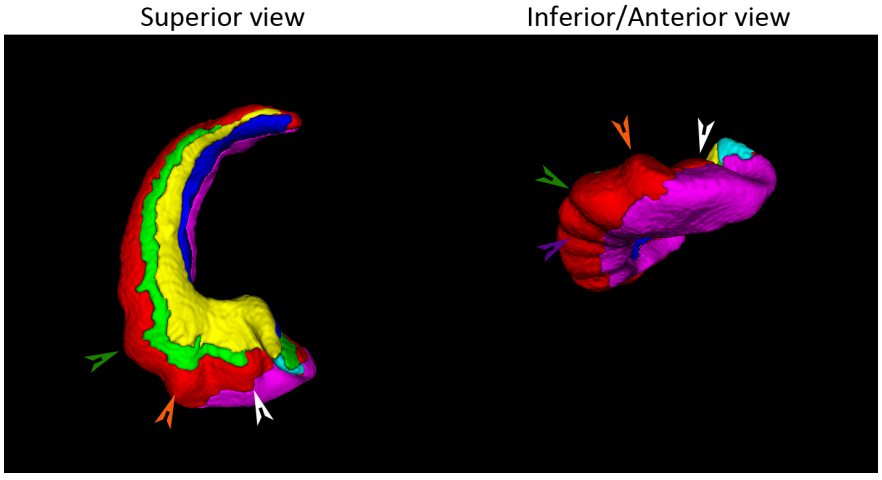


Figure S13: **Hippocampal Digitation Morphology.** 3D rendering from the hippocampus from the superior view (left) and the anterior/inferior view (right). The digitations and folds are indicated by colored arrows, matched between the two renderings. The white and orange arrows point to digitations in the hippocampal head and the purple arrow points to an inferior fold. The green arrow points to a laterally located digitation or fold, suggesting a continuity between the typical head digitations and the typical inferiorly located body folds.

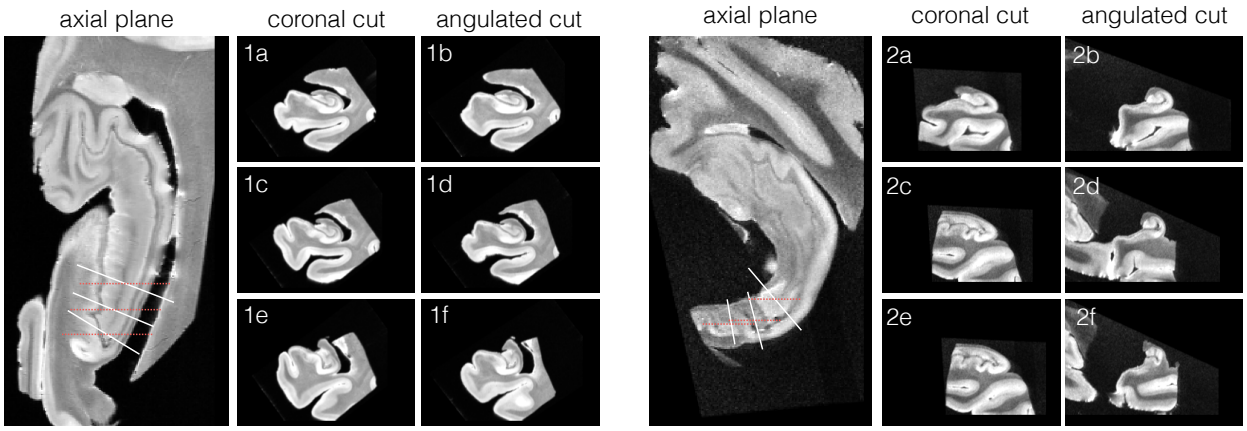


Figure S14: **Hippocampal Tail Morphology.** Examples of tail sections of two specimens are shown, resliced perpendicular to the hippocampal long axis at three locations (white lines in the axial plane) as well as sliced in the typical coronal cut at approximately the same location for comparison (red dashed lines in the axial plane). It can be observed that in the two specimens with differing angulations of the tail, the resliced sections have a "body-like" appearance, in some sections with a slight clockwise rotation (e.g. 1f). The comparison of these angulated cuts with the typical coronal cut shows only small differences in example 1 in which the tail only shows a slight angulation. The differences are more striking for example 2, where the angulation of the tail approximates 90 degrees. It was observed that in some specimens in the most posterior resliced section, there is little DG present, an example is presented in 2f. The appearance of the hippocampus in such sections slightly deviates from the "body-like" appearance.

