A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology

Iglesias, Juan Eugenio; Insausti, Ricardo; Lerma-Usabiaga, Garikoitz; Bocchetta, Martina; Van Leemput, Koen; Greve, Douglas N.; van der Kouwe, Andre; Fischl, Bruce; Caballero-Gaudes, César; Paz-Alonso, Pedro M.

Published in:
NeuroImage

Link to article, DOI:
10.1016/j.neuroimage.2018.08.012

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology

Juan Eugenio Iglesias, Ricardo Insauti, Garikoitz Lerma-Usabiaga, Martina Bocchetta, Koen Van Leemput, Douglas N. Greve, Andre van der Kouwe, the Alzheimer's Disease Neuroimaging Initiative, Bruce Fischl, Cesar Caballero-Gaudes, Pedro M. Paz-Alonso

ARTICLE INFO

Keywords:
Thalamus
Atlasing
Histology
Ex-vivo MRI
Segmentation
Bayesian inference

ABSTRACT

The human thalamus is a brain structure that comprises numerous, highly specific nuclei. Since these nuclei are known to have different functions and to be connected to different areas of the cerebral cortex, it is of great interest for the neuroimaging community to study their volume, shape and connectivity in vivo with MRI. In this study, we present a probabilistic atlas of the thalamic nuclei built using ex vivo brain MRI scans and histological data, as well as the application of the atlas to in vivo MRI segmentation. The atlas was built using manual delineation of 26 thalamic nuclei on the serial histology of 12 whole thalami from six autopsy samples, combined with manual segmentations of the whole thalamus and surrounding structures (caudate, putamen, hippocampus, etc.) made on in vivo brain MR data from 39 subjects. The 3D structure of the histological data and corresponding manual segmentations was recovered using the ex vivo MRI as reference frame, and stacks of blockface photographs acquired during the sectioning as intermediate target. The atlas, which was encoded as an adaptive tetrahedral mesh, shows a good agreement with previous histological studies of the thalamus in terms of volumes of representative nuclei. When applied to segmentation of in vivo scans using Bayesian inference, the atlas shows excellent test-retest reliability, robustness to changes in input MRI contrast, and ability to detect differential thalamic effects in subjects with Alzheimer's disease. The probabilistic atlas and companion segmentation tool are publicly available as part of the neuroimaging package FreeSurfer.

1. Introduction

The thalamus is a diencephalic structure located between the cortex and the midbrain. Traditionally, the thalamus has been considered primarily a link in the flow of sensory signals (i.e., analogous to a relay station or transponder), through its white matter connections to virtually the entire cortex (Johansen-Berg et al., 2005). However, current views suggest that the thalamus is more than a simple transponder, and continues to contribute to the processing of information within cortical hierarchies (Sherman, 2007, 2016). Among other functions, the thalamus is involved in the regulation of consciousness, sleep and alert states; the motor system; and spoken language (Sherman and Guillery, 2001). The study of these functions with MRI has attracted wide attention from the neuroimaging community (Fernández-Espejo et al., 2010; Czisch et al., 2004; Guye et al., 2003; Binder et al., 1997), and so has the in vivo study of pathologies associated with the thalamus, such as schizophrenia.
Segmentation of the whole thalamus in structural MRI is a prerequisite for most MRI-based studies of this structure, and many methods have been developed to produce automated segmentations. Fischl et al. (2002) used a voxel-wise probabilistic atlas of anatomy and MRI intensities to segment the thalamus, along with a number of other brain structures; this method is implemented in the widespread, open-source package FreeSurfer (Fischl, 2012). Patenaude et al. (2011) used a combined model of shape and appearance, also to segment a set of brain structures including the thalamus; an implementation of this method (FIRST) is available as part of the popular FSL package (Smith et al., 2004). A number of standard segmentation algorithms have also been applied to thalamus segmentation in structural MRI, such as multi-atlas segmentation (Fischl et al., 2006), fuzzy clustering (Amini et al., 2004), voxel classification (Zikri et al., 2013) or Bayesian segmentation (Puonti et al., 2016).

However, the thalamus is not a homogeneous structure; it consists of several nuclear masses, which serve multiple and highly specific functions. While many studies agree on the existence of 14 major nuclei, these can be subdivided histologically into many more subnuclei, so the exact number depends on the level of detail of the classification (Jones, 2012; Morel, 2007; Mai and Forutan, 2012). In vivo segmentation of these nuclei in MRI can enable neuroimaging studies of the thalamus (e.g., morphometry; structural and functional connectivity) at a much higher level of specificity, and also has the potential to provide more accurate surgical planning and more precise placement of deep brain stimulation (DBS) devices. These applications have sparked the interest of the neuroimaging community in automated segmentation algorithms for the thalamic nuclei.

Many thalamic nuclei segmentation methods have been based on applying clustering techniques to diffusion MRI data, which provide more contrast between the nuclei than structural MRI scans – despite their lower resolution. A subset of these techniques have focused on the local diffusion properties of the tissue. For example, Mang et al. (2012) clustered the voxels inside the thalamus into 21 predefined groups using the direction of the leading eigenvector of the diffusion tensor. Duan et al. (2007) used the mean shift algorithm (Fukunaga and Hostetler, 1975) with the Frobenius distance between the tensors. Wiegell et al. (2003) used the k-means algorithm to create 14 clusters, using a distance function that was a linear combination of the Mahalanobis voxel distance and the Frobenius tensor distance. Jonasson et al. (2007) also used k-means, but only to initialize surfaces that then evolved with the level set method; the cluster prototypes were given by the tensors minimizing the variation within the groups. Recently, Battistella et al. (2017) moved away from the tensor model and used a spherical harmonic representation of the full orientation distribution function, to cluster the voxels with k-means.

Local diffusion information is typically insufficient to discriminate between thalamic nuclei. In their pioneering work, Behrens et al. (2003) and Johansen-Berg et al. (2005) used probabilistic tractography to parcellate the thalamus according to the connectivity of its voxels with predefined target regions of the cerebral cortex. This approach and its extensions (e.g., Rasenburgh et al., 2016; Abbivardi and Bach, 2017) have been used in multiple applications, such as the definition of targets in DBS (Akram et al., 2018; Middlebrooks et al., 2018). Other approaches have relied on clustering connectivity patterns to parcellate the thalamus; the connectivity can be structural (i.e., derived from diffusion MRI), as in Lambert et al. (2017), or functional (i.e., estimated with resting state functional MRI), as in Ji et al. (2016); Hale et al. (2015).

Other thalamic parcellation efforts have relied on supervised machine learning techniques. For example, Stough et al. (2014) used local measures (fractional anisotropy, eigenvector directions) and connectivity with cortical regions as input features to a random forest (Breiman, 2001) in order to divide the thalamus into six major nuclei. Supervision beyond the use of cortical regions for connectivity-based parcellation ameliorates the main disadvantage of unsupervised methods, which is the lack of correspondence of the clusters with the underlying anatomy – even if an expert attempts to manually map clusters to nuclei, as in Wiegell et al. (2003).

