2014

Derivation of high-resolution MRI atlases of the human cerebellum at 3 T and segmentation using multiple automatically generated templates

Min Tae M. Park
Centre For Addiction And Mental Health, Toronto, Canada

Jon Pipitone
Centre For Addiction And Mental Health, Toronto, Canada

Lawrence H. Baer
Concordia University

Julie L. Winterburn
Centre For Addiction And Mental Health, Toronto, Canada

Yashvi Shah
Centre For Addiction And Mental Health, Toronto, Canada

See next page for additional authors

Derivation of high-resolution MRI atlases of the human cerebellum at 3 T and segmentation using multiple automatically generated templates

Abstract

The cerebellum has classically been linked to motor learning and coordination. However, there is renewed interest in the role of the cerebellum in non-motor functions such as cognition and in the context of different neuropsychiatric disorders. The contribution of neuroimaging studies to advancing understanding of cerebellar structure and function has been limited, partly due to the cerebellum being understudied as a result of contrast and resolution limitations of standard structural magnetic resonance images (MRI). These limitations inhibit proper visualization of the highly compact and detailed cerebellar foliations. In addition, there is a lack of robust algorithms that automatically and reliably identify the cerebellum and its subregions, further complicating the design of large-scale studies of the cerebellum. As such, automated segmentation of the cerebellar lobules would allow detailed population studies of the cerebellum and its subregions. In this manuscript, we describe a novel set of high-resolution in vivo atlases of the cerebellum developed by pairing MR imaging with a carefully validated manual segmentation protocol. Using these cerebellar atlases as inputs, we validate a novel automated segmentation algorithm that takes advantage of the neuroanatomical variability that exists in a given population under study in order to automatically identify the cerebellum, and its lobules. Our automatic segmentation results demonstrate good accuracy in the identification of all lobules (mean Kappa [κ] = 0.731; range 0.40–0.89), and the entire cerebellum (mean κ = 0.925; range 0.90–0.94) when compared to "gold-standard" manual segmentations. These results compare favorably in comparison to other publically available methods for automatic segmentation of the cerebellum. The completed cerebellar atlases are available freely online (http://imaging-genetics.camh.ca/cerebellum) and can be customized to the unique neuroanatomy of different subjects using the proposed segmentation pipeline (https://github.com/pipitone/MAGeTbrain).

Keywords

3, human, mri, cerebellum, resolution, t, high, segmentation, derivation, multiple, automatically, generated, templates, atlases

Disciplines

Education | Social and Behavioral Sciences

Publication Details

Park, M. M., Pipitone, J., Baer, L. H., Winterburn, J. L., Shah, Y., Chavez, S., Schira, M. M., Lobaugh, N. J., Lerch, J. P., Voineskos, A. N. & Chakravarty, M. M. (2014). Derivation of high-resolution MRI atlases of the human cerebellum at 3 T and segmentation using multiple automatically generated templates. Neuroimage, 95 217-231.

Authors

Min Tae M. Park, Jon Pipitone, Lawrence H. Baer, Julie L. Winterburn, Yashvi Shah, Sofia Chavez, Mark M. Schira, Nancy J. Lobaugh, Jason P. Lerch, Aristotle N. Voineskos, and M Mallar Chakravarty

Derivation of high-resolution MRI atlases of the human cerebellum at 3T and segmentation using multiple automatically generated templates.

Min Tae M. Park[1]*, Jon Pipitone[1], Larry Baer[1], Julie L. Winterburn[1], Yashvi Shah[1], Sofia Chavez[2,3], Mark M. Schira[4,5], Nancy J. Lobaugh[2,6], Jason P. Lerch[7], Aristotle N. Voineskos[1,3], M. Mallar Chakravarty[1,3,8*].

1- Kimel Family Translational Imaging Genetics Research Laboratory, Research Imaging Centre, Centre for Addiction and Mental Health, Toronto, Canada
2- MRI Unit, Research Imaging Centre, Centre for Addiction and Mental Health, Toronto, Canada
3- Department of Psychiatry, University of Toronto, Toronto, Canada
4- School of Psychology, University of Wollongong, Wollongong, NSW, Australia
5- Neuroscience Research Australia, Sydney, NSW, Australia
6- Division of Neurology, Department of Medicine, University of Toronto, Toronto, Canada
7- Program in Neuroscience and Mental Health, Hospital for Sick Children, Toronto, Canada
8- Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada

correspondence to:

Min Tae M. Park and M. Mallar Chakravarty
Kimel Family Translation Imaging Genetics Research Laboratory
Research Imaging Centre
Centre for Addiction and Mental Health
250 College St.
Toronto, Canada
M5T 1R8
mtpark89@gmail.com; mallar.chakravarty@camh.ca
1-416-535-8501 ext. 34064

Abstract

The cerebellum has classically been linked to motor learning and coordination. However, there is renewed interest in the role of the cerebellum in non-motor functions such as cognition and in the context of different neuropsychiatric disorders. The contribution of neuroimaging studies to advancing understanding of cerebellar structure and function has been limited, partly due to the cerebellum being understudied as a result of contrast and resolution limitations of standard structural magnetic resonance images (MRI). These limitations inhibit proper visualization of the highly compact and detailed cerebellar foliations. In addition, there is a lack of robust algorithms that automatically and reliably identify the cerebellum and its subregions, further complicating the design of large-scale studies of the cerebellum. As such, automated segmentation of the cerebellar lobules would allow detailed population studies of the cerebellum and its subregions. In this manuscript, we describe a novel set of high-resolution in vivo atlases of the cerebellum developed by pairing MR imaging with a carefully validated manual segmentation protocol. Using these cerebellar atlases as inputs, we validate a novel automated segmentation algorithm that takes advantage of the neuroanatomical variability that exists in a given population under study in order to automatically identify the cerebellum, and its lobules. Our automatic segmentation results demonstrate good accuracy in the identification of all lobules (mean Kappa [κ]=0.731; range 0.40-0.89), and the entire cerebellum (mean κ=0.925; range 0.90-0.94) when compared to "gold-standard" manual segmentations. These results compare favorably in comparison to other publically available methods for automatic segmentation of the cerebellum. The completed cerebellar atlases are available freely online (http://imaging-genetics.camh.ca/cerebellum) and can be customized to the unique neuroanatomy of different subjects using the proposed segmentation pipeline (https://github.com/pipitone/MAGeTbrain).

Keywords

MRI

High resolution

Atlas

Cerebellum

Lobule

Automatic

1.0 Introduction

The cerebellum is located below the cerebral hemispheres and posterior to the pons of the brainstem, forming a fan-like structure with intricate and tightly packed foliations of grey matter radiating out from the central white matter, the corpus medullare. Classically, the cerebellum has been extensively linked to motor coordination and motor learning on the basis of anatomical, functional, and neuroimaging studies (Thach et al., 1992, Thach, 1996, Bastian, 2006, Mottolese et al., 2013). This traditional portrait of the cerebellum is supported by clinical findings that show motor dysfunction in patients with cerebellar damage (Bastian et al., 1998).

There has been renewed interest in the role of cerebellum in normal brain function (Schmahmann, 1991, Kim et al., 1994, Middleton and Strick, 1994, Strick et al., 2009, O'Halloran et al., 2012) and many studies have observed cognitive and emotional impairments in patients with cerebellar damage (Schmahmann and Shermann, 1998, Riva and Giorgi, 2000, Schmahmann, 2004, Schmahmann et al., 2007), specific activations localized to the cerebellum in cognitive tasks (Stoodley and Schmahmann, 2009, Stoodley, 2012), and concurrent findings of cerebellar integration in cortico-cerebellar networks that may be responsible for cognitive processes (Krienen and Buckner, 2009, Habas et al., 2009, O'Reilly et al., 2010, Buckner et al., 2011). Additional findings in anatomical, clinical, and neuroimaging studies further motivate these associations by illustrating novel cerebellar involvement in various domains of cognition such as language processing (Petersen et al., 1989), memory (Andreason et al., 1995, Andreasen et al., 1999), attention (Allen et al., 1997) and emotional control (Habel et al., 2005, Turner et al., 2007). Specific lobules of the cerebellum are implicated in these processes (Stoodley and Schmahmann, 2009), demonstrating functional specialization of individual lobules. Such structure-function relationships involving the cerebellar lobules are further supported by studies that link even specific motor functions to individual lobules as well (Schlerf et al., 2010, Mottolese et al., 2013). Taken together, these findings provide motivation for developing tools to analyze cerebellar neuroanatomy with greater specificity.

The wide spectrum of cerebellar involvement in diverse forms of brain functions has led to the re-examination of its roles in various neuropsychiatric conditions. For example, there is substantial evidence of cerebellar involvement in autism from both structural (Courchesne et al., 1988, Murakami et al., 1989, Piven et al., 1992) and functional neuroimaging studies (Muller et al., 2001, Mostofsky et al., 2009). However in other disorders such as schizophrenia, reports of structural abnormalities of the cerebellum are inconsistent; some studies have reported reduced cerebellum vermis size in patients with schizophrenia (Jacobsen et al., 1997, Nopoulos et al., 1999, Ichimiya et al., 2001, Loeber et al., 2001, Okugawa et al., 2002), while others presented contradictory or inconclusive results (Levitt et al., 1999, James et al., 2004, Nenadic et al., 2010). Examinations in other neuropsychiatric disorders such as Alzheimer's disease (Fukutani et al., 1996, Wegiel et al., 1999, Thomann et al., 2008, Chen et al.,

2010), bipolar disorder (Baldacara et al., 2011), dyslexia (Kibby et al., 2008) report similar inconsistent results. This heterogeneity in the literature may, in part, be caused by the lack of large-scale neuroimaging studies targeting the cerebellum as well as the limits of typical structural magnetic resonance images (MRIs) for the detailed visualization of cerebellar foliations. Unlike other brain structures, there are only a few methods available for automated analysis of cerebellar anatomy. The cerebellum is understudied, in part, due to these methodological shortcomings.

A limited number of automated segmentation algorithms have been developed and validated for the entire cerebellum (Hartmann et al., 1999, Isambert et al., 2008, Zhao et al., 2010, Hwang et al., 2011, Ljin et al., 2012) and its lobes (Powell et al., 2008). Surprisingly, only a few fully automated methods (Diedrichsen, 2006, Diedrichsen et al., 2009, Bogovic et al., 2013b) segment the cerebellar lobules using a well-known standard for cerebellar anatomy (Schmahmann et al., 1999, Schmahmann et al., 2000). One approach for automatic segmentation of the cerebellar lobules would be to use multiple candidate atlases to model anatomical variability in a multi-atlas based segmentation framework (Heckemann et al., 2006). However, this class of methods requires a large (60-80) number of atlases for accurate segmentations and the manual derivation of such atlases is time consuming and difficult. Bogovic et al. (2013a) have also evaluated the use of multi-atlas based segmentation, showing poor automatic segmentations (mean Kappa 0.66 over all lobules when compared to manual segmentations) even with the use of 15 manually derived atlases as input.

The goals of this manuscript were three-fold: the first was to develop novel in vivo atlases of the cerebellum, cerebellar lobules, and deep cerebellar nuclei (DCN) based on high-resolution T1- and T2-weighted data. Unlike other structures, publicly available atlases of the cerebellum are minimal. The atlases presented in this manuscript are one of the few that provide detailed high-resolution representations of the cerebellum. By contrast, the Diedrichsen probabilistic atlas (Diedrichsen et al., 2009) is also available in different image processing toolkits (SPM, FSL, and AFNI). However, to the best of our knowledge, none of the data from individual subjects used to derive the atlas is currently available. A second goal of the present study was to provide an extensively detailed description of our manual segmentation protocol to enable replication by other research groups. A third goal was to validate automated segmentation within the MAGeT Brain framework (Multiple Automatically Generated Templates; Chakravarty et al., 2013). MAGeT is a novel automated segmentation algorithm recently proposed by our group that requires only a minimal number of atlases to produce accurate segmentations of brain structures (Chakravarty et al., 2013). Using our cerebellar atlases as input, we present an evaluation of MAGeT Brain by applying automatic segmentation on a standard resolution (1.5T) validation set consisting of 20 T1-weighted images, and comparing the resultant segmentations to “gold standard" manual segmentations. We additionally compare the performance of MAGeT Brain to the SUIT toolbox, a publically available method for cerebellum lobule segmentation (Diedrichsen, 2006, Diedrichsen et al., 2009). Our five cerebellar atlases are publically available online at

(http://imaging-genetics.camh.ca/cerebellum), and the source code for the MAGeT Brain is publicly available at https://github.com/pipitone/MAGeTbrainand an alternative version for mouse imaging data at https://wiki.phenogenomics.ca/display/MICePub/Home.

2.0 Methods

In the following we describe the image acquisition parameters for the high-resolution MRIs used to create the cerebellar atlases, and the standard T1-weighted images used for the validation set images. We then describe the manual segmentation protocol in great detail, used for both the atlases and the validation set. Finally, we present the methods for evaluating the automatically generated MAGeT Brain and multi-atlas based cerebellar segmentations.

2.1 Image acquisition

High-resolution MR images for atlases

Previous work using these data has been published for the hippocampus (Winterburn et al., 2013), and we have repeated the acquisition details in this manuscript for completeness.

Cerebellar atlases were developed based on high-resolution MR images acquired from five healthy volunteers (2 male, 3 female, aged 29-57), free of neurological and neuropsychiatric disorders and no history of head injury with loss of consciousness. The study was conducted in keeping with the Declaration of Helsinki, was approved by the Centre for Addiction and Mental Health Research Ethics Board, and all subjects provided written, informed consent for data acquisition and sharing.

T1- and T2-weighted images were acquired on a GE Discovery MR 750 3T system (General Electric, Milwaukee, WI) using an 8-channel head coil. High-resolution T1-weighted images were acquired using the 3D inversion-prepared fast spoiled gradient-recalled echo acquisition, FSPGR-BRAVO, in a scan time of ~20 min with the following parameters: TE/TR= 4.3 ms/9.2 ms, TI=650 ms, α=8°, 2NEX, FOV= 22 cm, slice thickness= 0.6 mm, 384x384 in-plane steps for an approximate isotropic resolution of 0.6 mm dimension voxels. High-resolution T2-weighted images were acquired using the 3D fast spin echo acquisition, FSE-CUBE, in a scan time of ~16 min, with the following parameters: TE/TR= 95.3 ms/2500 ms, ETL=100 ms, 2NEX, FOV=22 cm, slice thickness=0.6 mm, 384×384 in-plane steps for approximately isotropic 0.6 mm dimension voxels. Reconstruction filters, ZIPX2 and ZIP512, were also used resulting in isotropic 0.3 mm^3 voxels. All 2NEX scans were repeated three times and averaged for a total of 6NEX for both image contrasts (i.e.: T1 and T2). T1 and T2 images were acquired in two separate sessions for each volunteer to minimize discomfort to the subject.