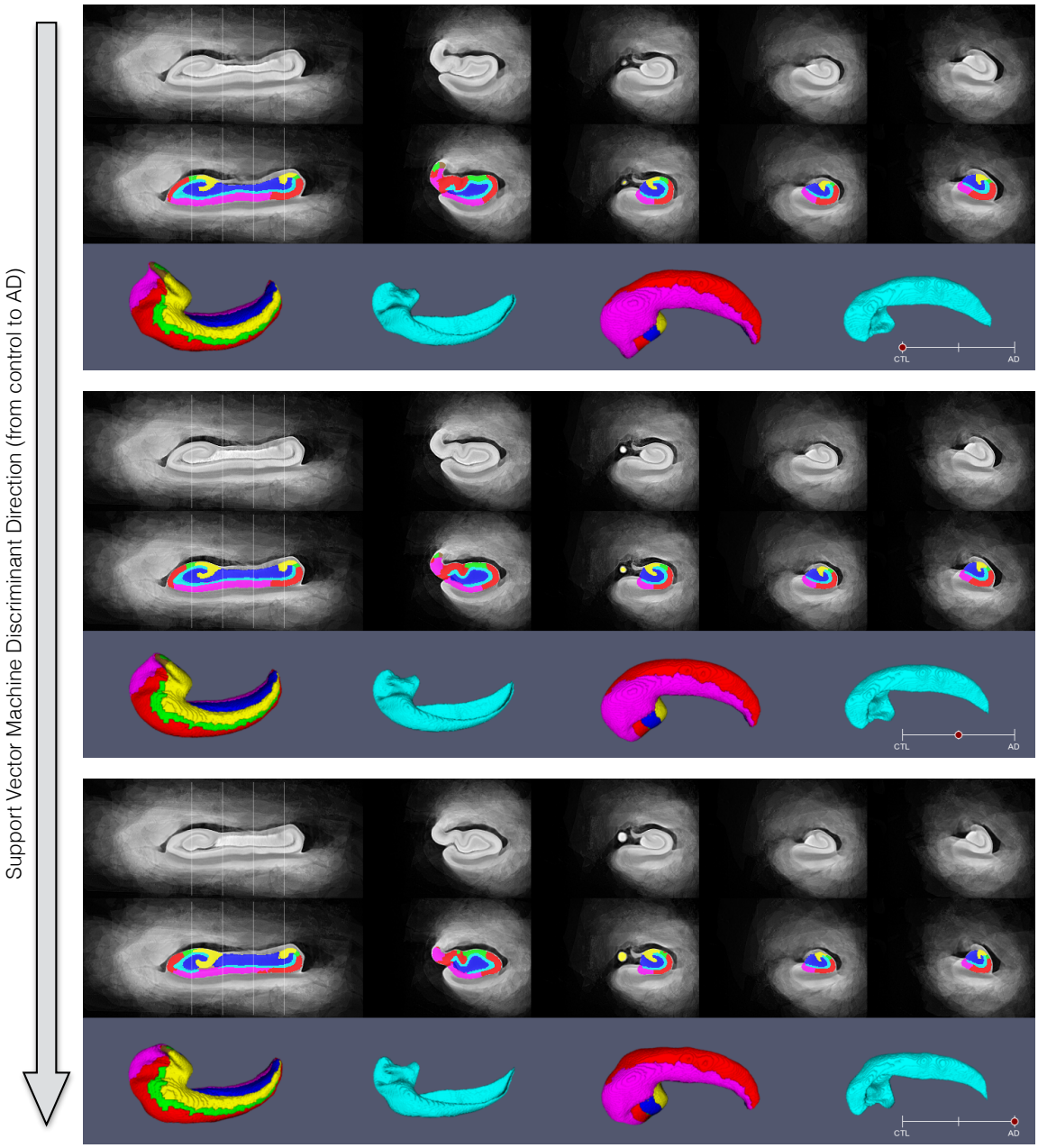


Figure S15: Transformation of the MRI atlas corresponding to moving along the discriminant direction between control and AD specimens.

