Automated morphometry for mouse brain MRI through structural parcellation and thickness estimation

Da Ma

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Centre for Medical Image Computing
Centre for Advanced Biomedical Imaging
University College London, United Kingdom

Primary supervisor: Prof. Sebastien Ourselin
Secondary supervisor: Prof. Mark Lythgoe
Tertiary supervisor: Dr. M Jorge Cardoso
Tertiary supervisor: Dr. Marc Modat
University College London

February 15, 2016

I, Da Ma, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

To my parents

Abstract

Quantitative morphometric analysis is an important tool in neuroimaging for the study of understanding the physiology of development, normal aging, disease pathology and treatment effect. However, compared to clinical study, image analysis methods specific to preclinical neuroimaging are still lacking. The aim of this PhD thesis is to achieve automatic quantitative structural analysis of mouse brain MRI. This thesis focuses on two quantitative methods which have been widely accepted as quantitative imaging biomarkers: brain structure segmentation and cortical thickness estimation.

Firstly, a multi-atlas based structural parcellation framework has been constructed, which incorporates preprocessing steps such as intensity non-uniformity correction and multi-atlas based brain extraction, followed by non-rigid registration and local weighted multi-atlas label fusion. Validation of the framework demonstrated improved performance compared to single-atlas-based structural parcellation, as well as to global weighted multi-atlas label fusion methods.

The framework has been further applied to *in vivo* and *ex vivo* data acquired from the same cohort so that the respective volumetric analysis can be compared. The results reveal a non-uniform distribution of volume changes from the *in vivo* to the post-mortem brain. In addition, volumetric analysis based on the segmented structures showed similar statistical power on *in vivo* or *ex vivo* data within the same cohort.

Secondly, a framework to segment the mouse cerebellar cortex sublayers from brain MRI data and estimate the thickness of the corresponding layers has been developed. Application of the framework on the experimental data demonstrated its ability to distinguish sublayer thickness variation between transgenic strains and their wild-type littermate, which cannot be detected using full cortical thickness measurements alone.

In conclusion, two quantitative morphometric analysis frameworks have been pre-

sented in this thesis. This demonstrated the successful application of translational quantitative methods to preclinical mouse brain MRI.

Acknowledgements

The four years' PhD study is a precious treasure for me. It won't be possible without so many kind support, selflessly guidance and continuous encouragement that surrounded me.

Firstly, I would like to thank my primary supervisor, Prof. Sébastien Ourselin, for his valuable advice and guidance throughout my PhD. His continuous support and encouragement also helped me build confidence, learn fast, and develop the courage to tackle challenging tasks.

My gratitude also goes to my secondary supervisor, Prof. Mark Lythgoe, for the various advices he gave me and opportunities he provided that opened my mind and made me learn how to put what I learned into practice.

My grateful appreciation to the two tertiary supervisors in the Centre for Medical Image Computing: Dr. Jorge Cardoso and Dr. Marc Modat, for their thoughtful guidance and help throughout my PhD. I've learned so much from them through countless discussions.

I would like to thank all my colleagues and friends at CMIC (Centre for Medical Image Computing) and CABI (Centre for Advanced Biomedical Imaging). A lot of the joy of the four years study comes from the beautiful time working together with so many lovely people. Specially thanks to Pankaj and Maria, who shared their valuable experience with me as well as their invaluable jokes, both light and dark; and to Abi, Alex, Carla, Carole, Chris and Jonas for the work, chat and joy we've shared together shoulder to shoulder in CMIC; and also to Holly, Jack, James, Ian, Nick and Oz for all the struggling an strive we've gone through together as a team in CABI.

I am also very thankful to everyone I collaborated with over these last years within the Centre for Medical Image Computing, the Centre for Advanced Biomedical Imaging, the Institute of Neurology, the Crick Institute and the CUBRIC in Cardiff.

Finally, I want to express my thanks to my dear family members, who have always been along with me. Special Thanks to my cousin Cheng, who studies in the same University as me and makes my study life abroad much more fun. My dearest thanks to my parents, who always bore with me with patience, support and care for me no matter what happened.

List of Publications

Peer-reviewed Journal Papers

1. **D. Ma**, M.J. Cardoso, M. Modat, N. Powell, J. Wells, H.E. Holmes, F. Wiseman, V. Tybulewicz, E. Fisher, M.F. Lythgoe, S. Ourselin: Automatic structural parcellation of mouse brain MRI using multi-atlas label fusion. PloS one, 9(1), e86576, 2014.

2. B.A. Duffy, K.P. Chun, **D. Ma**, M.F. Lythgoe, R.C. Scott: Dexamethasone exacerbates cerebral edema and brain injury following lithium-pilocarpine induced status epilepticus. Neurobiology of disease, 63, 229-236, 2014.

3. H.E. Holmes, N. Colgan, O. Ismail, **D. Ma**, N.M. Powell, J.M. OĆallaghan, I.F. Harrison, R.A. Johnson, T.K. Murray, Z. Ahmed, M. Heggenes, A. Fisher, M.J. Cardoso, M. Modat, S. Walker-Samuel, E.M.C. Fisher, S. Ourselin, M.J. OŃeill, J.A. Wells, E.C. Collins, M.F. Lythgoe: Imaging the accumulation and suppression of tau pathology using multi-parametric MRI. Neurobiology of Aging, 2015.

Peer-reviewed Conference Papers

1. **D. Ma**, M.J. Cardoso, M.A. Zuluaga, M. Modat, N. Powell, F. Wiseman, V. Tybulewicz, E. Fisher, M.F. Lythgoe, S. Ourselin: Grey matter sublayer thickness estimation in the mouse cerebellum. MICCAI 2015.

2. **D. Ma**, M.J. Cardoso, M. Modat, N. Powell, H. Holmes, M.F. Lythgoe, and S. Ourselin: Multi-atlas segmentation applied to in vivo mouse brain MRI. MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling.

Peer-reviewed Conference Abstracts Proceeding

1. N. Powell, **D. Ma**, F. Prados, M. Modat, J. Cardoso, H. Holmes, O. Ismail, N. Colgan, M. O'Neill, E. Collins, M.F. Lythgoe, S. Ourselin: Longitudinal whole-brain atrophy measurement in a mouse model of tauopathy using the Generalised Boundary Shift Integral. ISMRM, 2015.

2. N. M Powell, H. Holmes, **D. Ma**, M. Modat, J. Cardoso, F. Wiseman, V. Tybulewicz, E. Fisher, M.F. Lythgoe, S. Ourselin: Tensor-Based Morphometry reveals structural differences between Down's Syndrome and Alzheimer's disease mouse model brains. ISMRM, 2015.

3. H. Holmes, N. Colgan, O. Ismail, **D. Ma**, J. Wells, N. Powell, J. O'Callaghan, I. Harrison, M. Cardoso, M. Modat, E. Fisher, S. Ourselin, M. O'Neill, E. Collins, M. Lythgoe: A multi-scale MRI approach to investigate novel drug treatment strategies in mouse models of Alzheimer's disease. ISMRM, 2015.

4. O. Ismail, H. Holmes, N. Colgan, D. Ma, J.A. Wells, N.M. Powell, J.M. O'Callaghan, I.F. Harrison, S. Walker-Samuel, J.M. Cardoso, M. Modat, E. Fisher, S. Ourselin, T.K. Murray, Z. Ahmed, M.J. O'Neill, R.A. Johnson, E.C. Collins, M.F. Lythgoe: Imaging the accumulation and suppression of tau pathology using multi-parametric MRI. Alzheimer's & Dementia 11.7 (2015): P106-P107.

5. J. Steventon, **D. Ma**, M.J. Cardoso, M. Modat, M.F. Lythgoe, S. Ourselin, R. Trueman, A.E. Rosser, D.K. Jones: A Longitudinal Study In Huntington's Disease Reveals Differential Macro- and Micro-structural Effects ISMRM, 2014.

6. **D. Ma**, M.J. Cardoso, M.F. Lythgoe, S. Ourselin: Cortical thickness map: an automatic quantification of cerebral cortex for in vivo mouse brain MRI. British Chapter ISMRM, 2013.

7. H. Holmes, J Wells, N. Colgan, J.M. O'Callaghan, S. Richardson, B. Siow,N. Powell, **D. Ma**, M. Modat, S. Ourselin, E. Fisher, M. F. Lythgoe: Tensor-Based Morphometry as a Sensitive Biomarker of Alzheimer's Disease Neuropathology in a Tau Transgenic Mouse ISMRM, 2013.

8. H. Holmes, N. Powell, J Wells, N. Colgan, J.M. O'Callaghan, **D. Ma**, M. Modat, M.J. Cardoso, S. Richardson, B.M. Siow, M.J. O'Neill, E.C. Collins, E. Fisher, S. Ourselin, M. F. Lythgoe: Morphometric Genomics: in vivo microMRI for 3D structural imaging of transgenic mice. ISMRM, 2013.

9. **D. Ma**, M.J. Cardoso, M. Modat, N. Powell, H. Holmes, M.F. Lythgoe, and S. Ourselin: Multi-atlas structural parcellation for in vivo quantification of mouse brain anatomy. British Chapter ISMRM, 2012.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Noninvasive medical imaging techniques, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT) allow us to "visualise" the inside of the body noninvasively by collecting signals that represent intrinsic physical, chemical or biological properties of its tissues, and reconstructing those signals into human or computer interpretable images. Depending on the source of the obtained image signal, the medical image can be categorised into structural and functional imaging. Specifically, modern neuroimaging techniques provide vast information about the nervous system (mainly the brain), such as the water (proton) density, tissue microstructure, nerve cell myelination, metabolism status, etc. Understanding this information can help us track the development, aging, pathological processes in the brain, as well as response to treatment.

The signal difference between neighbouring tissues or structures with different properties creates image contrast which can be used in image analysis for finding tissue/structural boundaries. The longitudinal change of signal intensities is an indication of underlying tissue property change with time in response to pathological, physiological or treatment effects. However, it is not a trivial task to interpret the complex information embedded in various types of neuroimages. Proper strategies have to be developed to analyse imaging data accurately and robustly, and extract the appropriate information of interest.

Among various imaging techniques, magnetic resonance imaging (MRI) is often the method of choice because of its soft tissue contrast. In order to detect structural variations among different pathophysiology, or longitudinal structural changes across

time, it is necessary to have automated and accurate quantitative morphometric analyses of the MR images. Various quantitative morphometric methods have been developed as surrogate biomarkers for image-based diagnosis, such as structural volume analysis, voxel-level morphometry [4], statistical shape analysis [5, 6], boundary shift integral [7, 8] and cortical thickness analysis [9].

1.1 Challenges of brain morphological analysis for preclinical neuroimaging

Genetically modified mice are widely used in preclinical studies of human diseases as they share more than 85% of their genes with humans [10]. As a result, mice are usually the animal model of choice in preclinical research for studying human diseases because of the similar biochemistry and physiology properties they shared with humans, limited variation between inbred individuals, as well as their fast growth and reproduction rate. Since the completion of mouse/human genome sequencing, research focus has been shifted from genotyping to phenotyping, which focuses on the characterisation of various physiological properties expressed by specific functional genomes. Huge efforts have been made involving large scale and collaborative works. Examples of these efforts are consortia such as International Mouse Phenotyping Consortium (IMPC[1] [11]), EuroPhenome[2] and EMPReSS[3] [12]. Neuroimaging is a standard technique in clinical environment for disease screening, diagnosis and prognosis, as well as for treatment monitoring, and has also become one important method for phenotyping studies.

Several difficulties arise when imaging small animals in preclinical studies that are not present when imaging humans. These include limited image resolution, signal-to-noise ratio (SNR) and image contrast which are due to much smaller brain size, as well as less control of image artefacts such as motion and flow [13, 14, 15]. With the development of high field MRI (7-16.4 Tesla compared to clinical MRI which is most commonly 1.5T and 3T) [16, 17], the advancing of coil and sequence design [18], improved anesthesia techniques [19] and contrast enhanced techniques [20, 21], small animal MR imaging using genetically modified mice has grown rapidly in recent years, and

[1] http://www.mousephenotype.org
[2] http://www.europhenome.org
[3] http://empress.har.mrc.ac.uk

has been widely adopted as an important tool to study the physiology of diseases' onset and progression, as well as to monitor recovery outcomes following potential treatment. This is especially the case for neurodegenerative diseases such as Alzheimer's disease, Huntington's disease and Down's syndrome. Fig 1.1 compares the *in vivo* T2 MRI of a human (acquired with a standard 3T clinical scanner) and a mouse brain (acquired with a 7T preclinical MRI scanner). There is good image contrast in both images to distinguish brain structures as well as to drive the image registration. The partial volume effect is more prominent for mouse brain due to the relatively limited image resolution. The problem is however partially reduced by the lack of convoluted cortical regions.

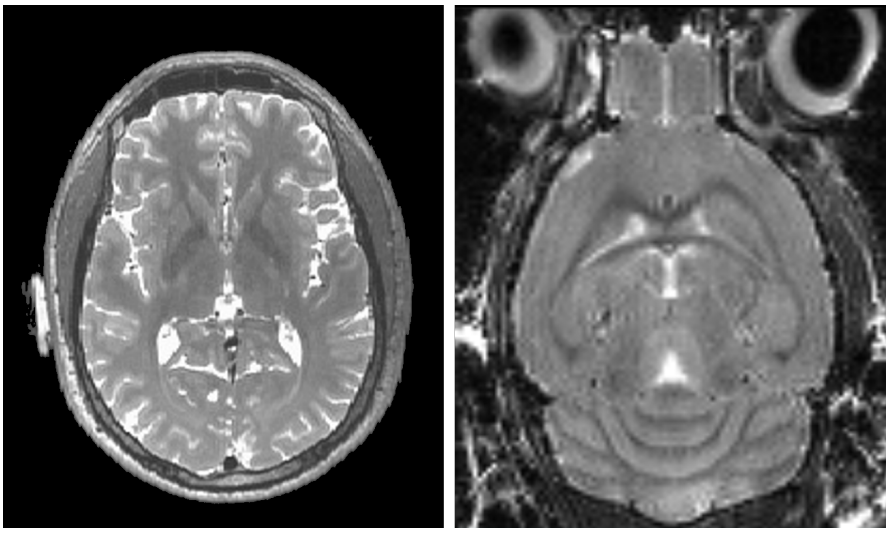

Figure 1.1: Comparison of an *in vivo* T2W of human brain from a 3T clinical scanner (Left) and an *in vivo* T2W mouse brain from a 7T preclinical scanner (Right). The image of human brain MRI is taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [22], and the mouse brain MRI is obtained from the study presented Chapter 4.

Despite the continuous efforts to improve the data acquisition, research on analysis for preclinical image studies is still relatively sparse compared to that for clinical studies. With the ever increasing amount of preclinical data, there is a strong and urgent need for proper tools that are tailored specifically for preclinical image data to achieve high throughput automated image interpretation and analysis for phenotyping studies [23].

1.2 Brain image structural parcellation

Structural parcellation, sometimes called structural segmentation, is an important initial step in the analysis of brain images, usually structural MR images (e.g. T1 weighted, T2 weighted or proton density images), and is typically required before further morphometric analysis (e.g. volumetric, shape and thickness). In the studies including multispectral MRI (e.g. Diffusion MRI and Function MRI) or multiple image modalities (e.g. CT, PET, SPECT), parcellation labels from the the structural MRI can be used to define the regions of interest (ROI) in other image types and confine the analysis of effects to relative structures.

Automated parcellation algorithms try to group voxels in brain images into several different anatomically or functionally relevant regions (Fig. 1.2). The most critical information to guide the parcellation is the image contrast, where neighbouring anatomical structures possess distinguishable signal intensities reflecting the underlying tissue property disparity (e.g. subcortical structures or hippocampus subregions). Other criteria have to be adopted in cases where no apparent image contrast is available between neighbouring structures. For example, the boundaries between distinct functional cortical subregions are determined from the underlying anatomical morphologies such as cortical gyri and sulci. In cases where there is no image feature to distinguish structures, the structural boundaries can be inferred from other image modalities in which boundaries are visible, e.g. histological slides stained by immunohistochemistry.

Generally, manual parcellation from expert human raters is considered to be the gold standard, and the goal of automatic structural parcellation is to achieve accuracy approaching that standard. Prior information such as anatomical locations (i.e. brain structural atlas) and intensity distribution (i.e. tissue prior maps), are necessary to train the algorithm properly. Various studies have been conducted to improve the structural parcellation accuracy with different approaches. Details of these approaches and current progress on mouse brain image structural parcellation will be reviewed in this thesis, and the application of the automated structural parcellation method to preclinical neuroimages will be explored.

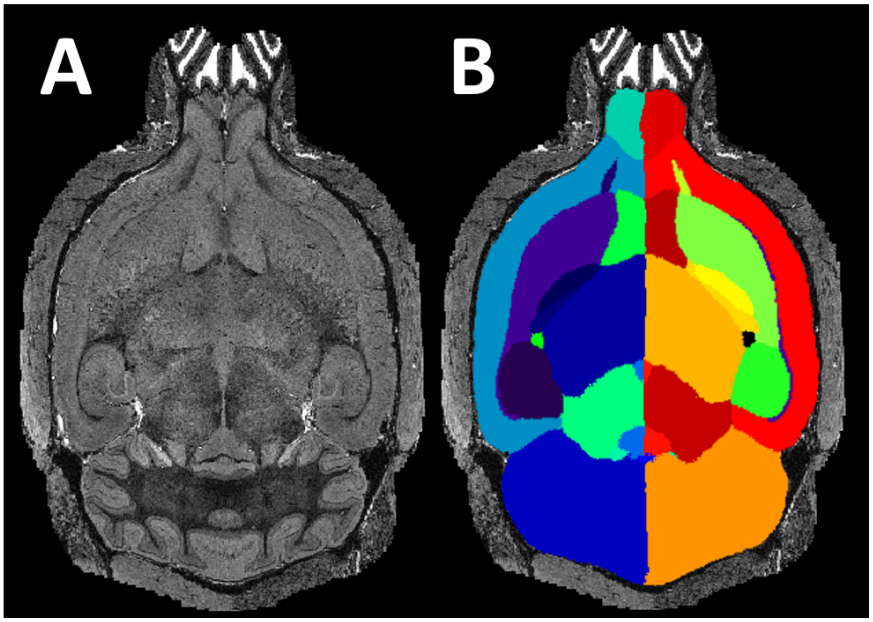

Figure 1.2: (A) An *en vivo* T2 mouse brain MRI, and (B) It's corresponding automatic structural parcellation result.

1.3 Comparing *in vivo* and *ex vivo* images

When designing preclinical neuroimaging experiments, a crucial choice to make is whether to image the animals alive (*in vivo*) or postmortem (*ex vivo*). Imaging the animal *in vivo* can ensure a better preservation of tissue properties and enables monitoring of longitudinal morphological changes; it does however, suffer from limited image resolution as well as motion and flow artefacts. On the other hand, the *ex vivo* imaging can guarantee higher image quality because of the prolonged motion-free acquisition time and much reduced physiological artefacts; but it suffers from tissue degradation and morphological changes after animal sacrifice as well as experimental procedures such as perfusion and fixation (Fig. 1.3). Currently, it is not entirely clear whether those advantages and drawbacks would affect the quantitative morphometric analysis of brain structures in preclinical neuroimaging. This issue will be explored in this thesis, specifically focusing on the effect on the structural volumetric quantification by adopting the framework I developed.

1.4 Cortical thickness measurement

Volumetric measurement is the most fundamental and commonly used imaging biomarker based on structural parcellation. Other quantitative biomarkers are also be-

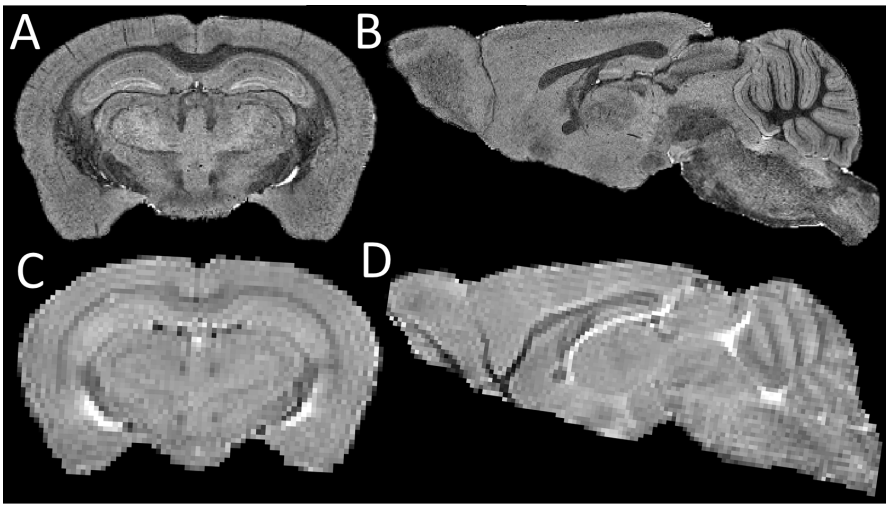

Figure 1.3: Top row: (A) coronal view and (B) sagittal view of an *ex vivo* T2 mouse brain MRI; Bottom row: (C) coronal view and (D) sagittal view of an *in vivo* T2 mouse brain MRI acquired from the same sample. The *ex vivo* images have better image quality compared to *in vivo* image, in terms of resolution, contrast and SNR. It however suffers from post-mortem structural change, such as the collapse of the entire ventricle (the high intensity regions). The parallel stripes appeared in the *in vivo* image is the truncation (Gibb's) artifact, which is due to the reconstruction from finite sampled signal in the k-space due to the limited matrix size during the acquisition.

ing investigated to study more specific morphological properties. Cortical thickness is one of the widely adopted measurements among them.

The cortex is a thin laminar structure of grey matter covering the outermost area of the brain, both in the cerebrum and cerebellum. Research into cortical cytoarchitecture has revealed that the cortex is a laminar structure [24, 25]. The neocortex, which is the major part and phylogenetically newest part of the cerebral cortex, consists of six cortical layers [25, 26, 2], while the cerebellar cortex consists of three cortical layers [27]. According to various studies of developmental and comparative neuroanatomy, the primary functional units of the cortex are the cortical microcolumns (or minicolumns) that run orthogonal through the cortical layers (Fig 1.4) [28, 25, 29, 30], which varies in function and morphology across different cortical areas due to different neuronal subtypes [31, 32, 33]. Based on this understanding of the cortical microcytoarchitecture, researchers have constructed mathematical models of cortical thickness from MRI [34, 35, 33, 36, 2, 37]. The morphological variation of cortical thickness, at both the global and local scale, has been shown to be associated with various neurodegenerative diseases [38, 39, 40, 41, 42].

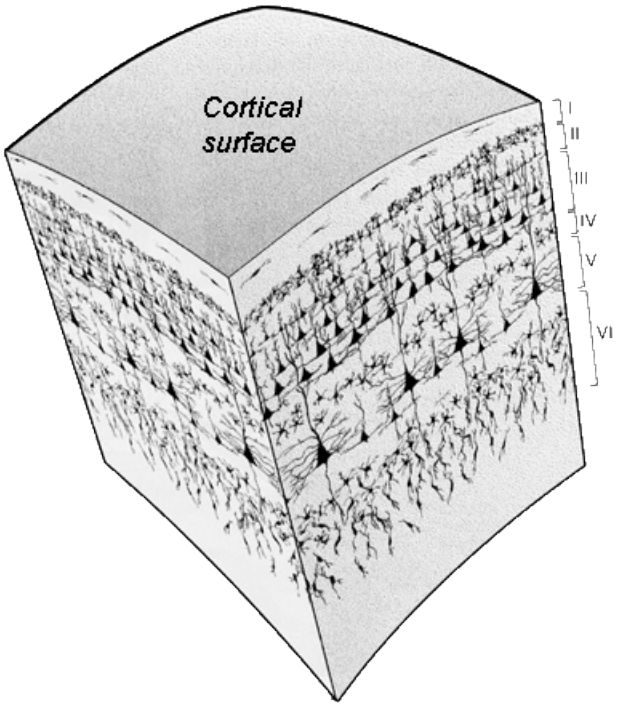

Figure 1.4: Schematic diagram of cortical column architecture, demonstrating the laminated morphology of the cortical sublayers, with microcolumns passing through the layers perpendicularly. Image taken from NeuWrite San Diego (http://www.neuwritesd.org)

The WM/GM ratio of human is much larger than that of the mouse, as the axons increase disproportionally faster than the neocortex during revolution [43], and the human cortex is more convoluted both in cerebrum and cerebellum. Meaningful cortical thickness measurement relies on accurate cortical extraction, so that the inner and outer boundaries of the cortex can be accurately defined from the surrounding tissues. However, the accuracy of the cortical surface extraction is hindered by the partial volume effect due to the cortical convolution, as well as the thin nature of cortex, (2-3mm for normal healthy human cortex vs. 1mm voxel resolution for standard *in vivo* MRI, and 200-400μm for normal wildtype mouse cortex vs. 150μm voxel resolution for standard *in vivo* MRI and 40μm voxel resolution for standard *ex vivo* MRI). A proper model of the cortical thickness, such as the cortical layers and microcolumn structures, is necessary to make an anatomically plausible measurement.

The human and mouse cortices have a distinctive morphology, both in cerebrum and cerebellum, as shown in Fig. 1.5. In human, current standard MRI is able to capture the morphology of sulci and gyri in the cerebral cortex, but not those in the cerebellar

cortex, which is more convoluted (Fig. 1.5A,B). Compared to that of the human, the mouse cerebral cortex is almost flat, which makes it relatively easier to extract and quantify. The mouse cerebellum is also less convoluted than human cerebellum, and can be well captured by *ex vivo* MRI (Fig. 1.5C,D). In addition, different cell layers in the mouse cerebellar cortex are also observable in the corresponding MRI, which makes it possible to investigate the cerebellar cortical sublayer quantitatively. It would be ideal to incorporate such image-contrast-based layer information into the cortical thickness model, and estimate the sublayer thickness of the mouse cerebellum.

1.5 Genetically modified animal models

To validate the translational frameworks I developed and presented in this thesis, mouse brain image data acquired from two inbred strains of transchromosomic/transgenic mice were used. The strains respectively model two neurodegenerative diseases: Down's Syndrome (DS) and Alzheimer's disease (AD). For the sake of the completeness, a description of these two mouse models is provided here.

Firstly, the Tc1 mouse model - a transchromosomic mouse strain that models Down's syndrome (DS) - was used in the validation of structural parcellation (in Chapter 3) as well as the cerebellar cortical sublayer thickness estimation (in Chapter 5). The Tc1 mouse line is a trans-species aneuploid model of mouse carrying a freely segregating copy of human chromosome 21 (Hsa21) and thus is functionally trisomic for 75% of Hsa21 genes, which has the clinical manifestation and similar phenotypes of DS in humans [44, 45].

The rTg(Tau_{P391L})4510 mouse model for human tauopathy was used to compare volumetric analysis using *in vivo* and *ex vivo* imaging (in Chapter 4). rTg4510 mouse overexpresses the tau protein (4R0N) due to the mutation (P301L) in exon 10. This abnormal processing of tau will lead to the neurofibrillary tangles (NFTs), which is associated with Alzheimer's Disease (AD) and other neurodegenerative diseases such as Parkinson's disease and frontotemporal dementia [46]. In the rTg4510 model, the transgene overexpression is induced through the Tetracycline-controlled transcriptional activation, a reversible gene transcription method in which the level of overexpression can be controlled through the treatment of doxycycline.

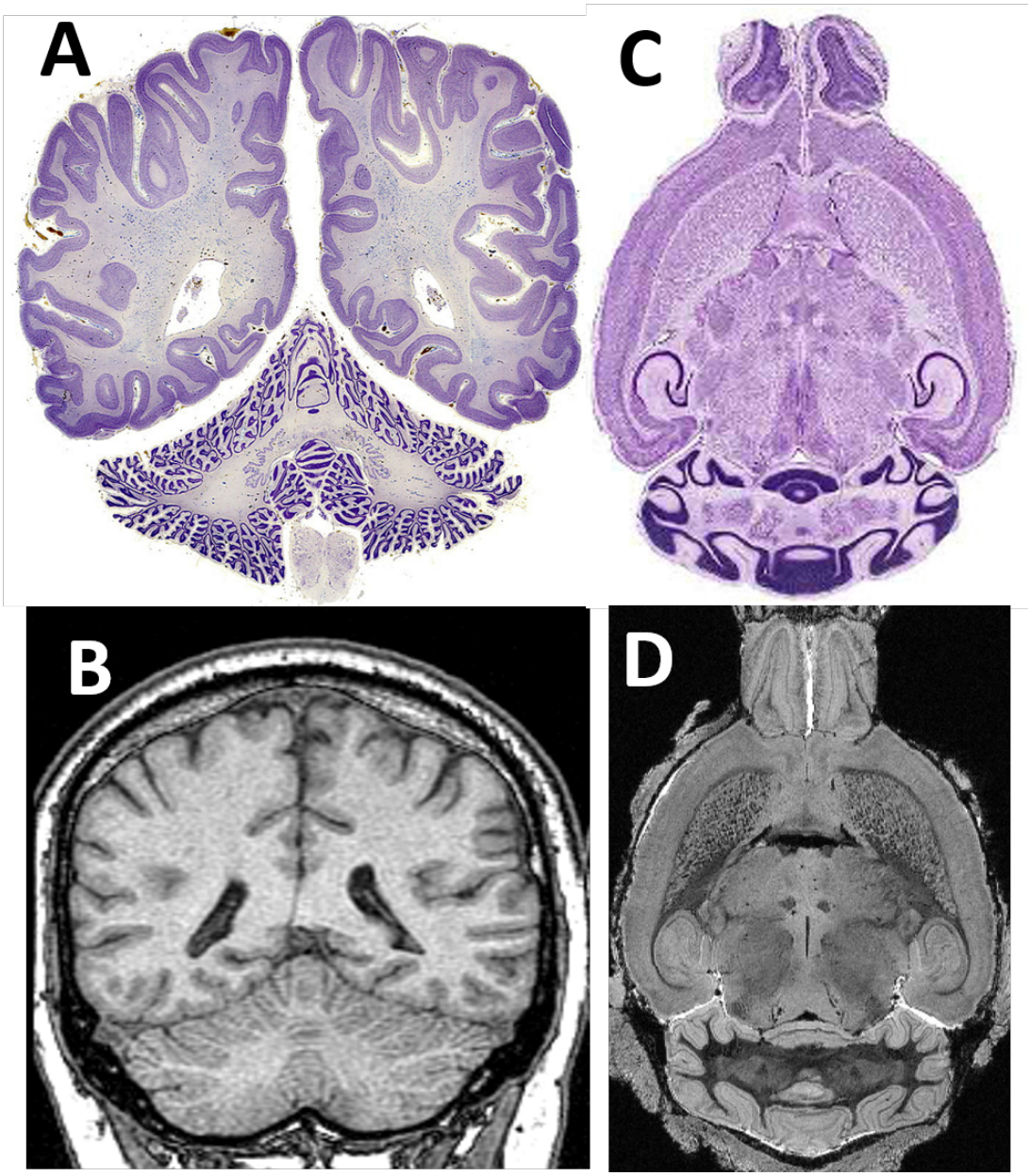

Figure 1.5: Comparison between human and mouse cortex, both in cerebrum and cerebellum. (A) A slide of the human brain with hematoxylin and eosin (H&E) stain at the axial view. (Image taken from the "The Human Brain Atlas" at: http://www.msu.edu) (B) An *in vivo* MRI of human brain at similar slice position. (Image taken from the "The whle brain atlas" at: http://www.med.harvard.edu/aanlib/) (C) A slide of the mouse brain with hematoxylin and eosin (H&E) stain at the sagittal view (Image taken from www.neura.edu.au) (D) The corresponding sagittal view in *ex vivo* MRI of mouse brain.

1.6 Thesis contributions

The aim of this thesis is to develop automated quantitative structural analysis frameworks for mouse brain MRI, which can be used to study the pathology of neurodegen-

erative diseases and the effect of potential treatments. Specifically, this thesis focuses on the construction and validation of two quantitative methods - brain structural parcellation and cerebellar cortical sublayer thickness estimation.

- Firstly, a multi-atlas-based structural parcellation framework is constructed. The framework incorporates preprocessing steps such as intensity non-uniformity correction and multi-atlas-based brain extraction, followed by label propagation based on non-rigid registration and locally weighted multi-atlas label fusion. The parameters in the label fusion step are optimised through leave-one-out cross validation, and the parcellation framework is evaluated on both *in vivo* and *ex vivo* MRI data. This is the first multi-atlas-based brain parcellation framework applied on both *in vivo* and *ex vivo* mouse brain MRI. The proposed framework outperforms single-atlas-based structural parcellation, as well as global-weighted multi-atlas label fusion algorithm. When applied to the *ex vivo* images of a Down's Syndrome mouse model, the proposed framework also successfully detects phenotypes in terms of structural volumetric difference between groups with different genetic backgrounds.

- To provide some quantitative guidance for the experimental design of small animal neuroimaging studies, I applied the structural parcellation framework to both the *in vivo* and *ex vivo* data acquired from the same wildtype cohort, and compared the volumetric analysis result of the two measurements. The results showed a non-uniform distribution of brain volume shrinkage from *in vivo* to *ex vivo* data. When determining the effect of a treatment compound, both *in vivo* and *ex vivo* images demonstrated similar statistical analysis result using the parcellated brain structural volumes as a quantitative biomarker.

- Based upon the structural parcellation, a multi-layer thickness estimation framework is developed for morphometric characterisation of mouse cerebellar cortical sublayers from the high-resolution, contrast-enhanced *ex vivo* MRI. A layer model which follows the laminar features of the highly convoluted cortex is adopted which, when combined with an anisotropic contrast enhancement filter, segments and estimates the thickness of the cerebellar sublayers. Evaluation of the framework on mouse models of Down's syndrome demonstrated its ability to

detect regional cortical sublayer thickness variations, which cannot be revealed using full cortical thickness measurements alone.

1.7 Thesis organisation

This thesis introduces the development and translational applications of two quantitative medical image analysis frameworks I developed in my PhD: the automated structural parcellation for volumetric analysis, as well as thickness estimation for cerebellar cortex.

Chapter 2 presents a literature review about the current state-of-the-art in automated structural parcellation and cortical thickness measurements, both in clinical and preclinical scenarios. I also review the current research on the relation between the structural analysis of *in vivo* and *ex vivo* images.

Chapter 3 presents a multi-atlas structural parcellation framework for mouse brain MRI. The framework is evaluated on various datasets, both *in vivo* and *ex vivo* data. This chapter also demonstrated the successfully application of the framework in different studies.

In **Chapter 4**, the structural parcellation framework is applied to compare the volumetric analysis on images acquired either *in vivo* or *ex vivo* from the same cohort. The evaluations include both direct comparison of volumetric analysis and the application of the framework to discriminate between two animal groups.

In **Chapter 5**, a framework of cerebellar subcortical layer segmentation and thickness estimation is introduced. This incorporate both intrinsic image contrast and a mathematical model of the cortical lamination. The framework is validated on the *ex vivo* mouse brain MRI.

Finally, an overall discussion is presented in **Chapter 6**, and the thesis is concluded in **Chapter 7**, in which the plan for the future direction of my work is also discussed.

Chapter 2

State-of-the-art

2.1 Automatic structural parcellation for mouse brain

Structural parcellation is the process of extracting the labels of brain structures from images like MRI, CT, ultrasound, etc. In medical images, different anatomical structures might share similar tissue intensity properties, and the same structure may contain several different tissue classes (Fig. 2.1) [1]. As a result, it is not possible to parcellate brain structures automatically based solely on the tissue intensity distribution as in brain tissue segmentation - a process in which the brains are segmented into grey matter (GM), white matter (WM) and cerebralspinal fluid (CSF). Manual labeling by experts is still the current gold standard of structural parcellation in MRI studies of mouse brains, despite being expert-dependent and labour intensive [3, 47].

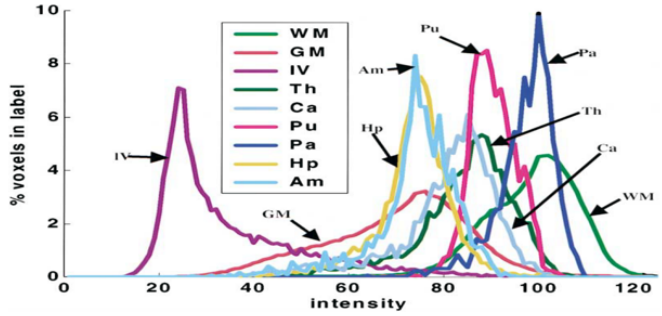

Figure 2.1: Overlapping brain structural tissue intensity histograms of MRI (human brain) make it difficult to distinguish different tissues using only the intensity distribution profile. Figure taken from [1]

.

2.1.1 Atlas-based label propagation

There are two types of structural information that can be derived from MRI images: shape- (or edge-) based labels and volume- (or voxel-) based labels. The shape-based labels define structures with a contour or mesh based labels, sometimes along with various visual appearance feature statistics. Several methods have been developed to extract the shape information of one or several structures, such as hippocampus and cerebellum, for rodent brain through active shape model [48], active appearance model [49, 50] and active volume model [51]. This type of automated structural extraction technique shows promising result to segment some specific structures. Its performance is however relying on the image gradient information which is sensitive to noise and spurious edges. As a result, using the shape-based method along is inadequate to simultaneously parcellate large number of structures with variable size and shapes as well as structures with complex intensity distribution profile.

Another type of structural information is voxel-based structural labels, where all the voxels in the same structure are assigned the same label. The manually delineated structural labels for existing MR images can be used as prior information to automatically parcellate new images, in a process called atlas-based label propagation. Here an atlas is defined as a pair of images containing both the original MR data (either from a single sample or a groupwise average) and its corresponding manually labelled anatomical structures. The atlas based label propagation method has shown parcellation accuracy when compared to manual labeling for both clinical [7, 52] and preclinical images [53, 54]. There are increasing number of brain atlas databases developed and released for human brain [55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 66, 67, 68, 69, 70] and for mouse brain [71, 72, 73, 74, 75, 76, 77, 78, 3, 79, 47, 80, 81, 82]. A typical atlas-based label propagation consists of two steps: image registration and label propagation (Fig. 2.2).

Image registration Atlas-based label propagation relies on image registration, which is one of the fundamental techniques in medical image processing (Fig. 2.2). In image registration, a transformation matrix T which deforms a floating image I_F to the space of a reference image I_R is optimised to minimise the image differences (or maximise image similarities depending on the objective function adopted) between the deformed floating image $T(I_F)$ and target image I_T (Fig. 2.2 ①).

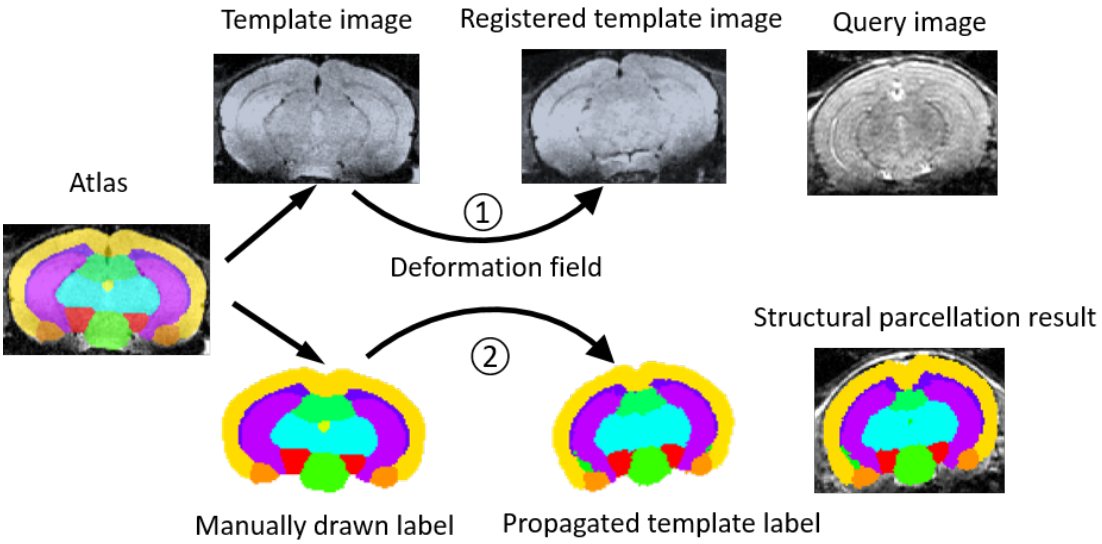

Figure 2.2: Schematic diagram of atlas-based label propagation. The template image from the atlas is firstly registered to the enquiry image. The resulted deformation matrix is then applied to the atlas labels to derive the structural parcellation result.

When performing the image transformation T, the deformed floating images $T(I_F)$ are resampled into the space of the reference image I_R: $\forall \vec{x} \in T(I_F)$. The intensity of $T(I_F)$ is calculated at its corresponding coordinates (determined from T) in the space of I_F through intensity interpolation (e.g. linear, spline, sinc, etc).