Creating ground truth segmentations on in vivo scans can be difficult, particularly when one tries to segment thalamic nuclei defined at finer levels of division – when the resolution (especially in diffusion MRI) and contrast are insufficient. For this reason, some works have used histology to create atlases of the thalamus. Krauth et al. (2010) used manual segmentations on the histology of six cases (Morel, 2007) in order to compute an average thalamus with 42 nuclei. In follow-up work, Jakab et al. (2012) mapped their average to MNI space to be able to use it in registration-based segmentation. Sadikot et al. (2011) used a single labeled case, which they also mapped to a reference space (Collin27, Holmes et al., 1998) to be able to segment new cases. These approaches inherit the limitations of registration-based segmentation: the inability of a single subject-segmentation pair to cover the spectrum of variability of a larger population, and the inability to accurately register across MRI contrasts.

In this work, we present a probabilistic atlas of the human thalamus and its nuclei, as well as surrounding anatomy (the latter is important to enable segmentation of thalamic nuclei using Bayesian inference). The atlas was derived from manual segmentations of the thalamic nuclei on the histological images of 12 whole thalamus from six autopsy samples, as well as delineations of the whole thalamus and surrounding structures (e.g., caudate, putamen, hippocampus) in 39 T1 scans acquired at standard resolution (i.e., 1 mm). The 3D reconstruction of the histology was assisted by high resolution ex vivo MRI scans and blockface photographs. Compared with previous histology-based atlases of the thalamus (Krauth et al., 2010; Sadikot et al., 2011), our proposed atlas is probabilistic, models surrounding anatomy, and can be used in combination with Bayesian inference in order to directly segment MRI scans of arbitrary contrast.

2. Materials and methods

2.1. Ex vivo specimens

In order to build the proposed atlas, we used data from six post mortem cases from the body donor program of the University of Castilla - La Mancha (UCLM) Medical School (Albacete, Spain). Informed consent was obtained from the donors, following the Declaration of Helsinki. The use of this brain tissue for research purposes was approved by the Ethics Committee of the University Hospital of Albacete.

The demographic data of the cases is shown in Table 1. None of the donors had a history of disease that affected the morphology of the brain (or the thalamus): their clinical history did not include any confounding factors such substance abuse or pathology (e.g., dementia), and the weight of the specimens was approximately average. We acknowledge that the age range of these specimens is older than that of most neuroscience experiments. However, this issue is mitigated by the lack of pathology in the samples, as well as by the fact that our probabilistic atlas

<table>
<thead>
<tr>
<th>Case</th>
<th>Age at death</th>
<th>Gender</th>
<th>Brain weight</th>
<th>PMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNL4_13</td>
<td>97</td>
<td>male</td>
<td>1.238 Kg</td>
<td>9h</td>
</tr>
<tr>
<td>HNL7_14</td>
<td>98</td>
<td>female</td>
<td>1.168 Kg</td>
<td>6h</td>
</tr>
<tr>
<td>HNL5_13</td>
<td>59</td>
<td>male</td>
<td>1.020 Kg</td>
<td>N/A</td>
</tr>
<tr>
<td>HNL8_14</td>
<td>61</td>
<td>female</td>
<td>1.409 Kg</td>
<td>10h</td>
</tr>
<tr>
<td>HNL14_15</td>
<td>87</td>
<td>male</td>
<td>1.100 Kg</td>
<td>2h 30 m</td>
</tr>
<tr>
<td>HNL16_16</td>
<td>84</td>
<td>male</td>
<td>1.264 Kg</td>
<td>3h 30 m</td>
</tr>
</tbody>
</table>
also includes an in vivo dataset of brain scans of 39 subjects (including several younger controls), as explained in Section Atlas Construction below. This strategy of combining these 39 in vivo scans with ex vivo images from elderly subjects has been proven successful by our earlier works on the hippocampus (Iglesias et al., 2015a) and the amygdala (Saygin et al., 2017).

The fixation of all the brain samples was performed in situ by personnel of the UCLM Human Neuroanatomy Laboratory, by neck dissection of both primitive carotids in the lower third of the neck, followed by cannulation of the carotids. The fixation started with a flush of 4 l of saline, followed by 8 l of 4% paraformaldehyde in phosphate buffer (pH 7.4). In order to allow the fixative to flow, the internal jugular vein was sectioned on one side. After perfusion, the brain was left in situ for 48 h, and subsequently extracted following standard autopsy procedures. Postfixation until scanning was carried out by storage in a container filled with 4% paraformaldehyde. This in situ fixation method better preserves the shape of the individual brain, fitting exactly the intracranial shape (as opposed to a generic container), and minimizes the impact of the extraction procedure on the probabilistic atlas to be built.

2.2. MRI scanning and processing

The fixed samples were transferred to the Basque Center on Cognition, Brain and Language (BCBL, Donostia - San Sebastian, Spain) for MRI scanning. Ex vivo MR images of the whole brains were acquired on a 3 T Magnetom TIM Trio scanner with a 12 channel receiver coil. Despite its reduced efficiency compared with the 32 channel counterpart, the 12 channel coil enables acquisition at higher resolution without running out of RAM in the image reconstruction. The brains were scanned in vacuum bags filled with Fluorinert FC-3283 (3M, Maplewood, MN, U.S.A.), in order to minimize the negative impact of air bubbles and susceptibility artifacts. The images were acquired with a 3D multi-slab balanced steady-state free precession sequence (McNab et al., 2009) with TE/TR = 5.3/10.6 ms and flip angle 35°. Four axial slabs with 112 slices each were used to cover the whole volume of the brains, and 57% slice oversampling was used in order to minimize slab aliasing. The resolution of the scans was 0.25 mm isotropic, with matrix size 720 × 720 × 448 voxels (axial). MR images were acquired with four different RF phase increments (0, 90, 180, 270°) and averaged to reduce banding artifacts. The time of acquisition per phase was 90 min. Ten repetitions of this protocol were acquired for increased signal-to-noise ratio (SNR). The total length of the protocol was thus 60 h.

Combined with the 12 channel receiver coil, the multi-slab acquisition described above enabled us to bypass the memory limitations of our clinical scanner when reconstructing the images, while preserving the SNR efficiency of 3D acquisitions. However, this type of acquisition also introduces slab boundary artifacts at the interfaces between the slabs. After computing a brain mask with simple Otsu thresholding (Otsu, 1975), such artifacts were corrected simultaneously with the bias field using a Bayesian method (Iglesias et al., 2016). Sample slices of the MRI scans are shown in Fig. 1.