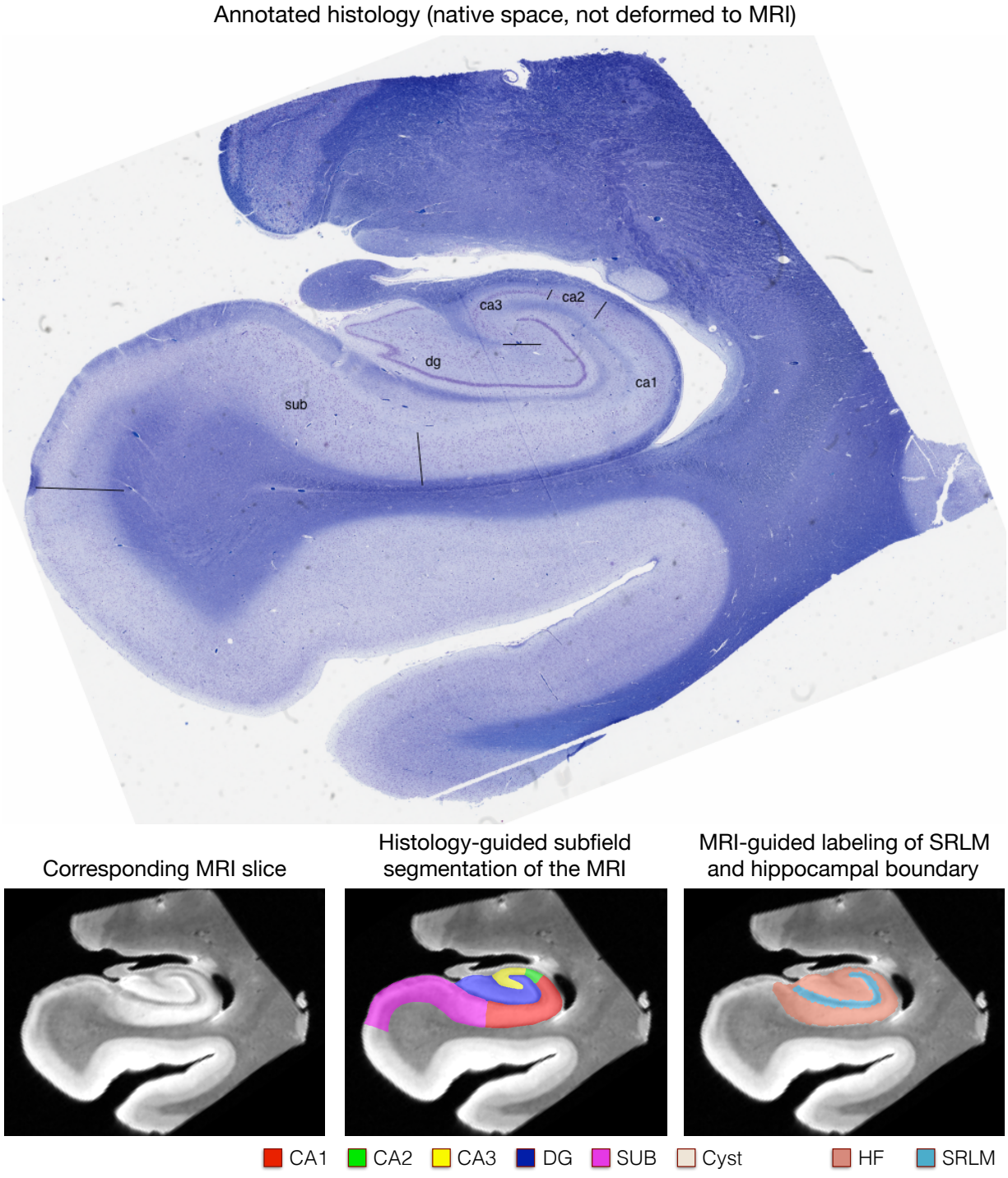


Figure S16: **Histology-Guided MRI Segmentation Example 1.** As a complement to Figures S5 and S6, this figure and the following show a large-scale view of the annotated histology slide (as in Figure S5), and the corresponding MRI slice overlaid with the histology-guided manual tracing of the hippocampal subfields in MRI space (as in Figure S6), as well as with the MRI-guided semi-automatic segmentation of the SRLM and whole hippocampal boundary (Figures S1 and S6). The large-scale histology slice is shown in its native space, before deformation to the MRI slice.