Current state-of-the-art image registration framework normally involves two steps: the global-transformation-based step followed by a local-transformation-based step [83]. In the global registration step, the same image transformation T is applied to each pixel (in the case of 2D) or voxel (in the case of 3D) in an image, by either rigid transformation (including translation and rotation) or affine transformation (adding shearing and scaling to a rigid transformation). In the local (also called elastic or non-rigid) registration step, a deformation field which is capable of performing local wrapping is generated to account for the unevenly distributed local image dissimilarities through either parametric or non-parametric deformation models [84]. Regularisation or penalty terms are used to ensure a smooth deformation and anatomical plausibility.

Label propagation After the deformation field is determined through image registration, the same transformation is then applied to the manually labelled anatomical structures in order to match the unlabelled image's morphology (Fig.2.2 ②). To preserve the integral nature of the resulted label, the resulted labels are resampled through nearest neighbour interpolation, in which the label values were determined from its corre-

sponding coordinates (determined from T) in the space of atlas label L_F similar to the interpolation procedure in the registration steps, but are assigned the label value of the voxels in L_F that have the shortest distance to the transformed positions.

As the performance of structural parcellation is directly related to the image registration accuracy, this single-atlas label propagation technique is also referred to as registration-based structural parcellation. Lee *et al.* [53] evaluated two different non-rigid registration algorithms to achieve label propagation on mouse brain, the fluid-model [85] and free-form B-spline[86], and found a similar performance both in terms of accuracy and stability. However, although image registration algorithms are constantly advancing, it is by itself an ill posed problem [87]. Compared to human patients, the subject variations within same mouse strains is relatively small. However, local misalignment will still occur due to the large morphological variability across strains, imaging artefacts, low signal-to-noise-ratio (SNR) or contrast-to-noise ratio (CNR), resulting in poor parcellation accuracy.

2.1.2 Probabilistic-atlas-based structural parcellation

To account for the morphological variation within populations, probability-based atlas priors have been proposed instead of discretised ones, where the probability represents the occurrence of different labels for each voxel.

Scheenstra *et al.* [88] parcellated the structures from mouse brain MRI through an edge-based Bayesian clustering after single-atlas label propagation using only affine registration. The algorithm uses prior information derived from the atlas regarding the intensity distribution within each label class to guide the edge-based Markov Random Field (MRF) clustering on the query image. Evaluation of the algorithm in both *in vivo* and *ex vivo* mouse brain images showed comparable performance to parcellation methods based on non-linear registration [89]. This edge-refinement methods reduced the required computational time significantly compared to non-linear-registration-based parcellation, while maintaining comparable parcellation accuracy. However, this method requires a reasonable initialisation from affine registration, and therefore cannot account for large local morphological variations between the atlas and query images. The statistical prior information also relies on a similar intensity distribution between the atlas image and the query image, which is not always the case.

Ali *et al.* [90] developed an automatic parcellation framework using multi-spectral MR microscopy (including T2, proton density and diffusion weighted images) for the mouse brain, which is a follow up to the previous study on human brain MRI by Fischl *et al.* [1]. The probabilistic atlas database is constructed from 6 intensity-normalised brain images affinely registered to a common space, each with 21 manually labelled structures. The intensity information from the multispectral MR images forms a multi-dimensional feature space, and is modelled as classes of multivariate Gaussian distribution. The contextual information is modeled under a spatially variant first-order MRF model. The final labels are obtained through maximum a posteriori (MAP) estimation. This framework has been applied to mouse brain MRI with image contrast enhanced by active staining technique [91] to characterise the intra-strain [92], and inter-strain [93] structural volumetric variation. However, the multispectral imaging protocols increase the scanning time, making them unfeasible for use in all studies.

Bae *et al.* [94, 95] and Wu *et al.* [96] extended this model by adopting a support vector machine (SVM) classifier (which has been extended to multiclass classification) to model the posterior label probabilistic function instead of using the Gaussian probability distribution. This extended MRF outperforms the original methods proposed by Ali *et al.* [90] when evaluated on the same multispectral data, but requires a much longer training time. (For a training dataset of 5 mouse brains manually segmented into 20 structures, it tooks 290 to 364 minutes to train the algorithm excluding the registration time, and 75 minutes to segment one brain, on a 3.4-GHz PC.) This however is not a big problem if the training set is fixed, as the training procedure only need to be performed once offline.

Compared to the registration-based single-atlas label propagation, the probabilistic-atlas-based structural parcellation demonstrated improved accuracy, especially with the help of MRF to consider contextual voxel information. However, this probabilistic approach still can not fully capture large local morphological variations across subjects, and the improvement over small-structure parcellation is still limited.

2.1.3 Multi-atlas label propagation and fusion

With the availability of the database which contains more than a single atlas, another approach has become increasingly popular recently which improves the structural parcel-

lation accuracy by combining information coming from multiple atlases in a database into a consensus result - a method normally called "label fusion" (Fig. 2.3) [97, 98, 99]. This multi-atlas label fusion approach has gained increasing attention in recent years due to its promising performance, and has multiple applications in both clinical and preclinical neuroimaging [100].

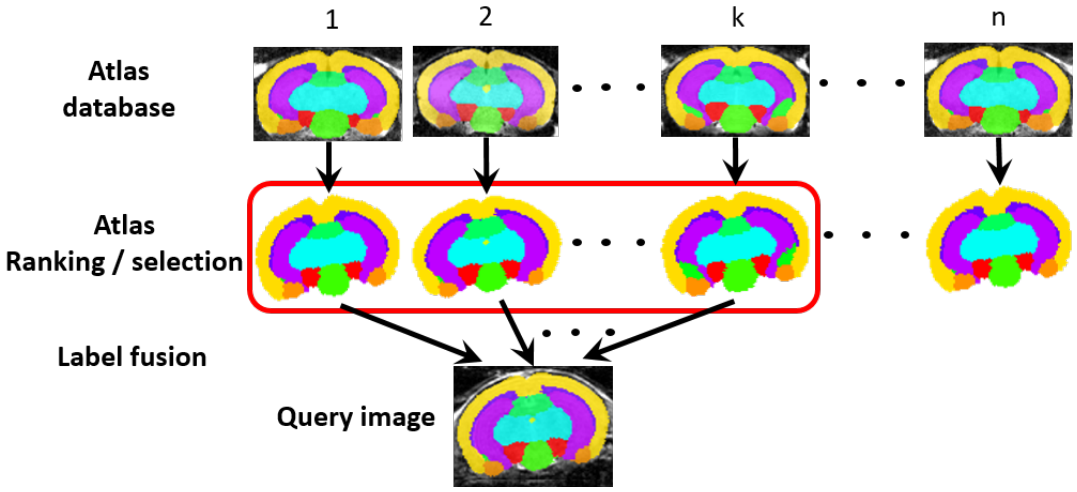

Figure 2.3: Schematic diagram of multi-atlas structural parcellation. The labels of each atlas in the database are propagated to the enquiry image through image registration and ranked. A subset of top-ranked atlas candidates are selected and combined through label fusion to obtain the final parcellation result.

In a multi-atlas structural parcellation, the first step is to perform image registration and label propagation for each individual atlas in the database as in the single-atlas-based method. The propagated structural label candidates are used as a training set to obtain the best final structural label for the query image (rather than treat the candidates equally as in the probabilistic atlas). Fig. 2.3 shows the basic workflow of a label fusion algorithm. The performance or accuracy of the candidate labels are first estimated or ranked (either globally or locally). A subset of top ranked labels are then selected and combined to estimate the underlying true structural labels.

2.1.3.1 Majority voting

An intuitive way to fuse the labels is through majority voting, in which each voxel is assigned to the label that receives the largest number of votes from the propagated candidate atlases [101, 102]. Suppose there are N atlases in the database, and we denote S_n as the structural parcellation candidate for each registered atlas template image, where $n \in 1...N$. For $x \in \Omega$ where Ω is the computational domain in the query image

space, the final label $S(x)$ can be obtained from:

$$S(x) = \underset{l \in \{1,\ldots,L\}}{\arg\max} \sum_{n=1}^{N} \frac{1}{N} p(l|S_n, x) \,, \tag{2.1}$$

where L is the total number of labels, $p(l|S_n,x)$ is the probability of parcellation candidate S_n vote for label l at voxel x, which is 1 if $S_n(x) = l$ and 0 otherwise, and $\sum_{l=1}^{L} p(l|S_n,x) = 1 \; \forall n \in 1,\ldots,N$.

Given the fact that the propagated parcellation of some atlas templates performs better than others, studies have been focused on ranking the atlases in order to assign different weights to them during the label fusion process, or only select a small subset of atlases with higher rank for label fusion, to improve the parcellation accuracy. The updated voting equation becomes:

$$S(x) = \underset{l \in \{1,\ldots,L\}}{\arg\max} \sum_{n=1}^{N} w_n(x) p(l|S_n, x) \,, \tag{2.2}$$

in which the $w_n(x)$ is the assigned weight to the n^{th} atlas, where $\sum_{i=1}^{N} w_n(x) = 1$.

The weight w can be simple metadata [103], such as age or sex [104], or can be the image similarity between transformed atlas and the test image using methods such as Summed Square Distance (SSD) [105], Mutual Information (MI) [106], Normalized Mutual Information (NMI), or transforming distance [107, 108]. Wu *et al.* [104] attempted to find the single most appropriate atlas within the database through the NMI image similarity measurement, and demonstrated improved parcellation results. Wang *et al.* [102] managed to optimise the weight $w_n(x)$ by minimising the total error calculated as the correlation between each atlas image and the query image in a covariance matrix M_x and solved through a Lagrange multiplier.

2.1.3.2 Atlas selection

Besides weighting, studies have managed to select limited atlases from the database based on different ranking scheme and demonstrated improved parcellation accuracy. Aljabar *et al.* [109] selected the atlas subset randomly without ranking, and showed that the accuracy of the majority voting label fusion depends on the number of atlases selected, but reaches the plateau when the number of atlas selected reaches a certain number. Wu *et al.* [104] selected the single best template (determined by measuring

the NMI between the registered atlas image and the query image) within the database and achieved a parcellation accuracy better than simply averaging all the parcellation. Aljabar *et al.* [103] later demonstrated that the label fusion accuracy can be significantly improved by using a subset of best ranked candidate labels, with the atlases ranked by the global normalised cross correlation (GNCC) between the registered template and the query image over a specific region of interest. Sabuncu *et al.* [105] extended the study by Aljabar *et al.* [109] and concluded that, when selecting the top ranked atlases instead of selecting them randomly, the parcellation accuracy is less dependent on the number of atlas selected. Shen *et al.* [108] suggested another atlas ranking scheme using the Least Angle Regression (LAR), a variable selection method in regression, and demonstrated improvement in both parcellation accuracy as well as robustness when compared to the other ranking scheme using image-similarity-based measurement including NMI, cross correlation (CC), and Mean Square Difference (MSD).

2.1.3.3 Selective and Iterative Method for Performance Level Estimation (SIMPLE)

Compared to image similarity, a better way to compute the weight for the parcellation candidates S_i would be estimating the performance of each candidate by calculating its agreement to the underlying ground truth parcellation (e.g. Dice similarity coefficient - DSC, Jacard coefficient, or sensitivity and specificity). Although this performance level is not readily available due to the lack of such ground truth, it can be approximated as the current estimation of the fused label and improved iteratively.

Langerak *et al.* [110] devised an algorithm - Selective and Iterative Method for Performance Level Estimation (SIMPLE) - which combines the advantage of both atlas selection and performance estimation in an iterative manner and showed further improvement in parcellation accuracy. In the SIMPLE algorithm, the initial true parcellation estimation S_T^0 is obtained from weighted majority voting, with the weight determined from NMI. At each iteration j, the performance of each parcellate candidate S_i is calculated as the binary label overlap measure $E(S_i, S_T^j)$. A subset of S_i with performance above a threshold $E(S_i, S_T^j) > \theta^j$ are selected to estimate the updated S_T^{j+1} using weighted majority voting weighted by $E(S_i, S_T^j)$ until S_T converges. The threshold at iteration j is determined as: $\theta^j = \mu^j - \alpha \times \sigma^j$, where μ^j and σ^j are the mean

and standard deviation of the performance of all the parcellation candidates, and α is a scalar variable which controls the number of candidates discarded in each iteration.

Leave-one-out validation on the SIMPLE algorithm demonstrated promising performance, which is not sensitive to the choice of either the performance estimator or the initial estimation of the true parcellation S_T^0. However, the fact that the underlying true parcellation is determined based on weighted majority voting from the current available parcellation candidates is not perfect, which will incur bias when calculating the subsequent parcellation performance.

2.1.3.4 Simultaneous Truth and Performance Level Estimation (STAPLE)

Warfield *et al.* [98] developed an algorithm - simultaneous truth and performance level estimation (STAPLE) - which fuse a set of structural parcellation candidates of a test image (from either expert labelling or automatic label propagation algorithm) by computing the probabilistic estimate of the underlying ground truth parcellation S_T (rather than the actual S_T as implemented in the SIMPLE algorithm [110]). The algorithm treats the S_T as missing/hidden data in a complete dataset which includes both the parcellation candidate and the true parcellation: $(S_{(1,...,N)}, S_T)$, and optimises the estimation of S_T through an iterative expectation maximization (EM) framework. The EM framework is initialised by assigning equal weights to each parcellation candidate (label assignment), and iteratively updates the sensitivity p (proportion of true positive) and specificity q (proportion of true negative) to obtain the optimal weights to derive the combined parcellation. The performance level here is defined as the p and q of propagated parcellations for each individual atlases. The parameters p and q are defined as the conditional probability:

$$\begin{cases} p_n^l = P(S_n = l | S_T = l) \text{ (true positive fraction)} \\ q_n^l = P(S_n \neq l | S_T \neq l) \text{ (true negative fraction)} \end{cases} \quad \forall n \in \{1,...,N\} . \tag{2.3}$$

In order to estimate the optimised S_T, it is necessary to maximise the complete dataset $(S_{(1,...,N)}, S_T)$ from the parameter set (p,q), which can be transformed as to optimise

the cost function of the log likelihood of $(P(S_{(1,\ldots,n)}, S_T|p,q)$:

$$(\hat{p},\hat{q}) = \arg\max_{p,q} log(P(S_{(1,\ldots,n)}, S_T|p,q)) \ . \tag{2.4}$$

With the assumption that the candidate parcellation are independent to each other (thus $S_i \perp S_j$, $p_i \perp p_j$ and $q_i \perp q_j$, $\forall i \neq j$), Eq. 2.4 can be optimised in an Expectation-Maximisation fashion. The posterior probability of a voxel x belonging to label $l \in \{1,\ldots,L\}$ is denoted as $W_l(x)$ (Expectation step), and is estimated as:

$$W_l^{i+1} \equiv P(S_T = l|S_{\{1,\ldots,N\}}, p^i, q^i) = \frac{\alpha}{\alpha+\beta} \ , \tag{2.5}$$

where

$$\alpha = P(S_T = l)\prod_n P(S_n|S_T = l, p_n, q_n) \ , \tag{2.6}$$

$$\beta = P(S_T \neq l)\prod_n P(S_n|S_T \neq l, p_n, q_n) \ . \tag{2.7}$$

And the performance level parameter p, q that maximise the log likelihood of the estimated underlying ground truth for each label l and each parcellation candidate n are updated in each iteration (Maximisation step):

$$p_n^{i+1} = \frac{\sum_{S=l} W_l^i}{\sum_L W_l^i} \ , \tag{2.8}$$

$$q_n^{i+1} = \frac{\sum_{S\neq l}(1-W_l^i)}{\sum_L (1-W_l^i)} \ . \tag{2.9}$$

This expectation-maximisation framework is solved recursively following Bayes' rule until the final true parcellation S_T converges.

Rohlfing *et al.* [97] generalised this algorithm to multiple labels fusion by replacing the parameters p and q in the original STAPLE algorithm by a "confusion matrix" M_n, an idea introduced by Xu *et al.* [111], in which each entry $m_{n,i,j}$ in M_n represent the joint-occurrence of candidate parcellation label decision i and the corresponding true label j:

$$m_{n,i,j} = \#\{x|S_n = i, S_T = j\} \ , \tag{2.10}$$

The sensitivity p_n is equivalent to the corresponding row normalised cross-parcellation

coefficient λ_n, and can be computed as:

$$p_n \equiv \lambda_n = P(S_n(x) = l_n | S_T(x) = l_n) = \frac{m_{n,i,j}}{\sum_j m_{n,i,j}} , \tag{2.11}$$

while the specificity q_n are spread over the off-diagonal element of M_n and can be expressed as:

$$q_n \equiv 1 - \frac{\sum_{i' \neq i} m_{n,i',i}}{\sum_{i' \neq i} \sum_j m_{i',j}} . \tag{2.12}$$

Validation of the STAPLE algorithm showed significant improvement of parcellation accuracy over voting strategies when applied on digital/brain phantom data [98] and real data acquired and manually segmented from 3D confocal microscopy images of the honeybee brain [112]. Leung *et al.* [113] applied STAPLE algorithm to brain extraction of human brain MRI and demonstrated improvement compared to other frameworks. Leung *et al.* [114] also demonstrated successful application of STAPLE algorithm for hippocampal segmentation, which achieved good distinction between control, mild cognitive impaired patients (MCI) and Alzheimer's disease patients (AD), as well as the ability to predict disease progression for MCI patients.

2.1.3.5 Locally-weighted label fusion strategies

The parcellation candidates for label fusion come from the registration-based label propagation. The parcellation errors of individual candidates may originate from the local morphological variations, locally inaccurate registration due to noise or regularisation, or resampling error due to partial volume effect in images with limited resolution. In cases where the registration algorithm performs well in some local regions but poorly in some other regions (as shown in 2.4), a global ranking/weighting is no longer fully representative of the local parcellation quality and sometimes misleading, and may affect the final parcellation accuracy. This effect is more apparent when the number of atlas candidates is limited. Methods have since been proposed to weight and fuse the atlas locally, thus improving the accuracy of the parcellation compared to globally weighted voting strategies, especially in areas with bad tissue contrast (Fig. 2.4) [115, 105, 102].

Wang *et al.* [102] introduced a locally-weighted majority voting method and managed to reduce the expectation of the combined error correlation between parcellation candidate by optimising the local weight with a neighbourhood specified by a radius

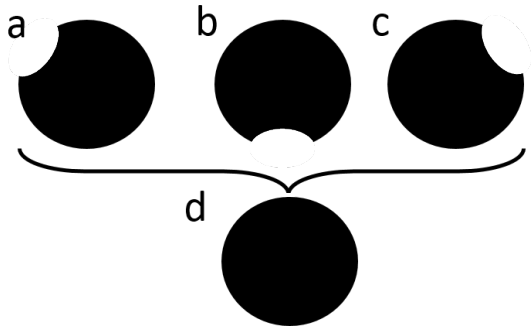

Figure 2.4: Schematic example to shows the parcellation candidate from single-atlas label propagation (a-c) each with local inaccurate parcellation, but can be fused together with a more accurate consensus final parcellation (d) by locally-weighted label fusion strategy.

r through covariance matrix. Agarwal *et al.* [116] proposed a locally-weighted and multi-label SIMPLE method for Chest CT data by dividing the images into overlapping local cubes and deriving different weight to different local regions, which showed improvements over the original SIMPLE algorithm with global strategy. Artaechevarria *et al.* [115] further conducted an experiment to compare the parcellation accuracy by using various global and local fusion algorithms - including majority voting, global weighted voting based on normalised cross correlation (NCC), MI or MSD, locally-weighted voting based on NCC, MI or MSD, as well as STAPLE - and concluded the best label fusion strategy to be the locally-weighted voting based on MSD.

Finally, Cardoso *et al.* [117] recently proposed an algorithm - multi-label Similarity and Truth Estimation for Propagated Segmentations (STEPS) - which incorporates a local ranking strategy for template selection based on a locally Normalised Cross Correlation (LNCC) to the original STAPLE algorithm for atlas selection. STEPS algorithm introduces a new image similarity based model variable to represent the observed cluster assignment $O(x)$ which is equal to 1 if the registered atlas image $T(I_{F_n})$ is within the selected top ranked image at voxel x and equal to 0 if otherwise. The original STAPLE framework (Equation 2.4) is thus transformed to:

$$(\hat{p},\hat{q}) = \arg\max_{p,q} log(P(S_{(1,\ldots,n)}, S_T, O)|p,q))) \; . \tag{2.13}$$

And the expectation maximisation optimisation scheme in the original STAPLE framework (Equation 2.5) now becomes restricted maximum likelihood (REML) which fo-

cuses only on the likelihood of a subset of the data:

$$W_l \equiv P(S_T = l | S_{\{1,\ldots,N\}}, O = 1, p^i, q^i) = \frac{\alpha}{\alpha + \beta} \,. \tag{2.14}$$

In addition, the STEPS algorithm also integrated the MRF regularisation into the optimisation scheme that Cardoso *et al.* previously developed [118] which is updated iteratively with a mean field approximation on the probabilistic labels. This is in contrast to the MRF regularisation presented in the original STAPLE algorithm, which was implemented as a post processing step. Similarly to Rohlfing *et al.* [97], Cardoso *et al.* also extended this "local ranked STAPLE" algorithm to multiple label fusion through a confusion matrix M_n (Equation 2.10) and its corresponding row normalised equivalent λ_n (Equition 2.11). Performance validation of the STEPS algorithm on human brain MRI showed significant improvement in parcellation accuracy compared with the ranking using global cross correlation (GCC), and is also less dependent on the size of the atlas database and the number of selected templates [117].

2.1.3.6 Learning-based mislabel correction and joint label fusion

The structural parcellation errors consist of random errors and consistent errors [119, 103, 102]. The random errors normally come from image noise and anatomical variation, and can be reduced through the label fusion methods reviewed above. The consistent error is a systematical bias when transferring manual segmentation into automatic segmentation which however, cannot be corrected by the the conventional label fusion methods. To address this problem, Wang *et al.* [102] developed a learning-based wrapper to detect the mislabeled voxels which are affected by the consistent error between the automatic and manual segmentation, and replace that with corrected labels classifier. The classifiers combines the information from of image feature (normalised intensity), contextual feature (label of the neighbouring voxels) as well as coordinate feature (spacial information). Application of this learning-based wrapper on the local-weighted multi-atlas label fusion methods [115, 105] demonstrated a reduction of consistent segmentation bias and lower mislabeled voxels.

In addition, within a multi-atlas database, the segmentations from some atlas candidates $\delta^i(x)$ are more similar than the rest in the database. Wang *et al.* [120] further proposed a method, the joint label fusion, to take such atlas similarity into account

when combining labels. Instead of estimating the weight for each atlas independently, the joint label fusion algorithm also calculate the correlation between each atlas pairs, where the pairwise dependency is defined as the joint probability of segmentation error. The total expected error E between the the consensus segmentation $\overline{S}(x)$ and the true segmentation $S_T(x)$ is given by $E = w_x^T M_x w_x$ where $w_x = [w_1(x); ...; w_n(x)]$ are the estimated weights, and M_x is a pairwise dependency matrix estimating the likelihood that two atlases both produce wrong segmentation: $M_x(i, j) = p(\delta^i(x)\delta^j(x) = 1)$. The voting weights are estimated to minimise the total expectation error, i.e.

$$w_x^* = \arg\min w_x^t M_x w_x \text{ subject to } \sum_{i=1}^{n} w_i(i) = 1. \tag{2.15}$$

Validation of the joint label fusion algorithm on hippocampus segmentation and hippocampal subfields segmentation showed its ability to eliminate the effect of the redundant information from similar atlases when assigning the weight.

The combination of the learning-based mislabel correction and the joint label fusion method has demonstrated great segmentation accuracy [121], and the final implementation has won the first place of the MICCAI 2012 Multi-Atlas Labeling Challenge, and is also among the top performers in MICCAI 2013 segmentation: Algorithms, Theory and Applications (SATA) challenge.

2.1.3.7 Progressive label propagation

Several studies also explored the intrinsic variabilities within the atlas database or among the query images prior to the label fusion process. The inaccurate parcellation due to intrinsic morphological scattering among the images are resolved by propagating labels progressively through similar image pairs (Fig. 2.5). This learning procedure might require extra running time, but only one-time offline learning is necessary for each dataset.

Langerak *et al.* [122] introduces an atlas selection step prior to the step of registration between atlas images and query images, effectively saving large amount of time for image registration with little or no effect on the parcellation accuracy. This atlas preselection is achieved by pairwise pre-registration of all the images in the atlas database, and clustering the atlases as graph nodes through the "affinity propagation" [123]. The weight between the nodes of the graph is assigned based on the DSC, which is defined

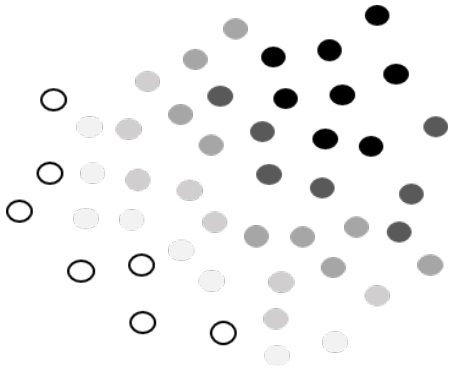

Figure 2.5: Schematic graph demonstrating the multistep progressive label propagation. The distance between the dots representing the relative similarity. Images which are more similar or close (represented in light dots) to the atlas images (represented in hollow dots) are parcellated first, which are then included to propagate the labels and parcellated the images less similar to the atlas images but are relatively closer to the current parcellated ones (represented in darker dots).

as the percent voxel overlap between two labels [124].

$$DSC = \frac{2|V_{manual} \cap V_{auto}|}{|V_{manual}| \cup |V_{auto}|} \tag{2.16}$$

. Rather than performing pairwise registration from atlas dataset to the query dataset and simply discarding dissimilar images or assigning lower weight, Wolz *et al.* [125] embedded a manifold learning on the querying dataset, and introduced a stepwise approach to first parcellate those query images most similar to the atlas database, and then propagate the labels successively to the less similar ones. This approach breaks the large deformation into sequential smaller ones and has been shown to improve both the accuracy as well as the robustness. Recently, Cardoso *et al.* [126] have proposed a similar approach that embeds the query images into a spatially-variant morphological graph, and progressively diffuses the structural labels from the atlas database to the target images through geodesic information flow, which demonstrated stable and accurate result when applied to various datasets.

2.1.4 State-of-the-art: Multi-atlas parcellation for mouse brain MRI

Structural parcellation based on multi-atlas label fusion have gained a lot of attention recently and lent itself to many applications in the clinical neuroimage studies [127, 128]. However, only a handful of studies have applied multi-atlas-based

structural parcellation techniques to preclinical data. Lancelot *et al.* [129] applied majority voting on rat brain and showed better performance in terms of dice score over the probabilistic-atlas-based approaches. Artaechevarria *et al.* [106] segmented *ex vivo* mouse brains using a mutual information-based weighted majority voting label fusion, which showed improvement in parcellation accuracy when compared to a simple majority voting method. Nie *et al.* [130] proposed a mouse brain structural parcellation using weighted voting label fusion, in which the diffeomorphic image registration is mainly driven by the voxels around the label boundary, and further improved through an SVM-guided surface deformation.

Compared to clinical studies, the number of mouse brain atlases databases that can be used for evaluating the multi-atlas study, as well as the number of atlases in each database, are much limited. A complete list of current mouse brain atlas is shown in table 2.1. In the case where only one atlas is available in the database, Chakravarty *et al.* [131] proposed a method which first propagates the atlas labels to a set of unlabelled query images using a conventional single-atlas label propagation approach, and subsequently the resulting set of structural labels were regarded as a multi-atlas database to parcellate each query image using majority voting, demonstrating improvements in terms of parcellation accuracy when compared with direct single-atlas parcellation propagation. However, a potential risk of this approach is the accumulation or aggravation of the registration errors when propagating the labels due to the morphological bias of the original single atlas of choice.

Furthermore, the current preclinical studies are largely dominated by *ex vivo* data sets. Researchers have been trying to shift the image acquisition protocols from *ex vivo* towards *in vivo* imaging, reducing artefacts from tissue preparation and enabling longitudinal studies [132, 3, 133]. However, *in vivo* studies inevitably generate images with much lower contrast/signal-to-noise ratio due to the shortened scanning time and the limited use of contrast agents. Scheenstra *et al.* [89] proposed an automatic structural parcellation of *in vivo* mouse brain MR images by first performing a single-atlas affine registration-based parcellation, followed by an edge-based clustering in order to achieve a fast parcellation. This method is shown to achieve the same level of parcellation accuracy compared to non-rigid registration. More recently, Bai *et al.* [79] conducted a study to compare structural parcellation accuracy on *in vivo* mouse brain

MRI using various methods and found no significant improvement when using a more advanced label fusion algorithm, the Simultaneous Truth and Performance Level Estimation (STAPLE) [97, 98], or a MRF approach alone, compared to a simple majority voting approach. There is a current need for robust methodologies tailored for *in vivo* mouse brain MRI, while current studies are definitely not adequate.

In Chapter 3, a structural parcellation framework applicable to mouse brain MRI is introduced which uses the above mentioned multi-atlas label fusion algorithm, STEPS, proposed by Cardoso *et al.* [117]. And the framework has been evaluated on both *in vivo* and *ex vivo* mouse brain MRI data.

Table 2.1: Current publicly available mouse brain MRI atlas database. Only two databases provide multiple atlases of individual animals - the MRM NeAt and the NUS atlas - which are selected to validate the multi-atlas label fusion of the proposed framework.

Data source	**Mouse Image Centre, Toronto Centre for,Phenogenomics** [1]	**John Hopkins University School of,Medicine** [2]	**Mouse Atlas Project 2.0, Laboratory of Neuro Imaging, UCLA** [3]	**Centre for *in vivo* microscopy, Duke,University Medical Centre** [4]	**Magnetic Resonance Microimaging Neurological Atlas, University of Florida** [5]	**National University of Singapore** [6]	**Australia mouse brain mapping consortium** [7]
Mouse type	12 weeks C57B1/6J + 8 week 129S1 / Svlmj, 129SV/S1	C57BL/6 (E12-18,P0-P80 and older)	Neonatal C57BL/6J,(p0); 100 days C57BL/6J	9-12 weeks C57BL/6J	12-14 week C57BL/6J	Adult C57BL/6J	12 weeks male C57BL/6J
Imaging type	T2W	T2W + DTI (FA, DEC)	T2W (neonatal);DWI (adult)	T1, T2W, T2*, DTI	T2* *ex vivo* / *in vivo*	T2W	T1/T2*
Atlas type	Average atlas	Average atlas	Average atlas	Average atlas	Individual, average and probabilistic atlas	Individual atlas	Average atlas
Animal and structure number	62 structure, 40 samples / 40 labels, 9 samples	Major gray/white structures; Embryonic,(E12-E18)+postnatal (P7,11,15,21,28,42,60); 6 samples for each time point	Neonatal: 13 structure (by averaging brain map/ probability maps threshold); Adult: not mentioned	9 white matter structure; 8 samples	Ex,vivo: 20 structures,10 samples,In,vivo: 20 structures,12 samples	In,vivo: 40 structures, 5 samples	Ex,vivo: 38 cerebellum subregions 74,cortical subregions,
Scanner field	7T (Magnex+Varian)	11.7T (Bruker)	11.7T (Bruker)	9.4T (GE)	Ex,vivo: 17.6T (Bruker),In,vivo: 9.4T (Magnex+Bruker)	In,vivo: 7T (Bruker)	Ex,vivo: 16.4T (Bruker)
Resolution	C57B1/6J: 32μm; 129S1/Svlmj: 54μm	Embryo: 80μm; Adult: 125μm	Neonatal: $50\times70\times70\mu$m;Adult: 60μm	43μm (voxel volume $\tilde{8}$0 pl)	*ex vivo*: 47μm; *in vivo*: 100μm	*in vivo*: 100μm	*ex vivo*: 30μm

[1] [71, 72, 73] http://www.mouseimaging.ca/technologies/mouse_atlas.html
[2] [75, 134] http://lbam.med.jhmi.edu
[3] [135] http://map.loni.ucla.edu http://www.birncommunity.org/data-catalog/mouse-3d-mr-minimum-deformation-atlas
[4] [136, 92] http://www.civm.duhs.duke.edu
[5] [78, 3] http://www.bnl.gov/ctn/mouse http://phenome.jax.org
[6] [79] http://www.bioeng.nus.edu.sg/cfa/atlas/mouse.html
[7] [80, 81, 82, 47] www.imaging.org.au/AMBMC/AMBMC

2.2 Volumetric analysis of *in vivo* and *ex vivo* mouse data

In neuroimaging studies, quantitative analysis of neuroanatomy, such as volumetric analysis of brain structures, has played a crucial role in the diagnosis of diseases at the early stages of the pathology before the onset of clinical symptoms [137]. However, the analytical power of volumetric measurement varies depending on the segmentation accuracy, which in turn depends on the image qualities such as resolution, signal to noise ratio (SNR) and contrast to noise ratio (CNR) [138, 139].

Controversial arguments exist regarding whether to acquire MR images alive or post-mortem[23]. For mice scanned *ex vivo*, there are no motion or flow artefacts, the prolonged scanning time enables increased image resolution, and the tissue contrast can be enhanced by perfusion with contrast enhancement agents such as Gadolinium [21]. On the other hand, *ex vivo* imaging suffers from direct or indirect morphological changes in brain tissues due to post-mortem tissue processing such as fixation and perfusion, while *in vivo* imaging ensures minimum distortion to the tissue morphology. Furthermore, with *in vivo* imaging, it is possible to trace the morphological change and/or other physiological changes of each individual animal longitudinally [140].

Schulz *et al.* [141] utilised the deformation field of temporal registrations between images of a single postmortem brain scanned at different time points after formalin fixation, to investigate the effect of fixation on the structural volume in the human brain. The affine image registration showed a 5.2% expansion of the total brain volume one day after brain extraction and fixation, and eventually an 8.1% shrinkage after 70 days fixation. Non-rigid image registration additionally revealed that the volume shrinkage was distributed unevenly across different structures, with a maximum 32% local volume shrinkage. On the contrary, a study by Kotrotsou *et al.*d [142] found a linear correlation between *in vivo* and *ex vivo* brain volumes (in the grey matter only) using structural MR imaging of human brain hemispheres, but with no significant change in the brain volume in the *ex vivo* images during the 6 months fixation perio.

Studies have also been conducted on preclinical imaging data. Ma *et al.* imaged the mouse brain (strain: C57BL/6J) both *in vivo* [3] and *ex vivo* [78], and segmented the brain into 20 functional structures (through manual labelling with subsequent semi-

automatic segmentation propagation). In their results, the total brain volume from the *ex vivo* data was 10.6% smaller than the *in vivo* data, where some parts of the grey matter shrank from *in vivo* to *ex vivo* and others expanded. The ventricles shrank by 78.6% due to the leakage of cerebrospinal fluid (CSF). However, the *in vivo* and *ex vivo* imaging data were acquired from different mouse populations (of the same strain). In addition, the *ex vivo* image was scanned after physical skull removal, with notable brain tissue loss from the image.

Zhang *et al.* [140] studied the variations in structural volume in mice through manual labelling. The post-mortem brain was kept inside the skull and fixed for a week. The results of their study showed a 4.5% brain shrinkage in wild type mice after perfusion, 8% brain shrinkage in Huntington's Disease model mice (genotype: R6/2), while the entire lateral ventricle collapsed in both normal (99% shrinkage) and transgenic (97.3% shrinkage) mice. Oguz *et al.* [143] also compared the accuracy of the *in vivo* and *ex vivo* imaging acquired from rats (male Wistar), by segmenting the brain structures with a single-atlas label propagation strategy, using an in-house atlas database which includes only *ex vivo* images. They found no significant change between *in vivo* and *ex vivo* measurements of total brain volume, regional volumes of 22 anatomically defined structures, or ventricles, and concluded that *ex vivo* scanning was a valid and reliable method of estimating volumetric rodent brain measurements.

Scheenstra *et al.* [89] developed a fast single-atlas label propagation algorithm based on affine registration followed by edge-based clustering, then applied it to both *in vivo* and *ex vivo* mouse brain MRI data, and concluded that the algorithm provided better results when segmenting *ex vivo* data in terms of agreement with the manually segmented gold standard (Dice score), except for the olfactory bulb and ventricles.

Finally, Lerch *et al.* [138]. assessed the difference in theoretical statistical powers between *in vivo* and *ex vivo* data by using a pre-determined variance value to perform simulated statistical analysis on the mouse hippocampus They concluded that *ex vivo* imaging provides better precision and should be preferred if the relative volume (normalised with brain volume) is measured; while *in vivo* measurements give better results on the absolute volume measurements, and are more sensitive to longitudinal changes than cross-sectional *ex vivo* measurements. However, studies comparing the statistical power of volumetric analysis on real data are needed to confirm these simulated

findings.

In Chapter 4, I investigate this issue by applying the structural parcellation framework developed in Chapter 3 to compare the difference between volumetric analyse of *in vivo* and *ex vivo* images acquired in the same cohort of animals.

2.3 Cortical thickness morphometric analysis

The morphometric characteristics of the cortex, such as volume, surface and thickness have been widely used to study brain function and physiology. Cortical structural abnormalities, such as variation in cortical thickness and surface area, have been shown to correlate with various neurological disorders as well as cognitive functions [144, 145, 146]. Volume-based cortical morphometry presents a mixed measurement of cortical thickness, folding and surface area, while other measurements like thickness estimation manage to investigate each specific factor separately [147].

2.3.1 Cortical thickness estimation

Cortical thickness estimation is a widely used morphological measurement to detect local cortical variation and provide quantitative analysis of neurodevelopment, normal aging or various neurodegeneration diseases (e.g., Huntington's disease, dementia such as Alzheimer's disease, or schizophrenia), both in clinical [148, 144, 146] and preclinical settings [40, 149, 42]. The cortical thinning is non-uniformly distributed across cortical surface and is regional dependent. Longitudinal studies on cortical thickness can reveal the disease progression as well as potential treatment response [150, 42].

All thickness estimation methods start from the definition of inner (white matter) surface and an outer (pial) surface of the cortex, obtained from an accurate cortical segmentation (either binary or probabilistic).-based on the implementation, current cortical thickness estimation methods can be generally categorised into two types: surface-based, and voxel-based approaches. Fig 2.6 shows a schematic diagram of these different approaches.

2.3.1.1 Surface-based approaches

Dale *et al.* [34] and Fischl *et al.* [35] introduced the surface-based cortical thickness estimation, which has been widely adopted with the distribution of the FreeSurfer package. In this approach, tessellated surfaces are first constructed for the WM boundary as

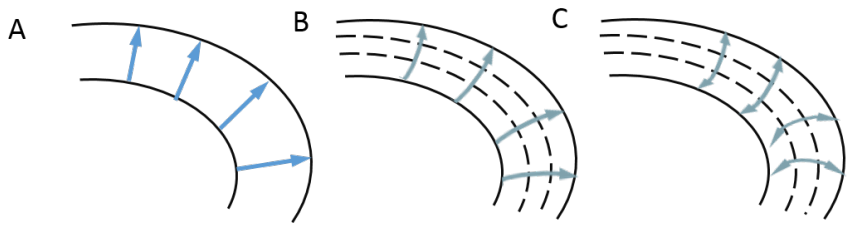

Figure 2.6: Schematic diagram of different cortical thickness estimation models. A: Surface based model; B: Voxel-based deformation approach; C: Voxel-based PDE approach

a triangulated mesh with subvoxel accuracy following a series of processing steps. An iterative deformable model is then adopted to evolve and refine the tessellated surfaces to construct the smoothed white matter surface and pial surface, which are both second order smooth, while maintaining the topology and local curvature (surface repositioning). The cortical thickness is then derived by firstly estimating the distance of each vertex point on the surface to the closest point on the opposite surface, and then taking the average of the distances measured from each side [150].

Han *et al.* [151] proposed another surface-based thickness estimation approach - Cortical reconstruction using implicit surface evolution (CRUISE) - which uses a topology-preserving geometric deformable surface model to evolve the surface in a level set function and is propagated implicitly through temporal evolution.

MacDonald *et al.* [152] also introduced a model which extracts the white matter surface and the pial surface simultaneously by deforming a set of polyhedral surfaces to minimise the objective function including an image term, a stretch term, and a bending term along with a vertex-vertex proximity constraint. Three definitions of the cortical thickness from the deformed inner/outer surfaces are tested and compared:

- T_{link} - derived directly from the linkage of the two surfaces in the algorithm;
- T_{near} - the distance to the nearest point on the other surface; and
- T_{normal} - the distance in the direction of the surface normal.

The three definitions showed similar cortical thickness map patterns, although their measurement significantly differs from each other in absolute terms: the T_{normal} shows largest standard deviation (doubled the other two measurements), while the T_{near} is considerably dissimilar to the other two after pairwise comparison among the three.

The surface-based methods have shown robust results in various applications and are resistant to image artefact such as noise along with the topology preservation and

smoothness terms. However, the explicit modeling requirement makes it computationally very expensive, especially in highly convoluted regions, and the sensitivity may be biased toward the parameters of choice.

2.3.1.2 Voxel-based approaches

In the voxel-based approach, the cortical thickness is determined progressively through all the voxels within the grey matter region. Three methods have been proposed: deformation-based model, projection-based model and partial-differential-equation (PDE)-based model.