Data from each subject were converted to the MINC file format (http://www.bic.mni.mcgill.ca/ServicesSoftware/MINC), corrected for RF intensity inhomogeneities

(Sled et al., 1998) and the second and third T1 and T2 images were rigidly aligned to the first image volume of homologous contrast (Collins et al., 1994). Each image was then normalized to a fixed intensity range (0-10,000), and intensity-averaged on a voxel-by-voxel basis to enhance signal and contrast (Holmes et al., 1998). All rigid-body transformations were applied using sinc interpolation to maintain contrast enhancement at the borders between structures.

Validation set based on standard T1 MR images

Manually labeled MR images from 10 healthy control and 10 patients with schizophrenia (15 male, 5 female, average 30.1 years age, range 25-35 years) were used to create a validation set. Imaging data were acquired using an eight-channel head coil on a 1.5 Tesla GE Echospeed system (General Electric Medical Systems, Milwaukee, WI), which permits maximum gradient amplitudes of 40 mT/m. Axial inversion recovery prepared spoiled gradient recall images were acquired: echo time (TE) = 5.3 ms, repetition time (TR) = 12.3 ms, time to inversion (TI) = 300 ms, flip angle = 20 º, number of excitations (NEX) = 1 (124 contiguous images, 1.5 mm thickness).

2.2 Manual segmentation

High-resolution MR images were manually segmented by two expert raters (MTP and YS, 2 and 3 atlases respectively). The presence of two manual raters allowed for both inter- and intra-rater comparisons of segmentations to validate the consistency of the manual segmentation protocol. Segmentations were done using Display software (also part of the MINC toolkit) to visualize the T1- and T2-weighted images. Display permits simultaneous viewing of the brain in all three directions (sagittal, coronal, and axial) facilitating an accurate and morphologically contiguous manual segmentation protocol.

2.2.1 Separating the cerebellum from the brain stem

While the cerebellum is distinct from the brainstem at the midline, the cerebellar white matter extends into the pons of the brainstem (Figure 1a, left). A virtual boundary between the cerebellar peduncles and the brainstem was defined by a plane perpendicular to the AC-PC line and tangent to the posterior border of the inferior colliculi. This plane was identified on the mid-sagittal slice of the cerebellum and superimposed on all sagittal slices (Figure 1a, centre) (Luft et al., 1998, Pierson et al., 2002, Hutchison et al., 2003). Cerebellar cortex that lay anterior to this plane was identified and segmented accordingly (Figure 1a, right).

2.2.2 Guidelines for segmentation

Variability in cerebellar morphology

There is significant variability in cerebellar lobule morphology between subjects, which poses a

challenge to both manual and automatic segmentation of the cerebellum. To simplify and maintain consistency in the manual segmentations, a number of general and lobule-specific rules were implemented.

1. Cerebellar lobules (I, II, III, IV, V, VI, Crus I, Crus II, VIIB, VIIIA, VIIIB, XI, X) and deep cerebellar nuclei (DCN) were segmented based on Schmahmann definitions (2000), which has consistently been used as one of the primary references for cerebellar lobule definitions, e.g., the original three-dimensional MRI atlas (Schmahmann et al., 1999) of the colin27 brain (Holmes et al., 1998), the Diedrichsen probabilistic atlas (Diedrichsen et al., 2009) and most recently in a multi-rater manual segmentation study (Bogovic et al., 2013a).

2. Intensity contrast on T1- and T2-weighted images were used to define lobule definitions. The most significant (and easily recognizable) intensity differences between adjacent lobules occur at the major fissures that divide them. As such, the lobules were defined as distinct branches of grey matter that surround white matter originating from the corpus medullare and separated from adjacent folia (lobule) by major fissures, as defined by Schmahmann et al. (2000) (Figure 1b). Fissure definitions were followed and confirmed by progressing slice-by-slice from medial to lateral, ensuring proper division of the lobules throughout from the vermis to the hemispheres.

3. White matter that extended into lobules was segmented as part of the lobule, while the corpus medullare was defined as white matter (Figure1c). The corpus medullare was defined using straight boundaries between the opposing limits of gray matter at the lobule-corpus medullare interface. Although in our high-resolution atlases the grey-white matter separation is distinct, this is generally not possible in standard T1-weighted MR data. We adapted this protocol to make our atlases widely and easily applicable to different imaging datasets with varying acquisition parameters.

4. Segmentations were performed slice-by-slice on sagittal and coronal views of the cerebellum. Consistency, both across the three planes (sagittal, axial, and coronal) as well as on the 3 dimensional surface representations (Figure1d) of the cerebellum, was evaluated to ensure accuracy and agreement with the reference literature and to maintain morphological consistency throughout.

A manual segmentation protocol of the cerebellum based on the Schmahmann definitions has been recently published (Bogovic et al., 2013a), yet it was not available during the development of our atlases and segmentation protocol. The manual segmentation protocol we have included is an independently developed protocol that is divergent to the Bogovic protocol in several ways:

1. High-resolution imaging data used: the use of high resolution imaging data allows for clear delineation of the lobule boundaries, while this is a difficult task in standard T1 images.

2. Consistent separation of cerebellum from the brainstem using previous methods.

3. Combination of cerebellar vermis and hemisphere: this approach allows for simpler definition of cerebellar anatomy, eliminating any variability due to an arbitrary definition of the vermis-hemispheric boundary.

These differences emphasize visible morphological boundaries to facilitate replication, and provide highly detailed illustrations to clearly identify differences in neuroanatomy between individuals. Finally, these atlases are available for public use. Thus, our manual segmentations can be used as a neuroanatomical baseline for groups that are interested in replicating our methodology.

2.2.3 Manual segmentation protocol- lobules and DCN

Cerebellar lobules were segmented lobule-by-lobule, starting with lobule I, II and proceeding counter-clockwise to lobule X (Figure 2a). In this section, we describe our manual segmentation protocol for identifying anatomical landmarks and the characteristics that were used to define cerebellar lobules and fissures.

Anterior lobules: I, II, III, IV, V

Lobules I and II are located most anterior of the anterior lobules and pressed close to the cerebellar peduncles posteriorly (Figure 2b). They can be seen as a single structure with no clear boundary and were segmented together as a single lobule. It is the smallest structure, with the narrowest boundaries, appearing as a thin and short branch of grey matter. Some subjects show lobules I and II as indistinguishable from the white matter near the midline (Figure 2c). In such cases, both white and grey matter were segmented as lobules I and II, and separated in more lateral sections where the white matter and lobules are distinct from one another (Figure 2c). Lobule III lies posterior to lobules I and II, between the precentral and preculminate fissures. It has a single branch with one folium that is smaller than both lobules IV and V (Figure 2d).

Lobules IV and V form a single tree-like structure sharing the same branch of white matter in the vermis of the cerebellum, and separate into distinct branches progressing laterally from the vermis. Therefore, on medial slices, the white matter was equally divided between the two lobules IV and V (Figure 2d) until clear separation was seen. The intraculminate fissure divides lobules IV and V, and was confirmed as the correct definition when the fissure expands laterally to divide lobules IV and V into one and two folia respectively (Figure 2d, right). In our atlases, the number of folia for IV and V range from 1-2 and 2-3 respectively, with lobule V generally consisting of one more folium than lobule IV.

Superior posterior lobules: VI, Crus I (VIIAf), Crus II (VIIAt), VIIB

Lobule VI is sizeable with 2-3 folia, posterior to the primary fissure and anterior to the superior posterior fissure (Figure 2e, f). The primary fissure is the most prominent fissure in sagittal midline views and separates the anterior and posterior lobules (Figure 2e, red line). Lobule VIIA is defined as consisting of VIIAf and VIIAt at the vermis which expands laterally to form Crus I and II respectively in the lateral, hemispheric divisions of the cerebellum (Schmahmann et al., 2000). In our segmentations, vermian lobules VIIAf and VIIAt were not distinguished from Crus I and II as it was not possible to derive a consistent intensity-based protocol to differentiate the vermian and hemispheric regions of the cerebellum.

Crus I lies posterior and inferior to lobule VI as a single folium (Figure 2f). In the midline Crus I is very thin but becomes more prominent laterally in the sagittal plane (Figure 2f). Crus I borders the horizontal fissure superiorly, which can be identified as a prominent fissure viewing the cerebellum in sagittal and coronal views (Figure 2g). Crus II lies below the horizontal fissure originating from a single branch of white matter, varying in size, number of folia, and relative location between subjects (Figure 2g, j, k)**.**Crus II was more prominent on the right side than on the left which is consistent with literature (Schmahmann et al., 2000). As well, it is often associated with lobule VIIB as part of the same branch of white matter in the left cerebellum (Figure 2h).Lobule VIIB can be identified as the last remaining lobule of the superior posterior lobe (Figure 2i). It is located superior to the prepyramidal/prebiventer fissure and anterior to the vermian part of Crus II in the midline, and spreads out as a single branch laterally (Figure 2i) varying in the number of folia (1-2) between the left and right hemispheres.

Inferior posterior and flocculonodular lobules- VIIIA, VIIIB, IX, X

Lobule VIIIA and VIIIB are located posterior to lobule IX and are part of the same branch at the midline (Figure 2i). Progressing laterally from the midline, lobule VIIIA and VIIIB are separated into distinct branches by the intrabiventer fissure. Lobule VIIIA varies in the number of folia across subjects. Three of our atlases were segmented with lobule VIIIA having two folia (1 male, 2 female) unilaterally, while the other two had one folium throughout—this is consistent with literature findings regarding variations in cerebellar lobule morphology in lobule VIIIA (Schmahmann et al., 2000, Bogovic et al., 2013a) (Figure 2j,k, red outlines).

The anterior portion of lobule VIIIB is seemingly concealed by lobule IX in the midline, and begins to appear in more lateral slices (Figure 2a, b). Lobules VIIIB and IX are closely associated physically and difficult to separate in the sagittal view (Figure 2a, j, k), often leading to inaccurate segmentations with lobule IX and VIIIB segmented as one lobule. To differentiate the two lobules correctly, the secondary fissure, which separates lobules VIIIB and IX, was identified in the coronal view as a clear and distinct fissure dividing the two lobules as a straight line (Figure 2l).

Lobule IX is the most prominent lobule in mid-sagittal slices and is pressed close to both lobules VIIIB and X. Lobule X is a small structure located anterior to lobule IX and appears in mid-sagittal slices as a round lobule and in more lateral slices takes on a triangular-shaped lobule inferior to the cerebellar peduncles (Figure 2 m).

Deep cerebellar nuclei- dentate, fastigial, interposed

T2-weighted MR images were used to segment the 3 DCN while viewing the cerebellum in the coronal and sagittal planes. In the T2-weighted images, the DCN stand out from the corpus medullare as a low signal intensity structure, with the surrounding brighter white matter (Figure 3a). The DCN were manually segmented in the high-resolution MRI data for completion but not used in the automatic segmentation algorithms as our validation set consisting of the standard T1-weighteddata does not possess sufficient contrast for the identification of these nuclei.

The dentate nucleus is the largest, most prominent nucleus and can be seen over most slices of the corpus medullare. It appears as a low intensity ovoid in the middle of the higher intensity corpus medulla with varying intensities in the middle. All intensities enclosed by the ovoid shape were segmented because there was no clear method to differentiate between the grey matter and white matter that compose the nuclei (Figure 3a, b). The interposed nucleus is a small round structure that appears in medial slices, centrally located in the cerebellum and close to the dentate (Figure 3b). It is located posterior to and larger than the fastigial nucleus. The fastigial nucleus spans ~12 coronal slices and appears as a small pea-shaped structure near the midline of the cerebellum (Figure 3c).

2.2.4 Protocol evaluation and validation metrics

To evaluate the inter-rater reliability of the manual segmentation protocol, three atlases produced by one rater (YS) were re-segmented by the other (MTP) in one hemisphere. The remaining two atlases originally segmented by one rater (MTP) were re-segmented in one hemisphere by the same rater to give intra-rater reliability measures. The validation set images were manually segmented for the entire cerebellum by a rater (YS) and the cerebellar lobules by another (MTP). Five images of the validation set were re-segmented (MTP) for the entire cerebellum to estimate inter-rater reliability. Three additional images were re-segmented (MTP) to give intra-rater reliability measures for the lobules and the entire cerebellum. A minimum two week interval was given between segmentation and re-segmentation for all test-retest evaluations.

Dice's volumetric Kappa was used to quantify inter- and intra-rater test-retest reliability and the accuracy of automatic segmentations with respect to manual segmentations:

$$\kappa = \frac{2a}{2a + b + c}$$

where a is the number of correctly labeled voxels and b + c is the sum of the voxels unique to each label set. A higher Kappa value denotes a higher degree of overlap, with a value of one representing perfect overlap. The kappa metric is sensitive to subtle differences in different segmentations and is a commonly used metric in the evaluation of different segmentation protocols (Chakravarty et al., 2008, Winterburn et al., 2013).

2.3 Automatic segmentation with multi-atlas based segmentation and MAGeT Brain.

Methodological considerations

Classically, atlas-based approaches use a single, manually labeled atlas of neuroanatomical structures and warp these labels to candidate subject using nonlinear registration methods(Collins et al., 1994, Chakravarty et al., 2008, Chakravarty et al., 2009). These methods may not be suitable in all cases since using a single atlas has limitations in modeling the neuroanatomical differences that exist between individuals and the segmentations are limited by the accuracy of the nonlinear registration technique used and by possible resampling errors. Multi-atlas based segmentation, where multiple manually segmented atlases are used to label a single subject (Heckemann et al., 2006), has been proposed as a way to ameliorate some of these issues. However, multi-atlas based algorithms typically require a large number of atlases as input (between 30-80), which are difficult and time consuming to produce particularly in structures such as the cerebellum where there is significant individual morphological variability. Our group has recently addressed this issue by developing the MAGeT Brain algorithm, an enhanced multi-atlas based segmentation approach optimized for use with a limited number of atlases (Chakravarty et al., 2013). In MAGeT Brain, a subset of images from the dataset to be segmented (typically 20) are chosen, which are automatically segmented using our atlases. The resultant segmentations are named the "template library", which is then used for automatic segmentation of the entire dataset.