2.3. Histological analysis

After MRI scanning, the specimens were transferred back to the UCLM Human Neuroanatomy Laboratory for histological analysis. First, a brain sectioning knife was used to make a transverse cut perpendicular to the intercommisural line. Using this plane of section, parallel blocks (thickness 10–14 mm) were extracted, from the frontal to the occipital pole. For each case and hemisphere, and in order to avoid incomplete sectioning of the whole thalamus, the anterior limit of the first block was sectioned at the level of the anterior hypothalamus and head of the caudate nucleus. The blocks containing thalamic tissue (three or four, depending on the case) were selected for histological analysis, numbered and subsequently immersed in cryoprotectant (Glycerol 10% and subsequently 20%) during 120 h at 4°C temperature.

After this procedure, the tissue blocks were sectioned at 50 μm intervals using a sliding microtome, which was coupled to a freezing unit and covered in dry ice. A blockface photograph was taken before each cut using an Olympus E-420 camera mounted on a shelf over the microtome; these photographs are useful for post-hoc 3D reconstruction. One every ten sections (i.e., every 0.5 mm) was selected for Nissl staining with thionin and posterior cytoarchitectonic analysis and segmentation. The other nine were preserved for future analyses with complementary techniques. The selected, Nissl-stained sections were digitized at 4 μm resolution using an Epson Perfection V800 Photo flatbed scanner. A corresponding pair of blockface and histological images are shown in Fig. 2.

2.4. Blockface photograph processing

In order to be useful in histology reconstruction, the blockface images need to be registered and perspective corrected, such that, when stacked, they render a volume that is 3D consistent. Registration is needed to correct for the perspective distortion introduced by small movements of the microtome and the camera setup. In order to co-register the images, we used as reference a photograph of the microtome without any sample on the block holder. On this photograph, we manually marked the corners of the block holder, and used them to define a binary mask covering the whole image domain except for the block holder – so that the registration is not influenced by the brain samples. The registration was performed by detecting salient points inside the mask with SURF (Bay et al., 2006), matching their (SURF) feature vectors, and robustly optimizing a homography transformation with RANSAC (Fischler and Bolles, 1987). In order to introduce very salient points and thus ease the registration, we glued a checkerboard pattern and round stickers in different colors to the microtome (see Fig. 2a).

After registration, we used an homography transform to correct for scaling and geometric perspective distortion. The homography was computed by matching the four corners of the block holder in the reference image to a rectangular grid of size equal to the holder’s physical dimensions, defined at 0.1 mm resolution. The transform was pre-computed using the reference photograph, and then applied to all the other blockface images in order to extract perspective corrected images of known resolution. A sample blockface photograph and its corrected counterpart are shown in Fig. 2a and b.

Finally, we segmented the tissue from the underlying block holder (which was surrounded by dry ice) using a random forest pixel classifier
tion within a block, in order to correct for the tissue shrinkage, estimating two sets of transforms: 1. nonrigid registrations of each section (which are stained and mounted on glass slides; and 2. rigid registrations that align the blocks with each other. Computing these transforms – particularly the nonlinear – using only histological data is an ill-posed problem: without any additional information, we can only resort to registering each section with its neighbors, which is well known to cause geometric distortions such as the “banana effect” (straightening of curved structures) and “z-shift” (accumulation of errors along the stack); see further details in Pichat et al. (2018).

A better alternative is to guide the reconstruction with 3D consistent data acquired with another modality, typically MRI. However, using only the MRI volume to 3D reconstruct the histology is still complicated, since solving for the pose of the blocks and the nonlinear registrations simultaneously is also ill posed (Malandain et al., 2004). More precise solutions can be achieved by using 3D consistent images of the whole blocks as intermediate target, as these images can be linearly registered to the MRI to obtain the pose of the blocks, i.e., playing the role of “stepping stones” between the histology and the MRI.

For example, Adler et al. (2018) used MRI scans of the blocks as intermediate data. This approach makes the registration to the original MRI scan of the whole brain easier, since it is an intra-modality problem. However, it has the disadvantage that it still requires estimating a rigid transform between the MRI scan of the block and the histology stack, which might not be straightforward. Here, we followed Amunts et al. (2013) instead, and used the stacks of (perspective corrected, segmented) blockface photographs. This choice has the advantage that the exact correspondence between the histological sections and photographs in the stack is known. On the other hand, the registration between the intermediate images and the original MRI is slightly harder than in Adler et al. (2018) because it is inter-modality, but, since it is a rigid registration problem, mutual information works well.

More specifically, we first rigidly aligned the stacks of photographs to the whole brain MRI, using the segmentations produced by the Otsu thresholding (MRI) and the pixel classifier (photographs) to mask the cost function, which used mutual information. For this purpose, we used an iterative algorithm, in which we alternately: updated a global registration of the whole brain MRI to the set of blocks; and refined the set of individual block transforms that aligned them with the MRI. The pose of the whole brain MRI was initialized by coarse manual alignment. The rigid co-registration algorithm is illustrated with an example in Fig. 3.

Given the pose of the blocks, it is still necessary to compute the nonlinear registration of each stained section. For this purpose, we first took the photograph to which the section corresponded, resampled the MRI volume onto it with the corresponding rigid transform, and masked it with the corresponding mask, given by the pixel classifier. This resampled MRI was used as target to register the corresponding histological section. We used Elastix (Klein et al., 2010) for the registration, combining mutual information with a B-spline transform (control point spacing: 3 mm).

We found the registration based solely on image intensities not to be robust enough for our application, particularly when tears and folds had occurred when mounting the tissue on the slide. To increase the robustness, pairs of corresponding landmarks were manually placed by J.E.I. on the images (between 6 and 18 per pair of images). Quality control on landmark placement was based on outlier detection as follows.

---

Fig. 2. Histological analysis: (a) Sample blockface photograph of case HNL4_13. (b) Homography (i.e., perspective), corrected photograph. (c) Automated segmentation with random forest classifier. (d) Digitized histology of section corresponding to (a). (e) Close-up of the region inside the red square in (d).

Fig. 3. Rigid alignment of stacks of blockface photographs to MRI. (a) Initialization: axial view across the thalamus. (b) MRI. (c) Aligned stacks. We have overlaid a grid on subfigures (b) and (c) to ease the visual assessment of the quality of the registration.
First, an affine transform was robustly fitted to the pairs of landmarks using RANSAC (Fischler and Bolles, 1987). Second, the transform was applied to the landmarks on the MRI. Third, the distance between the actual and predicted landmark locations on the histology was computed. And fourth, if any of these distances was larger than 5 mm, the user was prompted to revise the placement of the corresponding landmark pairs.

Once the landmarks were in place, we optimized a cost function equal to the sum of the mutual information and the mean distance between corresponding landmarks after registration. A preliminary experiment, in which 1000 registrations from randomly selected sections were computed leaving one (randomly selected) landmark out, revealed that the registration error of this approach was 0.92 ± 0.64 mm, which we considered sufficient for the purposes of this article. The nonrigid registration is illustrated with an example in Fig. 4. Sample orthogonal slices of the reconstructed histology stacks are shown in Fig. 5.

