

# Automated morphometric analysis and phenotyping of mouse brains from structural $\mu$ MR images

Nicholas Morton Powell

Submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy of University College London

January 2016

## Supervisors

Professor Sebastien Ourselin

Professor Mark Lythgoe

University College London

UK



# Declaration

I, Nicholas Morton Powell, confirm that the work presented in this thesis is my own, except where stated otherwise in the text. This work is based on research conducted by me, during the time period November 2011 to December 2015 at University College London.

Nicholas Morton Powell

19<sup>th</sup> January 2016





# Abstract

In light of the utility and increasing ubiquity of mouse models of genetic and neurological disease, I describe fully automated pipelines for the investigation of structural microscopic magnetic resonance images of mouse brains – for both high-throughput phenotyping, and monitoring disease.

Mouse models offer unparalleled insight into genetic function and brain plasticity, in phenotyping studies; and neurodegenerative disease onset and progression, in therapeutic trials. I developed two cohesive, automatic software tools, for Voxel- and Tensor-Based Morphometry (V/TBM) and the Boundary Shift Integral (BSI), in the mouse brain. V/TBM are advantageous for their ability to highlight morphological differences between groups, without laboriously delineating regions of interest. The BSI is a powerful and sensitive imaging biomarker for the detection of atrophy.

The resulting pipelines are described in detail. I show the translation and application of open-source software developed for clinical MRI analysis to mouse brain data: for tissue segmentation into high-quality, subject-specific maps, using contemporary multi-atlas techniques; and for symmetric, inverse-consistent registration. I describe atlases and parameters suitable for the preclinical paradigm, and illustrate and discuss image processing challenges encountered and overcome during development.

As proof of principle and to illustrate robustness, I used both pipelines with *in* and *ex vivo* mouse brain datasets to identify differences between groups, representing the morphological influence of genes, and subtle, longitudinal changes over time, in particular relation to Down syndrome and Alzheimer's disease. I also discuss the merits of transitioning preclinical analysis from predominately *ex vivo* MRI to *in vivo*, where morphometry is still viable and fewer mice are necessary.

This thesis conveys the cross-disciplinary translation of up-to-date image analysis techniques to the preclinical paradigm; the development of novel methods and adaptations to robustly process large cohorts of data; and the sensitive detection of phenotypic differences and neurodegenerative changes in the mouse brain.

# Acknowledgements

I am especially grateful to my supervisors, Professor Sebastien Ourselin and Professor Mark Lythgoe, for giving me the chance to pursue research at the interface between two fast-advancing, involving disciplines.

Thanks to members of Centre for Medical Image Computing and the Translational Imaging Group, past and present, who were welcoming, friendly and open to explaining everything. I am indebted to Marc Modat and Jorge Cardoso, whose endless patience, intuition and knowledge I exploited quite mercilessly, and to Da Ma for his camaraderie. Thank you.

Thank you to members of the Centre for Advanced Biomedical Imaging, past and present, for their friendship and helpful discussion, many of whom worked extra-long and unsociable, windowless hours acquiring MRI data for me to analyse, including Holly Holmes, Niall Colgan, Ozama Ismail, Yichao Yu, Ian Harrison, Francesca Norris, James O’Callaghan, Simon Richardson, Tom Roberts, Arun Niranjana, Yanan Zhu, Isabel Christie, Bernard Siow, Simon Walker-Samuel, Jack Wells, Jon O. Cleary and Ben Sinclair. I am also very grateful to Ben and Jon for their introductions to morphometry analysis methods.

Thank you to my collaborators from the Institute of Neurology, Professor Elizabeth Fisher, Frances Wiseman, Adrian Isaacs, and from Eli Lilly, Emily Collins and Mike O’Neill. And to Ferran Prados and Kelvin Leung for their helpful discussions and encouragement around the BSI chapter. And to the mice, without whom this work would not have been possible.

Thank you to Katy and Baptiste for housing me while I found some feet in London; the staff and vice-wardens of UCL Goldsmid House; and understanding housemates and

house/floor/couch-lenders Matt & Aileen A., Alex S., James A., Chai P., Katie H., Mia H., Sam F., Kevin G., Maria T., Rachel A., Marielle M., Dave M., Sam & Phil S. You are all wonderful and I am so, so lucky.

The many opportunities for public outreach and socialising made UCL a great place to be. Nothing would have been possible without the wonderful and generous Dominique Draï, Ron Gaston, Jenny Nery, Morium Ali, Liz Zuzikova and Rupy Ghatrora.

And my parents, and Bess, for her company in the final few weeks.

*For my parents*

# Contents

	Declaration.....	3
	Abstract.....	5
	Acknowledgements.....	7
	Contents .....	9
	List of Figures .....	14
	List of Tables.....	17
	Publications arising from this work.....	18
	Abbreviations.....	22
<b>1</b>	<b>Introduction.....</b>	<b>25</b>
	1.1 The promise of MRI .....	25
	1.2 Phenotyping mice.....	26
	1.3 Morphometry in preclinical studies .....	28
	1.4 From <i>ex vivo</i> to <i>in vivo</i> .....	30
	1.5 Aims and contributions of this research and thesis.....	31
	1.6 Overview of chapters.....	33
<b>2</b>	<b>Background theory and state-of-the-art practices .....</b>	<b>37</b>
	2.1 MR images of the mouse brain .....	37
	2.2 Morphometry.....	41
	2.2.1 Atlases .....	43
	2.2.2 Mouse brain atlases .....	44
	2.2.3 Multi-atlas label fusion for parcellation .....	47
	2.2.4 Tissue segmentation .....	49
	2.2.5 Mouse brain segmentation and tissue probability maps .....	52
	2.2.6 Image registration.....	56
	2.2.7 Voxel- and Tensor-Based Morphometry .....	62
	2.2.8 Preclinical use of V/TBM.....	72
	2.3 Alzheimer's disease .....	78

2.3.1	Biomarkers .....	82
2.3.2	Outcome measures .....	87
2.3.3	Treatment .....	89
2.3.4	Mouse models of Alzheimer's disease .....	90
2.4	Down syndrome .....	92
2.4.1	Mouse models of Down syndrome .....	93
<b>3</b>	<b>Development of a high-throughput software pipeline for phenotyping preclinical subjects .....</b>	<b>97</b>
3.1	Objectives and motivation .....	98
3.2	Automated extraction of individual subjects from multiple-subject images .....	101
3.2.1	Multiple subject scanning .....	101
3.2.2	Challenges for automated identification and extraction .....	103
3.2.3	Solutions for automated identification and extraction .....	106
3.3	Orientation correction .....	109
3.3.1	Initialisation: principal axes .....	109
3.3.2	Using symmetry to test rotation .....	112
3.3.3	Final choice of orientations .....	115
3.3.4	Orientation of embryos .....	117
3.4	Intensity non-uniformity correction .....	117
3.5	Brain masking (skull-stripping) .....	120
3.5.1	Multi-Atlas masking with STEPS .....	122
3.5.2	Problems for automated mouse brain masking .....	123
3.6	Segmentation of brain tissues .....	124
3.7	Intensity standardisation .....	128
3.8	Group-wise registration .....	130
3.8.1	Assessing GWR quality .....	133
3.9	Segmentation propagation for Jacobian integration .....	136
3.10	Transformation of Jacobian determinant maps .....	136
3.11	Modulation of tissue class maps .....	137
3.12	Statistical tests .....	137
3.13	Discussion .....	138
3.14	Conclusions .....	141
<b>4</b>	<b>Application of the pipeline to <i>ex vivo</i> mouse brains .....</b>	<b>143</b>

4.1	Morphometric analysis of the Tc1 model of Down syndrome.....	144
4.1.1	The Tc1 model.....	144
4.1.2	Mice and multiple brain imaging.....	144
4.1.3	Voxel-wise statistical tests.....	147
4.1.4	Group-wise registration assessment.....	147
4.1.5	Global volume results.....	149
4.1.6	Voxel-wise analysis.....	152
4.1.7	Tensor-Based Morphometry.....	155
4.1.8	Individual voxel analysis of TBM results.....	156
4.1.9	Voxel-Based Morphometry.....	158
4.1.10	Discussion: morphometry of Tc1 brains.....	160
4.1.11	Comparison with earlier results.....	165
4.1.12	Corroboration of V/TBM with cohort 2.....	165
4.1.13	Comparison with parcellation.....	165
4.1.14	Hyperintense rim.....	167
4.1.15	Conclusions.....	169
4.2	Tc1xJ20 cross-breed study.....	169
4.2.1	Introduction.....	169
4.2.2	Methodology.....	170
4.2.3	Results.....	171
4.2.4	Discussion.....	176
<b>5</b>	<b><i>In vivo</i> morphometry.....</b>	<b>181</b>
5.1	The rTg4510 mouse model of tauopathy.....	182
5.2	Study design and methodology.....	184
5.2.1	Animals.....	184
5.2.2	Treatment.....	186
5.2.3	Image acquisition.....	186
5.3	Adjustments to the pipeline for <i>in vivo</i> data.....	188
5.3.1	Non-uniformity correction.....	188
5.3.2	Brain masking with manual intervention and bootstrapping.....	190
5.3.3	Group-wise registration.....	191
5.4	Single time-point morphometric investigation of the rTg4510 mouse brain.....	193
5.5	Multiple time-point rTg4510 study.....	196

5.5.1	Cross-sectional analysis .....	197
5.5.2	Longitudinal analysis .....	203
5.6	The TIV covariate .....	208
5.7	Comparing <i>ex vivo</i> morphometry to <i>in vivo</i> .....	212
5.7.1	TBM results compared, <i>in vivo</i> and <i>ex vivo</i> .....	212
5.7.2	Local structural differences between <i>in vivo</i> and <i>ex vivo</i> .....	217
5.7.3	Power analysis .....	220
5.7.4	Fundamentally differing morphologies .....	224
5.8	Discussion .....	226
5.8.1	Cross-sectional and longitudinal TBM .....	226
5.8.2	<i>In vivo</i> versus <i>ex vivo</i> .....	228
<b>6</b>	<b>Longitudinal atrophy measurement using the Boundary Shift Integral in the mouse brain .....</b>	<b>231</b>
6.1	Introduction .....	232
6.2	Methodology .....	236
6.2.1	Study design .....	236
6.2.2	Data quality control, image selection and animal attrition .....	238
6.2.3	Image processing pipeline .....	239
6.2.4	Probabilistic XOR boundary shift region .....	246
6.2.5	Double-window approach for T <sub>2</sub> W mouse brains .....	248
6.2.6	Intensity standardisation and window selection .....	251
6.2.7	Boundary Shift Integral calculation .....	253
6.3	Results .....	256
6.4	Evaluation .....	263
6.4.1	Comparison with segmentation .....	264
6.4.2	Transitivity .....	268
6.4.3	Linear regression .....	270
6.4.4	Power calculation .....	273
6.5	Discussion .....	275
6.5.1	Brain volumes in human and healthy mouse ageing .....	276
6.5.2	Treatments and power .....	277
6.5.3	BSI of substructures .....	279
6.5.4	Limitations and potential future improvements .....	280



6.6	Conclusions.....	283
7	<b>General discussion, conclusions and future work.....</b>	<b>285</b>
7.1	Can <i>in vivo</i> replace <i>ex vivo</i> ?.....	288
7.2	Future work .....	289
7.2.1	Further phenotyping investigations.....	291
7.2.2	Mouse brain atlasing .....	294
7.2.3	The future of animal studies.....	296
	<b>References.....</b>	<b>299</b>

# List of Figures

Figure 2.1	Example <i>ex vivo</i> and <i>in vivo</i> images of the same mouse brain. ....	38
Figure 2.2	3D surface rendering of a mouse brain.....	42
Figure 2.3	Example human brain average atlas images.....	44
Figure 2.4	Example slices from a single brain of the NUS atlas, with parcellations.....	46
Figure 2.5	Example T <sub>1</sub> W MRI human brain intensity histogram.....	50
Figure 2.6	Example <i>ex vivo</i> mouse brain T <sub>2</sub> W image intensity histogram.....	51
Figure 2.7	SPMMouse tissue prior maps.....	54
Figure 2.8	Examples of published mouse brain tissue classification results.....	55
Figure 2.9	Iterative group-wise registration illustrated using sagittal images from in-skull mouse brain MRI. ....	61
Figure 2.10	Local expansion and contraction in a deformation field.....	64
Figure 2.11	Explanation of modulation in VBM. ....	67
Figure 2.12	The pathological spread of tau in human Alzheimer’s disease.....	81
Figure 2.13	Theoretical progression of AD biomarker abnormality with time. ....	83
Figure 2.14	Human T <sub>1</sub> W MRI showing structural changes with MCI and AD.....	85
Figure 2.15	VBM-derived GM loss at different Braak stages.....	86
Figure 2.16	Illustration of the Boundary Shift Integral in the human brain.....	88
Figure 2.17	Some common trisomic mouse models and their syntenic regions with Hsa21.....	94
Figure 3.1	Overview of the modular pipeline steps.....	100
Figure 3.2	Mouse embryo holders. ....	102
Figure 3.3	High-throughput embryo images. ....	102
Figure 3.4	Touching subjects in multi-subject scans without separators.....	105
Figure 3.5	3D-printed brain separator and resulting MR image. ....	106
Figure 3.6	Extraction steps in coronal and sagittal views of a downsampled Tc1 brain image.....	108
Figure 3.7	Exploded CAD view of updated, 6-brain holder.....	108
Figure 3.8	Illustration of a mouse brain in RAS orientation.....	110
Figure 3.9	Determining optimal orientation via reflection.....	114
Figure 3.10	Overview of orientation correction steps, with coronal views. ....	116
Figure 3.11	Illustration of a mouse embryo in RAS orientation .....	117
Figure 3.12	NUC of multiple-subject brain and embryo images prior to extraction.....	118
Figure 3.13	Non-uniformity over-correction in an <i>ex vivo</i> brain. ....	119
Figure 3.14	Mouse brain masks after the STEPS procedure.....	123

Figure 3.15	Representative slices from a Tc1 brain illustrating probabilistic tissue segmentations. .	126
Figure 3.16	Improved tissue classifications. ....	128
Figure 3.17	Intensity standardisation results in the Tc1 (cohort 1, §4.1) dataset. ....	130
Figure 3.18	Group-wise registration average images ....	131
Figure 3.19	<i>Ex vivo</i> mean positional distance image from a cohort scanned in skull. ....	133
Figure 3.20	Plot of the overall mean standard deviation (SD) of voxel intensities, within the original target brain mask, of the resampled structural Tc1 and WT images at each stage of GWR.....	135
Figure 3.21	Final Tc1 NRR average image with UFL atlas parcellations overlaid. ....	136
Figure 4.1	Group-wise registration assessment for the Tc1 and WT groups.....	149
Figure 4.2	Total intracranial and brain volumes (TIV, BV) of WT (blue) and Tc1 (green).....	150
Figure 4.3	Parcellated WT and Tc1 region volumes from Jacobian integration.....	152
Figure 4.4	Mean Tc1 voxel-wise volume difference from mean WT volume.....	154
Figure 4.5	Tc1 TBM results.....	156
Figure 4.6	Peak t-statistic voxel values within selected significant clusters.....	157
Figure 4.7	Unmodulated VBM results ....	159
Figure 4.8	VBM in the cerebellum and hippocampus, with a narrower smoothing kernel.....	160
Figure 4.9	Comparing label propagation with tissue segmentation.....	166
Figure 4.10	Example of hyperintense rim, and its segmentation, in the Tc1 dataset. ....	167
Figure 4.11	Box plots of TIV and BV for all four groups, from tissue segmentation.....	172
Figure 4.12	TBM results: Tc1 vs. WT (cohort 2).....	173
Figure 4.13	TBM results: J20 vs. WT .....	174
Figure 4.14	TBM results: J20 vs. Tc1xJ20.....	175
Figure 5.1	Prior TBM results and parcellation in the rTg4510 mouse brain. ....	183
Figure 5.2	Longitudinal rTg4510 study design, showing mouse ages at each TP.....	186
Figure 5.3	Single-TP WT brain, <i>in vivo</i> , before and after initial NUC. ....	189
Figure 5.4	Intensity non-uniformity in an rTg4510 <i>in vivo</i> mouse brain image. ....	189
Figure 5.5	Sample group-wise registration parameter testing results. ....	192
Figure 5.6	Single time-point TBM results from 8-month-old rTg4510 mice.....	194
Figure 5.7	Visible atrophy in the rTg4510 brain, compared with relative preservation of WTs.....	196
Figure 5.8	TBM results at TP <sub>1</sub> of the cross-sectional multi-time-point study.....	198
Figure 5.9	TBM results for WT, UT and doxycycline-treated mice at TP <sub>2</sub> .....	200
Figure 5.10	Cross-sectional TBM results at TP <sub>3</sub> .....	202
Figure 5.11	Illustration of the registration steps for longitudinal TBM. ....	204
Figure 5.12	Longitudinal TBM results for UT vs. WT and UT vs. TR comparisons.....	206
Figure 5.13	Comparing the TIV measurement with the GWR affine determinant.....	209
Figure 5.14	TBM results in rTg4510 brains, comparing affine and TIV covariates. ....	210
Figure 5.15	<i>In</i> and <i>ex vivo</i> TBM, comparing UT rTg4510s with WTs. ....	214
Figure 5.16	<i>In</i> and <i>ex vivo</i> TBM, comparing TR rTg4510s with UTs.....	216
Figure 5.17	Mean positional distance image showing the mean distances moved by voxels during pair-wise registration from <i>ex vivo</i> to <i>in vivo</i> data. ....	218

Figure 5.18	Power analysis results in rTg4510 data: Cohen's d and sample sizes.....	222
Figure 5.19	Selected anatomical features of the meninges and skull in an <i>ex vivo</i> rTg4510 brain (10 months of age), after severe atrophy due to tauopathy.....	224
Figure 5.20	GWR averages of an <i>ex vivo</i> rTg4510 and WT littermate cohort (10 months of age), after one iteration of.....	225
Figure 6.1	Progressive atrophy in the human brain revealed with difference images. ....	232
Figure 6.2	Progressive atrophy clearly visible in T <sub>2</sub> W rTg4510 mouse brain images.....	235
Figure 6.3	Overview of pre-processing steps prior to BSI calculation .....	239
Figure 6.4	Binary brainstem region in average space.....	243
Figure 6.5	Illustration of symmetric registration to a middle space, for three TPs. ....	244
Figure 6.6	Probabilistic masks, and pXOR region.....	247
Figure 6.7	Comparing T <sub>1</sub> W human and T <sub>2</sub> W mouse brain images .....	248
Figure 6.8	Idealised 1D boundary intensity profiles of (a) T <sub>1</sub> W single window BSI and (b) T <sub>2</sub> W double window BSI. ....	250
Figure 6.9	Intensity windows overlaid upon baseline and repeat image histograms. ....	252
Figure 6.10	WT and UT brain BSI result visualisations .....	257
Figure 6.11	Mean absolute brain volume loss (ml), from the BSI, per 30 days.....	260
Figure 6.12	Mean segmentation volume change measurements.....	261
Figure 6.13	Overall standardised PBVC per group, per 30 days.....	262
Figure 6.14	Effect sizes (difference of means from UT) for each treatment group.....	263
Figure 6.15	Segmented volumes vs. mouse age.....	266
Figure 6.16	BSI vs. segmentation-measured volume differences (in ml).....	267
Figure 6.17	Scatter plot illustrating BSI transitive error.....	269
Figure 6.18	Transitive error rate, per group.....	270
Figure 6.19	Linear regression of BSI measurements with scan interval between TPs. ....	271
Figure 6.20	Susceptibility artefact from the aural cavity causes false BSI results in two registered TP images of a WT brain.....	282

# List of Tables

<b>Table 2.1</b>	Selection of morphometry papers and pipeline features (previous page).....	78
<b>Table 3.1</b>	Manual binary classification of NUS atlas labels used to initialise EM.....	127
<b>Table 4.1</b>	Parcellated volume results for Tc1 and WT groups. ....	151
<b>Table 4.2</b>	Tissue volumes after segmentation in WT, Tc1, Tc1xJ20 and J20 mice.....	172
<b>Table 5.1</b>	Number of rTg4510 mice, and their mean ages, for the longitudinal study. ....	185
<b>Table 5.2</b>	Mean SNR and CNR for all rTg4510 scans in the longitudinal study.....	197
<b>Table 6.1</b>	Number of baseline and follow-up image pairs, and days between time-points, for the BSI investigation. ....	237
<b>Table 6.2</b>	WT and UT cerebellar volumes for C <sub>1</sub> , C <sub>2</sub> . ....	258
<b>Table 6.3</b>	Standardised BSI results (ml/30 days). ....	260
<b>Table 6.4</b>	Standardised segmentation-based volume change (ml/30 days).....	261
<b>Table 6.5</b>	Percentage brain volume change (PBVC) per group, per 30d. ....	262
<b>Table 6.6</b>	BSI and segmentation's ability to separate WT and UT groups.....	265
<b>Table 6.7</b>	Mean transitive error measurements, for each group.....	269
<b>Table 6.8</b>	Linear regression of BSI (ml) with scan interval (days), for each group.....	272
<b>Table 6.9</b>	Linear regression of segmentation difference (ml) with scan interval (days).....	272
<b>Table 6.10</b>	Sample sizes for a significant effect, using BSI or segmentation.....	274

## Publications arising from this work

### ***Peer-reviewed journal papers***

- **Powell, NM.**, Modat, M., Cardoso, MJ., Ma, D., Holmes, HE., Yu, Y., O'Callaghan, J., Cleary, JO., Sinclair, B., Wiseman, FK., Tybulewicz, VLJ., Fisher, EMC., Lythgoe, MF., Ourselin, S. Fully-automated  $\mu$ MRI morphometric phenotyping of the Tc1 model of Down syndrome. (In press.)
- **Powell, NM.**, Prados, F., Cardoso, MJ., Modat, M., Ma, D., Colgan, N., O'Neill, M., Collins, EC., Lythgoe, MF., Ourselin, S. The Boundary Shift Integral sensitively reveals longitudinal whole-brain atrophy in a mouse model of tauopathy. (Submitted to journal.)
- Ma, D., Cardoso, MJ., Modat, M., **Powell, N.**, Wells, J., Holmes, H., Wiseman, F., Tybulewicz, V., Fisher, E., Lythgoe, MF., Ourselin, S. (2014). Automatic structural parcellation of mouse brain MRI using multi-atlas label fusion. *PLOS ONE*, 9(1), e86576. doi:10.1371/journal.pone.0086576
- Wells, JA., O'Callaghan, JM., Holmes, HE., **Powell, NM.**, Johnson, RA., Siow, B., Torrealdea, F., Ismail, O., Walker-Samuel, S., Golay, X., Rega, M., Richardson, S., Modat, M., Cardoso, MJ., Ourselin, S., Schwarz, AJ., Ahmed, Z., Murray, TK., O'Neill, MJ., Collins, EC., Colgan, N., Lythgoe, MF. (2015). In vivo imaging of tau pathology using multi-parametric quantitative MRI. *NeuroImage*, 111, 369–378. doi:10.1016/j.neuroimage.2015.02.023
- Holmes, HE., Colgan, N., Ismail, O., Ma, D., **Powell, NM.**, O'Callaghan, JM., Harrison, IF., Johnson, RA., Murray, TK., Ahmed, Z., Heggenes, M., Fisher, A., Cardoso, MJ., Modat, M., Walker-Samuel, S., Fisher, EMC., Ourselin, S., O'Neill, MJ., Wells, JA., Collins, EC., Lythgoe, MF. (2016). Imaging the accumulation and suppression of tau pathology using multi-parametric MRI. *Neurobiology of Aging*. doi:10.1016/j.neurobiolaging.2015.12.001
- Holmes, HE.\* , **Powell, NM.\***, Ismail, O., Ma, D., Harrison, IF., Wells, JA., Colgan, N., O'Callaghan, JM., Johnson, RA., Murray, TK., Ahmed, Z., Heggenes, M., Fisher, A., Cardoso, MJ., Modat, M., O'Neill, MJ., Collins, EC., Fisher, EMC., Ourselin, S., Lythgoe, MF. Comparison of in vivo and ex vivo MRI for the detection of structural abnormalities in a mouse model of tauopathy. (\* Joint first authors. In preparation.)
- Clayton, EL., Mancuso, R., Nielsen, TT., Mizielinska, S., Holmes, HE., **Powell, NM.**, Norona, F., Larsen, JO., Lythgoe, MF., Ourselin, S., Nielsen, JE., 7, Johannsen, P., Holm, I.,

Collinge, J., Frontotemporal Dementia Research in Jutland Association, Oliver, P., Gomez-Nicola, D., Isaacs, AM., Early microglial activation precedes neuronal loss and behavioural impairment in mice with a frontotemporal dementia causing CHMP2B mutation. (In preparation.)

### ***Peer-reviewed conference abstracts and papers***

- **Powell, NM.,** Modat, M., Cardoso, MJ., Ma, D., Holmes, H., Norris, F., Lythgoe, MF., Ourselin, S. (2012). Automated high-throughput morphometric phenotyping of mouse brains and embryos. Proc. British Chapter of the International Society for Magnetic Resonance in Medicine (ISMRM), (Cambridge). Oral O5. *And* British Chapter ISMRM Postgraduate Symposium (London). Poster.
- **Powell, N.,** Modat, M., Cardoso, MJ., Ma, D., Lythgoe, MF., Ourselin, S. MouseMorph: A complete, automatic pipeline for processing preclinical  $\mu$ MRI subjects for morphometric phenotyping. (2013). European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), (Toulouse). Software exhibit 699.
- **Powell, NM.,** Modat, M., Cardoso, MJ., Ma, D., Holmes, HE., Norris, F., Lythgoe, MF., Ourselin, S. (2013). A software toolbox for the automated high-throughput  $\mu$ MRI phenotyping of mouse brains and embryos with VBM and TBM. Proc. British Chapter ISMRM (York). Poster.
- **Powell, NM.,** Modat, M., Cardoso, MJ., Ma, D., Tybulewicz, VLJ., Wiseman, F., Fisher, EMC., Lythgoe, MF., Ourselin, S. (2014). Fully automatic phenotyping of the Tc1 mouse model of Down syndrome with TBM and VBM. British Chapter ISMRM Postgraduate Symposium (Cardiff). Oral O16.
- **Powell, NM.,** Holmes, HE., Ma, D., Modat, M., Cardoso, J., Wiseman, FK., Tybulewicz, VLJ., Fisher, EMC., Lythgoe, MF., Ourselin, S. (2015). Tensor-Based Morphometry reveals structural differences between Down syndrome and Alzheimer's disease mouse models. Alzheimer's Research UK Conference (London). Poster.
- ‡ **Powell, NM.,** Ma, D., Prados, F., Modat, M., Cardoso, J., Holmes, HE., Ismail, O., Colgan, N., O'Neill, M., Collins, EC., Lythgoe, MF., Ourselin, S. (2015). Longitudinal whole-brain atrophy measurement in a mouse model of tauopathy using the Generalised Boundary Shift Integral. Proc. ISMRM (Toronto). Poster 6268.
- **Powell, NM.,** Holmes, HE., Ma, D., Modat, M., Cardoso, J., Wiseman, FK., Tybulewicz, VLJ., Fisher, EMC., Lythgoe, MF., Ourselin, S. (2015). Tensor-Based Morphometry reveals structural differences between Down syndrome and Alzheimer's disease mouse model brains. Proc. ISMRM (Toronto). Poster 6388.

- Holmes, HE., Wells, JA., Siow, B., O'Callaghan, JM., Richardson, S., **Powell, N.**, Ma, D., Fisher, EM., Ourselin, S., Lythgoe, MF. (2012). Optimisation of the in vivo sequence for structural imaging of a mouse model of Alzheimer's disease. British Chapter ISMRM (Cambridge). Poster. *And* Postgraduate Symposium (Bristol). Poster.
- Ma, D., Cardoso, MJ., Modat, M., **Powell, N.**, Holmes, H., Lythgoe, MF., Ourselin, S. (2012). Multi-atlas structural parcellation for in vivo quantification of mouse brain anatomy. British Chapter ISMRM.
- Ma, D., Cardoso, MJ., Modat, M., **Powell, N.**, Holmes, H., Lythgoe, MF., Ourselin, S. (2012). Multi-atlas segmentation applied to in vivo mouse brain MRI. MICCAI Workshop on Multi-Atlas Labeling.
- Holmes, HE., **Powell, N.**, Wells, JA., Colgan, N., O'Callaghan, JM., Ma, D., Modat, M., Cardoso, MJ., Richardson, S., Siow, B., O'Neill, MJ., Collins, EC., Fisher, EM., Ourselin, S., Lythgoe, MF. (2013). Tensor-based morphometry as a sensitive biomarker for Alzheimer's disease neuropathology. Proc. ISMRM (Salt Lake City). E-poster 3679. *And* British Chapter ISMRM (York). *And* Postgraduate Symposium (London). Poster P10.
- Holmes, HE., Wells, JA., O'Callaghan, JM., **Powell, N.**, Siow, B., Ma, D., Richardson, S., Ismail, O., Cardoso, MJ., Modat, M., Fisher, EM., Ourselin, S., Collins, EC., O'Neill, MJ., Colgan, N., Lythgoe, MF. (2013). ASL, DTI and structural MRI: multiparametric biomarkers of Alzheimer's disease neuropathology in the Tg4510 mouse model of tauopathy. Proc. ESMRMB (Toulouse). Oral 375.
- Holmes, HE., **Powell, NM.**, Wells, J., Colgan, N., O'Callaghan, JM., Ma, D., Modat, M., Cardoso, MJ., Richardson, S., Siow, BM., O'Neill, MJ., Collins, EC., Fisher, E., Ourselin, S., Lythgoe, MF. (2013). Morphometric Genomics: in vivo microMRI for 3D structural imaging of transgenic mice. Proc. ISMRM.
- Yu, Y., **Powell, NM.**, Ma, D., Bryson, J., Greensmith, L., Fisher, E., Lythgoe, M. (2013). Investigation of brain structural changes seven days after sciatic nerve transection. Proc. British Chapter ISMRM.
- Yu, Y., Kenny, G., Zhang, L., **Powell, NM.**, Ourselin, S., Gaspar, HB., Lythgoe, MF. (2014). Mouse model of ADA deficiency shows similar brain abnormalities as human patients. Proc. ISMRM-ESMRMB Joint Annual Meeting.
- Goodburn, R., **Powell, N.**, O'Callaghan, J., Walker-Samuel, S., Shmueli, K. (2015). Characterising and Modelling Susceptibility Artifacts in the Mouse Brain at 9.4 T. Proc. British Chapter ISMRM (London).
- Holmes, HE., Ismail, O., Wells, JA., O'Callaghan, JM., **Powell, N.**, Murray, TK., Ahmed, Z., O'Neill, MJ., Collins, EC., Johnson, RA., Colgan, N., Lythgoe, MF. (2015). An imaging strategy to characterise tau pathology in a mouse model of Alzheimer's disease using multi-



parametric MRI. Alzheimer's Assoc. Int'l. Conference (Washington DC). Poster P158. *And* (2014). British Chapter ISMRM (Edinburgh).