In multi-atlas based segmentation, manually labeled atlas images are non-linearly registered to input images (Heckemann et al., 2006, Collins and Pruessner, 2010), and then atlas labels are warped into subject space by the resulting transformation. Labels for a subject are fused, via voxel-wise majority voting, to produce a unified segmentation for a single image (Figure 4a). For example, using five atlases would yield five possible segmentations for each image in the dataset, which would be fused to a single label (Figure 4a). The MAGeT Brain algorithm first performs non-linear registration of the each atlas image to each template library image (Figure 4b). A second registration step then occurs between each template library images and each input image. Lastly, the atlas labels are propagated to the input image by stepwise application of the atlas-to-template and template-to-subject non-linear transformations (Figure 4b). The resulting labels, or possible segmentations for a given input image are fused using voxel-wise majority voting into a final segmentation (Figure4b). As an example, using 5

atlases and a template library of 20 subjects, 100 possible segmentations (5 x 20) are fused to produce the final segmentation for an individual subject.

In the following sections, we describe validation experiments to determine the following:

1. Evaluate multi-atlas and MAGeT Brain segmentations with respect to a wide range of parameter settings from varying the number of atlases and templates.

2. Using the high-resolution atlases, quantify MAGeT Brain segmentation accuracy over all lobules and the entire cerebellum.

3. Compare MAGeT Brain with existing methods for lobule segmentation.

For all linear and non-linear registrations, we use the Automatic Normalization Tools (ANTS). ANTS provides great flexibility over the choice of transformation model, objective function, and the consistency of transformations (Avants et al., 2008). The following options were used for registrations using ANTS:

number-of-affine-iterations 10000x10000x10000x10000x10000
affine-gradient-descent-option 0.5x0.95x1.e-4x1.e-4
use-Histogram-Matching
MI-option 32x16000
-r Gauss[3,0], -t SyN[0.5] -i 100x100x100x20.

These options have been previously tested with MAGeT Brain for the hippocampus (Chakravarty et al., 2013, Pipitone et al., submitted) and showed robust registrations. All source code used for development and use of MAGeT Brain is available at https://github.com/pipitone/MAGeTbrain/.

2.3.1 Validation of multi-atlas based segmentation and MAGeT Brain

Using the validation set of twenty subjects and the five high-resolution cerebellar atlases, we evaluated multi-atlas segmentation and MAGeT Brain segmentation through a set of validation experiments. We additionally compared performance of MAGeT Brain to the SUIT toolbox (described below) (Diedrichsen, 2006, Diedrichsen et al., 2009).

Monte-Carlo validation using 9 atlases

Monte Carlo Cross-Validation (MCCV) was conducted to rigorously evaluate the effect of number of atlases and templates on the accuracy of segmentations. The experiment allows for an analysis of all possible parameter settings given a set of manually labeled atlases and a validation cohort of twenty subjects, comparing automatic segmentations to manual segmentations. This establishes a theoretical

validation for the rest of the experiments conducted with MAGeT Brain. We tested both multi-atlas and MAGeT Brain on the validation set using nine manually segmented cerebella(derived from the same data) as input atlases. The development of these atlases based on standard T1 images are described in sections 2.1 (image acquisition) and 2.2.4 (test-retest evaluation).

Ten validation rounds were completed where in each round 171 trials were conducted per image, reflecting all possible parameters settings varying the atlas library size (1-9), and the template library size (1-19). A total of 20 (subjects) x 9 (atlases) x 10 (validation rounds) = 1800 multi-atlas based segmentations are compared to 20 (subjects) x 9 (atlases) x 19 (templates) x 10 (validation rounds) = 34200 MAGeT Brain segmentations. For a given round of validation, random subsamples are taken from the atlas and template library for segmentation. In this way, we thoroughly assessed the impact of both the number of atlases and templates on segmentation quality. Segmentations from both methods were compared to manually segmented labels of the validation set. We evaluated whole cerebellum and lobule-specific Kappa scores, and the resulting difference in accuracy and precision between multi-atlas based segmentation and MAGeT Brain.

Leave-one-out cross validation using high-resolution atlases

Five high-resolution atlases of the cerebellum described in this manuscript were used to evaluate MAGeT Brain. The highly compact structure of the cerebellar foliations prevents clear viewing of lobule boundaries in standard resolution MRI. This poses problems for both manual and automatic segmentations of the cerebellum, while giving motivation for atlases based on higher resolution imaging data. The validation set of twenty subjects was automatically segmented using MAGeT Brain in a leave-one-out design where for a given image, the other nineteen images acted as the template library for segmentation. We evaluated lobule-wise and whole cerebellum segmentation accuracy of MAGeT Brain in this single setting, comparing MAGeT Brain segmentations with the twenty manual segmentations in the validation cohort.

Comparison between MAGeT Brain and SUIT toolbox

Currently, the SUIT toolbox (Diedrichsen, 2006, Diedrichsen et al., 2009) remains the sole publically available method for cerebellum lobule segmentation. We compare MAGeT Brain segmentations to those generated with the SUIT toolbox on the validation set.

The SUIT toolbox was used to automatically segment the validation set, following standard procedure for processing using MATLAB (http://www.icn.ucl.ac.uk/motorcontrol/imaging/suit_function.htm). The cerebellum is first isolated (suit_isolate) using automatic tissue segmentation and cropped for the cerebellum in an isolation map. Manual correction was not performed on the isolation map to strictly evaluate completely automatic segmentations between the methods. This also puts both techniques on equal footing for subsequent evaluation. Next, the isolation maps generated were used for

normalization (suit_normalize) to the SUIT atlas template (Spatially Unbiased Infra-tentorial Template; Diedrichsen, 2006). The resulting deformation maps are then used to resample (suit_reslice_inv) the probabilistic atlas of the cerebellum (Diedrichsen et al., 2009) into native space for each subject. The final segmentations are compared to manual segmentations of the validation set.

To minimize discrepancy between our segmentations and SUIT definitions, segmentations were converted to contain same or similar labels. The probabilistic atlas used in the SUIT toolbox divides the cerebellum into vermal and hemispheric regions (left and right). We combined lobule segmentations across the left, right hemispheres and the corresponding vermal regions for each lobule. For segmentations generated with MAGeT Brain, we combine left and right hemisphere divisions for comparison. Additionally, lobules I-II, III, and IV in MAGeT Brain segmentations were combined to match the segmentations generated from SUIT. The resulting converted labels consist of the following subregions: lobules I-IV, V, VI, Crus I, Crus II, VIIB, VIIIA, VIIIB, IX, and X. Crus I and lobule IX were excluded from comparison due to irreconcilable discrepancies, as the SUIT segmentations excluded portions of white matter within the lobules while our atlases did not. We compare the methods quantitatively in terms of Kappa over subregions and correlations between manual and automatically derived total cerebellar volume. We additionally evaluate both methods qualitatively through visual inspection.

3.0 Results

3.1 Manual segmentation protocol validation- atlases and validation set

Approximately 150 hours of training time was required for each rater to achieve consistent reliability. Segmentation time for the high-resolution data used for the atlases was a minimum of 40 hours per image over the ~450 sagittal slices in any given image. Segmentation times for the standard validation set images was 2-3 hours per image, with each image consisting of ~50 sagittal slices. In the high-resolution atlas set, high inter-rater Kappa was observed at the level of the individual lobules for the manual segmentation of the lobules $\kappa = 0.755$ to 0.929 (Table 1). Lobules I and II show the least consistency in manual segmentations ($\kappa = 0.755$). This is due to the narrow and small size of lobules I and II, as well as the variable morphology across subjects. Excluding lobules I and II, inter-rater Kappa for lobules ranged from $\kappa = 0.882$ to 0.929. Intra-rater reliability for individual lobules ranged from $\kappa = 0.762$ to 0.924, with the lowest being lobules I-II (0.762). Overall, the high test-retest validation results shows strong consistency of the manual segmentation protocol for the lobules and the entire cerebellum. Similar high reliability was found for the validation set images (Table 1), illustrating consistency of the manual segmentation protocol in standard T1-weightedMRI data as well.

3.2.1 Multi-atlas and MAGeT Brain cross-validation

The results of the Monte-Carlo Cross-Validation are presented in Figures 5, 6, Supplementary Figure 1, and Supplementary Tables 1 and 2.Segmentation accuracy is evaluated over the whole cerebellum and at the level of individual lobules. For both definitions, multi-atlas and MAGeT Brain segmentation accuracy is compared at each parameter setting across the ten rounds of validation. Kappa values are treated as distributions across validation rounds, where we compare mean difference in Kappa and the p-values of t-tests between multi-atlas and MAGeT Brain for each parameter setting. Figure 5a illustrates mean difference in Kappa between multi-atlas and MAGeT Brain with increasing template library size, while Figure 5b shows corresponding p-values of the t-tests between the two methods.

Whole cerebellum segmentation

Segmentation accuracy shows significant improvement with increased number of atlases used. This effect diminishes as the number approaches 8-9 for atlas library size, suggesting a plateau effect (Figure 5a, Supplementary Table 1). For MAGeT Brain segmentations, the use of the template library consistently results in greater accuracy over multi-atlas, although the difference is no longer significant at the 9 atlas setting, and approaches significance for 7-8 atlases (Figure 5b). In addition, the size of the template library contributes incrementally to improve segmentation accuracy (Figure 5).

Lobule segmentation

Similar trends exist at the lobule level- the effect of the number of atlases diminishes approaching 9 atlases (Figure 6, Supplementary Table 2). In general, MAGeT Brain improves lobule segmentation for low number of atlases used (less than five). The effect reaches significance for 1-3 atlases used for some of the lobules (Figure 6).With increased number of atlases, improvements in segmentation accuracy associated with the template library varies between lobules. Using nine atlases and 19 templates, MAGeT Brain performs worse than multi-atlas for the segmentation of lobules I-II, III, Crus I, Crus II, VIIB, VIIIA, VIIIB, while performing equally well or better for lobules IV, V, VI, IX, X (Figure 6, Supplementary Figure 1). Neither of these differences are significant ($p > 0.5$, Supplementary Figure 1), for both cases where MAGeT Brain performs better than multi-atlas and vice versa.

3.2.2 MAGeT Brain leave-one-out segmentation with high-resolution atlases

Results from the leave-one-out validation are presented in Figure 7 and Supplementary Table 3. For most of the lobules examined, MAGeT Brain accuracy ranges from acceptable (κ=0.7) to excellent (κ approaching 0.9), with a few exceptions (lobules I-II and VIIB). Lobules I-II and VIIB show the lowest accuracy (mean κ= 0.396, 0.599 respectively, Supplementary Table 3) due to relatively large variations in morphology (Figure 2b, c, h). Lobules VI, Crus I, and IX show highest Kappa with mean κ= 0.871, 0.887, 0.856 respectively. The remaining lobules III, IV, V, Crus II, VIIIA, VIIIB and X show variable

performance indicated by larger error. For the whole cerebellum, MAGeT Brain shows high Kappa: mean $\kappa = 0.925$ (Figure 7, Supplementary Table 3).

3.2.3 Comparing MAGeT Brain and SUIT segmentations

The evaluation between MAGeT Brain and SUIT are presented in Figures 8, 9, and 10. We evaluated automatic segmentations from both methods in the validation cohort of twenty subjects. Excluding Crus I and lobule IX, MAGeT Brain performs significantly better than SUIT over five of the eight subregions compared (Figure 8). MAGeT Brain performs significantly better for lobules IV and X ($p < 0.0001$), I-IV and V ($p < 0.05$) and marginally for Crus II ($p = 0.093$) and VIIIB ($p = 0.1344$). The differences are not significant for lobules VIIB and VIIIB ($p=0.52$ for both).

When comparing correlations between manually and automatically derived total cerebellar volumes, MAGeT Brain segmentations show greater correlation to manual volumes ($r = 0.976$, Figure 9) although the sample size examined is small. Qualitative visual examination of the same subject's manual (Figure 10b), MAGeT Brain (Figure 10c), SUIT (Figure 10d) segmentations, and the SUIT isolation map (Figure 10e) show that SUIT definitions do not differ significantly from our manual segmentations with the exception of Crus I and lobule IX (Figure 10d). SUIT is shown to produce over- and under-estimation of the cerebellum and its lobules (Figure 10d) due to the initial isolation map that was used for registration (Figure 10e).

4.0 Discussion

We developed five novel high-resolution in-vivo atlases of the cerebellum and validated the MAGeT Brain automated segmentation algorithm for the cerebellar lobules. The atlases are based on a consistent manual segmentation protocol, validated at multiple levels through test-retest reliability. The atlases, based on high-resolution imaging data, allow for clear viewing of the detailed foliations of the cerebellum and accurate identification of the cerebellar lobules. Using these atlases, we evaluated a segmentation framework that allows for robust automatic segmentation of the cerebellar lobules.

Comparing MAGeT Brain and multi-atlas based segmentation

Through the ten-fold Monte-Carlo Cross-Validation, we demonstrated the effects of the template library in improving segmentation accuracy for the whole cerebellum, and partly for the cerebellar lobules. The overall trends we observed were consistent with previous experiments with MAGeT Brain (Pipitone et al., submitted) in that MAGeT Brain provides improvements over multi-atlas based segmentation for relatively low number of atlases used, while the boost decreases with increased number of atlases. Based on Figure 5, it is likely that MAGeT Brain will continue to improve segmentation accuracy for the entire cerebellum past nine atlases, along with diminishing returns observed with increased number of atlases used. Regarding the cerebellar lobules, similar trends are

observed with increasing the atlas library while effectively reaching a plateau at 8-9 atlases used (Figure 6). At this setting, a possible concern may be the variable results observed in the accuracy of some lobule segmentations where increasing the template library slightly reduces Kappa. However, these lobule-wise differences are not significant (Supplementary Figure 1b, $p > 0.5$ for all lobules where mean Kappa for multi-atlas is higher than MAGeT Brain). Taken together, these results illustrate benefit of the template library over the traditional multi-atlas based segmentation framework. In this ten-fold validation, the atlases used are based on standard T1-weighted data where it was difficult to clearly visualize the detailed foliations of the cerebellum. It is possible that our accuracy was limited by the quality of the imaging data used for the atlases.