Deformation-based model: Das *et al.* [153] proposed to estimate the thickness through deformation/registration similar to the surface-based approach, but on the volumetric image space instead. In this method, the white matter segmentation is diffeomorphically deformed outwards to the entire brain (including both white matter and grey matter) through a velocity field $v(x,t) : \Omega \times t \in [0,1] \rightarrow \mathbb{R}^d$ where Ω is the region within grey matter, v is the regularised vector field, t is the "time" parameter and d is the dimension. The thickness is determined on each voxel on the initial surface as the distance travelled within the deformation field, and is propagated across the entire cortical area. The corresponding differentiable map (diffeomorphism) on the transformation space $\phi(x,t) : \Omega \times t \in [0,1] \rightarrow \Omega$ can be calculated as:

$$\phi(x,t) = x + \int_0^t v(\phi(x,t),t)dt \tag{2.17}$$

Projection-based model: Dahnke *et al.* [154] proposed a project-based model, in which the cortical thickness of each grey matter voxel is estimated as the accumulated distance towards the white matter surface (WMD). Distance to the boundary is measured with an Eikonal equation with a non-uniform speed function $F(x)||\nabla D(x)|| = 1$ for $x \in \Omega$, which finds the closest white matter boundary voxel $B(x)$. This method progressively searches the largest local white matter distance (WMD) within the non-boundary grey matter regions which is readily the cortical thickness, avoiding the need to explicitly reconstructing the sulcal lines on the pial surface through methods such as skeletonisation. However, the definition of thickness as the WMD derived from Eikonal equation is still oversimplified, and may not represent the underlying morphology of microcolumn arrangement in highly convoluted cortical regions.

Partial-differential-equation-(PDE-)based model: The PDE-based method is another popular approach for voxel-based thickness estimation. In this approach, the laminar layers between the two surfaces were constructed through PDE, and the thickness is defined as the length of the streamlines (representing the underlying microcolumns) passing through each voxel which is orthogonal to the laminar layers [155, 33, 36]. The original idea was proposed by Jones *et al.* [33], which models the stratification in the laminar layers of the cortex as a Laplacian field T, and can be solved using the Jacobi method:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0 \,. \tag{2.18}$$

The thickness is defined on each voxel, and can be estimated either through explicit integration along the unit vector $\overrightarrow{F} = \frac{\nabla T}{\|\nabla T\|}$ with a specific step size δd [33]. The partial volume effect on the boundary voxels can be solved with methods such as sub-sampling and interpolation [155]. Yezzi *et al.* [36] introduced an Eulerian PDE-based method to solve the PDE implicitly, in which the length towards the white matter surface L_0 and the length towards the pial surface L_1 can be updated iteratively through the following upwind scheme of finite difference:

$$L_0[i,j,k] = \frac{1 + |T_x| L_0[i \mp 1, j, k] + |T_y| L_0[i, j \mp 1, k] + |T_z| L_0[i, j, k \mp 1]}{|T_x| + |T_y| + |T_z|} \,, \tag{2.19}$$

$$L_1[i,j,k] = \frac{1 + |T_x| L_1[i \pm 1, j, k] + |T_y| L_1[i, j \pm 1, k] + |T_z| L_1[i, j, k \pm 1]}{|T_x| + |T_y| + |T_z|} \,, \tag{2.20}$$

where $T_x[i,j,k]$, $T_y[i,j,k]$ and $T_z[i,j,k]$ denotes the vector component of $\overrightarrow{T}$ at voxel coordinate (i,j,k), and the L_0 and L_1 can be updated simultaneously.

Waehnert *et al.* [2] has recently proposed an improved anatomically motivated cortical laminae model, which is initialised with a geometric deformable model, and imposed a constraint to maintain constant volume between laminae by using the curvature information on the boundaries. The initial geometric deformable model is solved through a narrow band level set method which evolves the inner (WM) surface θ_{in} and the outer (pial) surface θ_{out} to a certain level set θ when giving a specific cortical depth d:

$$\frac{\partial \theta}{\partial t} + (\theta - \theta_{\{in,out\}}) \cdot |\nabla \theta| = \varepsilon \kappa |\nabla \theta| \,, \tag{2.21}$$

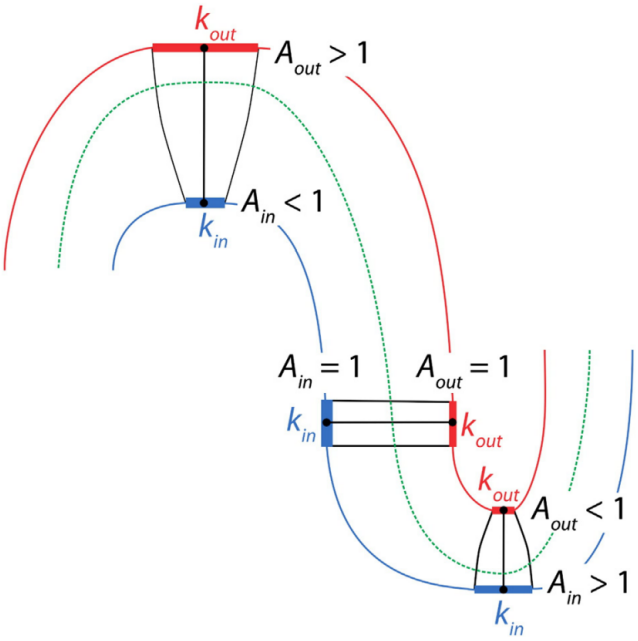

Figure 2.7: Schematic diagram of the equivolume model for cortical lamination. The cortical volume is constant across sublayers. As a result, layers with higher curvature becomes thinner. Figure taken from [2].

and the target level set θ_d is the weighted average of the inner cortical surface θ_{in} and the pial cortical surface θ_{out}:

$$\theta_d = (1-\rho)\cdot\theta_{in} - \rho\cdot\theta_{out}, \rho \in [0,1]\,. \tag{2.22}$$

The curvature-based equivolume constraint applied on the initial level set is based on the anatomical observation from studies of cortical cytoarchitecture, that (1) the laminae layers with higher curvature are relatively thicker and vice versa; (2) the volume fraction stays constant in a given cortical segment (micrcocolumn). The model assumes that the surface area grows linearly along the column segmentation from the inner surface where the height $h = 0$ to the pial surface where $h = d$:

$$A(h) = A_{in} + (A_{out} - A_{in})\cdot\frac{h}{d}\,, \tag{2.23}$$

where the surface areas A_{in} and A_{out} are determined by the corresponding surface curvatures, and the volumes are calculated by integrating the areas from $h = 0$ to $h = d$ (Fig. 2.7).

Based on the same equivolume principle, Leprince *et al.* [37] developed a similar

equivolume laminar model by integrating the volumes directly along the streamlines derived from the Laplacian field. In this model, a unit surface δS is defined for each cortical voxel and advected along on the streamline. The relative surface area changes of the adjacent points along the streamline are obtained from the divergence of the vector field $\nabla \cdot \overrightarrow{F}$:

$$\delta S_{n+1} = \delta S_n (1 + \delta l (\nabla \cdot \overrightarrow{F})) \,, \tag{2.24}$$

where δl is the step size. The δS can be integrated along the directions towards the two cortical surfaces to obtain the upwind volume towards the pial surface $V_{pial} = \int_{X_0}^{X_{pial}} \delta S$ and the downwind volume towards the white matter $V_{WM} = \int_{X_0}^{X_{WM}} \delta S$. The laminar information on each voxel is defined as the relative equivolumic depth d_{vol} as:

$$d_{vol} = \frac{V_{pial}}{V_{WM} + V_{pial}} \,. \tag{2.25}$$

2.3.1.3 Comparison between surface- and voxel-based approaches

In conclusion, the surface-based approach is very robust to the presence of image noise due to its implicit model parameter constraints of smoothness and local curvature, which however makes it computationally very expensive. In addition, these topological correction procedures might also introduce model bias and loss of sensitivity of local fine changes. On the other hand, the voxel-based method is relatively more prone to image noise which requires careful preprocessing steps. However, it is more sensitive to large local morphological variation and is computationally more efficient.

Clarkson *et al.* [9] conducted a cross sectional study to compare the surface-based and voxel-based approaches. Both methods demonstrated comparable estimation accuracy, with the surface-based method demonstrating better test-retest consistency while voxel-based approach showed higher sensitivity. Li *et al.* [156] further applied all three methods on a longitudinal stroke data, and find highly correlation between the two voxel-based methods: the registration-based method and the Laplacian based method.

2.3.2 Study of mouse cortical thickness estimation

2.3.2.1 Cortical thickness for mouse cerebrum

Cerebral cortical thickness has been investigated extensively in human brain MRI, and successfully applied in clinical to study the neurodevelopment [148, 157] neurodegen-

erative diseases [158, 159, 39, 160] in human cortex. Studies have also applied the cerebral cortical thickness estimation for preclinical images of mouse and rat as biomarkers for neurodegenerative diseases [40, 149], various genetic conditions [161, 162] or other non-genetic conditions [38, 163]. Compared to human cortex, there were no complex convolution and folding patterns in mouse cortex, and less variability between individuals, which results in a more accurate registration/deformation, and more plausible and consistent thickness estimations.

2.3.2.2 Sublayer thickness estimation in the mouse cerebellum

Current studies on cortical thickness for neurological disorders mainly focused on the cerebrum, even though the cerebellum is also affected. Cerebellar damage, deficit or volume change have been shown to correlate with various neurological deficiencies in clinical studies [164, 165, 166]. Although cerebellar cortical thickness analysis could provide more insights about neurodegenerative diseases and phenotype identification, these are currently limited by the resolution of clinical MRI and the highly convoluted nature of the human brain (as shown in Fig. 1.5).

Giving the genetic similarities and anatomical correlation between human and mouse brains, mice represents a promising model to understand further how the cerebellum is affected by the progression of neurodegenerative diseases. Currently, there is a limited number of studies focusing on the quantitative analysis on mouse cerebellar MRI, of which only volumetric analysis were conducted [80, 167, 168]. To the best of my knowledge, no previous studies have addressed the problem of estimating the cerebellar cortical thickness on 3D volumetric images such as MRI.

The cerebellar cortex consists of three sublayers - the granular layer, the Purkinje layer, and the molecular layer (ordered from the white matter towards the external CSF) - each comprised of different cellular type, corresponding to different functions [169, 170], and their pathology varies with different conditions [171, 172]. Studies from Eickhoff *et al.* [173] and Baxter *et al.* [174] showed variation in sublayer morphology using 2D histology data with various staining technique. However, studies about sublayer thickness variation in 3D MR structural images are still lacking, mainly due to the limited resolution as well as the lack of inter-layer contrast in the cortical regions. Cleary *et al.* [21] have recently applied the active staining technique on high

resolution structural mouse brain MRI, and demonstrated the possibility to produce contrast between different cellular layers within the mouse brain cortex. From the active stained gadolinium enhanced MRI, the irregular shape of the Purkinje layer within the cortex could not be captured through mathematical modelling of the laminar layer (e.g. in the paraflocculus in the cerebellar hemispheres - PF1, as shown in Fig.2.8). However, to the best of my knowledge, no study has been using these kinds of intrinsic contrasts available in MRI to investigate the thickness variations within the cortical sublayers.

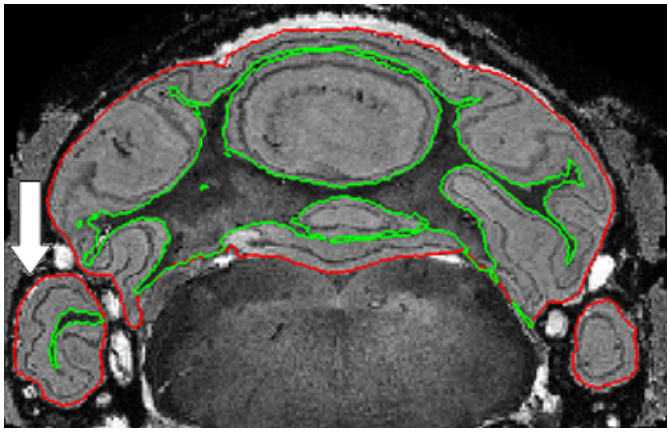

Figure 2.8: The uneven distribution of sublayer thickness due to morphological variation.

In Chapter 5, a framework is developed to investigate this issue, and estimate the mouse cerebellar cortical sublayer thicknesses from the contrast enhanced *ex vivo* mouse cerebellar MRI. The voxel-based thickness estimation approach is used with the cerebellar cortical laminar layers estimated using the equivolume model proposed by Leprince *et al.* [37]; the Eulerian PDE-based method by Yezzi *et al.* [36] is used to construct the thickness map; and a two-step contrast- and laminar-layer-model-based extraction method is proposed to segment the middle Purkinje layer from the cerebellar cortex, and estimate the thickness for the inner granular layer and the outer molecular layer.

2.4 Current framework available for preclinical image analysis

Current publicly available and widely used tools for brain morphometric analysis, including integrated pipeline packages such as SPM[8] [175], Freesurfer[9] [1], FSL [176], 3D Slicer[10] [177], Brainsuite[11] [178] and Brainvoyoger[12] [179], or individual tools like ANTs[13] [180] and IRTK[14] [181], are mainly for clinical data. Conversely, limited resources are available for preclinical image analysis, in terms of both analysis tools, and databases which can be used as prior knowledge to train algorithms. Efforts have been made in some studies to address this imbalance between huge demand and limited supply of preclinical image computing by adapting existing algorithms and methods of medical image computing to the preclinical field. Gerig *et al.* [14] created a synergistic cross-species image analysis pipeline. Budin *et al.* [182] introduced a rodent brain MR image processing platform (Midas server) with a web interface which provide optimised parameters for structural parcellation of preclinical data. But the structural parcellation tools provided in these studies only incorporate single atlas label propagation which will provide limited accuracy compared to the probabilistic-atlas-based and multi-atlas-based methods. And none of the mouse cortical thickness analysis tools provided in the these platforms provide solutions for mouse cerebral cortex, not to mention the cortical sublayer thickness.The focus of this thesis is to solve those limitations of the current preclinical neuroimage analysis framework.

[8] http://www.fil.ion.ucl.ac.uk/spm/
[9] http://freesurfer.net/
[10] http://www.slicer.org
[11] http://brainsuite.org/
[12] http://www.brainvoyager.com/
[13] http://picsl.upenn.edu/software/ants/
[14] http://biomedia.github.io/IRTK/

Chapter 3

Automatic structural parcellation of mouse brain MRI using multi-atlas label propagation and fusion

3.1 Introduction

Multi-atlas label propagation and fusion has evolved quickly in recent years, becoming the state-of-the-art methodology of automatic structural parcellation for medical images. However, few studies have applied this method to preclinical research. In this chapter, I present a fully automatic open source framework[1] for mouse brain MRI structural parcellation using multi-atlas label propagation and fusion. The framework consists of (1) several preprocessing steps (2) a non-rigid B-spline parameterised registration [86, 183, 184], and (3) the label fusion method STEPS [117] as described in Section 2.1.3.5. I investigated the parameters of the STEPS label fusion algorithm, and optimised those parameters for the *in vivo* mouse brain atlas database: the MRM Neurological Atlas (MRM NeAt), provided by Ma et al [3]. The performance of the proposed framework was evaluated using leave-one-out cross validation, and compared with a single-atlas-based method without any label fusion technique, as well as with the STAPLE label fusion algorithm, which is based on global image similarity measurement and is the state-of-the-art label fusion method at the moment this framework was developed. To demonstrate the ability of this framework to parcellate new unlabelled query images, I adopted a separate *in vivo* mouse brain MRI atlas - the NUS mouse

[1] https://github.com/dancebean/multi-atlas-segmentation

atlas [79] - as testing dataset, with the MR images in it regarded as unlabelled images for validation. Furthermore, I tested the ability of the framework to detect volumetric differences between the brain structures of two groups of mice with or without genetic modification. Optimised parameters were obtained for the STEPS label fusion algorithm to achieve the best parcellation accuracy. The proposed multi-atlas framework resulted in significantly higher parcellation accuracy compared to single-atlas-based structural parcellation, as well as compared to the original STAPLE framework.

3.2 Material and methods

In Section 3.2.1, the multi-atlas framework for automatic structural parcellation is introduced step by step. Section 3.2.2 describes the *in vivo* mouse brain atlas databases that are used for evaluation. In Section 3.2.3, I present the optimisation of the STEPS label fusion algorithm in the framework and evaluate its performance using the *in vivo* atlas. In Section 3.2.5, I assess the ability of the proposed framework to parcellate new unlabelled *in vivo* MRI data. In Section 3.2.6, I investigate the ability of this framework to identify structural differences between two groups of animals. Finally in Section 3.2.7, I introduce the process of adding the flipped version of the atlas to the database to improved the parcellation accuracy.

3.2.1 Automated multi-atlas structural parcellation framework construction

The proposed automated multi-atlas structural parcellation framework includes two pre-processing steps - brain extraction and bias field correction - followed by a series of non-rigid registrations and a final label fusion step. Fig. 3.1 shows a step-by-step summary of the pipeline.

Brain extraction (Fig. 3.1 A,B). Brain extraction is an important pre-processing step to confine the analysis of regions of interest (ROI) to areas specifically within the brain region. During this step, brain tissues in the unlabelled query image are masked thereby excluding non-brain tissues and background. The mask is automatically created from the atlas images through the following steps. Firstly, the query image is globally aligned to all atlas images through affine registration, with the cost function in the optimisation step calculated over only the voxels inside the atlas mask and their corresponding

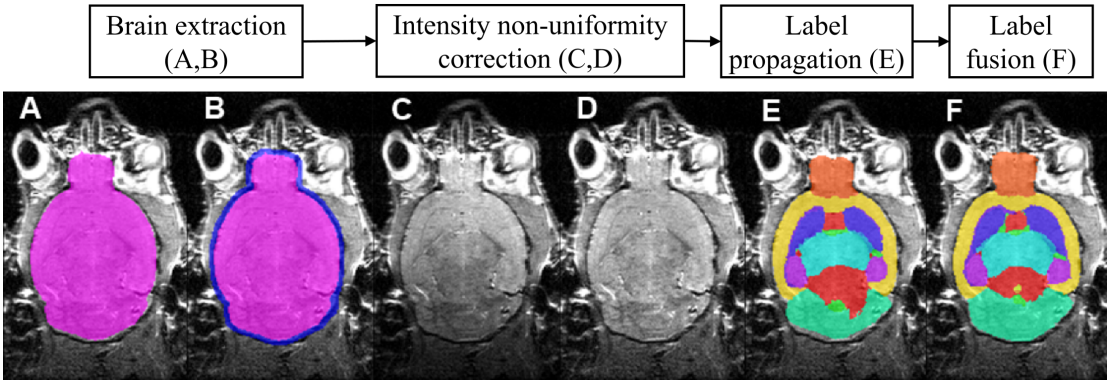

Figure 3.1: Step-wise summary of the framework. Pipeline of the framework is shown at the top of the image. Below the pipeline are representative images of results obtained after each processing step of the framework when applied to an unlabelled query image. (A) A brain mask is created for bias field correction; (B) The mask is dilated to include contrast of brain tissues and CSF for image registration; (C, D) Images before and after bias field correction; (E) Structural parcellation result after single-atlas label propagation; (F) Structural parcellation result after multi-atlas label fusion.

voxels in the warped query image. This affine registration step is performed using a symmetric block-matching approach [185, 186]. Secondly, the resulting transformation matrices are inverted and used to propagate all the atlas brain masks to the query image. Thirdly, all the brain masks propagated from the atlas database are fused using majority voting in order to obtain a consensus brain outline. Finally, the mask is slightly dilated to include the surrounding external CSF (e.g. 2 voxels for the testing image with 150µm voxel resolution) so that the contrast between brain tissue and the surrounding CSF can be captured by non-rigid image registration in a later step.

Intensity non-uniformity correction (Fig. 3.1 C,D). MR images are corrupted by intensity non-uniformity, caused by the inhomogeneity of both the RF excitation field and the receiver coil sensitivity. This may lead to misalignment in the registration process. An established standard method to correct the intensity non-uniformity is Non-parametric Non-uniformly intensity Normalisation method (N3) developed by Sled *et al.* [187]. The N3 algorithm models the observed image v as $v(x) = u(x)f(x) + n(x)$, where $u(x)$ is the true image, $n(x)$ is the additive noise and $f(x)$ is a smooth, slowly varying, multiplicative bias field, and is derived by maximising the frequency content of the distribution of the true image intensity. The $f(x)$ is estimated iteratively by progressively increasing the width of the distribution probability density of $f(x)$. We adopted the N4ITK, an improved version of the N3 algorithm proposed by Tustison *et*

al. [188]. The N4ITK introduced a new B-spline approximator for smoothing the bias field, which allows for smaller control points spacing to adapt greater field strength. It also include a modified optimisation scheme which speed up the convergence rate. We adjusted the default parameters to accommodate the resolution of mouse brain MRI. The characteristic distance over which the field varies set to 10 mm, the deconvolution kernel used to sharpen the histogram set to 0.15 mm.

Image registration and label propagation (Fig. 3.1 E). After the intensity non-uniformity correction was implemented, the affinely aligned atlas images obtained in the brain extraction step were non-rigidly registered to the query image. Similarly to the affine registration, the non-rigid registration aims at maximising image similarity between the reference and floating image. The atlas image and test image might be acquired with different image modalities, and the normalised mutual information (NMI) is selected as the image similarity measure which is a robust measurement of inter-modality similarity [189, 190]. A symmetric scheme based on a cubic B-Spline parameterisation of a stationary velocity field [86, 183, 184] is employed to model the transformation. A bending energy with weight of 0.005 is imposed as regularisation term to enforce a smooth deformation. The derived transformations were then used to deform the manually parcellated structural labels of the atlas images into the space of the query image. Nearest-neighbour interpolation was used to preserve the integer nature of the labels.

Label fusion (Fig. 3.1 F). After all the propagated structural labels were obtained from the image registration step, they were fused to generate the final result of structural labels for the query image. The STEPS algorithm developed by Cardoso et al. [117] is used to perform the label fusion. As described in Chapter 2 STEPS is an extension of the original STAPLE algorithm proposed by Warfield *et al.* [98], and extended by Rohlfing et al. [97]. The original STAPLE algorithm was developed with the purpose of fusing several expert-delineated manually labelled anatomical structures in order to obtain the hidden ground truth structural parcellation. Several improvements over STAPLE were introduced in STEPS: including a LNCC based local similarity measurement as ranking strategy for template selection and an integrated MRF regularisation [117]. The LNCC is less affected to the MRI intensity inhomogeneity compared to the NMI because the image similarity is computed over local region. The mathematical detail

of the STAPLE and STEPS algorithm has been described in Section 2.1.3.4 and Section 2.1.3.5 respectively. Cardoso *et al.* [117] have compared the STEPS algorithm to several other label fusion methods available at the moment when it's been developed, and demonstrated that STEPS resulted in the highest structural parcellation accuracy in many key internal structures, and was the most robust to reduction in the size of the database. Representative resulting images obtained after each processing step of the framework when applying to an query image are shown in Fig. 3.1.

3.2.2 Mouse brain atlas

To optimise the parameters for label fusion and evaluate the accuracy of the parcellation results, the publicly available mouse brain MRI atlas databases were used. To the best of my knowledge, there are 7 publicly available atlas databases [92, 79, 75, 72, 136, 73, 78, 78, 81, 81]. A detailed comparison of all the databases is presented in Table 2.1 in Chapter 1. Most of databases contain only one structurally labelled average atlas (the minimal deformation atlas). Only two of them contain the structural labels for each individual atlas sample, the Magnetic Resonance Microimaging Neurological Atlas (MRM NeAt) [78, 78] and the National University of Singapore (NUS) atlas (Bai *et al.*, 2012), which are suitable for the proposed multi-atlas-based framework (Fig. 3.2). Both of these atlas databases include *in vivo* image samples and the brain structures were both manually parcellated following the Franklin-Paxinos atlas [191]. The MRM NeAt database includes atlases of 12 individual T2*-weighted brain MR images of 12-14 weeks old C57BL/6J mice; each with 20 manually labelled anatomical structures. The NUS mouse atlas database includes 5 individual T2-weighted brain MR images of adult male C57BL/6J mice, each has 40 manual labelled anatomical structures. Detailed scanning parameters are described in [3] and [79]. Heckemann *et al.* [99] has previously shown that increasing the number of images in the atlas database can improve the accuracy of the label fusion. As a result, the MRM NeAt atlas database, which has the largest number of atlases, was selected for this part of the study. Due to missing labels in 2 of the 12 available atlases in the MRM NeAt database, only 10 images and associated structural labels were included.

For neurodegenerative diseases, the progression of pathology might vary between two hemispheres [114]. Furthermore, for studies interested in further estimating the

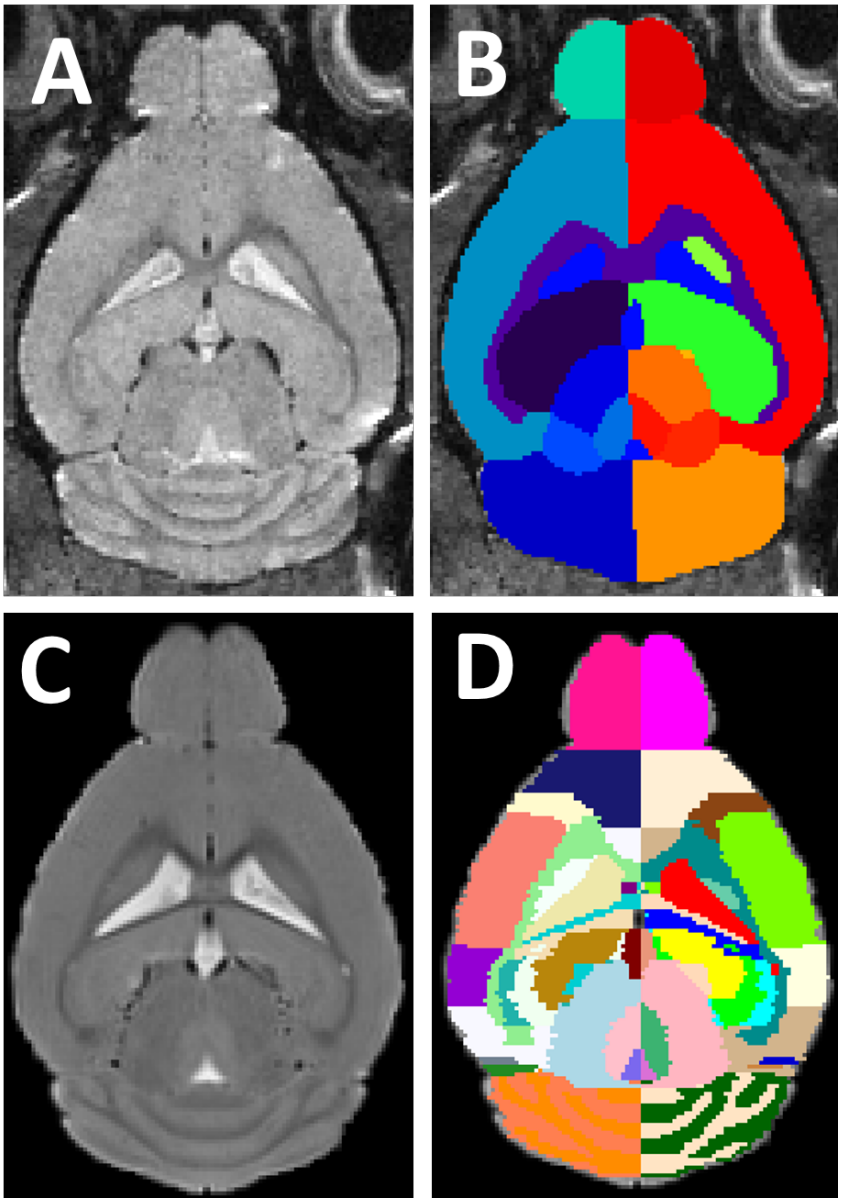

Figure 3.2: Sample images of current multi-atlas database. Images on the top show (A) a template image of the MRM NeAt atlas database, with (B) its corresponding manual delineated structural labels. The image at the bottom showed (C) a template image of the NUS atlas database, with (D) its corresponding manual delineated structural labels.

cortical thickness from the structural parcellation result, hemisphere separation can also help to identify and segment the intra-hemispheric cortical surface area [33, 40]. It is thus preferable to separate the structural labels of the original atlas into left/right hemispheres. I thus separated the brain images and their corresponding structural labels in the original atlas database into left and right hemispheres along the mid-sagittal plane. Maes *et al.* [192] achieved left and right hemisphere separation for asymmetry

measurement using registration-based label propagation. An alternative way to determine the inter-hemisphere separation plane is to exploit the symmetric nature of the MR data. This method aims at finding the reflective rigid-body transformation that minimizes the absolute distance of an image and its mirrored version [193, 194, 195]. This method is only valid for brain images from wild type mouse strains, for which no left/right asymmetries are induced by diseases. Since the mice in the atlas database used in this study are wild type animals, the latter method was selected by firstly flipping the atlas images and then using NMI as an asymmetry measurement to find the mid-sagittal plane.

3.2.3 Parameter optimisation

In the STEPS label fusion algorithm, the best local labels are selected after ranking based on the LNCC computed over a local Gaussian kernel [117]. As a result, the parcellation accuracy varies depending on two user-specified parameters. The first one is the size of the Gaussian kernel used to estimate the LNCC for locally ranking the propagated atlases. The second one is the number of top ranked atlases to include in the local label fusion. In this study, the parameters were optimised on the MRM NeAt atlas database after the labels have been separated into left and right hemisphere. The Gaussian kernel size varies from 1 to 6 voxels (incremental step of 0.5 voxel) and the number of atlases used ranges from 3 to 9. In total, 77 parameter combinations were calculated.

The parcellation accuracy is calculated in terms of Dice similarity coefficient (DSC) - the percent voxel overlap between two labels [124]. For each pair of parameters, the average DSC of every atlas across the entire database was calculated as an indicator of the structural parcellation performance. The average DSC is obtained in the following steps (known as a "leave-one-out cross-validation"). Firstly, each of the 10 images was regarded as an unlabelled test image, and the remaining 9 were used as the atlases. The structural labels in the atlases were propagated to the query image with multi-atlas label propagation and fusion scheme. Secondly, the DSCs between the automated and the manual structural parcellation were calculated for every image in the database. Finally, the averaged DSCs for all the images across all structures were calculated for each parameter combination. The combination that gave the highest average

DSC was selected as the optimal set of parameters.

It is worth noting that bias in the parameters might be introduced in the "leave-one-out cross-validation", a better way to optimisation the parameters have been proposed, called the "nested cross-validation" where each time a subset of the data were chosen as training data, and the rest were regarded as testing data to measure the performance on the training data. It would be recommended to use "nested cross-validation" in the step when the atlas database becomes larger.

3.2.4 Performance evaluation

In order to evaluate the performance of the multi-atlas label fusion part of the proposed framework, it was compared with a single-atlas-based label propagation method as well as with the commonly used multi-atlas label fusion method STAPLE. A leave-one-out cross-validation similar to that described in Section 3.2.3 was performed for both methods. For the single-atlas method, the labels were propagated from each of the 9 atlases and their resulting DSC were averaged for each structure.

3.2.5 Application to unseen images

MR images collected from different sites and studies may vary due to various factors such as differences in scanner/coil specifications and scanning sequences. As a result, in addition to the cross-validation within the same atlas database, I tested the performance of the proposed framework to parcellate images collected in another site different from that of the atlas database, to further evaluate its performance in the situation of a real application.

To quantitatively evaluate the performance of my multi-atlas framework when applied to a new dataset, a corresponding expert-delineated manual structural parcellation is required to be considered as the gold standard for the new dataset. Here I use the atlases in the NUS atlas database [79] as unlabelled test images. The corresponding manual structural labels were separated into left/right hemisphere as I did for the MRM NeAt atlas database (Section 3.2.2), and 40 structural labels were generated for each hemisphere.

I selected and grouped 24 structural labels (12 in each hemisphere), which were presented in the manual structural parcellation in both atlas databases to ensure a one-to-one structural correspondence between all atlases (Table 3.1). However, the inter-

Table 3.1: Structural label correspondence between MRM NeAt atlas and NUS atlas.

NeAT	**NUS**
Neocortex	Frontal Cortex
	Perirhinal cortex
	Entorhinal cortex
	Motor cortex
	Somatosensory cortex
	Autitory cortex
	Visual cortex
	Remaining cortical region
Hippocampus	CA1 region
	CA3 region
	Dentate Gyrus
	Remaining hippocampal region
Amygdala	Amygdala
Olfactory bulbs	Olfactory bulbs
Thalamus	Thalamus
Hypothalamus	Hypothalamus
Superior colliculi	Superior+inferior colliculi
Inferior colliculi	
Cerebellum	Cerebellar cortex
	General cerebellum
Brainstem	Pons
	Medulla
Internal capsule	Internal capsule
External capsule	External capsule
Anterior commissure	Anterior commissure
Ventricles	Lateral ventricle
	Forth ventricle

rater variability still needs to be taken into account. This is due to the differences in the manual structural parcellation protocols between the two databases and subsequent accuracy of the quantitative analysis [196]. For the two atlas databases I used, the manual structural parcellations were both following the Franklin-Paxinos atlas [191]. Nevertheless, giving the fact that there is no knowledge about the inter-rater variability between these two datasets, there is still a source of variability in the experiment. This limitation is discussed in more detail in section 4.4.

Similar to the procedure described in the performance evaluation step (Section 3.2.4), the structural labels from the MRM NeAt database were propagated to each of the query images from the NUS database using my multi-atlas framework, as well as the STAPLE algorithm and single-atlas label propagation method. I use the pa-

rameters previously obtained (Section 3.2.3) to better represent a real situation, where no manual labels are available as ground truth to optimise the parameters. Finally, for each of the three approaches, I calculated the DSC between the automatic parcellation results and the manual labels in the NUS database, and compared the results.

3.2.6 Application to the groupwise analysis

One of the main applications of structural parcellation is to detect and quantify volumetric changes in brain structures of different animal groups, which vary in terms of pathology or genetic background. To test the ability of the proposed framework to detect such statistical differences, a previously published data set of the Tc1 mouse model of Down's Syndrome was used [197]. *Ex vivo* mouse brains of 28 animals, 14 wild type and 14 transchromosomic, were selected and structurally parcellated. The MRM NeAt database also includes 10 *ex vivo* atlas images, which were manually parcellated into the same 20 structures [78]. I thus use these *ex vivo* atlases, again with structural labels separated into left and right hemisphere. Similarly to the *in vivo* database, I used a leave-one-out cross-validation strategy to obtain an optimised combination of parameters. I then applied the proposed framework to parcellate the structures of all 28 animals. Both the proposed framework and the single-atlas method were used to detect volume differences in all structures between the wild-type group and the transchromosomic group.

The absolute structural volumes differences are affected by the natural variation of brain size among individual subjects, especially for cross-sectional *ex vivo* studies [138]. This effect can be corrected through normalising the structural volume to the total brain volume (TBV) that improves the specificity of the statistical test [198]. Such normalisation procedure sometimes however, reduces the sensitivity to detect subtle structural volume changes [199]. As a result, the obtained structural volumes of the two groups were compared both with and without total brain volume (TBV) normalisation.

3.2.7 Mirroring process for doubling the atlas number

It has been shown that label fusion algorithms benefit from an increase in the number of atlases, as the accuracy increases with sample size [109]. The brain has structural layout across two hemispheres. Studies have shown that by including the flipped mirror

images of the atlas to double the database size, the structural parcellation result can be improved [117, 114, 102, 200]. Given the limited number of available atlas in the database, I hypotheses that the local atlas selection algorithm LNCC can benefit from the atlas doubling, and the final parcellation accuracy (in terms of the DSC) can be increased. To test this hypothesis, the database size was double by including a left/right flipped version of all the atlases following the previous studies. It is worth noticing that, by including the flipped version, the atlases are no longer necessarily independent, as there are potential correlations between the same structures in two hemisphere of the brain.

3.3 Results

The result of parameter optimisation for the STEPS algorithm in the proposed framework is presented in Section 3.3.1. The performance is evaluated using leave-one-out cross-validation for the *in vivo* atlas database MRM NeAt. Section 3.3.2 compares the parcellation accuracy obtained from my pipeline with that obtained from a single-atlas label propagation method, and with the STAPLE algorithm. In Section 3.3.3, the mouse brain MRI data from another atlas database, the NUS mouse atlas, is used as test images to validate the ability of the proposed multi-atlas framework to parcellate unlabelled new data from different site. In Section 3.3.4, the proposed framework is applied to MRI data from two groups of mice with different genetic background to evaluate the ability to detect volumetric differences between groups. Finally, the result of the mirroring process is presented in Section 3.3.5. The framework source code and a sample atlas database can be downloaded online[2].

3.3.1 Parameter optimisation

Optimal parameters were obtained for the MRM NeAt atlas database. I optimised the values of two parameters in the local ranking system LNCC of the STEPS algorithm: the size of the Gaussian kernel for image comparison and the number of top ranked labels to include in the label fusion. Between the two parameters, the number of top-ranked atlas selected for label fusion appears to have a dominant effect on the performance. Fig. 3.3 shows the DSC value obtained for the different parameter com-

[2]http://cmic.cs.ucl.ac.uk/staff/da_ma/multi_atlas/

binations. The parcellation accuracy was estimated as the average DSC across all the structures, varying from 0.79 to 0.83. A number of selected atlases of 8 and a Gaussian kernel size of 5 voxels returned the highest average DSC. The number of selected atlas almost reaches the maximum (8 out of 9, as shown in Figure 3.3), which indicates that the atlases number in the training data has not yet reached the optimal. Therefore, including the flipped version of the atlases could potentially addressed such limitation, effectively doubling the database size.

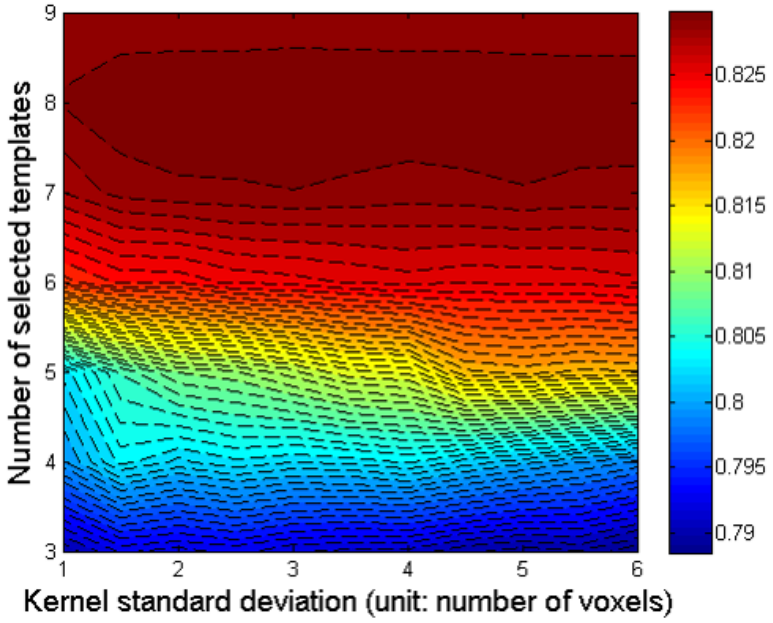

Figure 3.3: Parameter optimisation for atlas database with left/right hemisphere separated. The colour map represent the average DSC over all structures with different parameter sets. The y-axis represent different number (from 3 to 9) of top-ranked atlases selected for label fusion. The x-axis represent different Gaussian kernel size in the LNCC image similarity measurement (from 1 to 6 with 0.5 step increments). The optimal parameter combination was obtained when the number of atlases for label fusion = 8 and Gaussian kernel standard deviation = 5 voxels, with a corresponding average DSC of 0. 84. The small variation along the x-axis indicates the limited effect of the Gaussian kernel size towards the overall parcellation accuracy.

Sample images of the cross-validation of the proposed framework on the original atlas database as well as the atlas with left-right hemisphere separation are shown in Fig. 3.4. This combination of parameters was used to assess the differences between the proposed framework and two other approaches, as reported in the following section.