- Holmes, HE., Colgan, N., Ismail, O., Ma, D., Wells, JA., **Powell, N.**, O'Callaghan, JM., Harrison, I., Cardoso, MJ., Modat, M., Fisher, EM., Ourselin, S., O'Neill, MJ., Collins, EC., Lythgoe, MF. (2015). A multi-scale MRI approach to investigate novel drug treatment strategies in mouse models of Alzheimer's disease. Proc. ISMRM (Toronto), poster 2221.
- Holmes, HE., Ramasawmy, R., Ma, D., **Powell, N.**, Cardoso, MJ., Modat, M., Walker-Samuel, S., Ourselin, S., Siow, B., Lythgoe, MF. (2015). Is 1T the new 9.4T? A tool for morphological phenotyping and regional brain volume analysis. Proc. ISMRM (Toronto), E-poster 3518. *And* British Chapter ISMRM (London), poster P18.
- † Holmes, HE., **Powell, N.**, Wells, JA., Colgan, N., Ismail, O., O'Callaghan, JM., Ma, D., O'Neill, MJ., Collins, EC., Cardoso, MJ., Modat, M., Fisher, EM., Ourselin, S., Lythgoe, MF. (2015). Comparing *in vivo* and *ex vivo* imaging in an Alzheimer's model using tensor-based morphometry. Proc. ISMRM (Toronto), oral O394. *And* British Chapter ISMRM (London), oral O26.
- Ismail, O., Holmes, H., Colgan, N., Ma, D., Wells, JA., **Powell, NM.**, O'Callaghan, JM., Harrison, IF., Walker-Samuel, S., Cardoso, J., Modat, M., Fisher, E., Ourselin, S., Murray, TK., Ahmed, Z., O'Neill, MJ., Johnson, RA., Collins, EC., Lythgoe, MF. (2015). A Multiscale MRI Approach to Investigate Novel Drug Treatment Strategies in Mouse Models of Alzheimer's Disease. Alzheimer's Association International Conference. *And* British Chapter ISMRM Postgraduate Symposium.
- Ismail, O., Harrison, IF., Holmes, H., Colgan, N., Wells, JA., O'Callaghan, JM., Walker-Samuel, S., **Powell, NM.**, Ma, D., Ourselin, S., Murray, TK., Fisher, A., Ahmed, Z., O'Neill, MJ., Johnson, RA., Collins, EC., Lythgoe, MF. (2015). Imaging the Efficacy of Microtubule Stabilizing Agent Epothilone D in the rTg4510 Mouse Model of Tauopathy. Alzheimer's Association International Conference. *And* British Chapter ISMRM (London).
- Ma, D., Cardoso, MJ., Zuluaga, MA., Modat, M., **Powell, N.**, Wiseman, F., Tybulewicz, V., Fisher, E., Lythgoe, MF., Ourselin, S. (2015). Grey matter sublayer thickness estimation in the mouse cerebellum. Proc. Medical Image Computing and Computer Assisted Intervention 2015 (648).

† *Summa cum laudae*

‡ *Magna cum laudae*

# Abbreviations

2/3D	two-/three-dimensional	
3Rs	replacement, refinement and reduction of research animals (NC3Rs)	
AD	Alzheimer's disease	§2.3
AN(C)OVA	analysis of (co)variance	
APP	amyloid precursor protein	§2.3, §4.2
BG	background (signal)	
BSI	boundary shift integral	§6
BV	brain volume	
CAD	computer-aided design (software)	
CNR	contrast-to-noise ratio	E4.2
CPP	control point positions	
CPU	central processing unit	
CSF	cerebrospinal fluid; eCSF: external; vCSF: ventricular	
CT	computed tomography; $\mu$ CT: microscopic	
DBM	deformation-based morphometry	
DOF	degrees of freedom	
DS	Down syndrome	§2.4
DTI	diffusion tensor imaging	
EM	expectation maximisation (segmentation)	§2.2.4
FDR	false discovery rate	
FFD / F <sub>3</sub> D	free-form deformation / fast FFD	
FSE	fast spin echo	
FSL	The Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library	
FTD	frontotemporal dementia	
FOV	field of view	
FWHM	full width at half maximum	

Gd-DTPA	gadolinium-diethyltriamine pentaacetic acid (Magnevist)	
GE(3D)	(3D) gradient echo	
GLM	general linear model	§2.2.7
GM	grey matter	
GPU	graphics processing unit	
GWR	group-wise registration	§2.2.6
$J_{det}$	Jacobian determinant	E2.1
LNCC	locally normalised cross-correlation	
MCI	mild cognitive impairment	
MR(I)	magnetic resonance (imaging); $\mu$ MRI: microscopic	
MRF	Markov random field	
NFTs	Neurofibrillary tangles	§2.3, §5.1
NIfTI(-1)	Neuroimaging Informatics Technology Initiative (data format version 1)	
NMI	normalised mutual information	
NRR	non-rigid registration	
NUC	non-uniformity correction (of intensities)	§3.4
PBVC	percentage brain volume change	§6; E6.10
PV	partial volume	
ROI	region of interest	
SD	standard deviation	
SNR	signal-to-noise ratio	E4.1
SPM	statistical parametric map (or Statistical Parametric Mapping software, <a href="http://www.fil.ion.ucl.ac.uk/spm/software">www.fil.ion.ucl.ac.uk/spm/software</a> )	
STEPS	Similarity and Truth Estimation for Propagated Segmentations	§2.2.3
$T_1$ , $T_2$ , $T_2^*$	Longitudinal, transverse, and transverse (gradient echo) magnetisation relaxation time constants; $T_1W$ : images weighted by $T_1$ (etc.)	
TBM	tensor-based morphometry	§2.2.7
TIV	total intracranial volume	E3.13
TP	time-point	
TPM	tissue probability map	§2.2.5
VBM	voxel-based morphometry	§2.2.7
WM	white matter	
WT	wild-type	

### ***Brain atlases frequently referenced***

NUS	National University of Singapore – mouse brain, <i>in vivo</i> . See: <a href="http://www.bioeng.nus.edu.sg/cfa/mouse_atlas.html">www.bioeng.nus.edu.sg/cfa/mouse_atlas.html</a> and <i>Bai et al. (2012)</i> .
UFL	University of Florida (USA) Magnetic Resonance Microimaging Neurological Atlas (“MRM NeAr”) – mouse brain, <i>in</i> and <i>ex vivo</i> . See: <a href="http://brainatlas.mbi.ufl.edu">brainatlas.mbi.ufl.edu</a> , <i>Ma et al., (2005)</i> and <i>Ma et al., (2008)</i> .

### ***Acronyms for organisations***

ADNI	Alzheimer’s Disease Neuroimaging Initiative, <a href="http://adni.loni.usc.edu">adni.loni.usc.edu</a>
CABI	UCL Centre for Advanced Biomedical Imaging, <a href="http://www.ucl.ac.uk/cabi">www.ucl.ac.uk/cabi</a>
CMIC (TIG)	UCL Centre for Medical Image Computing, (Translational Imaging Group), <a href="http://cmictig.cs.ucl.ac.uk">cmictig.cs.ucl.ac.uk</a>
FDA	United States Food and Drug Administration, <a href="http://www.fda.gov">www.fda.gov</a>
IKMC	International Knockout Mouse Consortium
IMPC	International Mouse Phenotyping Consortium, <a href="http://www.mousephenotype.org">www.mousephenotype.org</a>
MRC	Medical Research Council, <a href="http://www.mrc.ac.uk">www.mrc.ac.uk</a>
NC3Rs	National Centre for the Replacement, Refinement & Reduction of Animals in Research, <a href="http://www.nc3rs.org.uk">www.nc3rs.org.uk</a>

### ***A note on notation***

Throughout, §1.1 refers to Chapter 1, section 1; E1.1 to Chapter 1, equation 1; Fig 1.1 to figure 1 in Chapter 1 (etc.); footnotes are numbered as<sup>1</sup> per chapter and *references* are ordered alphabetically and all included at the end.

# 1 Introduction

This chapter introduces the context in which this thesis rests – including the motivations behind genetic and disease mouse model phenotyping, and for translating clinical image analysis techniques to the preclinical paradigm.

## 1.1 The promise of MRI

Apart from the satisfaction of scientific curiosity, the chief motivation behind medical imaging technologies – pioneered by Wilhelm Röntgen’s x-rays in 1895 – is to noninvasively diagnose pathologies and help plan for their treatment. The desire to see within the body without cutting tissues – causing additional damage and perhaps exposing the inner organs to infection (heightened risks in a population of ageing patients) – has driven billions of pounds from governments and companies into imaging research over the past century. Endoscopy, positron emission tomography (PET), ultrasound and x-ray computed tomography (CT), each with unique advantages, have cemented themselves in hospitals worldwide.

To the horror of radiology’s first practitioners, high doses of x-rays were linked directly with terminal cancer (*Sansare et al., 2011*). Thus, although still used routinely thanks to its low cost and the relative safety of low exposures, and although capable of very high tissue contrast and detail (for example, *Wong et al., 2012*), the use of ionising radiation for imaging and treatment remains a balance of risks with benefits.

Magnetic resonance imaging (MRI), developed in the 1970s, is an attractive alternative, despite its relatively high cost (a high field strength, 9.4T small animal scanner may cost several million pounds). The extraordinary resolution and contrast achievable with MR (§2.1), and its diagnostic utility, have led to increased use over the past 20 years (*Lang et al., 2013*). MR techniques and biomarkers are an established facet in the study and diagnosis of one of the most important contemporary diseases: dementia (§2.3).

Among MR's advantages, its use of non-ionising radiation must be one of the greatest: the modality is truly non-invasive and, provided a patient is not fitted with metallic implants – safe. Different structural MR contrasts give complementary information on the physical and chemical properties of tissue, and various techniques, such as diffusion tensor imaging (DTI, for measuring connectivity); arterial spin labelling (ASL, for blood flow and volume); and functional MRI (fMRI, via blood oxygenation) can provide quantitative assessments of *in vivo* physiology. The digital, three-dimensional (3D) images can be interpreted by clinicians directly or automatically analysed using computational techniques, providing objective, quantitative assessments of subtle pathologies, such as lesions in multiple sclerosis and atrophy in the early stages of Alzheimer's disease.

The advance of high magnet field strengths and availability of hardware, such as signal receiver coils, specifically designed for preclinical use, has enabled microscopic magnetic resonance imaging ( $\mu$ MRI) with fine resolutions (10—100 $\mu$ m per voxel) and high contrasts to probe smaller structures, such as those in the brains of small animal models, and to detect subtle pathological and developmental effects (*Benveniste & Blackband, 2002; Nieman et al., 2005; 2011*).

## 1.2 Phenotyping mice

The mouse – quick to mature; cheap to house; with a genome we can engineer to model human disorders, and orthologs of 99% of human genes among its own 20—25,000 (*Gunter & Dhand, 2002; Rosenthal & Brown, 2007*) – is a crucial human surrogate in the

study of gene function, the observation of disease progression, and the search for effective drugs (*McConville et al., 2005; Collins et al., 2007; Henkelman, 2010*).

Physical and behavioural traits – determined both by genetics and environment – define an organism's phenotype. Mouse genetic, disease and behavioural models are phenotyped to confirm accurate recapitulation of human pathology, to determine behavioural correlates, such as neuroplasticity in learning, and to observe the effects of abnormalities at different developmental stages (e.g. *Cryan & Holmes, 2005; Cleary et al., 2009; Sawiak et al., 2009b; Lerch et al., 2011; Norris et al., 2013*). The validity of models is vital for testing therapeutics and evaluating the risks of inheriting certain genes.

Owing to its importance as a model organism, after *Homo sapiens* (*IHGSC et al., 2004*), only the second mammalian genome to be fully sequenced was that of the mouse (*Mus musculus* strain C57Bl/6J – *Church et al., 2009*). Knowledge of this sequence, however, does not elucidate individual genes' interactions, or their function. This challenge is the target of the International Mouse Phenotyping Consortium (IMPC), a large-scale ongoing study involving 18 centres worldwide, with aims to comprehensively phenotype mouse strains in which genes are individually disabled (knocked out), thereby determining that gene's effects under controlled conditions. A phenotype may encompass a broad spectrum of characteristics: two thirds of strains have pleiotropic phenotypes (with multiple effects). A phenotype may express differently at different ages. Easily-assessed features (such as weight, litter size, fur or eye colour, gait, behaviour, etc.) are captured by 'first-line' phenotyping, but subtler structural biomarkers may warrant 'second-line' approaches, such as histology or  $\mu$ MRI, potentially increasing the rate of phenotype discovery (*Fuchs et al., 2010; Brown & Moore, 2012*).

This study is primarily concerned with structural features of the mouse brain, although embryos are also briefly considered. In the brain, structure is thought to be inherently tied to function, with structural changes such as atrophy in neurodegeneration severely affecting cognitive ability (*Fox et al., 1999*), and plasticity evident with learning (*Maguire et al., 2000; Draganski et al., 2004; Lerch et al., 2011*).

The vast number of genetic knock-outs, knock-ins, disease models, large cohorts, and the increasing usefulness of  $\mu$ MRI as a phenotyping tool, motivate a ‘high-throughput’ approach to phenotyping (*Norris et al., 2013*), such as simultaneously scanning multiple brains. Automated analysis is essential for the large, high-resolution datasets produced, and image processing tools must be designed to cope. There is currently an unmet need for such tools.

## 1.3 Morphometry in preclinical studies

Morphological variability can arise from mouse age, sex, genetic background, and environment. In-bred mice (of more than 20 generations) are isogenic, maintaining consistent characteristics. This reduced and manipulable gene pool eases comparisons between studies (*Beck et al., 2000*); in-bred mice exhibit less variability than humans. Mice are heterogeneous between strains – exhibited in behavioural phenotypes (*Rogers et al., 1999*) and brain structure, with much smaller differences within-strain (*Chen et al., 2005*). Within C57Bl/6J (“black six”) mice, one of the most commonly-used strains, *Kovačević et al. (2005)* found over 90% of brain structures exhibited under 180 $\mu$ m positional variability (even after physical extraction from the skull, a potentially damaging procedure, *Badea et al., 2007a*).

Physical changes in the brain and developing embryo are detectable with traditional techniques. Histology offers unparalleled detail, sensitivity and specificity: often, individual cells may be stained, visualised, and counted (*Cleary et al., 2011b*). However, the technique is prohibitively time-consuming when considering the whole brain, so is limited to small pre-defined regions of interest within a small number of animals. It is also reliant upon intact, delicately-preserved tissues, but is necessarily destructive, and therefore non-repeatable. Phenotypes can be partially penetrant and not appear in every animal. Rather than comparing individuals with a known “wild-type” strain, therefore, statistical techniques can be wielded to compare the average phenotype, across many animals.



In concert with  $\mu$ MRI, powerful statistical morphological analysis techniques such as Voxel- and Tensor-Based Morphometry (V/TBM, described in more detail in §2.2.7) enable non-invasive, hypothesis-free structural investigations covering an entire organ or organism. VBM employs maps of grey and white matter (GM/WM) to reveal tissue density changes in the brain. TBM is used to detect volumetric differences in physical structures (*Ashburner & Friston, 2000*). Both offer exploratory, whole-brain assessment, without operator bias – particularly pertinent to phenotyping, genetic, or disease studies. They reduce the requirement for histology or the laborious expert delineation of regions of interest (ROIs), which are vulnerable to intra- and inter-rater variability and may miss unexpected changes in unexplored regions (*White et al., 2003; Carducci et al., 2013*). These computational techniques have much to offer the field of preclinical phenotyping.

Despite their potential, and although V/TBM enjoy widespread use in human studies with popular software packages such as SPM<sup>1</sup>, FSL<sup>2</sup> and Freesurfer<sup>3</sup>, morphometry is not routinely employed preclinically (*Sawiak et al., 2013*; *Table 2.1* includes a selection). The image processing burden for large datasets remains high. Standard protocols for phenotyping mice with MRI have yet to be developed (*McConville et al. 2005; Scheenstra, 2011*), and several pre-processing steps, unique to high-throughput preclinical studies, must currently be performed manually, with disparate software. Few atlases – datasets containing structural MR mouse brain images and corresponding anatomical labels which may be used as prior knowledge – are publicly available (§2.2.1; *Ma et al., 2014*). These limitations form a bottleneck to automated analysis techniques including V/TBM, segmentation propagation, or the examination of individual substructures.

---

1: <http://www.fil.ion.ucl.ac.uk/spm>

2: <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki>

3: <http://freesurfer.net>

## 1.4 From *ex vivo* to *in vivo*

Despite its many advantages for phenotyping mouse brains, MRI, as discussed in the first half of this thesis and in the majority of preclinical morphometry studies, cannot be considered non-invasive; the animal is sacrificed to achieve greater spatial resolution, or contrast, or to preclude motion artefacts. While histology at the end of a study – perhaps preceded by *ex vivo* MR – might remain essential for many years, MR still has much to offer *in vivo*, where tissues and physiological processes remain intact (*Benveniste & Blackband, 2002*). Indeed, there are numerous advantages of longitudinal, multi-time-point studies over cross-sectional, single time-point investigations (§5). It is feasible to follow mouse development and ageing throughout the animal’s entire lifespan – difficult in humans both because of the time required and the number of uncontrollable lifestyle factors (*Fjell et al., 2014*) – and monitor the trajectory of disease and treatment effects without confounding inter-individual variability. The Boundary Shift Integral (BSI, §6, *Freeborough & Fox, 1997*) is one technique for doing so, by measuring voxel intensity differences between scans.

Ethical justifications for sometimes controversial animal research rest upon the lack of viable alternatives for both phenotyping and drug studies. Mice are intimately involved in the drug development pipeline: both for identifying and validating drug targets, and eventual testing of new therapeutic compounds (*Hall & Roberson, 2012*). To ensure their safety, animal studies are mandatory for all drugs to be approved by the Food and Drug Administration (FDA) in the United States<sup>4</sup>. Over four million animals were used for scientific research in both 2012 and 2013 in the UK<sup>5</sup>. The National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) aims to reduce this number, by improving the quality and reproducibility of such studies. One method for doing so in neurological studies which investigate disease progression, or a gene’s effects, over time, is to observe the same animals longitudinally, with multiple

---

4: “The FDA’s Drug Review Process: Ensuring Drugs Are Safe and Effective”, <http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm>

5: <http://www.understandinganimalresearch.org.uk/animals/numbers-animals>

time-points – rather than culling different animals at different ages (as, for example, [Badea et al., 2010](#); [Chuang et al., 2011](#)). Being non-invasive, MRI is well-suited to such investigations.

There remain some advantages to *ex vivo* MR ([§5.7](#)): extended scan times enable higher resolutions and signal-to-noise ratios (SNR). Maintaining animals on anaesthesia, *in vivo*, is difficult ([Balaban & Hampshire, 2001](#)). Contrast agents – which may otherwise be neurotoxic ([Lerch et al., 2012](#)) – enable improved tissue contrast and image quality. Multiple brains may be easily scanned in parallel ([§3.2.1](#)), and motion artefacts are eliminated. In [§5](#), the phenotyping pipeline developed for *ex vivo* MRI is applied to *in vivo*, longitudinal data, with appropriate parameter adjustments. Morphometric results from *ex vivo* and *in vivo* scans of the same animals are compared. If the latter are able to show the same phenotypic effects of disease, despite the image contrast and resolution trade-offs, a compelling case may be made for the exclusive use of *in vivo* MRI for investigations of brain structure in preclinical studies.

## 1.5 Aims and contributions of this research and thesis

The aim of this thesis is to present the development, translation and application of two software pipelines, for Voxel- and Tensor-Based Morphometry, and the Boundary Shift Integral – both established clinical MR image processing techniques – to the preclinical paradigm. As proof of principle, I show and discuss results from each in mouse models of Down syndrome and tauopathy in Alzheimer’s disease. The main contributions of this work are:

1. The development of a cohesive, automated software pipeline for analysing multiple mouse brain and embryo images, using advanced, up-to-date MR image analysis techniques employed in the clinical paradigm to reduce bias and measurement error, such as multi-atlas parcellation and segmentation, and symmetric registration. I describe

novel techniques for multiple subject extraction, global orientation correction, and improved tissue segmentations, prior to voxel-wise statistical morphometry (§3). I demonstrate the application of this pipeline to preclinical genetic phenotyping studies and disease models, as proof both of principle and of high-throughput robustness, anticipating large cohorts.

2. Using this pipeline with TBM and VBM, I show and discuss novel structural findings in the brains of several mouse models, including the Tc1 model of Down syndrome (DS, §4); the Tc1xJ20 model (a DS model cross-bred with an Alzheimer's disease model), and the rTg4510 model of tauopathy (§5). Appropriate software parameters for investigative phenotyping with both *ex vivo* and *in vivo* data are described, and the statistical power of preclinical morphometry is investigated. The rTg4510 study involved a large, cross-sectional and longitudinal analysis, comparing untreated animals with wild-types (WTs) and those treated with two drugs, and afforded the opportunity to compare both drug effectiveness and *in vivo* and *ex vivo* morphometry in the same animals, and to assess at which ages the technique could differentiate groups.
3. Continuing the analysis of longitudinal data, I adapted the Boundary Shift Integral (BSI) algorithm for measuring whole-brain atrophy in T<sub>2</sub>-weighted mouse brain images, and show and discuss the results of its application to a large *in vivo* mouse brain dataset. This is the BSI's first application to a non-human animal. In evaluating the technique, I show that the method is a sensitive and promising measure for the early detection and quantification of atrophy in the mouse brain, and for future use in preclinical therapeutic trials (§6).

### ***Ethics statement***

All studies in this thesis were conducted following approval from an internal UCL ethical review, and were authorised by the UK Home Office, Animals (Scientific Procedures) Act 1986 under the relevant Project Licence authority. I strove for high standards in the design and conduct of animal-related work, and took full consideration of the 3Rs (as introduced above).

## 1.6 Overview of chapters

The structure of the following six chapters comprising this thesis is laid out below. There are two introductory chapters, four methodological and experimental chapters, and a concluding discussion. Each chapter has its own introduction and discussion. A bibliography of all references follows the final conclusions.

### Chapter 2: Background theory and state-of-the-art practices

This chapter summarises established voxel-wise methods for the analysis of clinical data, Voxel- and Tensor-Based Morphometry, and the Boundary Shift Integral. Both techniques harness the wealth of information present in high-resolution, microscopic MR images, by considering data from all voxels. Image processing theory, atlases, and state-of-the-art methods for registration and segmentation are also described, upon which these methods are dependent, and some preclinical studies employing morphometric techniques are discussed. Finally, to provide motivating context for the mouse models employed in this thesis, I introduce the aetiology and current MRI-based study of Alzheimer’s disease and Down syndrome.

### Chapter 3: Development of a high-throughput software pipeline for phenotyping preclinical subjects

This chapter describes the translation and application of the techniques introduced in §2 to the preclinical paradigm, for the development of a high-throughput, fully automatic morphometric image processing pipeline designed to phenotype large cohorts of mouse brains and meet the needs described in §1. The considerations necessary for high-throughput embryo data are briefly included. I also discuss suitable atlases, parameters, and some of the existing literature concerning each pipeline step.

I incorporate an accessible approach for simultaneously scanning multiple *ex vivo* brains, requiring only a 3D-printed brain holder. The software toolbox I developed includes novel pre-processing methods for the separation of multiple subjects from a single MRI

scan, their orientation to a standard space, skull-stripping based upon publicly-available atlases and tissue segmentation using accurate subject-specific tissue prior maps. These steps are followed by group-wise registration and the generation of statistical parametric maps based upon the resulting deformation fields.

## Chapter 4: Application of the pipeline to *ex vivo* mouse brains

Here, I present the application of the morphometric analysis pipeline developed in §3 to *ex vivo* mouse brain datasets, describe some of the considerations necessary in each case, and show and discusses results.

I report novel brain morphological phenotypes in the Tc1 mouse model of Down syndrome, as compared with wild-type littermates. I also analyse the Tc1xJ20 Down syndrome and APP (Alzheimer's disease) cross-bred mouse model, to investigate how the two phenotypes interact in the brain.

## Chapter 5: *In vivo* morphometry

This chapter describes the application and refinement of the pipeline for *in vivo* data, specifically the rTg4510 mouse model of tauopathy in Fronto-Temporal Dementia and Alzheimer's disease, in collaboration with Eli Lilly. Whereas the previous chapter necessarily dealt with cross-sectional, single time-point analyses, and while *ex vivo* scans may provide higher image resolution and contrast, *in vivo* scans introduce the possibility of longitudinal study. The structural integrity of brains is also preserved. This has the potential to reduce the number of animals used in preclinical research, and hence lower costs. Over 250 scans were analysed, including a cross-sectional study and 87 animals with up to three time-points. I examined whether *in vivo* morphometry could sufficiently capture the findings of *ex vivo*, by comparing TBM in the same brains.

## Chapter 6: Longitudinal atrophy measurement using the Boundary Shift Integral in the mouse brain

Finally, continuing the large longitudinal study of §5, I describe a translation of the Boundary Shift Integral, a sensitive and objective clinical measure of atrophy, to the brain of the rTg4510 mouse model. This is the first application of the BSI to a non-human animal, and its first application to T<sub>2</sub>-weighted MR data, and required adaptations to the code and clinical pipeline. I evaluated the power of the technique, using sample size estimates, to detect changes in atrophy rate. The results indicate the BSI's ability to measure subtle longitudinal brain volume changes and differentiate groups, which may be of significant use in preclinical therapeutic drug trials.

## Chapter 7: General discussion, conclusions and future work

A summary and conclusion drawn from all the findings and work completed for this thesis. Possible future work, limitations, and recommendations for moving forward with preclinical image analysis are also discussed, including the continuation of phenotyping investigations, the development of atlases, and the further use of image analysis tools to investigate preclinical models.



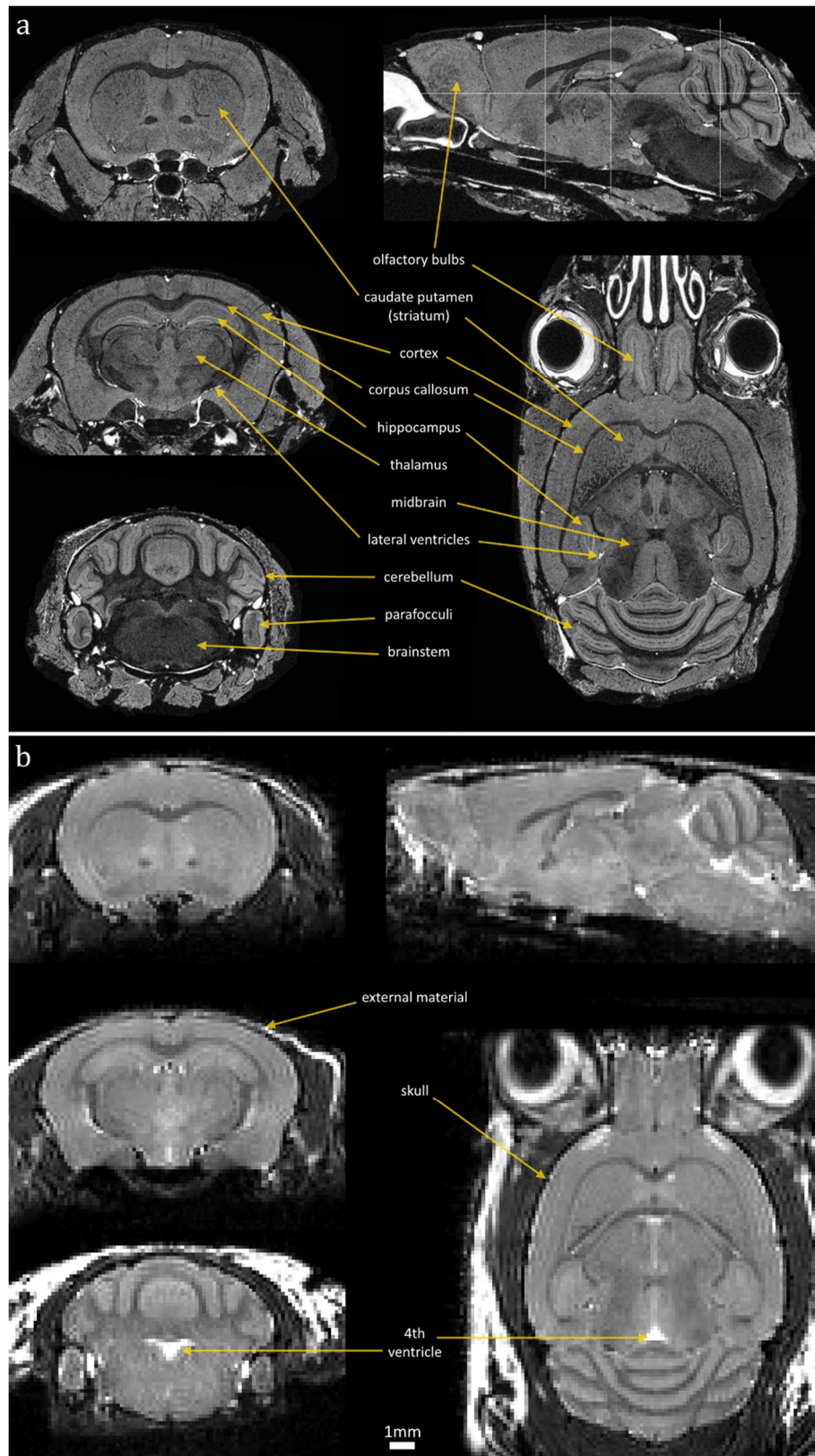


## 2 Background theory and state-of-the-art practices

This chapter describes the theoretical principles and existing state-of-the-art concepts and methods upon which the research in this thesis is built. I show examples of mouse brain images analysed, whose acquisitions have previously been optimised, and discuss the MR contrast and image quality as it pertains to the processing techniques I employed. Existing methods to analyse MR images are next discussed, including registration, segmentation and statistical morphometry methods. I include a summary of existing pipelines for morphometry in preclinical studies, covering the steps detailed in §3. My work concerned the investigation of mouse models of Alzheimer’s disease and Down syndrome, hence, finally, both conditions are briefly introduced, along with biomarkers and treatments – investigated further in §6. The mouse models I investigated are introduced in their respective chapters: Tc1 in §4.1; Tc1xJ20 in §4.2; rTg4510 in §5.1.

### 2.1 MR images of the mouse brain

Researchers at CABI, including Jon Cleary, Ben Sinclair, Francesca Norris, Holly Holmes, James O’Callaghan and Yichao Yu all worked to develop and optimise the high-resolution, high-contrast mouse brain and embryo fixation and scan protocols used to acquire each image processed as part of this thesis. Fig 2.1 shows example *ex vivo* and *in vivo* scans acquired from the same brain.



**Figure 2.1** Example *ex vivo* and *in vivo* images of the same mouse brain. Equivalent coronal (left column), sagittal (top right) and transverse (lower right) slices from a wild-type brain (from §5, age 7.5 months), scanned at CABI *ex vivo* (a: 40μm isotropic

voxels with contrast agent, 11hr 4min spoiled GE3D sequence; TE 4.54ms; TR 17ms; flip angle 51°; 6 averages; further parameters in §4.1.2) and *in vivo* (b: 150µm isotropic voxels, no contrast agent; 1hr 30min 3D fast spin echo (FSE) sequence; TE<sub>eff</sub> 43ms; TR 2500ms; further parameters in §5.2.3). High resolution and excellent contrast from 9.4T T<sub>2</sub>W scans, optimised for high GM (light)/WM (dark) contrast. Slice locations indicated in top right. Note the ventricles (containing bright CSF) have mostly collapsed, *ex vivo*, and the brain has shrunk slightly (see §5.7). Most flesh at the top of the skull was removed prior to the *ex vivo* scan, but *in vivo*, increased partial volume reduces the brain's apparent separation from external material. Cerebellar and cortical layers are visible *ex vivo*, but not *in vivo*. Fig 5.19 includes labels for the meninges. Note also the relative homogeneity *in vivo*, compared with *ex vivo*, of the striatum and midbrain. See also Fig 2.2.

Thanks to the high field strength magnet (9.4T), the liberty to perform long (overnight, 11-12 hour) scans, and optimised protocols (Cleary *et al.*, 2011b), an extraordinary level of detail is obtainable *ex vivo*, including cortical and cerebellar layers, the Purkinje cell layer, and the thin external capsule. Magnevist<sup>1</sup> was used to improve contrast and shorten scan times. External tissues have a similar intensity as the brain, here, owing to their uptake of the contrast agent.

The commonly-used *in vivo* contrast agent, manganese, also shortens scan times – thus increasing the likelihood for anaesthetised mice to survive several longitudinal scans, and enabling increased *in vivo* throughput (Benveniste & Blackband, 2002). However, manganese is believed to have neurotoxic effects, which could interfere with observations of pathology (§5; Lerch *et al.*, 2012). A contrast agent was therefore not employed for *in vivo* scans; however, acquisition sequences were optimised for high grey matter (GM) and white-matter (WM) contrast. Detail and contrast are lower *in vivo*, but structures such as the hippocampus, corpus callosum, and cerebellar folds are clearly distinguishable. Contrast-to-noise and signal-to-noise ratios (CNR, SNR) are reported in §4.1.5 (*ex vivo*) and §5.5 (*in vivo*).

---

1: Gd-DTPA, containing paramagnetic gadolinium.