The high-resolution atlases of the cerebellum provide consistent delineation of the foliations due to its superior resolution. Using these atlases, MAGeT Brain produced accurate segmentations of the cerebellar lobules in a leave-one-out design (Figure 7). We demonstrate that most lobules achieve Kappa values in the range of 0.7-0.9, while lobules I-II and VIIB show low Kappa due to irreconcilable differences in morphology between subjects. This is noted in the manual segmentation protocol where lobules I-II and VIIB morphology were shown to differ amongst the high-resolution atlases (Figure 2b, c, h). Despite this, the rest of the lobules show high segmentation accuracy providing further support for MAGeT Brain as a segmentation tool for automatic identification of the cerebellum and its lobules.

Comparing MAGeT Brain and existing methods

The greatest benefit of MAGeT Brain and the use of the template library would be its adaptability in novel datasets where manual segmentation are not readily available. This is especially apparent in the comparison between MAGeT Brain and the SUIT toolbox. The SUIT toolbox utilizes a probabilistic atlas of the cerebellum and its lobules based on data from 20 healthy individuals, which is then used to automatically segment the cerebellum. A recent study evaluated the use of the SUIT algorithm in patients with cerebellar degeneration and reported inaccurate segmentations of the cerebellar lobules (mean Kappa of 0.55; Bogovic et al., 2013a). While we cannot make claims with respect to the algorithmic implementation details from this group, there are several reasons that may have led to the lower levels of overlap reported from this group. As suggested by one of the authors of the SUIT algorithm (Diedrichsen, personal communication), we speculate that the sources of error may be attributable due to differences in the definition of the "ground truth" between labs, thereby leading to extremely low overlap. Nonetheless, it has been noted by our group and others that it can be difficult to evaluate segmentation algorithms across laboratories as the definition of a “gold standard” or a “ground truth” may be marginally (or even significantly) different across groups (Chakravarty et al, 2008, 2009; Klein et al., 2009). Although we have not evaluated our methodology in subjects suffering from neurodegeneration in this manuscript (as Bogovic et al. did), our analysis demonstrates far superior accuracy of SUIT in comparison to what was previously reported by Bogovic and colleagues. However,

we note that it is possible that our cerebellum and lobule definitions may inadvertently differ from the Diedrichsen group's. As such, we have laboriously attempted to minimize discrepancies between labeling schemes between our atlases and the probabilistic atlas used in the SUIT toolbox by converting both sets of segmentations into a more common, comparable set of definitions. Our hope is that such discrepancies should be minor since we have also modeled the definitions of our lobules using the Schmahmann definitions (Schmahmann et al., 1999).

Excluding subregions where the definitions differed significantly between MAGeT Brain and SUIT, MAGeT Brain performed significantly better across most of the lobules (Figure 8) and for the whole cerebellum (Figure 9). This difference may be due to the following factors:

1. Use of multiple atlases and templates over a single probabilistic atlas:

- The probabilistic atlas used in the SUIT toolbox is derived from the anatomical data from 20 manually segmented labels, where the probabilities are essentially preset in a single template. The SUIT toolbox then depends on a single registration to bring subjects into template space, while MAGeT Brain performs exhaustive registrations between the atlases, template library, and the subject images to minimize different sources of error through each successive step.

- As such, the lack of additional registrations points to a possible inflexibility of the SUIT toolbox to adapt to different populations (such as those studied in Bogovic et al., 2013a) and image acquisitions where scan quality may not be optimal. This is especially noted in Bogovic et al., where it was noted that the probabilistic atlas is biased towards a single (SUIT) template and not applicable to subjects with cerebellar degeneration.

2. Registration algorithms employed:

- The SUIT toolbox uses a registration algorithm (SPM normalization) that was shown to perform poorly compared to the ANTS registration used in MAGeT Brain (Klein et al., 2009). Klein et al. note the SPM normalization step used in SUIT ranking as one of the least effective, while ANTS was among the most effective.

3. Need for manual intervention:

- Another concern with the SUIT toolbox is that the segmentation errors we presented may be due to the lack of manual correction in the first isolation step. The SUIT toolbox recommends hand correction at this stage, where the user would need to manually correct the cerebellar isolation map before proceeding to the next steps. However, one of the goals of this manuscript is to compare fully-automated segmentation methods. The type of manual segmentation suggested may not be possible on large datasets where thousands of images need to be

segmented without manual intervention. We also note here that we did not perform any manual corrections after this stage of the processing pipeline. We also note that we believe this to be a critical issue as all steps after the initial isolation stage depend on the accuracy of the cerebellum isolation. As such, if the isolation is inadequate, downstream processing such as linear- and non-linear registrations would be inaccurate.

Despite this, the SUIT toolbox may have benefits over MAGeT Brain considering factors such as ease of platform and implementation, computing time, and performance in normal populations. The SUIT toolbox is built to work with MATLAB and provides a simple graphical user interface for image processing. Current MATLAB and SPM users may find SUIT easier to implement into their analyses than MAGeT Brain, which relies on a command line interface in a linux terminal. Most importantly, the SUIT toolbox requires less than 10 minutes for fully processing a single MR image, while MAGeT Brain requires up to 6 hours using a parallel computing interface.

To the best of our knowledge, only two previous methods, the SUIT algorithm (Diedrichsen, 2006, Diedrichsen et al., 2009) and ACCLAIM (Bogovic et al., 2013b) performed automatic segmentations that included the cerebellar lobules. ACCLAIM is a recently published method that has shown to produce significantly more accurate segmentations than the SUIT and multi-atlas algorithms. In this study, MAGeT Brain showed comparable performance when compared to ACCLAIM across all lobules (Bogovic et al., 2013b). However, in contrast to both SUIT and ACCLAIM, MAGeT Brain allows for a customizable template and atlas library that would allow the algorithm to take advantage of the existing anatomical variations in the dataset to better model neuroanatomical characteristics.

Compared to other previous methods with respect to whole-cerebellum automatic segmentation methods, our results show that MAGeT Brain is comparable to, or in most cases exceeds previous methods reported in literature (Table 2). The average and maximum Kappa values in our study exceed most methods, while only a few studies (Heckemann et al., 2006, Lijn et al., 2011) presented higher Kappa values ($\kappa = 0.95$) with regard to the entire cerebellum. However, it is important to note that our method only used 5 atlases as opposed to other multi-atlas based approaches in literature that use a much larger atlas set (typically 20+) for multi-atlas segmentation methods.

Future steps and possibility of further improvement

It is critical to note that the voting schemes used in our validation experiments were the most basic schemes, where we take the highest occurring label at each voxel (otherwise known as majority-voting). There are more sophisticated voting methods available such as the STAPLE algorithm (Warfield et al., 2004) or joint label fusion (Wang and Yuskevich, 2013). We excluded cross-correlation weighted and normalized mutual information weighted voting from analyses as we did not see any significant benefit of using these voting schemes (data not shown). Recent studies from our group (Chakravarty et al.,

2013, Pipitone et al., submitted) has also shown that there is no significant benefit in the use of these types of weighted votes (at least in the context of MAGeT Brain).

Further, it is highly likely that additional pre- and post-processing steps would significantly improve segmentation accuracy. Although such topics are out of scope for this study, these type of image processing steps could such as brain extraction, intensity normalization, and post-processing including alternative voting schemes, bias correction, and segmentation refinement through other algorithms could possibly boost the accuracy of the our algorithm. For example, due to the large number of segmentations generated with MAGeT Brain in comparison to multi-atlas, it would be possible to employ strategies such as post-processing error correction (Hanson et al., 2012). This is made possible by the sheer number of registrations employed in MAGeT Brain, providing more opportunities for "truth" propagation between each successive stage. A recent study (Hanson et al., 2012) employed machine learning (Freund and Schapire, 1995) for segmentation error correction (Wang et al., 2011), where systematic biases were detected by comparison between manual and automatic segmentations and corrected through incorporating additional features of the imaging data such as intensity and spatial relationships. These options provide further support MAGeT Brain in its relative flexibility over multi-atlas based segmentation to integrate additional options in segmentation.

Another usage of MAGeT Brain to overcome variation in cerebellar lobule morphology in different populations would be to use MAGeT Brain in multiple iterations, where: 1) A subset of the dataset would be automatically segmented using MAGeT Brain. Resulting labels would be thoroughly checked for accuracy, and segmentations deemed as excellent would be chosen. 2) Selected segmentations would be manually corrected by an expert rater. 3) Resulting novel atlases would be added to the atlas set, and MAGeT Brain would then be run on the entire dataset. Developing novel cerebellar atlases based on the dataset using MAGeT Brain would allow for quick production of a larger number of atlases, while normalizing the inherent differences in imaging acquisition between the high-resolution atlas set and typically lower-resolution datasets. This would also allow for creation of atlases based on non-healthy subjects with differences in cerebellar anatomy (i.e. cerebellar degeneration), therefore increasing the accuracy of the segmentation by better modeling the neuroanatomical differences that exist in certain populations. Similar multi-iterative strategies have been previously suggested in other automatic algorithms where existing priors (as in manually segmented atlases) can be used to generate additional libraries to better represent large variations in neuroanatomy amongst subjects (Eskildsen et al., 2012).

Conclusions

We have developed a set of 5 novel, high-resolution in-vivo atlases of the cerebellum that are publically available. Using these atlases, we have validated an automatic segmentation method of the cerebellar lobules using the MAGeT Brain algorithm. Our method produced accurate segmentations of

the cerebellar lobules with the use of a template library, with greater accuracy than previously published techniques. MAGeT Brain produced robust lobule segmentations that are comparable to manual segmentations, validating MAGeT Brain as a useful tool for automatic segmentation of the cerebellum.

Acknowledgements

We wish to thank Anusha Ravichandran for help with image acquisition, and we acknowledge support from the CAMH Foundation, thanks to Michael and Sonja Koerner, the Kimel Family, and the Paul E. Garfinkel New Investigator Catalyst Award. MMC is funded by the W. Garfield Weston Foundation and ANV is funded by the Canadian Institutes of Health Research, Ontario Mental Health Foundation, NARSAD, and the National Institute of Mental Health (R01MH099167). Computations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto.

Figure Captions:

Figure 1. General overview of manual segmentation protocol for the cerebellum. a) Separation of the cerebellum from the brainstem. Red dotted line indicates AC-PC line, and the green is the line perpendicular to and posterior to the inferior colliculus. The plane used excludes parts of grey matter in the more lateral sections, which were included in the manual segmentations. 3D representation of the cerebellum indicates the approximate location in space of the sagittal slices illustrated. b) Cerebellar fissures (Schmahmann et al., 2000) used to define lobules: i) precentral ii) preculminate iii) intraculminate iv) primary v) superior posterior vi) horizontal vii) ansoparamedian viii) prepyramidal/prebiventer ix) intrabiventer x) secondary. Coronal view was used for confirmation of horizontal and secondary fissures. c) Cerebellar lobules defined according to fissures with the left and right hemispheres containing a different set of labels. White matter that is enclosed by grey matter of the lobules were segmented as part of the lobules. d) Final 3D surface representation used to confirm correct definition of lobules.

Figure 2. Detailed manual segmentation protocol of the individual lobules. a) Example of sagittal slices viewed in manual segmentation. The three images correspond to mid-sagittal, and 2 progressively lateral slices. b, c) Lobules I and II were segmented together as a single lobule. c) In some subjects, lobules I and II are indistinguishable from the white matter at the midline, then in more

lateral slices can be distinguished as a separate lobule. d) Separation of lobules IV and V at the midline. Red arrow indicates intraculminate fissure, which expands laterally to divide lobules IV and V into one and two folia respectively. e) The primary fissure is the most prominent fissure and divides lobules V and VI. f) Crus I is difficult to identify at the midline, but becomes prominent progressing laterally slice-by-slice. g) The horizontal fissure (in red) divides Crus I and Crus II, which is a significant fissure that was identified and confirmed in both sagittal and coronal views. h) Crus II is more prominent in the right hemisphere of the cerebellum than the left. In both hemispheres, Crus II is closely associated physically with lobule VIIB, often as part of the same branch of white matter. i) Lobule VIIB is the last lobule of the superior posterior lobe, and located above lobule VIIIA/B and the prepyramidal/prebiventer fissure. Lobule VIIB spreads out as a single folium laterally starting from the midline, while lobules VIIIA and VIIIB form a single branch at midline but divide into respective branches progressing laterally. j, k) Lobule VIIIA differs in the number of folia between subjects. Cerebellum that have two folia of lobule VIIIA seem to have previous lobules (superior posterior lobules) pushed upwards, therefore altering the relative location of lobules such as Crus II and VIIB. l) The secondary fissure is best identified in coronal view as a straight line dividing lobules VIIIB and IX. Lobule IX is the largest of the cerebellar lobules and pressed together with lobule VIIIB. m) Progressing from medial to lateral sections of the cerebellum slice-by-slice, lobule X seems to appear in the midline and reappear in more lateral slices as a triangular-shaped lobule.

Figure 3. Segmentation of the deep cerebellar nuclei (DCN) in high-resolution T2-weighted MR data. a) Sagittal: dentate nucleus (green) appears as a low-intensity oval in the centre of the corpus medullare and cerebellum. b) The interposed nuclei (purple) is centrally located and close to the dentate nuclei medially. c) The fastigial nuclei (teal) are small, pea-shaped structures located anterior and superior to the interposed nuclei.

Figure 4. Automatic segmentation with multi-atlas and MAGeT Brain, illustrated with three atlases and small template library for clarity. a) Multi-atlas segmentation labels the dataset by nonlinear registration of atlases onto a single image, in which the resulting labels are fused using voxel-wise majority voting to produce a final segmentation for a single image. b) The MAGeT Brain algorithm (Chakravarty et al., 2012) is a modified multi-atlas algorithm:

1) The algorithm first labels a subset of the dataset (template library) by applying the transformation resulting from nonlinear registration between each atlas and each template.