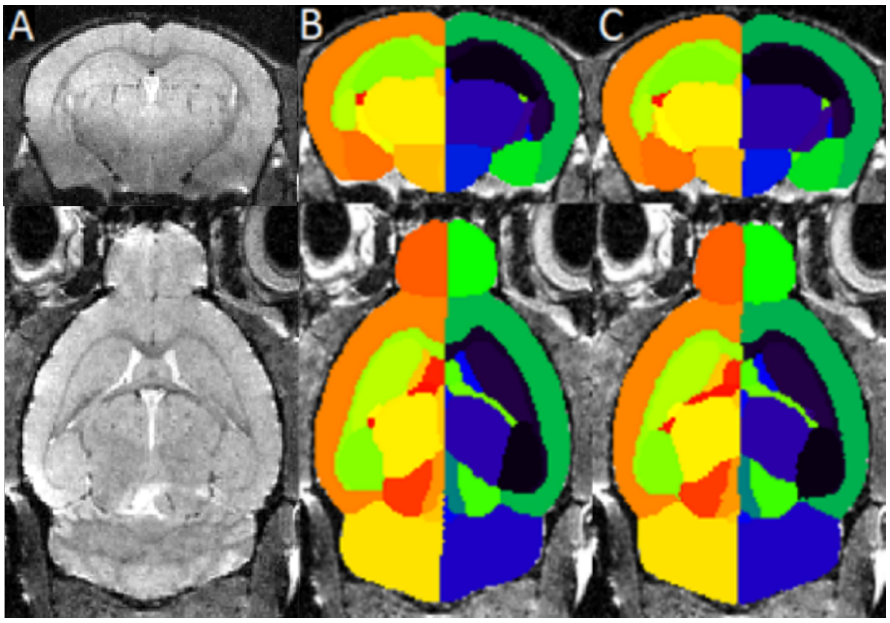

Figure 3.4: Sample images from the cross-validation result of the pipeline on the atlas databases. Parcellation results obtained with the proposed method and parameters. (A) The original MR image from the atlas (B) The MR image from the atlas overlaid with corresponding manually labelled anatomical structures which is considered as gold standard. (C) The same MR images overlaid with the structural parcellation result after applying my multi-atlas framework. Top row: coronal view, bottom row: axial view.

3.3.2 Statistical comparisons

I compared the parcellation accuracy of the proposed framework with the single-atlas label propagation and STAPLE label fusion methods using a leave-one-out cross-validation on the MRM NeAt atlas database. For each atlas image, I averaged the DSC of all the propagated atlases. Both approaches were compared with the STEPS algorithm using an optimised parameter combination (Fig. 3.5). A one-way analysis of variance (ANOVA) with Dunnett method to correct for multiple group comparison was performed on each of the 40 structures, and multiple comparisons across all the structures was corrected with False Discovery Rate (FDR) set to q=0.05. Compared to the single-atlas method, the proposed multi-atlas framework achieved significantly higher parcellation accuracy for most structures except some big structures (the brain stem, left hypothalamus, hippocampus, right caudate putamen, cerebellum, neocortex and thalamus) in which case the single-atlas method already performs well. Compared to the STAPLE algorithm, significantly higher parcellation accuracies are achieved in the left/right anterior commissure, left central gray and the remaining left/right mid-

brain and left superior colliculi. It can be observed that the improvement from the STEPS algorithm is more significant for small structures. This is because the DSC is a percentage measurement which is relative to the structural size.

3.3.3 Application to unseen images

In order to evaluate the ability of the proposed multi-atlas framework to parcellate new data, I used the mouse brain MR images in the NUS atlas database as unlabelled test images. I propagated the structural labels in the MRM NeAt atlas database to the MR images in the NUS atlas database with the optimised parameter combination obtained previously for the MRM NeAt atlas database. Fig. 3.6 shows the sample images of the test MRI data overlaid with corresponding manual labels as well as the automatic structural parcellation after applying my multi-atlas framework. The 24 manually delineated structural labels (12 in each hemisphere) that were present in both atlas databases were selected and grouped. Fig. 3.7 shows the statistical comparison of the resulting DSC derived from my multi-atlas framework as well as that from the STAPLE algorithm and the single-atlas label propagation. One should note that the Dice similarity may be of limited use due to the intrinsic variability between the manual labeling protocols in the two atlas databases. The DSC obtained here should neither be compared with that derived in the parameter optimisation part nor to the results from Bai *et al.*'s study [79], as the comparison here is between two manual segmentations of two different image dataset acquired at two sites. A one-way ANOVA with the Dunnett method to correct for multiple group comparison was performed on each of the 24 structures, and multiple comparisons across all the structures was corrected with FDR (q=0.05). When compared to the single-atlas method, the proposed multi-atlas framework achieved significantly higher parcellation accuracy to parcellate left/right external capsule, left/right internal capsule, left/right amygdala and right anterior commissure. Compared to the STAPLE algorithm, significantly higher differences are achieved in the external capsule, anterior commissure and cerebellum for both hemispheres, and amygdala for right hemisphere. Similar to the result in Section 3.3.2, the improvement from the STEPS algorithm is more significant for small structures.

Interestingly, the DSC of STAPLE is lower than that of the single atlas method when parcellating the anterior commissure. It could be due to the fact that the STAPLE

algorithm assumes that the parcellation errors, for each individual candidate parcellation, are due to random human rater error. However, a large portion of the parcellation errors here are due to image registration. The anterior commissure is a small structure, resulting in a relatively low impact on the registration algorithm compared to the contrast from surrounding tissues. On the other hand, the STEPS algorithm reduced the registration-derived parcellation error by taking into account the local image similarity in the atlas selection [117].

3.3.4 Application to groupwise analysis

Fig. 3.8 shows sample images of parcellation results using the proposed framework (Fig. 3.8b) as well as using the single-atlas-based method with misregistrations occurring in some regions (Fig. 3.8c, d). The STEPS label fusion algorithm successfully obtained the correct labels at both regions whereas both single-atlas-based methods produce inaccurate labels (shown by the red arrow). Statistical analysis between Tc1 Down's Syndrome and wild type mouse was performed on the volumetric data both with and without TBV normalisation. A two-tailed paired t-test was performed on each of the 40 structures. Multiple comparisons were corrected with a false discovery rate q=0.05 (Fig. 3.9). I compared the statistical result of the proposed framework with the result of the single-atlas-based method for each of the 9 atlases in the database (Fig. 3.2 and Table. 3.2).

For the unnormalised volumetric data, the proposed framework detected significant volume increase in the transchromosomic group in hippocampus, caudate putamen, thalamus cerebellum neocortex and rest of the midbrain in both left and right hemispheres. Conversely, in the single-atlas method, five out of nine atlases (A1 A4 A5 A6 A8) failed to detect all significant volume increases as shown in the proposed framework, one (A3) showed the same significant result, and one (A9) showed a significant result on the olfactory bulb which is neither detected by the proposed framework nor the Tensor-Based Morphometry (TBM) analysis in the original study[197]. This was possibly due to a larger variance in the single-atlas-based method. On the other hand, two atlases (A2 A7) showed a significant increase in external capsule which was not picked up by the proposed framework, although they failed to detect significant differences either for the thalamus on the left hemisphere (A2) or for the rest of the

midbrain on the right hemisphere (A7).

For the volume normalised by the TBV, the proposed framework detected significant volume shrinkage in the cerebellum and olfactory bulb that coincides with the TBM results reported by the original study, while all the single-atlas-based methods detected less or no significant volume differences, suggesting less statistical power. It is worth noting that for the above mentioned single-atlas-based method on A2 and A7, which revealed additional statistical group differences on external capsule in the unnormalised data, no statistical differences were observed on any of the structures after TBV normalisation.

Overall, the proposed framework obtained better statistical power to detect structural group volume differences when compared with the single-atlas method. Nevertheless, some of the structural volume differences detected by the TBM analysis in the original study, such as superior colliculus and hypothalamus in the unnormalised data, and external capsule (posterior part of the corpus callosum) for the normalised data, were not captured using the proposed framework. This is possibly due to the voxelwise nature of TBM techniques, which can detect very local changes, as opposed to the proposed technique, which can only detect changes in regional volume. Furthermore, an accurate structural parcellation is not only important for regional volume analysis, but also for further quantitative analyses such as thickness or shape analysis.

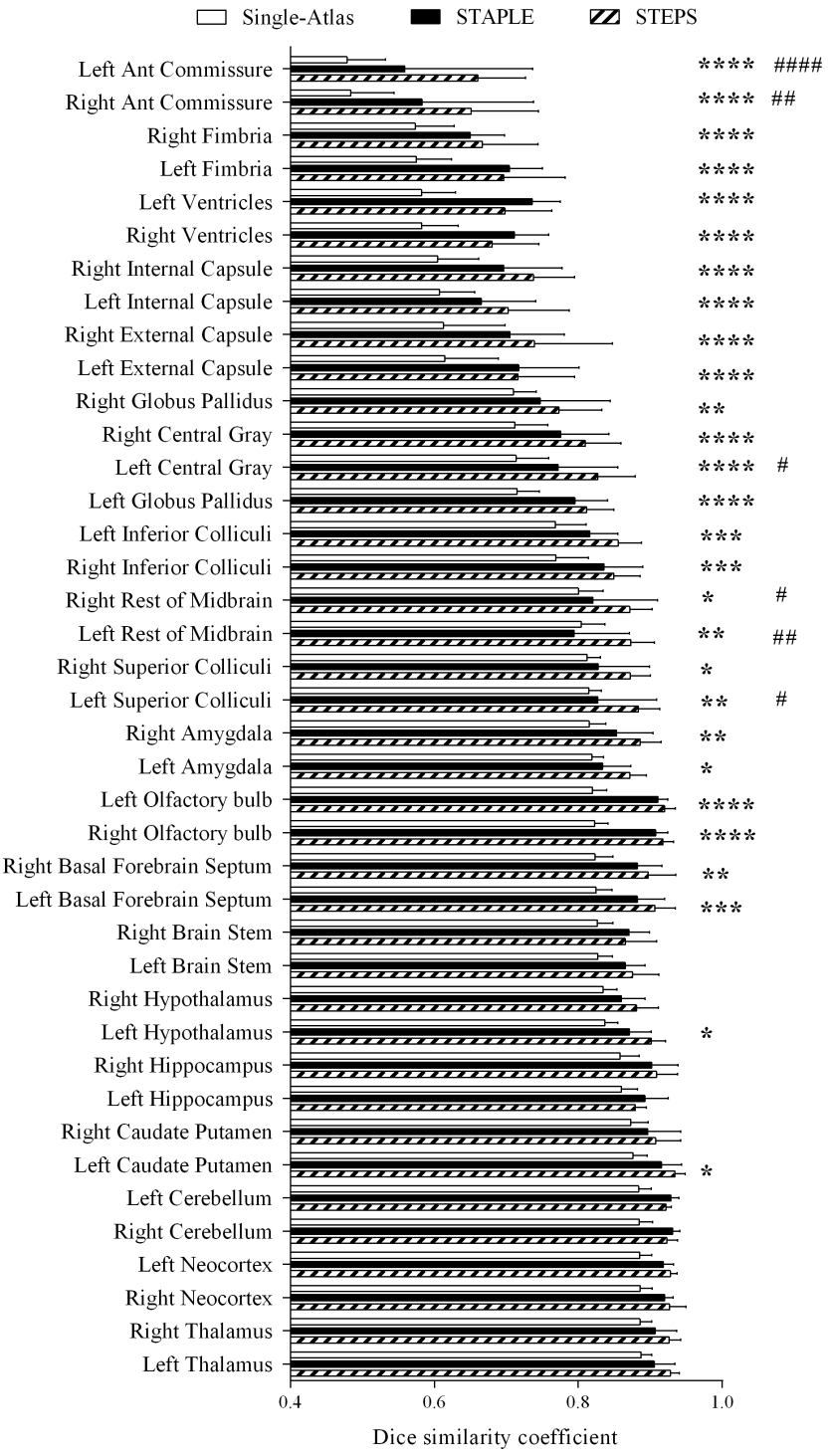

Figure 3.5: cross-validation result on the *in vivo* mouse brain atlas MRM NeAt [3]. Comparison of the average DSC using the proposed framework, a single-atlas label propagation method and the STAPLE algorithm. ANOVA were performed, with multiple comparisons of 40 structures corrected with false discovery rate set to 5%. Error bars represent the standard deviations. (For single-atlas vs. STEPS, *: $0.01<p\leq0.05$; **: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$; ****: $p\leq0.0001$; for STAPLE vs. STEPS #: $0.01<p\leq0.05$; ##: $0.001<p\leq0.01$; ###: $0.0001<p\leq0.001$; ####: $p\leq0.0001$.)

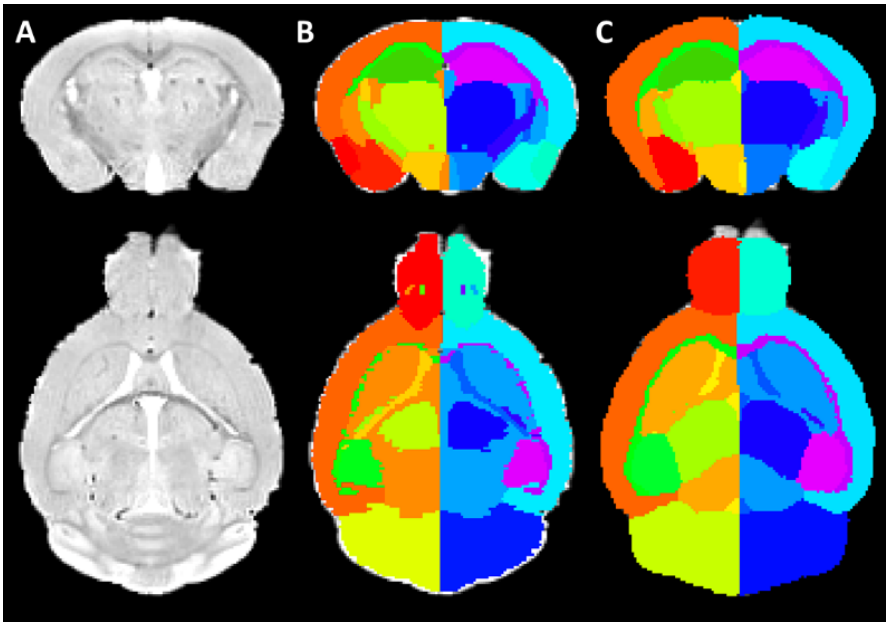

Figure 3.6: The structural parcellation result of applying my multi-atlas framework to a new dataset. (A) The MR image from an NUS mouse atlas which is treated as a new dataset. (B) MR image of the query image overlaid with corresponding manually labelled anatomical structures considered as gold standard. (C) The same MR images overlaid with the structural parcellation result after applying my multi-atlas framework. Top row: coronal view, bottom row: axial view. The discontinuity shown in axial view of the manual segmentation in (B) is due to the fact that the manual segmentations are performed in the other view. More discussion about the lack of manual segmentation protocal for preclinical data comparing to the clinical data are elaborated in session 3.6.

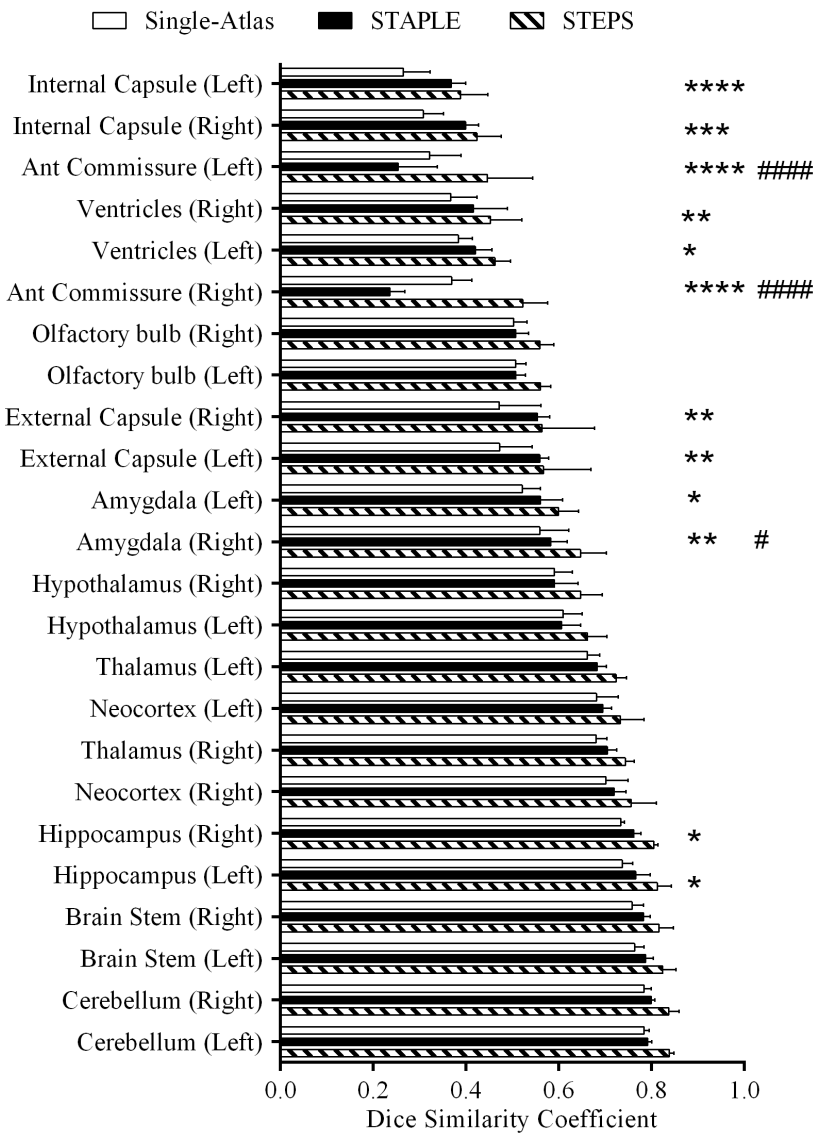

Figure 3.7: Validation on the ability of the multi-atlas framework to parcellate structures of the new dataset. The new dataset is taken from the NUS mouse atlas[79] with the corresponding manual labels regarded as gold standard. 12 manually parcellated structural labels were included in the comparison which appeared in both of the two atlas databases. Previously obtained optimised parameter combination for the MRM NeAt atlas database were used to calculate the DSC. ANOVA were performed, with multiple comparisons of 40 structures corrected with false discovery rate set to 5%. Error bars represent the standard deviations. (For single-atlas vs. STEPS, *: $0.01<p\leq0.05$; **: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$; ****: $p\leq0.0001$; for STAPLE vs. STEPS #: $0.01<p\leq0.05$; ##: $0.001<p\leq0.01$; ###: $0.0001<p\leq0.001$; ####: $p\leq0.0001$.)

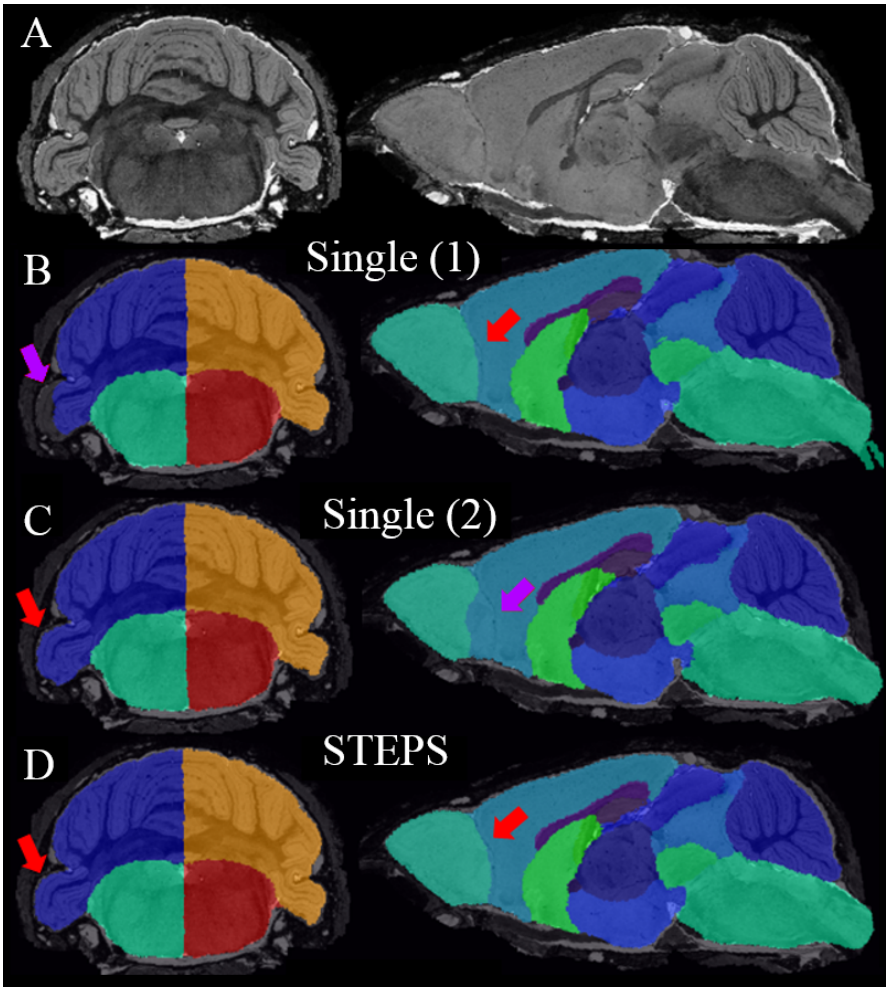

Figure 3.8: Sample images comparing the parcellation result of the framework and the single-atlas-based methods. The selected slices demonstrated that despite some local misalignments in the single-atlas-based method (as shown in red arrows). The STEPS label fusion algorithm in the framework successfully preserved the correct local registration in different regions. Structural parcellations are overlaid on the original image (in both coronal and sagittal view, A). (B) Structural parcellation result of a single-atlas-based method with part of the cerebellum misclassified. (C) Another structural parcellation result of single-atlas-based method with the edge between olfactory bulb and cortex misclassified. (D) Structural parcellation using the proposed framework.

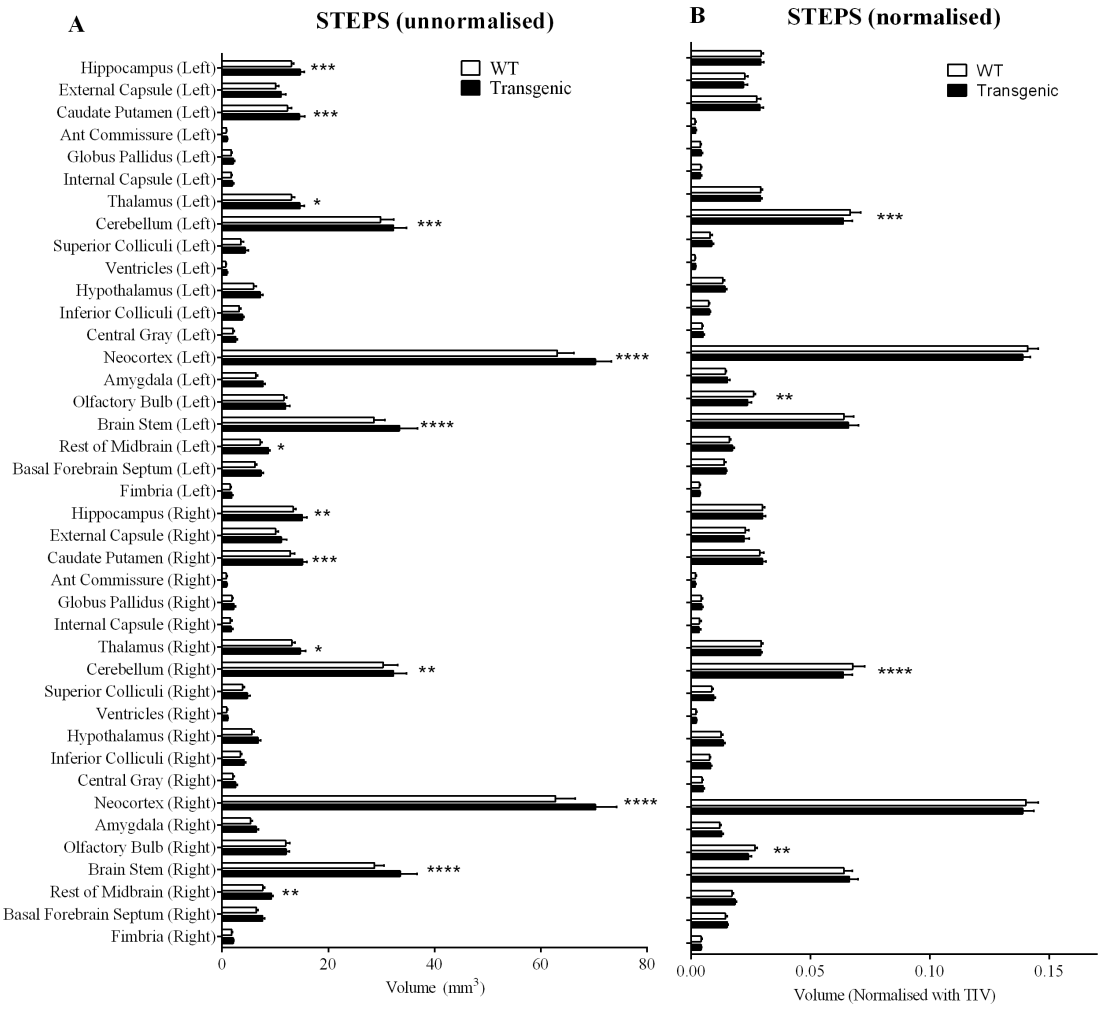

Figure 3.9: Statistical comparison of the structural volume difference between groups of Tc1 Down's Syndrome mouse and wild type. A) Volumetric comparison of the unnormalised data; B) To eliminate the fact of the head size variation, the structural volumes were also normalised by total intracranial volume. Two-tailed paired t-tests were performed to compare the structural volumes of the wild type and the transchromosomic group for each of the 40 structures. Multiple comparisons are corrected with false discovery rate q=0.05. Error bars represent the standard deviations. (*: $0.01<p\leq0.05$; **: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$; ****: $p\leq0.0001$)

Table 3.2: Statistical significant result of the volumetric comparison between the Tc1 groups and wild type. * represents a significant volume different between groups. The STEPS algorithm consistently preserves the statistical significance while for the single-atlas-based method(A1-A7), some atlas performs worse than others, which becomes more severe after normalisation.

	Before normalisation										After normalisation									
Structure	STEPS	A1	A2	A3	A4	A5	A6	A7	A8	A9	STEPS	A1	A2	A3	A4	A5	A6	A7	A8	A9
Hippocampus (Left)	*		*	*				*		*										
External Capsule (Left)			*					*												
Caudate Putamen (Left)	*		*	*			*	*	*	*										
Ant Commissure (Left)																				
Globus Pallidus (Left)																				
Internal Capsule (Left)																				
Thalamus (Left)	*			*				*												
Cerebellum (Left)	*		*	*	*			*	*	*	*									
Superior Colliculi (Left)																				
Ventricles (Left)																				
Hypothalamus (Left)																				
Inferior Colliculi (Left)																				
Central Gray (Left)																				
Neocortex (Left)	*	*	*	*	*	*	*	*	*	*				*						*
Amygdala (Left)																				
Olfactory Bulb (Left)										*	*	*								
Brain Stem (Left)	*			*				*												
Rest of Midbrain (Left)																				
Basal Forebrain Septum (Left)																				
Fimbria (Left)	*		*	*				*												
Hippocampus (Right)			*																	
External Capsule (Right)	*		*	*	*		*	*	*	*										
Caudate Putamen (Right)																				
Ant Commissure (Right)																				
Globus Pallidus (Right)																				
Internal Capsule (Right)	*			*				*		*										
Thalamus (Right)	*		*	*				*	*	*	*	*		*	*					*
Cerebellum (Right)																				
Superior Colliculi (Right)																				
Ventricles (Right)																				
Hypothalamus (Right)																				
Inferior Colliculi (Right)																				
Central Gray (Right)	*	*	*	*	*	*	*	*	*	*										*
Neocortex (Right)																				
Amygdala (Right)										*	*	*			*					
Olfactory Bulb (Right)	*		*	*																
Brain Stem (Right)																				
Rest of Midbrain (Right)																				

3.3.5 Mirroring process for doubling the atlas number

I investigated the parcellation accuracy of my multi-atlas framework by applying it to the original left/right hemisphere separated atlas database as well as to the database including the mirrored atlases, with the optimal parameter combination for each database. The parcellation accuracy was presented in terms of DSC for each structure (Fig. 3.10). A two-tailed paired t-test was performed on each of the 40 structures. multiple comparisons across all the structures is corrected with False Discovery Rate (FDR) set to q=0.05.

Although there is a tendency of increase in the DSC, no significant differences remain after correcting for multiple comparisons (Fig. 3.10). This result shows that the improvement related to the inclusion of the flipped version of the atlases is not obvious, probably due to the small number of available atlas in the database.

3.4 Open source efforts

This section lists the essential tools developed within the multi-atlas label fusion framework. The intention is to make the code open source, in order to make the research output sustainable and transferable.

The tools in this package has been made available online[3], and publicly available on *github*[4]. The code is built upon two software packages in the NifTK platform - NiftyReg[5] and NiftySeg[6] - which are also open source softwares. The code has been distributed in two versions. Version one is for running on a single workstation; version two is for high-throughput parallel computing on the cluster (deployed with Oracle® grid engine). The codes have been tested on both Unix/Linux-based environment and Windows-based environment. Following is the list of the essential commands in the open source tool package, along with their corresponding function and usage.

"mask" : Performing the initial rough brain extraction based on affine registration and STAPLE label fusion algorithm.

- Input parameter 1: Name of the query image file

- Input parameter 2: Name of the atlas database folder

[3] http://cmic.cs.ucl.ac.uk/staff/da_ma/Multi_Atlas/

[4] https://github.com/dancebean/multi-atlas-segmentation

[5] http://sourceforge.net/projects/niftyreg/

[6] http://sourceforge.net/projects/niftyseg/

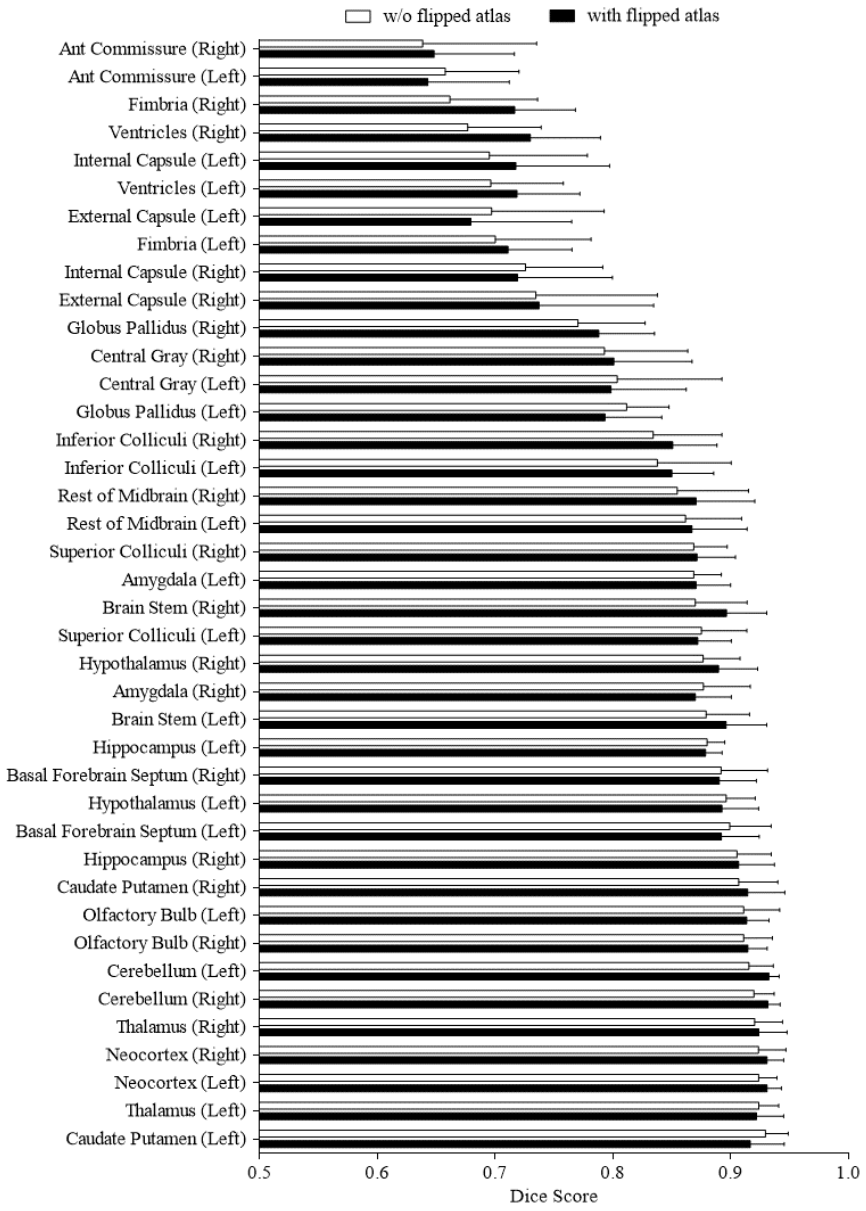

Figure 3.10: Comparison of atlas database with or without the flipped atlases. DSC of each structures was obtained by applying the framework to the database with only left/right hemisphere separated atlases and the database with mirrored atlases, using the optimal parameter combination for each database. Two-tailed paired t-tests were performed, with multiple comparisons of 40 structures corrected with false discovery rate set to 5%, Error bars represent the standard deviations.

- Input parameter 3 (Optional): Name of the file containing user-defined non-default parameter

- Output: the resulted mask file will be saved in the folder named as “mask”.

“parcellation” : Performing non-rigid image registration and label fusion.

- Input parameter 1: Name of the query image file

- Input parameter 2: Name of the brain mask for the query image file (can be manually delineated mask or result from the "mask" command). - Input parameter 3: Name of the atlas database folder

- Input parameter 4 (Optional): Name of the file containing user-defined non-default parameter

- Output: the resulted mask file will be saved in the folder named as "label" under the current directory.

"STEPS_optimisation" : Performing the leave-one-out validation to parameter optimisation for the atlas database of choice based on the dice similarity measurement between the manual label ground truth and the parcellated labels from the "paracellation" command. - Input parameter 1: Name of the atlas database folder

- Output: the result table of dice score will be saved in the folder named "Dice_Score" under the current directory.

3.5 Sample applications of the framework

The multi-atlas-based structural parcellation framework for mouse brain MRI has been successfully applied to various studies, either in disease progression or drug treatment.

3.5.1 Application to the longitudinal accessment of disease progression and potential treatment

In the first example, the structural parcellation framework is applied in a study to determine the structural volumes as a morphological indicator of disease progression and potential drug treatment. The image data were acquired and provided by my colleagues Jack Wells and Holly Holmes at Centre for Advanced Biomedical Imaging (CABI). Three mouse groups were included in the study: 1) wild type mouse, 2) transgenic mouse (rTg4510), which overexpresses the Tau protein and under no treatment, and 3) rTg4510 mouse with continuous oral treatment of doxycycline. Each animal was scanned at 3 time points: 3.5 months, 5.5 months and 7.5 months after birth. The structural parcellation result successfully captured the progressive volume shrinkage in both the region with high Tau burden (cortex and hippocampus), as well as the reduced volume shrinkage in the doxycycline treated group (Fig. 3.11).

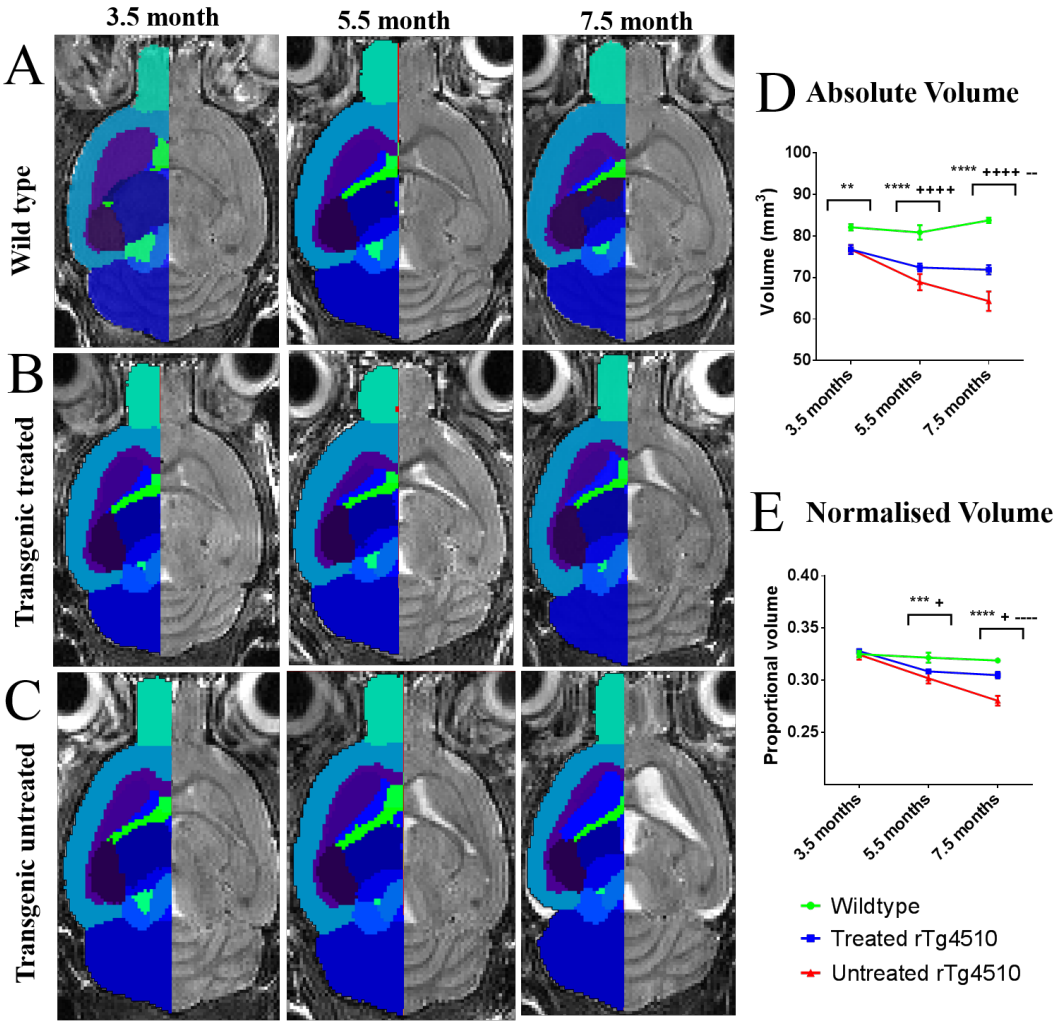

Figure 3.11: (Left) Sample images of each group and its corresponding parcellated structural labels at each time point for (A) wild type group, (B) transgenic group with treatment, and (C) transgenic group without treatment. (Right) The longitudinal volume change of the region including hippocampus and cortex, both (D) with and (E) without normalisation to the TBV). Error bars represent the standard error of the mean. Wildtype vs. untreated rTg4510s: **: $0.001<p\leq0.01$; ***: $0.0001<pzleq0.001$; ****: $p\leq0.0001$. Wildtype vs. treated rTg4510s: + = $0.01<p\leq0.05$; ++++ = $p\leq0.0001$. Treated rTg4510s vs. untreated rTg4510s: – = $0.001<p\leq0.01$; —- = $p\leq0.0001$.

3.5.2 Application to rat hippocampus segmentation to access effect of treatment

The other example is a study conducted by my colleague Duffy *et al.* [201] to evaluate the effect of dexamethasone on the cerebral edema and brain injury from lithium-pilocarpine induced status epilepticus. The structural parcellation framework was used to measure the total brain volume, as well as the hippocampus volume using rat. Four animal groups were included in the study: 1) control group, 2) group with sta-

tus epilepticus (SE); 3) SE rat receiving 10 mg/kg dexamethasone sodium phosphate (SE-DEX10) and 4) SE rat receiving 2 mg/kg dexamethasone sodium phosphate (SE-DEX2). The results showed that the group with dexamethasone treatment are associated brain volume shrinkage and hippocampus volume reduction (Fig. 3.12), which is in-line with other findings in the study. This lead to the conclusion that the dexamethasone actually exacerbates the cerebral edema and brain injury in the case of lithium-pilocarpine induced.

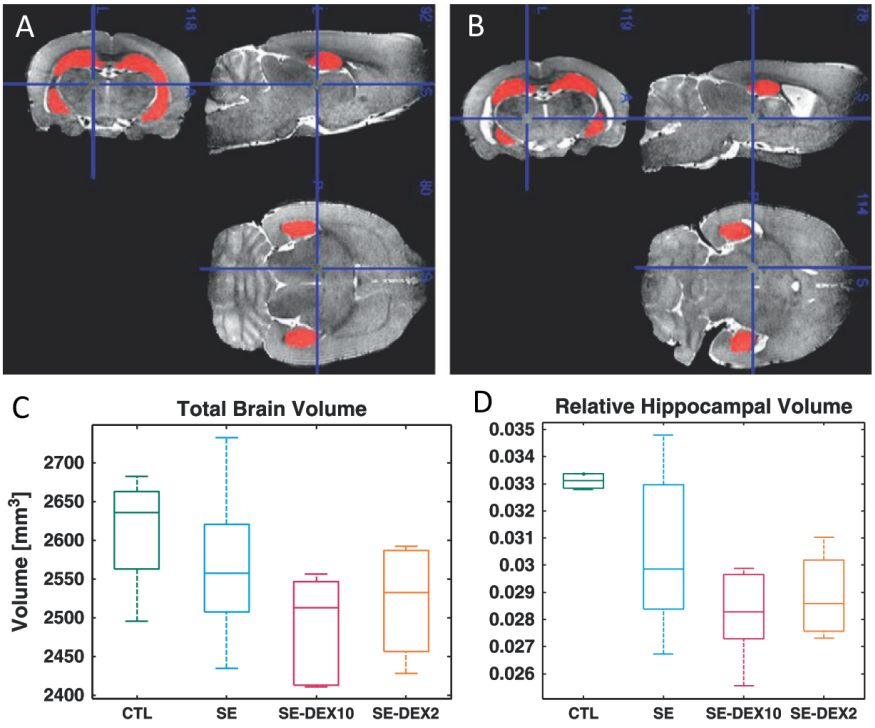

Figure 3.12: Application of the framework on rat hippocampus segmentation in study drug treatment. Top Row: representative images of automated parcellation of the hippocampus in a (A) control rat and (B) post-status epilepticus rat. Bottom row: quantitative result of the structural pipeline including the (C) total brain volume and (D) hippocampal volume of control groups (green) and different treatment (blue, red and orange). Figure taken from [201].