To increase throughput, this *ex vivo* scan involved three subjects simultaneously (§3.2), necessitating a large field of view (FOV). To reduce the FOV and so shorten acquisition times, some studies (e.g. *Keifer et al., 2015*) omit the olfactory bulbs or cerebellum. Such cropping is acceptable when the region of interest is known, but this may compromise inter-subject registration (if the FOV is variable) and registration to atlases (which include the whole brain, §2.2.2). Additionally, a substantial benefit of investigative morphometry is its ability to identify unexpected differences (*Sawiak et al., 2009b*).

Ideally, images should have high resolution, to give better detail of fine structures. Contemporary  $\mu$ MRI of mouse brains is limited to resolutions of about  $15\mu\text{m}^3$  or greater (*Janke & Ullmann, 2015; Lerch, 2010*). As this does not reach the cellular scale, each voxel constitutes a shared signal from various sources. GM consists of neuronal cell bodies, axons, and supporting glial cells. WM is predominately composed of myelinated axonal tracts connecting GM neurons. Cerebrospinal fluid (CSF), thought to be protective and to facilitate waste clearance, fills the ventricles, aqueducts and cisternae. (The mouse skull fits very tightly around the brain; there is little external CSF visible *in vivo*.) Each voxel therefore necessarily exhibits a degree of partial volume (PV). Larger voxels have more, and therefore lower structural contrast, but smaller voxels compromise SNR (which is proportional to voxel volume). *Kale et al. (2008)* showed that an SNR of around 20 was ideal for morphometry.

In-plane resolution can be significantly increased at the expense of slice thickness, while decreasing scan times, but this has several disadvantages. A chief advantage of morphometry is the ability to construct 3D atlases and view features from any angle. Registration algorithms (§2.2.6) typically rely upon information from all dimensions. Inevitable out-of-plane transformations and warps will result in the corruption of in-plane voxels with severe PV effects (*Lerch, 2010*). Isotropic voxels are therefore preferable.

The human brain measures approximately  $140 \times 167 \times 93\text{mm}$ . Total intracranial volume (TIV) is approximately 1700ml: 1400ml brain tissue; 150ml blood; 150ml CSF (*Rengachary & Ellenbogen, 2005*). Clinical structural whole-brain MR resolution from the most common 1.5—3T scanners is typically  $1 - 1.5\text{mm}^3$ , giving around 1.7 million voxels per image. With 4 bytes per voxel, images thus occupy under 10MB.

A healthy adult mouse brain occupies approximately 0.5ml. Structures of interest, such as the hippocampus, are relatively large (compared with the overall brain volume, and the same ratio in humans). The *ex vivo* and *in vivo* brains shown above, respectively, occupy around 31MB and 0.6MB. (Given the FOV and presence of external tissues, however, more typical image file sizes are 150MB and 4MB.)

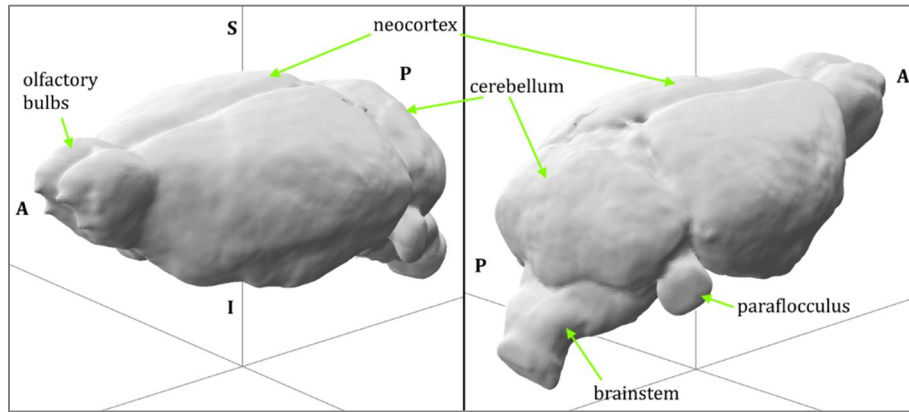
## 2.2 Morphometry

The mouse brain images described above each represent a rich source of data. Hippocampal volume may be useful for the assessment of Alzheimer's disease (§2.3), for example, but individual voxels' intensities and neighbourhoods are also valuable. The aim of morphometry is to quantify this morphological information, allowing the derivation of meaningful conclusions relating to phenotype or disease.

To compare the volume or shape of specific sub-regions, which may have functional, disease, or phenotypic significance, brains can be parcellated using manual or automated, atlas-based techniques. *Redwine et al. (2003)* and *Delatour et al. (2006)*, for example, both used manual delineation of structures to measure atrophy in AD mouse models. Manual volumetry is thought to be the most accurate in both human and preclinical studies, but is very time-consuming (taking hours per scan, even for experienced raters<sup>2</sup>, and ideally requiring more than one rater), becoming prohibitive for all except very small cohorts (*Teipel et al., 2013*). Manual measurements require strict protocols, yet are still prone to inter- and intra-rater variability (*Riegler et al., 2010*). Automated methods are repeatable, and now approach manual accuracy, so are preferred.

---

2: *Leung et al. (2010a)* noted that it took trained raters 45 minutes to manually delineate the hippocampus with good reproducibility (<5% volume difference within and between raters).



**Figure 2.2** 3D surface rendering of a mouse brain.

Two oblique views of a WT brain mask following automated extraction from the skull (§3.5), with prominent features labelled. A: anterior; P: posterior; S: superior; I: inferior.

I refer to “parcellation” here to mean the classification of distinct anatomical brain regions into probabilistic or binary spatial maps (hippocampus; cerebellum; corpus callosum). I use “segmentation” to refer to probabilistic or binary tissue classification (into GM, WM, CSF, etc.), which is generally intensity-based and not confined to individual anatomical regions – for example, both GM and WM exist in the cerebellum (Fig 2.1).

The majority of this thesis is concerned with measurements taken at every voxel. These are either compared, individually, between animals (V/TBM, §4, §5) – or combined into summary measurements for a single subject (BSI, §6). In each case, large groups may be required to detect statistically significant differences between populations, and both require image registration, to ensure anatomically equivalent regions are compared between animals: for morphometry, to deform scans from individuals to achieve voxel-level correspondence across the group; and for the BSI, to align the same animal’s serial scans between time-points.

Automated methods are essential, requiring suitable prior information (atlases with accurate labels) and unbiased registration. State-of-the-art techniques for clinical data should be adopted. In this section, I discuss recent developments for atlas-based parcellation (§2.2.1); tissue segmentation (§2.2.4); and alignment with registration (§2.2.6). These key steps are fundamental to morphometry and segmentation

propagation. However, several other pre-processing steps are required to initialise image data and ensure accuracy and low bias – including multiple subject extraction (from multi-subject scans); orientation (from arbitrary initial alignments in those scans); non-uniformity correction and intensity standardisation. These are often omitted from descriptions of preclinical morphometry (§2.2.8; [Table 2.1](#)). I discuss them, along with current methods, in more detail in §3.

### 2.2.1 Atlases

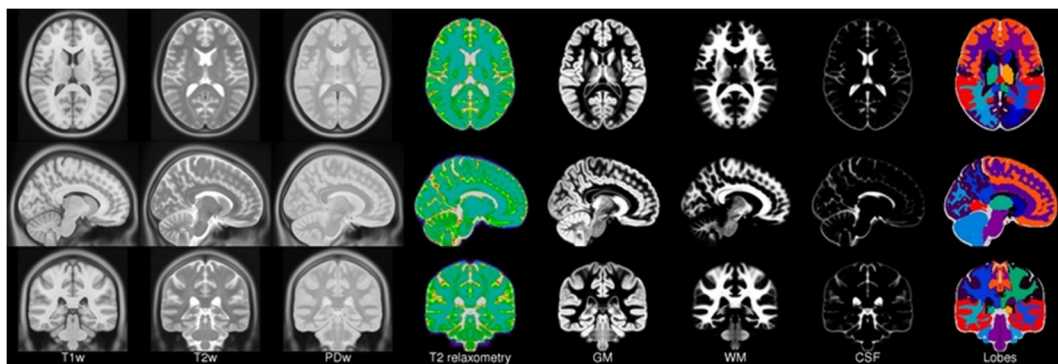
Neuroimaging atlases provide a common reference space, including a standardised orientation and set of coordinates, against which researchers may compare findings. One of the earliest atlases of the human brain was the Talairach atlas (published 1967), which consisted of histological slices from a single female brain, to assist surgery. MRI atlases emerged in the 1990s. The Colin27 atlas, an early example, consists of 27 T<sub>1</sub>W images from separate scans of the same subject, aligned and averaged to improve SNR and contrast ([Holmes et al., 1998](#)).

Advanced registration algorithms allow near-exact alignment between atlas images and new patient images. The transformations calculated to map one image to the other allow the propagation of atlas information, such as anatomical labels. Atlas databases comprising more than one image help to account for natural inter-subject morphological variation ([Aljabar et al., 2009](#)), e.g. between sexes; right and left-handedness; across age-groups, and so on. There is a greater chance a new image will match at least some of the atlas images, hence improving alignment and the ability to propagate information. For example, [Leung et al.](#) used such a “template library” for automated hippocampal ([2010a](#)) and whole-brain ([2011](#)) delineation.

Contemporary human brain atlases thus include many subject images and multi-dimensional data, such as corresponding parcellations, probabilistic tissue class maps, and image weightings, to provide as much information as possible. For example, the



MNI<sup>3</sup>152 atlas contains 152 brains and individual tissue maps aligned to a stereotaxic “MNI space”. *Fig 2.3* shows average structural images created from this database, with parcellations and tissue maps. These are included with software analysis packages such as SPM and FSL. Using smoothed or blurred structures, average or probabilistic atlases attempt to balance natural variability with enough spatial specificity to be anatomically useful when registered to new data (*Hammers et al., 2003*). Average images also have higher SNR and improved contrast over individuals (*Fig 2.9 inset*). Furthermore, domain-specific atlases exist for developing, ageing or diseased brains (*Evans et al., 2012*).



**Figure 2.3** Example human brain average atlas images.

Equivalent transverse (top row), sagittal and coronal slices through aligned atlas images, with multiple weightings and segmentations, from the “ICBM 2009a” nonlinearly registered, symmetric template (1mm<sup>3</sup> isotropic resolution). By registering data to a relevant atlas image, parcellations and tissue class maps can be applied. Columns from left: T<sub>1</sub>W structural average; T<sub>2</sub>W; proton density weighted; T<sub>2</sub> relaxometry map; probabilistic GM, WM and CSF maps; and binary anatomical parcellations. Image from [www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009](http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009).

## 2.2.2 Mouse brain atlases

For years, early editions of *Paxinos & Franklin (2012)*, containing histology slice images and derived diagrams, served as the primary mouse brain atlas. As mouse studies have

---

3: Montreal Neurological Institute; sometimes also “ICBM”, the North American International Consortium for Brain Mapping.



proliferated, digital mouse MRI atlases have been released to ease the localisation of functional and anatomical findings. Several mouse brain atlases exist and are in development, but there are fewer preclinical atlases than human. Given the extraordinary number of mouse strains (*Beck et al., 2000*), it is important to select a relevant atlas for a given study. Ventricle sizes and shapes are variable, and, just as in humans, the brain changes shape in the first few months after birth.

Single-subject atlases (for example, *MacKenzie-Graham et al., 2004*) may exhibit deviations from normal shape or suffer from image artefacts at particular locations. *Kovačević et al. (2005)* created an atlas via group registration of 9 in-bred, *ex vivo* brains. The average image was manually parcellated and its variability measured using deformation fields. *Dorr et al. (2008)* used the same technique to label 62 structures in a 40-mouse average image with isotropic,  $32\mu\text{m}^3$  voxels. This atlas has been subsequently employed many times for annotating morphometry results (e.g. *Ellegood et al., 2010*; *Lerch et al., 2011*; *Badhwar et al., 2013*).

A promising equivalent of MNI space for mouse brain atlasing is Waxholm space (*Johnson et al., 2010*), a standardised orientation atlas, with average and probabilistic parcellations, aligned to a histology correlate. The Australian Mouse Brain Mapping Consortium<sup>4</sup> has generated high resolution, detailed atlases of the C57Bl/6J cortex, cerebellum, hippocampus and other sub-regions ( $30\mu\text{m}$  isotropic voxels), from averaged *ex vivo* images (e.g. *Ullman et al., 2012*; *2013*). The group continues to release further parcellations and recommendations for atlas-building (*Janke & Ullmann, 2015*; *Ullmann et al., 2015*). Efforts are also underway to construct embryo (*Cleary et al., 2011a*; *Wong et al., 2012*; *Norris et al., 2013*) and developmental (*Aggarwal et al., 2009*; *Chuang et al., 2011*) atlases.

*Ex vivo* images, especially when skull-stripped (such as in *Kovačević et al., 2005* and *Ma et al., 2005*), may exhibit damage to the cortical surface, and so deviate from natural mouse anatomy (*Chen et al., 2005*; *Dorr et al., 2008*). *In vivo* atlases are usually lower resolution and have lower contrast. The mouse brain exhibits a high degree of partial

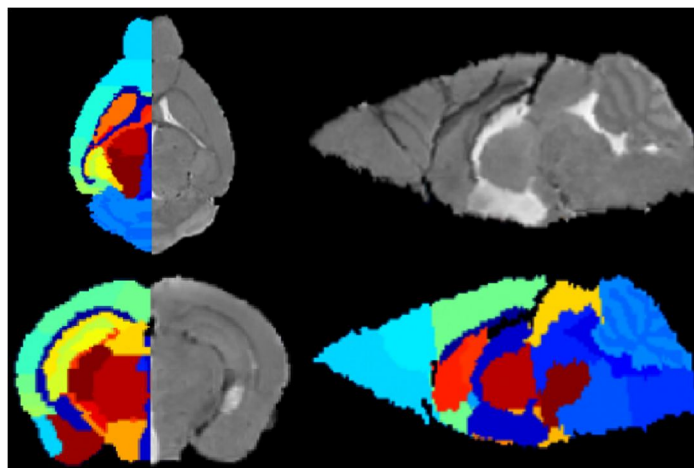
---

4: [www.imaging.org.au/AMBMC](http://www.imaging.org.au/AMBMC)

volume (PV) – a mixture of tissue classes within individual voxels. Contrast between structures is often low, especially in the midbrain: both manual and automatic parcellation methods may struggle to accurately, reproducibly delineate boundaries. [Bae et al. \(2009\)](#) and [Wu et al. \(2012\)](#) used a support vector machine with multi-spectral MR data (T<sub>2</sub>W, proton density-weighted and DTI) to parcellate brains, together with a Markov random field (MRF), and atlas-based priors for spatial consistency.

Probabilistic atlases incorporate some natural anatomical variability (e.g. [Ali et al., 2005](#)); however, the weight of each voxel is determined by its probability of appearing in the original dataset used to create the atlas; thus, both local variability and resolution may be reduced when there are a large number of subjects ([Leung et al., 2010a](#)). When segmenting, each image is initialised using the same probabilities; there is no ability to weight similar images more highly in particular regions.

Multi-subject atlas databases are preferred in the human paradigm, but have only recently been implemented preclinically. Only three multi-atlas mouse brain databases are presently available, from the National University of Singapore (NUS, [Bai et al., 2012, Fig 2.4](#)); and the University of Florida (UFL, [Ma et al., 2005, 2008](#)); only one of which is *ex vivo* (UFL, [Ma et al., 2005](#)).



**Figure 2.4** Example slices from a single brain of the NUS atlas, with parcellations. Transverse, coronal and two sagittal views, with labels overlaid. This *in vivo* atlas has large ventricles, relatively low GM/WM tissue contrast, and is skull-stripped, but it has more detailed labels than the UFL atlas – for example, the cerebellar WM is delineated. Images

in this atlas were rigidly aligned to that of *MacKenzie-Graham et al. (2004)*, and hence the Paxinos & Franklin atlas.

### 2.2.3 Multi-atlas label fusion for parcellation

To automatically parcellate a new image, after atlas registration, labels are propagated to the image space using, for example, nearest neighbour or trilinear interpolation. Registration accuracy thus plays an important role in the quality of automatic parcellation. Large atlas databases can aid the process by providing images with variable morphology (encompassing natural variation between individuals and across age groups, or in disease). Registration is more likely to succeed in at least some cases, even to new images with outlier morphology, such as hydrocephalus (grossly enlarged ventricles); severe atrophy, or cranial damage. Each atlas image is registered to the new image, the labels propagated, and an appropriate label chosen for a given region. Multi-atlas techniques thereby reduce bias and increase accuracy over single-atlas and probabilistic approaches (*Barnes et al., 2008; Leung et al., 2010a*).

One method for deciding upon a consensus label is voxel-wise majority voting. Label decision fusion techniques attempt to be less naïve. A subset of templates from the atlas database can be chosen based upon metadata (such as age or sex), or the registered atlas images can be ranked by global or local similarity with the data, a subset of top-ranked atlases chosen, and their corresponding labels' contribution to the voting weighted by this similarity (*Wu et al., 2007; Aljabar et al., 2009; Iglesias et al., 2013*).

To segment hippocampi and whole brains, *Leung et al. (2010a), (2011)* used a method developed by *Warfield et al. (2004)*, STAPLE<sup>5</sup>, which provides a probabilistic estimate of the true underlying parcellation given a set of registered templates and labels, finding good performance (using the Jaccard index) compared with manual labelling and the sensitive distinguishing of diseased and healthy brains. *Cardoso et al. (2013b)* proposed an extension, Similarity and Truth Estimation for Propagated Segmentations (STEPS),

---

5: Simultaneous Truth and Performance Level Estimation.

which ranks classifiers (combinations of the registration and atlas label) by local similarity to the data, using locally normalised cross-correlation. This improves robustness to local deficiencies either in atlas quality, or registration accuracy, and the authors showed improved performance, using a Dice score, while requiring a smaller atlas database. In a recent technique, [Wang et al. \(2013\)](#) showed improved parcellations could be achieved when atlases' contributions were negatively weighted based upon their local similarity to one another – such that atlases with similar errors contributed less.

[Chakravarty et al., 2013](#) showed the automatic creation of atlas databases for both human and mouse brains, from a single starting atlas (for mice, that of [Dorr et al., 2008](#)), enabling multi-atlas label fusion techniques and improvement over single atlases. However, this is likely to have similar drawbacks as single-atlas label propagation, as any errors in the original atlas, and arising from multiple registrations, will bias the final result.

Perhaps due to the scarcity of multi-subject mouse brain atlas databases, and the availability of higher-resolution single-subject and probabilistic atlases, very few studies have employed multi-atlas label fusion techniques to segment data. [Bai et al. \(2012\)](#) found that multi-atlas techniques, including STAPLE, outperformed single-atlases in terms of a volume overlap percentage, compared with manual labels. Both [Ma et al. \(2014\)](#) and [Badea et al. \(2012\)](#) also reported improved accuracy using multi-atlas techniques. [Nie et al. \(2013\)](#) used region of interest (ROI, or parcellation) boundaries in the atlas (UFL, [Ma et al., 2005, 2008](#)) to inform the registration step (reasoning that these are the most important regions to correctly register, for parcellations). The resulting surfaces were then updated using a support vector machine.

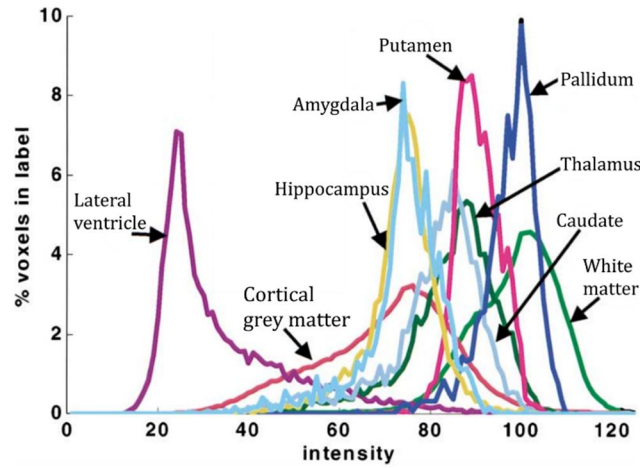
Many mouse studies use manual parcellation ([Table 2.1](#)) or single-atlas label propagation to delineate structures ([Chen et al., 2005](#); [Scheenstra et al., 2009](#); [Badhwar et al., 2013](#)). [Badea et al. \(2007a\)](#) used parcellations' surface deformations to measure shape differences in C57Bl/6J mice. Whole-brain parcellation (skull-stripping, [§3.5](#)) is an important preprocessing step for excluding extraneous material prior to tissue segmentation or morphometry. Parcellation of the cortex and adjacent structures enables measurement of cortical thickness and, with sufficient contrast, sublayer thickness (e.g. [Lerch et al., 2008](#); [Sawiak et al., 2012](#); [Hébert et al., 2013](#)), which is applicable to neurological conditions

including AD. More recently, this has been extended to the cerebellar cortex, a significantly more complex structure with tighter laminar folds (*Ma et al., 2015*).

The mouse brain has large regions of homogeneous and gradually-varying intensity (such as the brainstem, midbrain, colliculus, thalamus and striatum). Even at high resolutions below 50 $\mu$ m, there is a large degree of PV. In these low-contrast areas, parcellation techniques may struggle to accurately delineate anatomical sub-regions, resulting in imprecise or low-confidence parcellations (*Fischl et al., 2002*; *Fig 2.5*).

## 2.2.4 Tissue segmentation

Brain tissue segmentation aims to classify image voxels by their predominant, underlying type (GM; WM; CSF), rather than anatomical location, enabling analysis based upon physical – hence, functional – properties, and with less dependence upon parcellation, or registration, accuracy. Tissue volumes can be highly sensitive to diseases such as dementia (*Teipel et al., 2004*; *Serra et al., 2010*). Segmentation into binary or probabilistic labels proceeds via contrasting signals: physically similar tissues exhibit comparable intensities in MRI. Manual tracing of regions has the same drawbacks as anatomical parcellation, and is especially subjective for PV voxels. Automated segmentation can rest simply upon thresholding (*Wright et al., 1995*), but this requires assumptions about the arbitrary MR intensity scale. Artefacts, such as noise; inhomogeneity; magnetic susceptibility; low contrast; or unexpected pathological features can corrupt these assumptions. Intensity distributions of each tissue type can vary between scans, patients and scanners. To add robustness, the majority of segmentation techniques require some prior information, such as the number of tissue classes, and probabilistic maps of their location, provided in atlases (*Fig 2.3*) and software packages such as SPM, FSL, and Freesurfer (*Klauschen et al., 2009*).



**Figure 2.5** Example T<sub>1</sub>W MRI human brain intensity histogram.

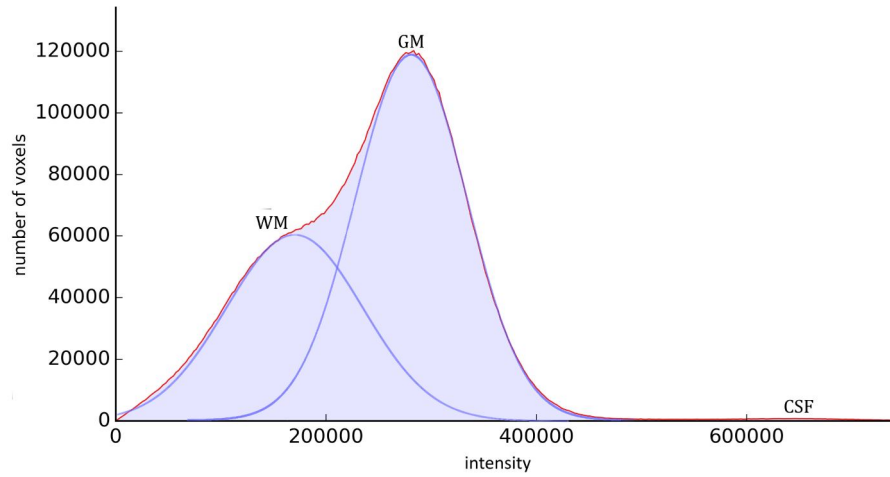
Structures labelled to illustrate their overlapping intensities, and the broad range of intensities in some structures. Adapted from *Fischl et al. (2002)*.

In this work, an expectation maximisation (EM) algorithm, implemented in the open-source *NiftySeg* software<sup>6</sup> is used to fit a probabilistic model to the image data intensity histogram. It is assumed that the log-transformed histogram may be represented as a mixture of Gaussian distributions with different means and standard deviations; a Gaussian Mixture Model (GMM). Each Gaussian is parameterised by its mean, standard deviation, and the number of voxels. The number of classes is given *a priori*, as are initial estimates of the mean intensity and variance of each tissue. For segmentation of brain tissues, these are obtained from a tissue probability map (TPM) aligned with structural images, which can be registered to new data (*Fig 2.3*).

Image voxels are classified probabilistically based upon these initial parameter estimates (expectation step), and new parameters estimated based on this classification (maximisation). This proceeds iteratively, and advantageously, the parameter estimation continuously improves (*Do & Batzoglou, 2008*). The EM method has been adapted into several segmentation frameworks for human brains.

---

6: Available from <http://sourceforge.net/projects/niftyseg>



**Figure 2.6** Example *ex vivo* mouse brain  $T_2W$  image intensity histogram.

Approximate grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) peaks noted, assuming a Gaussian mixture (blue lines). Most CSF is expelled during fixation. The MR intensity scale (x-axis) is arbitrary. The GM and WM peaks are not well-separated: there is a high degree of partial volume (PV), particularly in the midbrain.

*Van Leemput et al. (1999a)* included iterative correction for smoothly-varying image intensity non-uniformity. The bias field is assumed to be multiplicative in MRI, but after log-transformation of the intensities it is additive, and can be modelled using a polynomial function, whose parameters are estimated after the maximisation step based upon tissues with a narrow intensity distribution. *Van Leemput et al. (1999b)* and *Zhang et al. (2001)* added a Markov random field (MRF) weighting to increase spatial consistency. Voxels' classifications are assumed to depend statistically upon the content of their immediate neighbours, avoiding the misclassification of noise and voxels surrounded by other tissue types. In SPM, segmentation, non-uniformity correction and registration with TPMs are unified in a single iterative framework, allowing for PV effects by assuming each tissue can be represented by more than one Gaussian, and constraining classifications based upon the spatial priors (*Ashburner & Friston, 2005*). TPMs supplement the image intensity information and provide a spatial context for the expected classification of each voxel.

TPMs are used to initialise segmentations. They are smoothed or blurred to account for population variance, and spatially correspond to template images from an atlas. Within

TPMs, voxels are probabilistically assigned one of a number of classes: GM, WM, CSF and ‘non-brain’, or ‘background’ (BG). Human studies utilise extensive public TPM databases such as MNI305 and ICBM152 (*Evans et al., 2012; Fig 2.3*), which represent a range of naturally-occurring morphologies, balancing specificity for the group under study with natural variability.

TPMs are typically based upon healthy brains. Natural variability (for example, of human cortical folds) and pathology (such as severely enlarged ventricles or tissue damage), or even healthy ageing, may complicate the use of priors, even when warped to the data using non-rigid registration: segmentation results may be biased towards healthy, young morphology. If the priors are smoothed to incorporate more variability, they become less useful. *Cardoso et al. (2011)* therefore introduced a prior probability relaxation factor. After EM has converged to voxels’ initial classification, the TPMs are multiplied by this factor dependent on whether the classified voxels neighbour a different class, thereby generating a new ad hoc, image-specific prior for subsequent iterations of EM. This was shown to improve segmentation accuracy in an Alzheimer’s disease dataset.

### 2.2.5 Mouse brain segmentation and tissue probability maps

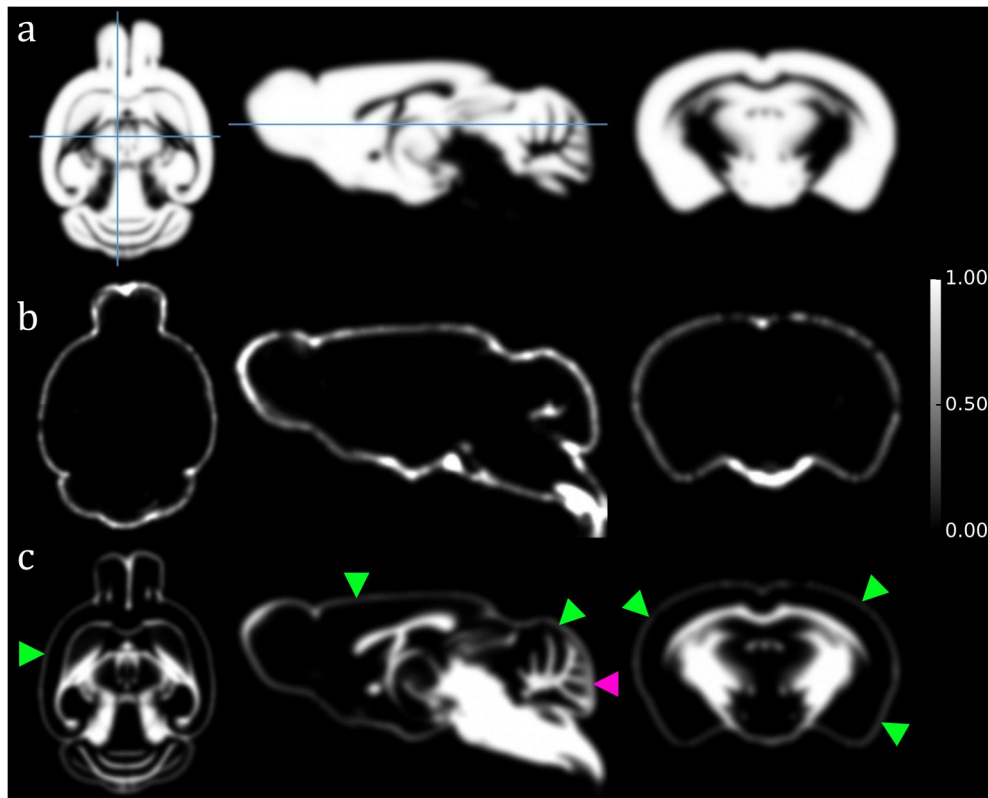
Different approaches have been adopted for segmenting mouse brain data. To avoid an atlas, *Wang et al. (2010)* used a threshold-based method to separate the skull and brain, followed by k- and fuzzy c-means clustering of brain tissues based upon intensity. This required manual initialisation and would be sensitive to signal inhomogeneity and artefacts. Noting the sensitivity of EM to converge to incorrect maxima with inappropriate atlases, *Tohka et al. (2007)* combined EM with a genetic algorithm to segment human and mouse brain data without priors. In mice, the technique showed improved results over naïve, prior-free EM; however, it required setting a lower bound on the amount of GM in different structures (60%, overall). The mixture model had classes: {cerebellar GM; cortical GM; WM; CSF} and three PV classes. Noting the considerable degree of intensity overlap in mouse brains, the authors stated, “segmentation between [GM] structures does not appear possible without applying spatial information”.



Owing to the variable and reduced MR contrast of the still-myelinating newborn human brain, *Prastawa et al. (2005)* employed atlas-based priors to initialise segmentations, noting that histogram-based techniques could not reliably differentiate tissues. Because of the increased PV content, a similar issue faces mouse brain segmentation. Atlas-based techniques will likely be advantageous.

Several groups have performed VBM in rodents, although the technique is more complex than TBM as it requires prior knowledge of each voxel's fractional tissue content. A substantial source of rodent brain TPMs – a database equivalent to the MNI152 human atlas – does not yet exist. Though TPMs based on non-representative atlases may misclassify voxels (*Cardoso et al., 2011*), it is unrealistic to create atlases for each of the vast number of mouse strains available. Additional difficulties of tissue classification lent by the smaller structures and greater partial volume (PV) proportion in mouse brains, especially at resolutions above 150 $\mu$ m (*Natt et al., 2002*) has meant a slow uptake of preclinical VBM.