2) A second registration step takes each of the labels in the template library, which are propagated to the subject image via nonlinear registration and fused using voxel-wise majority voting to produce a final segmentation. By using 5 atlases and a template library of 20, 100 possible segmentations (5 X 20) would be fused to produce the final segmentation for an individual subject.

Figure 5. Comparison between MAGeT Brain and multi-atlas segmentation for the entire cerebellum in a ten-round Monte-Carlo Cross-Validation. a) Mean differences in Kappa between MAGeT Brain and multi-atlas with varying atlas library (1-9) and template library size (1-19). b) p-values of the t-tests between MAGeT Brain and multi-atlas at each parameter setting (atlases, templates). Points are shown where mean Kappa values for MAGeT Brain were higher than multi-atlas.

Figure 6. Comparison between MAGeT Brain and multi-atlas segmentation for cerebellar lobules in a ten-round Monte-Carlo Cross-Validation. Mean differences in Kappa between MAGeT Brain and multi-atlas with varying atlas library (1-9) and template library size (1-19) are shown.

Figure 7. MAGeT Brain lobule and whole cerebellum segmentation accuracy in a leave-one-out design in the validation set of twenty subjects. High-resolution atlases derived in this manuscript were used.

Figure 8. Kappa comparison between MAGeT Brain and the SUIT toolbox (Diedrichsen, 2006, Diedrichsen et al., 2009) over eight subregions. Crus I and lobule IX were excluded due to discrepancies in lobule definitions between the methods. Asterices indicate level of significance for t-tests conducted between Kappa values for each subregion: *- $p < 0.01$, **- $p < 0.05$, ***- $p < 0.0001$.

Figure 9. Linear regression and pearson r coefficients between manually derived and automatically derived whole cerebellum volumes for MAGeT Brain and SUIT.

Figure 10. Sample images from a single subject spanning equivalent ten slices from the medial to lateral, illustrating: a)Matching T1 images, b) Manual segmentations, c) MAGeT Brain segmentations, d) SUIT segmentations, e) SUIT isolation map. Red arrows indicate regions where the isolation step led to inaccurate registration and segmentations. Blue arrows indicate regions where the anatomical definition between MAGeT Brain and SUIT segmentations differ significantly (Crus I, lobule IX), and was excluded from analysis.

Supplementary Figure 1. p-values of the t-tests between MAGeT Brain and multi-atlas at each parameter setting (atlases, templates) for the individual lobules, where a) Points are shown where mean

Kappa values for MAGeT Brain were higher than multi-atlas, and b) Points are shown where mean Kappa values for MAGeT Brain were lower than multi-atlas.

Supplementary Table 1. Mean Kappa and standard deviation for whole cerebellum segmentation Kappa values, comparing MAGeT Brain (using 19 templates) and multi-atlas. Atlases range from one to nine atlases in the ten-round Monte-Carlo Cross-Validation.

Supplementary Table 2. Mean Kappa and standard deviation for lobule segmentation Kappa values, comparing MAGeT Brain (using 19 templates) and multi-atlas. Atlases range from one to nine atlases in the ten-round Monte-Carlo Cross-Validation.

Supplementary Table 3. Lobule and whole cerebellum segmentation results for MAGeT Brain, using five high-resolution atlases in the leave-one-out experiment.

References

Allen, G., Buxton, R., Wong, E., Courchesne, E., 1997. Attentional activation of the cerebellum independent of motor involvement. Science 275, 1940–1943.

Andreasen, N.C., O'Leary, D.S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G.L., Boles-Ponto, L.L., Hichwa, R.D., 1995. PET studies of memory: novel versus practiced free recall of word lists. NeuroImage 2, 296–305.

Andreasen, N.C., O'Leary, D.S., Paradiso, S., Cizadlo, T., Arndt, S., Watkins, G.L., Ponto, L.L., Hichwa, R.D., 1999. The cerebellum plays a role in conscious episodic memory retrieval. Human Brain Mapping 8, 226–234.

Balda çara, L., Nery-Fernandes, F., Rocha, M., Quarantini, L.C., Rocha, G.G., Guimar ães, J.L., Ara újo, C., Oliveira, I., Miranda-Scippa, A., Jackowski, A., 2011. Is cerebellar volume related to bipolar disorder? Journal of Affective Disorders 135, 305–309.

Bastian, A.J., Mink, J.W., Kaufman, B.A., Thach, W.T., 1998. Posterior vermal split syndrome. Annals of Neurology 44, 601–610.

Bastian, A.J., 2006. Learning to predict the future: the cerebellum adapts feed forward movement control. CurrOpin Neurobiol 16, 645–9.

Bogovic, J.A., Jedynak, B., Rigg, R., Du, A., Landman, B.A., Prince, J.L., Ying, S.H., 2013a. Approaching expert results using a hierarchical cerebellum parcellation protocol for multiple inexpert human raters. Neuroimage 64, 616-29.

Bogovic, J.A., Bazin, P.L., Ying, S.H., Prince, J.L., 2013b. Automated Segmentation of the Cerebellar Lobules using Boundary Specific Classification and Evolution. 23rd Conference on Information Processing in Medical Imaging (IPMI 2013).

Buckner, R.L., Krienen, F.M., Castellanos, A., Diaz, J.C., Yeo, B.T., 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. Journal of Neurophysiology 106(5), 2322-45.

Courchesne, E., Yeung-Courchesne, R., Press, G.A., Hesselink, J.R., Jernigan, T.L., 1988. Hypoplasia of cerebellar vermal lobules VI and VII in autism. N Engl J Med 318(21):1349–1354

Chakravarty, M.M., Steadman, P., van Eede M.C., Calcott, R.D., Gu, V., Shaw, P., Raznahan, A., Louis Collins D., Lerch, J.P., 2012. Performing label-fusion-based segmentation using multiple automatically generated templates. Human Brain Mapping, doi: 10.1002/hbm.22092.

Chakravarty, M.M., Sadikot, A.F., Germann, J., Hellier, P., Bertrand, G., Collins, D.L., 2009. Comparison of piece-wise linear, linear and nonlinear atlas-to-patient warping techniques:

Analysis of the labeling of subcortical nuclei for functional neurosurgical applications. Human Brain Mapping 30(11):3574-95.

Chakravarty, M.M., Sadikot, A.F., Germann, J., Bertrand, G., Collins, D.L., 2008. Towards a validation of atlas warping techniques. Medical Image Analysis 12(6):713–26.

Chen, J., Cohen, M.L., Lerner, A.J., Yang, Y., Herrup, K., 2010. DNA damage and cell cycle events implicate cerebellar dentate nucleus neurons as targets of Alzheimer's disease. Molecular Neurodegeneration, 20, 60.

Collins, D. L., Neelin, P., Peters, T. M., Evans, A. C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18(2), 192-205.

Collins, D. L., Pruessner, J. C., 2010.Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Neuroimage, 52(4), 1355-1366.

Diedrichsen, J., 2006. A spatially unbiased atlas template of the human cerebellum. Neuroimage, 33(1):127-38.

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., Ramnani, N., 2009. A probabilistic MR atlas of the human cerebellum. Neuroimage, 2009 46(1):39-46.

Eskildsen, S.F., Coupé, P., Fonov, V., Manjón, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Østergaard, L.R., Collins, D.L., Alzheimer's Disease Neuroimaging Initiative, 2012. BEaST: brain extraction based on nonlocal segmentation technique. Neuroimage, 59(3):2362-73.

Fukutani, Y., Cairns, N.J., Rossor, M.N., Lantos, P.L., 1996. Purkinje cell loss and astrocytosis in the cerebellum in familial and sporadic Alzheimer's disease. Neurosci Lett, 214:33-36.

Freund,Y., and Schapire, R.,1995. A decision-theoretic generalization of online learning and an application to boosting. *Computational Learning Theory*, ed. P. Vitányi (Berlin/Heidelberg: Springer), 23–37.

Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C.F., Menon, V., Greicius, M.D., 2009. Distinct cerebellar contributions to intrinsic connectivity networks. Journal of Neuroscience 29(26):8586-94.

Habel, U., Klein, M., Kellermann, T., Shah, N.J., Schneider, F., 2005.Same or different? Neural correlates of happy and sad mood in healthy males. NeuroImage 26, 206–214.

Hanson, J.L., Suh, J.W., Nacewicz, B.M., Sutterer, M.J., Cayo, A.A., Stodola, D.E., Burghy, C.A., Wang, H., Avants, B.B., Yushkevich, P.A., Essex, M.J., Pollak, S.D., Davidson, R.J.,

2012. Robust automated amygdala segmentation via multi-atlas defeomorphic registration. *Front Neurosci*, doi: 10.3389/fnins.2012.00166.

Hartmann, S.L., Parks, M.H., Martin, P.R., Dawant, B.M., 1999. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part II, validation on severely atrophied brains. IEEE Trans Med Imaging, 18(10):917-26.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage, 33(1):115–126.

Hutchinson, S., Lee, L.H., Gaab, N., Schlaug, G., 2003.Cerebellar volume of musicians. Cereb Cortex, 13(9):943-9.

Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C., 1998. Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr, 22(2):324–333.

Hwang, J., Kim, J., Han, Y., Park, H., 2011. An automatic cerebellum extraction method in T1-weighted brain MR images using an active contour model with a shape prior. Magn Reson Imaging, 29(7):1014-22.

Ichimiya, T., Okubo, Y., Suhara, T., Sudo, Y., 2001.Reduced volume of the cerebellar vermis in neuroleptic-naive schizophrenia. Biological Psychiatry 49, 20–27.

Isambert, A., Dhermain, F., Bidault, F., Commowick, O., Bondiau, P.Y., Malandain, G., Lefkopoulos, D., 2008. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. Radiother Oncol, 87(1):93-9.

Jacobsen, L.K., Giedd, J.N., Berquin, P.C., Krain, A.L., Hamburger, S.D., Kumra, S., Rapoport, J.L., 1997. Quantitative morphology of the cerebellum and fourth ventricle in childhood-onset schizophrenia. Am J Psychiatry, 154:1663–1669.

James, A.C., James, S., Smith, D.M., Javaloyes, A., 2004. Cerebellar, prefrontal cortex, and thalamic volumes over two time points in adolescent-onset schizophrenia. Am J Psychiatry, 161:1023–1029.

Kibby, M.Y., Fancher, J.B., Markanen, R., Hynd, G.W., 2008. A quantitative magnetic resonance imaging analysis of the cerebellar deficit hypothesis of dyslexia. Journal of Child Neurology 23, 368–380.

Kim, S.G., Ugurbil, K., Strick, P.L., 1994. Activation of a cerebellar output nucleus during cognitive processing. Science 265, 949–951.

Krienen, F.M., Buckner, R.L., 2009. Segregated fronto-cerebellar circuits revealed by intrinsic functional connectivity. Cereb Cortex19(10):2485-97.

Levitt, J.J., McCarley, R.W., Nestor, P.G., Petrescu, C., Donnino, R., Hirayasu, Y., Kikinis, R., Jolesz, F.A., Shenton, M.E., 1999. Quantitative volumetric MRI study of the cerebellum and vermis in schizophrenia: Clinical and cognitive correlates. Am J Psychiatry, 156:1105-1107.

Loeber, R.T., Cintron, C.M., Yurgelun-Todd, D.A., 2001. Morphometry of individual cerebellar lobules in schizophrenia. Am J Psychiatry, 158:952–954.

Luft, A.R., Skalej,M., Welte, D., Kolb, R., Bürk, K., Schulz, J.B., Klockgether, T., Voigt, K., 1998. A new semi automated, three-dimensional technique allowing precise quantification of total and regional cerebellar volume using MRI. Magn Reson Med, 40(1):143-51.

Middleton, F.A., Strick, P.L., 1994. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science, 266(5184):458-461.

Mostofsky, S.H., Powell, S.K., Simmonds, D.J., Goldberg, M.C., Caffo, B., Pekar, J.J., 2009.Decreased connectivity and cerebellar activity in autism during motor task performance. Brain 132, 2413–2425.

Mottolese, C., Richard, N., Harquel, S., Szathmari, A., Sirigu, A., Desmurget, M., 2013. Mapping motor representations in the human cerebellum. Brain, 136(1):330-42.

Muller, R.A., Pierce, K., Ambrose, J.B., Allen, G., Courchesne, E., 2001. Atypical patterns of cerebral motor activation in autism: a functional magnetic resonance study. Biol Psychiatry 49(8):665–676.

Murakami, J.W., Courchesne, E., Press, G.A., Yeung-Courchesne, R., Hesselink, J.R., 1989. Reduced cerebellar hemisphere size and its relationship to vermal hypoplasia in autism. Arch Neurol 46(6):689–694.

Nenadic, I., Sauer, H., Gaser, C., 2010. Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. Neuroimage, 49:1153–1160.

Nopoulos, P.C., Ceilley, J.W., Gailis, E.A., Andreasen, N.C., 1999. An MRI study of cerebellar vermis morphology in patients with schizophrenia: Evidence in support of the cognitive dysmetria concept. Biol Psychiatry, 46:703–71

O'Halloran, C.J., Kinsella, G.J., Storey, E., 2012. The cerebellum and neuropsychological functioning: a critical review. Journal of Clinical and Experimental Neuropsychology 34, 35–56.

O'Reilly, J.X., Beckmann, C.F., Tomassini,V., Ramnani, N., Johansen-Berg, H., 2010. Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity. Cereb Cortex, 20(4):953-65

Okugawa, G., Sedvall, G., Nordstrom, M., Andreasen, N.C., Pierson, R., Magnotta, V., Agartz, I., 2002. Selective reduction of the posterior superior vermis in men with chronic schizophrenia. Schizophr Res, 55:61– 67.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E., 1989. Positron emission tomographic studies of the processing of single words. J CognNeurosci, 1:153–170.

Pierson, R., Corson, P.W., Sears, L.L., Alicata, D., Magnotta, V., Oleary, D., Andreasen, N.C., 2002. Manual and semiautomated measurement of cerebellar subregions on MR images. Neuroimage, 17(1):61-76.