The framework has also been used to compare the structural volumetric analysis between *in vivo* and *ex vivo* imaging, and will be discussed in more detail in Chapter 4.

3.6 Discussion

Previous studies have shown successful applications of such multi-atlas label propagation and fusion techniques in a clinical context, to detect volumetric variation of brain structures such as ventricles and hippocampi [7, 114]. In this study, the results denote

that with a relatively limited number of available atlases in the database when compared with clinical studies, the structural parcellation of pre-clinical data can be significantly improved when compared to previous approaches [97, 98]. I have also demonstrated the ability of the proposed framework to use existing atlas databases to parcellate images acquired at different sites. I also tested its ability to detect brain structure volumetric changes in genetically modified pathological animal model.

In order to assess the pre-clinical relevance of the proposed framework, further work could include statistical power analysis. For example, would the improved method be able to detect the same amount of change by using fewer samples, or what level of subtle change can be detected using the new method with the same amount of data? van Eede *et al.* [202] recently proposed a method to generate an artificial deformation field to test registration sensitivity, which could also be used to generate simulated volume changes.

3.6.1 Parameter optimisation related issues

The optimised STEPS parameters were chosen based on the average DSC over all the structures and across all samples in the atlas database. However, one should note that the DSC is intrinsically biased towards large structures (e.g. hippocampus and neocortex), while small structures (e.g. external capsule, anterior commissure) are more sensitive to local registration errors and inter-atlas morphological variation. However, as shown in Fig. 3.3B & C, studies interested in parcellating only certain structures can obtain the optimised parameter combinations following this framework by considering only the DSC for the specific structures of interest.

As demonstrated in section 2.5 and 3.3, when dealing with new data, there is likely to be no manually delineated labels associated with the data. As a result, it is impossible to further improve the parameters for the pipeline. Moreover, as shown in Fig. 3.3, the parameter optimisations reach a plateau around the optimal value. This indicates that the proposed framework is robust to variation in this region of the parameter space. As a result, small deviations of the parameter values have a small impact on the overall parcellation performance. When applying the pipeline to a set of unseen data from another study, the parcellation error derived from image registration would increase. Considering the two parameters of interest, the “Kernel standard deviation”

of LNCC is directly related to the image registration error, while the number of selected atlas is more related to the statistical power of the label fusion. It is shown in Fig. 3.3 that the "Kernel standard deviation" has less effect in the parcellation accuracy when compared to the number of selected atlases. As a result, the optimised parameters obtained through the leave-one-out cross-validation are a reasonable approximation when parcellating the new data.

3.6.2 Image registration related issues

Most image similarity measurements used in registration algorithms are governed by high contrast edges. The registration accuracy in regions with low contrast is limited. For the neighbouring anatomical regions that lack contrast in between, the registration algorithm will have to rely on the regularisation term rather than on image features for accurate matching [203, 86]. This can lead to a decrease in parcellation performance. In addition, the atlases used for the proposed multi-atlas framework are limited in number and are T2* weighted, which might impede their direct application to images acquired with different contrast. However, the NMI used in this framework for image similarity measurement has been shown to be less dependent on image contrast, and is currently commonly used to compare image similarity between multi-modal images [204].

3.6.3 Current limitations in mouse brain studies

Compared to human brain MRI structural parcellation studies [109, 117, 114], the availability of mouse brain atlas databases is limited, and subsequently, the performance of label fusion techniques is reduced. To the best of my knowledge, there are currently only two *in vivo* mouse multi-atlas databases that are publicly available [79, 3]. The amount of available databases, as well as the number of atlases in each database, is far from ideal. Although Chakravarty *et al.* [205] improved the parcellation accuracy by introducing an artificial intermediary multi-atlas database from a single-atlas, it does not address the problem of insufficient data and morphometric variability.

Previous studies have shown improved structural parcellation result by including the flipped mirror images of the atlas to double the database size [117, 114, 102, 200]. However, when testing on the MRM NeAt atlas database, the improvement of such process is limited (Fig. 3.10). This might be due to the small number of atlases available in the database, which reduces the chance to get better local morphological match

from the flipped images for the label fusion algorithm to benefit from, while additional registration error is introduced in other regions due to the brain asymmetry. Furthermore, MRI has been used to measure the asymmetry of adult mouse brains [206], and a recent study using optogenetics conducted by Michael *et al.* [207] also showed differences between the left and right hippocampal plasticity. Given such lateralisation of the brain, further validation is still necessary to assess the anatomical viability of such flipping process.

Within the field of clinical research, there are well documented and standardised protocols for manual labelling of brain structures [208, 209, 210], but equivalent information lacks in the case of mouse brain MRI study. The unclear nature of the anatomical standardisation and vague definitions of the structure labeling protocol also reduces consistency between human raters. Furthermore, manual labeling is considered as the gold standard to evaluate parcellation accuracy, and is used for comparison to assess the performance of automatic parcellation methods. This makes the intra- and inter-rater labelling variability crucial as it represents the theoretical upper bound of any automatic method performance. Such variability has not been fully assessed in mice, which makes it difficult to determine the potential improvement that an algorithm can achieve. Most of the available publications about mouse brain MRI atlas construction, either *in vivo* or *ex vivo*, single-atlas-based or multi-atlas-based, lack clear guidance about the protocol for manual labeling of structures [73, 78, 3, 91]. However, efforts are being made to address this. Bai *et al.* [79] included a detailed protocol for manual labelling of every structure that is parcellated in the *in vivo* atlas they released (in supplementary material). More recently, Ullmann *et al.* [80, 81, 82] and Richards *et al.* [47] from the Australian Mouse Brain Mapping Consortium described a detailed structure labeling protocol on the minimal deformation atlas for *ex vivo* MR images on the C57BL/6J mouse, which provided further information for parcellating sub-regions of the neocortex, hippocampus, cerebellum, and basal ganglia; this could be a safer guideline for future investigations. Those studies will eventually lead to a standardised consensus protocol for manual structural labelling of mouse brain MRI.

3.7 Conclusion

In conclusion, this chapter presents a fully automated multi-atlas framework for structural parcellation of mouse brain data. The proposed work includes pre-processing techniques such as brain extraction and intensity non-uniformity correction using N3, an efficient non-rigid registration algorithm, and a locally-weighted multi-atlas label fusion method - STEPS. Validation of the framework showed consistent and accurate parcellation performance compared to single-atlas-based method as well as global-weighted multi-atlas-based method, and has also been successfully applied to several different studies.

Chapter 4

Comparison of *in vivo* and *ex vivo* imaging biomarkers

4.1 Introduction

Brain structural volume is a widely accepted neuroimaging biomarker to study brain physiopathology. During the experimental design of the preclinical imaging studies, the choice has to be made whether to scan animals post-mortem or alive. While the former is usually superior in terms of image quality thanks to the prolonged scanning time, lack of motion artefact and sometimes contrast enhancement agent, the latter has the advantage of better representing the underlying true physiology without distortion from the post-mortem process, and the potential to follow longitudinal structural change. In order to understand how these differences in the image quality affect the sensitivity of structural parcellation, I compared the volumetric analysis of *in vivo* and *ex vivo* brain MRI both acquired from the same cohort of adult mice with the optimised acquisition protocols. I used a multi-atlas structural parcellation framework developed in Chapter 3 to determine whether volumetric analysis is able to detect the presence of structural changes from *in vivo* to *ex vivo* data due to animal death and fixation/perfusion processes, and how these differences affect the statistical analysis. The result demonstrated uneven distribution of volumetric changes from *in vivo* brains to *ex vivo* brains. In addition, the volumetric analysis of the *in vivo* and *ex vivo* data showed similar statistical power to differentiate between two animal groups.

4.2 Methods

4.2.1 Experimental data

I used image data acquired from 8 female FVB/NCrl wildtype mice. Images were acquired and provided by my colleagues Jack Wells and Holly Holmes at Centre for Advanced Biomedical Imaging (CABI). For the sake of completeness, imaging acquisition protocol and scanning parameters are described in detail here.

Imaging was performed with a 9.4T VNMRS horizontal bore scanner (Agilent Inc). Animals were firstly scanned *in vivo* at 7.5 months after birth. For the *in vivo* imaging, a 72 mm inner diameter birdcage volume coil (Rapid Biomedical) was used for FR transmission and a 4 channel array head coil (Rapid Biomedical) was used for signal receiving. T2 weighted images were obtained using a 3D fast spin-echo sequence. A 3D T2-weighted sequence was employed with parameters: FOV=19.2 mm $\times$ 16.8mm $\times$ 12.0 mm; resolution 150 μm $\times$ 150 μm $\times$ 150 μm; TR = 2500 ms, TEeff = 43 ms, ETL = 4, NSA = 1. (The *in vivo* image data have been published by Wells *et al.* [211]).

The animals were sacrificed immediately after the *in vivo* imaging. An active staining technique was employed to enhance the contrast for *ex vivo* imaging, by perfusing the animal and soaking the decapitated head in buffered formal saline and 8mM Magnevist solution for 9 weeks [21]. An imaging gradient set with a 60 mm inner diameter (SGRAD 115/60/HD/S, Agilent Technologies UK Ltd., Berkshire, UK) was used. A 35mm birdcage RF coil was used for RF transmission and signal detection. A custom-build three brain holder (developed by my colleague Yichao Yu at Centre for Advanced Biomedical Imaging (CABI)) was used to acquire high resolution *ex vivo* images of multiple brains simultaneously. A 3D spoiled-gradient echo sequence was implemented for structural imaging with parameters: FOV = 32 mm $\times$ 25 mm $\times$ 25 mm; resolution = 40 μm $\times$ 40 μm $\times$ 40 μm; TR = 17 ms; TE = 4.54 ms; flip angle = 51°; NSA = 6.

Since the *in vivo* and *ex vivo* images were scanned using different RF coils, the shift of the coil gradient was calibrated to eliminate the scaling effect [212]. To mimic the real case scenario, the acquisition protocols for *in vivo* and *ex vivo* imaging are designed and optimised separately to get the best possible images for both cases. How-

ever, the fact that different acquisition protocols generate different image contrast may introduce bias towards the image registration and have potential effect to the final quantification of structural volume. This issue will be discussed in more detail in section 4.4.

To study and compare the power of *in vivo* and *ex vivo* measurement to detect the volume differences between groups, image data from genetically modified mice (rTg4510) is used which model Alzheimer's disease by overexpressing mutant human tau protein, leading to the Tau neurofibrillary tangles (NTF) pathology [213]. The rTg4510 mice, which are littermates of the FVB/NCrl wildtype mice, were licensed from the Mayo Clinic (Jacksonville Florida, USA) and bred for Eli Lilly by Taconic (Germantown, USA) [211]. The rTg4510 dataset consisted of 17 animals, of which 10 received no intervention (untreated group), and the remaining 7 were orally fed with doxycycline mixed chow to reduce the overexpression of mutant (human) tau from 4.5 months (treated group). Animals in both the treated and untreated groups were underwent the same *in vivo* and *ex vivo* scanning protocols as their wildtype littermates.

4.2.2 Automatic structural parcellation

The brain structures were extracted using the multi-atlas segmentation propagation framework I previously developed which has been evaluated extensively on both *in vivo* and *ex vivo* mouse brain image data [214]. I used a publicly available atlas database created by Ma *et al.* which includes both *in vivo* [3] and *ex vivo* atlases [78]. The images from the atlas were firstly globally registered to the test images with a symmetric block-matching approach [185, 186], followed by a local registration step with a symmetric scheme based on a cubic B-Spline parametrisation of a stationary velocity field [86, 183, 186]. Image similarities are measured using normalised mutual information. The registered structural labels were then fused based on ranking using local normalised cross correlation similarity measurements [117].

4.2.3 Volumetric analysis

4.2.3.1 Structural volumetric analysis of *in vivo* and *ex vivo* images

As mentioned in the introduction, previous clinical and preclinical studies showed different or even controversial results regarding volume shrinkage from *in vivo* to *ex vivo* brain. With the structural volume information readily available from the automatic structural parcellation framework, I aimed to validate such structural volume changes

in the mouse brain, when imaged either *in vivo* or *ex vivo* in the same cohort. The metrics of interest comprise correlations between structural volumes of *in vivo* and *ex vivo* data, here defined as the Pearson product-moment correlation coefficient.

4.2.3.2 Relative structural volume difference

The correlation test alone is not capable of detecting systematic bias when assessing the consistency of two measurements. In medical research, a Bland-Altman plot is often used to illustrate the agreement or difference of two measurements, which plots the measurement differences on the y axis with respect to the mean measurement value on the x axis [215, 216, 217]. Here, I use the Bland-Altman plot to investigate the distribution of the difference between the *in vivo* and *ex vivo* measurement for each individual structure. It is recommended by Pollock *et al.* [218] that, when the variability of the measurement differences is related to the magnitude of the measurements, one should plot the proportional difference in the magnitude of the measurements on the y axis of the Bland-Altman plot instead of the absolute difference. Here in this study, since the structural volume difference is indeed directly related to the structural size, I plotted the difference in structural volume in proportion to the mean structural volume against the mean structural volume, $D(x,y)$, for each structure, where:

$$D(x,y) = \left(\frac{V_{in}+V_{ex}}{2}, \frac{V_{in}-V_{ex}}{(V_{in}+V_{ex})/2}\right) \tag{4.1}$$

in which V_{in} is the volume of the structure in the *in vivo* image, and V_{out} is the volume of the structure in the corresponding *ex vivo* image.

To further investigate whether the observed volume difference was statistically significant, I further examined the volume difference between the *in vivo* and *ex vivo* measurements through paired statistical t-test for all the parcellated structured. All tests were corrected for multiple comparisons with a false discovery rate (FDR) of 0.05.

4.2.3.3 Statistical power comparison

As an analytical imaging biomarker, it is important to assess the ability of the volumetric analysis to characterise and classify subjects from different groups (e.g. distinguishing normal population from genetically altered animals or diseased animals, or detecting drug efficiency on disease animals with or without treatment). In this sec-

tion, I aim to investigate whether the improved image quality (e.g. resolution) of the *ex vivo* data can and increase the power of the volumetric analysis to discriminate between groups. The images from the rTg4510 mouse were adopted, and an unpaired statistical t-test were performed on the volumes of all the parcellated structures between the treated group and untreated group, for both the *in vivo* and *ex vivo* data. I also compared the structural volumes normalised by the total brain volume, which have been shown previously to increase the statistical power in the study by Lerch *et al.* [138]. The statistical tests were performed with and without normalisation. All tests were corrected for multiple comparisons with a false discovery rate (FDR) of 0.05.

4.3 Results

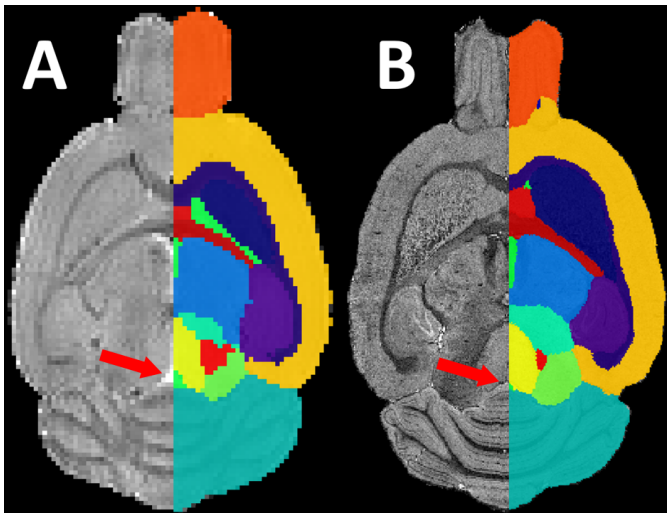

Figure 4.1: Representing images of the (A) *in vivo* image and (B) *ex vivo* image. The *ex vivo* images possess better resolution and contrast, which could potentially lead to less partial volume effect as well as better image registration. Red arrow indicates an example of collapsed ventricles.

4.3.1 Automatic structural parcellation

After the structural parcellation, both the *in vivo* and *ex vivo* wildtype mouse brain images are segmented into 37 functional structures as defined in the mouse brain atlas (Fig. 4.1). The *ex vivo* image quality is superior both in terms of resolution and signal contrast. However, it can also be observed from the images that the ventricles, which can be clearly observed in the *in vivo* images, collapse in the post-mortem brain as shown in the *ex vivo* images, which are captured by the structural parcellation (as shown in Fig. 4.1 red arrow). It is worth noting that, in some extreme cases in which

the entire ventricles collapses, there are potentially underestimated deformation during image registration because of the topological change, resulting underestimated volume changed.

4.3.2 Volumetric analysis

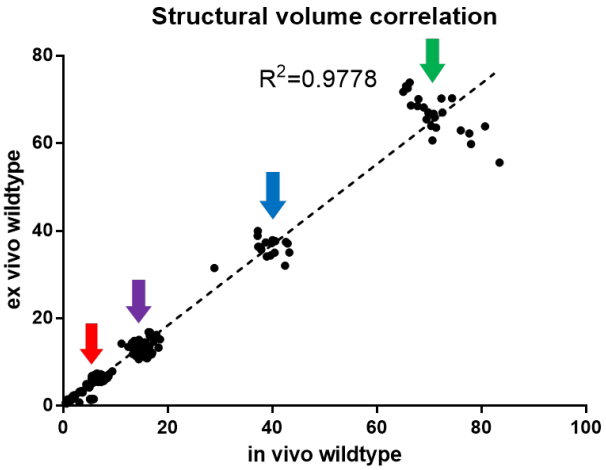

Figure 4.2: The structural volumes correlation of the *in vivo* data versus the *ex vivo* data. The Pearson correlation coefficient (R2) is 0.98. Red arrow: ventricle; purple arrow: olfactory bulb; blue arrow: cerebellum; green arrow: neocortex.

The next step is to investigate whether these confounding factors between the two volume measurements (*in vivo* and *ex vivo*) affect the data analysis and conclusion. As expected, there is a strong correlation between the structural volumes of the *in vivo* parcellation result and the *ex vivo* parcellation result. (R2=0.98), (Fig. 4.2).

However, the correlation coefficient is not sensitive to systematic bias when assessing the consistency of two measurements. When comparing the brain volumes between *in vivo* and *ex vivo* (with the ventricle excluded), a significant (paired t-test, $p<0.0001$) average total brain volume loss of 9.5% was observed in the wildtype animal, which is in line with previously reported findings [3]. (The total brain volume loss is 9.6% for the untreated group, and 9.3% for the treated group, both differences are significant, paired t-test, $p<0.0001$.) To evaluate whether or not this volume change is a systematic effect that is consistent across different structures, I assigned the value of the proportional structural volume difference to the corresponding structural labels for better visualisation (Fig. 4.3). It can be observed that most regions shrank by a small amount, with ventricles being the most affected region. For some small structures such

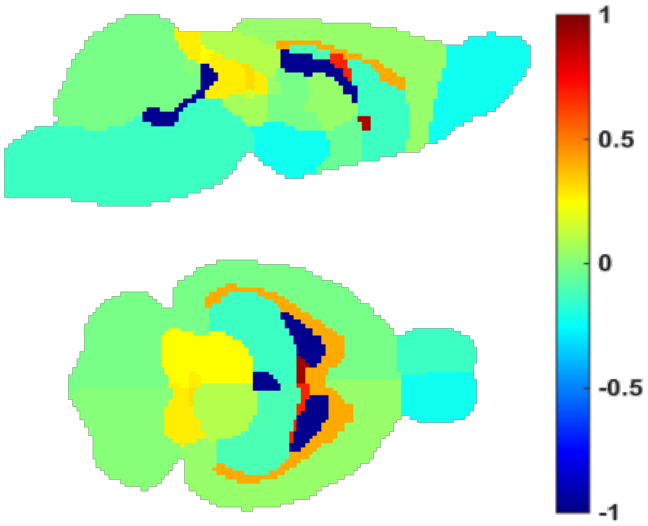

Figure 4.3: Colour-map showing the normalised structural volume difference on each of the corresponding labelled structures. Note that for anterior commissure (dark red), fimbria (light red), internal capsule (orange) and central grey (yellow), the *ex vivo* volumes are significantly larger than the *in vivo* volume.

as the central gray, globus pallidus, internal capsule, fimbria and anterior commissure, the *ex vivo* measure of the volume size appeared to be bigger in *ex vivo* images when compared to *in vivo* measurements.

I then used the Bland-Altman plot to investigate the proportional difference of the structural volume measured from *in vivo* and *ex vivo* data. The Bland-Altman plot (Fig. 4.4) demonstrated the volume difference of the *in vivo* and *ex vivo* volume measurements proportional to the mean structural volume for the wildtype animals. The -110% volume change of the ventricles (indicated by the red arrow) reflected the collapse of the ventricles from *in vivo* to *ex vivo*. If the volume shrinkage is uniformly distributed with ventricles excluded, the volume difference of each structure should be proportional to its volume, and all the values would be similar to each other. However, the Bland-Altman plot shows variation of structural difference which indicates non-uniform distribution of the volume shrinkage from *in vivo* to *ex vivo*. Specifically, the volume of the olfactory bulb (purple arrow) shrinks around 30% from *in vivo* to *ex vivo*, the shrinkage of the cerebellum (blue arrow) is 10%, while there is negligible volume shrinkage in the cortex (green value). The relatively larger proportional volume difference for small structures might be due to the partial volume effect which becomes more predominant when the structure size is smaller.

When examining the volume difference between the *in vivo* and *ex vivo* measure-

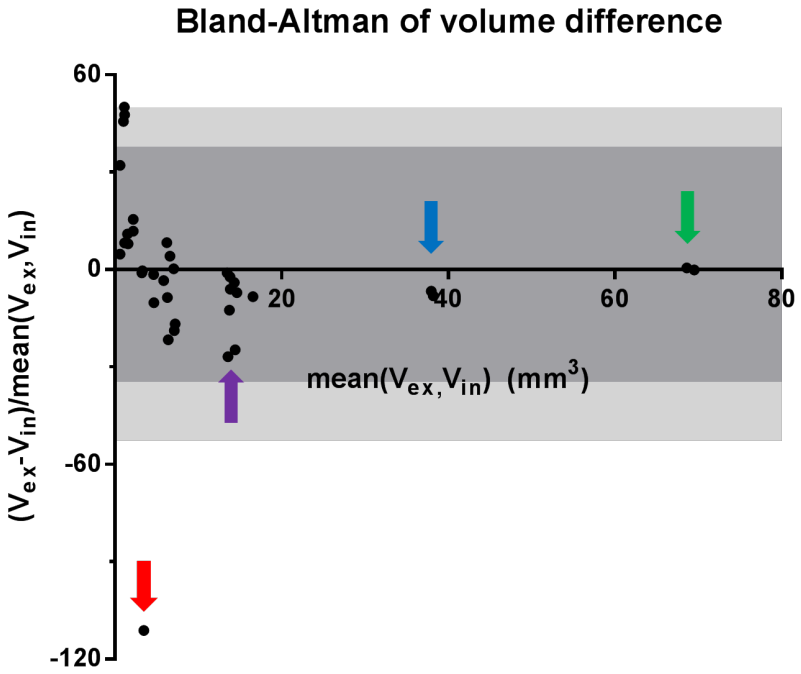

Figure 4.4: Bland-Altman plot showing the mean proportional volume difference between *in vivo* and *ex vivo* measurements with regard to the structural size for the wild type animals. Light grey area: 95% limit of agreement between *in vivo* and *ex vivo* measurement (with ventricles included); dark grey area: 95% limit of agreement (with ventricles excluded). Red arrow: ventricle; purple arrow: olfactory bulb; blue arrow: cerebellum; green arrow: neocortex.

ments through paired statistical t-test (Fig. 4.5), difference in the neocortex, inferior colliculi, caudate putamen and the rest of the midbrain on both sides of the brain were not found to be the statistically significant, which confirms the conclusion that the volume change is not equally distributed across structures. There is a noticeable large volume shrinkage of the olfactory bulb. A possible explanation of such volume difference might be due to the connected olfactory nerves, which is attached to the olfactory bulb without clear boundary, can can be over-segmented in the *in vivo* images with lower resolution. In addition, the statistically significant volume shrinkage of the caudate putamen and amygdala appears only on left side of the brain, indicating that the level of brain shrinkage might not be symmetrical for these two structures, which might further reflect some intrinsic asymmetry property of the brain. For the small structures such as central gray, globus pallidus, internal capsule, fimbria and anterior commissure, the volume increases from *in vivo* to *ex vivo* measurement are significant.

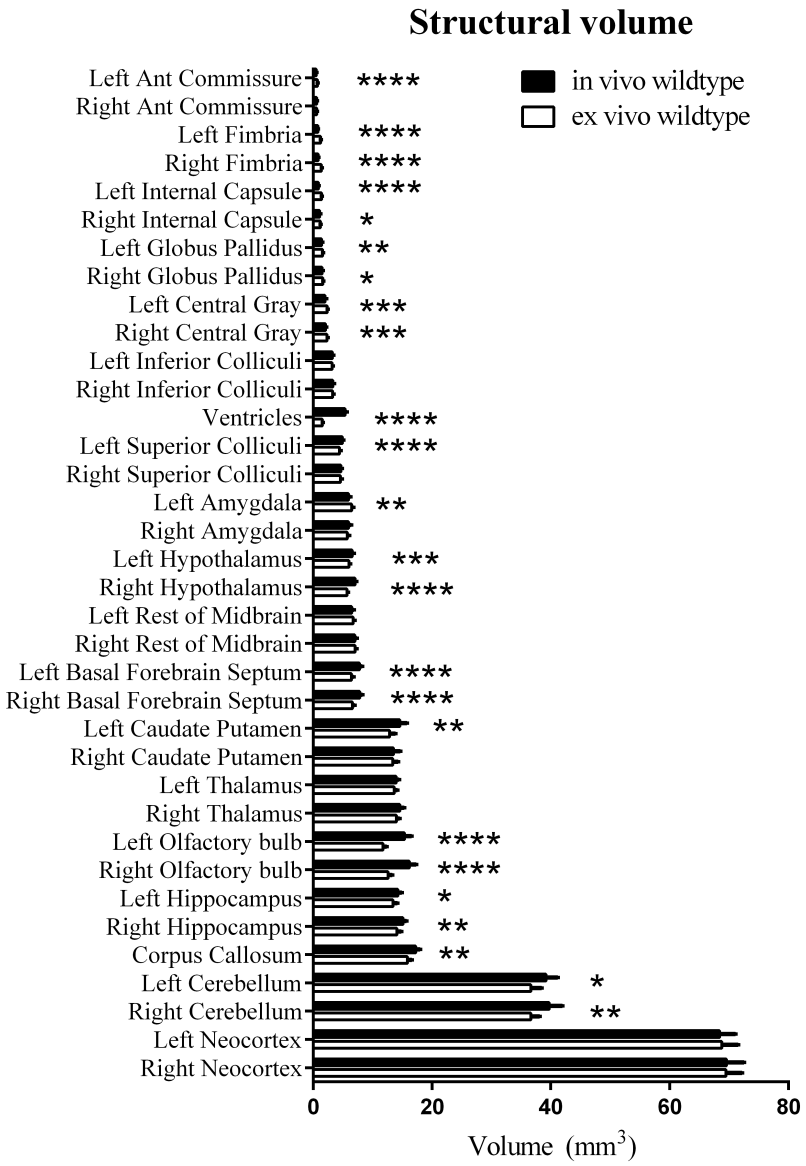

Figure 4.5: Statistical comparison of *in vivo* and *ex vivo* structural volumetric measurement. Error bars represent the standard deviations. (*: $0.01<p\leq0.05$; **: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$; ****: $p\leq0.0001$.)

4.3.3 Statistical power comparison

In order to compare the statistical power of the volumetric analysis between *in vivo* and *ex vivo* measurement, I used the rTg4510 transgenic mouse which overexpresses mutant human Tau leading to neurofibrillary tangles to model AD, and compared the brain structures of animals with or without doxycycline treatment to suppress tangle pathology. Unpaired two-tailed statistical t-tests for both measurements were used to compare the individual structural volumes between the treated and untreated groups (Fig. 4.6 and Fig. 4.7). Both the unnormalised and normalised volumes were tested.

For the unnormalised data, both the *in vivo* and *ex vivo* results showed significant volume differences between the treated and untreated groups for neocortex and hippocampus (both left and right) and left external capsule. Significant difference for left caudate putamen was detected in *in vivo* image but not in *ex vivo*, and for cerebellum was detected in *ex vivo* image but not in *in vivo*. Additionally, significant ventricular shrinkage was detected by the *in vivo* measurement, while the ventricles collapsed in the *ex vivo* data for both the untreated and treated groups, resulting in no significant difference between the groups (Fig. 4.6).

Similar results were obtained after the structures were normalised: comparable volumetric differences were found in both *in vivo* and *ex vivo* measurements (Fig. 4.7), with significant ventricular shrinkage was detected by the *in vivo* measurement but not the *ex vivo* measurement.

4.4 Discussion

When studying brain imaging, one has to choose whether to scan animals *in vivo* or *ex vivo* when designing the experiments. It is sometimes a controversial choice as each of the paradigm has its own strengths and weaknesses. In this study, I aimed to investigate how this choice of paradigm will affect the volumetric analysis, by using the previously extensively evaluated accurate multi-atlas-based automated mouse brain structural parcellation framework.

I first parcellated sets of *in vivo* and corresponding *ex vivo* images of wildtype mouse brain, using the method developed in Chapter 3. I then compared the *in vivo* and *ex vivo* volume measures for each individual structure. The result showed that the total brain volume change from *in vivo* to *ex vivo* mouse brain is not uniformly distributed across different structures: while some regions shrank to various degrees, others increased in size and some had little or no change in volume.

I also applied the structural parcellation to transgenic mice modelling AD, where one group underwent no treatment, while the other group treated with doxycycline to suppress the expression of the transgene. Both the *in vivo* and *ex vivo* measurements provide comparable statistical significance on structural volume to distinguish the treated and untreated groups.

The ventricular expansion is an important neuroimaging biomarker for neurode-

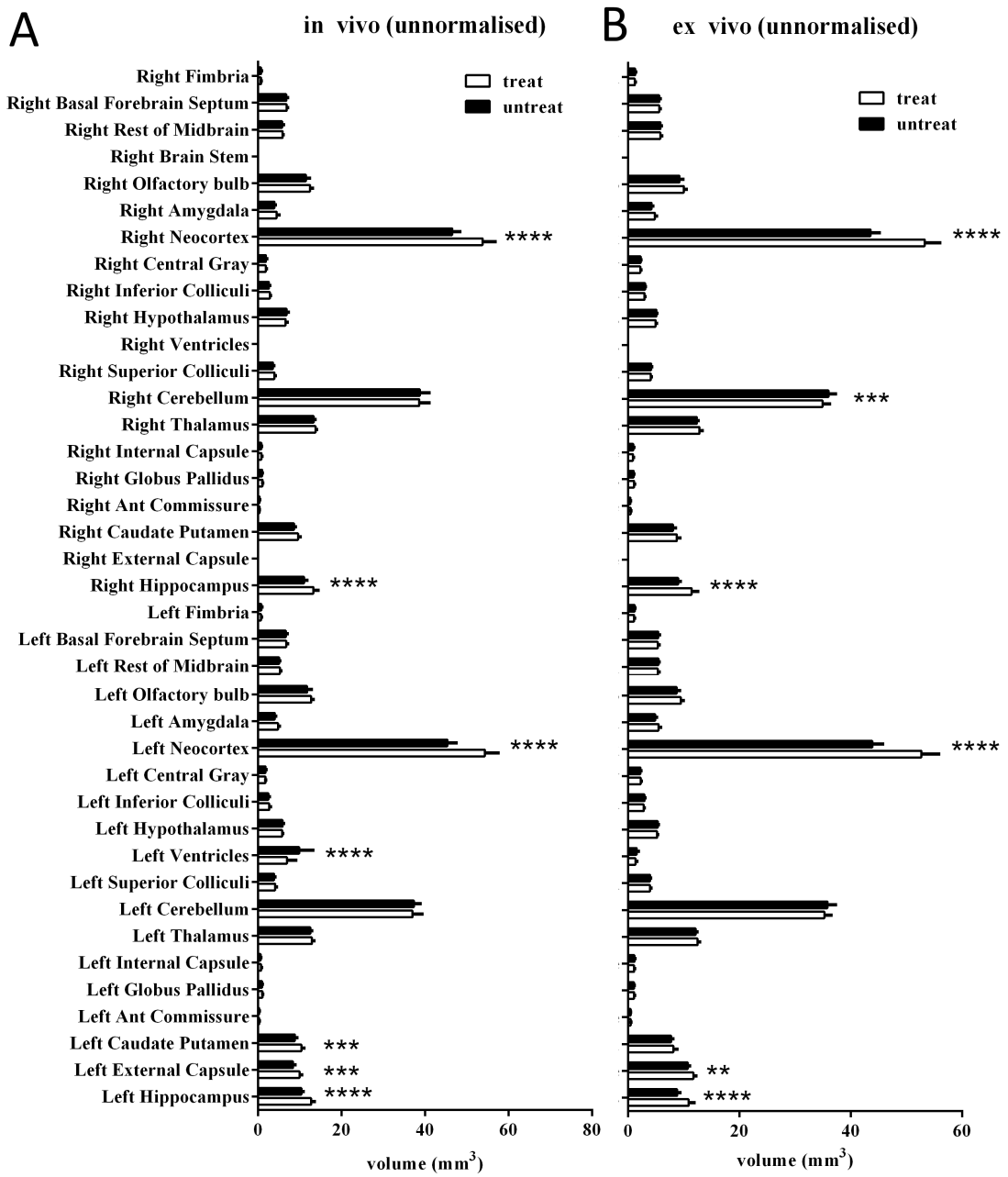

Figure 4.6: Statistical analysis to compare the unnormalised individual structural volume of the treated and untreated group on both (A) *in vivo* and (B) *ex vivo* data. All tests were corrected for multiple comparisons with a false discovery rate (FDR) of 0.05. Error bars represent the standard deviations. (*: $0.01<p\leq0.05$; **: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$; ****: $p\leq0.0001$.)

generative diseases such as Alzheimer's Disease and potential treatment [219, 220]. The result of this study showed that the *in vivo* volumetric measurement can detect the significant volume difference between treated and untreated groups, while the *ex vivo* measurement failed to do so due to the ventricular collapse and the loss of ventricular CSF in the post-mortem brain tissue preservation process.

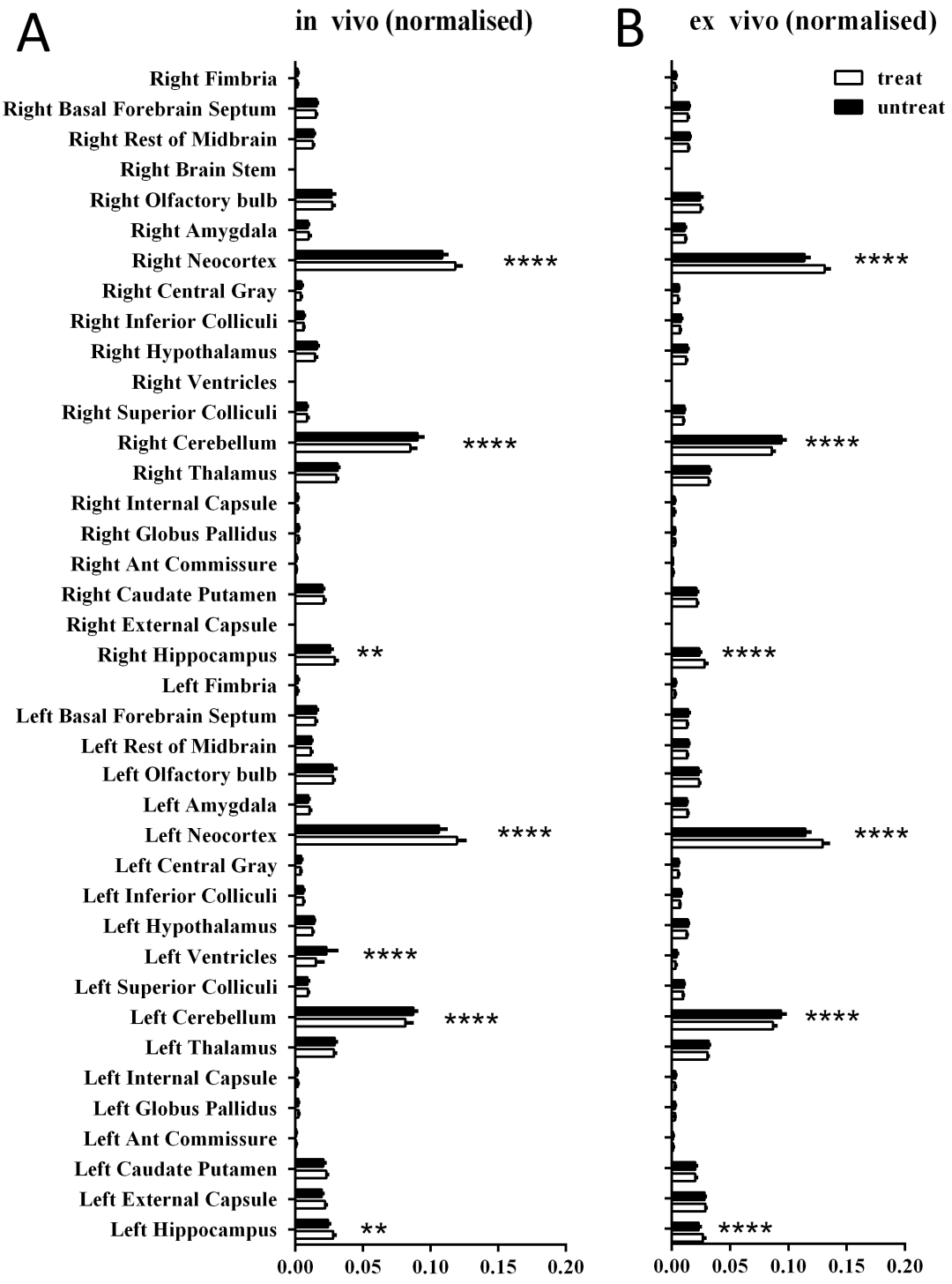

Figure 4.7: Statistical analysis to compare the normalised individual structural volume of the treated and untreated group on both (A) *in vivo* and (B) *ex vivo* data. All tests were corrected for multiple comparisons with a false discovery rate (FDR) of 0.05. *: Significant difference between the wildtype and the transchromosomic group. Error bars represent the standard deviations. (*: $0.01<p\leq0.05$; **: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$; ****: $p\leq0.0001$)

Previous clinical and preclinical studies showed contradictory conclusions about the volumetric loss from *in vivo* to *ex vivo* brain. No significant volumetric loss was detected in the study by Kotrotsou *et al.* [142] and Oguz *et al.* [143], while the studies of Schulz *et al.* [141], Ma *et al.* [3] and Zhang *et al.* [140] showed various levels of volumetric loss. The results of this study agree with the study by Schulz *et al.*, Ma *et al.* and Zhang *et al.*. Specifically, all three studies reported large volumetric loss of the ventricles due to the CSF leakage (85.7% in this study, 78.6% from Ma *et al.* and 97.3% from Zhang *et al.*). In addition, the results of this study also agree with the findings of Schulz *et al.* [141] that the volumetric change is unevenly distributed. The volumetric change during the post-mortem fixation and the perfusion process varies in different studies, possibly due to factors such as animal strain differences, pathological model diversity and protocol variation. These various factors might impact on the accuracy and reliability of the quantitative measurement of the *ex vivo* data.

Given the advantage and disadvantage of both the *in vivo* and *ex vivo* images, one should not be treated as the gold standard over the other. In the current study, both *in vivo* and *ex vivo* images were acquired under different imaging protocols with different scanning sequences and coils to acquire images with optimised quality in both cases. The measured volume difference might be due to the actual biological difference, but may also be affected by the differences coming from the image acquisition protocols. The artefacts coming from the spin echo sequence for the *in vivo* imaging and that from the gradient echo sequence for the *ex vivo* imaging inevitably have different effects toward the acquired images and also the final volumetric analysis results. In this study, the gradient for each coil are calibrated to reduce the corresponding bias. However, in order to have a bias-free comparison to accurately assess the volume change, the same acquisition protocol (same coil and acquisition sequence) can be used for both *in vivo* and *ex vivo* imaging for volumetric comparison, which will eliminate some confounding factors from images, such as resolution-varying partial volume effect and gradient bias. Seperate experiments are needed to evaluate the degree of the effect.