### 2.6. Manual segmentation of nuclei on histology

The analysis of the histological sections was carried out by an expert neuroanatomist (R.I.), using a stereo microscope (Leica EZ4) with 50× magnification, an optic microscope (Nikon Eclipse E600), and the digitized histology. R.I. manually delineated the nuclei with ITK-Snap (http://www.itksnap.org), following the rostrocaudal axis of the thalamus, from the anterior thalamus to the end of the pulvinar nuclei. The delineation protocol was based on the characterization of the human and mammalian thalamus by Jones (2012), and is summarized in Table 2. Further details on the protocol and criteria for delineation will be described in an additional publication in a specialized journal.

### 2.7. 3D reconstruction of manual segmentations: filling the gaps

The rigid and nonrigid deformations that were computed to recover the 3D structure of the histological images can be directly applied to the manual segmentations in order to warp them to the space of the MRI scan. However, the directly warped labelings do not immediately yield usable 3D segmentations due to the gaps between blocks and the inconsistencies between the manual segmentations of adjacent sections (see Fig. 6a). To refine the segmentation in MRI space, we used the following automated approach. First, we performed a 2D erosion on each section and label (including the background) with a small circular kernel (radius: 1 mm), in order to generate a thin uncertainty zone around the edges of the segmentation – which models the registration error. Next, we deformed these eroded segmentations to the space of the MRI scan. Finally, we automatically assigned labels to the eroded voxels, as well as to the voxels in the gaps between blocks, by minimizing the following cost function:

\[
E(S) = -\sum_j \log p(l_j; \theta_\ell) + \alpha \sum_j D(l_j; S_j) + \beta \sum_{i \neq j} \delta(S_i \neq S_j),
\]

where \(S = \{S_j\}\) is the segmentation we want to compute (\(S_j\) is the segmentation of voxel \(j\)); \(l_j\) represents the MRI intensity at voxel \(j\); \(\theta_\ell\) is a set of Gaussian parameters associated with each label \(\ell\); \(D(l_j; S_j)\) represents the (physical) distance of voxel \(j\) to the initial segmentation of label \(S_j\); \(\delta(\cdot)\) is the Kronecker delta; \(\Gamma(j)\) is the 6-neighborhood of voxel \(j\); and \(\alpha\) and \(\beta\) are nonnegative weights.

The first term in Equation (1) encourages voxels with similar intensities to share the same label. The parameter set \(\theta_\ell\) correspond to the weights, means and variances of a Gaussian mixture model associated with label \(\ell\). These were estimated using the voxels labeled as \(l\) in the initial, eroded segmentation. The second term penalizes distance from the original segmentation; we set the distance \(D(l_j; S_j) = -\infty\) for voxels that are inside the segmentation, which effectively preserves these labels in the minimization. The third term is a Markov random field that penalizes pairwise differences in labels of neighboring voxels, thus ensuring the smoothness of the final result.

In order to minimize the cost in Equation (1), we used \(\alpha\)-expansion (Boykov et al., 2001). The number of Gaussian components was set to three for the background, and one for all other labels. The relative weights of the terms were set to \(\alpha = \beta = 1\). A preliminary experiment, in which 100 reconstructions from randomly selected cases were computed leaving one (randomly selected) deformed section out, revealed that the pixel classification accuracy of this approach was 89.1 ± 2.6%, which we considered sufficient for the purposes of this article. A sample output of the algorithm is shown in Fig. 6.
2.8. Atlas construction

In order to build a probabilistic atlas of the thalamic nuclei and surrounding tissue, we used our previously presented atlas construction method (Iglesias et al., 2015a). Compared with its conventional counterpart (Van Leemput, 2009), which cannot mix manual delineations at different levels of detail, this method relies on a generative model in which joint segmentations of the thalamic nuclei and the surrounding structures are assumed to exist, but are hidden. Instead, we observe a modified version in which sets of labels have been deterministically merged into coarser labels: in the in vivo scans, all the thalamic nuclei are merged into a generic thalamus label, whereas in the ex vivo images, the extrathalamic labels are fused into a single, generic background. Within this framework, Bayesian inference can be used to derive the most likely atlas of the thalamic nuclei and surrounding (whole) structures that generated the observed coarse labels, effectively combining the segmentations made on the in vivo and ex vivo images.

In this study, we combined our reconstructions of the thalamic nuclei with our previously presented atlas construction method (Iglesias et al., 2015a). Compared with its conventional counterpart (Van Leemput, 2009), which cannot mix manual delineations at different levels of detail, this method relies on a generative model in which joint segmentations of the thalamic nuclei and the surrounding structures are assumed to exist, but are hidden. Instead, we observe a modified version in which sets of labels have been deterministically merged into coarser labels: in the in vivo scans, all the thalamic nuclei are merged into a generic thalamus label, whereas in the ex vivo images, the extrathalamic labels are fused into a single, generic background. Within this framework, Bayesian inference can be used to derive the most likely atlas of the thalamic nuclei and surrounding (whole) structures that generated the observed coarse labels, effectively combining the segmentations made on the in vivo and ex vivo images.

In this study, we combined our reconstructions of the thalamic nuclei with our previously presented atlas construction method (Iglesias et al., 2015a). Compared with its conventional counterpart (Van Leemput, 2009), which cannot mix manual delineations at different levels of detail, this method relies on a generative model in which joint segmentations of the thalamic nuclei and the surrounding structures are assumed to exist, but are hidden. Instead, we observe a modified version in which sets of labels have been deterministically merged into coarser labels: in the in vivo scans, all the thalamic nuclei are merged into a generic thalamus label, whereas in the ex vivo images, the extrathalamic labels are fused into a single, generic background. Within this framework, Bayesian inference can be used to derive the most likely atlas of the thalamic nuclei and surrounding (whole) structures that generated the observed coarse labels, effectively combining the segmentations made on the in vivo and ex vivo images.

In this study, we combined our reconstructions of the thalamic nuclei with our previously presented atlas construction method (Iglesias et al., 2015a). Compared with its conventional counterpart (Van Leemput, 2009), which cannot mix manual delineations at different levels of detail, this method relies on a generative model in which joint segmentations of the thalamic nuclei and the surrounding structures are assumed to exist, but are hidden. Instead, we observe a modified version in which sets of labels have been deterministically merged into coarser labels: in the in vivo scans, all the thalamic nuclei are merged into a generic thalamus label, whereas in the ex vivo images, the extrathalamic labels are fused into a single, generic background. Within this framework, Bayesian inference can be used to derive the most likely atlas of the thalamic nuclei and surrounding (whole) structures that generated the observed coarse labels, effectively combining the segmentations made on the in vivo and ex vivo images.