Piven, J., Nehme, E., Simon, J., Barta, P., Pearlson, G., Folstein, S.E., 1992. Magnetic resonance imaging in autism: measurement of the cerebellum, pons, and fourth ventricle. Biol Psychiatry, 31(5):491–504

Powell, S., Magnotta, V.A., Johnson, H., Jammalamadaka, V.K., Pierson, R., Andreasen, N.C., 2008. Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. Neuroimage, 39(1):238-47.

Riva, D., Giorgi, C., 2000. The cerebellum contributes to higher functions during development: evidence from a series of children surgically treated for posterior fossa tumours. Brain, 123(5):1051–61.

Schlerf, J.E., Verstynen, T.D., Ivry, R.B., Spencer, R.M., 2010. Evidence of a novel somatopic map in the human neocerebellum during complex actions. J Neurophysiol, 103(6):3330-6.

Schmahmann, J.D., 1991. An emerging concept: the cerebellar contribution to higher function. Archives of Neurology 48, 1178–1187.

Schmahmann, J.D., Sherman, J.C., 1998. The cerebellar cognitive affective syndrome. Brain 121, 561–579.

Schmahmann, J.D., Doyon, J., McDonald, D., Holmes, C., Lavoie, K., Hurwitz, A.S., Kabani, N., Toga, A., Evans, A., Petrides, M., 1999. Three- dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. NeuroImage, 10:233-260.

Schmahmann, J.D., Doyon, J., Toga, A., Petrides, M., Evans, A., 2000. MRI Atlas of the Human Cerebellum. Academic Press, San Diego, CA.

Schmahmann, J.D., 2004. Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. J Neuropsychiatry Clin Neurosci, 16(3):367-378.

Schmahmann, J.D., Weilburg, J.B., Sherman, J.C., 2007. The neuropsychiatry of the cerebellum: insights from the clinic. Cerebellum, 6(3): 254-267.

Sled, J. G, 1998.A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE transactions on medical imaging, 17(1):87-97.

Stoodley, C.J., Schmahmann, J.D., 2009. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. NeuroImage, 44(2):489--501.

Stoodley, C.J., 2012. The cerebellum and cognition: evidence from functional imaging studies. Cerebellum, 11(2): 352-65.

Strick, P.L., Dum, R.P., Fiez, J.A., 2009. Cerebellum and nonmotor function. Annual Review of Neuroscience 32, 413–434.

Thach W.T., Goodkin, H.P., Keating, J.G., 1992. The cerebellum and the adaptive coordination of movement. Annu Rev Neurosci, 15:403–422.

Thach, W.T., 1996. On the specific role of the cerebellum in motor learning and cognition: clues from PET activation and lesion studies in humans. Behavioral and Brain Sciences 19, 411-431.

Thomann, P.A., Schläfer, C., Seidl, U., Dos Santos, V., Essig, M., Schröder, J., 2008.The cerebellum in mild cognitive impairment and Alzheimer's disease—a structural MRI study. J. Psychiatr. Res. 42, 1198–1202.

Turner, B.M., Paradiso, S., Marvel, C., Pierson, R., Ponto, L.L.B., Hichwa, R.D., Robinson, R.G., 2007.The cerebellum and emotional experience. Neuropsychologia 45, 1331-1341.

van der Lijn, F., de Bruijne, M., Klein, S., den Heijer, T., Hoogendam, Y.Y., van der Lugt, A., Breteler, M.M., Niessen, W.J., 2012. Automated brain structure segmentation based on atlas registration and appearance models. IEEE Trans Med Imaging. 2012 Feb;31(2):276-86.

Wang, H., Das, S.R., Suh, J.W., Altinay, M., Pluta, J., Craige, C., Avants, B., Yuskevich, P.A.; Alzheimer's Disease Neuroimaging Initiative. A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation. Neuroimage 55, 968–985.

Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging, 23(7):903-21.

Wegiel, J., Wisniewski, H.M., Dziewiatkowski, J., Badmajew, E., Tarnawski, M., Reisberg, B., Mlodzik, B., De Leon, M.J., Miller, D.C., 1999.Cerebellar atrophy in Alzheimer's disease-clinicopathological correlations. Brain Res, 818(1):41-50.

Winterburn, J. L., Pruessner, J. C., Chavez, S., Schira, M., Lobaugh, N. J., Voineskos, A. N., & Chakravarty, M.M., 2013. A novel in vivo atlas of human hippocampal subfields using high-resolution 3T magnetic resonance imaging. NeuroImage, 74:254-65

Zhao, L., Ruotsalainen, U., Hirvonen, J., Hietala, J., Tohka, J., 2010. Automatic cerebral and cerebellar hemisphere segmentation in 3D MRI: adaptive disconnection algorithm. Med Image Anal, 14(3):360-72.

Table 1. Inter- and intra-rater test-retest validation for the cerebellar atlases and validation set measured over individual lobules and the entire cerebellum. Values are given as Kappa mean and range. a) Three of the high-resolution atlases were re-segmented to give inter-rater reliability measures, and two for intra-rater measures. b) Three subjects in the validation set were re-segmented on one hemisphere for the cerebellar lobules for intra-rater reliability.

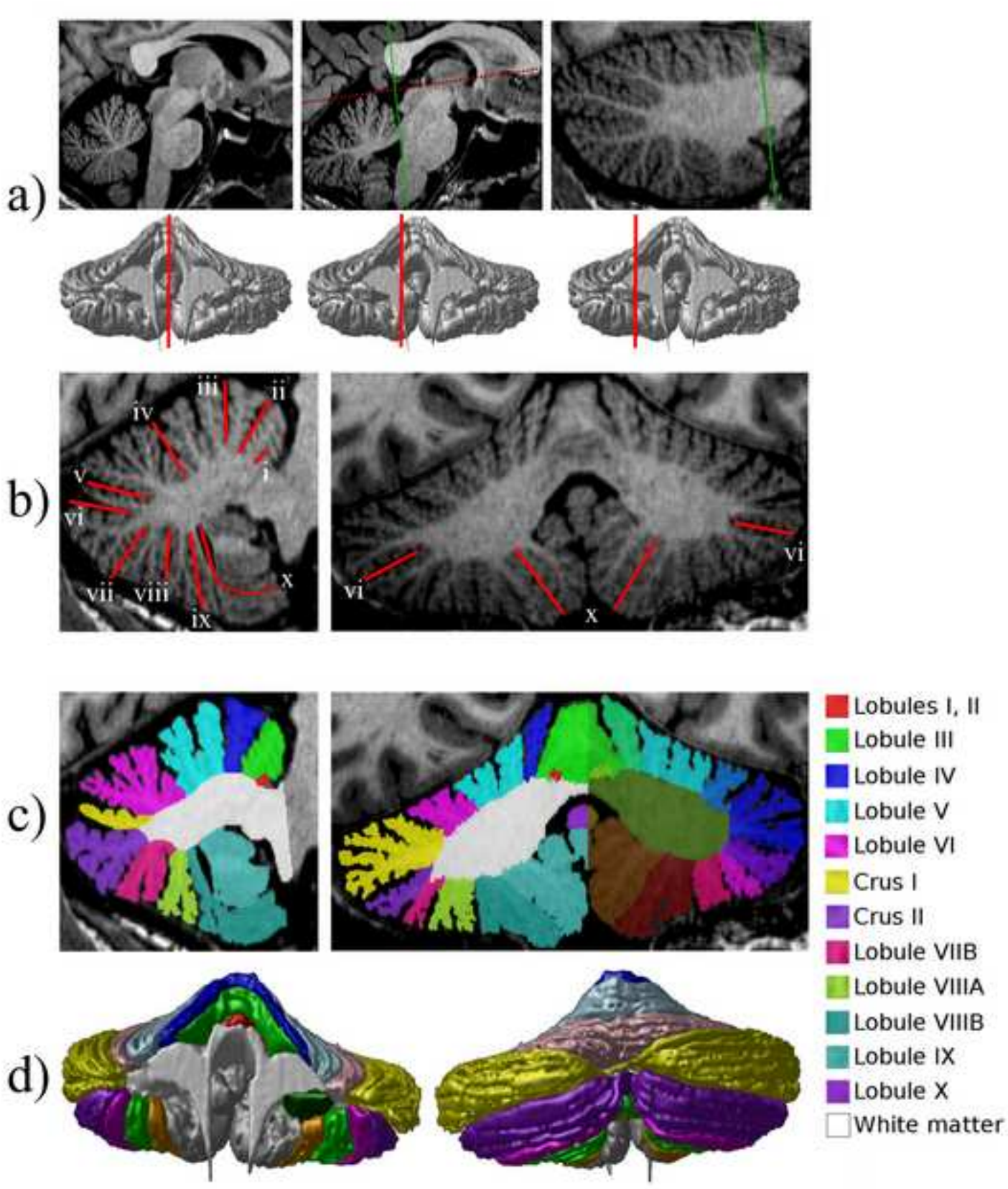
	a) Atlases				**b) Validation set**	
	Inter-rater		**Intra-rater**		**Intra-rater**	
Lobule	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**
Lobule I, II	0.755	(0.544-0.874)	0.762	(0.641-0.883)	0.639	(0.590-0.704)
Lobule III	0.900	(0.893-0.91)	0.904	(0.889-0.919)	0.751	(0.655-0.852)
Lobule IV	0.907	(0.887-0.927)	0.890	(0.872-0.907)	0.818	(0.689-0.893)
Lobule V	0.908	(0.881-0.928)	0.923	(0.913-0.932)	0.881	(0.870-0.894)
Lobule VI	0.921	(0.886-0.943)	0.924	(0.912-0.935)	0.912	(0.898-0.922)
Crus I	0.928	(0.892-0.947)	0.892	(0.865-0.919)	0.904	(0.889-0.920)
Crus II	0.929	(0.916-0.954)	0.910	(0.906-0.913)	0.900	(0.894-0.906)
Lobule VIIB	0.898	(0.870-0.926)	0.899	(0.894-0.903)	0.863	(0.833-0.886)
Lobule VIIIA	0.900	(0.862-0.935)	0.854	(0.823-0.885)	0.860	(0.841-0.887)
Lobule VIIIB	0.923	(0.894-0.946)	0.849	(0.788-0.909)	0.833	(0.809-0.866)
Lobule IX	0.926	(0.919-0.936)	0.892	(0.845-0.939)	0.874	(0.810-0.912)
Lobule X	0.882	(0.866-0.894)	0.848	(0.840-0.856)	0.760	(0.736-0.789)
Entire cerebellum	0.951	(0.944-0.955)	0.932	(0.929-0.935)	0.941	(0.932-0.948)

Table 3. Review of previous automatic segmentation methods for the cerebellum.

Year	Author	Method	Segmented structures (Dice overlap)
1999	Hartmann et al.	Automated, single atlas-based segmentation	Whole cerebellum (0.94-95)
2002	Pierson et al.	Semi-automated, artificial neural network (ANN), multi-atlas with 30 atlases	Lobe-based (0.79-87)
2006	Heckemann et al.	Automated, multi-atlas based segmentation with 29 atlases	Whole cerebellum (0.95-97)
2008	Powell et al.	Automated, four methods: template, probabilistic, ANN, support vector machine (SVM algorithm).	Lobe-based: template (0.81-0.91), probabilistic (0.82-0.91), ANN (0.87-0.93), SVM (0.86-0.93)
2008	Isambert et al.	Automated, single atlas-based	Whole cerebellum (mean 0.84, range 0.79-0.86)
2009	Diedrichsen et al.	Automated, probabilistic atlas based on 20 healthy scans	Lobules (mean 0.55, from Bogovic et al, 2013a)
2010	Zhao et al.	Automated, Adaptive Disconnection algorithm	Whole cerebellum (no Dice overlap reported)
2011	Hwang et al.	Automated, multi-atlas based cerebellum extraction with 5 atlases	Whole cerebellum (0.882-885)
2012	Gui et al.	Automated, morphology-driven (intensity-based) segmentation of neonatal brains	Whole cerebellum (0.87)
2012	Ljin et al.	Automated, multi-atlas (17-18) combined with appearance-based segmentation	Whole cerebellum (0.937-0.954)