Through simulated statistical analysis, Lerch *et al.* [138] concluded that *ex vivo* imaging is preferable for study measuring relative volume (normalised to the TBV) while *in vivo* experiments is a better choice for absolute volume measurements. The result of this chapter shows comparable statistical analysis results between *in vivo* and

ex vivo brain MRI image data for volumetric measurements both with and without normalisation.

The statistical result in this study includes structures with different size, with various volume change from *in vivo* to *ex vivo*. However, the conclusions might be different for data with smaller effect size in terms of volumetric difference. A thorough power analysis is necessary to determine the appropriate sample size required for a specific effect size.

In the field of preclinical imaging research, the choice of experimental design is expected to shift from histology to *ex vivo* imaging, and later to *in vivo* imaging. Such a shift in imaging paradigm will not only enable longitudinal assessment of the neuroanatomical changes, but will also help reduce the unnecessary sacrifice of animals in preclinical studies.

It has been shown in previous histology studies that both the fixation and perfusion process caused tissue shrinkage [221, 222]. Recently, Halbach *et al.* [223] proposed an alternative method to conduct post-mortem *ex vivo* imaging directly after sacrifice, to prevent the potential volume change in the fixation and perfusion process. While this is an interesting alternative solution, the necessity to counteract fast post-mortem tissue degradation imposes experimental complexity as well as uncertainty [224]. Further difficulties arise when we consider the long scanning time of the *ex vivo* imaging which may amount to several hours, and the reusability of the samples for potential histology study afterwards.

I used an atlas-based method to extract the entire brain from the original *in vivo* and *ex vivo* images, which necessitates some manual correction to exclude remaining CSF surrounding the brain tissue as well as intracranial nerves. A better, more accurate way to extract the brain would be to use a tissue classification technique, which uses expectation maximisation to update the tissue probability for each voxel, including grey matter, white matter, CSF, and non-brain tissues [225, 226, 227]. However, a tissue probability map is necessary as prior information to initialise the expectation-maximisation procedure. The current limitation in mouse brain MRI studies is the lack of an accurate tissue probability map [228]. A tissue classification framework with accurate tissue probability map would be beneficial for future preclinical studies.

4.5 Conclusion

In conclusion, the results of this chapter demonstrate non-uniformly distributed structural volume change from *in vivo* to *ex vivo* measurements across the entire brain. However, interestingly, both *in vivo* and *ex vivo* results demonstrated similar discriminatory power when comparing the volume difference between two groups of mice, except that *in vivo* measurements showed ventricular shrinkage in the treatment group, in which the effects of the mutant Tau gene are suppressed, while *ex vivo* measurements did not due to the ventricular collapse in the post-mortem brain tissue preparation.

Chapter 5

Mouse cerebellar cortical sublayer thickness estimation through Purkinje layer extraction

5.1 Introduction

The morphology of the cortex has been widely used as a quantitative imaging biomarker in studies of neurodegenerative diseases, such as Alzheimer's disease or Down's Syndrome, for both clinical preclinical studies. Most studies about cortical thickness analysis focused on the cerebellar cortex, while little attention has been drawn to the cerebellum.

Recent studies comparing high-field MRI of cortical grey matters with histological staining [229, 230] showed that the variation of iron and myelin content in different cortical layers produced MR contrasts which reflect the local laminar architectures. With the administration of Gd-DTPA such inter-layer contrast can be further enhanced [231]. Marques *et al.* [232] also demonstrated the ability of T2*-weighted images of cerebellar cortex to reveal the contrast between the cortical sublayers - the grannular layer, the Purkinje layer and the molecular layer. It is thus interesting to utilise such information revealed in MR to infer detailed morphological information of each cerebellar cortical sublayers.

This chapter introduces the framework I developed to estimate the layer thickness of mouse cerebellar cortex from the high resolution *ex vivo* mouse brain MRI. This framework takes advantage of the information embedded in the data, and obtains ac-

curate extraction of the middle Purkinje layer through surface segmentation and, when not visible, extrapolates it from the laminar layer model of the cortex. The framework explores and extends the conventional Laplace-equation-based cortical thickness estimation method to multiple sublayers. The extracted Purkinje layer enables the estimation the thickness of two other sublayers (*i.e.* the granular, and the molecular layer).

The framework has been evaluated on high resolution *ex vivo* T2* MRI data acquired from mouse models of Downs' Syndrome (TC1) with gadolinium contrast enhancement to get the optimised tissue contrast among cortical sublayers (detailed acquisition protocol are described in section 5.3), and found reduced cortical and layer thicknesses in the transchromosomic group, which agrees with previous histological study of such disease model [174]. However, this framework cannot be applied to the *in vivo* MRI yet due to the limitation of the resolution and contrast, even with optimised acquisition protocol (Figure 4.1 A in section 4.3).

To the best of my knowledge, this is the first time the cortical thickness analysis of MRI has been used to study the mouse cerebellum, and is the first study that managed to extract and measure the thickness of the cortical sublayers.

5.2 Methods

The overall pipeline of the proposed framework is presented in Fig. 5.1. The first part of the framework includes the extraction of the cerebellum cortex (Section 5.2.1), the fissure extraction (Section 5.2.2) and the parcellation of the cerebellum based on the functional characteristics of the tissues (Section 5.2.6). The second part includes the extraction of the Purkinje layer (Section 5.2.3), which enables the estimation of the sublayer thickness (Section 5.2.5). The Purkinje layer extraction is achieved through three substeps: the planar structure filtering, the laminar layer estimation and the surface extrapolation.

In the following sections, I will describe each step of the framework's pipeline in detail.

5.2.1 Cerebellar cortex extraction

Cerebellum extraction. First of all, the cerebellum is extracted from the original MR images. I use the multi-atlas structural parcellation framework described in Chapter 4

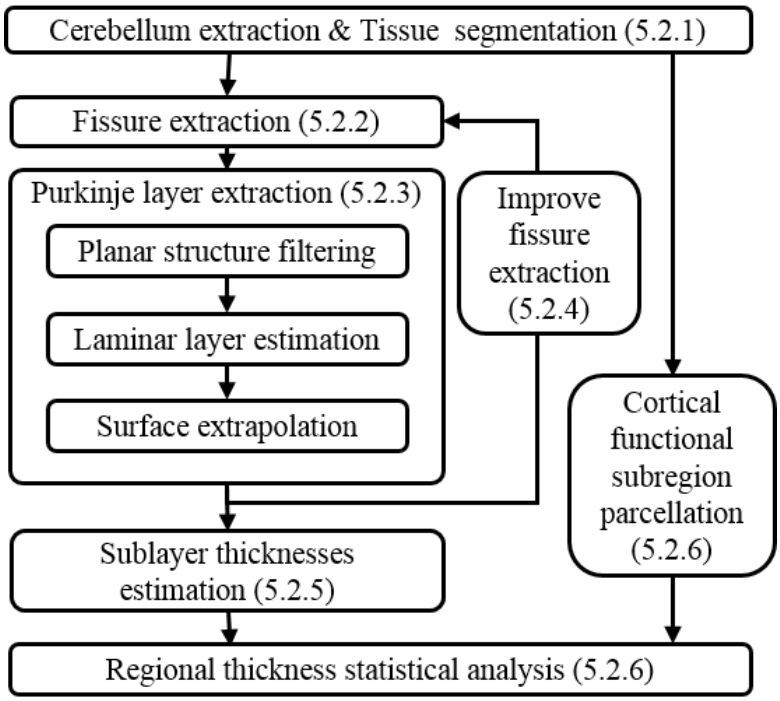

Figure 5.1: Schematic diagram of the proposed framework.

with a publicly available *ex vivo* atlas database created by Ma *et al.* [78] to obtain an accurate structural parcellation (Fig. 5.2 red). A symmetric block-matching approach [185, 186] is used for the global registration steps, a symmetric scheme based on a cubic B-Spline parameterisation of a stationary velocity field [86, 183, 184] is employed for the local registration stages. Normalised mutual information is used to measure the image similarity in both the global and local image registration steps. Structural labels are fused based on a ranking stragegy using local similarity measurements [117].

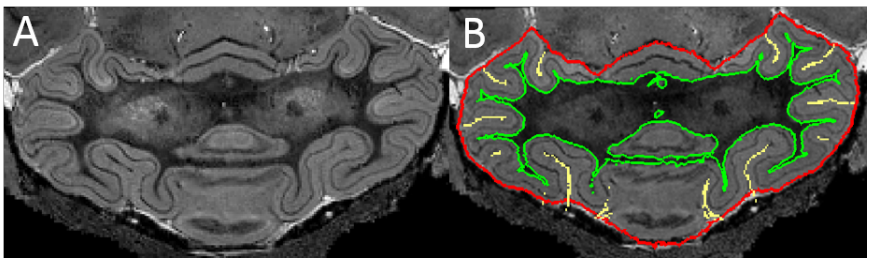

Figure 5.2: Representative images of (A) the cerebellum in the original gadolinium enhanced *ex vivo* mouse brain MRI, (B) overlaid with the result of the cerebellum extraction (red), the tissue segmentation (green) and the fissure extraction (yellow).

Tissue segmentation. The next step is to obtain an accurate segmentation of the WM and GM tissues (Fig. 5.2 green). I follow the approach introduced by Ashburner & Friston based on a Gaussian mixture model of tissue classes, and segmented the cere-

bellum into grey matter and white matter within an expectation-maximisation framework [233]. This approach gains in robustness by using prior spatial information in the form of tissue probabilistic maps.

A tissue probability map is a necessary prior information for tissue segmentation. There is no currently available tissue probability map for mouse cerebellum. To address this issue, a semi-automatic approach is used to first generate the tissue probability maps, and then segment the tissues.

Prior to creating the tissue probability map, the intensity distribution within the cerebellar area are standardised for all MR images in the study using a landmark-based piecewise linear intensity scaling introduced by Nyúl *et al.* [234], with 11 landmarks automatically defined as equally distributed percentiles of the histogram.

An average image is then created using an iterative groupwise scheme. In the first iteration, images are rigidly aligned to the space of a reference image randomly selected from the data. Subsequently, ten iterations of affine registrations and ten iterations of non-rigid registrations were performed to correct the global and local misalignment respectively (Fig. 5.3). In each iteration, the average image from the previous iteration were treated as the reference image, and repeated until the average intensity standard deviation across all voxels converges [235]. Similar to the cerebellum extraction step, a symmetric block-matching approach [185, 186] is used for the global registration steps. a symmetric scheme based on a cubic B-Spline parameterisation of a stationary velocity field [86, 183, 184] is employed for the local registration stages

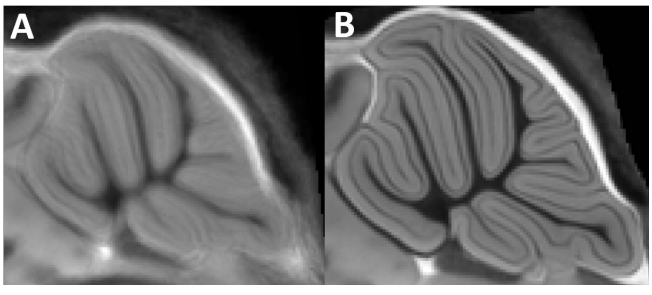

Figure 5.3: The group average image after (A) the groupwise affine registration steps, and (B) the groupwise non-rigid registration steps. It can be observed that the final average image are much sharper than the affine-registered average image.

The group average image is then segmented into four tissues types (one for WM and three for GM due to the intensity variation between cortical layers), using a Gaus-

sian mixture model of tissue classes without using anatomical priors. An anatomically correct WM segmentation is then obtained by manually disconnecting the Purkinje layer voxels which have been misclassified as WM, and keep the largest connected component. This step only needs to be done once, as the groupwise tissue segmentation can be used to segment new subjects.

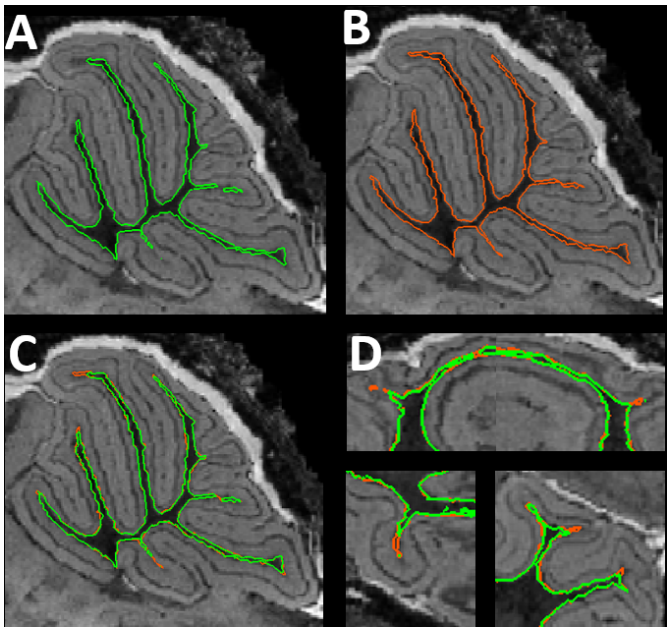

Figure 5.4: The WM segmentation (A) after label propagation from the group average image. (B) After applying the leave-one-out framework within the testing image group, followed with a further Gaussian-mixture-model-based tissue segmentation (C) two results overlaid. (D) Other representative images of WM segmentation improvement.

The segmented WM and GM are propagated back to the initial input images using the backward transformations obtained during the groupwise registration step. As proposed by Chakravarty *et al.* [131], all input images and their newly generated tissue segmentations are considered as a template database for a multi-atlas label propagation and fusion scheme to further improve the parcellation accuracy for each image. Using a leave-one-out framework, for each image, a spatial anatomical prior is generated by first propagating all the tissue segmentations from every other image to the target dataset, followed by a fusion step [214]. Each image is then segmented using a Gaussian mixture model of tissue classes combined with the image specific probability maps (Fig.5.4).

5.2.2 Fissure extraction

The morphology of mouse cerebellum consists of folia lobes separated by fissures, making the cerebellar cortex a convoluted folding structure. Most of these thin fissures are subjected to partial volume (PV) effect in MRI, resulting in incorrect tissue segmentations. In order to appropriately extract the fissures, even when highly corrupted by PV, a distance-based skeletonisation technique [151] is used here. A geodesic distance transformation $D(x)$ is obtained by solving the Eikonal equation $F(x)|\nabla D(x)| = 1$ from the white/grey matter boundary. Here, $D(x)$ is the geodesic distance and $F(x)$ is the speed function, in the computational domain Ω, defined as $F = I * \mathscr{G}_\sigma$, with I being the image and $\mathscr{G}_\sigma$ being a Gaussian kernel with $\sigma = 1.5$ voxels. The fissures are then extracted by first finding the local maxima of $D(x)$ only along the direction of $\nabla D(x)$, followed by a recursive morphological thinning in order to ensure single voxel thick fissure segmentations [236].

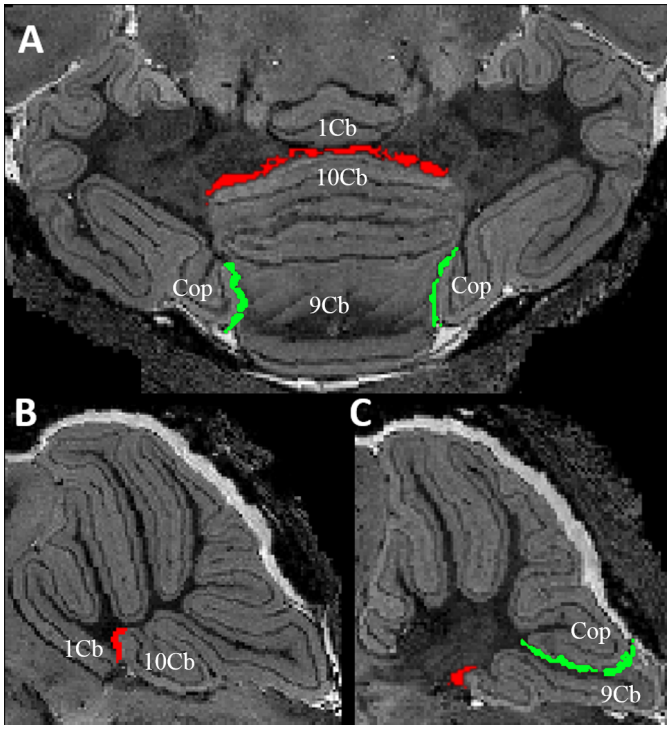

Figure 5.5: Two boundaries are defined to separate the cerebellar cortical lobes which are not anatomically connected. One (red) is defined between the cerebellar vermis lobule 1 (1Cb) and lobule 10 (10Cb); the other one (in green) is defined between cerebellar vermis lobule 9 (9Cb) and Copula of the pyramis in the cerebellar hemispheres (Cop) (A) Axial view of the two resistant layers, (B,C) Sagittal view on two different planes of the resistant layers.

In the cerebellar vermis, the lobule 1 (1Cb) and lobule 10 (10Cb) touches each

other while they were not anatomically connected (Fig.5.5 A,B). This is also true for the cerebellar vermis lobule 9 (9Cb) and Copula of the pyramis in the cerebellar hemispheres (Cop) (Fig.5.5 A,C). To prevent these touching lobules from creating false positive fissure lines when calculating the geodesic distance, boundaries were manually delineated between these lobes from the groupwise average image, and propagated back to each individual image (Fig.5.5).

Representative images of the extracted fissure lines are shown in Fig. 5.2 (yellow).

5.2.3 Purkinje layer extraction

In this step, I extract the narrow layer that sits in the middle of the cerebellar cortex - the Purkinje layer, which is a mono-cell layer consisting of Purkinje cells. This will make it possible to obtain the thickness of the other two grey matter sublayers - the molecular layer and the granular layer.

The extraction of the Purkinje layer is a challenging task due to the intensity homogeneity and partial volume effect and its highly convoluted nature. Simple thresholding or even Gaussian-distribution-based tissue classification methods failed to extract the Purkinje layer without heavy manual intervention. The method I propose here segments the Purkinje layer by exploiting both the intensity information and its laminar nature.

Planar structure filtering. Giving the surface nature of the Purkinje layer, a modified Frangi vesselness filter [237] is used here to find and enhance the image contrast of planar structures *P(s)* (instead of tubular ones) within the cortical region at a fixed scale $s = 0.04$:

$$P(s) = \begin{cases} 0 & \text{if } \lambda_3 < 0 \\ exp(-\frac{R_A^2}{2\alpha^2})exp(-\frac{R_B^2}{2\beta^2})(1-exp(-\frac{S^2}{2c^2})) & \end{cases} \tag{5.1}$$

where α, β and c are thresholds, $R_A = \frac{|\lambda_2|}{|\lambda_3|}$, $R_B = \frac{|\lambda_1|}{\sqrt{|\lambda_2\lambda_3|}}$, and λ_k are the k-*th* smallest eigenvalue decomposition ($|\lambda_1| \leqslant |\lambda_2| \leqslant |\lambda_3|$). Areas with significant filter response are used as an initial estimates of the Purkinje layer M_{P0} (Fig. 5.6 A).

Cortical laminar layer estimation. The above method does not capture the Purkinje in its entirety. Regions with high curvature cannot be captured entirely. Thus, the initial estimation of the Purkinje layer is used to extrapolate the remaining locations by

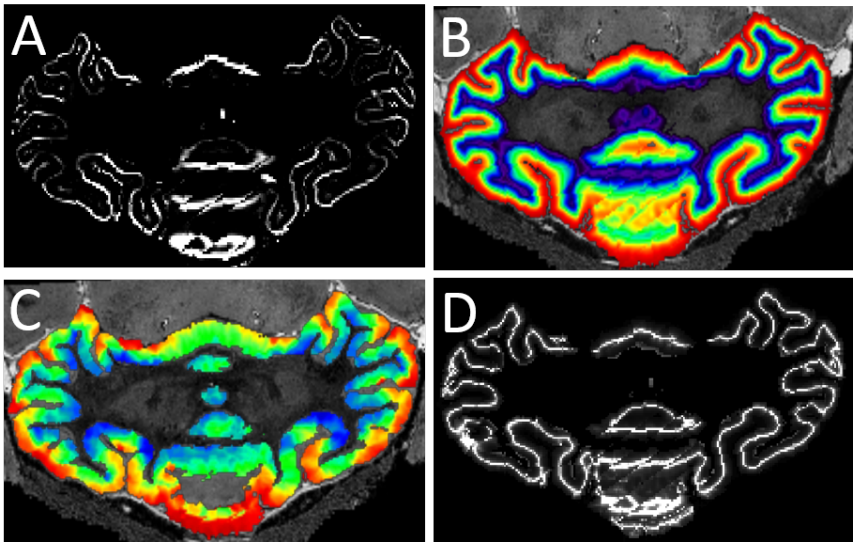

Figure 5.6: Two-step extraction of the middle cerebellar Purkinje layer. (A) The initial extraction is obtained from adopting a Frangi filter to enhance the contrast of planar structure. (B) Cortical laminar map is generated with Laplacian-equivolume-combined model R_{vol}. (C) Gaussian smoothed map for the laminar map on Purkinje layer R_{P_S}. (D) The final extraction is obtained from the map of λ.

mathematical modeling of the laminar layer.

The anatomical laminar layer of the cortex is modeled following the Laplacian-equivolume model proposed by Leprince et al. [37]. For the ease of reading, I briefly describe the Laplacian-equivolume model below, which has been reviewed in Section 2.3.1.2.

I first reconstruct the pial surface by combining the extracted fissures with the outer boundary of cerebellum mask. An initial layer estimation is derived from a Laplacian field level-set analogues between the pial surface and the WM/GM boundary following the implementation introduced by Jones *et al.* [33] and Yezzi *et al.* [36].

A unit surface δS is defined for each cortical voxel on the streamline following the normalised unit vector $\overrightarrow{F}$ of the Laplacian field. The relative surface area change of the adjacent points along the streamline are obtained from the divergence of the vector field $\nabla \cdot \overrightarrow{F}$.

$$\delta S_{n+1} = \delta S_n (1 + \delta l (\nabla \cdot \overrightarrow{F})) \,. \tag{5.2}$$

An upwind volume towards pial surface V_{pial} and a downwind volume towards white matter V_{WM} are calculated by accumulating the relative surface areas of the current voxel X_0 along the streamline following $\overrightarrow{F}$ and $-\overrightarrow{F}$ relatively. The laminar information

on each voxels is defined as the relative volumetric ratios R_{vol}.

$$R_{vol} = \frac{V_{pial}}{V_{pial} + V_{WM}} \,. \tag{5.3}$$

The final laminar layer is shown in Fig. 5.6 B.

Surface extrapolation. The model-derived laminar information obtained from the last step is used to extrapolate the missing parts from the initially extracted Purkinje layer. The volume ratio are defined at each Purkinje layer voxel as $R_P = R_{vol} * M_{P0}$. A multi-level Gaussian smoothing is applied on the volume-ratio-based laminar map R_{vol} to all voxels $x \in R_{vol} \cap x \notin M_{P0}$. Ten Gaussian smoothing levels with exponentially decreasing variances between 15 and 1 voxels were used in this work. This multi-level Gaussian smoothly propagates and averages the value of R_P to neighbouring regions, providing an estimate of R_P for voxels outside M_{P0}. The smoothed version of R_P is here denoted R_{P_S} (Fig. 5.6 C). If the Gaussian smoothed map R_{P_S} is equal to R_{WM} at location x, then the voxel x should be part of the Purkinje layer. I thus define the $\delta = |R_{P_S} - R_{WM}|$ as a measurement of distance between the two maps (Fig. 5.6 D).

In order to robustly find locations where $\delta \approx 0$, the local minima of δ is found only along the direction of the vector field $\overrightarrow{F}$. The recovered local minima δ_{min} are then added to the initial Purkinje layer mask, i.e. the final segmentation of the Purkinje layer is given by $M_{P^F} = M_{P0} \cup \delta_{min}$. This process can be repeated to gradually fill the gaps from the initial extraction until convergence (Fig. 5.7).

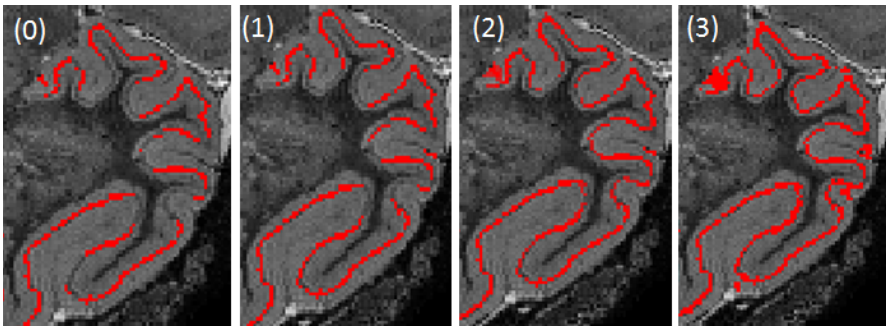

Figure 5.7: Iterative multi-kernel Gaussian smoothness gradually recovers the Purkinje layer. (0) Initial extracted Purkinje layer. (1-3) Recovered Purkinje layer after each iteration.

5.2.4 Improved fissure extraction after Purkinje layer removal

In the original fissure extraction step (Section 5.2.2), the grey matter intensity is used as the speed function to calculate the geodesic distance to the white matter surface, which is affected by the presence of the intensity from both the external CSF and Purkinje layer (Fig. 5.8 A,B).

In this step, we improve the fissure extraction by removing the intensity information from the Purkinje layer, effectively providing a speed function dominated mainly by the presence of external CSF, which represents the location of the fissures. This is achieved by applying another multi-level Gaussian smoothing to the normalised image intensities I for all voxels $x \in M_{P0}$, to replace the original intensity of the voxels at the Purkinje layer with an averaged value of the intensities in the surrounding cortical voxels (Fig. 5.8 C,D).

With the extracted fissure line updated, the laminar modelling and Purkinje layer extraction are updated accordingly as well with the updated fissure line to get a more accurate result.

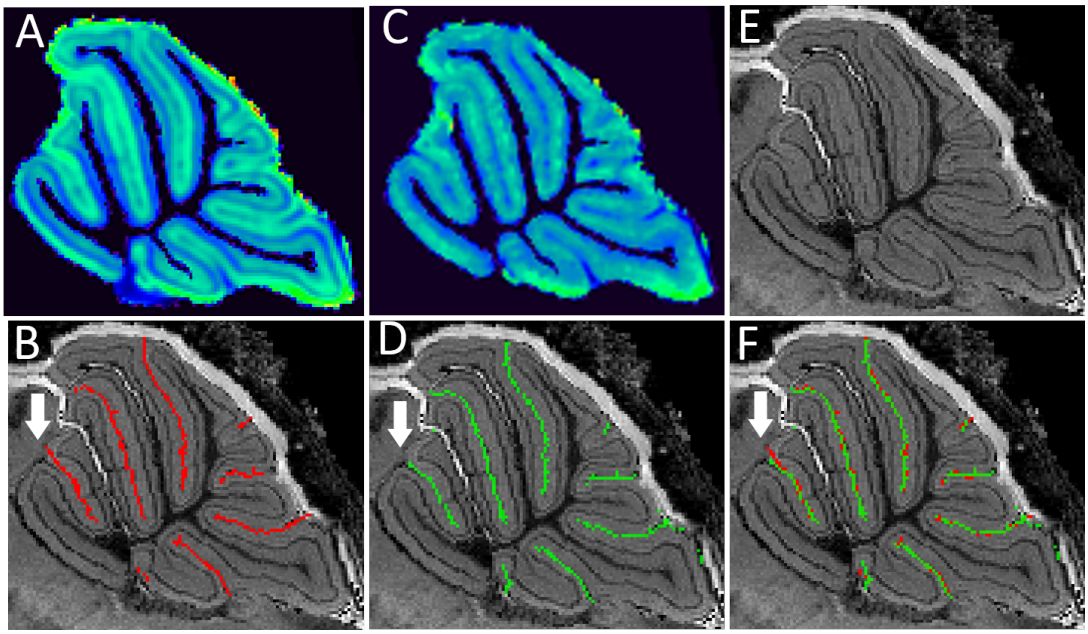

Figure 5.8: The the accuracy of fissure extraction is improved by removing the middle Purkinje surface. (A) Original speed function based on a Gaussian smoothed image intensity. (B) Original extracted fissure, with arrows indicating the inaccurate extraction (C) Speed function after removing the Purkinje layer and substitute with an averaged intensity of surrounding cortical areas (D) Updated extracted fissure. Arrows indicate the improved accuracy. (E) Original image. (F) Original and improved fissure overlaid on top of the original image

5.2.5 Cerebellar cortical sublayer thickness estimation

Lastly, all relevant thicknesses, including the entire cortical thicknesses T_{GM}, the granular layer thicknesses T_{Gran}, and the molecular layer thickness T_{Mol} are measured on voxels at the Purkinje layer location M_{PF}. For each voxel within M_{PF}, its distance towards the cortical surfaces are defined as the length of the streamlines passing through it, which is always perpendicular to the cortical layers. This is achieved by integrating the streamlines along the forward and backward directions of the normalised vector field $\overrightarrow{F}$, and are calculated using the Eulerian PDE method proposed by Yezzi and Prince [36]. At each Purkinje layer voxel, the length of the streamlines integrated along the direction $\overrightarrow{F}$ towards the pial surface is denoted by D_{Pial}, and the streamlines integrated along the $-\hat{V}$ towards the WM is denoted by D_{WM}.

The granular layer thickness is defined as the distance from the WM, given by $T(x)_{Gran} = D_{WM}(x)\ \forall x \in M_{PF}$, the molecular layer thickness is defined as the distance from the pial layer, given by $T(x)_{Mol} = D_{Pial}(x)\ \forall x \in M_{PF}$, and the total GM thickness at the Purkinje layer voxels, denoted by $T_{GM_{PF}}$, is given by $T_{GM_{PF}}(x) = T_{GM}(x) \forall x \in M_{PF}$. The thickness of the entire cerebellar cortex x can thus be estimated by adding $D_{WM}(x)$ and $D_{\text{Pial}}(x)$, i.e. $T_{GM}(x) = D_{WM}(x) + D_{\text{Pial}}(x)$.

5.2.6 Cortical functional subregional parcellation

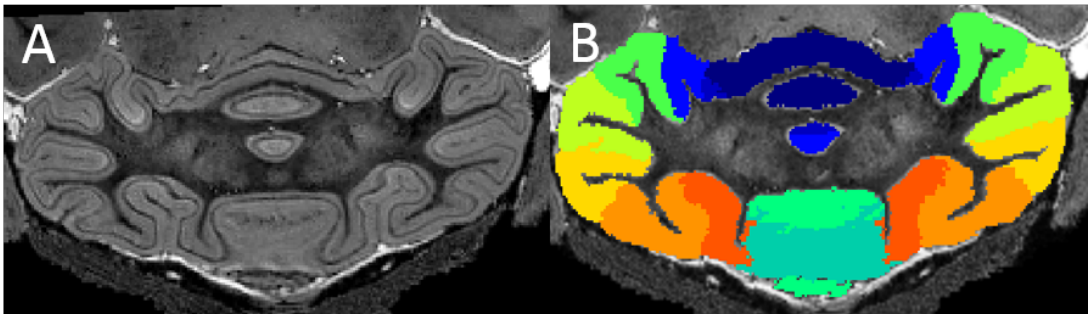

Figure 5.9: Representative images of (A) the cerebellum in the original gadolinium enhanced *ex vivo* mouse brain MRI, (B) overlaid with the parcellated function aware cortical substructures

Finally, the subregions of the cortex are parcellated automatically using the high resolution mouse cerebellum atlas database published by Ullmann *et al.* [80]. This atlas contains a single average image which divides the cerebellum into multiple regions based on their neuronal function. As previously, I use the approach by Chakravarty *et*

al. [131] to obtain image specific parcellations based on a leave-one-out label propagation scheme, and apply the multi-atlas label fusion framework to improve the parcellation accuracy [214] (Fig. 5.9). Similarly to studies on human cortical thickness [238], I measure the average thickness (T_{GM}, T_{Gran} and T_{Mol}) in regions of interest.

5.3 Experimental data and validation

The proposed method is assessed on a data set which includes 28 active stained gadolinium-enhanced *ex vivo* T2* MRI of mouse brain, scanned at the age 18-21 weeks. 14 of them are transchromosomic mice that model Down's Syndrome (Tc1 model) [44]; the other 14 are wildtype littermate controls. Mice were perfusion-fixed and decapitated, with intact skull post-fixed in a solution of 4% formal saline and 8mM Gd-GTPA for 9 weeks [21].

The image data was acquired and provided by my colleague Ben Sinclair at Centre for Advanced Biomedical Imaging (CABI) and has been previously published [197]. For the sake of completeness, the imaging acquisition protocol and scanning parameters are included and described in detail here. Imaging was performed with a 9.4T VNMRS horizontal bore scanner (Agilent Inc) with a 26mm quadrature volume coil using a 3D spoiled gradient echo sequence. The scanning parameters are: $TE = 4.03ms$, $TR = 17ms$, $FA = 52^o$, $FOV = 20.48 \times 13.04 \times 13.04mm^3$, $matrix = 512 \times 216 \times 326$, $average = 6$.

All theses images have been processed through the pipeline of the framework described in the previous sections. Full cortical grey matter thickness as well as the thickness of granular layer and molecular layer were defined on the Purkinje layers for each image. I compare the thicknesses between the wild type and the transchromosomic groups for each parcellated function region. Furthermore, to understand the relationship between layer thickness, surface area and grey matter volume, I also measured the grey matter volume and the Purkinje layer surface area of each parcellated grey matter region.

It has been shown that normalisation to the TBV for certain morphological analysis is necessary for regressing out factors such as normal development and strain difference [239, 138], and may improve the classification power of quantitative analysis of image studies for neurodegenerative diseases such as AD and DS [240, 241]. As a

result, I also normalised the results to the TBV in this study to regress out the effect of the brain size vaiability. The multiple comparisons are corrected with a false discovery rate set to $q = 0.1$.

5.4 Results

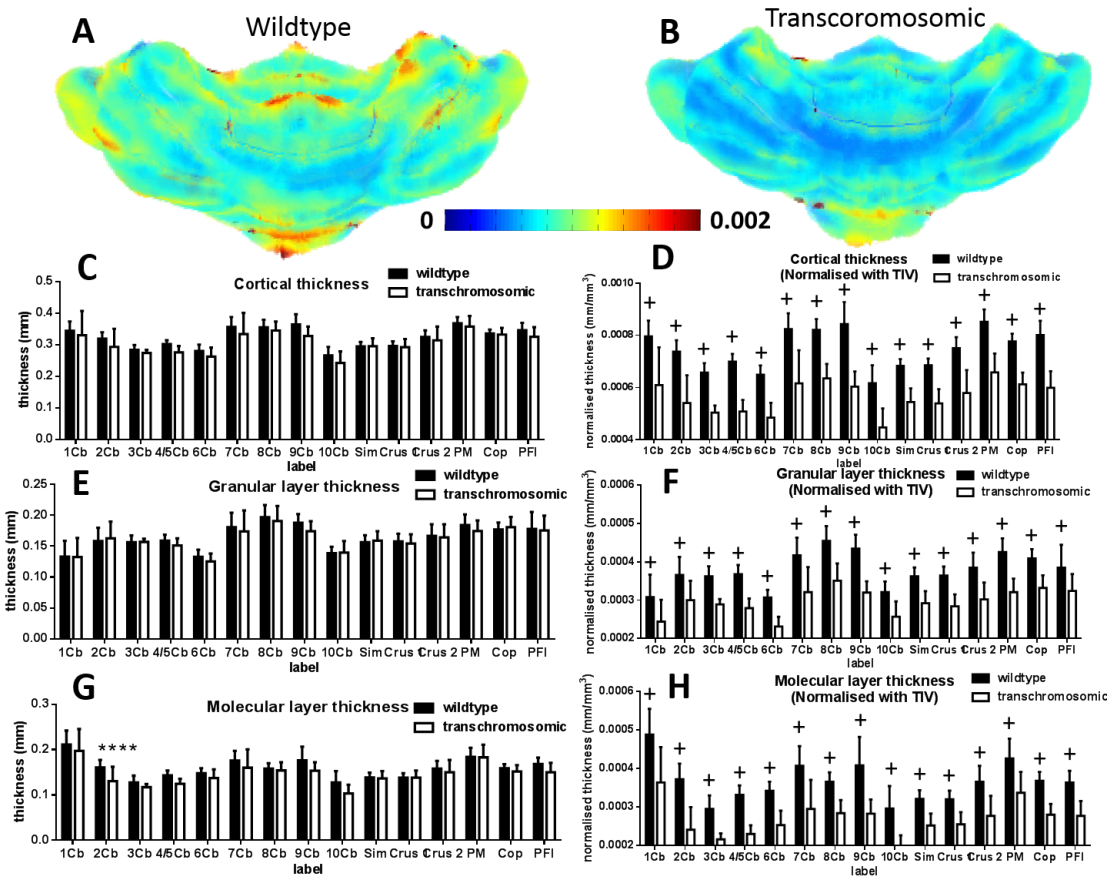

Figure 5.10: Normalised grey matter thickness map of (A) wildtype group, and (B)transgenic group, as well as the thicknesses of parcellated regions of (C,D) the full cortical, (E,F) granular layer and (G,H) the molecular layer before and after normalisation with TBV. Error bars represent the standard deviations. (**: $0.001 < p \leq 0.01$; +: $p \leq 0.0001$)

To quantitatively evaluate the framework, I compared the regional average thickness between the transchromosomic group and the wild type group, including the full cortical thickness, as well as the granular and molecular sublayer thicknesses (Fig. 5.10). The full name of the corresponding cerebellar structure is shown in Table 5.1. There is a tendency of thinner cortex for the transchromosomic group compared to that of the wild type group, although the difference is not significant (Fig. 5.10 C). When looking at the individual sublayers, a significant thickness difference was

Table 5.1: The abbreviations and their corresponding cerebellar structures.

Abbreviation	Cerebellar structures
Lobules of cerebellar vermis	
1Cb	Lobule 1
2Cb	Lobule 2
3Cb	Lobule 3
4/5Cb	Lobule 4/5
6Cb	Lobule 6
7Cb	Lobule 7
8Cb	Lobule 8
9Cb	Lobule 9
10Cb	Lobule 10
Lobules of cerebellar hemispheres	
Sim	Simple lobule
Crus 1	Crus 1 of the ansiform lobule
Crus 2	Crus 2 of the ansiform lobule
PM	Paramedian lobule
Cop	Copula of the pyramis
PFl	Paraflocculus

found at the lobule 2 of the cerebellar vermis (2Cb) (Fig.5.10 E) in the molecular layer.

I also normalised the regional average thicknesses with TBV to regress out the effect of the gross brain difference. The TBV of the transchromosomic mice is larger than the wild type littermate. After the normalisation, the thickness difference becomes significant over all subregions for the full cortex as well as both sublayers.

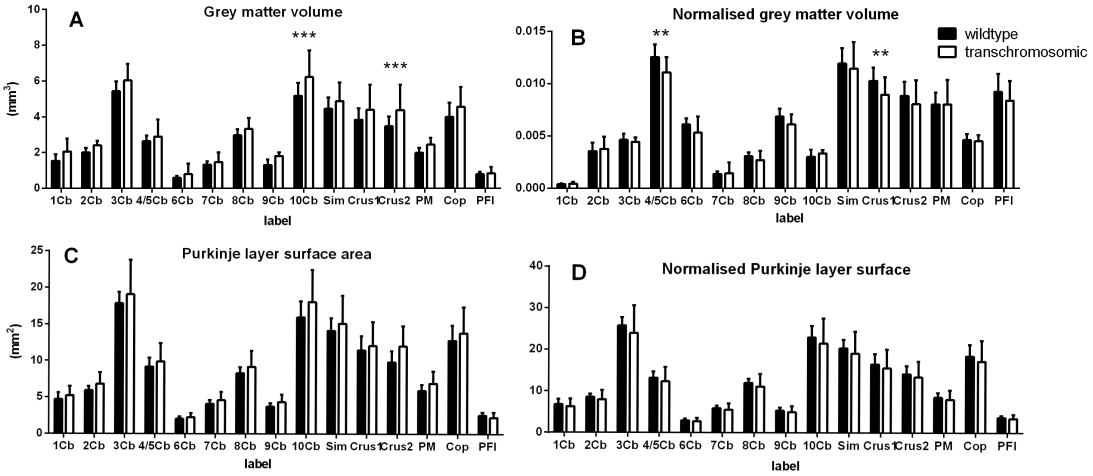

Figure 5.11: (A,B) Volume of the parcellated grey matter subregions (C,D) Surface area of Purkinje layer. Both measurements includs absolute value (on the left) and the value normalised with TBV (on the right). Error bars represent the standard deviations. (**: $0.001<p\leq0.01$; ***: $0.0001<p\leq0.001$)

In order to understand the relationship between the cortical thicknesses, cortical volume and cortical surface area, I also estimate the results of cortical volumes and surface areas of the extracted Purkinje layer Fig. 5.11. The absolute cortical volume of the transchromosomic group appears to be larger than their wildtype littermate, with significant different found at the Lobule 10 of cerebellar vermis (10Cb) and the Crus 2 of the ansiform lobule (Crus2). The larger volumes and longer surface areas of the Purkinje layer in the mice that model Down's Syndrome suggests that the morphology of the cerebellar grey matter of the transchromosomic mice could be more convoluted than the one of the wild type mice. After normalisation to the TBV, the cortical volume becomes smaller with significant difference found at the Lobule 4/5 of cerebellar vermis (4/5Cb) and the Crus 1 of the ansiform lobule (Crus1), and no significant difference remains for Purkinje layer surface after normalisation.