In this study, we combined our reconstructions of the thalamic nuclei with our previously presented atlas construction method (Iglesias et al., 2015a). Compared with its conventional counterpart (Van Leemput, 2009), which cannot mix manual delineations at different levels of detail, this method relies on a generative model in which joint segmentations of the thalamic nuclei and the surrounding structures are assumed to exist, but are hidden. Instead, we observe a modified version in which sets of labels have been deterministically merged into coarser labels: in the in vivo scans, all the thalamic nuclei are merged into a generic thalamus label, whereas in the ex vivo images, the extrathalamic labels are fused into a single, generic background. Within this framework, Bayesian inference can be used to derive the most likely atlas of the thalamic nuclei and surrounding (whole) structures that generated the observed coarse labels, effectively combining the segmentations made on the in vivo and ex vivo images.
with manual delineations at the whole structure level made on 39 in vivo, 
T1-weighted scans, acquired at $1 \times 1 \times 1.25$ mm resolution (sagittal) on a 
Siemens 1.5 T platform with an MP-RAGE sequence. Thirty-six structures, 
including the left and right whole thalamus, were manually 
delineated using the protocol described by Caviness et al. (1989). We 
note that this is the dataset that was used to build the atlas in Friesen 
(Fischl et al., 2002; Fischl, 2012); further details on the acquisition can be 
found in the original publication. Both the in vivo and ex vivo datasets 
were augmented with left-right flips to increase their effective size. The 
atlas is represented as a tetrahedral mesh, which is locally adaptive to the 
complexity of the anatomy (Van Leemput, 2009). Sample slices of the 
atlas are displayed in Fig. 7.

### 2.9. Segmentation of in vivo MRI

Given a probabilistic atlas and a generative model of MRI scans, 
segmentation can be posed as a Bayesian inference problem within the 
model. As in previous work (Van Leemput et al., 1999; Iglesias et al., 
2015a,b; Puonti et al., 2016; Saygin et al., 2017), and following the 
literature of Bayesian segmentation (Wells et al., 1996; Zhang et al., 
2001; Ashburner and Friston, 2005; Pohli et al., 2006), we assumed the 
following forward model: first, the atlas is spatially warped following a 
deformation model (Ashburner et al., 2000). Second, a segmentation is 
drawn for each voxel independently, following the categorical distribu-
tion specified by the (deformed) atlas at each location. And third, image 
densities are drawn independently at each voxel, as independent samples 
of Gaussian mixture models conditioned on the underlying seg-
mamentation, i.e., each label has an associated set of Gaussian parameters 
describing the distribution of the intensities of its voxels.

In order to obtain an automated segmentation using the atlas, we first 
compute point estimates of the model parameters, namely the deforma-
tion of the atlas and the Gaussian parameters. This is done with a coor-
dinate ascent strategy, in which the deformation and the Gaussian 
parameters are alternately updated while keeping the other fixed. The 
deformation is updated with a conjugate gradient algorithm, whereas the 
Gaussian parameters are estimated with the Expectation Maximization 
(EM) algorithm (Dempster et al., 1977). The deformation is initialized by 
fitting the atlas to the automated segmentation of brain structures pro-
vided by the FreeSurfer main recon-all stream (aseg.mgz, Fischl et al., 
2002). Once the point estimates have been computed, the posterior 
probability of the segmentation is obtained as a by-product of the EM 
algorithm. Further details are given in Iglesias et al. (2015a); Puonti et al. 
(2016).

An important design choice of the segmentation algorithm is which 
classes are grouped in tissue types. Forcing different labels with similar 
intensity characteristics (e.g., gray matter structures such as the cerebral 
cortex, the hippocampus and the amygdala) to share Gaussian param-
eters improves the robustness of the algorithm. Here, we chose to group 
the thalamic nuclei into three different sets, representing different tissue 
types. First, the reticular nucleus was grouped with the rest of cerebral 
white matter structures in the atlas, as the large amount of fibers that 
cross it gives it a very similar appearance to that of white matter. A 
second set includes the mediodorsal and pulvinar nuclei (i.e., MDm, MDl, 
PuA, PuM, PuL, and PuL). All other nuclei are grouped into a third set. The 
division of nuclei between the second and third sets reflects the internal 
boundary that can be observed in in vivo MRI, even at standard resolution 
(see top row in Fig. 9). Fitting the atlas to this internal boundary provides a 
more reliable estimate of the segmentation.

### 3. Experiments and results

To validate the built atlas and its application to segmentation, we 
performed four different sets of experiments. As a first basic check, we 
conducted a volume comparison of the thalamic nuclei derived with the 
proposed atlas with those obtained with the atlas described in Krauth 
et al. (2010) (also derived from histology) using registration-based seg-
mentation. The other three experiments aimed at evaluating the perform-
ance of the proposed segmentation method indirectly – as direct 
evaluation would require labeling in vivo scans with the ex vivo protocol, 
which is not feasible. Specifically, the second experiment evaluated the 
reliability of our segmentation over time using an in vivo test-retest T1 
MRI dataset. The third set of experiments evaluated the robustness of the 
proposed atlas with respect to changes in MRI contrast of the input scan, 
using a heavily multimodal MRI dataset. The fourth and last set of ex-
periments assessed the ability of the proposed method to detect differ-
ential effects in the thalamus in a group study of Alzheimer’s disease, 
using the publicly available ADNI dataset.

#### 3.1. Volumetric comparison with Krauth’s atlas

To initially examine our proposed atlas in relation to existing 
thalamic atlases derived from histology, we selected Krauth et al. (2010). 
We conducted volumetric analysis for six representative nuclei (one per 
thalamic group, see Table 2), which are also present in Krauth’s atlas, and 
which are well characterized in terms of functional and structural con-
nectivity: anteroventral (AV), lateral posterior (LP), centromedian (CM), 
mediodorsal (MD, equal to the union of MDj and MDM), ventral lateral 
(VL, equal to the union of VLa and VLp), and pulvinar (PU, equal to the 
union of PuA, PuM, PuL and PuL).

Comparing the volumes of the nuclei in the two atlases directly is 
problematic, particularly given that Krauth’s model is a non-probabilistic 
mean derived from a small number of subjects. Instead, we compared the 
distributions of volumes of the nuclei when the atlases were applied to 
the automated segmentation of 66 subjects. For the proposed atlas, we 
used the method described above in this paper. For Krauth, we used 
direct registration-based segmentation: we took the MNI template with 
thalamic labels overlaid (see Jakab et al., 2012 for details); deformed it to 
the target subjects with ANTS (Avants et al., 2008); and used the 
resulting deformation fields to warp the labels to target space and create 
the segmentations.

The 66 subjects (age $24.31 \pm 3.70$ years; 40 females) were right-

d-handed healthy adults, with no history of psychiatric, neurological, 
attention or learning disorders. They were scanned at the BCBL on a 
3T Magnetom TIM Trio scanner, using a 32-channel head coil. T1-

![Fig. 7. Probabilistic atlas, with tetrahedral mesh superimposed. The color of each voxel is a linear combination of the colors in Table 2, weighted by the corresponding label probabilities. For the surrounding structures, we used the standard FreeSurfer color map. (a) Sample coronal slice. (b) Axial slice. (c) Sagittal slice.](image-url)
TI = 1100 ms, $\alpha = 7^\circ$, FOV = 256 mm, 176 slices, resolution 1 mm isotropic.