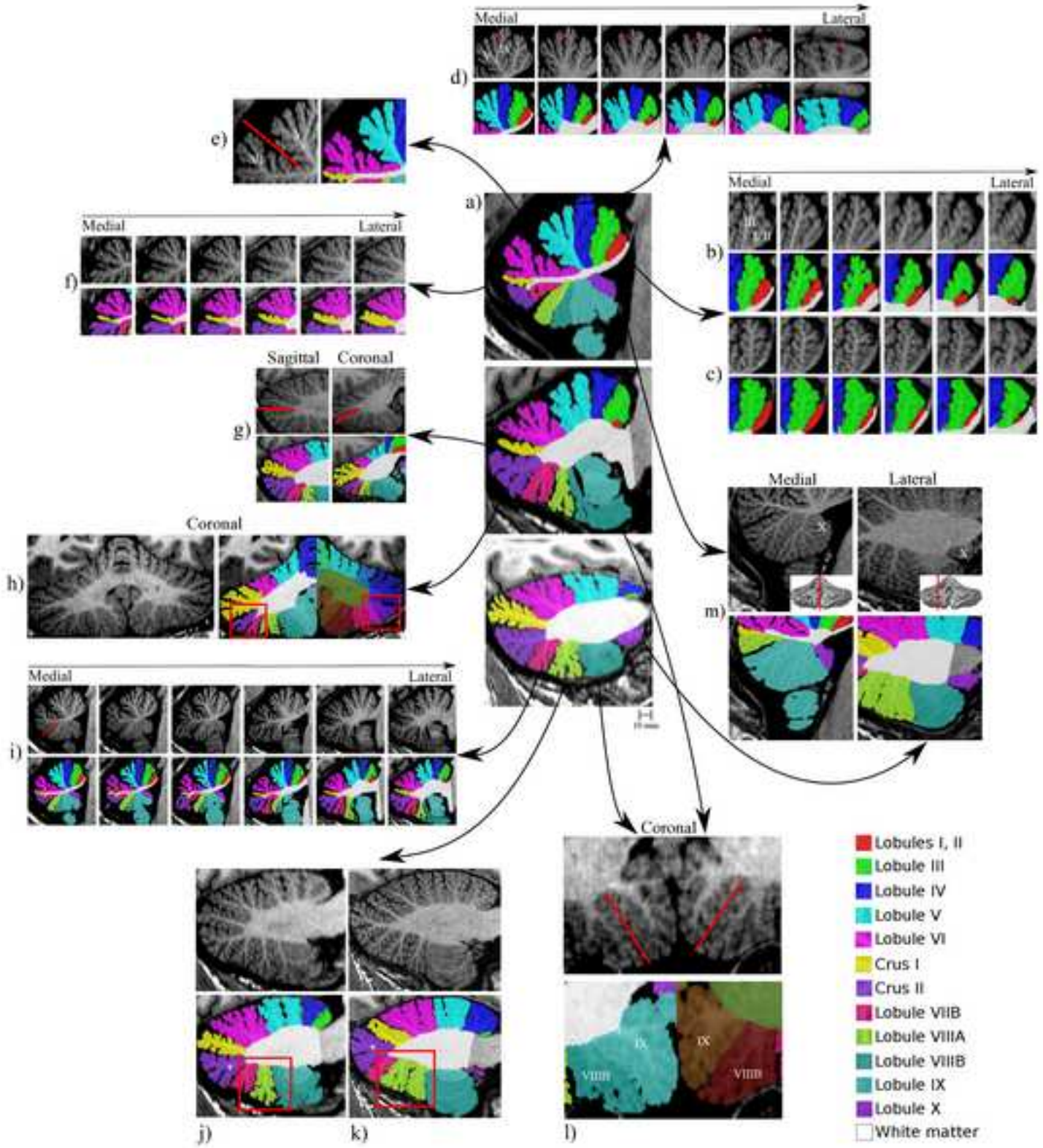
9. Figure

Click here to download high resolution image



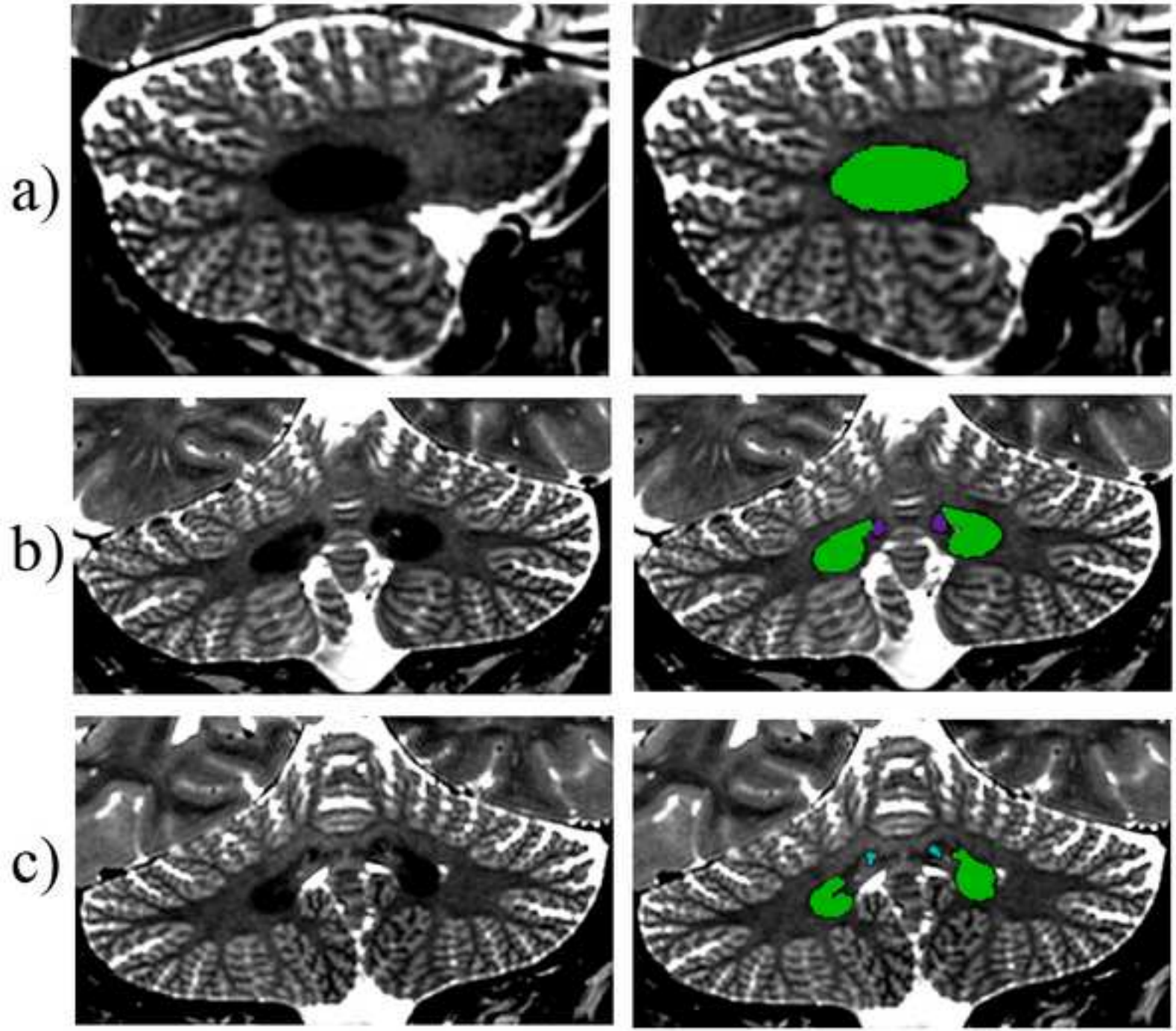
9. Figure

Click here to download high resolution image



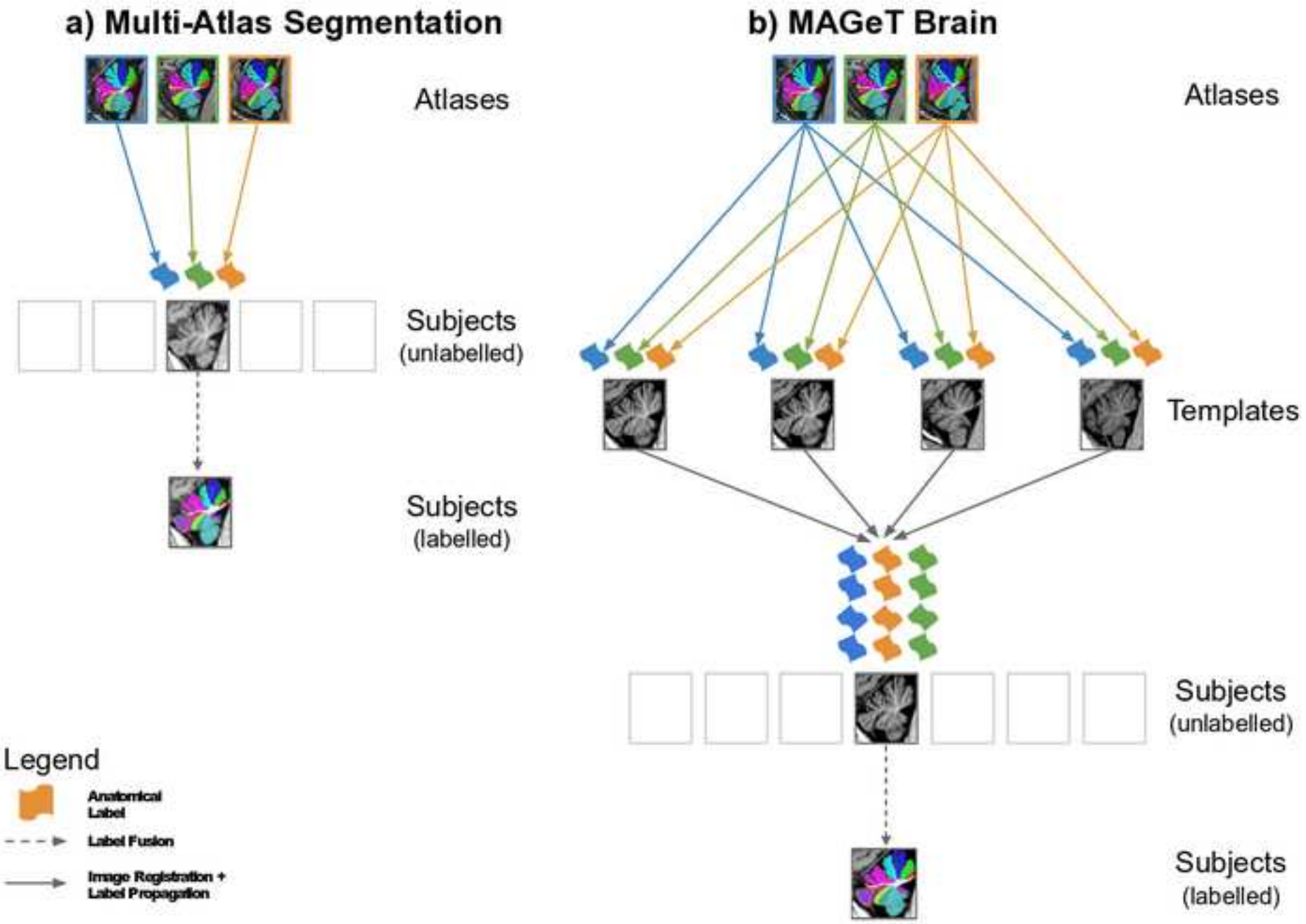
9. Figure

Click here to download high resolution image



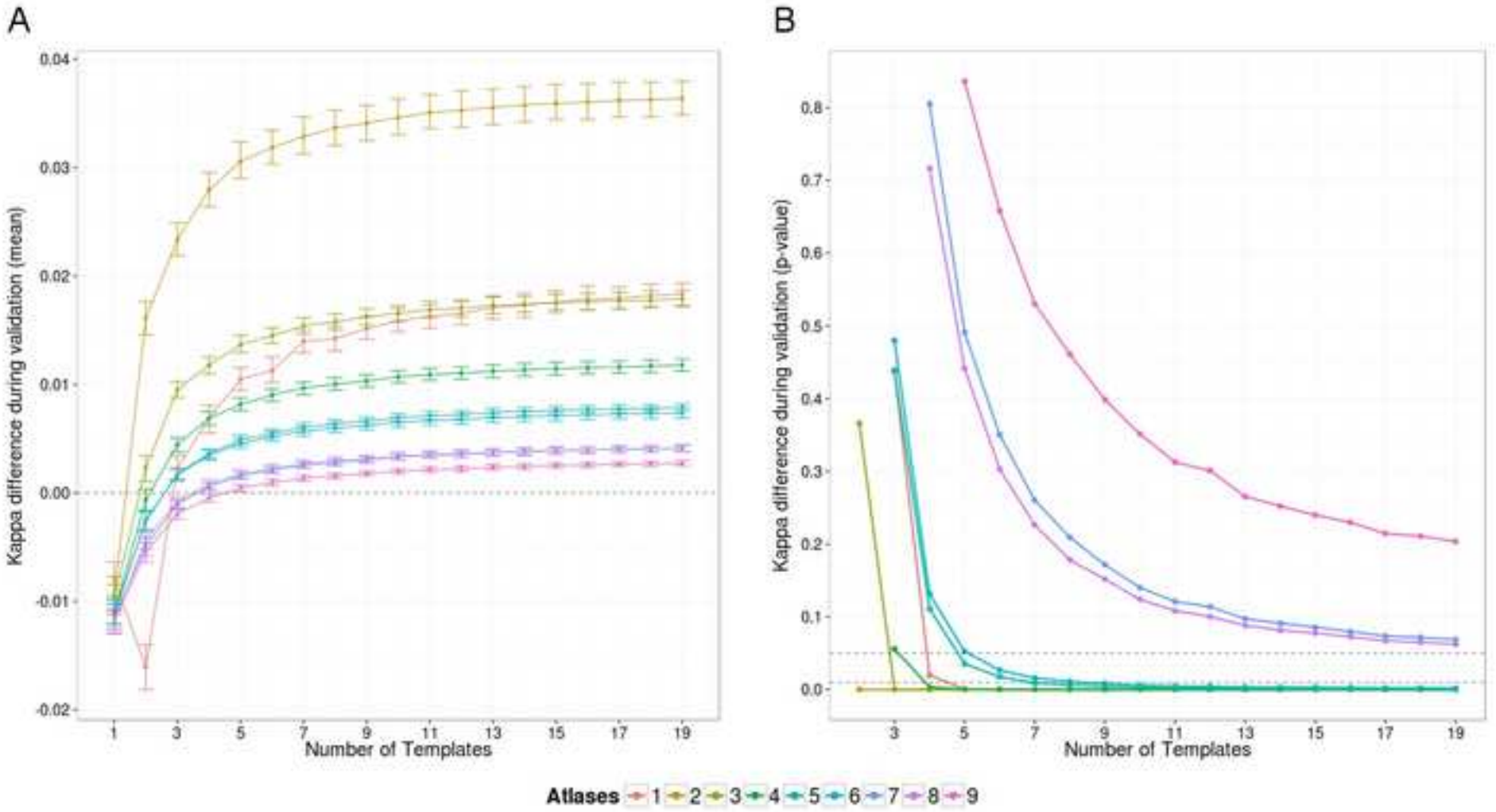
9. Figure

Click here to download high resolution image



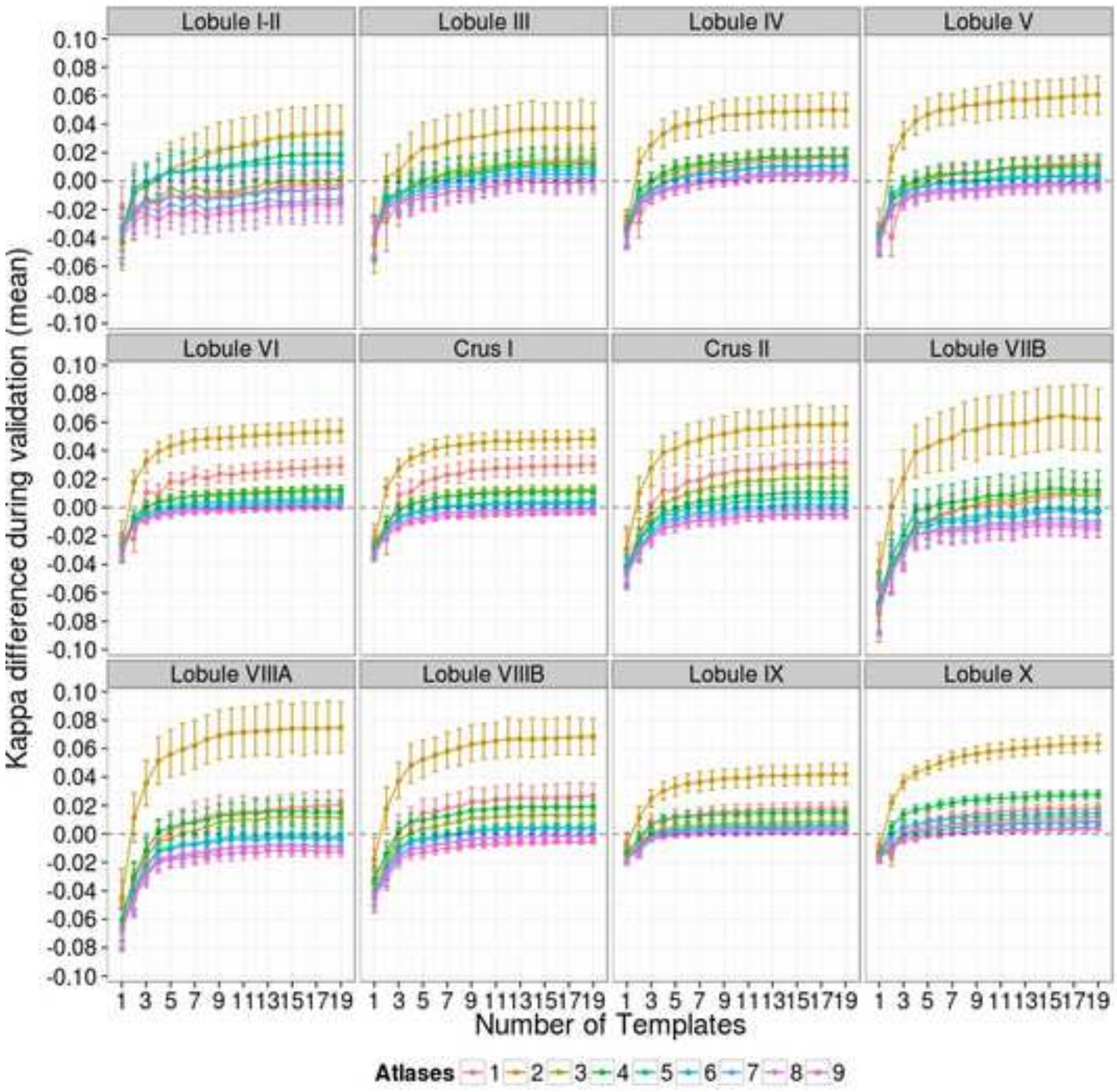
9. Figure