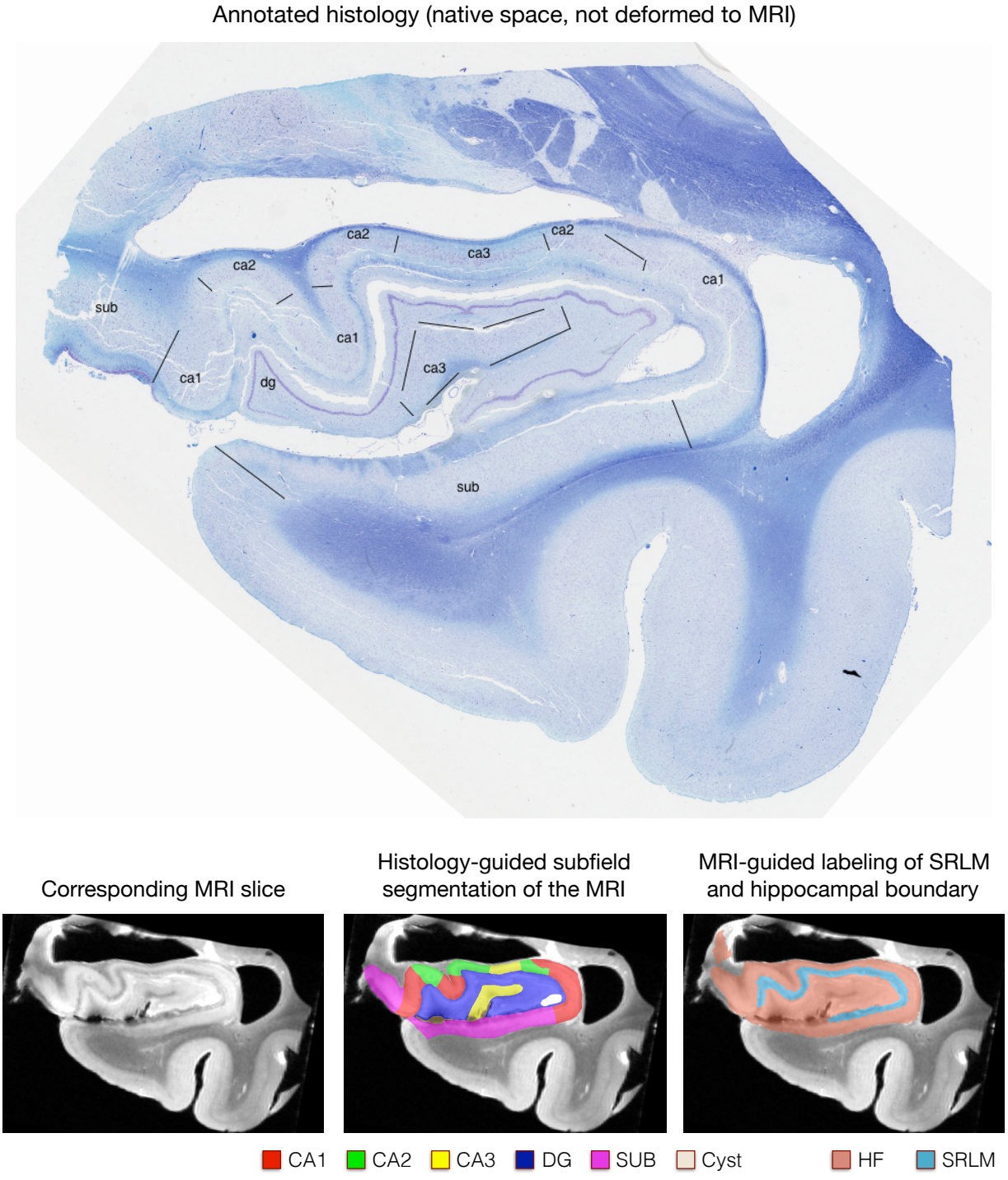


Figure S17: **Histology-Guided MRI Segmentation Example 2.**

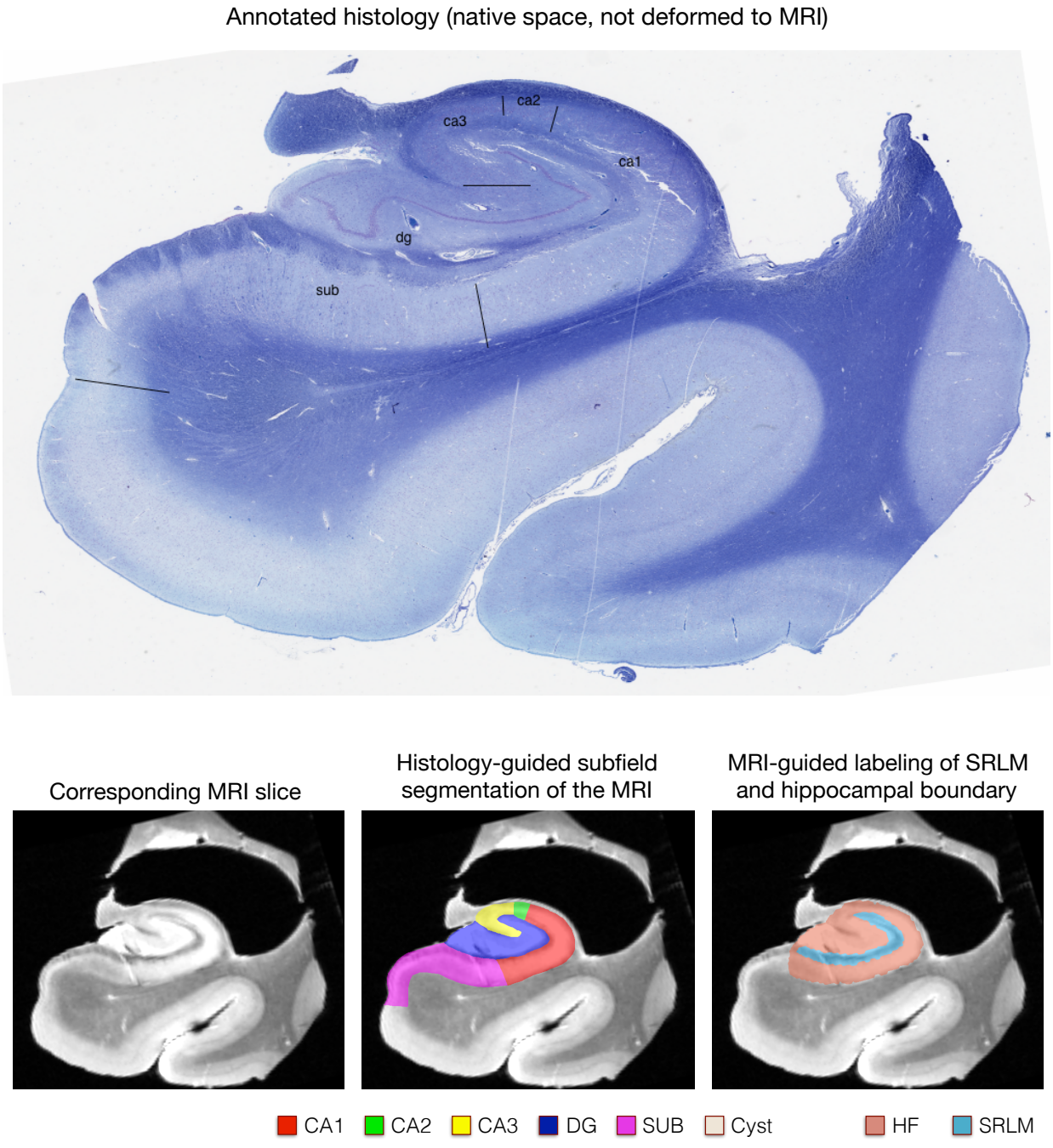


Figure S18: **Histology-Guided MRI Segmentation Example 3.**

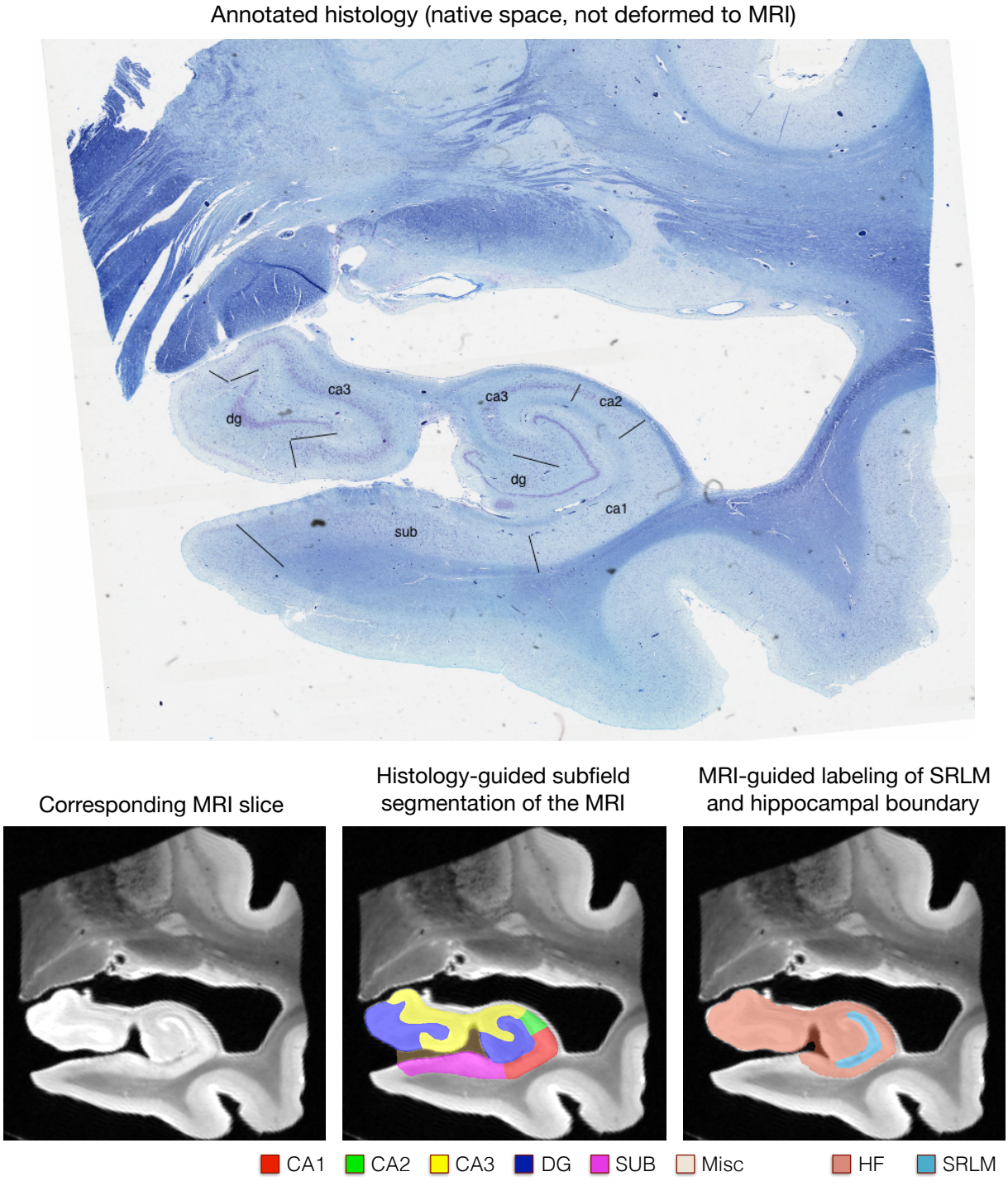