Fig. 8 displays violin plots (box plots with a rotated kernel density plot on each side) comparing the distribution of volumes of the representative nuclei. In qualitative terms, the agreement between the nuclei is good, even though Krauth’s atlas yields slightly larger volumes for the anteroventral and lateral posterior nuclei (bilateral); and the left pulvinar nucleus. This good agreement is also apparent in Fig. 9, which displays a case segmented with both atlases. We note that direct comparison of whole thalamic volumes is not straightforward, due to different inclusion criteria, e.g., Krauth’s atlas covers the red nucleus, subthalamic nucleus and mammillothalamic tract, while the proposed atlas does not (see Fig. 9).

### 3.2. Test-retest reliability

In order to evaluate the test-retest reliability of the proposed segmentation method, we used MRI data from 31 of the 66 subjects (age $24.34 \pm 2.96$ years; 17 females), who also participated in a second session between seven and 10 days after the first session. The scanning protocol was the same as in the first experiment. The intraclass correlation coefficients (ICC) between the volumes derived from the scans in the first and second session are displayed in Table 3. Despite the fact that the thalamus is notoriously difficult to segment due to its faint lateral boundary, our algorithm produces very high ICCs for the left and right whole thalamus (above 0.97). Moreover, and despite the fact that the internal boundary used by the algorithm to fit the atlas is also faint, the ICCs for the individual representative nuclei are also excellent – all scores are over 0.85, and most are over 0.90 and even 0.95.

### 3.3. Robustness against changes in MRI contrast

Compared with discriminative approaches, a general advantage of Bayesian segmentation methods is their robustness against changes in the MRI contrast of the input scans. In order to test this robustness, we used a separate dataset consisting of multimodal MRI data from 11 subjects, acquired on a 3T Prisma scanner (32 channel head coil) at the Martinos Center for Biomedical Imaging. For each subject, we first acquired two T1-weighted, MP2RAGE scans (Marques et al., 2010) with parameters: $\text{TI} = 700 \text{ ms and } 2500 \text{ ms}; \alpha = 4^\circ \text{ and } 5^\circ$, $\text{TE} = 2.98 \text{ ms}$, image size $= 256 \times 240 \times 176$, GRAPPA acceleration factor = 3, bandwidth 240 Hz/pixel, resolution 1 mm isotropic. The two T1s were motion corrected and averaged. The two inversion times were used to compute the quantitative T1 relaxation time at each voxel. Given the quantitative T1 data and the individual MPRAGEs, we computed the proton density. Using these data, we synthesized the k-space for a single-inversion MPRAGE with $\alpha = 7^\circ$, $\text{TR} = 2530 \text{ ms}$, $\text{TE} = 0 \text{ ms}$, with inversion times ranging from 300 ms to 940 ms using a Bloch simulator (Ma et al., 2013). Within the simulation, the data were acquired instantaneously (i.e., infinite bandwidth). The simulated k-space data was then reconstructed using an FFT and the absolute value taken.

![Fig. 8](image8.png) Violin plots (box plots with a rotated kernel density plot on each side) for volumes of representative nuclei, computed with our proposed atlas and with Krauth’s. MD, VL and PU represent the whole mediiodorsal, ventral lateral and pulvinar nuclei, i.e., MD is the union of MDm and MDl; VL is the union of VLa and VLp; and PU is the union of PuA, PuM, PuL and PuI.

![Fig. 9](image9.png) Sagittal (left), coronal (middle), and axial (right) slices of a brain MRI scan segmented with the proposed atlas and Krauth’s. The color map for our atlas is that described in Table 2. For Krauth, we attempted to match the color map as much as possible; we note that their atlas includes the red nucleus (in red, pointed by yellow arrows) and subthalamic nucleus (in green, pointed by blue arrows), in the inferior region, which are not part of our proposed atlas. On the left thalamus of the input scan, we have overlaid a manually delineated boundary between the mediiodorsal and pulvinar nuclei, and the rest of the nuclei; this is the main feature we use to fit the internal boundaries of the thalamus.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anteroventral (AV)</td>
<td>0.86</td>
<td>0.93</td>
</tr>
<tr>
<td>Lateral posterior (LP)</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td>Centromedian (CM)</td>
<td>0.94</td>
<td>0.92</td>
</tr>
<tr>
<td>Mediodorsal (MD)</td>
<td>0.89</td>
<td>0.95</td>
</tr>
<tr>
<td>Ventral lateral (VL)</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>Pulvinar (PU)</td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td>Whole thalamus</td>
<td>0.97</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 3

Intraclass correlation coefficients for representative thalamic nuclei and whole thalamus. As in Fig. 8, MD is the union of MDm and MDl; VL is the union of VLa and VLp; and PU is the union of PuA, PuM, PuL and PuI.
We also acquired two additional volumes for each subject. First, a T2-weighted volume with the following parameters: TR = 3200 ms, TE = 564 ms, acceleration factor = 2, bandwidth = 651 Hz/pixel, echo spacing 3.66 ms, resolution 1 mm isotropic. And second, an additional MPRAGE with a contrast that is typically used in DBS: TR = 3000 ms, TE = 3.56 ms, TI = 406 ms, $\alpha = 8^\circ$, phase field of view = 81.3%, resolution 0.8 mm isotropic.

The segmentation algorithm for the alternative MRI contrasts is the same as for the T1 data, and also uses the automated segmentation aseg.mgz in the initialization. Since this segmentation is computed by FreeSurfer from the T1 images, the results of this experiment are positively biased by the common initialization. However, we note that this will be the scenario of the public release of the algorithm in FreeSurfer, in which the availability of a T1 scan is always assumed.

An example of a coronal slice with all available MRI contrasts and associated segmentations is shown in Fig. 10. The segmentation is quite stable across contrasts, though some differences can be observed both in the internal and external boundaries of the thalamus. Quantitative results are displayed in Fig. 11, which shows, for each representative nucleus, the Dice overlap between the segmentations for each pair of MRI contrasts. The agreement between the segmentations of the whole thalamus is quite high, near or above 0.90 in almost all cases. For the individual nuclei, the overlaps are moderately high, given that we are considering substructures, particularly for CM and VL (Dice approximately between 0.75 and 0.85). The overlaps are slightly lower for AV, MD and PU, with Dice scores approximately between 0.65 and 0.75. In terms of MRI contrast, the best agreement is observed (as expected) between the synthetic MPRAGEs with inversion times over 600 ms – since TI < 600 ms produces a contrast flip. The least consistent MRI contrast is the T2, in which the boundary between MD/PU and the rest of nuclei is almost invisible (see example in Fig. 10). In general, the agreement is good between most pairs of contrasts.