*Li et al. (2009)* classified tissues based solely upon intensity, using FSL FAST. *Biedermann et al. (2012)* initialised SPM segmentations by assigning equal probabilities to every brain voxel, and improved classifications via iteration. In both approaches, the observed data is fitted without anatomical prior information, which can result in unreliable segmentations due to morphological variability, PV, intensity inhomogeneity, image artefacts and natural intensity variation within tissues (*Cardoso et al., 2011*). Other mouse TPM creation strategies have included manually thresholding DTI mean diffusivity images to create probabilistic priors (*Oguz et al., 2011*), and segmenting and smoothing the GWR average of 87 brains using SPM (*Sawiak et al., 2009c*). To my knowledge, these comprise the only publicly available mouse TPMs. They are included with SPMMouse, a plugin for SPM version 5 (*Sawiak et al., 2009a; Fig 2.7*). However, there is no anatomical reference to which external data may be registered; alignment – necessary for priors to initialise segmentations – relies entirely upon the TPMs. This may involve some manual adjustment, which for large datasets would be prohibitive to high-throughput analysis. Nevertheless, *Sawiak et al. (2012)* manually aligned 399 mouse brain images to these TPMs. *Kielar et al. (2012)* did the same for 30 brains.

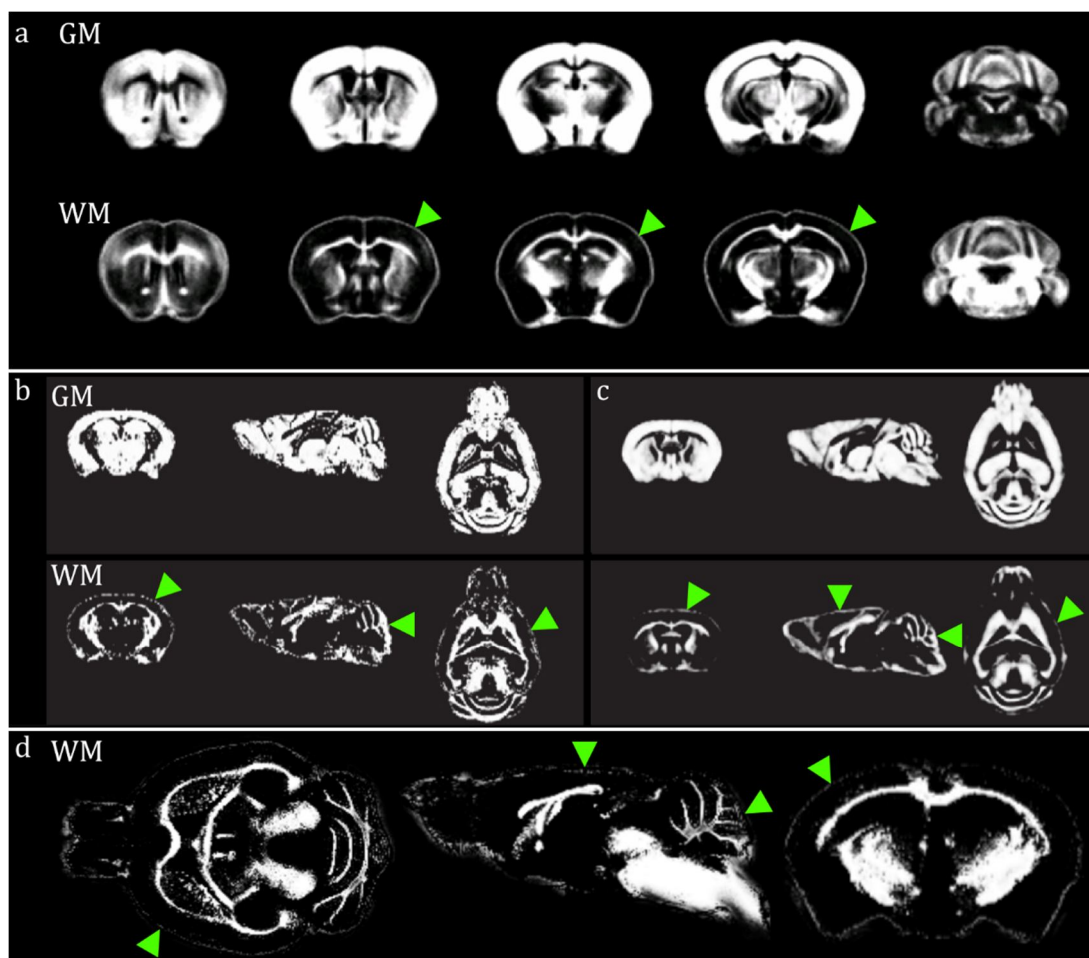


**Figure 2.7** SPMouse tissue prior maps

Sample transverse, sagittal and coronal slices (positions indicated by blue lines) through GM (row a), CSF (b) and WM (c) maps provided in the SPMouse package ([Sawiak et al., 2009a](#)). These are the only publicly-available mouse brain TPMs of which I am aware. The scale represents average tissue class proportion. The WM map exhibits a ‘rim’ of WM around the cortex (green arrows) and caudally at the cerebellum (pink). Although only initialisations for EM in SPM, these misclassifications may cause errors in subsequent segmentations, and mismeasurement of tissue volumes (see [Fig 2.8](#)).

TPMs published so far appear to possess several defects, which could propagate to any dataset which uses them for *a priori* information. Those of [Sawiak et al. \(2009a\)](#), [Li et al. \(2009\)](#), and [Biedermann et al. \(2012\)](#) exhibit an artificial WM ‘rim’ around the brain, where GM/BG PV voxels with lower signal intensity are apparently misclassified. Signal drop-off at the cerebellum causes a similar artefact ([Fig 2.8](#)). This could cause overestimation of BV in probabilistic segmentations (by classifying PV voxels as WM with more confidence than deserved); could misattribute volume changes at the brain surface to the wrong tissue type; or could cause over- or under-estimation of tissue loss

to atrophy. *Good et al. (2001)* noted similar artefacts in GM segmentations of human brains. This was the motivation for their introduction of an “optimised” VBM protocol. In *Sawiak et al. (2012)*, the SPMouse priors were used to initialise the creation of a set of study-specific TPMs. The segmentation was repeated using the averaged results of this initial segmentation as new priors. The results appear to inherit this problematic rim.



**Figure 2.8** Examples of published mouse brain tissue classification results.

The green arrows, in each case, indicate erroneously classified WM at the cortical surface or cerebellum. (a-c): GM and WM slices; only WM is shown for the last example. Results from an individual brain are shown in (b); (c) is the combined result after group registration, averaging, and smoothing. Adapted from, (a): *Li et al. (2009)*; (b,c): *Biedermann et al. (2012)*; (d): *Sawiak et al. (2012)*.

## 2.2.6 Image registration

Registration algorithms seek to align images with corresponding features, such as anatomy, by performing global transformations and local warps. Registration can be used to fuse images from different modalities: for example, PET and MRI (*Rowland et al., 2005*), which provide complementary functional and structural information. To monitor changes in a single patient, it can be useful to align baseline and follow-up MR images minutes (such as for MR mammography, *Rueckert et al., 1999*) or years apart (detecting subtle atrophy rates in the brain, *Freeborough & Fox, 1997*), so that anatomy can be easily compared. Surgery may require “real-time” registration, for motion correction.

Registration provides aligned images (interpolated into the same space) and the deformations necessary to repeat the alignment. This is vital for segmentation propagation, the transfer of labels from one image (such as an atlas) to another. Furthermore, by aligning groups of images, statistical tests can be performed between anatomically equivalent regions, either comparing those deformations, or the images themselves (§2.2.7).

Medical image registration algorithms face several difficulties. First, there may not be a true, biologically meaningful match between two different images either of the same, or different, subjects: in pathology, structures may have atrophied, or be damaged or absent. Second, there can be more than one possible “good” solution (a global maximum similarity). Third, noise, image non-uniformity and other artefacts may corrupt the available information.

Registration requires: a model describing the transformations between images; a measure of similarity or error; a way to interpolate images; and a method for adjusting the transformation to maximise similarity or minimise error (an optimiser, *Crum et al., 2003*). These are introduced briefly below.

Algorithms may rely upon the alignment of landmarks or surfaces, principal axes, or the centre of mass, assuming that their alignment imputes the alignment of other structures (*Modersitzki, 2004*). However, manual landmark identification is time-consuming for

large datasets and subject to human error (*Crum et al., 2003*), and automatic feature extraction requires an additional step. Instead, voxel intensities can be used to measure similarity. Within-modality, a direct relationship between images' intensities may be found. A more robust technique, normalised mutual information (NMI, *Studholme et al., 1999*), assumes a statistical relationship between intensities, and allows inter-modality registration. NMI involves the minimisation of information (joint entropy) in a joint histogram, on which two well-aligned images should produce little dispersion.

A transform is computed between source images  $S_i$  and a target,  $T$ , which specifies a mapping of voxels in  $T$  to those of  $S_i$ . The transformation model might be rigid or affine (for global adjustments), or non-rigid: elastic, fluid or basis spline (B-spline, for local warps). Transformations are usually updated iteratively, with small changes applied, the similarity tested, and the optimiser deciding whether an ideal match has been found (halt iterations, if a threshold is reached, or the measure has converged on a maximum) or to continue improving the accuracy.

3D rigid-body registration involves 6 parameters, or degrees of freedom (DOF): independent translations along, and rotations about, the  $(x, y, z)$  axes. Affine transformations include these, as well as scaling and shears along the  $(x, y, z)$  axes – thus, 12 DOF.

Non-rigid registration (NRR) is performed after global alignment, and involves many more degrees of freedom to compute a deformation field mapping voxels in  $T$  to those in  $S_i$ : either every voxel in the image can move independently (non-parametric), or a subset of voxels is used. A successful, popular implementation is free-form deformation (FFD, *Rueckert et al., 1999*), which parameterises deformations using a mesh of regularly-spaced control points (CP), several voxels apart. This mesh is deformed, moving the CPs (to improve the similarity metric) and calculating intervening voxels' displacements, using a cubic B-spline. As one CP moves, the four nearest CPs' positions (in each dimension) are updated. In 3D, this involves  $4\delta_x \times 4\delta_y \times 4\delta_z$  voxels, where  $\delta$  is the CP spacing. Reducing  $\delta$  can improve alignment, but drastically increase the required number of computations.

Regularisation constraints can be placed upon the warps to ensure plausible deformations and preserve topology. For example, “folding” of the mesh will result in a deletion of parts of the image, so should be prevented. This can be achieved using a penalty term on the Jacobian determinant,  $J_{det}$  (calculated from partial first derivatives of the deformation field – see below – and corresponding to expansion or contraction, which should never be zero or negative). Also, in some applications of intra-subject registration (for example, to correct for inhalation and exhalation), between time-point images  $TP_1$ ,  $TP_2$ ,  $TP_3$ , composing the transformations  $TP_1 \rightarrow TP_2$  and  $TP_2 \rightarrow TP_3$  should equal  $TP_1 \rightarrow TP_3$ , assuming no tissue loss (such as atrophy).

A regulariser in FFD is bending energy, the second derivative of the deformation field (the rate expansion/contraction varies, spatially). Constraining bending energy results in smoothed, more evenly-distributed expansion or contraction, which allows whole structures to deform evenly, rather than homogeneous regions exhibiting no change, but high-contrast boundary regions exhibiting extreme deformations.

For this study, I used the open-source *NiftyReg*<sup>7</sup> software, which implements rigid, affine and non-rigid registration. Many other tools are available, implementing different methods, including ANIMAL<sup>8</sup>, ANTs<sup>9</sup>, FLIRT/FNIRT<sup>10</sup>, and SPM’s DARTEL<sup>11</sup>.

The global transformations are computed using a block-matching technique (*Ourselin et al., 2001*), in which sub-regions of  $T$  are independently, affinely transformed to match equivalent regions, given a search space, in  $S_i$ . A subset of these blocks, with the best similarity performance, is used to compute the overall transformation – the method is thus robust to outliers.

As preclinical MR images may contain millions of voxels, high-throughput and large-cohort studies may involve millions of control points. FFD is thus computationally expensive. In *NiftyReg*, NRR can be computed using an efficient version of FFD, fast

---

7: Available from <http://sourceforge.net/projects/niftyreg>.

8: Automated Nonlinear Image Matching and Anatomical Labelling

9: Advanced Normalization Tools, <http://stnava.github.io/ANTs>

10: FMRIB (FSL)’s Linear/Nonlinear Image Registration Tool

11: Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra, (*Ashburner, 2007*).

free-form deformation (FFD, *Modat et al., 2010*), which was optimised for multi-core graphics processing units (GPUs). Several mouse papers have employed FFD (e.g. *Maheswaran et al., 2009a; Badea et al., 2010; Cleary et al., 2011a; Lebenberg et al., 2011*). *Lee et al. (2009)* compared FFD with a fluid registration-based technique for mouse brain parcellation. The methods produced comparable performance and were both suitable for mouse brains, although the latter required prior intensity standardisation between images. *Bai et al. (2012)* also compared registration methods for parcellation accuracy. Non-rigid registrations performed best and produced comparable results. FFD was outperformed by a technique known as LDDMM<sup>12</sup>, although an asymmetric implementation of FFD was used.

In both the block-matching and FFD approaches, a coarse-to-fine scale “pyramidal” scheme is used, to increase speed and reduce the chance of local minima. For block-matching, the registration is first performed using downsampled images. The resulting transformation is used to initialise a second registration, with a lower degree of downsampling – and so on, until the native image resolution is reached. For FFD, initial CP spacing can be doubled a number of times, providing coarse initial alignment. At successive levels,  $\delta$  is then halved, until the desired spacing is reached.

Following registration, each  $S_i$  is resampled into the space of  $T$ , which may have a different number of – or slightly offset – voxels. The interpolation scheme provides a method for treating discrete image data (voxels) as continuous, by computing new intensities in the target space. The simplest method is nearest neighbour (used, for example, for binary masks and labels); other methods include (tri-)linear and splines, which result in sub-voxel smoothing and hence a loss of information. Interpolation thus introduces a directional bias, by lowering the quality of the resampled  $S_i$ , but not the (static)  $T$ .

As discussed throughout this thesis, clinical studies require sensitivity to both disease progression and treatment effect: this lowers costs, and increases the efficiency and quality

---

12: large deformation diffeomorphic metric mapping.

of therapeutic research. Through its employment in morphometry and other techniques, registration plays an important role in such studies.

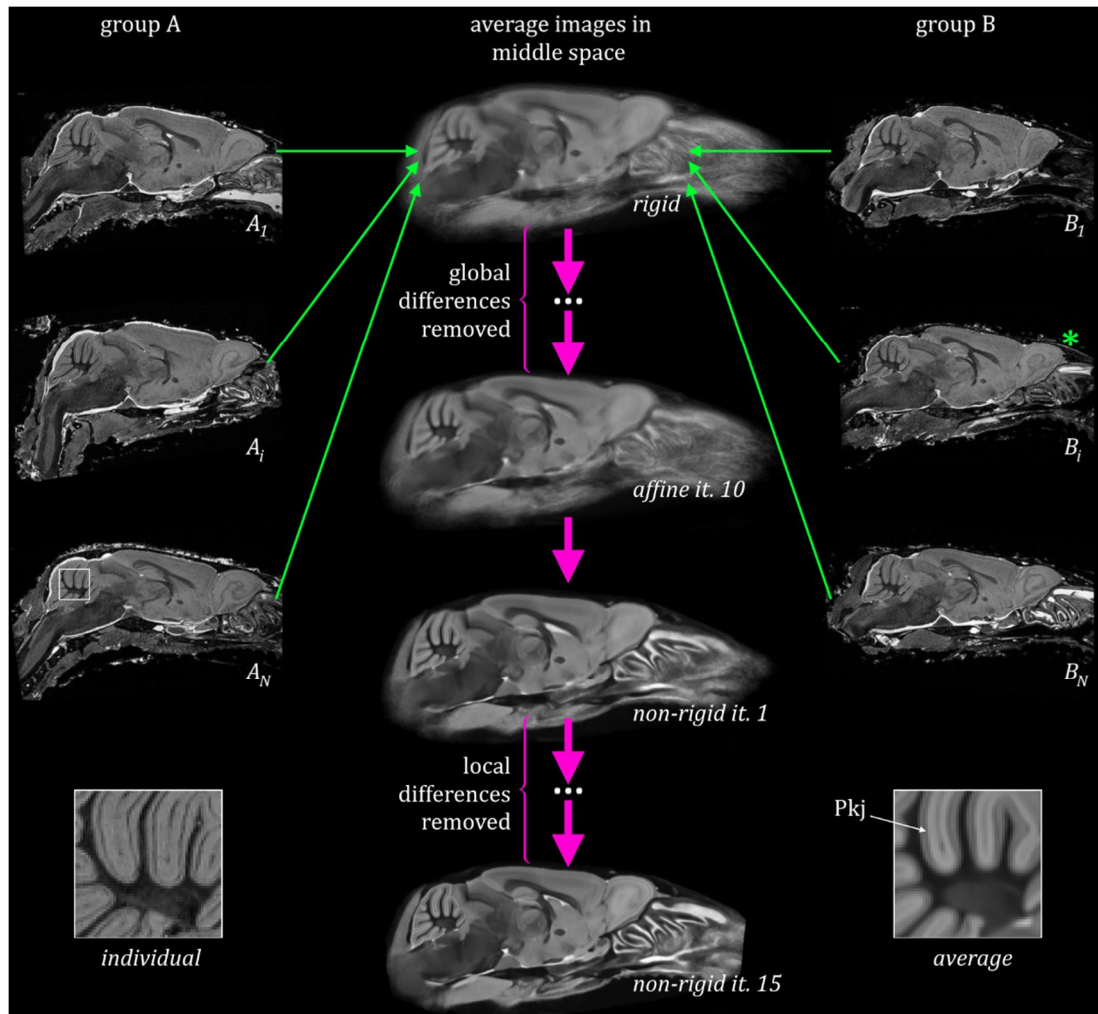
*Yushkevich et al. (2010)*, for example, found that directional bias (the asymmetric interpolation of only one image in a registered pair) from both global and local registrations induced overestimations of hippocampal atrophy in a morphometric study of Alzheimer’s disease. *Thompson & Holland (2011)* showed that such biases in TBM could severely influence power calculations and sample size estimates for clinical trials. *Leung et al. (2012)* showed that symmetric registration provided improved estimates of atrophy over asymmetric, using the Boundary Shift Integral (§6). Recent developments in registration algorithms have therefore focussed on reducing biases.

Both the global block-matching and local NRR methods in *NiftyReg* have been updated to provide symmetry and inverse consistency (*Modat et al., 2014; 2012*), such that the transformation from registering image  $A \rightarrow B$  equals the inverse of that from  $B \rightarrow A$ . Although there are several other implementations of symmetric NRR (e.g. *Avants et al., 2008* – part of ANTs; *Tao et al., 2009*), there are very few inverse-consistent global registration techniques (*Modat et al., 2014* identified only one other). These robust methods are likely even more important for *in vivo* mouse brain MRI, as the larger voxels, greater PV and lower contrast mean there is less image information available to drive accurate registration.

### **Group-wise registration**

Group-wise registration (GWR) aims to align all images within a common coordinate space, resulting in an average “atlas” image representing the mean group morphology, and deformation fields mapping each voxel to points in the original image space. Parcellations of the average can thus be easily propagated to each original image, and group statistical comparisons performed at each voxel within the common space. The multi-level, iterative scheme (*Fig 2.9*) was described by *Rohlfing et al. (2004)* for bee brains, and then *Kovačević et al. (2005)* for mice – and adopted by many subsequent papers.





**Figure 2.9** Iterative group-wise registration illustrated using sagittal images from in-skull mouse brain MRI.

All images from each group are initially rigidly registered to a randomly-chosen target (\*). In the target space, the intensity average is found. For the second global iteration, 12 DOF affine registration is used to match each image to that average (thin green arrows), and a second average is computed. This continues for several iterations (pink arrows), after which NRR iterations begin, initially using the final affine average image as a target. The average images consist of the mean sample shape from all groups. See [Figs 3.18, 3.20](#) for more. The lower inset images show detail zooms of the same cerebellar region in  $A_N$  (indicated) and the NRR iteration (it.) 15 average. The average image has been smoothed from interpolation, but has lower noise than the individual. Fine detail, such as the Purkinje cell layer (dark, indicated Pkj), a single cell thick, is still clearly visible.

Registration is performed between a target,  $T$ , and the source images,  $S$ .  $T$  is initially randomly selected from the images under study, eliminating possible bias toward external features. During iteration  $n$ , each image  $S_i$  is registered to  $T$ , and the intensity average  $A_n$  of all transformed, resampled  $S_i$  in the space of  $T$  found. In the subsequent iteration  $n + 1$ ,  $A_n$  replaces  $T$ . For the first iteration, rigid registration is used (6 DOF only), and  $A_1$  is thus quite blurred and unbiased by features of  $T$  besides global position and rotation. For  $n = 2$  onward, affine (12 DOF) registration is used. As  $n$  increases, the alignment improves, and  $A_n$  sharpens. After several iterations of affine registration, non-rigid registration is employed (initialised by the final affine transformation), and  $A_n$  sharpens further and approaches the true average global morphology ([Figs 2.9, 3.18](#)).

The stopping point can be decided subjectively via observation of  $A_n$ , or using a measure (such as intensity standard deviation at each voxel, [Fig 3.20](#), or label overlap, as [Lau et al., 2008](#)) which plateaus or converges after a certain number of iterations. Only one interpolation is performed at each level, as the global and local transformations are composed together before final calculation of voxel positions ([Yushkevich et al., 2010](#)).

For mice, [Nieman et al. \(2006\)](#) adopted a different technique, registering mutant mouse brains to the GWR average of the control group, which is unlikely to provide the optimal deformations or an unbiased average morphology. [Biedermann et al. \(2012\)](#) used SPM to register brains, scaling images  $\times 10$  for human parameters. SPMMouse includes adjusted parameters and a mouse TPM for use as the initial target to which the group is registered. However, targets should have a similar background to the subjects studied; any features not reflected in the cohort may introduce bias toward external morphology. The large number of mouse strains thus constitutes a significant impediment to the construction of a universally applicable target.

### 2.2.7 Voxel- and Tensor-Based Morphometry

Following GWR into the space of an average image, the deformation fields are used in TBM to measure local volume differences between groups. In VBM, segmentations are derived from the intensity information embedded in voxels, and used to determine

changes in local tissue volume or concentration between groups or over time. Segmented tissues are aligned into the GWR average space, and compared at every voxel.

These techniques are non-invasive and hypothesis-free, covering the entire brain and thereby reducing the need for the laborious delineation of regions of interest. In an early VBM paper, [Wright et al. \(1995\)](#) registered GM and WM maps from 15 patients with schizophrenia. The authors found that GM density (in the temporal lobe) and WM density (in the corpus callosum) were correlated with patients' scores in tests of schizophrenia syndromes. In a well-cited work, [Maguire et al. \(2000\)](#) sensitively showed specific hippocampal volume increases in London taxi drivers, compared with controls, suggesting a structural correlate with spatial navigation skills. These studies highlighted the sensitivity of morphometry to group differences due to both disease and plastic learning and memory-related brain changes, and presaged the techniques' widespread utilisation in the clinical literature.

In animal studies, they ease the burden of exploratory histology, which is both destructive and time-consuming. Statistical tests of other values are possible in the GWR average space. For example, [Teipel et al. \(2011\)](#) measured brain  $T_2$  relaxation times on a voxel-wise basis in an AD mouse model, compared with controls. [Lebenberg et al. \(2011\)](#) used a GWR approach to register autoradiographic data of glucose uptake and metabolism in mouse brain tissue slices reconstructed into a 3D volume.

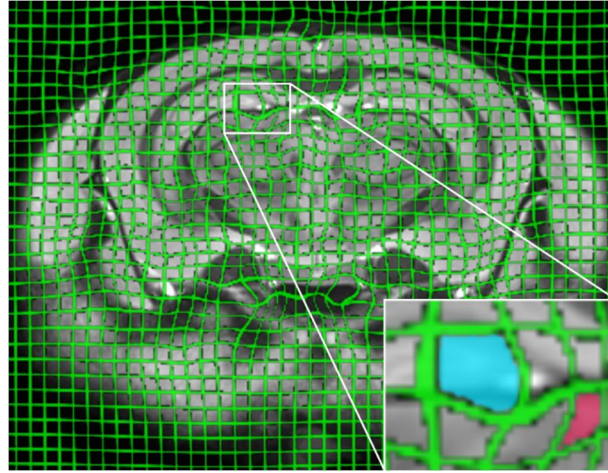
### ***Tensor-Based Morphometry***

In TBM, for each source image, the 3D displacement transformation  $\mathbf{T}(\vec{x}) = \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$  maps a voxel at position  $\vec{x} = (x, y, z)$  in the final GWR average (reference) space to its corresponding position in the original image. For 3D deformation fields, the Jacobian matrix  $J$  at each voxel is a  $3 \times 3$  tensor of displacement gradients ([Chung et al., 2001](#)), given by the first-order partial differentials of each component of the transformation. Its determinant is:

E2.1

$$J_{det} = \begin{vmatrix} \frac{\partial T_x(\vec{x})}{\partial x} & \frac{\partial T_x(\vec{x})}{\partial y} & \frac{\partial T_x(\vec{x})}{\partial z} \\ \frac{\partial T_y(\vec{x})}{\partial x} & \frac{\partial T_y(\vec{x})}{\partial y} & \frac{\partial T_y(\vec{x})}{\partial z} \\ \frac{\partial T_z(\vec{x})}{\partial x} & \frac{\partial T_z(\vec{x})}{\partial y} & \frac{\partial T_z(\vec{x})}{\partial z} \end{vmatrix}$$

Where  $T_x(\vec{x})$  is the first component of the deformation field which transforms voxel position  $\vec{x}$  in the reference image to the equivalent position in the floating image.  $J$  contains information about the elongation and contraction, in each direction (Ashburner & Friston, 2003). The determinant  $J_{det}$  is a scalar value which represents the relative expansion (when  $>1$ ) or contraction (when  $<1$ ) of that unit voxel to encompass an equivalent region in the original image (Fig 2.10). If  $J_{det} = 1$ , there is no volume change, and if  $J_{det} < 0$ , as previously stated, there is “folding” of the voxel grid (Ashburner & Friston, 2000).



**Figure 2.10** Local expansion and contraction in a deformation field.

A regular grid overlaid on an average mouse brain image (after several iterations of GWR) and warped following a deformation field which maps the average to an individual image. The inset illustrates some regions expanding (blue;  $J_{det} > 1$ ), while others contract (red;  $J_{det} < 1$ ) from their average volume to encompass the same anatomy in the original.

The *NiftyReg* implementation of GWR includes the final affine (global) and NRR (local) iterations' transformations in  $J$ . If the affine were omitted, the  $J_{det}$  value can simply be multiplied by the determinant of the affine matrix, to give the overall scaling factor.

In deformation-based morphometry (DBM), the vector lengths themselves (or sometimes the directional components of  $J$ , giving shape change) are compared, between groups. In TBM,  $J_{det}$  values from each group are compared, at every voxel, often using a t-test. An assumption of t-tests is that values are normally distributed, but because  $J_{det}$  is always positive, the distribution is skewed, and assumed lognormal.  $J_{det}$  values are therefore usually log-transformed, prior to statistical tests, to render them more normally distributed (*Chung, 2012*). This renders, via symmetry about zero, expansions and contractions of equal magnitude to be equally likely (*Leow et al., 2007*).

Smoothing (with a Gaussian kernel with given full width at half maximum, FWHM) also aides this assumption of normality<sup>13</sup>, compensates for imprecise registration, and reduces the influence of noise. The FWHM is chosen to correspond to, and hence enhance, the expected spatial scale of structural differences between groups: smaller structures' volume changes are likely to be smoothed out<sup>14</sup> (*Ashburner & Friston, 2001*). In the human brain, this is generally of the order 8—12mm (e.g. *Good et al., 2001*; *Carducci et al., 2013*).

The result of performing these mass-univariate, voxel-wise t-tests is a statistical parametric map (SPM), conferring (after multiple testing correction) significance or non-significance of the mean local volume difference between groups. These maps are useful for exploratory or naïve morphometry, without prior assumptions about the location of volume differences. Thus, for example, regional growth or atrophy can be estimated between time-points (*Brambati et al., 2007*; *Maheswaran et al., 2009a*), or local volumes compared between transgenic and wild-type groups (*Ellegood et al., 2010*). Alternatively, the  $J_{det}$  values themselves can be integrated over parcellated regions of interest in the

---

13: Via the Central Limit Theorem: that is, with enough observations with a well-defined (unique) variance and expected value, the mean of observations tends towards a normal distribution.

14: The Matched Filter Theorem.

final average image, giving structural volumes for each subject (Jacobian integration, *Boyes et al., 2006*).

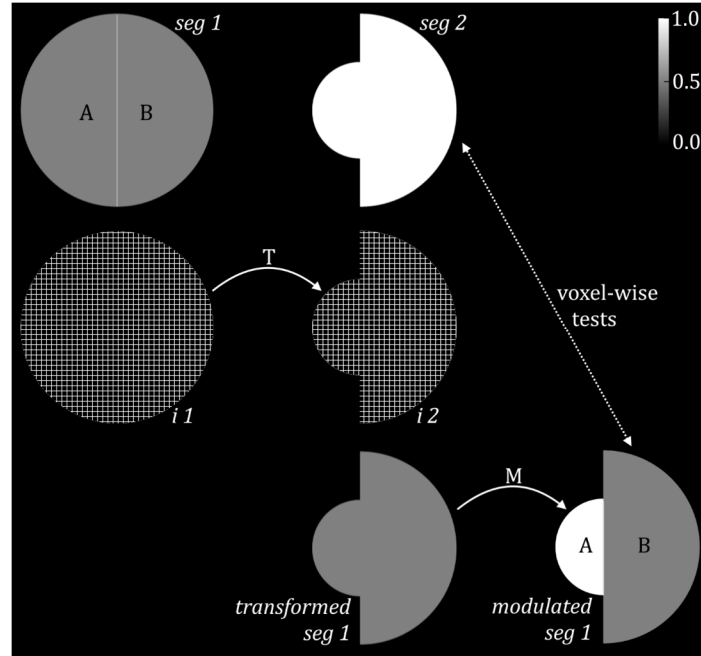
### ***Voxel-Based Morphometry***

In VBM, images are registered and aligned to GWR space, as above. They are either segmented beforehand (in which case, the segmented tissues are warped and interpolated into the final average space, using their images' respective transformations), or after alignment (it is assumed that interpolation and consequent PV does not corrupt the subsequent segmentation, *Ashburner & Friston, 2000*).

The aligned segmentation values, for each tissue class of interest, are then compared at each voxel. It is assumed that these accurately represent the underlying tissue, and that there are no systematic differences between image acquisitions, which might bias segmentations via, for example, different noise levels between scanners. VBM was developed prior to the advent and common availability of highly accurate, computer resource-intensive NRR algorithms, and thus transformations were not expected to achieve precise alignment between images (*Ashburner & Friston, 2001*). A smoothing step (as above) is applied to compensate for possible misalignments. Smoothing also introduces a PV continuum, and, by incorporating neighbouring voxels, constitutes a spatial averaging of segmented values – equivalent to assessing GM “density” or “concentration” within an ROI, or the proportion of GM relative to other tissue types, since at every voxel following segmentation, the proportions of all tissues sum to 1 (*Ashburner & Friston, 2000*).

Because smoothing introduces a spatial aspect to the segmentation, *Ashburner & Friston (2000; 2001)*, and *Good et al. (2001)* suggested applying a “modulation” step. This compensates for the change in total volume of the tissue within structures commensurate with the deformations from registration (*Fig 2.11*). Voxels' concentration values are multiplied by their  $J_{det}$  from TBM, thereby preserving tissue volume, and the statistical test is between regional absolute tissue volumes. Modulation allows this comparison having accounted for larger volume differences via registration. Without modulation, the comparison is not between local volume differences (GM or otherwise); it is directly

between tissue concentrations (from the original images' space – i.e., GM probability, or proportion). Hence *Ashburner & Friston (2001)* placed modulated VBM on a “continuum” with TBM: both examine local volume.



**Figure 2.11** Explanation of modulation in VBM.

If initial segmentations  $seg1$  and  $seg2$  of structural images  $i1$  and  $i2$  represent the relative volume of a particular tissue at each voxel, then modulation is a volume-preserving step. Here, greyscale represents intensity, hence, the voxel values after image segmentation. Following segmentation of  $i1$  and  $i2$ ,  $i1$  is registered to  $i2$ . The resulting transformation  $T$  is used to warp  $seg1$  into the space of  $i2$ . Here, region A shrinks and region B retains its volume. To preserve tissue volume at each voxel, transformed  $seg1$  is modulated ( $M$ ) by dividing voxel values by the relative local volumes of (transformed  $i1$ )/ $i1$  (equivalent to multiplying by  $J_{det}$ ). Intuitively, in this illustration, the density increases (modulated A is darker), thereby preserving volume. Region B does not change size, so intensities remain constant. Voxel-wise tests will show differences in B, but none in region A. If modulation was not performed, here they would show differences at all voxels.