Figure S19: **Histology-Guided MRI Segmentation Example 4.**

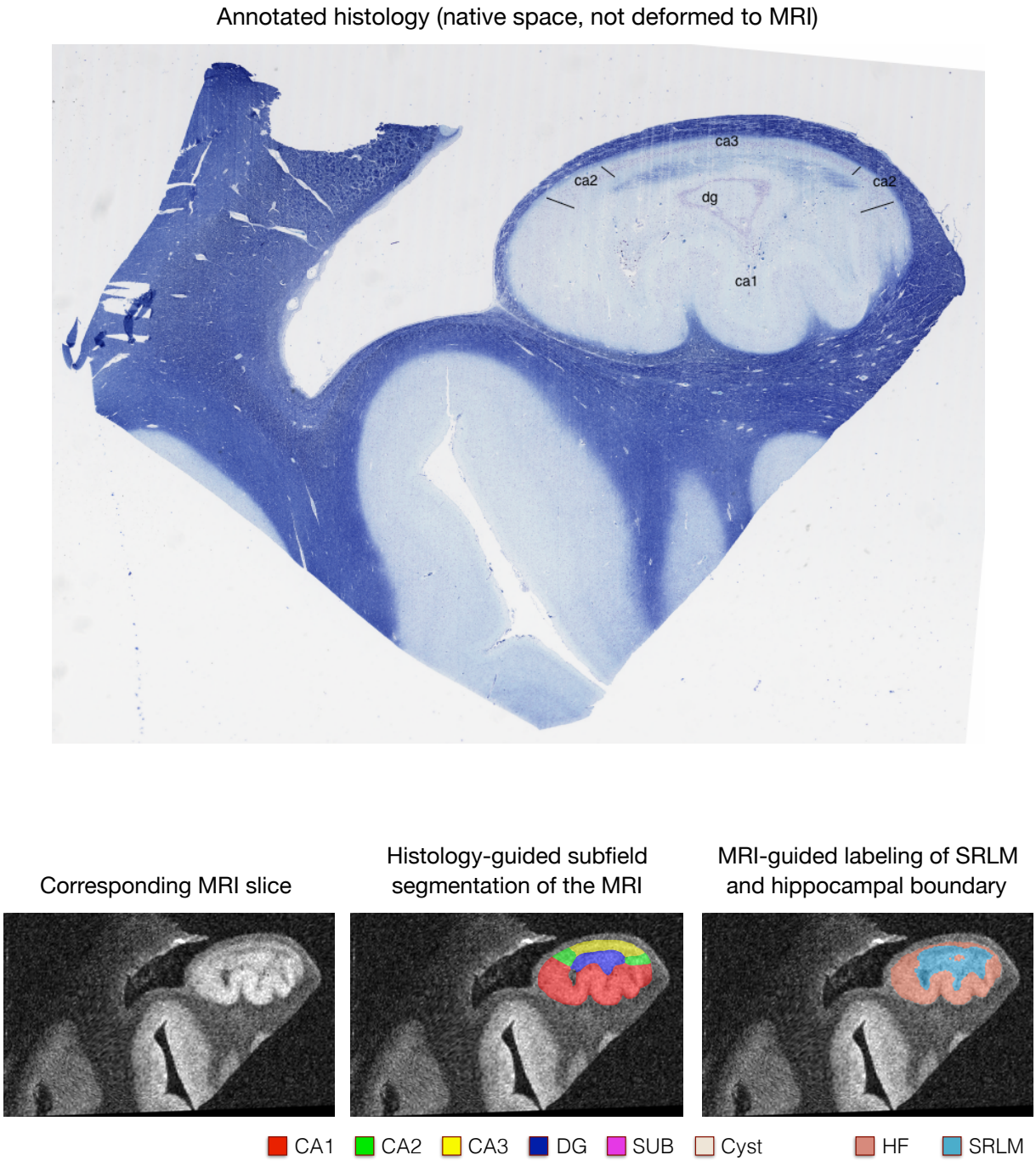


Figure S20: **Histology-Guided MRI Segmentation Example 5.**

References

[1] Daniel H Adler. *A Postmortem Computational Atlas of the Human Hippocampus*. PhD thesis, University of Pennsylvania, 2016.

[2] Daniel H Adler, John Pluta, Salmon Kadivar, Caryne Craige, James C Gee, Brian B Avants, and Paul A Yushkevich. Histology-derived volumetric annotation of the human hippocampal subfields in postmortem MRI. *Neuroimage*, 84:505–23, Jan 2014.

[3] Stéphanie Allassonnière, Alain Trouvé, and Laurent Younes. Geodesic shooting and diffeomorphic matching via textured meshes. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 365–381. Springer Berlin Heidelberg, 2005.