3.4. Alzheimer’s disease study

In order to assess the effectiveness in neuroimaging group studies of our proposed automated segmentation method based on the ex vivo atlas, we ran the algorithm on a subset of the publicly available ADNI dataset. The ADNI (adni.loni.usc.edu) 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 MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer’s disease. Here we considered T1-weighted scans from 213 subjects with Alzheimer’s and 161 age-matched controls (Alzheimer’s: 76.04 ± 5.42 years; controls: 75.58 ± 7.37 years); we note that these are the subjects that we have used in previous studies from our group (e.g., Iglesias et al., 2015a; Saygin et al., 2017). The resolution of the T1 scans was approximately 1 mm isotropic; further details on the acquisition can be found in the ADNI website. The volumes of the thalamic nuclei and of the whole thalamus were corrected by age and intracranial volume (as estimated by FreeSurfer) and left-right averaged in all analyses.

To discriminate the subjects into the two classes, we considered three different approaches. First, thresholding of the whole thalamic volume, as estimated with the main FreeSurfer recon-all stream (i.e., aseg.mgz). Second, thresholding of the whole thalamic volume, as estimated by the proposed atlas (i.e., summing all the nuclei). And third, thresholding of the likelihood ratio given by a linear discriminant analysis (LDA, Fisher, 1936), computed...
consistent with the literature (Aggleton et al., 2016; Pini et al., 2016) and with much more accuracy than the whole thalamus. These results are boundary of the thalamus, even if faint (and hence prone to segmentation
tation by the number of nuclei. The table shows that
signi
Table 4, which shows the p-values for the individual nuclei that display
whole thalamus alone. The reason for this increase is apparent from
represents a very large increase with respect to using the volume of the
whole thalamus to the volumes of its subregions. Looking at indi-
vidual nuclei, large differences were found in mediodorsal, anteroventral
areas. This is consistent with previous results in
imaging studies (Aggleton et al., 2016; Pini et al., 2016; Step
kowska et al., 2014; Zarei et al., 2010). Moreover, a three-year
longitudinal study reported that thalamic atrophy was
first localized in the ventromedial regions, and spread to anterior regions in later stages
(Cho et al., 2014), which are well represented in the ADNI dataset. Mediodorsal regions have also been reported as foci of atrophy: for
example, a voxel-based morphometry study in genetic Alzheimer's dis-
ee showed that mutation carriers exhibit a decreased gray matter
density localized in the medial-dorsal regions of the thalamus within 5
years of symptoms onset (Cash et al., 2013).

Our results are also supported by neuropathological studies. For
example, Braak and Braak (1991) and Xuereb et al. (1991) have shown
that the primary site of Alzheimer's disease degeneration in the thalamus
is the anterodorsal nucleus, which showed severe neuronal loss and
tangle formation. In our proposed atlas, the anterodorsal nucleus (due to
its small size) has been included in the anteroventral (AV) nucleus. The
AV nucleus not only shows a strong effect between Alzheimer's and
controls, but is also the nucleus with the strongest ICC in the repeatability
experiment. These anterior thalamic nuclei are densely and directly
connected to the medial temporal lobe structures (e.g., hippocampus,
amygdala and entorhinal cortex), in addition to the relay in the
mammillary nuclei of the hypothalamus, all of which are typically
affected in Alzheimer's disease and linked to its episodic memory deficits
(Aggleton et al., 2016). Moreover, Zarei et al. (2010) showed that
practically the entire cortex, the availability of accurate and reliable
automated segmentation algorithms for the thalamic nuclei is of high
interest for the neuroimaging community. Here we have introduced a
probabilistic atlas of 26 human thalamic nuclei built upon 3D recon-
structed histological data from 12 thalami; presented a Bayesian seg-
mentation method that applies the atlas to the automated segmentation of
thalamic nuclei in in vivo brain MRI scans of arbitrary contrast; and
validated the atlas and segmentation with four different sets of experi-
ments, whose results are discussed next.

First, we compared the proposed atlas with a previously presented
histology-based atlas (Krauth et al., 2010) by segmenting the thalamic
nuclei in a population of 66 subjects and comparing the distribution of
the volumes. To produce the segmentations, our proposed atlas was
combined with a Bayesian segmentation technique presented in this
article. For Krauth, we used direct registration-based segmentation.
Leaving aside discrepancies in the anatomical definition of nuclei, the
two models yielded very similar volumes, which is reassuring in terms of
scientific reproducibility. The advantage of our model lies in its proba-
bilistic nature, which enables segmentation of scans of arbitrary MRI
contrast within a Bayesian framework.

Second, we conducted a test-retest reliability study of our segmenta-
tion tool using 1 mm T1 scans acquired approximately one week apart.
This experiment revealed excellent repeatability, with ICC scores mostly
above 0.90 (which is higher than that of some whole structures in
FreeSurfer; see Morery et al., 2010). This result reassures that the obtained
volumes are reliable, i.e., they are an accurate representation of a mea-
surement, rather than attributed to random fluctuations.

A third experiment assessed the robustness of the tool to changes in
the MRI contrast of the input scan. This experiment was also successful,
as the agreement between segmentations is good across a wide array of
contrasts. This result supports the hypothesis that our structural seg-
mentation algorithm will be able to take advantage of MRI pulse se-
quences that produce good contrast in the thalamus in vivo, such as fast
grey matter acquisition T1 inversion recovery scans (FGATIR).

The most intriguing results are those from the experiment with Alz-
heimer's disease subjects, as a large boost in classification performance
(increase of 0.20 in AUC) is observed when switching from the volume of
the whole thalamus to the volumes of its subregions. Looking at indi-
vidual nuclei, large differences were found in mediodorsal, anteroventral
and ventral anterior areas. This is consistent with previous results in
neuroimaging studies. Antero-ventro-medial regions have been reported
as main foci of impairment in Alzheimer's disease in several neuro-
imaging studies (Aggleton et al., 2016; Pini et al., 2016; Stepán-Buksa-
kowska et al., 2014; Zarei et al., 2010). Moreover, a three-year
longitudinal study reported that thalamic atrophy was first localized in
the ventromedial regions, and spread to anterior regions in later stages
(Cho et al., 2014), which are well represented in the ADNI dataset. Mediodorsal regions have also been reported as foci of atrophy: for
example, a voxel-based morphometry study in genetic Alzheimer's dis-


4. Discussion

Human cortical information flow and dynamics cannot be fully un-
derstood without taking into account thalamocortical interactions. Due
to its critical functions and widespread structural connections with

Fig. 12. ROC curves for Alzheimer's vs. controls classification based on left-
right averaged thalamic volumes. The red curve is for the whole thalamic vol-
ume estimated by Freesurfer's recon-all stream (AUC = 0.632). The blue curve is
for the whole thalamic volume, estimated as the sum of the volumes of the nuclei
given by our proposed atlas (AUC = 0.645). The black curve corresponds to
a leave-one-out LDA classifier that simultaneously considers the volumes of
all nuclei, as estimated by the proposed atlas (AUC = 0.830). The volumes were
left-right averaged and corrected by age and intracranial volumes in all cases.
in a leave-one-out fashion. LDA is a simple linear analysis, which enables us
to consider all nuclei simultaneously, while ensuring that the performance is
mostly determined by the input data, rather than stochastic variations in a
more complex classifier.