Click here to download high resolution image



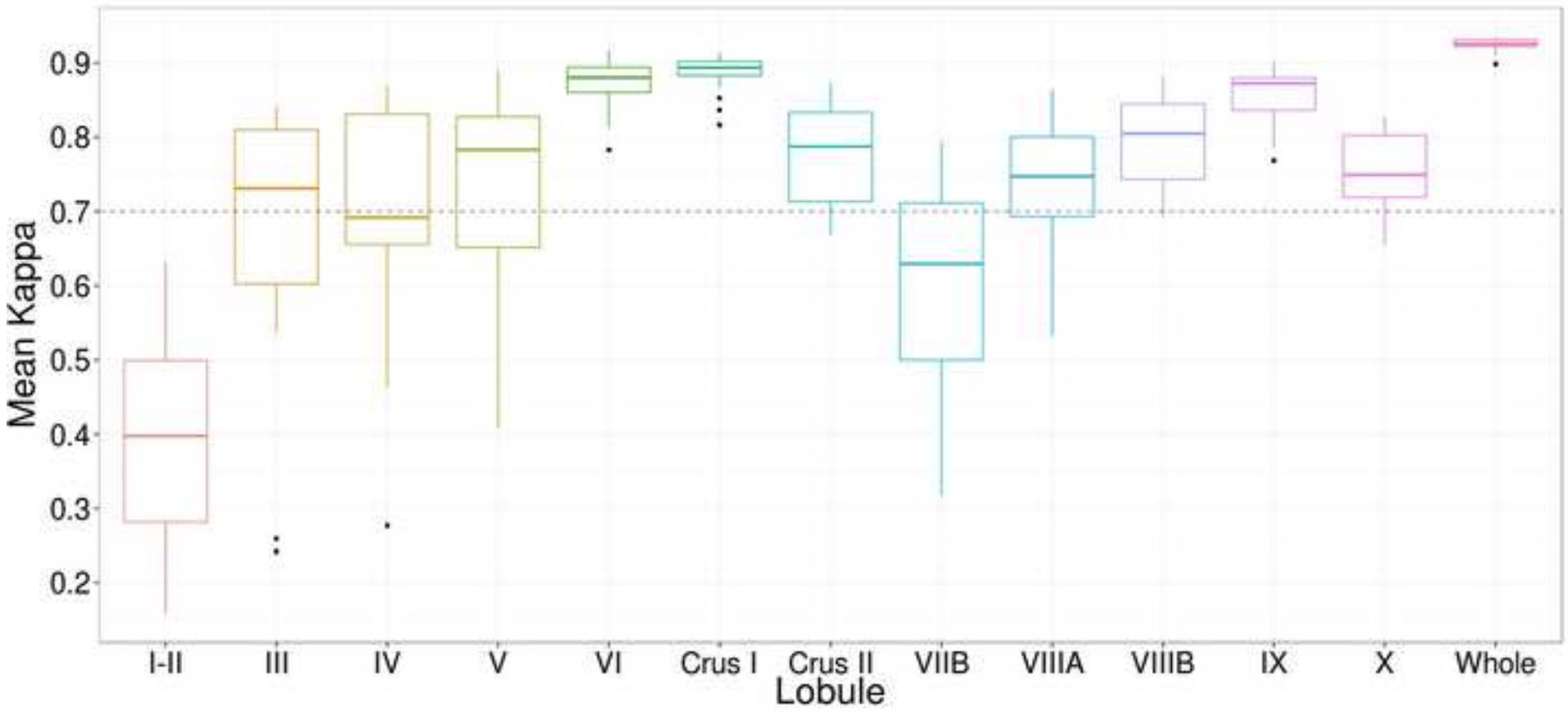
9. Figure

Click here to download high resolution image



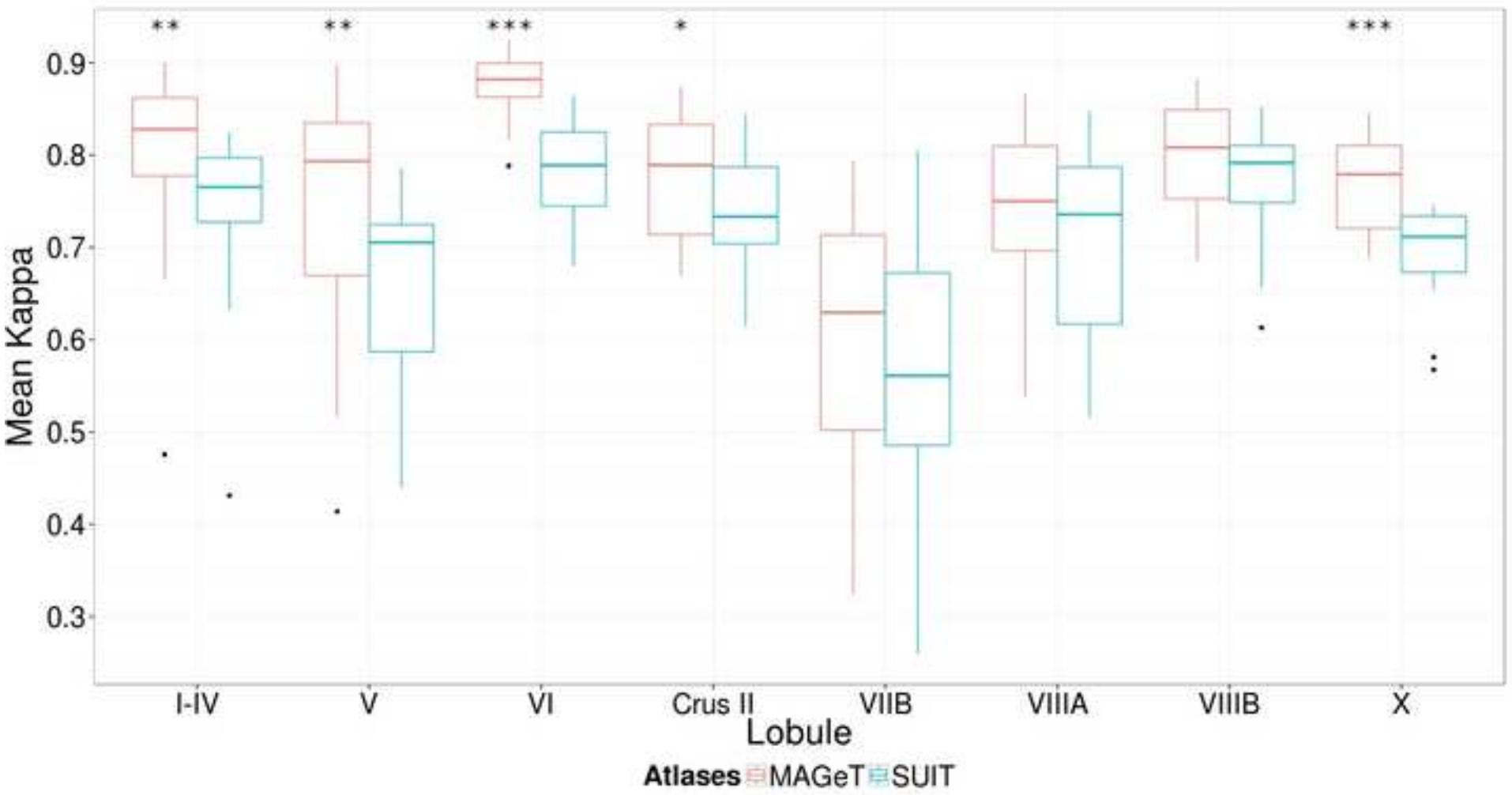
9. Figure

Click here to download high resolution image



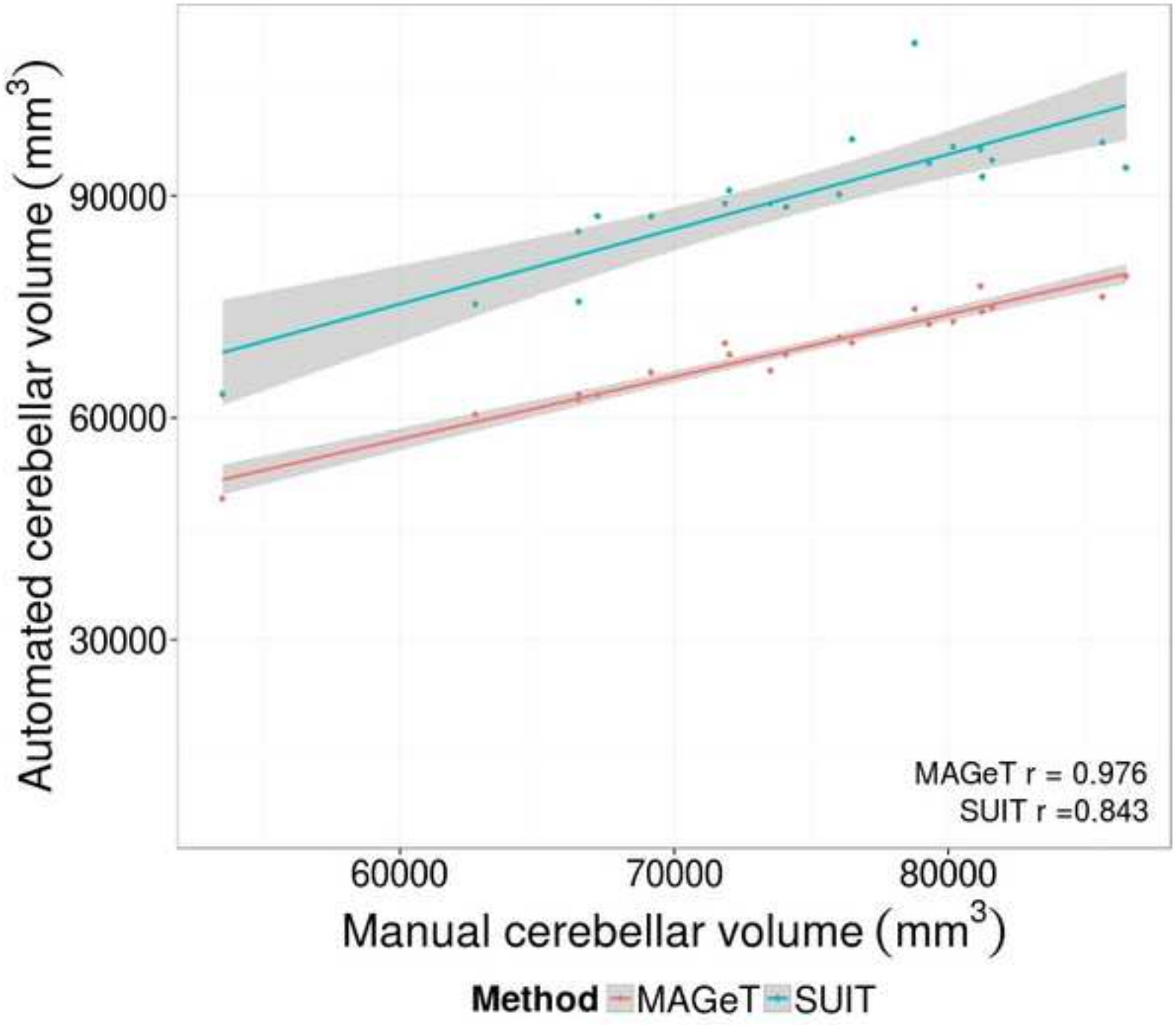
9. Figure

Click here to download high resolution image



9. Figure

Click here to download high resolution image



9. Figure

Click here to download high resolution image