[4] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*, 12(1):26–41, Feb 2008.

[5] Brian B. Avants and James C. Gee. Shape averaging with diffeomorphic flows for atlas creation. In *Proc IEEE Int Symp Biomed Imaging*, pages 595–598, 2004.

[6] Brian B Avants, Paul Yushkevich, John Pluta, David Minkoff, Marc Korczykowski, John Detre, and James C Gee. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage*, Oct 2009.

[7] Massimiliano Corsini, Paolo Cignoni, and Roberto Scopigno. Efficient and flexible sampling with blue noise properties of triangular meshes. *IEEE Transactions on Visualization and Computer Graphics*, 18(6):914–924, 2012.

[8] William R Crum, Oscar Camara, and Derek L G Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging*, 25(11):1451–61, Nov 2006.

[9] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, July 1945.

[10] Song-Lin Ding and Gary W Van Hoesen. Organization and detailed parcellation of human hippocampal head and body regions based on a combined analysis of cyto- and chemoarchitecture. *J Comp Neurol*, 523(15):2233–53, Oct 2015.

[11] I. Dryden and K. Mardia. *Statistical shape analysis*. John Wiley & Sons, New York, 1998.

[12] H. Duvernoy. *The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections with MRI*. Berlin Heidelberg: Springer, fourth edition, 2013.

[13] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Interventation—MICCAI'98*, pages 130–137. Springer, 1998.

[14] Alexandre Guimond, Jean Meunier, and Jean-Philippe Thirion. Average brain models: a convergence study. *Comput. Vis. Image Underst.*, 77(9):192–210, 2000.

[15] S. Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23 Suppl 1:S151–S160, 2004.

[16] H. Kluever and E. Barrera. A method for the combined staining of cells and fibers in the nervous system*. *Journal of Neuropathology & Experimental Neurology*, 12(4):400, 1953.

[17] Vladimir Kolmogorov and Ramin Zabin. What energy functions can be minimized via graph cuts? *IEEE transactions on pattern analysis and machine intelligence*, 26(2):147–159, 2004.

[18] W. E. Lorensen and H. E. Cline. Marching cubes: a high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.

[19] Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *J Math Imaging Vis*, 24(2):209–228, Jan 2006.

[20] James O'Callaghan, Jack Wells, Simon Richardson, Holly Holmes, Yichao Yu, Simon Walker-Samuel, Bernard Siow, and Mark F Lythgoe. Is your system calibrated? MRI gradient system calibration for pre-clinical, high-resolution imaging. *PLoS One*, 9(5):e96568, 2014.

[21] R. L. Ogniewicz and O. Kübler. Hierarchic Voronoi skeletons. *Pattern Recognit*, 28(3):343–359, 1995.

[22] Robert Clay Prim. Shortest connection networks and some generalizations. *Bell Labs Technical Journal*, 36(6):1389–1401, 1957.

[23] Gabriel Taubin. Curve and surface smoothing without shrinkage. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 852–857. IEEE, 1995.

[24] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*, 29(6):1310–1320, Jun 2010.

[25] M Vaillant, M I Miller, L Younes, and A Trouvé. Statistics on diffeomorphisms via tangent space representations. *Neuroimage*, 23 Suppl 1:S161–9, 2004.

[26] V. N. Vapnik. *Statistical learning theory*. John Wiley & Sons, 1998.

[27] Hongzhi Wang, Jung W Suh, Sandhitsu R Das, John Pluta, Caryne Craige, and Paul A Yushkevich. Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Anal Mach Intell*, 35(3):611–623, Mar 2013.

[28] Laurent Younes, Felipe Arrate, and Michael I. Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1, Supplement 1):S40 – S50, 2009. Mathematics in Brain Imaging.

[29] Paul A Yushkevich, Brian B Avants, John Pluta, Sandhitsu Das, David Minkoff, Dawn Mechanic-Hamilton, Simon Glynn, Stephen Pickup, Weixia Liu, James C Gee, Murray Grossman, and John A Detre. A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 t. *Neuroimage*, 44(2):385–98, Jan 2009.

[30] Paul A Yushkevich, Yang Gao, and Guido Gerig. Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3342–3345. IEEE, 2016.

[31] Paul A Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, Jul 2006.