The receiver operating characteristic (ROC) curves are shown in
Fig. 12. The area under the curve (AUC), which is a threshold-
independent measure of performance for a classifier, was 0.632 for the
thalamic volumes given Freesurfer's main stream. When using the whole
thalamic volumes given by the proposed atlas, the AUC was slightly
higher (AUC = 0.645), though the difference was not statistically sig-
nificant (p = 0.4) according to a paired DeLong test (DeLong et al.,
1988). However, when using all nuclei simultaneously, the AUC
increased considerably to 0.830 (p < 10^{-10}) against the whole thalamic
volumes, given by recon-all or our proposed atlas.

While AUC = 0.830 is modest in terms of Alzheimer's classification, it
represents a very large increase with respect to using the volume of the
whole thalamus alone. The reason for this increase is apparent from
Table 4, which shows the p-values for the individual nuclei that display
significant differences between the two groups, after Bonferroni corre-
tion by the number of nuclei. The table shows that the fitting the internal
boundary of the thalamus, even if faint (and hence prone to segmentation
mistakes), enables some thalamic structures to separate the two classes
with much more accuracy than the whole thalamus. These results are
consistent with the literature (Aggleton et al., 2016; Pini et al., 2016) and
may be explained by the distinct pattern of connections and functions of
these thalamic regions, and by the fact that not all thalamic nuclei are
equally affected by the disease – see further details below under Section
4.

4. Discussion

Human cortical information flow and dynamics cannot be fully un-
derstood without taking into account thalamocortical interactions. Due
to its critical functions and widespread structural connections with
atrophy of the mediodorsal thalamus (which shows the second to largest effect in our experiment) corresponds to changes in connectivity with cortical and subcortical areas. Finally, Braak and Braak (1991) also found amyloid deposition in the anterointernal, laterodorsal, and the central medial nucleus; the first two of these regions (anterointernal and laterodorsal) correspond with our volumetric findings.

While these neuroimaging and neuropathological studies support the outcome of our experiment, our results could also be a partial correlate of the expansion of the neighboring ventricles in Alzheimer’s disease. For example, the thalamic atrophy detected by Zarei et al. (2010) through a shape analysis was characterized by an inward movement of vertices in the dorsomedial and ventral aspects of the thalamus, which may be associated with ventricular enlargement. However, we may also argue that such ventricular effect should also affect the volume of the whole thalamus, which is only the case to a much lesser extent in our experiment. Another aspect that requires further inspection is the atrophy detected in the lateral and medial geniculate nuclei (LGN/MGN), for which there is little evidence in the literature. The measured atrophy might be a false positive due to small nuclei sizes, lack of contrast with neighboring cerebral white matter, or both. However, it could also be a true effect, e.g., a correlate of the visual impairments associated with Alzheimer’s disease (Kirby et al., 2010), as the LGN is a major relay of the visual pathway; a similar effect is potentially possible for the MGN, which is linked to auditory processing. We note that, even when the LGN (the single most discriminating nucleus; see Table 4) and the MGN are removed from the LDA analysis, the AUC of the classifier is 0.783, which is still considerably higher than the values given by the whole thalamus (AUC = 0.645).

5. Conclusion

We have presented a probabilistic atlas of the human thalamus based on ex vivo imaging techniques (ex vivo MRI, histology), and a companion tool that enables segmentation of thalamic nuclei from in vivo MRI of arbitrary contrast. At the technical level, future work will focus on integrating diffusion MRI data in the segmentation algorithm. While segmentation based solely on structural MRI enables analysis of large amounts of legacy datasets that do not include diffusion data, the local diffusion information and structural connectivity are strong signatures of the divisions between different thalamic nuclei. Therefore, including these data in the generative model should greatly inform the Bayesian segmentation algorithm, which currently relies on faint boundaries and prior knowledge to fit the atlas to the input images.

Improvements in the segmentation will also enable more advanced studies of multiple disorders, e.g., Alzheimer’s disease, Parkinson’s disease, dyslexia, schizophrenia, etc. Future analyses will include: further investigation of the results of the group study reported in this article; correlation of thalamic nuclei with clinical scores, subtypes of the disease, and disease duration; investigation of specific thalamic networks (functional and structural) using the segmented nuclei as seeds; or cluster analysis of nuclei volumes for subtyping the disease.

Our proposed atlas and segmentation tool are publicly available as part of the neuroimaging software package FreeSurfer (https://surfer.nmr.mgh.harvard.edu), and the ex vivo data used for building the atlas will be published in the future along with a paper focused on the morphology of the human thalamus. Our proposed atlas will enable neuroimaging studies of the human thalamus at sites that do not possess the expertise or staff resources to manually delineate the thalamic nuclei in 3D MRI data.

Acknowledgements

The authors would like to thank Professor Karla Miller (Oxford) for her help with the design of the ex vivo MRI acquisition; Ms. Mercedes Iglesias de Onzorio and Mr. Francisco Romero (UCLM) for their careful technical laboratory help; and Mr. Gonzalo Artacho (UCLM) for his help with the digitization and curation of his organization of histological data.

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 654911 (project “THALAMODEL”) and by the European Research Council (ERC) Starting Grant agreement No 677697 (“BUNGEETOOLS”). It was also funded by the Spanish Ministry of Economy and Competitiveness/MINECO TEC-2014-51882-P, RYC-2014-15440, PSI2015-65696, and SEV-2015-0490, the Basque Government (PI2016-12), and UCLM Internal Research Groups grants.

Support for this research was also provided in part by the National Institute of Biomedical Imaging and Bioengineering (P41EB015896, 1R01EB023281, R01EB006758, R21EB018907, R01EB019956), the National Institute on Aging (5R01AG008122, R01AG0160495), the National Institute of Diabetes and Digestive and Kidney Diseases (1-R21-DK-108277-01), the National Institute of Neurological Disorders and Stroke (R01NS0525851, R21NS072652, R01NS070963, R10NS083534, R01NS086625), and was made possible by the resources provided by Shared Instrumentation Grants 1S1ORR023401, 1S1ORR019307, and 1S-1ORR023043. Additional support was provided by the NIH Blueprint for Neuroscience Research (U01-MH093765), part of the multi-institutional Human Connectome Project. In addition, B.F. has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. B.F.’s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (National Institutes of Health Grant U01 AG024904) and DOD ADNI (DOD 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, Alzheimers Association; Alzheimers Drug Discovery Foundation; Arclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Elai; Eli Lilly and Company; Elan Pharmaceuticals, Inc.; El 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 Alzheimers Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.