5.5 Discussion

The results of this study agree with the finding from the previous study by Baxter *et al.* [174] using a different mouse model of Down's Syndrome, in which they also found thickness reduction of granular layer and molecular layer in the transchromosomic group by measuring histological slides. Baxter *et al.* used the Ts65Dn mouse model in their study which is a segmentally trisomic for the distal 12-15 Mb of mouse chromosome 16, while in this study, the Tc1 mouse model is used which has an aneuploid mouse line that stably transmits an integrated copy of human chromosome 21 (Hsa21). A known morphological difference of the two mouse models is that, the TBV of mice of Ts65Dn model is smaller than that of the wild type littermate, while TBV of the mice of Tc1 model has a larger TBV [197].

5.5.1 Cortical laminar layer modeling

Studies of cortical cytoarchitecture and myeloarchitecture revealed the cortex (both in cerebrum and cerebellum) as a laminar structures [242, 243]. In order to get a good estimation of cortical thickness, it is necessary to derive a cortical laminar model which follows the actual anatomical arrangement. Since the study by Jones *et al.* [33], in which he introduced a model of the laminar layer using Laplacian field, studies have been trying to create better laminar models to improve the representation of the under-

lying true anatomical arrangement of the cortex [2, 37].

5.5.2 Cortical surface representation

In order to perform quantitative statistical analyses, many studies of surface-based cortical morphology (most of which use Freesurfer) represent the cortical thickness/area on the GM/WM surface and/or pial surface [146, 244, 145]. In studies using voxel-based cortical thickness, the voxel-wised thickness map is commonly defined on the central surface (or the mid-cortical surface), which is the half-way of the WM/GM surface and GM/CSF surface [151, 154]. However, these models are limited by the spatial variability in sublayer thickness within the grey matter (see in Fig. 5.12 A).

The information of the Purkinje layer location could provide robust information consistently across subjects, when performing groupwise analyses (Fig. 5.12 B). It is reasonable to conclude that the Purkinje-layer-based surface representation provide a more plausible data-driven anatomical-based representation of the cortical surfaces compared to previous studies [245, 154].

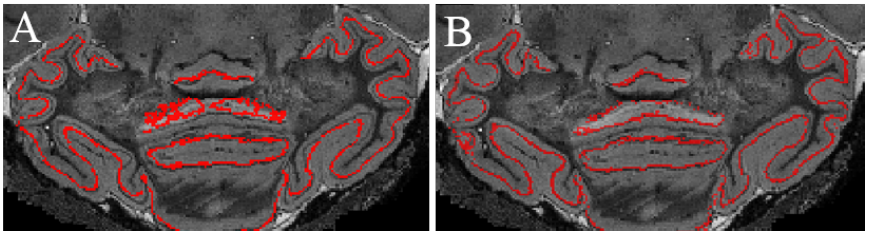

Figure 5.12: Comparison between the (a) central surface and (b) the Purkinje layer.

5.5.3 The effect and choice of normalisation for quantitative morphological analysis

There are controversial views regarding whether the morphological measurement should be regressed out through normalisation with TBV. Westman *et al.* [240] have used Freesurfer pipeline to investigate the choice normalisation with TBV for different morphological analysis in a study of Alzheimer's disease (AD). The conclusion of the study supports normalisation of volume measurement with TBV, but does not support the normalisation of thickness measurement. On the other hand, in another AD study, Zhou *et al.* [241] came to the entirely opposite conclusion that, to obtain the best classification power, the cortical thickness should be normalised with TBV, while such

regression is not necessary for volumetric analysis.

5.5.4 Volumetric, areal and surface analysis

Finally, the result in Fig. 5.11 showed interesting relationship among cortical volume, surface area and cortical thickness. The study by Meyer *et al.* [145] showed that the cortical surface area and cortical thickness demonstrated different structural properties. Winkler *et al.* [246, 247] compared the quantitative analysis of grey matter volume, surface area and cortical thickness and concluded that they are genetically and phenotypically independent biomarkers. The morphology of the cortex could vary even when the local volume remains constant. For example, the cortical thickness would be decreased if the cortical surface area increases. Meyer *et al.* [145] reviewed controversial opinions from different studies regarding the effect of between cortical thickness change towards cortical function and conjecture that, the thinning of the grey matter might be due to the increased volume of white matter as a need to establish denser connection between different functional regions (primary and secondary auditory regions in this case).

5.6 Conclusion

In conclusion, in this chapter, I presented a framework of thickness estimation for mouse cerebellar cortical sublayers - the granular layer and the molecular layer. The sublayer measurements were achieved through a 2-step extraction of the middle cell layer that sits in between them - the Purkinje layer. The proposed method managed to extract the layer information from both the image contrast, as well as the prior anatomical information. The framework has been evaluated on a gadolinium enhanced *ex vivo* MRI data of a mouse model of Down's Syndrome. The result suggests a reduction in regional average thickness in mice that model Down's Syndrome. The framework also managed to detect a significant local thickness reduction in the region of 2Cb in the molecular layer, which could not be found through gross estimation of the full cortical thickness, probably due to its limited sensitivity.

Chapter 6

Discussion

Quantitative analysis techniques for medical imaging have been continuously advancing in recent years, playing increasingly important roles in translational medical and biomedical research. Here the word "*translational*" represents the procedure of applying theoretical methods to medical applications. Specifically, there have been continuous efforts to establish "*quantitative imaging biomarkers*" which utilise medical image computing as a powerful tool for diagnosis, monitoring of disease progression or treatment effect in the clinic [248, 249]. However, when it comes to disease prognosis (e.g. predicting disease stage of Alzheimer's Disease), big training datasets with prior classification knowledge are required to compensate for large variations within the population.

On the other hand, the preclinical neuroimaging provide an alternative choice when investigating specific disease properties or phenotypes. Preclinical neuroimaging in phenotyping research is the T_0 phase of a translational study. Its advantage is the specificity, where inbred genetically modified mice (through gene knock-in/knock-out, or transchromosome modification, such as the Tc1 mouse that models Down's syndrome and rTg4510 mouse that models the Tau pathology in Alzheimer's Disease discussed in this thesis) ensure minimal morphological variation from uncontrolled factors. The animal models used in studies of this thesis (Tc1 and rTg4510) use human genomic segment or chromosome to model the syndromes and phenotypes of human disease. However, giving the complexity of the biological system, the inter-species anatomical and functional variations can affect the efficacy of using animal model to represent human diseases. Consequently, drug treatments which works on animal might become less effective in the later phases of translational study or produce additional

side effects. However, animal model did provide the best alternative solution prior to the clinical trials.

In addition, the morphological difference between the animal brain and human brain sometimes imposes challenges to adapt the current clinical image analysis tools to study the preclinical neuroimages. With the increasing amount of preclinical image data, there is a growing demand for image processing and analysis tools specifically aimed at small animals for high-throughput phenotyping.

Structural volume is one of the fundamental quantitative imaging biomarkers investigated in imaging studies, which requires structural parcellation. The study presented in Chapter 3 is one of the first studies to demonstrate the successful application of multi-atlas label fusion methods to the parcellation of mouse brain MRI, evaluated on both *in vivo* and *ex vivo* data. The resulting framework has been successfully applied to studies of disease progression (Section 3.5.1), and potential drug treatment (Section 3.5.2 and [201]). However, the accuracy of current preclinical image structural parcellation frameworks are still limited by the small number of atlases available in a database (five for NUS *in vivo* atlas [79], ten for the MRM NeAt atlas both *in vivo* [3] and *ex vivo* atlas [78]), and by the fact that each data database contains only a single imaging modality ((T2 weighted for NUS atlas and T2 weighted for MRM NeAt atlas) and a single strain (C57BL/6J). Constructing additional atlas databases including more individual samples and with diverse phenotypes and disease models would be beneficial. However, it would be very difficult to build specific atlases for each disease types or phenotype. A potential solution worth exploring would be the multi-step progressive label propagation methods I reviewed in Section 2.1.3.7, which explore the intrinsic distance within atlas or study groups to improve the parcellation accuracy of preclinical images.

Cortical thickness is another imaging biomarker that has been heavily investigated, which is based upon cortical segmentation from the result of structural parcellation or tissue segmentation. In current cortical thickness studies, the cerebellum has always been neglected despite its correlation with various neurodegenerative diseases, due to the highly convoluted nature of human cerebellum, coupled with the limitation of resolution in current clinical scanners. The morphology of the mouse cerebellum is relatively less convoluted, similar to the morphology of the human cerebral cortex. In

addition, compared to clinical neuroimaging, it is more feasible to acquire *ex vivo* images in preclinical studies, ensuring higher resolution images. Both of these advantages make the mouse brain *ex vivo* MRI a perfect candidate to study the cerebellar cortical thickness phenotyping.

The use of active staining techniques [20, 21] changes the intensities of specific tissue or cell types (in this case, the middle Purkinje layer of the cerebellar cortex) and enhance its image contrast to the surrounding tissues, which is beneficial for processing steps such as image registration, and also makes it possible to extract sublayers in the cerebellar cortex. However, the resulting signal intensity of the Purkinje layer cells in contrast enhanced MRI is very similar to the white matter intensity, which adversely complicates the intensity distribution of grey/white matter, and special care has to be taken when segmenting the white matter in an expectation maximisation approach. As presented in Chapter 5, this issue is solved by assigning multiple tissue classes to the grey matter in the Gaussian mixture model of tissue intensity distribution to avoid misclassification of Purkinje layer voxels to the white matter. The tissue classification within the grey matter will not be accurate using this method, due to the limited thickness and convoluted nature. As a result, the actual segmentation of the middle Purkinje layer is derived in a separate step through an entirely different approach. In addition, the tissue map of grey matter and white matter created and presented in Chapter 5 will also be useful as prior information for future studies.

There are many other problems specific to preclinical imaging. For instance, when designing an experiment, the choice has to be made whether to scan images *ex vivo*, favouring the benefit of higher resolution and SNR/CNR, or instead choose *in vivo* imaging instead, exploiting its advantages of better representation of the physiological situation in living organs, the ability to perform longitudinal studies and reduced animal usage. Part of this issue has been addressed from the perspective of volumetric analysis in Chapter 4.

There are other examples of challenges specific to preclinical imaging which is not the focus of this thesis but are also potentially interesting to address. These include the unique gradient distortion profile and intensity inhomogeneity in the cases of parallel multiple animal scans for high-throughput imaging [250], or the effect of anaesthesia on the animals' physiology and functionality due to the lack of conscious movement [19].

The studies presented in this thesis are specifically focused on mouse brain MRI. However, the application of the underlying methods is not restricted. Successful application of the current framework on rat brain has been demonstrated in Chapter 3. It can also be used for applications such as mouse embryo full body organ segmentation [251]. The sublayer thickness estimation method can also potentially be applied to other 3D biomedical images with contrast revealing multi-layer structures, such as retina layers in high resolution MRI or optical coherence tomography (OCT) [252].

Nevertheless, the studies introduced in this thesis only cover small aspects of preclinical neuroimage analysis for translational studies. Studies should also investigate other image analysis methods which could potentially be used as quantitative imaging biomarkers, such as voxel-/tensor-based morphology (V/TBM) [4], the boundary shift integral (BSI) [253], or the statistical shape models [254]. It would also be advantageous to include more MRI imaging modalities for preclinical image analysis, such as structural MR images (T1, T2, diffusion) and functional MR data (BOLD and ASL). In addition, other medical imaging techniques also play important roles in basic science research. These include molecular imaging methods including SPECT, PET/CT and PET/MRI, 3D volumetric histological slices reconstruction technique such as serial block face scanning electron microscopy (SBFSEM) [255] and high-resolution electron microscopy imaging (HREM) [256], and also optical imaging techniques like Optical projection tomography (OPT)[257] and CLARITY [258]. Studies about multimodality information fusion would then be necessary to extract more information from images acquired from those different modalities.

Chapter 7

Conclusion

This thesis presents the development and application of methods and tools targeting the purpose of automated quantitative morphometry for mouse brain MRI. With the development of phenotyping research in preclinical studies, the amount of neuroimage data is constantly increasing. There is an urgent demand for high-throughput image processing and quantitative analysis tools to extract useful information from these data. However, most of the current efforts in the neuroimage computing field focus on the clinical setting. The purpose of the studies presented in this thesis is to provide solutions tailored for preclinical neuroimage data, particularly focusing on structural parcellation and cortical thickness estimation.

Chapter 3 presented an integrated framework for automatic structural parcellation of mouse brain MRI, which assembles the state-of-the-art algorithms of label propagation (through affine registration based on symmetric block-matching and non-rigid registration based on symmetric fast free form deformation) and label fusion (with local-similarity-based atlas ranking and underlying true segmentation estimation). Validation on both *in vivo* and *ex vivo* data showed high accuracy of the framework, even with the limited amount of available mouse brain databases. An accurate structural parcellation is the fundamental step for many morphological analyses, such as volumetric analysis and cortical thickness estimation.

In Chapter 4, the thoroughly evaluated mouse brain structural parcellation framework was used to investigate the effect of one of the most important choices during the experimental design of preclinical neuroimaging - whether to scan the animal alive or post-mortem - by comparing the volumetric analysis result on *in vivo* and *ex vivo* mouse brain MRI data obtained from the same cohort. Volumetric analysis is a widely

used biomarker of neuroimage analysis for neurodevelopment and neurodegeneration. Imaging animals alive provides better correspondence to the actual *in vivo* physiology and enables longitudinal studies, but suffers from limited resolution and artefacts such as motion distortion; while post-mortem animals scans are inevitably affected by the structural volume change due to tissue processing despite their "high image quality" in terms of resolution and limited artefacts. The parcellation results indeed demonstrated a systematic tissue shrinkage from *in vivo* to *ex vivo* image data. However, the shrinkage is not uniformly distributed across all the brain structures. In particular, the ventricles collapse almost entirely from *in vivo* to *ex vivo*. However, both *in vivo* and *ex vivo* data demonstrated similar statistical power when analysing the parcellated structural volumes of transgenic animals that models human Tau pathology to differentiate between those with or without treatment.

With the brain structure labels available from the automated structural parcellation process, Chapter 5 describes a framework to investigate the cortical sublayer thickness of a specific structure - mouse cerebellum. The cortical thickness has been shown to provide more specific morphological insight than simple volumetric analysis, but the cerebellar cortex has often been ignored in previous research, although studies have shown evidence of its correspondence to various neurological functions. The framework managed to take advantage of the image contrast between the three sublayers of cerebellar cortex to segment the middle Purkinje layer and estimate the thickness of the outermost molecular layers and the innermost granular layer. Application of the framework to the experimental data demonstrated sensitivity to distinguish sublayer thickness variations between transgenic strains and their wildtype littermates, which cannot be detected from the full cortical thickness alone.

7.1 Future work

7.1.1 Structural parcellation using multi-modal imaging

Several approaches have been reviewed in Chapter 2 to improve the accuracy of atlas-based structural parcellation. Among them, the multi-atlas label fusion method has been explored in Section 2.1.2, and demonstrated improved parcellation accuracy. Another approach, which uses multispectral MR images as prior information to improve the parcellation result [1, 90], hasn't been explored in this thesis. Recent research has

emerged to integrate multi-modality images into multi-atlas frameworks, to explore the potential to feed information from different image types to label fusion and further improve the parcellation accuracy [100, 259, 260]. Some of the current available mouse brain atlases include multiple modalities: MacKenzie-Graham *et al.* [76] introduced an atlas database which includes MRI, block face histology and immunohistochemistry data; Badea *et al.* [92] also published an atlas database include T1, T2, T2* and Nissl staining. However, all these atlas only include a single sample. This can be improved either through building new multi-atlas multi-modality datasets, or when not possible, using the method introduced by Chakravarty *et al.* [205] to automatically build an intermediate template.

7.1.2 Cortical cytoarchitecture and myelination

In this thesis, I explored the segmentation of cortical laminar layers by utilising the intensity profile obtained with an active staining technique using a Gadolinium-based contrast enhancing agent. Recently, there have been an increasing number of studies which focus on investigating the intrinsic MR intensity profile of the cortical tissue. Some of these studies show that both the T1W/T2W map or the R1 map (1/T1) can reveal the cortical cytoarchitecture and myelination (sometimes referred to as the myelin map) [261, 262, 263, 243]. Other studies have also confirmed that the myelin density varies across cortical areas, and is independent of the cortical folding pattern [264, 242]. In addition, incorporating myelin maps can also improve the cortical surface registration [265]. However, with the current typical image resolution of structural MRI (1mm for clinical MRI, $100\mu m$ for *in vivo* mouse brain MRI and $40\mu m$ for *ex vivo* mouse brain MRI), image morphometric methods can only make sense on data following a top-down approach: processing and analysing the signals at the macrostructure level to infer the underlying microstructure features. Dinse *et al.* [266] introduced a bottom-up approach to model the MR image intensity information from the microstructural priors, and demonstrated the ability of this method to distinguish area specific signal profiles in human cortex in ultra-high resolution *in vivo* brain MRI. Enlightened by these studies, it would be interesting to extend the current cerebellar sublayer cortical thickness study to further investigate the underlying cortical cytoarchitecture and myelination.

Appendix A

List of Abbreviations

AD: Alzheimer's Disease

ANOVA: analysis of variance

CC: Cross Correlation

CNR: Contrast to Noise Ratio

CSF: Cerebralspinal Fluid

CRUISE: Cortical reconstruction using implicit surface evolution

CT: computed tomography

DSC: Dice Similarity Coefficient

DS: Down's Syndrome

DTPA-Gd: Diethylenetriaminepentaacetic acid-gadolinium

EM: Expectation Maximisation

FDR: False Discovery Rate

FSL: FMRIB's Software Library

GCC: global cross correlation

GM: Grey Matter

GNCC: Global Normalised Cross Correlation

H&E: hematoxylin and eosin

LAR: Least Angle Regression

NMI: Normalised Mutual Information

NCC: Normalised Cross Correlation

MI: Mutual Information

MRF: Markov Random Field

MRI: Magnetic Resonance Imaging

MRM NeAt: Magnetic Resonance Microimaging Neurological Atlas

MSD: Mean Square Difference

NUS: National University of Singapore

OCT: optical coherence tomography

PDE: Partial Differential Equation

PET: Positron Emission Tomography

ROI: Region Of Interest

SIMPLE: Selective and Iterative Method for Performance Level Estimation

SNR: Signal to Noise Ratio

SPECT: Single-Photon Emission Computed Tomography

SSD: Sum of Square Difference

SPM: Statistical Parametric Mapping

STAPLE: Simultaneous Truth and Performance Level Estimation

STEPS: multi-label Similarity and Truth Estimation for Propagated Segmentations

TBM: Tensor-Based Morphometry

TIV: Total Intracranial Volume

VBM: Voxel-Based Morphometry

WM: White Matter

WMD: distance towards the white matter surface

Bibliography

[1] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. V. D. Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, A. M. Dale, and A. van der Kouwe, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–55, 2002.

[2] M. Waehnert, J. Dinse, M. Weiss, M. Streicher, P. Waehnert, S. Geyer, R. Turner, and P.-L. Bazin, "Anatomically motivated modeling of cortical laminae," *Neuroimage*, vol. 93, pp. 210–220, 2014.

[3] Y. Ma, D. Smith, P. R. Hof, B. Foerster, S. Hamilton, S. J. Blackband, M. Yu, and H. Benveniste, "In vivo 3D digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy," *Frontiers in neuroanatomy*, vol. 2, no. 1, 2008.

[4] J. Ashburner and K. J. Friston, "Voxel-based morphometry – the methods," *NeuroImage*, vol. 11, no. 6, pp. 805–21, 2000.

[5] K.-K. Shen, J. Fripp, F. Mériaudeau, G. Chételat, O. Salvado, and P. Bourgeat, "Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models," *NeuroImage*, vol. 59, no. 3, pp. 2155–2166, 2012.

[6] M. Styner, I. Oguz, S. Xu, C. Brechbühler, D. Pantazis, J. J. Levitt, M. E. Shenton, and G. Gerig, "Framework for the Statistical Shape Analysis of Brain Structures using SPHARM-PDM," *Insight Journal*, no. 1071, pp. 242–250, 2006.

[7] G. Calmon and N. Roberts, "Automatic measurement of changes in brain volume on consecutive 3d MR images by segmentation propagation," *Magnetic Resonance Imaging*, vol. 18, no. 4, pp. 439–453, 2000.

[8] K. K. Leung, M. J. Clarkson, J. W. Bartlett, S. Clegg, C. R. Jack, M. W. Weiner, N. C. Fox, and S. Ourselin, "Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection," *NeuroImage*, vol. 50, pp. 516–523, 2010.

[9] M. J. Clarkson, M. J. Cardoso, G. R. Ridgway, M. Modat, K. K. Leung, J. D. Rohrer, N. C. Fox, and S. Ourselin, "A comparison of voxel and surface based cortical thickness estimation methods," *Neuroimage*, vol. 57, no. 3, pp. 856–865, 2011.

[10] S. Batzoglou, L. Pachter, J. P. Mesirov, B. Berger, and E. S. Lander, "Human and mouse gene structure: comparative analysis and application to exon prediction," *Genome Research*, vol. 10, no. 7, pp. 950–958, 2000.

[11] S. D. M. Brown and M. W. Moore, "The International Mouse Phenotyping Consortium: Past and future perspectives on mouse phenotyping," *Mammalian Genome*, vol. 23, no. 9-10, pp. 632–640, 2012.

[12] A. M. Mallon, A. Blake, and J. M. Hancock, "EuroPhenome and EMPReSS: Online mouse phenotyping resource," *Nucleic Acids Research*, vol. 36, no. Suppl. 1, 2008.

[13] E. A. Boudreau, G. Chen, X. Li, and C. D. Kroenke, "Magnetic resonance imaging approaches for studying alcoholism using mouse models," *Alcohol Research & Health*, vol. 31, no. 3, pp. 247–248, 2008.

[14] G. Gerig, I. Oguz, S. Gouttard, J. Lee, H. An, W. Lin, M. McMurray, K. Grewen, J. Johns, and M. A. Styner, "Synergy of image analysis for animal and human neuroimaging supports translational research on drug abuse," *Frontiers in Psychiatry*, vol. 2, 2011.

[15] O. Beuf, F. Jaillon, and H. Saint-Jalmes, "Small-animal MRI: Signal-to-noise ratio comparison at 7 and 1.5 T with multiple-animal acquisition strategies,"

Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 19, no. 4, pp. 202–208, 2006.

[16] H. Benveniste and S. Blackband, "MR microscopy and high resolution small animal MRI: Applications in neuroscience research," *Progress in Neurobiology*, vol. 67, no. 5, pp. 393–420, 2002.

[17] X. Hu and D. G. Norris, "Advances in high-field magnetic resonance imaging," *Annual Review of Biomedical Engineering*, vol. 6, no. 1, pp. 157–184, 2004.

[18] F. D. Doty, G. Entzminger, J. Kulkarni, K. Pamarthy, and J. P. Staab, "Radio frequency coil technology for small-animal MRI," *NMR in Biomedicine*, vol. 20, no. 3, pp. 304–325, 2007.

[19] I. J. Hildebrandt, H. Su, and W. A. Weber, "Anesthesia and other considerations for *in vivo* imaging of small animals," *ILAR journal*, vol. 49, no. 1, pp. 17–26, 2008.

[20] A. Petiet and G. Johnson, "Active staining of mouse embryos for magnetic resonance microscopy," in *Histology Protocols* (T. D. Hewitson and I. A. Darby, eds.), vol. 611 of *Methods in Molecular Biology*, pp. 141–149, Humana Press, 2010.

[21] J. O. Cleary, F. K. Wiseman, F. C. Norris, A. N. Price, M. Choy, V. L. J. Tybulewicz, R. J. Ordidge, S. Brandner, E. M. C. Fisher, and M. F. Lythgoe, "Structural correlates of active-staining following magnetic resonance microscopy in the mouse brain," *NeuroImage*, vol. 56, no. 3, pp. 974–83, 2011.

[22] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, Gunnar Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, pp. 685–691, Apr. 2008.

[23] R. S. Balaban and V. A. Hampshire, "Challenges in small animal noninvasive imaging," *ILAR journal*, vol. 42, no. 3, pp. 248–262, 2001.

[24] J. G. F. Baillarger, *Recherches sur la structure de la couche corticale des circonvolutions du cerveau*. J.-B. Baillière, 1840.

[25] C. F. von Economo, G. N. Koskinas, and L. C. Triarhou, *Atlas of cytoarchitectonics of the adult human cerebral cortex*. Karger Basel, 2008.

[26] D. Barazany and Y. Assaf, "Visualization of cortical lamination patterns with magnetic resonance imaging," *Cerebral Cortex*, vol. 22, no. 9, pp. 2016–2023, 2012.

[27] A. Sudarov and A. L. Joyner, "Cerebellum morphogenesis: the foliation pattern is orchestrated by multi-cellular anchoring centers," *Neural development*, vol. 2, no. 26, pp. 8306–8319, 2007.

[28] V. B. Mountcastle, "Modality and topographic properties of single neurons of cat's somatic sensory cortex," *Journal of neurophysiology*, vol. 20, no. 4, pp. 408–434, 1957.

[29] W. Calvin, "Cortical columns, modules, and hebbian cell assemblies," in *The Handbook of Brain Theory and Neural Networks* (M. A. Arbib, ed.), pp. 269–272, MIT Press, 1995.

[30] L. Cruz, S. V. Buldyrev, S. Peng, D. L. Roe, B. Urbanc, H. E. Stanley, and D. L. Rosene, "A statistically based density map method for identification and quantification of regional differences in microcolumnarity in the monkey brain," *Journal of Neuroscience Methods*, vol. 141, no. 2, pp. 321–332, 2005.

[31] J. DeFelipe, L. Alonso-Nanclares, and J. I. Arellano, "Microstructure of the neocortex: comparative aspects," *Journal of neurocytology*, vol. 31, no. 3-5, pp. 299–316, 2003.

[32] D. P. Buxhoeveden and M. F. Casanova, "The minicolumn hypothesis in neuroscience," *Brain*, vol. 125, no. Pt 5, pp. 935–951, 2002.

[33] E. G. Jones, "Microcolumns in the cerebral cortex," *Proceedings of the National Academy of Sciences*, vol. 97, no. 10, pp. 5019–5021, 2000.

[34] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis. I. Segmentation and surface reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.

[35] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system," *Neuroimage*, vol. 9, no. 2, pp. 195–207, 1999.

[36] A. J. Yezzi and J. L. Prince, "An Eulerian PDE approach for computing tissue thickness," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1332–1339, 2003.

[37] Y. Leprince, F. Poupon, T. Delzescaux, D. Hasboun, C. Poupon, and D. Riviere, "Combined Laplacian-equivolumic model for studying cortical lamination with ultra high field MRI (7 T)," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 580–583, 2015.

[38] J. Stewart and B. Kolb, "The effects of neonatal gonadectomy and prenatal stress on cortical thickness and asymmetry in rats," *Behavioral and Neural Biology*, vol. 49, no. 3, pp. 344–360, 1988.

[39] A. Y. Hardan, S. Muddasani, M. Vemulapalli, M. S. Keshavan, and N. J. Minshew, "An MRI study of increased cortical thickness in autism," *American Journal of Psychiatry*, vol. 163, no. 7, pp. 1290–1292, 2006.

[40] J. P. Lerch, J. B. Carroll, A. Dorr, S. Spring, A. C. Evans, M. R. Hayden, J. G. Sled, and R. M. Henkelman, "Cortical thickness measured from MRI in the YAC128 mouse model of Huntington's disease," *NeuroImage*, vol. 41, no. 2, pp. 243–51, 2008.

[41] S. A. Meda, J. R. Pryweller, and T. A. Thornton-Wells, "Regional brain differences in cortical thickness, surface area and subcortical volume in individuals with Williams syndrome," *PLoS ONE*, vol. 7, p. e31913, 2012.

[42] M. Grand'maison, S. P. Zehntner, M.-K. Ho, F. Hébert, A. Wood, F. Carbonell, A. P. Zijdenbos, E. Hamel, and B. J. Bedell, "Early cortical thickness changes predict β-amyloid deposition in a mouse model of Alzheimer's disease," *Neurobiology of Disease*, vol. 54, no. 0, pp. 59–67, 2013.

[43] K. Zhang and T. J. Sejnowski, "A universal scaling law between gray matter and white matter of cerebral cortex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 10, pp. 5621–6, 2000.

[44] A. O'Doherty, S. Ruf, C. Mulligan, V. Hildreth, M. L. Errington, S. Cooke, A. Sesay, S. Modino, L. Vanes, D. Hernandez, J. M. Linehan, P. T. Sharpe, S. Brandner, T. V. P. Bliss, D. J. Henderson, D. Nizetic, V. L. J. Tybulewicz, and E. M. C. Fisher, "An aneuploid mouse strain carrying human chromosome 21 with down syndrome phenotypes," *Science*, vol. 309, no. 5743, pp. 2033–2037, 2005.

[45] S. M. Gribble, F. K. Wiseman, S. Clayton, E. Prigmore, E. Langley, F. Yang, S. Maguire, B. Fu, D. Rajan, O. Sheppard, C. Scott, H. Hauser, P. J. Stephens, L. A. Stebbings, B. L. Ng, T. Fitzgerald, M. A. Quail, R. Banerjee, K. Rothkamm, V. L. J. Tybulewicz, E. M. C. Fisher, and N. P. Carter, "Massively parallel sequencing reveals the complex structure of an irradiated human chromosome on a mouse background in the Tc1 model of Down syndrome," *PLoS ONE*, vol. 8, no. 4, 2013.

[46] M. Ramsden, L. Kotilinek, C. Forster, J. Paulson, E. McGowan, K. SantaCruz, A. Guimaraes, M. Yue, J. Lewis, G. Carlson, M. Hutton, and K. H. Ashe, "Age-dependent neurofibrillary tangle formation, neuron loss, and memory impairment in a mouse model of human tauopathy (P301L)," *The Journal of neuroscience*, vol. 25, no. 46, pp. 10637–10647, 2005.

[47] K. Richards, C. Watson, R. F. Buckley, N. D. Kurniawan, Z. Yang, M. D. Keller, R. Beare, P. F. Bartlett, G. F. Egan, G. J. Galloway, G. Paxinos, S. Petrou, and D. C. Reutens, "Segmentation of the mouse hippocampal formation in magnetic resonance images," *NeuroImage*, vol. 58, no. 3, pp. 732–40, 2011.

[48] J. Zhou, S. Chang, Q. Liu, G. Pappas, V. Boronikolas, M. Michaelides, N. D. Volkow, P. K. Thanos, and D. Metaxas, "A novel learning based segmentation method for rodent brain structures using MRI," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI. IEEE International Symposium on*, pp. 61–64, IEEE, 2008.

[49] J. Zhou, S. Chang, S. Zhang, G. Pappas, M. Michaelides, F. Delis, N. Volkow, P. Thanos, and D. Metaxas, "Segmentation of rodent brains from MRI based on a novel statistical structure prediction method," in *Biomedical Imaging: From Nano to Macro. ISBI 2009. IEEE International Symposium on*, pp. 498–501, 2009.

[50] F. Van Der Lijn, M. De Bruijne, S. Klein, T. Den Heijer, Y. Y. Hoogendam, A. Van Der Lugt, M. M. B. Breteler, and W. J. Niessen, "Automated brain structure segmentation based on atlas registration and appearance models," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 276–286, 2012.

[51] S. Zhang, J. Huang, M. Uzunbas, T. Shen, F. Delis, X. Huang, N. Volkow, P. Thanos, and D. N. Metaxas, "3D segmentation of rodent brain structures using hierarchical shape priors and deformable models," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 611–618, Springer, 2011.

[52] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, "A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI," *IEEE Transactions on Medical Imaging,*, vol. 29, no. 9, pp. 1612–1625, 2010.

[53] J. Lee, J. Jomier, S. Aylward, M. Tyszka, S. Moy, J. Lauder, and M. Styner, "Evaluation of atlas based mouse brain segmentation," in *SPIE International Symposium on Medical Imaging*, pp. 725943–725943, International Society for Optics and Photonics, 2009.

[54] J. O. Cleary, M. Modat, F. C. Norris, A. N. Price, S. A. Jayakody, J. P. Martinez-Barbera, N. D. E. Greene, D. J. Hawkes, R. J. Ordidge, P. J. Scambler, S. Ourselin, and M. F. Lythgoe, "Magnetic resonance virtual histology for

embryos: 3D atlases for automated high-throughput phenotyping," *NeuroImage*, vol. 54, no. 2, pp. 769–78, 2011.

[55] T. Greitz, C. Bohm, S. Holte, and L. Eriksson, "A computerized brain atlas: construction, anatomical content, and some applications," *Journal of computer assisted tomography*, vol. 15, no. 1, pp. 26–38, 1991.

[56] R. Kikinis, M. E. Shenton, D. V. Losifescu, R. W. McCarley, P. Saiviroonporn, H. H. Hokama, A. Robatino, D. Metcalf, C. G. Wible, C. M. Portas, R. M. Donnino, and F. A. Jolesz, "A digital brain atlas for surgical planning, model-driven segmentation, and teaching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 2, no. 3, pp. 232–241, 1996.

[57] C. A. Cocosco, V. Kollokian, R. K. Kwan, G. B. Pike, and A. C. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," *NeuroImage*, vol. 5, no. 4, p. 425.

[58] M. J. Ackerman, "The visible human project," *Proceedings of the IEEE*, vol. 86, no. 3, pp. 504–511, 1998.

[59] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *Medical Imaging, IEEE Transactions on*, vol. 17, no. 3, pp. 463–468, 1998.

[60] R. Bajcsy, "Digital anatomy atlas and its registration to MRI, fMRI, PET: the past presents a future," in *Biomedical Image Registration*, vol. 2717, pp. 201–211, 2003.

[61] A. C. Evans, D. L. Collins, and B. Milner, "An MRI-based stereotactic atlas from 250 young normal subjects," *Society for Neuroscience Abstracts*, vol. 18, p. 408, 1992.

[62] A. C. Evans, D. L. Collins, S. Mills, E. Brown, R. Kelly, and T. M. Peters, "3D statistical neuroanatomical models from 305 MRI volumes," in *Nuclear Science Symposium and Medical Imaging Conference, 1993 IEEE Conference Record*, pp. 1813–1817, IEEE, 1993.

[63] C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *Journal of computer assisted tomography*, vol. 22, no. 2, pp. 324–333, 1998.

[64] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-d model-based neuroanatomical segmentation," *Human brain mapping*, vol. 3, no. 3, pp. 190–208, 1995.

[65] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM)," *NeuroImage*, vol. 2, no. 2, pp. 89–101, 1995.

[66] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, J. Feidler, K. Smith, D. Boomsma, H. H. Pol, T. Cannon, R. Kawashima, and B. Mazoyer, "A Four-Dimensional Probabilistic Atlas of the Human Brain," *Journal of the American Medical Informatics Association*, vol. 8, no. 5, pp. 401–430, 2001.

[67] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, no. 1, pp. 313–327, 2011.

[68] V. S. Caviness, J. Meyer, N. Makris, and D. N. Kennedy, "MRI-Based Topographic Parcellation of Human Neocortex: An Anatomically Specified Method with Estimate of Reliability," *Journal of Cognitive Neuroscience*, vol. 8, no. 6, pp. 566–587, 1996.

[69] B. Fischl, M. I. Sereno, R. B. H. Tootell, and A. M. Dale, "High-resolution intersubject averaging and a coordinate system for the cortical surface," *Human Brain Mapping*, vol. 8, no. 4, pp. 272–84, 1999.

[70] R. P. Woods, M. Dapretto, N. L. Sicotte, A. W. Toga, and J. C. Mazziotta, "Creation and use of a talairach-compatible atlas for accurate, automated, nonlinear

intersubject registration, and analysis of functional imaging data," *Human brain mapping*, vol. 8, no. 2-3, pp. 73–79, 1999.

[71] E. Chan, N. Kovacevic, S. Ho, R. Henkelman, and J. Henderson, "Development of a high resolution three-dimensional surgical atlas of the murine head for strains 129S1/SvImJ and C57Bl/6J using magnetic resonance imaging and micro-computed tomography," *Neuroscience*, vol. 144, no. 2, pp. 604–615, 2007.

[72] A. Dorr, J. P. Lerch, S. Spring, N. Kabani, and R. M. Henkelman, "High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult c57bl/6j mice," *Neuroimage*, vol. 42, no. 1, pp. 60–69, 2008.

[73] N. Kovacević, J. T. Henderson, E. Chan, N. Lifshitz, J. Bishop, a. C. Evans, R. M. Henkelman, X. J. Chen, N. Kovac, M. I. Centre, and M. B. Imaging, "A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability," *Cerebral cortex*, vol. 15, no. 5, pp. 639–45, 2005.

[74] M. Aggarwal, J. Zhang, M. I. Miller, R. L. Sidman, and S. Mori, "Magnetic resonance imaging and micro-computed tomography combined atlas of developing and adult mouse brains for stereotaxic surgery," *Neuroscience*, vol. 162, no. 4, pp. 1339–1350, 2009.

[75] N. Chuang, S. Mori, A. Yamamoto, H. Jiang, X. Ye, X. Xu, L. J. Richards, J. Nathans, M. I. Miller, A. W. Toga, R. L. Sidman, and J. Zhang, "An MRI-based atlas and database of the developing mouse brain," *NeuroImage*, vol. 54, no. 1, pp. 80–9, 2011.

[76] A. MacKenzie-Graham, E.-F. Lee, I. D. Dinov, M. Bota, D. W. Shattuck, S. Ruffins, H. Yuan, F. Konstantinidis, A. Pitiot, Y. Ding, G. Hu, R. E. Jacobs, and A. W. Toga, "A multimodal, multidimensional atlas of the C57BL/6J mouse brain," *Journal of anatomy*, vol. 204, no. 2, pp. 93–102, 2004.

[77] G. A. Johnson, A. Badea, J. Brandenburg, G. Cofer, B. Fubara, S. Liu, and J. Nissanov, "Waxholm space: an image-based reference for coordinating mouse brain research," *NeuroImage*, vol. 53, no. 2, pp. 365–72, 2010.

[78] Y. Ma, P. R. Hof, S. C. Grant, S. J. Blackband, R. Bennett, L. Slatest, M. D. McGuigan, and H. Benveniste, "A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy," *Neuroscience*, vol. 135, no. 4, pp. 1203–15, 2005.

[79] J. Bai, T. L. H. Trinh, K.-H. Chuang, and A. Qiu, "Atlas-based automatic mouse brain image segmentation revisited: model complexity vs. image registration," *Magnetic resonance imaging*, vol. 30, no. 6, pp. 789–98, 2012.

[80] J. F. P. Ullmann, M. D. Keller, C. Watson, A. L. Janke, N. D. Kurniawan, Z. Yang, K. Richards, G. Paxinos, G. F. Egan, S. Petrou, P. Bartlett, G. J. Galloway, and D. C. Reutens, "Segmentation of the C57BL/6J mouse cerebellum in magnetic resonance images," *NeuroImage*, vol. 62, no. 3, pp. 1408–1414, 2012.

[81] J. F. P. Ullmann, C. Watson, A. L. Janke, N. D. Kurniawan, and D. C. Reutens, "A segmentation protocol and MRI atlas of the C57BL/6J mouse neocortex," *NeuroImage*, vol. 78, no. 0, pp. 196–203, 2013.

[82] J. F. Ullmann, C. Watson, A. L. Janke, N. D. Kurniawan, G. Paxinos, and D. C. Reutens, "An MRI atlas of the mouse basal ganglia," *Brain Structure and Function*, pp. 1–11, 2013.

[83] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in medicine and biology*, vol. 46, no. 3, pp. R1–45, 2001.

[84] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[85] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. 23, pp. S151–S160, 2004.

[86] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 8, pp. 712–721, 1999.

[87] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, 2009.

[88] A. E. H. Scheenstra, J. Dijkstra, R. C. G. van de Ven, L. van der Weerd, and J. H. Reiber, "Automated segmentation of the *ex vivo* mouse brain," in *Medical Imaging*, p. 651106, International Society for Optics and Photonics, 2007.

[89] A. E. H. Scheenstra, R. C. G. Van De Ven, L. van der Weerd, A. M. van den Maagdenberg, J. Dijkstra, and J. H. C. Reiber, "Automated segmentation of *in vivo* and *ex vivo* mouse brain magnetic resonance images," *Molecular Imaging*, vol. 8, no. 1, pp. 35–44, 2009.