Contemporary fine-scale NRR methods, with high degrees of freedom, likely account for the majority of group volume differences within the deformations (as predicted by *Ashburner & Friston, 2000*). Modulated VBM may therefore closely approximate TBM.



Unmodulated VBM may still be used to test for voxel-wise differences in the underlying tissue concentration. Each method is valuable and likely to reveal group effects in different regions due to distinct aspects of neuroanatomy (*Keller et al., 2004; Mechelli et al., 2005*).

### ***Criticisms of VBM***

VBM originally incorporated allowances for some uncertainty in the alignment achieved between brains. Registration would model “macroscopic” volume and shape, while segmented values could reveal smaller-scale, regional volume differences. Several criticisms arose of the formalised VBM methodology proposed by *Ashburner & Friston (2000)*. *Bookstein (2001)*, adopting the “continuum” analogy, noted the difficulty of quantifying volume differences (from registration) using the same scale as intensities (from segmentation). However, this is accounted for by modulation (as above). The same paper indicated, and *Ashburner & Friston (2001)* agreed, that significant apparent group differences may arise between tissues at boundaries where there is a misregistration. This is particularly concerning when the registration is known to be imprecise, such as when low-dimensional warps are used to align images (as in the original VBM implementations), and when there is no true one-to-one correspondence between brains, as in the highly variable human cortex (*Crum et al., 2003*). Systematic misregistrations may also be caused by biological differences between groups (*Teipel et al., 2013*). Contemporary NRR algorithms, with many degrees of freedom, allow improved alignments, especially between the relatively simple structures of mouse brains, and these can be checked visually. Using VBM to examine residual volume differences may therefore be redundant (*Crum et al., 2003*).

*Keller et al. (2004)* noted that unmodulated VBM appeared less sensitive for detecting hippocampal GM atrophy than modulated VBM, when using ‘optimised’ VBM (*Good et al., 2001*), which improves alignment using customised GM templates. Both techniques revealed bilateral patterns of hippocampal volume loss, but this suggests that the majority of group differences were volumetric rather than in tissue concentration.



While the physical meaning of TBM results (volume differences) is unambiguous, the effects underlying statistical maps from VBM require interpretation; rejection of a null hypothesis does not confer the underlying cause. At the beginning of the century, the technique's clinical usefulness was unclear. VBM is a test of either local segmented tissue concentration or (after modulation) absolute volume. The past decade has provided accumulating evidence that VBM measures are reproducible (*Ewers et al., 2006*) and correlate well with pathology, such as atrophy in Alzheimer's disease (*Karas et al., 2003*; *Chételat et al. 2005*; *Whitwell et al., 2008*; *Teipel et al., 2013*). Several hypotheses exist regarding the tissue changes underlying VBM, including atrophy, altered neurogenesis or cell-specific changes such as size and density (*Keifer et al., 2015*). These will affect tissue  $T_1$  and  $T_2$  relaxation times, and hence the MRI signal and its segmentation (*Zatorre et al., 2012*).

Additional criticisms arise from the fact that segmentation methods which produce variable or unpredictable estimates in the presence of PV may produce unreliable VBM results in these regions (*Thacker, 2003*). This might be an important concern in the mouse brain, wherein the degree of PV is high, and may vary regionally between strains.

### **Statistical tests**

In the final GWR average space, at each brain voxel (and thus, at anatomically equivalent locations across all images), groups' tissue segmentations or log-transformed  $J_{det}$  values may be statistically compared to derive the most significant regions of difference, and hence of most scientific interest, visualised using a statistical parametric map. A common method for doing so is with t-tests to compare group means, using the general linear model (GLM) embedded in SPM, FSL, or other analysis software. The null hypothesis  $H_0$  is that means are equal. It is assumed that every voxel in image  $i$  can be tested independently, and that after fitting the model, the residuals are normally distributed (*Ashburner & Friston, 2000*). Each voxel represents a response variable  $y_i$  which is linearly dependent upon a summed combination of  $k$  independent variables  $x_i$ , and their respective weights  $b$ , which represent the contribution of the variable to  $y_i$ :

E2.2

$$y_i = b_0 + b_1x_{i1} + \cdots + b_kx_{ik} + \varepsilon_i$$

Here,  $b_0$  is an intercept term and  $\varepsilon_i$  is the residual error. In general form, E2.2 can be expressed as:

E2.3

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$$

Here,  $\mathbf{Y}$  is a matrix whose rows correspond to each image and columns to aligned voxels;  $\mathbf{X}$  is a design matrix coding group memberships (such as WT or transgenic) and covariates describing the independent variables (such as TIV; age; sex);  $\mathbf{B}$  is a vector of parameters to be estimated;  $\mathbf{XB}$  is their dot product; and  $\boldsymbol{\varepsilon}$  is a vector of assumed normally-distributed errors, such as noise.  $\mathbf{B}$  is estimated ( $\hat{\mathbf{B}}$ ) to give the smallest sum of squares of the residuals,  $\sum_i \varepsilon_i^2$ . In SPM this is done using the pseudoinverse of  $\mathbf{X}$ ,  $\mathbf{X}^+$  (Kiebel & Holmes, 2006):

E2.4

$$\hat{\mathbf{B}} = \mathbf{X}^+\mathbf{Y}$$

The model predictions (fitted values  $\hat{\mathbf{Y}}$ ) and residual sum of squared errors  $R$  are then:

E2.5

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{X}\hat{\mathbf{B}} \\ R &= \sum_i (\mathbf{Y} - \hat{\mathbf{Y}})^2 \end{aligned}$$

To isolate the contribution of certain parameters  $b$  to  $y$ , for example if genotype is of interest, and a confounding covariate such as age is not,  $\hat{\mathbf{B}}$  may be multiplied using a vector of appropriate contrasts,  $\mathbf{c}$ . The following equations show the calculation of the t-statistic using the mean residual sum of squares  $MRSS$ , and the standard error,  $SE$ :

$$MRSS = \frac{R}{DOF}$$

$$SE = \sqrt[2]{MRSS \times \mathbf{c}((\mathbf{X}^T\mathbf{X})^+ \mathbf{c}^T)}$$

$$t = \frac{\mathbf{c}\hat{\mathbf{B}}}{SE}$$

The DOF is the number of images minus the number of independent regressors (columns) in  $\mathbf{X}$ . Thus, mass-univariate tests are performed over the entire brain, fitting parameters to each voxel.

Non-parametric permutation tests (as implemented in FSL's *Randomise*) may be more robust to non-normally distributed data (*Ashburner & Friston, 2000*), but require a long time to test each possible permutation of every voxel (or a large enough number of permutations to achieve an estimate of the distribution, *Groppe et al., 2011*). Under the assumption that there is no effect of, for example, genotype ( $H_0$ ), for each voxel genotype labels are exchanged amongst images and a test statistic calculated comparing groups. A distribution is constructed by repeating the test after permuting label combinations. A voxel's p-value is the probability, over the image, of observing a more extreme test statistic (*Nichols & Holmes, 2001*).

### ***The multiple testing problem and false discovery rate***

In a 0.5ml mouse brain with  $40\mu\text{m}^3$  voxels, using conventional mass-univariate t-tests, there shall be  $7.8 \times 10^6$  tests of  $H_0$ . A Type I error rate  $\alpha = 0.05$  therefore implies 390,625 false positives (the familywise error rate, FWER, over the whole image): this is the multiple testing problem (*Nichols & Hayasaka, 2003*). There are various methods for compensating.

Decreasing  $\alpha$  at every voxel increases specificity ( $1 - \alpha$ ), lowering the probability of a single false positive. This simultaneously increases the false negative (Type II error) rate,  $\beta$  (thereby lowering sensitivity, or power,  $1 - \beta$ ), and is the reason the highly conservative Bonferroni correction ( $\alpha_{Bonf} = \alpha/N$ , where  $N$  is the number of tests) is avoided in morphometry with thousands of voxels.

Given an SPM with some voxels declared significant, cluster-based approaches use the size and “mass” of regions of significance to distinguish less significant “noise” such as single, isolated significant voxels (*Smith & Nichols, 2009*). Clusters' locations can be compared with a Gaussian random field to estimate which might arise randomly. This is helpful because smoothing introduces spatial correlation between neighbouring voxels, reducing the number of independent observations (*Brett et al., 2003*). There are

difficulties applying this to structural data, however, as there is greater variance at boundaries, where smaller clusters might be expected than in large, homogeneous regions (*Thompson et al., 2003*).

The false discovery rate (FDR) is a popular and simple alternative control of false positives in functional and structural experiments. If there are  $R$  total rejected null hypotheses,  $R_F$  of which are false, controlling FDR limits the expected mean of  $R_F/R$  over many tests (*Groppe et al., 2011*). If the chosen FDR rate  $q = 0.05$ , on average at most 5% of the  $H_0$  rejections will be false. The procedure assumes tests are independent and is as follows (*Benjamini & Hochberg, 1995; Genovese et al., 2002*):

1. Sort p-values in ascending order, such that for  $m$  tests,  $p_i$  is the  $i$ th smallest.
2. Let  $k$  be the greatest  $i$  for which  $p_i \leq \left(\frac{i}{m}\right)q$ . If  $k$  does not exist, stop.
3. All null hypotheses  $1 \dots k$  are rejected; the remainder are not.

Because structures' variance between subjects is higher in some regions (such as the ventricles), controlling the overall  $\alpha$  at an arbitrary level can lower sensitivity. Controlling the proportion of  $H_0$  rejections which are false is more powerful.

## 2.2.8 Preclinical use of V/TBM

In humans, inter-subject variability in the cortical gyri and sulci make true anatomical correspondence impossible, even in healthy individuals (*Crum et al., 2003*). Hence, techniques such as the Boundary Shift Integral (§2.3; §6) rely upon affine registration alone. However, the mouse cortex is far simpler. The low variability within in-bred strains even allows one-to-one correspondence within cerebellar folds (*Fig 2.9 inset*). Morphometric techniques are therefore especially valid: registration exploits this reduced variability to precisely align large groups, allowing statistical techniques to search for subtle structural changes at small spatial scales.

Though not comprehensive, *Table 2.1* provides an overview of preclinical mouse V/TBM, as used in the literature. DBM (deformation based morphometry) refers to the direct comparison of deformations' magnitude or direction, used for example to compare

*in vivo* to *ex vivo* ([Ma et al., 2008](#)) or within-group local variation ([Kovačević et al., 2005](#)). TBM refers to the voxel-wise analysis of Jacobian determinants (expansion or contraction).

Both TBM and DBM have been used to identify neuroanatomical phenotypes in diseases such as Huntingdon's ([Sawiak et al., 2009b](#); [Zhang et al., 2010](#)), multiple sclerosis ([MacKenzie-Graham et al., 2006](#)); and Alzheimer's ([Lau et al., 2008](#); [Maheswaran et al., 2009a](#)). Once these physical traits have been identified, localised histology may be used to confirm them ([Sawiak et al., 2009b](#); [Keifer et al., 2015](#); [Wells et al., 2015](#)). In light of the above-mentioned taxi driver study, several groups have focussed on plastic changes in the brain resulting from learning or abnormal behaviour. For example, [Nieman et al. \(2007\)](#) summarised findings from various MRI protocols, finding 13/15 mouse behavioural mutants had imaging-identified physical brain alterations. [Lerch et al. \(2011\)](#) found that just 5 days of maze training was enough to induce volume changes specific to the hippocampus or striatum. [Carey et al. \(2013\)](#) used VBM to confirm local GM reductions in the brains of mice with an inducible HIV<sup>15</sup>-linked gene, hypothesised from behavioural differences. [Chen et al. \(2005\)](#) noted that structural differences were not always predictive of functional changes, but also used morphometry to show intra-strain differences were smaller than those between strains. Morphometry techniques are robust to any structural imaging modality with sufficient resolution and contrast: [Nieman et al. \(2006\)](#) used  $\mu$ CT to analyse skull shape, and [Wong et al. \(2014\)](#) found  $\mu$ CT images possessed sufficient contrast to study mouse embryos.

All 44 studies listed describe a registration protocol to align the group of brain images under investigation. 16 include multiple-subject scans. The majority of these were performed with a custom array of receiver coils, so subjects did not require separation (see §3.2). Only 7 use VBM. 14 use *in vivo* brains, but only 5 take advantage of longitudinal measurements. By culling mice at different ages and scanning at high resolution, cross-sectional studies may measure longitudinal development ([Zhang et al., 2005](#); [Sussman et al., 2013](#)). However, this requires more mice and is subject to inter-

---

15: Human immunodeficiency virus.

individual variability: serial measurements in the same animals would improve statistical power. It is clearly important to establish whether advanced image analysis techniques may provide sufficient detail and power *in vivo*.

The relative infrequency of mouse VBM studies may be due to several factors. TBM measures an intuitive, easily-identified metric from deformation fields (local physical expansion or contraction). VBM requires more pre-processing steps (tissue segmentation; transformation of those segmentations; and sometimes modulation) and is less easily interpreted (see *criticisms*, above). Although histological tissue volume has been correlated with TBM results (e.g. *Kielar et al., 2012*), it is not clinically established what underlying tissue changes are reflected in VBM – perhaps because of the difficulty of performing neuronal counts or obtaining human tissue post-mortem (*Teipel et al., 2013*). A recent paper, however, suggested that dendritic spine density may cause VBM changes (*Keifer et al., 2015*).

Far more papers are published exploiting morphometric techniques in human brain MRI<sup>16</sup>. In part, this is likely thanks to the availability, and ease of use, of widely-available, integrated software packages, customised for human data with parameters, atlases and tissue prior maps. These include 3D Slicer<sup>17</sup>; AtlasWerks<sup>18</sup>; BrainVoyager<sup>19</sup>; SPM; Freesurfer; and FSL.

As discussed in §3.1, these packages omit, or are not suitable for, some of the preprocessing steps necessary for morphometry in preclinical images. For example, there is no provision for multiple subject scans, and no way to orient brains automatically to standard space if they are not already in a predefined orientation. Tissue priors and registration parameters must be customised for the scale and shape of mouse brains; and non-uniformity correction may be necessary at several stages, to account for surface coils and the higher field strengths of preclinical scanners.

---

16: In a review, *Piras et al. (2015)* found 156 VBM papers addressing obsessive-compulsive disorder, alone.

17: [www.slicer.org](http://www.slicer.org), *Fedorov et al. (2012)*.

18: [www.sci.utah.edu/software/atlaswerks.html](http://www.sci.utah.edu/software/atlaswerks.html)

19: [www.brainvoyager.com](http://www.brainvoyager.com)

Very few of the studies listed in [Table 2.1](#) describe the processing steps in detail. Although not distributed as integrated software packages, some pipelines do exist for processing preclinical images. [Kovačević et al. \(2005\)](#) included a method for intensity standardisation – a necessary prior step to remove bias (§3.7) – based upon the GM histogram peak.

[Lee et al. \(2009\)](#) used the UFL atlas ([Ma et al., 2005](#)) as a basis for skull-stripping, tissue segmentation, and parcellation. [Gerig et al. \(2011\)](#) developed a pipeline for human and *in vivo* rodent images, incorporating a standard atlas space; EM-based tissue segmentations and parcellations similar to [Lee et al. \(2009\)](#), with DTI and cortical thickness analysis. They generated tissue priors for rat data using the NUS mouse atlas. [Badea et al. \(2012\)](#) described a pipeline for skull-stripping, GWR and atlas-based parcellation for multi-spectral (T<sub>1</sub>W; T<sub>2</sub>W; DTI) MR data, using the standard Waxholm space and an average atlas. [Budin et al. \(2013\)](#) described a web-based interface to simplify these steps. In each of these, procedures to accommodate multi-subject scans were omitted. Uptake of such standardised methods has been slower in the preclinical paradigm. Their use is especially important because the small scale of mouse brains will magnify the importance of errors.

The aim of §3 is to show extensions to these pipelines, integrated with up-to-date clinical registration, segmentation and multi-atlas parcellation techniques, to perform all the necessary image processing steps fully automatically, to meet the need for image processing in high-throughput preclinical studies.

<b>Reference</b>	<b>target</b>	<b>life</b>	<b>longitudinal</b>	<b>resolution (<math>\mu\text{m}^3</math>)</b>	<b>statistical morphometry</b>	<b>multi-subject scans</b>
<i>Ali et al. (2005)</i>	C57Bl/6J	ex.		90	-	
<i>Chen et al. (2005)</i>	WT	ex.		60	DBM & TBM	
<i>Kovačević et al. (2005)</i>	in-bred	ex.		60	DBM	2
<i>Ma et al. (2005)</i>	C57Bl/6J	ex.		47	DBM	
<i>Verma et al. (2005)</i>	age	ex.		93 & 120	- (FA)	
<i>MacKenzie-Graham et al. (2006)</i>	MS	ex.		60	-	
<i>Nieman et al. (2006)</i>	GJA1	in. & ex.		110 & 80	DBM & TBM	•
<i>Badea et al. (2007b)</i>	Reln	ex.		21.5	shape	
<i>Clapcote et al. (2007)</i>	schizophrenia	ex.		-	TBM	
<i>Nieman et al. (2007)</i>	behaviour	in. & ex.		32-115	DBM & TBM	•
<i>Spring et al. (2007)</i>	sex	ex.		32	DBM & TBM	3
<i>Lau et al. (2008)</i>	AD	in.	L	156	TBM <sup>D</sup>	
<i>Lerch et al. (2008)</i>	HD	ex.		32	DBM & TBM	3
<i>Ma et al. (2008)</i>	C57Bl/6J	in.		100	DBM	
<i>Aggarwal et al. (2009)</i>	C57Bl/6J	in. & ex.		50-125	DBM	
<i>Lee et al. (2009)</i>	FXS	ex.		100	-	
<i>Li et al. (2009)</i>	schizophrenia	in.		98-250	VBM	
<i>Maheswaran et al. (2009a)</i>	AD	in.	L	78-156	TBM <sup>D</sup>	
<i>Mercer et al. (2009)</i>	Magel2	ex.		32	TBM	3
<i>Sawiak et al. (2009c)</i>	HD	ex.		70	VBM	
<i>Sawiak et al. (2009b)</i>	HD	ex.		70	-	
<i>Badea et al. (2010)</i>	AD	ex.		21.5, 43	TBM <sup>D</sup>	
<i>Ellegood et al. (2010)</i>	FXS	ex.		32	TBM	3
<i>Spring et al. (2010)</i>	C57Bl/6J	ex.		32	TBM	3
<i>Xie et al. (2010)</i>	AD (rTg4510)	in.		125-300	TBM	
<i>Zamyadi et al. (2010)</i>	embryo in-bred	ex.		~49	DBM & TBM	32
<i>Zhang et al. (2010)</i>	HD	in. & ex.	L	80-250	TBM <sup>D</sup>	
<i>Cheng et al. (2011)</i>	HD	in.	L	100-250	TBM <sup>D</sup>	
<i>Chuang et al. (2011)</i>	age	ex.		-	DBM	
<i>Ellegood et al. (2011)</i>	autism	ex.		32	TBM	3
<i>Lerch et al. (2011)</i>	learning	ex.		32	TBM <sup>D</sup>	3
<i>Teipel et al. (2011)</i>	AD	in.		31.25-600	VBM	
<i>Yu et al. (2011)</i>	Fgf17	in.		100	TBM <sup>D</sup>	
<i>Biedermann et al. (2012)</i>	exercise	in.		78-156	VBM	
<i>Carey et al. (2013)</i>	learning/memory	ex.		39-500	VBM	
<i>Kielar et al. (2012)</i>	ataxia	ex.		70	TBM	
<i>Badhwar et al. (2013)</i>	AD	ex.		32	TBM <sup>D</sup>	3
<i>Ellegood et al. (2013)</i>	autism	ex.		16	TBM	56
<i>Sawiak et al. (2013)</i>	HD	in. & ex.		70	VBM & TBM	
<i>Sussman et al. (2013)</i>	diet	ex.		130	TBM	
<i>van Eede et al. (2013)</i>	C57Bl/6J	ex.		56	TBM <sup>D</sup>	16
<i>Wong et al. (2014)</i>	embryo age	ex.		13.4 ( $\mu\text{CT}$ )	TBM <sup>D</sup>	
<i>Allemang-Grand et al. (2015)</i>	AD	in & ex.	L	125, 78 & 56	TBM	7
<i>Keifer et al. (2015)</i>	learning	ex.		81-162	VBM	9



<i>multi-subject extraction</i>	<i>orientation to standard space</i>	<i>skull stripping</i>	<i>non- uniformity correction</i>	<i>tissue segmentation</i>	<i>intensity standardisation</i>	<i>(group) registration</i>	<i>atlas-based parcellation</i>	<i>manual parcellation</i>
						•		•
						•	•	
			•	•	•	•		•
						•	•	•
		•	•			•		•
		•			•	•		•
						•		•
			•			•		•
				•		•	•	•
					•	•	•	•
	•	•	•	•		•	•	•
				•		•		•
	•					•		•
		•			•	•	•	
	•					•	•	
		•	•		•	•		•
•			•			•		•
	•				•	•		•
	•	•			•	•	•	
	•					•		•
		•	•	•		•	•	
		•			•	•		•
		•	•	•		•	•	
			•	•		•		•
			•	•		•	•	
				•		•	•	
				•		•	•	
				•		•		•
•				•		•		•

**Table 2.1** Selection of morphometry papers and pipeline features (previous page).

Publications (sorted by year) including voxel-wise brain (41) or embryo (2) morphometry measures found during a literature search encompassing mouse MRI (one  $\mu$ CT) morphometry or phenotyping. ‘•’ indicates the processing step (detailed in §3) is at least mentioned. Age: ageing, development, or maturation; C57BL/6(J): in-bred WT mice; Fgf17: Fibroblast growth factor 17 (brain organisation in early development); FXS: fragile X syndrome; GJA1: oculodentodigital dysplasia; HD: Huntington's disease; Magel2: Prader-Willi syndrome; MS: multiple sclerosis; ReIn: Reeler mouse neurodevelopmental model; sex: sexual dimorphism; rTg4510: tauopathy in AD (see §5); WT: comparison of WT in-/out-bred variability. TBM<sup>D</sup>: referred to by authors as DBM; shape: comparison of brain and substructure surface mesh shapes between groups; FA: comparison of registered FA maps. Resolution: smallest-largest voxel dimensions; isotropic otherwise.

## 2.3 Alzheimer's disease

Among the many genetic and sporadic ailments which we may wish to model in animals, neurodegenerative diseases, with their devastating, debilitating effects, are perhaps the most sinister. Thanks to several revolutions in healthcare knowledge and practice of the past 200 years – including in anaesthesia, blood transfusion, germ and antiseptic theories, vaccines, artificial organs and imaging – humans, in increasing populations, can expect to live longer. With our advancing years, natural brain tissue loss is inevitable (termed “healthy ageing”). However, in some individuals, age confers a startlingly increased risk of dementia. It is not fully understood why.

The emotional and economic burden of dementia – loss of cognitive function caused by accelerated neurodegeneration – is extremely high, affecting individuals, families, society, and countries' finances, as individual or institutional care must be provided (*Ferri et al., 2005*). The annual cost to the UK is estimated as £26.3Bn<sup>20</sup>; the majority is paid by individuals and families, rather than by the state. One of the strongest risk factors is age: in the UK, 1 in 700 people under 65 have dementia; but over 65, 1 in 14; and over 80,

---

20: This is 1.3% of UK gross domestic product (*World Bank, 2014*).

1 in 6. Worldwide, by 2050, the population of those over 60 years of age will increase by 1.25Bn, and 115 million people will have dementia if disease-modifying treatments are not found (*Brookmeyer et al., 2007; Prince et al., 2013; 2014*).

Alzheimer's disease (AD) is the most common cause of dementia, representing 62% of cases in the UK. Vascular dementia (17%), dementia with Lewy bodies (4%) and frontotemporal dementia (FTD, 2%) follow (*Prince et al., 2014*). Patients may present for diagnosis after family members notice a slow onset of memory loss, arising over several months, or behavioural changes. Diagnosis relies upon a large number of factors, including history, performance on cognitive tests such as the Mini-Mental State Examination (MMSE)<sup>21</sup>, and clinical judgment. Mild cognitive impairment (MCI) may be suggested if the patient retains social or occupational independence; dementia, if they do not (*Gauthier, 2007; Albert et al., 2011*). Brain imaging, such as MRI, can provide valuable, complementary, differentiating information to distinguish early dementia or MCI (*Frisoni et al., 2003*). The patterns of atrophy revealed by high-resolution structural MRI can also aide the diagnostic separation of AD, FTD, dementia with Lewy bodies and vascular dementia (*Frisoni et al., 2010*).

Cognitive decline in AD is strongly correlated with neurodegenerative atrophy, the cumulative damage, synapse loss and neuronal death detectable by MRI as cortical thinning, hippocampal shrinkage, and ventricular enlargement (*Terry et al., 1991; Vemuri et al., 2009; Vemuri & Jack, 2010*). These effects are downstream of underlying, driving pathophysiological brain changes.

Diagnosis of AD is “possible” or “probable” until “definite” confirmation with histology via biopsy or autopsy (*Gauthier, 2007*). The hallmarks in tissue are extracellular plaques of amyloid-beta ( $A\beta_{40}$  and  $A\beta_{42}$ : respectively, soluble and hydrophobic amino acid chains) and intracellular neurofibrillary tangles (NFTs) of hyperphosphorylated tau. These have characterised the disease since their first description, in 1911 (*Goedert &*

---

21: A score below 24/30 implies early dementia, but “normal” scores also vary between individuals.

*Spillantini, 2006*), and are found in both early- and late-onset cases (respectively, pre- and post-age 65).

The majority of cases (>95%) are late-onset. The remainder usually progress faster and are often from families with a history of AD (*Gauthier, 2007; Reitz & Mayeux, 2014*). Both inherited genes and environmental factors are thought to play a role in AD pathogenesis. A subset of early-onset, familial cases (<1% overall) are autosomal dominant<sup>22</sup> (ADAD), in which rare mutations in three genes are inherited. Their penetrance (the chance of the phenotype arising, given the mutation) is >85% (*Bertram & Tanzi, 2008; Bateman et al., 2011; Reitz & Mayeux, 2014*). As these account for only a small proportion of familial AD cases, there are likely further, undiscovered genetic risk factors. Only a few genes, most identified in genome-wide association studies, have so far been implicated in late-onset AD. Of these, the strongest effect is from an allele of ApoE<sup>23</sup>,  $\epsilon 4$ . Carrying one copy of this variant increases the risk of developing AD at some point by 2—3 times; two copies increase it by 5 or more (*Reitz & Mayeux, 2014*).

$A\beta$  is formed by cleavage of the amyloid precursor protein APP, by the proteases  $\beta$ - or  $\gamma$ -secretase (*Giacobini & Gold, 2013*). All three ADAD mutations (in APP, presenilin 1 and presenilin 2 genes) are related to this process, by increasing  $A\beta$  production or the  $A\beta_{42}/A\beta_{40}$  ratio (*Bertram & Tanzi, 2008*). ApoE mediates  $A\beta$  clearance from the brain, and of its three alleles,  $\epsilon 4$  is the least efficient (*Tarasoff-Conway et al., 2015*). These observations reinforced the “amyloid cascade hypothesis”, which posits that an imbalance between  $A\beta$  build-up and clearance precipitates synaptic damage via toxicity; the formation of NFTs; and subsequent neurodegeneration (*Hardy & Selkoe, 2002; Giacobini & Gold, 2013*).

Tau is a highly soluble intracellular protein which binds microtubules and contributes to axonal integrity. Insoluble, hyperphosphorylated tau ends its stabilising function and aggregates as NFTs, which impair normal axonal transport and are implicated in both

---

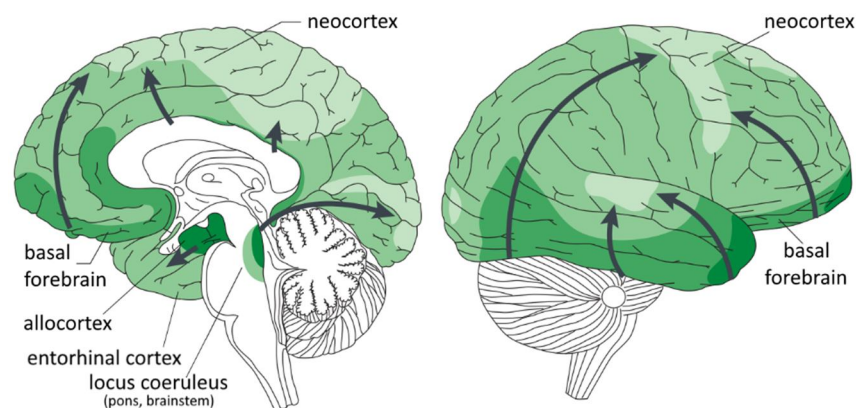
22: Only one copy of the gene is required for an effect. The chance of inheriting the mutated gene from a carrying parent is 50%.

23: Apolipoprotein E.

the onset and progression of neurodegeneration in tauopathies, including AD, FTD, and Parkinson's disease. Neither FTD nor Parkinson's patients exhibit amyloid plaques, hence, NFTs are sufficient to cause neurodegeneration by themselves. Around 5% of FTD cases are caused by mutations in the tau-encoding gene MAPT<sup>24</sup> (*Lee et al., 2001; Ballatore et al., 2007; Barten et al., 2012; Spillantini & Goedert, 2013*). Elevated NFT levels correlate with cognitive scores better than amyloid plaques (*Giannakopoulos et al., 2003*), and, unlike plaques, whose distribution throughout the brain varies between individuals, they spread in a recognised, characteristic pattern which correlates with grey matter loss (*Fig 2.12; Braak & Braak, 1995; Whitwell et al., 2008*). "Braak stages" are:

- I-II      transentorhinal: entorhinal cortex and hippocampus;
- III-IV    limbic: amygdala, basolateral temporal lobe;
- V-VI      isocortical: association areas of neocortex.

Stages III-IV coincide with initial memory loss and MCI; V-VI with dementia (*Vemuri & Jack, 2010*).



**Figure 2.12** The pathological spread of tau in human Alzheimer's disease.

NFTs spread from dark (earlier, where the disease is, eventually, most severe) to light (later). Tau deposits are thought to originate in the locus coeruleus. The allocortex includes the hippocampus. Adapted from *Brettschneider et al. (2015)*, with permission.

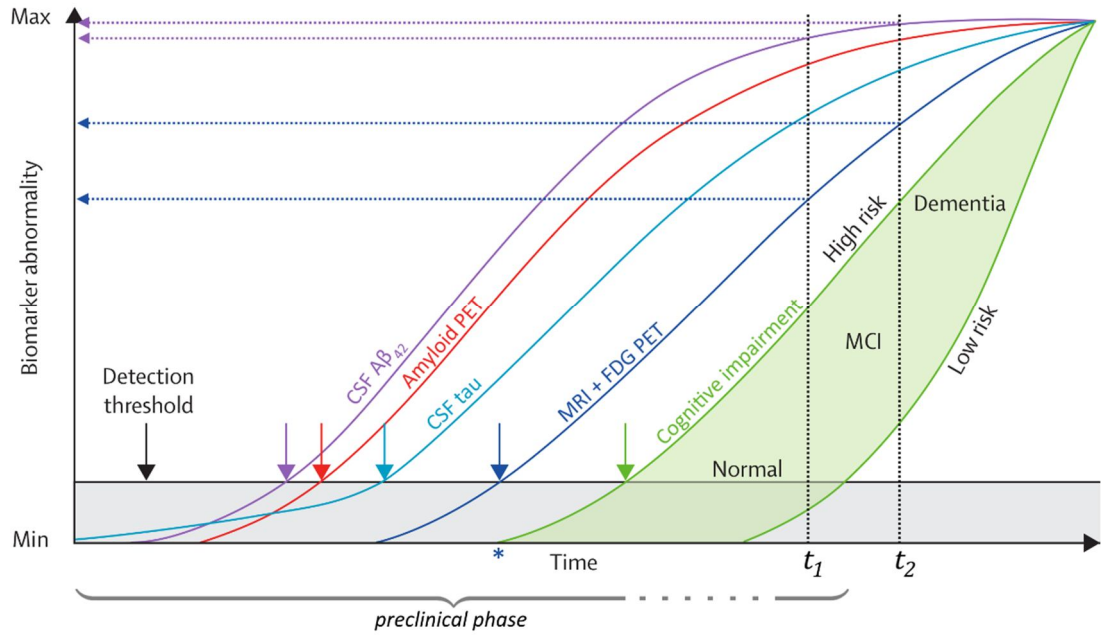
24: Microtubule-associated protein tau.

$A\beta$  and tau are misfolded proteins, which may act as “corruptive templates” for future misfolding and hence aggregation. Their spread may be governed by the strength of neuronal connections between regions (*Brettschneider et al., 2015*). Although tauopathic neurodegeneration occurs in FTD without  $A\beta$ , and in early-onset AD,  $A\beta$  pathology is a sufficient trigger, the two separate mechanisms are inexorably linked in late-onset AD – the vast majority of cases – which appears to result from a failure to clear these misfolded proteins or their toxic, soluble forms (*Jack et al., 2013*). A shared upstream defect has been suggested, for example the ApoE  $\epsilon 4$  genotype (*Small & Duff, 2008*).

### 2.3.1 Biomarkers

Neurodegenerative diseases can coexist and their clinically-presented phenotypes (patient history; symptoms; MMSE score) may overlap (*Clark et al., 2008*). Biomarkers should ideally be specific (to aide differential diagnosis); cheap; non-invasive; reproducible, unbiased and objective (to lend confidence); and predictive of future progression (to help patients and families plan). *Brookmeyer et al. (2007)* estimated that a short (one-year) delay to symptom onset would reduce worldwide cases in 2050 by 11.8 million. Therefore, they should also be sensitive to early disease onset and the effects of potential treatment (*Hampel et al., 2008; 2010*), which also reduces the costs of clinical trials, by improving power and requiring fewer subjects. The Alzheimer’s Disease Neuroimaging Initiative (ADNI), an international, longitudinal study, was formed in 2004 to assess and develop such biomarkers (*Weiner et al., 2013*). Promising candidates include cerebrospinal fluid (CSF) sampling, positron emission tomography (PET), and structural MRI.

*Jack et al. (2013)* proposed a sigmoidal model (*Fig 2.13*), now widely reproduced, in which biomarkers progress simultaneously, but are most dynamic, and reach maximum abnormality, in an ordered sequence.



**Figure 2.13** Theoretical progression of AD biomarker abnormality with time.

Each biomarker surfaces above the detection threshold, and reaches a peak rate of change, at different times prior to dementia, but all are dynamic simultaneously. Some “cognitive reserve” differentiates low- and high-risk patients with the same biomarker abnormality (*Vemuri et al., 2011*), and there is a long “preclinical phase” prior to diagnosis. Atrophy, detected with MRI, coincides with the onset of cognitive problems (\*). Two sampling time-points,  $t_1$  and  $t_2$  illustrate that MRI may be more useful than CSF  $A\beta_{42}$  at the MCI stage. Adapted from *Jack et al. (2013)*.

CSF biomarkers include reduced  $A\beta_{42}$  levels, reduced  $A\beta_{42}/A\beta_{40}$  ratio, and elevated phosphorylated tau. These follow directly from neurochemical brain changes. Additionally, elevated total tau concentration reflects neurodegeneration (*Hampel et al., 2008; 2010*). CSF markers are sensitive, track the severity of disease, and in combination are specific to AD. However, they omit spatial information, and sampling requires an invasive, often painful lumbar puncture (*Vemuri et al., 2009*).

For PET to map  $A\beta$  deposits or monitor brain metabolism, radiotracers are injected minutes before a scan.  $^{11}\text{C}$ -labelled Pittsburgh Compound-B (PiB) binds selectively to  $A\beta$ . PiB uptake very closely correlates with CSF  $A\beta_{42}$  decrease (*Vemuri et al., 2009*) and has detected high plaque burdens in MCI patients, but also in cognitively normal elderly

people: it is only loosely predictive of decline (*Hampel et al., 2010; Jack et al., 2013*).  $^{18}\text{F}$  fluorodeoxyglucose (FDG) may be injected to measure resting state glucose metabolism: its uptake, like neuronal activity, is locally reduced in neurodegeneration-afflicted regions. FDG-PET is highly predictive of conversion from MCI to AD (*Hampel et al., 2008; 2010*). The short half-lives<sup>25</sup> of the isotopes  $^{11}\text{C}$  and  $^{18}\text{F}$ ; their radioactivity; and PET's high cost and limited spatial resolution may, however, limit its usefulness, especially in small animal studies (*Clark et al., 2008; Götz & Ittner, 2008*).

In *Fig 2.13*, prodromal amyloid markers (CSF  $A\beta_{42}$ ; PiB-PET) are the first to surface above the current “detection threshold”. *Buchhave et al. (2012)* found that CSF  $A\beta_{42}$  levels reached maximum abnormality in MCI patients up to 10 years prior to conversion to AD. It is increasingly recognised that underlying pathophysiological processes, as measured by these techniques, begin a decade or more before cognitive decline and subsequent diagnosis. This asymptomatic “preclinical phase”, during which subjects can be cognitively normal (*Ewers et al., 2011; Sperling et al., 2011*), motivated the extension of ADNI to even earlier biomarkers for detecting and distinguishing newly-defined early MCI and subjective (worries of) memory impairment (SMI) stages, which carry an elevated risk of eventually developing AD (*Jessen et al., 2014; Beckett et al., 2015*).

There is histological evidence from autopsy of young brains (under 30 years) that pathologic, phosphorylated tau appears before amyloid deposits (*Spillantini & Goedert, 2013*). CSF total tau levels are directly related to age and tangles also form in healthy ageing. They may cause the slow atrophy which accompanies all healthy ageing, and underlie associated mild memory impairment (*Fjell et al., 2014*). Hence, in *Fig 2.13*, tau abnormality is postulated to precede amyloid deposits, but is only detectable at autopsy.

### **MRI biomarkers**

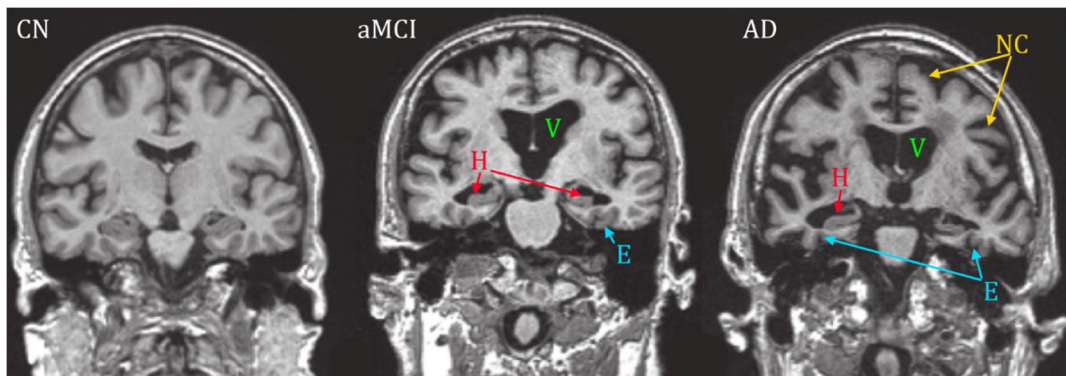
Non-invasively, and relatively cheaply, structural brain MRI provides high tissue contrast and resolution, enabling the sensitive and localised detection of atrophy, which can both differentiate dementias and objectively stage AD *in vivo* (*Frisoni et al., 2010; Vemuri et*

---

<sup>25</sup>:  $^{11}\text{C}$  has a half-life of approx. 20 minutes, requiring a cyclotron on-site;  $^{18}\text{F}$ : 110 mins.



*et al.*, 2010). Atrophy follows NFT pathology in the signature Braak stages, and its rate correlates with cognitive decline (Fox *et al.*, 1999). Vemuri *et al.* (2009) found that MRI performed better than CSF biomarkers at separating MCI and AD groups, and at predicting MMSE score. These advantages have propelled MRI to the early stages of clinical diagnosis. Other MR techniques, including diffusion tensor imaging (DTI, to measure white matter integrity) and functional MRI (fMRI, to assess the decline in connectivity), have been used as biomarkers of AD (Reitz & Mayeux, 2014), and may be acquired in the same patient visit, but these are currently not clinically established (Frisoni *et al.*, 2010), and this thesis is exclusively concerned with structural images.



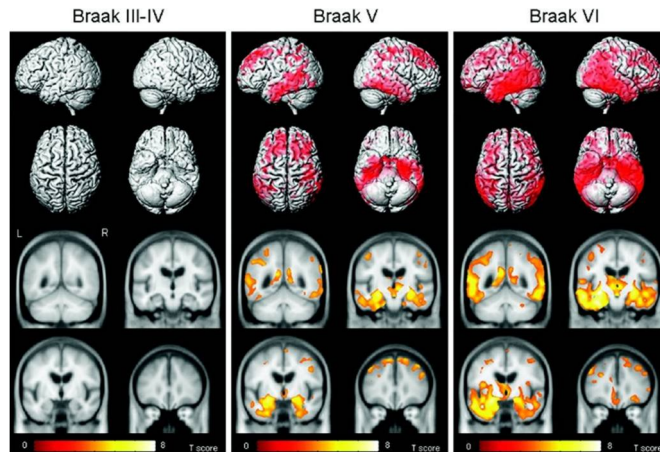
**Figure 2.14** Human T<sub>1</sub>W MRI showing structural changes with MCI and AD. Equivalent coronal T<sub>1</sub>W slices from cognitively normal (CN), amnesic MCI (aMCI) and AD brains, with ventricular enlargement (V), hippocampal volume loss (H), and atrophy of the entorhinal cortex (E) and neocortex (NC). Adapted from Vemuri *et al.* (2010).

The first MR atrophy measures involved visual assessment, or manual volumetry via delineation of vulnerable regions, such as the hippocampus (Fig 2.14). The advance of clinical scanner field strengths from 1.5T to 3 and 7T has improved resolution and contrast, and hence the feasibility of these techniques. However, visual inspection is subjective and manual measurements are extremely time-consuming, and variable.

To better-elucidate and quantify structural brain changes, advanced image processing techniques – including VBM and TBM – have emerged over the past two decades, exploiting increases in computer processing power, which have enabled high-dimensional

image registration and accurate automated anatomical parcellation. These techniques are widely available for human studies, thanks to free software packages (§2.2.7) and atlas databases (§2.2.1).

For example, in a longitudinal study, *Chételat et al. (2005)* showed VBM could distinguish MCI patients who converted to AD from those who did not. *Whitwell et al. (2008)* used VBM in a cross-sectional study of 82 human subjects with probable AD and different Braak stages, and showed regional GM loss proportional to the degree of NFT pathology (*Fig 2.15*). *Hua et al. (2013)* used an unbiased implementation of TBM and detected constant atrophy rates in AD over 6—24 months, with most changes localised to the cortical and hippocampal GM, as well as temporal lobe WM.



**Figure 2.15** VBM-derived GM loss at different Braak stages.

In a cross-sectional study, human brain MRI scans from patients with late Braak stages were compared with controls from earlier stages (Braak 0—II). VBM detected the characteristic pattern of progressive atrophy. 3D surface map (top) and four coronal slices for each set of stages. N=20 controls; N=23 III-IV; N=32 V; N=27 VI. FDR-corrected,  $q = 0.005$ . From *Whitwell et al. (2008)*, with permission.

The onset of NFT-related neurodegeneration likely precedes MCI by several years (*Teipel et al., 2013*). *Fig 2.13* illustrates that MRI-measured atrophy is highly dynamic during MCI and conversion to AD (between  $t_1 - t_2$ , the MRI biomarker's gradient is one of the largest). Meanwhile, CSF  $A\beta_{42}$  approaches its maximum abnormality, and the degree

of change between time-points is smaller. A multi-time-point study of atrophy rates, to stage disease, distinguish high risk from low risk patients, or to monitor drug effectiveness, may thus use more closely-spaced samples and require lower sensitivity or fewer subjects than CSF  $A\beta_{42}$  (*Frisoni et al., 2010*). While CSF  $A\beta_{42}$  may be abnormal and sensitive earlier, without symptoms or a policy of population screening, it is unlikely to be measured so early in a patient's life. Screening is expensive, may induce needless worry, and currently there are few treatment options, even with a positive diagnosis (§2.3.3). The side-effects of existing treatments also mean that confidence in diagnosis is vital.

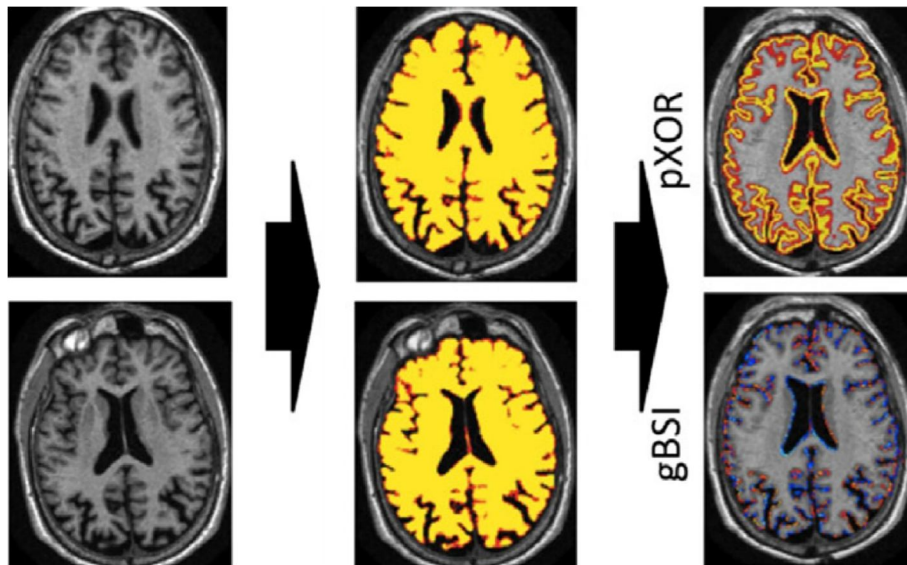
Recent efforts have therefore focused on using imaging to detect subtle changes prior to the onset of symptoms (lowering *Fig 2.13*'s "detection threshold"), by improving robustness (such as by using symmetric registration, with brain masks, for TBM: *Hua et al., 2013*), and focusing on volumes of the entorhinal cortex, hippocampus, ventricles and whole brain (*Frisoni et al., 2010*; *Vemuri et al., 2010*; *Teipel et al., 2013*). Using VBM, *Tondelli et al. (2012)* detected structural differences between the brains of healthy controls and people who would go on to develop AD symptoms up to 10 years before they did so.

### 2.3.2 Outcome measures

Measurable outcomes, relevant to both the disease and treatment, are required evidence for drug trials (*Frisoni et al., 2013*). Direct outcome measures include patient feeling, function, and survival. Imaging biomarkers, such as slowing of atrophy, can substitute for clinically meaningful trial endpoints. Whole-brain summary metrics can provide such surrogate markers of disease progression. The STAND<sup>26</sup> score summarises individual patients' GM densities by comparing with a database of clinically well-characterised images. This performs well in comparison to CSF sampling (*Vemuri et al., 2008; 2009*).

---

26: Structural Abnormality Index.



**Figure 2.16** Illustration of the Boundary Shift Integral in the human brain.

Left, two transverse slices of the same brain, from serial MR scans several months (or years) apart. Middle, probabilistic masks of each; right: probabilistic boundary region and, below, result map (blue indicating atrophy; red indicating growth). The BSI provides a summary whole-brain measure of longitudinal volume change. Adapted from [Prados et al. \(2014\)](#).

Early in AD, total tissue volume loss may fall within a broad range explicable by natural variation and healthy ageing ([Fjell et al., 2014](#)). The rate of change within an individual patient is thought to provide a better biomarker. This is the motivation behind the Boundary Shift Integral (§6; [Fig 2.16](#)). The BSI uses voxel intensity differences between aligned images from two or more time-points (e.g.,  $t_1$  and  $t_2$ ) to assess atrophy over the whole brain, or within delineated substructures. The method has developed from using absolute to, more recently, automated, unbiased and image-specific adaptive parameters and probabilistic segmentations ([Fox et al., 1996](#); [Prados et al., 2014](#)), and has been customised for the ventricles and hippocampus ([Schott et al., 2010](#)). Local measurements from each voxel are combined into a summary metric, which is easily compared between patients. The order of 1.9% of brain volume is lost per year to atrophy (§6.5.1).

SIENA<sup>27</sup> (*Smith et al., 2002*) also provides a summary metric from measurements of tissue boundary displacement, computed from intensity gradient profiles projected from the brain surface, over time. Such longitudinal measures can be used to assess therapeutic effectiveness. There is a need to improve the robustness of preclinical trials, but heretofore, none of these methods has been employed preclinically.

### 2.3.3 Treatment

There are presently no disease-modifying treatments for AD; contemporary drugs (5 are FDA-approved, *Mangialasche et al., 2010*) only mitigate behavioural and cognitive symptoms, and can delay the need for full-time care, but not progression or onset. Neuroinflammation, thought to be a response to amyloid plaques, is a (non-specific) feature of AD; some studies have suggested anti-inflammatory drugs may reduce the risk of, and delay, dementia (*Gauthier, 2007*).

Until recently, the amyloid cascade hypothesis focussed efforts on reducing  $A\beta$  levels, by suppressing secretases or increasing clearance (*Giacobini & Gold, 2013*). As MCI symptoms begin, amyloid therapy is possible, but by this point  $A\beta$  abnormality has already plateaued, whereas atrophy is still dynamic (*Jack et al., 2013*).

The interactions of  $A\beta$  and tau are still not fully understood. In human ADAD, APP-related mutations are sufficient to lead to all AD features, but not in mice (*Gauthier, 2007; Ashe & Zahs, 2010*).  $A\beta$  may “drive” or aggravate NFT pathology; tau may mediate the toxicity of  $A\beta$ , or each may amplify the effects of the other (*Ittner & Götz, 2011; Giacobini & Gold, 2013*). A recent call to reject the amyloid cascade hypothesis (*Herrup, 2015*) emphasised that  $A\beta$  can occur in healthy humans, and that its presence in mice does not cause severe neurodegeneration or permanent damage:  $A\beta$  clearance in mouse models has successfully restored earlier function, but human phase II and III clinical immunotherapy trials<sup>28</sup> have failed to replicate this effect (*Giacobini & Gold, 2013*).

---

27: Structural Image Evaluation, using Normalization, of Atrophy.

28: Phase II: several hundred patients and controls, to test for a disease-modifying effect and safety; phase III: several thousand patients and controls, when some effect is confirmed.

Several trials have been abandoned due to significant side-effects, such as exacerbated inflammation and atrophy (*Giacobini & Gold, 2013*). A recent paper showed (using the BSI) that Bapineuzumab, an  $A\beta$  antibody, increased the rate of ventricular enlargement in some patients (*Novak et al., 2016*). Some AD treatments have elicited severe side-effects in FTD patients (*Götz & Ittner, 2008*). That these can be serious further highlights the need for specific differential diagnosis, to enable correct, disease-specific, targeted prescriptions.

The slowing of atrophy may therefore be a better outcome measure (or surrogate endpoint, *Hampel et al., 2010*). Tau levels may begin their change earlier in AD, prior to MCI symptoms, and so tau represents a promising therapeutic target – for example, by inhibiting phosphorylation (*Jack et al., 2013*).

Many more drugs are currently in development. As tau and amyloid pathologies likely interact, effective AD treatments will have to target both (*Mangialasche et al., 2010; Ittner & Götz, 2011; Giacobini & Gold, 2013*). Mice are vital models both of the underlying genetic causes of AD, and for testing dosage and side effects in therapeutic trials.

#### 2.3.4 Mouse models of Alzheimer's disease

Mouse models of AD are broadly defined by their replication of either  $A\beta$  or tau pathology. Models may recapitulate only some aspects of the human disease, to enable targeted therapy with fewer confounding factors; no single mouse currently mimics the entire spectrum of AD features. AD progresses over many years in humans, so the short mouse lifespan may be an impediment (*McGowan et al., 2006*). The majority are genetic models, expressing mutant APP, presenilin 1 (PS1) or presenilin 2, and are thus models of early-onset, familial AD: a minority of clinical cases.

Transgenic human APP (hAPP) models have been shown to exhibit cognitive and memory deficits preceding or alongside amyloid plaques; however, few show signs of atrophy. For example, the PDAPP mouse overexpresses mutant hAPP and develops age-dependent hallmarks of AD including  $A\beta$  plaques and cognitive decline (*Schenk et al., 1999*). *Redwine et al. (2003)* found volume reductions in the hippocampus in this model

prior to plaque formation. The J20 mouse, which has two APP mutations, exhibits plaques from an early age, and synaptic dysfunction, but only minor neurodegeneration (*Mucke et al., 2000; Hall & Roberson, 2012; §4.2*).

Cross-breeding mice can give offspring with combined, interacting mutations and hence a more complete model. Mice expressing mutant PS1 exhibit elevated  $A\beta$  levels. Plaque aggregation is significantly increased, however, when these mice are cross-bred with a mutant human APP model (e.g. M146L and Tg2576, *Duff & Suleman, 2004*). The resulting PSAPP model exhibits cognitive deficits before plaques appear. The 5XFAD mouse combines five APP and PS1 mutations, exhibits severe AD pathology and also neurodegeneration. Despite their lack of aggregated NFTs, recent work has shown that tau is still tied to cognitive decline in hAPP models (*Hall & Roberson, 2012*).

Transgenic models of mutant human tau are currently less common, but advantageous as they show more severe atrophy than APP models. These models do not develop plaques independently, so are not full models of AD; rather, they better-model FTD, with select AD-like pathologies, including NFTs and neurodegeneration (*Hall & Roberson, 2012*). Such mice often utilise the most common FTD-linked human tau mutation, P301L. The prominent role of tau in AD, and the failure of APP models to replicate the severe atrophy which is the main correlate of cognitive decline, and monitorable with  $\mu$ MRI, motivated the study of the rTg4510 mouse (*§5, §6*), which exhibits severe atrophy due to tauopathy.

The theorised interaction of  $A\beta$  and tau has led to the development of cross-bred mice, which exhibit plaques and tangles. The 3xtg-AD mouse, for example, was bred from mice with PS1, APP and P301L mutations, and closely recapitulates AD pathology (*Duff & Suleman, 2004; Götz & Ittner, 2008*).

Few studies have quantified AD mouse models using advanced neuroimaging techniques (*Table 2.1*). Both *Lau et al. (2008)* and *Maheswaran et al. (2009a)* showed unexpectedly increasing brain volume, with age, in APP/PS1 cross-bred mice. *Badea et al. (2010)* found volume loss prior to amyloid formation in some regions of an APP model. Using SPMMouse for VBM, *Teipel et al. (2011)* found no GM decline in APP/PS1 mice. Using



TBM, *Badhwar et al. (2013)* found both volume increases and localised volume decreases in J20 mice (discussed further in §4.2.4).

As noted above, following the failure of treatments in humans which successfully halted  $A\beta$  accumulation and restored function in mice, there is concern about the validity of extrapolating treatments from mice – often models of familial and ADAD – to the majority of human cases, sporadic AD. Cognitive decline has been shown in APP mice prior to plaque formation; the reverse is found in humans. However, mice may still be good models of presymptomatic AD (*Hall & Roberson, 2012*). In preclinical drug trials, just as in humans, relevant outcome measures are vital.  $A\beta$  clearance does not appear to correlate with cognitive impairment (*Giacobini & Gold, 2013*). Thus, mice exhibiting atrophy, such as rTg4510 – or cross-bred mice with both amyloid plaques, NFTs, and subsequent neurodegeneration – may be better foci. Histology can be used to validate both plaque and tangle formation, via staining, and atrophy, via volume measurement. By measuring the slowing of atrophy, the Boundary Shift Integral (§6) can potentially provide such an outcome measure in these models *in vivo*, allowing longitudinal assessment.

## 2.4 Down syndrome

Down Syndrome (DS) is the most common genetic cause of human intellectual disability, affecting 0.11—0.13% of live births in the UK and USA, and characterised by physical and cognitive developmental deficits (*Morris et al., 2009; Parker et al., 2010*). DS is caused by trisomy (three copies) of human chromosome 21 (Hsa21), leading to the over-expression of genes encoded on this chromosome. As with AD, the majority of cases are sporadic, but risk factors include maternal age (over 37, which is rising on average), a previously affected pregnancy, and translocation of Hsa21 (a partial additional copy), which can be inherited. Incidence at conception has increased since the 1980s, but live births have remained relatively stable, thanks to improved access to prenatal screening



and termination: all mothers are now offered screening for chromosomal abnormalities<sup>29</sup> in the UK and USA. 92% of mothers in England and Wales with an antenatal DS diagnosis opt to terminate (*Roizen & Patterson, 2003; Morris et al., 2009*) – however, thanks to improved healthcare, life expectancy for DS individuals is also rising – to around 60 years today – increasing the DS population overall (*Zigman & Lott, 2007; Wiseman et al., 2015*).

Although the entire human genome has been sequenced, the (sometimes overlapping) function of many of these genes, and their interactions, is still unclear (*Roizen & Patterson, 2003*). The elevated gene dosage conferred by trisomy of Hsa21, and the consequent imbalance compared to those on other chromosomes, gives rise to various phenotypes, with differing prevalence and severity. All DS individuals exhibit craniofacial abnormalities and some learning disability. Around 40% have a congenital ventricular septal defect (a hole in the heart), and 38—78% have some hearing loss (*Roizen & Patterson, 2003*). Brain, cerebellar and hippocampal volumes are reduced (*Pinter et al., 2001a; Beacher et al., 2009*).

Of particular interest to AD researchers, among the 329 (or so) genes extant on Hsa21 is the gene encoding APP (*Roizen & Patterson, 2003*), which is consequently overexpressed in a dose-dependent manner (*Lott & Head, 2001*). After age 40, all people with DS develop amyloid plaques and NFTs. By age 60, 60—70% develop dementia in the form of early onset AD. DS may therefore represent a model of predementia AD (*Teipel et al., 2004*). Some never develop dementia, which hints at a possible protective effect from the remainder of Hsa21 (*Beacher et al., 2009; Wiseman et al., 2009; 2015*). The large number of genes is thought to contribute to the correspondingly numerous phenotypes of DS.

### 2.4.1 Mouse models of Down syndrome

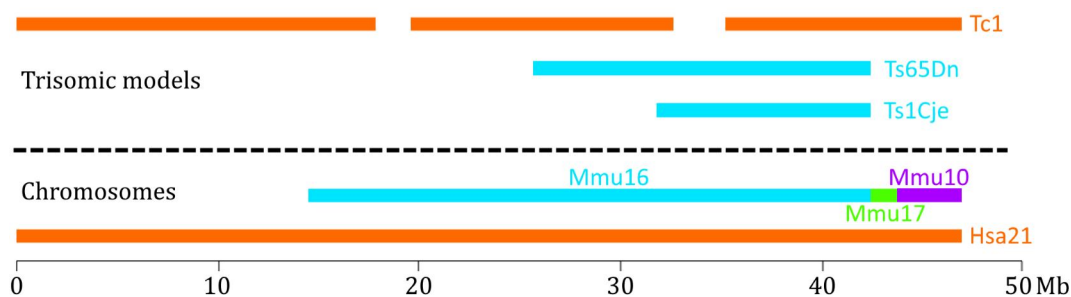
By attempting to replicate trisomy of Hsa21 genes, and hence the features of the human condition, mouse models assist investigation of the syndrome's underlying genetic causes. This is essential for the development of targeted therapeutics (*Wiseman et al., 2009*). An

---

29: The “combined test” offered between 10-14 weeks.

initial target was the “critical region” of Hsa21, isolated by comparing overlapping chromosomal regions of some patients with several DS phenotypes, but only partial Hsa21 trisomy. However, [Olson et al. \(2004\)](#) showed that this sub-region was not sufficient to cause craniofacial dysmorphism in mice.

The mouse has 20 chromosomes (humans have 23), and orthologous genes of Hsa21 are spread across mouse chromosomes Mmu10, 16 and 17. Orthologous genes may also have different functions between species ([Reeves, 2006](#)). Different models take different approaches to replicating human aneuploidy ([Fig 2.17](#)).



**Figure 2.17** Some common trisomic mouse models and their syntenic regions with Hsa21.

The Tc1 model is trisomic for more of Hsa21 genes than other popular models, Ts65Dn and Ts1Cj2. It has two deletions, shown as gaps in the orange bar. Adapted from [Wiseman et al. \(2009\)](#). The scale is in million base pairs (length).

The most popular model, Ts65Dn, carries orthologs of around 50% of Hsa21 genes by replicating part of Mmu16. The Ts1Cje model carries around 67% of these. [Hernandez et al. \(1999\)](#) described the creation of a new model, Tc1, via mouse embryonic stem cells and irradiation microcell-mediated chromosome transfer. Transchromosomal cells were injected into mouse blastocysts, creating chimeric mice with Has21 expressed in a subset of cells. The Tc1 mouse is thus transchromosomal, carrying a copy of 92% of Hsa21 ([Olson et al., 2004](#); [Reeves, 2006](#)). It is therefore thought to be an improved model of DS: initial studies have confirmed that it replicates a broad range of phenotypes ([O’Doherty et al., 2005](#); [Reeves, 2006](#)). It is discussed and analysed more fully in [§4.1](#).

The Tc1 mouse is not, however, functionally trisomic for APP. *Sheppard et al. (2012)* showed that older Tc1s nevertheless exhibited elevated levels of phosphorylated tau, without NFTs. Cross-breeding the Tc1 mouse with the J20 mouse allowed investigation of the DS phenotype in the presence of mutant APP. It is suspected that other Hsa21 genes also have a modulating effect (*Wiseman et al., 2015*). The brain morphometry of this Tc1xJ20 mouse is investigated in ([§4.2](#)).



### 3 Development of a high-throughput software pipeline for phenotyping preclinical subjects

This chapter describes, in detail, the complete set of image processing steps I developed, which are necessary for automated morphometric analysis, following MR data from scanner to statistical parametric map. Every dataset described in this thesis was processed with a version of this pipeline.

Throughout this chapter, datasets which are introduced in more detail later provide examples to illustrate the pipeline steps themselves, as well as the challenges encountered for making each step robust. The multi-atlas masking step is based upon work by Da Ma at CMIC, and published in *Ma et al. (2014)*. Animal preparation and MR data acquisition was performed by others, using parameters laid out in §4.1.2.

The datasets included in figures are of a mouse model of Down syndrome, Tc1 (§4.1); a model of DS cross-bred with an Alzheimer's disease APP model, Tc1xJ20 (§4.2); a mouse embryo model of spina bifida, GLDC, (scanned by Francesca Norris at CABI and originating from Professor Nicholas Greene of the UCL Institute of Child Health); and a mouse model of sciatic nerve injury, created and scanned by Yichao Yu.