[90] A. A. Ali, A. M. Dale, A. Badea, and G. A. Johnson, "Automated segmentation of neuroanatomical structures in multispectral MR microscopy of the mouse brain," *NeuroImage*, vol. 27, no. 2, pp. 425–35, 2005.

[91] A. A. Sharief, A. Badea, A. M. A. Dale, and G. A. Johnson, "Automated segmentation of the actively stained mouse brain using multi-spectral MR microscopy," *NeuroImage*, vol. 39, no. 1, pp. 136–45, 2008.

[92] A. Badea, A. A. Ali-Sharief, and G. A. Johnson, "Morphometric analysis of the C57BL/6J mouse brain," *NeuroImage*, vol. 37, no. 3, pp. 683–93, 2007.

[93] A. Badea, G. a. Johnson, and R. W. Williams, "Genetic dissection of the mouse brain using high-field magnetic resonance microscopy," *NeuroImage*, vol. 45, no. 4, pp. 1067–79, 2009.

[94] M. H. Bae, R. Pan, T. Wu, A. Badea, M. Hyeok, R. Pan, T. Wu, A. Badea, and M. H. Bae, "Automated segmentation of mouse brain images using extended MRF," *NeuroImage*, vol. 46, no. 3, pp. 717–25, 2009.

[95] M. H. Bae, T. Wu, and R. Pan, "Mix-ratio sampling: Classifying multiclass imbalanced mouse brain images using support vector machine," *Expert Systems with Applications*, vol. 37, no. 7, pp. 4955–4965, 2010.

[96] T. Wu, M. H. Bae, M. Zhang, R. Pan, and A. Badea, "A prior feature SVM-MRF based method for mouse brain segmentation," *NeuroImage*, vol. 59, no. 3, pp. 2298–306, 2012.

[97] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–42, 2004.

[98] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE transactions on medical imaging*, vol. 23, no. 7, pp. 903–21, 2004.

[99] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Multiclassifier fusion in human brain MR segmentation: modelling convergence," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 815–822, Springer, 2006.

[100] J. Eugenio Iglesias, M. Rory Sabuncu, and K. Van Leemput, "A unified framework for cross-modality multi-atlas segmentation of brain MRI," *Medical Image Analysis*, vol. 17, no. 8, pp. 1181–1191, 2013.

[101] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern analysis and Applications*, vol. 1, no. 1, pp. 18–27, 1998.

[102] H. Wang, S. R. Das, J. W. Suh, M. Altinay, J. Pluta, C. Craige, B. Avants, and P. a. Yushkevich, "A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation," *NeuroImage*, vol. 55, no. 3, pp. 968–985, 2011.

[103] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.

[104] M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein, "Optimum template selection for atlas-based segmentation," *NeuroImage*, vol. 34, no. 4, pp. 1612–8, 2007.

[105] M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1714–1729, 2010.

[106] X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz-de Solorzano, "Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the combination oracle," in *Medical Imaging*, p. 69141W, International Society for Optics and Photonics, 2008.

[107] O. Commowick and G. Malandain, "Efficient selection of the most similar image in a database for critical structures segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pp. 203–210, Springer, 2007.

[108] T. Shen, H. Li, and X. Huang, "Active volume models for medical image segmentation," *IEEE transactions on medical imaging*, vol. 30, no. 3, pp. 774–791, 2011.

[109] P. Aljabar, R. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Classifier selection strategies for label fusion using large atlas databases," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pp. 523–531, Springer, 2007.

[110] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, and J. P. W. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE transactions on medical imaging*, vol. 29, no. 12, pp. 2000–2008, 2010.

[111] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[112] T. Rohlfing, D. B. Russakoff, and C. R. Maurer, "Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation," *IEEE transactions on medical imaging*, vol. 23, no. 8, pp. 983–94, 2004.

[113] K. K. Leung, J. Barnes, M. Modat, G. R. Ridgway, J. W. Bartlett, N. C. Fox, S. Ourselin, A. D. N. Initiative, *et al.*, "Brain maps: an automated, accurate and robust brain extraction technique using a template library," *Neuroimage*, vol. 55, no. 3, pp. 1091–1108, 2011.

[114] K. K. Leung, J. Barnes, G. R. Ridgway, J. W. Bartlett, M. J. Clarkson, K. Macdonald, N. Schuff, N. C. Fox, and S. Ourselin, "Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease," *NeuroImage*, vol. 51, no. 4, pp. 1345–59, 2010.

[115] X. Artaechevarria, A. Munoz-Barrutia, C. Ortiz-de Solorzano, A. Muñoz Barrutia, and C. Ortiz-de Solórzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," *IEEE transactions on medical imaging*, vol. 28, no. 8, pp. 1266–77, 2009.

[116] M. Agarwal, E. Hendriks, B. Stoel, M. Bakker, J. Reiber, and M. Staring, "Local SIMPLE multi atlas-based segmentation applied to lung lobe detection on chest CT," in *SPIE International Symposium on Medical Imaging*, pp. 831410–831410, International Society for Optics and Photonics, 2012.

[117] M. J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N. C. Fox, S. Ourselin, and D. Neuroimaging, "STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation," *Medical Image Analysis*, vol. 17, no. 6, pp. 671–684, 2013.

[118] M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin, "LoAd: a locally adaptive cortical segmentation algorithm," *NeuroImage*, vol. 56, no. 3, pp. 1386–97, 2011.

[119] S. K. Warfield, K. H. Zou, and W. M. Wells, "Validation of image segmentation by estimating rater bias and variance," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 366, no. 1874, pp. 2361–75, 2008.

[120] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 611–623, June 2013.

[121] H. Wang and P. a. Yushkevich, "Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation," *Frontiers in neuroinformatics*, vol. 7, no. November, p. 27, 2013.

[122] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, F. F. Berendsen, and J. P. W. Pluim, "Label fusion in multi-atlas based segmentation with user-defined local weights," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1480–1483, Ieee, mar 2011.

[123] M. J. Brusco and H.-F. Köhn, "Comment on "Clustering by passing messages between data points"," *Science*, vol. 319, no. 5864, p. 726; author reply 726, 2008.

[124] F. van der Lijn, T. den Heijer, M. M. B. Breteler, and W. J. Niessen, "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts," *NeuroImage*, vol. 43, no. 4, pp. 708–720, 2008.

[125] R. Wolz, P. Aljabar, J. Hajnal, A. Hammers, and D. Rueckert, "LEAP: learning embeddings for atlas propagation," *NeuroImage*, vol. 49, no. 2, pp. 1316–25, 2010.

[126] M. J. Cardoso, M. Modat, R. Wolz, A. Melbourne, D. Cash, D. Rueckert, and S. Ourselin, "Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 9, pp. 1976–1988, 2015.

[127] P. Aljabar, D. Rueckert, and W. R. Crum, "Automated morphological analysis of magnetic resonance brain imaging using spectral analysis," *NeuroImage*, vol. 43, no. 2, pp. 225–35, 2008.

[128] R. A. Heckemann, S. Keihaninejad, P. Aljabar, K. R. Gray, C. Nielsen, D. Rueckert, J. V. Hajnal, and A. Hammers, "Automatic morphometry in Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 56, no. 4, pp. 2024–2037, 2011.

[129] S. Lancelot, R. Roche, A. Slimen, C. Bouillot, E. Levigoureux, J.-B. Langlois, L. Zimmer, and N. Costes, "A Multi-Atlas Based Method for Automated Anatomical Rat Brain MRI Segmentation and Extraction of PET Activity," *PLoS ONE*, vol. 9, no. 10, p. e109113, 2014.

[130] J. Nie and D. Shen, "Automated Segmentation of Mouse Brain Images Using Multi-Atlas Multi-ROI Deformation and Label Fusion," *Neuroinformatics*, vol. 11, no. 1, pp. 35–45, 2013.

[131] M. M. Chakravarty, P. Steadman, M. C. Eede, R. D. Calcott, V. Gu, P. Shaw, A. Raznahan, D. L. Collins, and J. P. Lerch, "Performing label-fusion-based segmentation using multiple automatically generated templates," *Human brain mapping*, vol. 34, no. 10, pp. 2635–2654, 2013.

[132] N. A. Bock, B. J. Nieman, J. B. Bishop, and R. Mark Henkelman, "*In vivo* multiple-mouse MRI at 7 Tesla," *Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1311–1316, 2005.

[133] O. Natt, T. Watanabe, S. Boretius, J. Radulovic, J. Frahm, and T. Michaelis, "High-resolution 3D MRI of mouse brain reveals small cerebral structures *in vivo*," *Journal of neuroscience methods*, vol. 120, no. 2, pp. 203–209, 2002.

[134] M. Aggarwal, J. Zhang, and S. Mori, "Magnetic resonance imaging-based mouse brain atlas and its applications," in *Magnetic Resonance Neuroimaging*, vol. 711, pp. 251–270, Humana Press, 2011.

[135] A. Mackenzie-Graham, "*In vivo* vs. *ex vivo* magnetic resonance imaging in mice," *Frontiers in neuroinformatics*, vol. 6, p. 19, 2012.

[136] Y. Jiang and G. A. Johnson, "Microscopic diffusion tensor atlas of the mouse brain," *NeuroImage*, vol. 56, no. 3, pp. 1235–43, 2011.

[137] L. McEvoy and J. Brewer, "Quantitative structural MRI for early detection of Alzheimer's disease," pp. 1675–1688, 2010.

[138] J. P. Lerch, L. Gazdzinski, J. Germann, J. G. Sled, R. M. Henkelman, and B. J. Nieman, "Wanted dead or alive? The tradeoff between *in-vivo* versus *ex-vivo* MR brain imaging in the mouse," *Frontiers in neuroinformatics*, vol. 6, p. 6, 2012.

[139] E. W. Montie, E. Wheeler, N. Pussini, T. W. K. Battey, J. Barakos, S. Dennison, K. Colegrove, and F. Gulland, "Magnetic resonance imaging quality and volumes of brain structures from live and postmortem imaging of California sea lions with clinical signs of domoic acid," *Diseases of aquatic organisms*, vol. 91, pp. 243–256, 2010.

[140] J. Zhang, Q. Peng, Q. Li, N. Jahanshad, Z. Hou, M. Jiang, N. Masuda, D. R. Langbehn, M. I. Miller, S. Mori, C. a. Ross, and W. Duan, "Longitudinal characterization of brain atrophy of a Huntington's disease mouse model by automated morphological analyses of magnetic resonance images," *NeuroImage*, vol. 49, no. 3, pp. 2340–51, 2010.

[141] G. Schulz, H. J. a. Crooijmans, M. Germann, K. Scheffler, M. Müller-Gerbl, and B. Müller, "Three-dimensional strain fields in human brain resulting from formalin fixation," *Journal of neuroscience methods*, vol. 202, no. 1, pp. 17–27, 2011.

[142] A. Kotrotsou, D. a. Bennett, J. a. Schneider, R. J. Dawe, T. Golak, S. E. Leurgans, L. Yu, and K. Arfanakis, "*Ex vivo* MR volumetry of human brain hemispheres," *Magnetic resonance in medicine*, vol. 71, no. 1, pp. 364–74, 2014.

[143] I. Oguz, R. Yaxley, F. Budin, and M. Hoogstoel, "Comparison of magnetic resonance imaging in live vs. post mortem rat brains," *PLoS ONE*, vol. 8, no. 8, 2013.

[144] H. G. Schnack, N. E. M. van Haren, R. M. Brouwer, A. Evans, S. Durston, D. I. Boomsma, R. S. Kahn, and H. E. Hulshoff Pol, "Changes in thickness and surface area of the human cortex and their relationship with intelligence," *Cerebral Cortex*, vol. 25, no. 6, pp. 1608–1617, 2015.

[145] M. Meyer, F. Liem, S. Hirsiger, L. Jäncke, and J. Hänggi, "Cortical surface area and cortical thickness demonstrate differential structural asymmetry in auditory-related areas of the human cortex," *Cerebral cortex (New York, N.Y. : 1991)*, vol. 24, no. 10, pp. 2541–2552, 2013.

[146] L. M. Rimol, R. Nesvåg, D. J. Hagler, Ø. Bergmann, C. Fennema-Notestine, C. B. Hartberg, U. K. Haukvik, E. Lange, C. J. Pung, A. Server, *et al.*, "Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder," *Biological psychiatry*, vol. 71, no. 6, pp. 552–560, 2012.

[147] C. Hutton, B. Draganski, J. Ashburner, and N. Weiskopf, "A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging," *NeuroImage*, vol. 48, no. 2, pp. 371–80, 2009.

[148] E. R. Sowell, P. M. Thompson, C. M. Leonard, S. E. Welcome, E. Kan, and A. W. Toga, "Longitudinal mapping of cortical thickness and brain growth in normal children," *The Journal of neuroscience*, vol. 24, no. 38, pp. 8223–8231, 2004.

[149] S. J. Sawiak, N. I. Wood, T. A. Carpenter, and a. J. Morton, "Huntington's Disease Mouse Models Online: High-Resolution MRI Images with Stereotaxic Templates for Computational Neuroanatomy," *PLoS ONE*, vol. 7, no. 12, p. e53361, 2012.

[150] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proceedings of the National Academy of Sciences*, vol. 97, no. 20, pp. 11050–11055, 2000.

[151] X. Han, D. L. Pham, D. Tosun, M. E. Rettmann, C. Xu, and J. L. Prince, "CRUISE: Cortical reconstruction using implicit surface evolution," *NeuroImage*, vol. 23, pp. 997–1012, 2004.

[152] D. MacDonald, N. Kabani, D. Avis, and A. C. Evans, "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI," *NeuroImage*, vol. 12, no. 3, pp. 340–356, 2000.

[153] S. R. Das, B. B. Avants, M. Grossman, and J. C. Gee, "Registration based cortical thickness measurement," *NeuroImage*, vol. 45, no. 3, pp. 867–879, 2009.

[154] R. Dahnke, R. A. Yotter, and C. Gaser, "Cortical thickness and central surface estimation," *NeuroImage*, vol. 65, pp. 336–348, 2012.

[155] C. Hutton, E. De Vita, J. Ashburner, R. Deichmann, and R. Turner, "Voxel-based cortical thickness measurements in MRI," *NeuroImage*, vol. 40, no. 4, pp. 1701–10, 2008.

[156] Q. Li, H. Pardoe, R. Lichter, E. Werden, A. Raffelt, T. Cumming, and A. Brodtmann, "Cortical thickness estimation in longitudinal stroke studies: A comparison of 3 measurement methods," *NeuroImage: Clinical*, vol. 8, no. 0, pp. 526–535, 2015.

[157] C. K. Tamnes, Y. Ostby, A. M. Fjell, L. T. Westlye, P. Due-Tø nnessen, and K. B. Walhovd, "Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure," *Cereb Cortex*, vol. 20, no. 3, pp. 534–548, 2010.

[158] J. P. Lerch and A. C. Evans, "Cortical thickness analysis examined through power analysis and a population simulation," *NeuroImage*, vol. 24, no. 1, pp. 163–173, 2005.

[159] J. P. Lerch, J. C. Pruessner, A. Zijdenbos, H. Hampel, S. J. Teipel, and A. C. Evans, "Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy," *Cerebral Cortex*, vol. 15, no. 7, pp. 995–1001, 2005.

[160] K. L. Narr, R. M. Bilder, A. W. Toga, R. P. Woods, D. E. Rex, P. R. Szeszko, D. Robinson, S. Sevy, H. Gunduz-Bruce, Y. P. Wang, H. DeLuca, and P. M. Thompson, "Mapping cortical thickness and gray matter concentration in first episode schizophrenia," *Cerebral Cortex*, vol. 15, no. 6, pp. 708–719, 2005.

[161] E. R. Sowell, B. S. Peterson, E. Kan, R. P. Woods, J. Yoshii, R. Bansal, D. Xu, H. Zhu, P. M. Thompson, and A. W. Toga, "Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age," *Cerebral Cortex*, vol. 17, no. 7, pp. 1550–1560, 2007.

[162] J. Lee, C. Ehlers, F. Crews, M. Niethammer, F. Budin, B. Paniagua, K. Sulik, J. Johns, M. Styner, and I. Oguz, "Automatic cortical thickness analysis on rodent brain," in *SPIE International Symposium on Medical Imaging*, pp. 796248–796248, International Society for Optics and Photonics, 2011.

[163] J. D. Gatewood, A. Wills, S. Shetty, J. Xu, A. P. Arnold, P. S. Burgoyne, and E. F. Rissman, "Sex chromosome complement and gonadal sex influence aggressive and parental behaviors in mice," *The Journal of neuroscience*, vol. 26, no. 8, pp. 2335–2342, 2006.

[164] E. H. Aylward, R. Habbak, W. AC, and E. Al, "Cerebellar volume in adults with down syndrome," *Archives of Neurology*, vol. 54, no. 2, pp. 209–212, 1997.

[165] S. M. Ravizza, C. a. McCormick, J. E. Schlerf, T. Justus, R. B. Ivry, and J. a. Fiez, "Cerebellar damage produces selective deficits in verbal working memory," *Brain*, vol. 129, no. Pt 2, pp. 306–20, 2006.

[166] J. D. Schmahmann and J. C. Sherman, "The cerebellar cognitive affective syndrome," *Brain*, vol. 121, no. 4, pp. 561–579, 1998.

[167] K. U. Szulc, B. J. Nieman, E. J. Houston, B. B. Bartelle, J. P. Lerch, A. L. Joyner, and D. H. Turnbull, "MRI analysis of cerebellar and vestibular developmental phenotypes in Gbx2 conditional knockout mice," *Magnetic Resonance in Medicine*, vol. 70, no. 6, pp. 1707–1717, 2013.

[168] P. E. Steadman, J. Ellegood, K. U. Szulc, D. H. Turnbull, A. L. Joyner, R. M. Henkelman, and J. P. Lerch, "Genetic effects on cerebellar structure across mouse models of autism using a magnetic resonance imaging atlas," *Autism research*, vol. 7, no. 1, pp. 124–37, 2014.

[169] E. D'Angelo and C. I. De Zeeuw, "Timing and plasticity in the cerebellum: focus on the granular layer," *Trends in Neurosciences*, vol. 32, no. 1, pp. 30–40, 2009.

[170] R. L. Buckner, "The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging," *Neuron*, vol. 80, no. 3, pp. 807–15, 2013.

[171] C. J. O'Halloran, G. J. Kinsella, and E. Storey, "The cerebellum and neuropsychological functioning: a critical review," *Journal of clinical and experimental neuropsychology*, vol. 34, no. 1, pp. 35–56, 2012.

[172] H. Lemaitre, A. L. Goldman, F. Sambataro, B. A. Verchinski, A. Meyer-Lindenberg, D. R. Weinberger, and V. S. Mattay, "Normal age-related brain morphometric changes: Nonuniformity across cortical thickness, surface area and gray matter volume?," *Neurobiology of Aging*, vol. 33, no. 3, 2012.

[173] S. B. Eickhoff, A. Schleicher, F. Scheperjans, N. Palomero-Gallagher, and K. Zilles, "Analysis of neurotransmitter receptor distribution patterns in the cerebral cortex," *NeuroImage*, vol. 34, pp. 1317–1330, 2007.

[174] L. L. Baxter, T. H. Moran, J. T. Richtsmeier, J. Troncoso, and R. H. Reeves, "Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse," *Human molecular genetics*, vol. 9, no. 2, pp. 195–202, 2000.

[175] K. J. Friston, A. P. Holmes, K. J. Worsley, J. Poline, C. D. Frith, R. S. Frackowiak, *et al.*, "Statistical parametric maps in functional imaging: a general linear approach," *Human brain mapping*, vol. 2, no. 4, pp. 189–210, 1994.

[176] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady, and P. M. Matthews, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23, pp. S208–S219, 2004.

[177] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[178] D. W. Shattuck and R. M. Leahy, "Brainsuite: An automated cortical surface identification tool," *Medical Image Analysis*, vol. 6, no. 2, pp. 129–142, 2002.

[179] R. Goebel, F. Esposito, and E. Formisano, "Analysis of Functional Image Analysis Contest (FIAC) data with BrainVoyager QX: From single-subject to cortically aligned group General Linear Model analysis and self-organizing group Independent Component Analysis," *Human Brain Mapping*, vol. 27, no. 5, pp. 392–401, 2006.

[180] B. Avants, N. Tustison, and G. Song, "Advanced Normalization Tools (ANTS)," *Insight Journal*, pp. 1–35, 2009.

[181] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, *et al.*, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, pp. 573–581, Springer, 2001.

[182] F. Budin, M. Hoogstoel, P. Reynolds, M. Grauer, S. K. O'Leary-Moore, and I. Oguz, "Fully automated rodent brain MR image processing pipeline on a midas server: from acquired images to region-based statistics," *Frontiers in neuroinformatics*, vol. 7, 2013.

[183] M. Modat, M. J. Cardoso, P. Daga, D. Cash, N. C. Fox, and S. Ourselin, "Inverse-consistent symmetric free form deformation," in *Biomedical Image Registration*, pp. 79–88, Springer, 2012.

[184] M. Modat, P. Daga, M. J. Cardoso, S. Ourselin, G. R. Ridgway, and J. Ashburner, "Parametric non-rigid registration using a stationary velocity field," in *Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on*, pp. 145–150, IEEE, 2012.

[185] S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block matching: A general framework to improve robustness of rigid registration of medical images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000*, pp. 557–566, Springer, 2000.

[186] M. Modat, D. M. Cash, P. Daga, G. P. Winston, J. S. Duncan, and S. Ourselin, "A symmetric block-matching framework for global registration," in *SPIE International Symposium on Medical Imaging*, pp. 90341D–90341D, International Society for Optics and Photonics, 2014.

[187] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *Medical Imaging, IEEE Transactions on*, vol. 17, no. 1, pp. 87–97, 1998.

[188] N. J. Tustison, B. B. Avants, P. a. Cook, Y. Zheng, A. Egan, P. a. Yushkevich, and J. C. Gee, "N4ITK: improved N3 bias correction," *IEEE transactions on medical imaging*, vol. 29, pp. 1310–20, June 2010.

[189] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE transactions on medical imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[190] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[191] G. Paxinos and K. B. Franklin, *The mouse brain in stereotaxic coordinates*. Gulf Professional Publishing, 2004.

[192] F. Maes, K. Van Leemput, L. E. DeLisi, D. Vandermeulen, and P. Suetens, "Quantification of cerebral grey and white matter asymmetry from MRI," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI'99*, pp. 348–357, Springer, 1999.

[193] D. M. Bowden, G. A. Johnson, L. Zaborsky, W. D. K. Green, E. Moore, A. Badea, M. F. Dubach, and F. L. Bookstein, "A symmetrical Waxholm canonical mouse brain for NeuroMaps," *Journal of neuroscience methods*, vol. 45, no. 3, pp. 170–175, 2011.

[194] Y. Liu, R. T. Collins, and W. E. Rothfus, "Automatic bilateral symmetry (midsagittal) plane extraction from pathological 3d neuroradiological images," in

SPIE International Symposium on Medical Imaging, pp. 1528–1539, International Society for Optics and Photonics, 1998.

[195] S. Prima, J.-P. Thirion, G. Subsol, and N. Roberts, "Automatic analysis of normal brain dissymmetry of males and females in MR images," in *Medical Image Computing and Computer-Assisted Interventation-MICCAI'98*, pp. 770–779, Springer, 1998.

[196] J. P. John, L. Wang, A. J. Moffitt, H. K. Singh, M. H. Gado, and J. G. Csernansky, "Inter-rater reliability of manual segmentation of the superior, inferior and middle frontal gyri," *Psychiatry Research: Neuroimaging*, vol. 148, no. 2-3, pp. 151–163, 2006.

[197] B. Sinclair, J. Cleary, M. Modat, V. Norris, F. Wiseman, V. Tybulewicz, E. Fisher, M. Lythgoe, and S. Ourselin, "Tensor-based morphometry on the Tc1 mouse model of Down syndrome highlights previously undetected phenotypes," in *Proc Intl Soc Mag Reson Med*, vol. 19, p. 702, 2011.

[198] J. L. Whitwell, W. R. Crum, H. C. Watt, and N. C. Fox, "Normalization of cerebral volumes by use of intracranial volume: Implications for longitudinal quantitative mr imaging," *American Journal of Neuroradiology*, vol. 22, no. 8, pp. 1483–1489, 2001.

[199] R. Steen, R. Hamer, and J. Lieberman, "Measuring Brain Volume by MR Imaging: Impact of Measurement Precision and Natural Variation on Sample Size Requirements," *American Journal of Neuroradiology*, vol. 28, no. 6, pp. 1119–1125, 2007.

[200] G. Wang, X. Zhang, Q. Su, J. Chen, L. Wang, Y. Ma, Q. Liu, L. Xu, J. Shi, and Y. Wang, "A heat kernel based cortical thickness estimation algorithm," in *Multimodal Brain Image Analysis*, pp. 233–245, Springer, 2013.

[201] B. A. Duffy, K. P. Chun, D. Ma, M. F. Lythgoe, and R. C. Scott, "Dexamethasone exacerbates cerebral edema and brain injury following lithium-pilocarpine induced status epilepticus," *Neurobiology of Disease*, vol. 63, pp. 229–236, 2014.

[202] M. C. van Eede, J. Scholz, M. M. Chakravarty, R. M. Henkelman, J. P. Lerch, and M. C. V. Eede, "Mapping registration sensitivity in MR mouse brain images," *NeuroImage*, vol. 82, no. 0, pp. 226–36, 2013.

[203] M. Modat, G. R. Ridgway, Z. a. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer methods and programs in biomedicine*, vol. 98, no. 3, pp. 278–84, 2010.

[204] D. Rueckert, C. Hayes, C. Studholme, P. Summers, M. Leach, and D. J. Hawkes, "Non-rigid registration of breast MR images using mutual information," in *Medical Image Computing and Computer-Assisted Interventation–MICCAI'98*, pp. 1144–1152, Springer, 1998.

[205] M. Chakravarty, M. van Eede, and J. P. Lerch, "Improved segmentation of mouse MRI data using multiple automatically generated templates," in *International Society for Magnetic Resonance in Medicine*, vol. 15, p. 2011, 2011.

[206] S. Spring, J. P. Lerch, M. K. Wetzel, A. C. Evans, and R. M. Henkelman, "Cerebral asymmetries in 12-week-old C57Bl/6J mice measured by magnetic resonance imaging," *NeuroImage*, vol. 50, no. 2, pp. 409–415, 2010.

[207] M. M. Kohl, O. A. Shipton, R. M. Deacon, J. N. P. Rawlins, K. Deisseroth, and O. Paulsen, "Hemisphere-specific optogenetic stimulation reveals left-right asymmetry of hippocampal plasticity," *Nature Neuroscience*, vol. 14, no. 11, pp. 1413–1415, 2011.

[208] M. Boccardi, M. Bocchetta, R. Ganzola, N. Robitaille, A. Redolfi, S. Duchesne, C. R. Jack, and G. B. Frisoni, "Operationalizing protocol differences for EADC-ADNI manual hippocampal segmentation," *Alzheimer's & Dementia*, pp. 1–11, 2013.

[209] M. J. Kempton, T. S. A. Underwood, S. Brunton, F. Stylios, A. Schmechtig, U. Ettinger, M. S. Smith, S. Lovestone, W. R. Crum, S. Frangou, S. C. R. Williams, and A. Simmons, "A comprehensive testing protocol for MRI neu-

roanatomical segmentation techniques: Evaluation of a novel lateral ventricle segmentation method," *NeuroImage*, vol. 58, pp. 1051–1059, 2011.

[210] S. M. Nestor, E. Gibson, F.-Q. Gao, A. Kiss, and S. E. Black, "A direct morphometric comparison of five labeling protocols for multi-atlas driven automatic segmentation of the hippocampus in Alzheimer's disease," *NeuroImage*, vol. 66C, pp. 50–70, 2012.

[211] J. A. Wells, J. M. O'Callaghan, H. E. Holmes, N. M. Powell, R. A. Johnson, B. Siow, F. Torrealdea, O. Ismail, S. Walker-Samuel, X. Golay, M. Rega, S. Richardson, M. Modat, M. J. Cardoso, S. Ourselin, A. J. Schwarz, Z. Ahmed, T. K. Murray, M. J. O'Neill, E. C. Collins, N. Colgan, and M. F. Lythgoe, "*In vivo* imaging of tau pathology using multi-parametric quantitative MRI," *NeuroImage*, vol. 111, pp. 369–78, 2015.

[212] J. O'Callaghan, J. Wells, S. Richardson, H. Holmes, Y. Yu, S. Walker-Samuel, B. Siow, M. F. Lythgoe, J. O. Callaghan, and S. Walker, "Is Your System Calibrated? MRI Gradient System Calibration for Pre-Clinical, High-Resolution Imaging," *PLoS ONE*, vol. 9, no. 5, p. e96568, 2014.

[213] K. Santacruz, J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingelsson, A. Guimaraes, M. DeTure, M. Ramsden, E. McGowan, *et al.*, "Tau suppression in a neurodegenerative mouse model improves memory function," *Science*, vol. 309, no. 5733, pp. 476–481, 2005.

[214] D. Ma, M. J. Cardoso, M. Modat, N. Powell, J. Wells, H. Holmes, F. Wiseman, V. Tybulewicz, E. Fisher, M. F. Lythgoe, and S. Ourselin, "Automatic Structural Parcellation of Mouse Brain MRI Using Multi-Atlas Label Fusion," *PLoS ONE*, vol. 9, no. 1, p. e86576, 2014.

[215] D. G. Altman and J. M. Bland, "Measurement in medicine: the analysis of method comparison studies," *The statistician*, pp. 307–317, 1983.

[216] J. Martin Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[217] P. Myles and J. Cui, "I. using the bland–altman method to measure agreement with repeated measures," *British journal of anaesthesia*, vol. 99, no. 3, pp. 309–311, 2007.

[218] M. A. Pollock, S. G. Jefferson, J. W. Kane, K. Lomax, G. MacKinnon, and C. B. Winnard, "Method Comparison - A Different Approach," *Annals of Clinical Biochemistry*, vol. 29, no. 5, pp. 556–560, 1992.

[219] S. M. Nestor, R. Rupsingh, M. Borrie, M. Smith, V. Accomazzi, J. L. Wells, J. Fogarty, and R. Bartha, "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain*, vol. 131, no. 9, pp. 2443–2454, 2008.

[220] M. W. Weiner, "Expanding ventricles may detect preclinical Alzheimer disease," *Neurology*, vol. 70, no. 11, pp. 824–825, 2008.

[221] B. Cragg, "Preservation of extracellular space during fixation of the brain for electron microscopy," *Tissue and Cell*, vol. 12, no. 1, pp. 63–72, 1980.

[222] S. L. Palay, S. M. McGee-Russell, S. Gordon, and M. A. Grillo, "Fixation of neural tissues for electron microscopy by perfusion with solutions of osmium tetroxide," *The Journal of cell biology*, vol. 12, pp. 385–410, 1962.

[223] O. v. B. und Halbach, M. Lotze, and J. P. Pfannmöller, "Post-mortem magnetic resonance microscopy (MRM) of the murine brain at 7 Tesla results in a gain of resolution as compared to *in-vivo* MRM," *Frontiers in Neuroanatomy*, vol. 8, no. June, pp. 1–7, 2014.

[224] J. Sun, C. Kil, M. C. Stankewich, Z. Yao, J. Li, and A. O. Vortmeyer, "A 10-minute prototype assay for tissue degradation monitoring in clinical specimens," *Experimental and molecular pathology*, vol. 99, no. 1, pp. 86–94, 2015.

[225] N. Acer, B. Sahin, O. Bas, T. Ertekin, and M. Usanmaz, "Comparison of three methods for the estimation of total intracranial volume: stereologic, planimetric, and anthropometric approaches," *Annals of plastic surgery*, vol. 58, no. 1, pp. 48–53, 2007.

[226] L. Lemieux, A. Hammers, T. Mackinnon, and R. S. Liu, "Automatic segmentation of the brain and intracranial cerebrospinal fluid in T1-weighted volume MRI scans of the head, and its application to serial cerebral and intracranial volumetry," *Magnetic resonance in medicine*, vol. 49, no. 5, pp. 872–884, 2003.

[227] G. Ridgway, J. Barnes, T. Pepple, and N. Fox, "Estimation of total intracranial volume: a comparison of methods," *Alzheimer's & Dementia*, vol. 7, no. 4, pp. S62–S63, 2011.

[228] S. J. Sawiak, N. I. Wood, G. B. Williams, a. J. Morton, and T. A. Carpenter, "Voxel-based morphometry with templates and validation in a mouse model of Huntington's disease," *Magnetic resonance imaging*, vol. 31, no. 9, pp. 1522–31, 2013.

[229] M. Fukunaga, T.-Q. Li, P. van Gelderen, J. a. de Zwart, K. Shmueli, B. Yao, J. Lee, D. Maric, M. a. Aronova, G. Zhang, R. D. Leapman, J. F. Schenck, H. Merkle, and J. H. Duyn, "Layer-specific variation of iron content in cerebral cortex as a source of MRI contrast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 8, pp. 3834–3839, 2010.

[230] S. Boretius, L. Kasper, R. Tammer, T. Michaelis, and J. Frahm, "MRI of cellular layers in mouse brain *in vivo*," *NeuroImage*, vol. 47, no. 4, pp. 1252–1260, 2009.

[231] T. Watanabe, J. Frahm, and T. Michaelis, "Cell layers and neuropil: contrast-enhanced MRI of mouse brain in vivo.," *NMR in biomedicine*, vol. 26, no. 12, pp. 1870–8, 2013.

[232] J. P. Marques, W. van der Zwaag, C. Granziera, G. Krueger, and R. Gruetter, "Cerebellar cortical layers: in vivo visualization with structural high-field-strength MR imaging.," *Radiology*, vol. 254, no. 3, pp. 942–948, 2010.

[233] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–51, 2005.

[234] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 143–150, 2000.

[235] J. C. Lau, J. P. Lerch, J. G. Sled, R. M. Henkelman, A. C. Evans, and B. J. Bedell, "Longitudinal neuroanatomical changes determined by deformation-based morphometry in a mouse model of Alzheimer's disease," *Neuroimage*, vol. 42, no. 1, pp. 19–27, 2008.

[236] L. Lam, S. W. Lee, and C. Y. Suen, "Thinning methodologies–A comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869–885, 1992.

[237] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Interventation–MICCAI'98*, pp. 130–137, Springer, 1998.

[238] M. Grand, S. P. Zehntner, M.-k. Ho, F. Hébert, A. Wood, F. Carbonell, A. P. Zijdenbos, E. Hamel, and B. J. Bedell, "Neurobiology of Disease Early cortical thickness changes predict β-amyloid deposition in a mouse model of Alzheimer's disease," *Neurobiology of Disease*, vol. 54, pp. 59–67, 2013.

[239] J. L. Whitwell, W. R. Crum, H. C. Watt, and N. C. Fox, "Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging," *American Journal of Neuroradiology*, vol. 22, no. 8, pp. 1483–1489, 2001.

[240] E. Westman, C. Aguilar, J.-S. Muehlboeck, and A. Simmons, "Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment," *Brain topography*, vol. 26, no. 1, pp. 9–23, 2013.

[241] Q. Zhou, M. Goryawala, M. Cabrerizo, W. Barker, R. Duara, and M. Adjouadi, "Significance of normalization on anatomical MRI measures in predicting Alzheimer's disease," *The Scientific World Journal*, vol. 2014, 2014.

[242] D. C. Van Essen and M. F. Glasser, "In vivo architectonics: A cortico-centric perspective," *NeuroImage*, vol. 93, pp. 157–164, 2013.

[243] M. F. Glasser, M. S. Goyal, T. M. Preuss, M. E. Raichle, and D. C. Van Essen, "Trends and properties of human cerebral cortex: Correlations with cortical myelin content," *NeuroImage*, vol. 93, pp. 165–175, 2013.

[244] A. B. Storsve, A. M. Fjell, C. K. Tamnes, L. T. Westlye, K. Overbye, H. W. Aasland, and K. B. Walhovd, "Differential Longitudinal Changes in Cortical Thickness, Surface Area and Volume across the Adult Life Span: Regions of Accelerating and Decelerating Change," *The Journal of neuroscience*, vol. 34, no. 25, pp. 8488–98, 2014.

[245] S. Bouix and K. Siddiqi, "Divergence-based medial surfaces," in *Computer Vision-ECCV 2000*, pp. 603–618, Springer, 2000.

[246] A. M. Winkler, P. Kochunov, J. Blangero, L. Almasy, K. Zilles, P. T. Fox, R. Duggirala, and D. C. Glahn, "Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies," *NeuroImage*, vol. 53, pp. 1135–1146, 2010.

[247] A. M. Winkler, M. R. Sabuncu, B. T. T. Yeo, B. Fischl, D. N. Greve, P. Kochunov, T. E. Nichols, J. Blangero, and D. C. Glahn, "Measuring and comparing brain cortical surface area and other areal quantities," *NeuroImage*, vol. 61, no. 4, pp. 1428–43, 2012.

[248] J. P. B. O'Connor, A. Jackson, M.-C. Asselin, D. L. Buckley, G. J. M. Parker, and G. C. Jayson, "Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives," *The lancet oncology*, vol. 9, no. 8, pp. 766–776, 2008.

[249] J. W. Prescott, "Quantitative imaging biomarkers: the application of advanced image processing and analysis to clinical and preclinical decision making," *Journal of digital imaging*, vol. 26, no. 1, pp. 97–108, 2013.

[250] M. Zamyadi, L. Baghdadi, J. P. Lerch, S. Bhattacharya, J. E. Schneider, R. M. Henkelman, and J. G. Sled, "Mouse embryonic phenotyping by morphometric analysis of MR images," *Physiological genomics*, vol. 42, no. 2, pp. 89–95, 2010.

[251] F. C. Norris, M. Modat, J. O. Cleary, A. N. Price, K. McCue, P. J. Scambler, S. Ourselin, and M. F. Lythgoe, "Segmentation propagation using a 3D embryo atlas for high-throughput MRI phenotyping: Comparison and validation with manual segmentation," *Magnetic Resonance in Medicine*, vol. 69, no. 3, pp. 877–883, 2013.

[252] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, Others, and A. Et, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[253] P. A. Freeborough and N. C. Fox, "The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI," *IEEE transactions on medical imaging*, vol. 16, no. 5, pp. 623–629, 1997.

[254] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: a review," *Medical image analysis*, vol. 13, no. 4, pp. 543–563, 2009.

[255] W. Denk and H. Horstmann, "Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure," *PLoS Biology*, vol. 2, no. 11, 2004.

[256] D. Li, D. Ping, H. Ye, X. Qin, and X. Wu, "HREM study of the microstructure in nanocrystalline materials," *Materials Letters*, vol. 18, no. 1, pp. 29–34, 1993.

[257] J. Sharpe, "Optical projection tomography," in *Advanced Imaging in Biology and Medicine*, pp. 199–224, Springer, 2009.

[258] K. Chung and K. Deisseroth, "CLARITY for mapping the nervous system," *Nature methods*, vol. 10, no. 6, pp. 508–13, 2013.

[259] K. Kasiri, P. Fieguth, and D. A. Clausi, "Cross modality label fusion in multi-atlas segmentation," in *Image Processing (ICIP), 2014 IEEE International Conference on*, pp. 16–20, 2014.

[260] M. P. Heinrich, O. Maier, and H. Handels, "Multi-modal Multi-Atlas Segmentation using Discrete Optimisation and Self-Similarities," *VISCERAL@ ISBI 2015 VISCERAL Anatomy3 Organ Segmentation Challenge*, p. 27, 2015.

[261] M. F. Glasser and D. C. Van Essen, "Mapping human cortical areas *in vivo* based on myelin content as revealed by T1- and T2-weighted MRI," *The Journal of neuroscience*, vol. 31, no. 32, pp. 11597–11616, 2011.

[262] A. Lutti, F. Dick, M. I. Sereno, and N. Weiskopf, "Using high-resolution quantitative mapping of R1 as an index of cortical myelination," *NeuroImage*, vol. 93, pp. 176–188, 2013.

[263] R. Shafee, R. L. Buckner, and B. Fischl, "Gray matter myelination of 1555 human brains using partial volume corrected MRI images," *NeuroImage*, vol. 105, pp. 473–485, 2015.

[264] J. Dinse, P. Martin, A. Schäfer, S. Geyer, R. Turner, and P.-L. Bazin, "Quantifying differences between primary cortical areas in humans based on laminar profiles in *in vivo* MRI data," in *Bildverarbeitung für die Medizin 2013*, pp. 39–44, Springer, 2013.

[265] C. L. Tardif, J. Dinse, A. Schäfer, R. Turner, and P.-L. Bazin, "Multi-modal surface-based alignment of cortical areas using intra-cortical T1 contrast," in *Multimodal Brain Image Analysis*, pp. 222–232, Springer, 2013.

[266] J. Dinse, N. Härtwich, M. Waehnert, C. Tardif, A. Schäfer, S. Geyer, B. Preim, R. Turner, and P.-L. Bazin, "A cytoarchitecture-driven myelin model reveals area-specific signatures in human primary and secondary areas using ultra-high resolution in-vivo brain MRI," *NeuroImage*, vol. 114, pp. 71–87, 2015.