## 3.1 Objectives and motivation

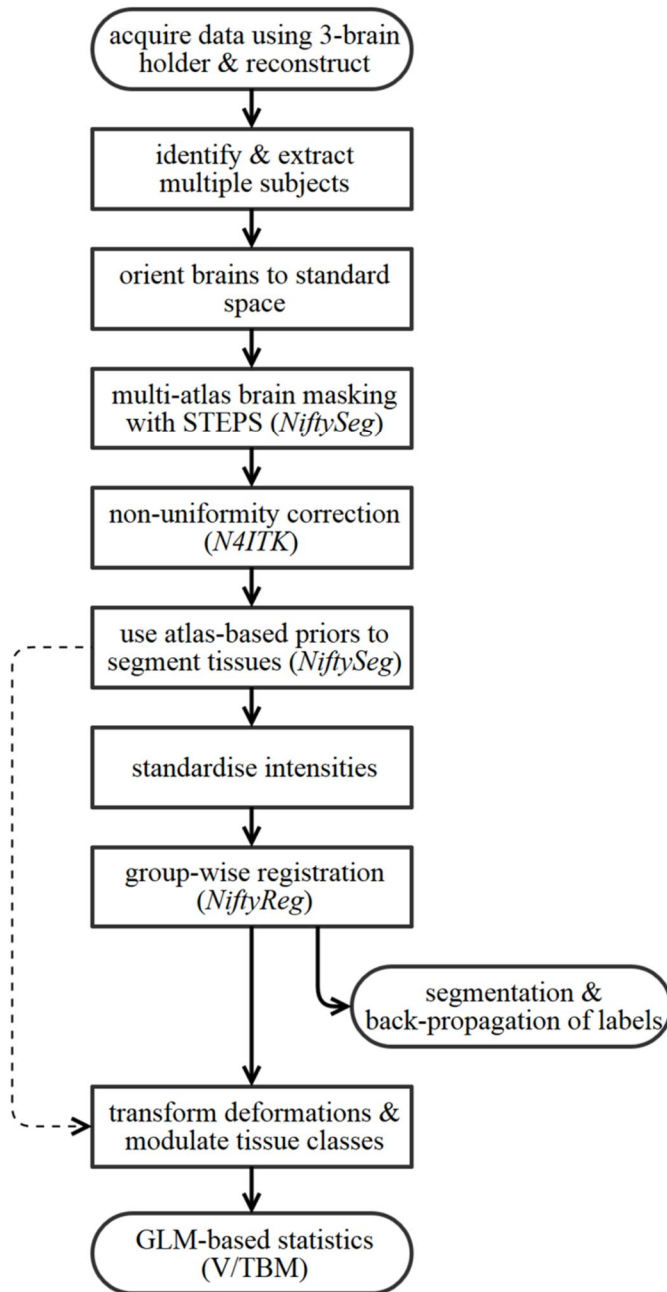
I describe the creation of a fully automatic software pipeline, named *MouseMorph*, for phenotyping large cohorts of mouse brains and embryos with Voxel- and Tensor-Based Morphometry (V/TBM). These powerful, non-destructive, automated investigative techniques can complement other analyses, allowing whole-organ (or organism), “hypothesis-free” investigations which highlight regions worthy of more time-consuming investigation, such as histology. The motivations for such a pipeline include:

1. High field-strength MR scanners, such as the 9.4T Agilent/Varian instrument at the UCL Centre for Advanced Biomedical Imaging (CABI), allow high resolution (40 $\mu$ m) *ex vivo* structural investigations, attractive to investigators with newly-developed transgenic mice and models of disease, behaviour or learning, who wish to interrogate tissue morphology. Taking advantage of this equipment, optimised  $\mu$ MRI protocols for exceptionally high contrast and isotropic resolution in mouse brains (both *in* and *ex vivo*) and embryos have recently been developed by Jon Cleary, Francesca Norris, Holly Holmes and Yichao Yu (§2.1). A large number of MRI datasets of transgenic and disease model mice had already been acquired at CABI, and more studies were planned.
2. V/TBM studies require significant image processing expertise, manual intervention, and parameter changes to existing, clinically-focused tools. No public software exists to seamlessly generate morphometric results from reconstructed preclinical MRI scans automatically; a large number of separate, time-consuming processing steps are required, some hitherto performed manually. These steps can also benefit – or are necessary to initialise – other automated analysis techniques, including segmentation propagation, atlas parcellation, cortical thickness estimation, and longitudinal atrophy measurement with the Boundary Shift Integral, which forms the final experimental §6. Such an automated pipeline would be useful to CABI and the wider phenotyping community, and would significantly contribute to the high-throughput phenotyping efforts laid out in §1.2. Many of the tools required are actively developed, for clinical datasets, at the Centre for Medical Image Computing (CMIC).

3. Human-centric toolboxes exist for morphometry (e.g. SPM, FSL) but are inappropriate for use with preclinical images: many parameters must be adjusted; assumptions accounted for, and several steps require manual intervention to correct non-human data. Several tasks unique to the high-throughput preclinical paradigm are not addressed by such software (e.g. multiple subject extraction; arbitrary orientation correction). Some existing customisations to clinically-focused tools, such as Rodent BET ([Wood et al., 2013](#)) and SPMMouse ([Sawiak et al., 2009a](#)), address only parts of the overall pipeline.
4. The remit was to develop a set of automated methods which completely mitigate manual intervention between a cohort of reconstructed structural MR images of mouse brains (*in* or *ex vivo*) or embryos, and a statistical parametric map (SPM) summarising the local morphological differences between groups, which would inform follow-up histology and strengthen high-throughput and large-cohort phenotyping and disease model studies.

I describe below a new open-source, customised image processing pipeline for high-throughput, large-cohort mouse brain (or embryo) phenotyping studies, designed to completely automate the steps between scanner and statistical parametric map ([Fig 3.1](#)). These include extraction from multiple-subject images; alignment from arbitrary orientations to a common, standard space; skull-stripping; tissue classification with accurate, subject-specific *a priori* tissue probability maps (TPMs); and intensity standardisation. I show how reference atlases may be used to automate several stages, and focus on the steps prior to statistical tests, which may be performed with existing packages such as SPM.

This chapter is predominately methodological: V/TBM results and analysis from both the Tc1 and Tc1xJ20 brains are included in [§4](#), where the MR acquisition parameters and the datasets themselves are also introduced ([§4.1](#), [4.2](#)). Some customisations necessary to process *ex vivo* embryos are included here. [§5](#) includes the customisations necessary for *in vivo* mouse brains. Images from the GLDC and nerve injury datasets are used to illustrate different stages of the pipeline, as well as some of the problems encountered in each paradigm, and how they were overcome.



**Figure 3.1** Overview of the modular pipeline steps.

After GWR, tissue segmentations are resampled into the space of the final average image, such that equivalent regions overlap. The segmentation and back-propagation of parcellation labels is an optional step for quantifying substructure volumes.

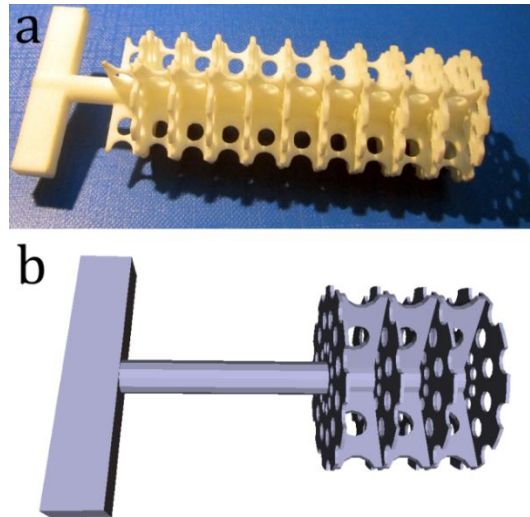


## 3.2 Automated extraction of individual subjects from multiple-subject images

### 3.2.1 Multiple subject scanning

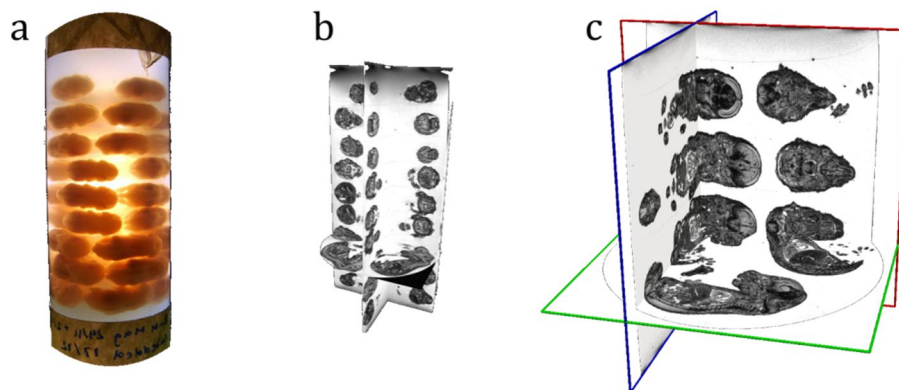
To achieve high throughput, [Bock et al. \(2005\)](#) described scanning up to 16 mice simultaneously *in* or *ex vivo* using a 7T clinical bore and a custom-built array of separate birdcage receiver coils, producing discrete, high-SNR images. However, the high field-strength clinical scanners required to produce high-resolution images are uncommon, and the large number of intricately-arranged coils may be prohibitive ([McConville et al., 2005](#)). The majority of morphometric studies implementing multiple subject scans include three brains per scan ([Table 2.1](#)).

Prior to 2012, *ex vivo* mouse brain images were acquired individually at CABI. To improve throughput, one dataset was acquired by suspending three brains, in-skull, in Fomblin inside a syringe and scanning overnight. The advent of a 3D printed 3-brain holder, designed by Yichao Yu, improved this overnight scanning protocol, as related below. Francesca Norris and Simon Richardson also designed and 3D printed a mouse embryo holder ([Fig 3.2a](#)) capable of holding up to 40 embryos. These plastic holders enabled “high-throughput” acquisitions, using a preclinical scanner and a single coil (parameters in [§4.1.2](#)) – a more prevalent setup necessitating subject separation into individual images before processing. This is a problem unique to preclinical imaging: humans and large animals are scanned individually.



**Figure 3.2** Mouse embryo holders.

(a) Original design by Francesca Norris and Simon Richardson: 3D printed in plastic (10 layers; 4 chambers per layer); (b) 3D CAD image of proposed holder design for future embryo studies: 3 layers only, to avoid signal drop-off and distortion away from the bore isocentre. 3 subjects per layer to duplicate the throughput of existing embryo phenotyping scans (*Fig 3.3c*). The numerous holes allow the suspension medium (e.g. Fomblin, or agarose) to move freely between compartments, and air bubbles to escape. The long neck allows precise positioning within the bore.



**Figure 3.3** High-throughput embryo images.

(a) Photo of multiple mouse embryos embedded in agarose, inside a falcon tube. (b) High throughput  $\mu$ MRI of up to 40 mouse embryos. The bright background signal is from the agarose. Note also the signal drop-off at the top of the image. (c) Moderate throughput (9 embryos). Photo and scans by Francesca Norris. Neither embryo holder was used to acquire these images.

### 3.2.2 Challenges for automated identification and extraction

The chief complications facing automated extraction are (1) touching subjects, (2) poor signal from parts of a subject leading to their exclusion, and (3) inclusion of unwanted signal or material, such as image artefacts, noise, bubbles, debris, tube features and identifying markers. If the first occurs, intricate image-specific erosion/dilation, or complex region-growing operations are required, involving a user-determined threshold, number of operations, seed locations, or manual masking. *Baghdadi et al. (2011)* adapted a deformable model algorithm to segment touching embryos, balancing region-growing and collision forces to prevent overlap. This was sensitive to seed locations, required training data, and sometimes excluded small features.

Isolation via thresholding alone, for example with an Otsu threshold (*Otsu, 1979*) is insufficient to identify subjects (*Baghdadi et al., 2011*): frequently, unwanted material or signals from image artefacts survive. Strong thresholding, conversely, may discard low-signal tissues, or create concavities where vessels or fissures extend into the subject.

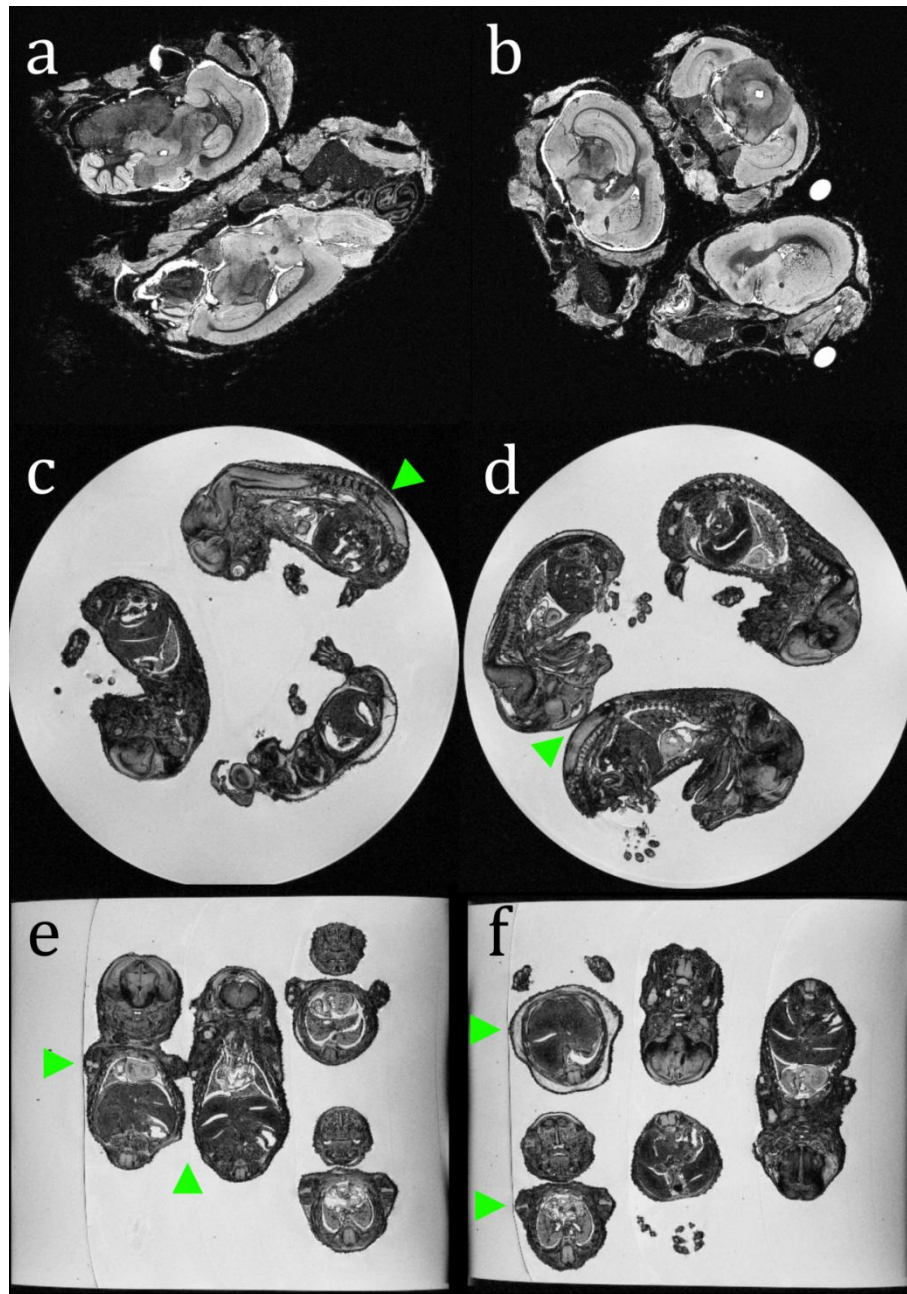
Image non-uniformity – arising from tissue susceptibility or signal drop-off away from the isocentre – may induce low signal at some locations. Significant displacement from the isocentre may also cause distortion (*O’Callaghan et al., 2014*). It is therefore prudent to perform some non-uniformity correction (NUC) prior to extraction in large-cohort scans, wherein subjects may be spread further from the isocentre. The large 40-embryo holder (*Fig 3.2a*) was not used for later phenotyping scans, as these artefacts were deemed too great. One brain and several embryo datasets were thus acquired at CABI without separators (*Fig 3.4*). I investigated solutions to the problems described above.

Without a holder, up to nine embryos were embedded in high-signal, viscous agarose, which set around them, precluding scanner-induced vibration or other movement. Unfortunately, without the subject separator, embryos were prone to sink or otherwise move slightly within the agarose before it set fully, so were not completely prevented from touching. Additional difficulties were lent by the process of embedding itself. First, a layer of agarose was poured into the syringe, which was allowed to set. A layer of 3 embryos was arranged on the surface, and more agarose poured in to surround them. This was

allowed to set before the next layer was added, and so on until 3 layers of embryos were set in the syringe. This process of intermittent agarose setting created distinct layer artefacts: thin, low-signal sheets in the reconstructed scans ([Fig 3.4e, f](#)). These sheets frequently touched – and hence joined – multiple subjects (also lower signal than the agarose). It was not possible, for example, to robustly use binary erosion/dilation operations in sequence to remove these artefacts or alleviate touching subjects, as narrow embryo features would not survive, and the number of operations to perform would depend upon image resolution.

With differing fixative and background intensities, an additional step to extract the high-signal agarose cylinder was necessary. I developed a convex hull operation, using *QuickHull* ([Barber et al., 1996](#)), to assist this and avoid omitting low-signal embryos in contact with the cylinder walls (and hence adjoining the low-signal background) – this was only a problem where the suspension medium had a different MR signal to the background ([Fig 3.4c](#)).

In several cases, I eventually settled upon manual masking as the most reliable (if non-repeatable and time-consuming) method for extracting these brains and embryos from their images. To mitigate the abovementioned issues, Simon Richardson and I adjusted the embryo holder design to include only three layers ([Fig 3.2b](#)), preventing significant signal drop-off or distortion, but allowing the use of low-signal, non-viscous Fomblin instead of agarose, as per the *ex vivo* brain protocol.

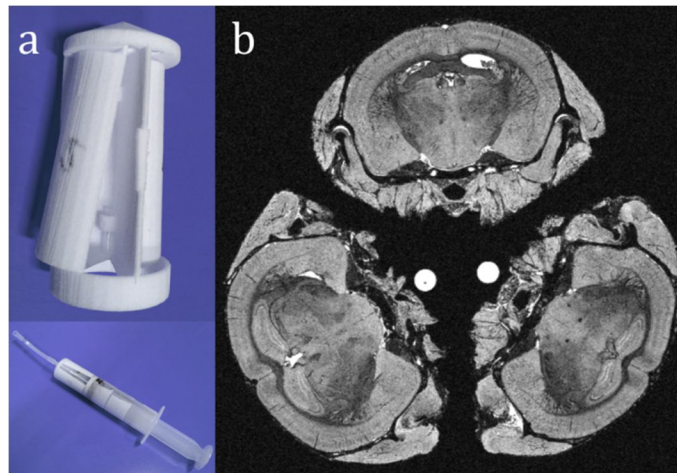


**Figure 3.4** Touching subjects in multi-subject scans without separators.

(a–b) mouse brains, in-skull, scanned in Fomblin without a brain holder; (c–f) slices of reconstructed syringe images from the GLDC embryo study, scanned in agarose. Arrows indicate (c) a subject in contact with the low-signal syringe rim; (d) touching subjects in a single layer; (e) an embryo in contact with dark agarose layer artefact and touching subjects between layers; (f) two otherwise-separated subjects joined by agarose layer artefact, as well as a surrounding layer of adipose tissue, containing high-signal fat (top arrow), which may additionally be confused with the agarose medium when using intensity-based segmentation methods.

### 3.2.3 Solutions for automated identification and extraction

The simplest solution to touching subjects is an adjustment to the scan protocol itself, to include a subject separator. The design, for brain scans, by Yichao Yu and Simon Richardson, secures three skulls inside a 50ml syringe. Its 1mm walls completely preclude touching or partial volume (PV) between neighbouring subjects ([Fig 3.5](#)). This enables unsupervised overnight scans (additional time accounting for the increased FOV), significantly reducing overall preparation time and necessary human intervention, and equals the most common multi-subject throughput of other studies ([Table 2.1](#)). Aligning brains in one layer, and embryos in three layers (along the z-direction) minimises signal drop-off and geometric distortion away from the bore isocentre. An adaptation ([Fig 3.7](#)) allows for six brains simultaneously, without significant distortion. I recommend that plastic separators are used in all future *ex vivo* multi-subject scans, both brain and embryo, together with a low-signal medium.



**Figure 3.5** 3D-printed brain separator and resulting MR image.

(a) 3-brain holder and syringe (design and photo by Yichao Yu); (b) three *ex vivo* mouse brains (2 Tc1, 1 WT in lower right) as they typically appear, suspended in low-proton Fomblin, in the axial view of a reconstructed  $\mu$ MR image. The plastic holder (25.4mm diameter, 44mm length) produces minimal signal. It may rotate within the syringe, so any orientation is possible. The hyperintense dots are agarose markers of different lengths, for identifying brains.



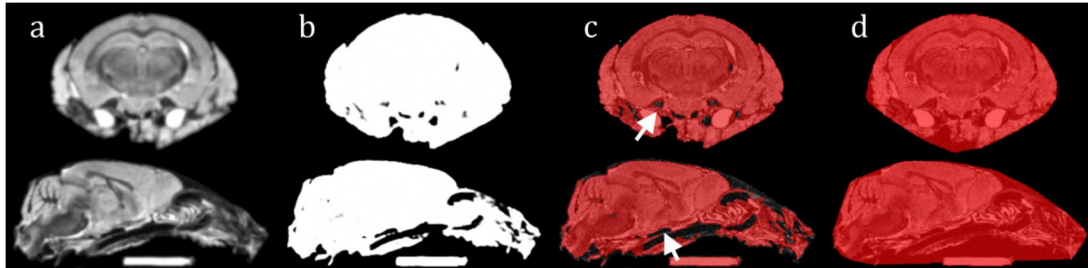
To identify and extract subjects automatically ([Fig 3.6](#)), I first standardised the intensities of all the volumes acquired to the approximate range 0—1 (see [§3.7](#)), smoothed with a Gaussian kernel (FWHM 0.2mm), and used *NiftySeg* to fit histograms with a two-component Gaussian Mixture Model (*GMM*, [Cardoso et al., 2011](#)), for subjects and background, omitting prior spatial information ([Fig 3.6b](#)). I discarded the background and thresholded the resulting probabilistic image at 0.8, producing binary masks of all image objects, which were then distinguished using a connected component algorithm with a 3D, 6-connected neighbourhood. These may include some extraneous objects and, despite standardisation and NUC, may not cover each subject completely ([Fig 3.6c](#)).

To ensure complete coverage, I therefore calculated a convex hull around each distinct binary object ([Barber et al., 1996](#); [Fig 3.6d](#)), and measured their resulting volumes. Together with smoothing, this ensured the inclusion of small external features – a particular problem with paws and the tail in embryo images, as encountered by [Baghdadi et al. \(2011\)](#) – as well as any parts of a subject omitted within concavities in the binary volume – accounting for hollow ventricles and the low-signal fissures sometimes present in fixed, *ex vivo* brains. Skull-stripping occurs later; the precision of this mask and the inclusion of superfluous features (such as parts of other embryos or brains) are inconsequential, provided the entire subject is covered.

Given the total number of study subjects  $N$  (across all images, known *a priori*), it is possible to robustly distinguish subjects from small fragments and large extraneous objects by choosing the  $N$  objects with the closest volumes to a set of training masks. These were produced, for brain images, by thresholding single-subject images (including skull), and creating a binary convex hull around the largest 6-connected component. The resulting mean mask volume from several such images provided an initialisation for the expected subject volumes. I labelled these  $N$  binary objects and used them to crop the original image.

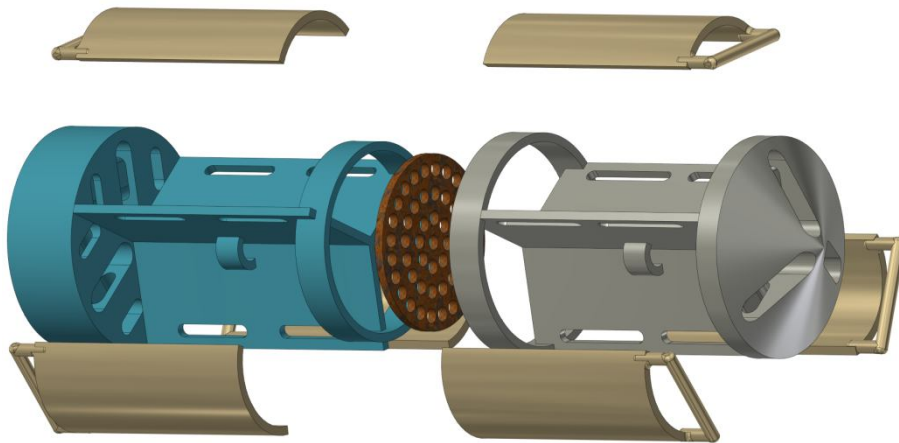
Although this similarity-based method has performed well in my experience, if the study includes a wide range of volumes it may not be robust: the selection method could easily be adapted to include other properties of the masked regions, such as texture, intensity range, ‘mass’ (voxel volume multiplied by intensity), shape, and so on. Atlas input could

also be used to initialise feature-based guesses as to which objects are subjects and which are of no interest. This step was not necessary for single-subject *in vivo* studies (§5, §6), nor some of the earlier *ex vivo* data acquired at CABI (e.g. cohort 1 in §4.1).



**Figure 3.6** Extraction steps in coronal and sagittal views of a downsampled Tc1 brain image.

(a) Smoothed; (b) probabilistic image after GMM fit; (c) binarised mask of all objects in image, including high-signal agarose marker, overlaid on the structural image. Examples of low-signal regions which might intrude into the mask and cause incomplete brain coverage are indicated; (d) convex hull mask.



**Figure 3.7** Exploded CAD view of updated, 6-brain holder

Designed with Simon Richardson and 3D printed for use in future scans. The holes in the middle insert enable the free flow of the scanning medium, and aid bubbles' escape. The pointed end (right) fits inside a 50ml syringe tip; the left end is concave to snugly accommodate a plunger. The gold leaves (shown detached and also present in the design in Fig 3.5a) are hinged so brains can be easily inserted. The leaves themselves hold the brains in place via contact with the syringe wall.



## 3.3 Orientation correction

Atlases are registered to the data to propagate brain masks, tissue priors, and parcellations (§2.2.1). To initialise this, as well as group-wise registration, subject brains must be approximately aligned to a standard orientation – a stereotactic space matching that of the atlas. Clinical software, such as SPM, assumes the human brain is already correctly oriented: typically, subjects lie supine within the bore during image acquisition. The rigid rotations necessary to align brains from the small range of possible rotations this posture affords are easily performed by the registration algorithms of SPM, FSL and *NiftyReg*, after a known rigid rotation orients the brain to standard space.

In this high-throughput scenario, this assumption is no longer safe: brains may be arranged arbitrarily to fit more into the scanner bore (Fig 3.4a, b; 3.5b), or move after placement. The registration algorithms of clinical software such as SPM and *NiftyReg* are unable to resolve the resulting large ( $>45^\circ$ ) rotations necessary for alignment. Registration between two subjects severely out of alignment is likely to converge on an incorrect local maximum similarity and fail. This was also the case when I performed tests with FSL FLIRT<sup>1</sup> (Jenkinson & Smith, 2001), which was designed to resolve large misalignments between human brain images. This problem is unique to the preclinical paradigm, with the exception perhaps of human foetal MRI, in which the head may be in any orientation within the womb (Kainz et al., 2014).

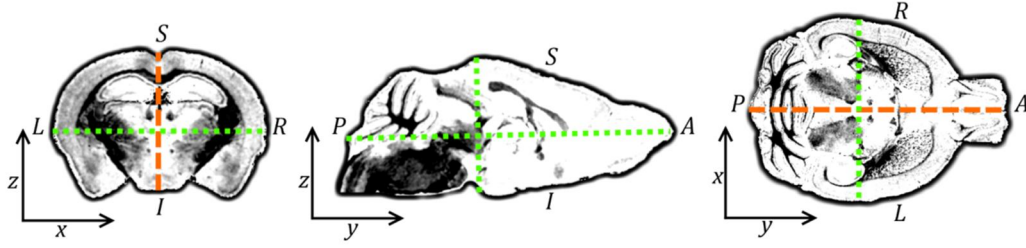
### 3.3.1 Initialisation: principal axes

To perform orientation correction automatically, I exploited the inherent 3D symmetric properties of mouse brains, about which some assumptions may be made, and referred to a set of structural images already in the desired orientation. I first assumed that the orthogonal principal axes of mouse brain structural images correspond approximately to the subjects' anatomical axes (Alpert et al., 1990; Fig 3.8): the antero-posterior (AP),

---

1: FMRIB's Linear Image Registration Tool

right-left (RL), and inferior-superior (IS) axes, and that  $AP \gg \{RL, IS\}$  (a safe first approximation both in- and ex-skull, though the relative lengths of  $\{RL, IS\}$  may be interchangeable). Extraneous material – such as skull tissue and external markers (Figs 3.5b; 3.6d) – may confound this assumption. Assuming this ellipsoidal shape, the image principal axes are the eigenvectors of its inertia matrix (E3.4; Arata et al., 1995; Marion & Thornton, 2004).



**Figure 3.8** Illustration of a mouse brain in RAS orientation

From left: coronal, sagittal, transverse views. Subject +Right Anterior Superior (RAS) parallel to image +x y z axes respectively; a “right-handed” orientation common to most human atlases (e.g. ICBM152: Ashburner & Friston, 2007; Evans et al., 2012) and to Waxholm space (Johnson et al., 2010). Approximate principal axes (RL, AP, SI) are shown, dotted green and dashed orange. The latter denotes the mid-sagittal plane, which coincides with the YZ plane of symmetry.

The intensity-weighted centroid of the image is located at  $C_{x,y,z}$ , where:

$$\text{E3.1} \quad C_x = \frac{\sum (x_i \times S_i)}{\sum S_i}; C_y = \frac{\sum (y_i \times S_i)}{\sum S_i}; C_z = \frac{\sum (z_i \times S_i)}{\sum S_i}$$

where voxel  $i$  is at position  $(x_i, y_i, z_i)$  and has intensity  $S_i$ , and  $\sum$  is over all  $i$ . The intensity-weighted moments of inertia (Arata et al., 1995) are then:

$$\begin{aligned} \text{E3.2} \quad I_{xx} &= \sum (S_i \times \{(y_i - C_y)^2 + (z_i - C_z)^2\}) \\ I_{yy} &= \sum (S_i \times \{(z_i - C_z)^2 + (x_i - C_x)^2\}) \\ I_{zz} &= \sum (S_i \times \{(x_i - C_x)^2 + (y_i - C_y)^2\}) \end{aligned}$$

Similarly, the products of inertia are:

$$\begin{aligned}
 I_{xy} &= \sum \{S_i \times (x_i - C_x) \times (y_i - C_y)\} \\
 I_{xz} &= \sum \{S_i \times (x_i - C_x) \times (z_i - C_z)\} \\
 I_{yz} &= \sum \{S_i \times (y_i - C_y) \times (z_i - C_z)\}
 \end{aligned}
 \tag{E3.3}$$

and the inertia matrix (Arata et al., 1995, Marion & Thornton, 2004) is thus:

$$I = \begin{pmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{pmatrix}
 \tag{E3.4}$$

The eigenvalues and eigenvectors of  $I$  may then be calculated. The eigenvector  $V$  with the smallest corresponding eigenvalue represents the greatest principal axis, AP (although large FOVs which included more external material, usually *in vivo*, sometimes confounded this assumption). In RAS space, we wish AP to align with the image  $y$  axis. I assumed that the two remaining eigenvectors correspond to the RL and IS axes, which we wish to align with the  $x$  and  $z$  axes respectively. If these eigenvectors are then arranged into a homogeneous rotation matrix:

$$H = \begin{pmatrix} V_{RL1} & V_{AP1} & V_{IS1} & 0 \\ V_{RL2} & V_{AP2} & V_{IS2} & 0 \\ V_{RL3} & V_{AP3} & V_{IS3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
 \tag{E3.5}$$

such that the first column corresponds to RL, the second to AP and the third to IS, the resulting affine matrix  $H$ , when applied to a NIfTI image header, will rotate it such that the AP axis aligns with the image  $y$  axis, the RL with the  $x$  axis, and the IS with the  $z$  axis. To rotate about an arbitrary point, such as the image centroid, we must translate to the origin, rotate and then apply the inverse translation:

$$M = T^{-1}(HT)
 \tag{E3.6}$$

where the translation matrix  $T$  translates the centroid coordinates to the origin. I thereby used *NiftyReg* to rotate each subject to align AP with the  $y$  axis (Fig 3.10a).

### 3.3.2 Using symmetry to test rotation

As mammals, and members of *bilateria*, the brains of both humans and mice exhibit approximate mid-sagittal plane symmetry (the YZ plane in RAS orientation). I used this feature to correct possible misalignment (as, e.g., [Liu et al., 1998](#); [Ruppert et al., 2011](#) did, with human brains). To measure symmetry, I reflected images in the YZ plane (along the x-axis, [E3.8](#)) and calculated the sample Pearson's product-moment correlation coefficient  $r$  between original and reflection, [E3.9-11](#) ([Kim & Fessler, 2004](#)). By composing with additional rotation matrices  $M_{rot}$ , I generated multiple rolls about the y axis within a  $180^\circ$  range, reflected each, and searched for the optimal rotation which maximised  $r$  and thus best aligned LR and IS with the x and z axes respectively ([Ruppert et al., 2011](#)) ([Fig 3.9](#); [Fig 3.10c](#)).

To test symmetry, one can insert an additional rotation matrix  $R$  into [E3.6](#), where  $R$  is found by adapting from [Arata et al. \(1995\)](#) for 3D homogeneous coordinates:

$$\begin{aligned}
 M_{rotate} &= T^{-1}HRT \\
 R &= R_z R_y R_x \\
 R_x &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 R_y &= \begin{pmatrix} \cos \psi & 0 & -\sin \psi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \psi & 0 & \cos \psi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 R_z &= \begin{pmatrix} \cos \varphi & \sin \varphi & 0 & 0 \\ -\sin \varphi & \cos \varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
 \end{aligned}$$

The matrix  $R_y$  rotates the image  $\psi$  radians about the y axis. To rotate about the y axis alone,  $\theta = \varphi = 0$ ,  $\psi = d \times \pi/D$ , and I varied  $d$  in integer steps from 0 to  $(D - 1)$ . The number of rolls to test,  $D$ , was determined to balance computational time and registration accuracy. As *NiftyReg* may reliably recover rotations up to  $28^\circ$  ([Ourselin et al., 2001](#)), I tested in steps of  $24^\circ/2 = 12^\circ$  (so that  $D = 180^\circ/12^\circ = 15$ :  $D$  must be an

integer). In the best case (*Fig 3.10di*), the reflection will therefore be at most  $12^\circ \times 2 = 24^\circ$  from the original rotated image.

The affine transformation matrix for reflection,  $F_{YZ}$  (along the x axis), is composed with the rotation in the following order:

$$M_{rotate\ and\ reflect} = T^{-1} H R F_{YZ} T$$

$$F_{YZ} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

I found Pearson's  $r$  between each of the  $D$  image pairs (*Kim et al., 2004; Rodgers & Nicewander, 1988*) using *E3.9*, where  $I_O$  is the original image and  $I_R$  the reflected; and the sample covariance  $Cov$  is given by *E3.10*:

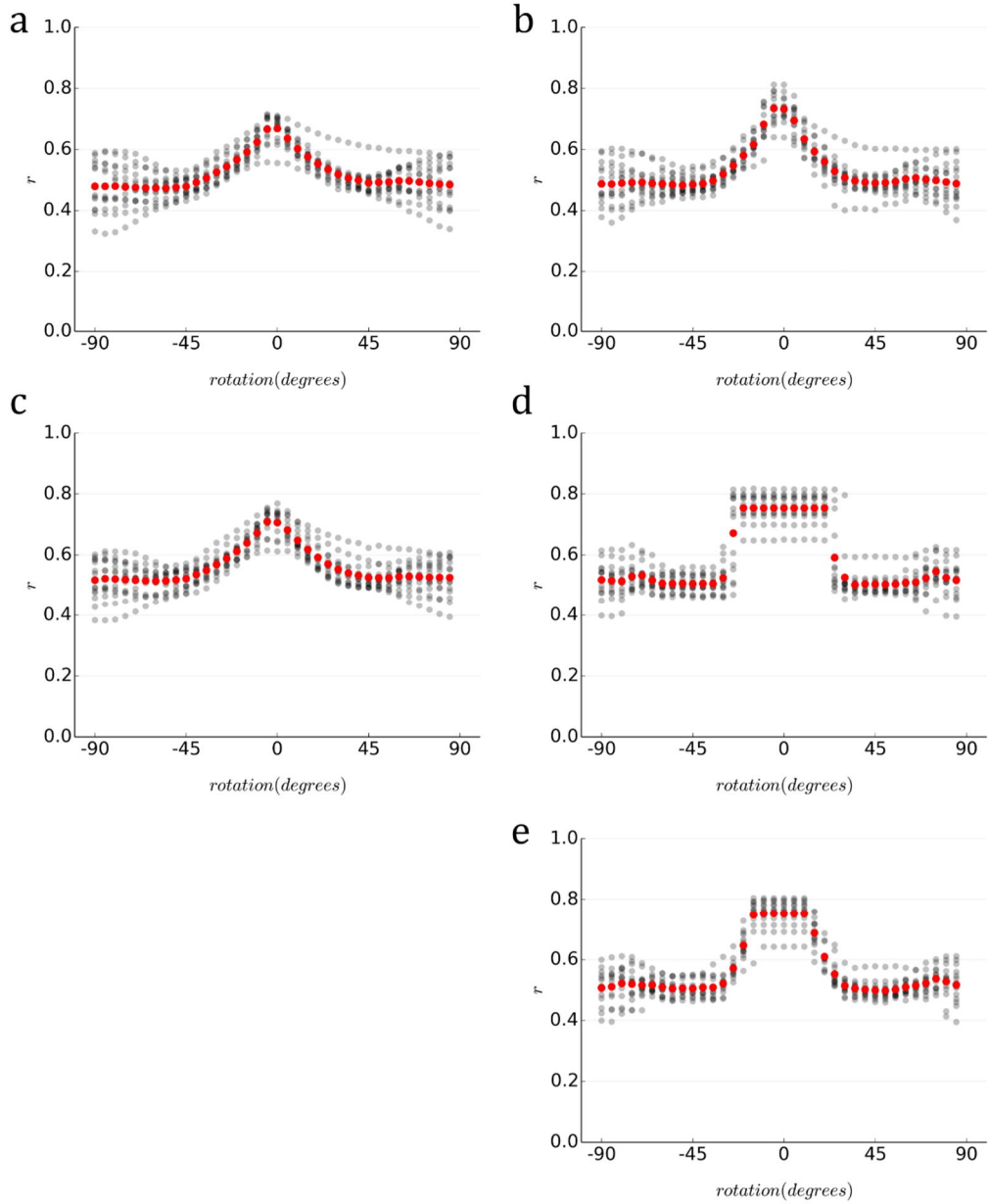
$$r_{sample} = \frac{Cov(I_R, I_O)}{\sqrt{\sigma_{I_R}^2 \sigma_{I_O}^2}}$$

$$Cov(I_R, I_O) = \frac{1}{N-1} \sum_{i=1}^N (I_{Ri} - \bar{I}_R)(I_{Oi} - \bar{I}_O)$$

$\bar{I}_R$  denotes the mean intensity value of all  $N$  voxels  $i$  in the reflected image, and the sample variances are given by:

$$\sigma_{I_R}^2 = \frac{1}{N-1} \sum_{i=1}^N (I_{Ri} - \bar{I}_R)^2$$

$$\sigma_{I_O}^2 = \frac{1}{N-1} \sum_{i=1}^N (I_{Oi} - \bar{I}_O)^2$$



**Figure 3.9** Determining optimal orientation via reflection.

I resampled 14 in-skull WT brain images into RAS space, so that the correct orientation was known. Rolls were applied about the Y axis ([Fig 3.10b](#)), in 5° increments from -90° to +85°. Each was reflected in the YZ plane and the reflection compared with the rotated original using Pearson's  $r$  (values in grey; mean in red). First/second column: comparison without/with prior 6-DOF rigid registration. First row: full-size images; second row: images downsampled by a factor of 4; (e): downsampled by a factor of 2.

The peaks at 0° illustrate the ability of this technique ([Fig 3.10b—d](#)) to correctly orient mouse brains using their symmetry, with robustness to extraneous material. Registration resulted in a plateau (b,d,e), which broadened with the degree of downsampling.

[Fig 3.9](#) shows the results  $r$  in tests with 14 *ex vivo* WT brains rotated at 5° increments from a ground truth correct orientation, demonstrating that this rotation, reflection and symmetry test technique was robust for many different brains. The plateaus demonstrate the ability of *NiftyReg* to correctly register reflected brains within a certain rotation range. However, this figure also demonstrates that it was better not to apply rigid registration after reflection, before calculating  $r$ , as this resulted in a broad plateau about ‘0’ degrees from the ground truth, and hence made the best alignment more ambiguous.

I corrected any remaining misalignments (due to performing a limited number of rotations and tests) by rigidly registering the image pair with the maximum  $r$ , giving an initial affine transformation matrix,  $A$ . Applying the (log Euclidean) half of  $A$ ,  $A_{half}$ , to the original, rotated image thus correctly aligned the anatomical axes to RAS ([Fig 3.8](#), [3.10d](#)).

E3.12

$$A_{half} = \expm\left(\frac{\logm(A)}{2}\right)$$

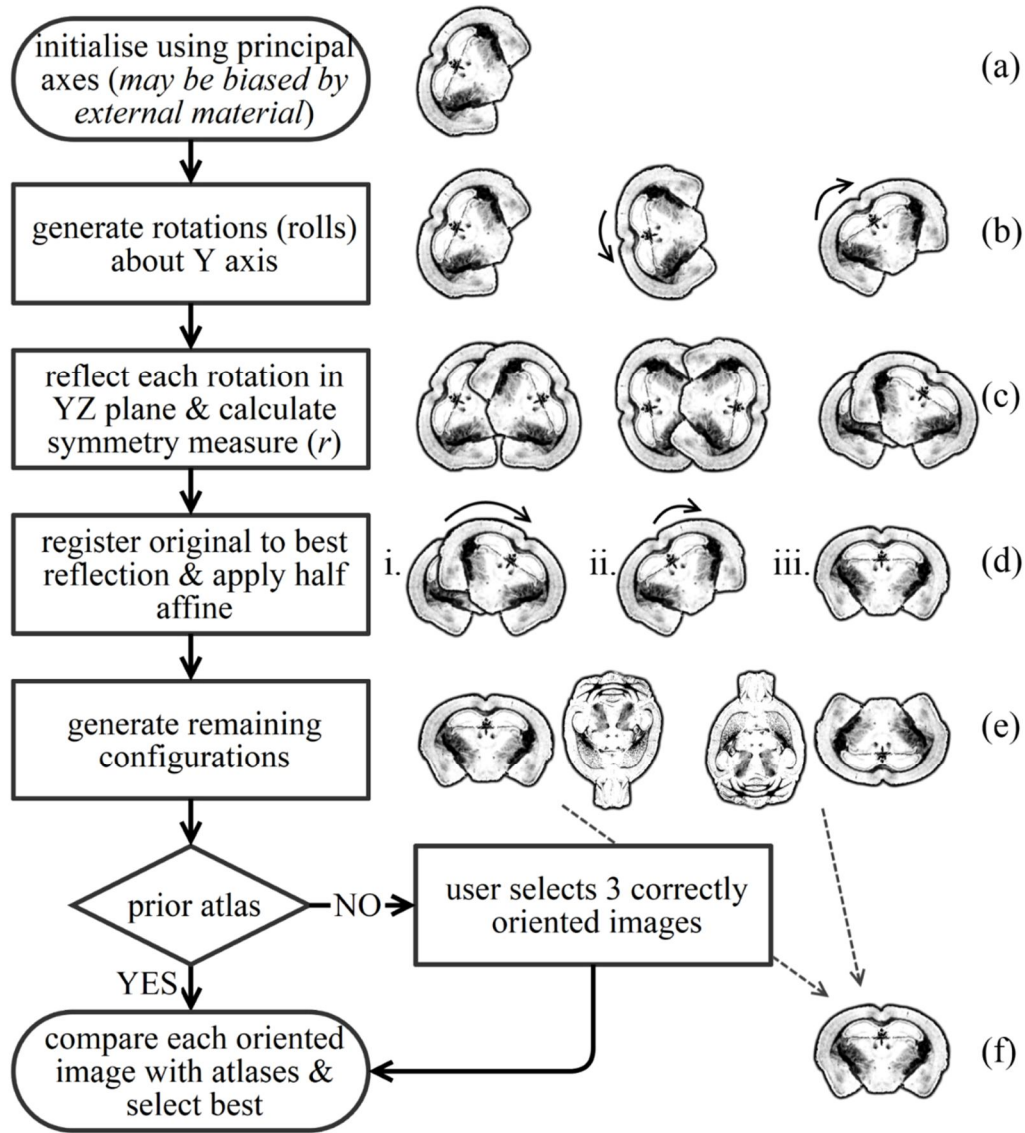
Where  $\logm$  and  $\expm$  denote, respectively, the matrix logarithm and the matrix exponential.

### 3.3.3 Final choice of orientations

As the principal axis calculation omits anatomical direction, AP is equivalent to PA, LR to RL and IS to SI. After correctly aligning principal axes, therefore, the subject could still be in one of four possible orientations, one of which conforms to RAS. I generated the three remaining candidates by composing with additional rotations: a 180° roll about y, inverting IS and RL; a 180° yaw about z, inverting AP and RL; and a 180° pitch about x, inverting IS and AP ([Fig 3.10e](#)). The  $M_{rot}$  and  $A_{half}$  which align one of these four will also align the others.

I rigidly registered each candidate to 3 correctly aligned images from the UFL *ex vivo* mouse brain atlas ([Ma et al., 2005](#)), and selected the candidate with the highest  $r$  in the most cases as the correct orientation ([Fig 3.10f](#)). A subset of correctly oriented candidates

could be manually chosen instead of external data. In practice, it was almost always possible to forgo this voting step and only test against one correctly oriented atlas image.



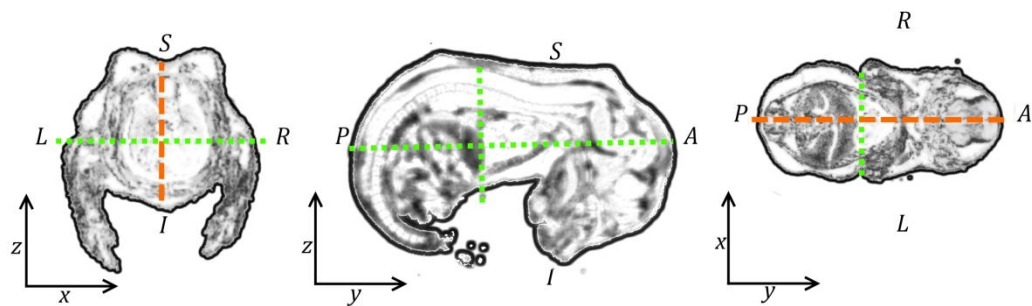
**Figure 3.10** Overview of orientation correction steps, with coronal views.

See [Fig 3.9](#) for validation of the symmetry measure technique (c). In the configurations illustrated (e), 4 orientations are possible: in the coronal views, the mouse is either facing the viewer or away (L and R interchangeable); in the transverse views, the view is from below or above.



### 3.3.4 Orientation of embryos

After extraction from the mother's gravid uterus, mouse embryos are typically found in the foetal position. Little extraneous material is included. Their gross morphology is thus, overall, largely symmetric and ellipsoidal (*Fig 3.11*) – with the exception of some internal organs, such as the heart, and the variable positions of their limbs and tail. Exactly the same orientation technique is therefore applicable, employing appropriate 'atlas' or target embryos already in the correct orientation for the final stage.



**Figure 3.11** Illustration of a mouse embryo in RAS orientation

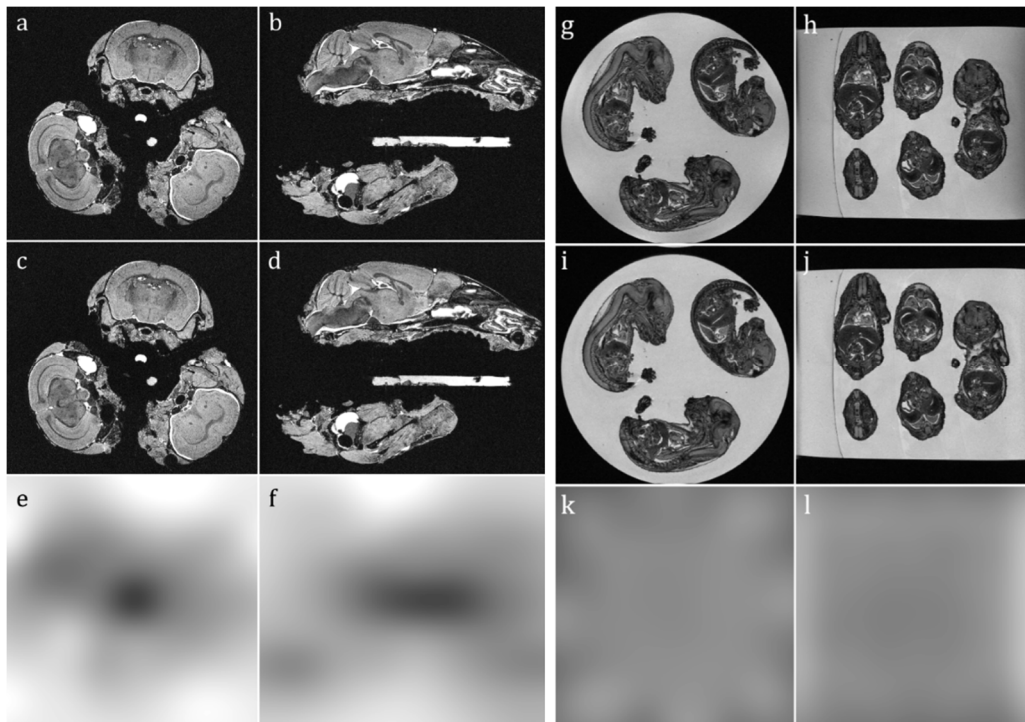
From left: coronal, sagittal, transverse views, as *Fig 3.8*. Approximate principal axes (RL, AP, SI) are shown, dotted green and dashed orange. The latter denotes the mid-sagittal plane, which, again, coincides with the YZ plane of approximate symmetry. In this case, AP corresponds with skull to tail.

## 3.4 Intensity non-uniformity correction

A smoothly-varying, hardware-induced, low-frequency intensity gradient often corrupts MR images (*Sled & Pike, 1998*). This is caused by inhomogeneity of the scanner's  $B_0$  field, variable sensitivity of the receiver coils, or inhomogeneity of the radiofrequency pulse, which can be induced via non-uniformity of the transmission coil or via electromagnetic interactions and distortions by the subjects themselves (*Lewis & Fox, 2004*). As this effect becomes especially pronounced at high field strengths (*Boyes et al., 2008*), intensity non-uniformity modelling is essential in preclinical studies. NUC

improves automated tissue classifications and image registration (*Sled & Pike, 1998; Van Leemput et al., 1999; Lewis & Fox, 2004*). Locally adaptive NUC algorithms, such as N3, have been shown to outperform nonadaptive methods (*Arnold et al., 2001*).

I initially employed N3 (*Sled et al., 1998*), and later the N4ITK (N4) algorithm (*Tustison et al., 2010*), found to reliably correct bias at high field strengths, using 200 iterations; 256 histogram bins; a 0.15mm FWHM Gaussian kernel to model the bias field; subsampling factor 4 (for high-resolution, *ex vivo* images) or 2 (for lower-resolution *in vivo* images). NUC is further refined during the iterative expectation maximisation steps of tissue segmentation. In neither technique is the image resampled; thus image quality is maintained.

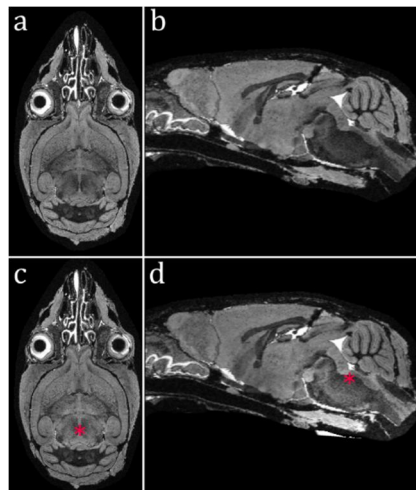


**Figure 3.12** NUC of multiple-subject brain and embryo images prior to extraction.

NUC has the potential to benefit multiple-subject extraction by improving subject identification. a-f: Three sciatic nerve injury study brains in Fomblin; g-l: nine GLDC embryos in agarose. First row: uncorrected original images; second: after NUC; third: calculated, inverted bias field (to illustrate, I enhanced the contrast in the final row by 25%).

Multiple-subject scans (with greater FOV than single-subject) may benefit from preliminary NUC prior to extraction, as signal drop-off away from the isocentre compounds non-uniformity ([Fig 3.3b](#)). In the 9-embryo and 3-brain scans analysed here, non-uniformity was not problematic for identifying subjects prior to extraction – however, if the larger embryo holder ([Fig 3.2a](#)) were to be used, to increase throughput, signal drop-off away from the bore isocentre would be more appreciable.

In *ex vivo* data, N4 occasionally “corrected” the lower WM intensities of the brainstem, particularly around the pons. The degree of GM/WM PV between the thalamus and brainstem yields a similar effect to a smoothly-varying intensity gradient, which may be misinterpreted by the NUC algorithm. Thus, after NUC, it is prudent to visually inspect the data for quality improvements or over-correction ([Fig 3.13](#)). In such *ex vivo* cases, NUC was not applied.



**Figure 3.13** Non-uniformity over-correction in an *ex vivo* brain.

Transverse and sagittal views showing the original data (a,b) and results of NUC using N4 (c,d). The WM of the midbrain and brainstem (red asterisks) is artificially brightened – a result of the misinterpretation of naturally smoothly-varying tissue intensities in the mouse brain.

## 3.5 Brain masking (skull-stripping)

To reduce damage, and hence improve the validity of morphometry results, it is better to scan mouse brains in-skull (*Sawiak et al., 2012*). Automatically distinguishing brain tissue from skull in humans, and localising the brain surface, is a prerequisite for many analysis techniques, and one of the most fundamental computational tasks in neuroimaging, addressed in many clinical papers.

Accurate brain masks benefit tissue classification, intensity standardisation, bias correction, registration and segmentation, by restricting the ROI and excluding variable extraneous material. This saves time and computer memory during registration, and helps constrain the multiple testing problem for voxel-wise statistical tests. Accurate brain surface localisation is important for the detection of cortical atrophy in V/TBM, as well as for measurement of brain volume and cortical thickness (*Smith, 2002; Leung et al., 2011*). As manual masking (e.g. *Cheng et al., 2011*) is extremely time-consuming at high resolutions, susceptible to inter- and intra-rater variability, and lacks repeatability, several automated techniques, most tailored to human brains, have arisen – for example, FSL’s brain extraction tool, BET (*Smith, 2002*); SPM.

FSL’s BET expands a spherical mesh ROI initialised within the human brain from half the estimated radius, until it reaches the surface, determined by an intensity threshold. In SPM, brains are skull-stripped using GM and WM tissue probabilities, requiring *a priori* maps with sufficient subject similarity. This probabilistic approach beneficially accounts for structural variability in the complex configuration of folding cortical gyri and sulci, which vary between subjects as well as with age and neurodegenerative disease (*Cardoso et al., 2009; Chakravarty et al., 2013*).

The mouse cortex is lissencephalic – smooth, without gyri – and less variable than a human’s. The cerebellum, proportionally larger, varies in shape with age and background. The brain overall exhibits greater morphological variation than a rat’s (*Paxinos & Franklin, 2012*). The brainstem also exhibits significant variability, depending on positioning during fixation in *ex vivo* scans or the alignment of the head in the scanner

when *in vivo* (see §5.7.2 and Fig 6.4). As in humans, overall brain shape changes with age. Additional variability arises between mouse strains, and in disease. *Oguz et al. (2014)* followed mathematical morphology operations with an intensity gradient-based graph cut method to constrain the outer surface of a rodent brain mask. This required varying smoothness constraints across the resulting brain surface, and a manually-specified upper bound on the brain volume. The method was fast, but would be sensitive to external material with a similar intensity as the brain, without clear separation, and omitted the paraflocculi and brainstem in an image with severe intensity inhomogeneity.

As mouse brains are much smaller and more ellipsoidal than humans', *Biedermann et al. (2012)* scaled their data  $\times 10$ , and then  $\times 0.5$  in the AP dimension, to fit BET's initial spherical ROI, and reversed this scaling when the optimal segmentation was achieved. Rodent BET – released by Tobias Wood at King's College London – provides modified parameters for rodent brains, including an ellipsoidal brain surface initialisation (*Wood et al., 2013*), but in extracting rat brains, generally omits the olfactory bulbs and paraflocculi, which do not conform to the assumed generic ellipsoid<sup>2</sup>. For investigative morphometry, it is better to include all parts of the brain.

*Li et al. (2009)* employed a 3D active surface model (using ITK-Snap<sup>3</sup>, *Yushkevich et al., 2006*), requiring manual seeding and corrections. Such edge-based region-growing approaches are vulnerable to regions where the brain appears to be in contact with external tissue with similar intensity – a problem in mice, as the skull tightly encases the brain – or non-uniformity and PV at the brain boundary, which obfuscate the exact surface location. *Uberti et al. (2009)* developed a level set method, requiring accurate user delineation of several slices to initialise an evolving 3D function, as well as constraint points placed at extrema to prevent expansion beyond the surface. *Chou et al. (2011)* employed Pulse-Coupled Neural Networks, expanding a 3D mask based upon the intensity values of neighbourhood voxels. Each voxel's inclusion in the mask is determined by the number of times it is 'activated', above a threshold. Voxels beyond

---

2: See image at <http://www.nitrc.org/projects/rbet> (accessed 2015-11-25)

3: <http://www.itksnap.org>

borders represented by steep intensity gradients require more iterations to activate; thus the mask volume plateaus. The result is reliable and includes the olfactory bulbs; however, the user must intervene to stop progression beyond the optimal segmentation, or provide parameters to search within a range for the centre of the plateau. *Janke & Ullmann (2015)* advocated a manual correction step, as “in many cases it is nearly impossible to prevent” the inclusion of extraneous material.

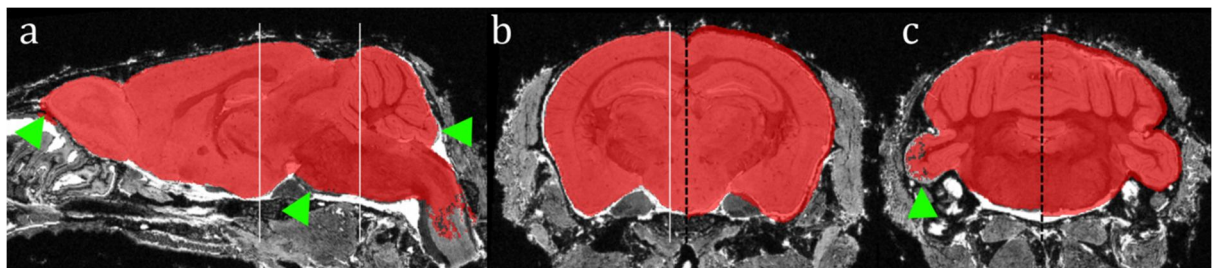
### 3.5.1 Multi-Atlas masking with STEPS

I adopted a multi-atlas technique, developed by Da Ma and Jorge Cardoso at CMIC (§2.2.3; *Ma et al., 2014*), relying upon registration accuracy and the applicability of atlas images, given potential morphological variability.

I created brain masks using Similarity and Truth Estimation for Propagated Segmentations (STEPS, *Cardoso et al., 2013b*) label fusion, implemented in *NiftySeg*, with the UFL atlas database of 10 skull-stripped *ex vivo* T<sub>2</sub>\*W structural images (47µm isotropic resolution; 20 corresponding labels) as a reference. Exploiting the approximate inherent symmetry of the atlas brains, I doubled the database size by reflecting structural and label images in the mid-sagittal plane (see §3.3.2, *Bowden et al., 2011* and *Ma et al., 2014*). I binarised and merged labels to create atlas brain masks  $A_M$ .

I registered each subject image  $S$  to each of the structural atlas images  $A_S$  using affine registration with 12 degrees of freedom (DOF), using *NiftyReg*. The inverse of the calculated affine matrix was used to transform each corresponding  $A_M$  to the space of  $S$ , resampling with nearest neighbour interpolation. These masks were dilated by 5 voxels to account for incomplete brain coverage (resulting from differing morphology of  $S$  and  $A_S$ ) and used to constrain a symmetric non-rigid registration (NRR, *Modat et al., 2012*) between  $S$  and each  $A_S$  (parameters: control point spacing 0.3mm, 4 levels of doubled CP spacing, bending energy penalty 0.1, inverse consistency penalty 0.1). The inverse deformation field was then used to propagate each  $A_M$  and  $A_S$  to the space of  $S$ . The resampled  $A_M$  were then combined.

In STEPS, the subject image  $S$  is compared with each propagated  $A_S$ , using locally normalised cross correlation (LNCC) over a Gaussian kernel. Atlas images are locally ranked based on this similarity metric, and the masks corresponding to the voxel-wise  $X$  highest-ranked atlas images are fused into a consensus mask (Cardoso *et al.*, 2013b). Poorly locally-matched atlases are thus excluded. The user defines  $X$  and the kernel standard deviation  $\sigma$  as algorithmic parameters (here for *ex vivo* data,  $X = 9$ ;  $\sigma = 8$  voxels). Optimised parameters were recently investigated using this atlas (Ma *et al.*, 2014). Instances of minor incomplete coverage (such as at the brainstem and paraflocculi) were all accounted for by dilating and filling the final consensus masks (Fig 3.14).



**Figure 3.14** Mouse brain masks after the STEPS procedure.

Original masks (left of dashed lines in coronal views) overlaid on representative slices (a, sagittal; b—c, coronal) from a single WT brain, and dilated by 5 voxels (right of dashed lines). Solid white lines indicate slice locations. Green markers indicate, from left, over-generous coverage of olfactory bulbs; minor incomplete coverage of the inferior brainstem, posterior cerebellum, and paraflocculi, where atlas registration was more imprecise – all of which are included after dilation.

### 3.5.2 Problems for automated mouse brain masking

Problematic regions for masking in-skull images include the area around the olfactory bulbs; the base of the brainstem; the lower cerebellum just superior to the cisterna cerebellomedullaris, and sometimes the posterior face of the cerebellum (Fig 3.14). In all these locations, the brain is in close contact with external tissues exhibiting similar signal intensities. The cerebellum, furthermore, is one of the regions of greatest morphological



variability in the studies conducted for this thesis; its shape changes with age, and the material surrounding it may be fully or partially excised during skull extraction.

The masks generated are dilated to include the intensity gradient at the brain boundary, informing registration, in which external tissues may thus play a role. A generous mask, advantageously, may later be refined via tissue classification. Indeed, such a two-step brain extraction technique may be preferable in whole-skull mouse MRI, as (for the abovementioned reasons) a parcellation- or region-growing-derived mask may include some external material. If tissue were to be cropped by an over-zealous or conservative mask, it cannot be recovered later, and the intensity gradient introduced at the boundary may bias registration.

## 3.6 Segmentation of brain tissues

I initialised intensity-based segmentation with subject-specific TPMs propagated from an atlas (§2.2.5), and used non-rigid registration to resolve structural differences. Custom, study-specific TPMs are more likely to reflect the anatomical variability of the data than external TPMs, and hence reduce bias toward the healthy morphology of atlases (*White et al., 2003*). I based the TPMs upon the NUS atlas (*Bai et al., 2012*), preferring its greater number of labels – including cerebellar WM – over the UFL atlas, which – for example – does not delineate the GM and WM of the cerebellum. *Lee et al. (2009)* nevertheless used the UFL atlas to initialise an itkEMS-based segmentation, subsequently used for probabilistic skull-stripping. In summary, the process was:

1. Classify anatomical atlas labels based upon their tissue content
2. Register each atlas structural image to each data image
3. Propagate the tissue labels to the data space, and calculate the average at each voxel
4. Use these normalised probabilistic labels as TPMs to initialise segmentation
5. Correct CSF classifications, using voxel intensities.



I manually classified 22 of the 39 NUS atlas parcellations as GM; 9 WM ([Table 3.1](#)), based upon their predominant tissue content. Following [Lee et al. \(2009\)](#), I initially classified 3 regions (caudate putamen/striatum, thalamus, superior and inferior colliculi) as a separate GM/WM mixture class (GWmix). I excluded ventricular CSF (vCSF), as the large *in vivo* atlas ventricles did not register well to *ex vivo* data, whose ventricles in most cases almost completely collapse upon perfusion-fixation (see [§5.7](#)).

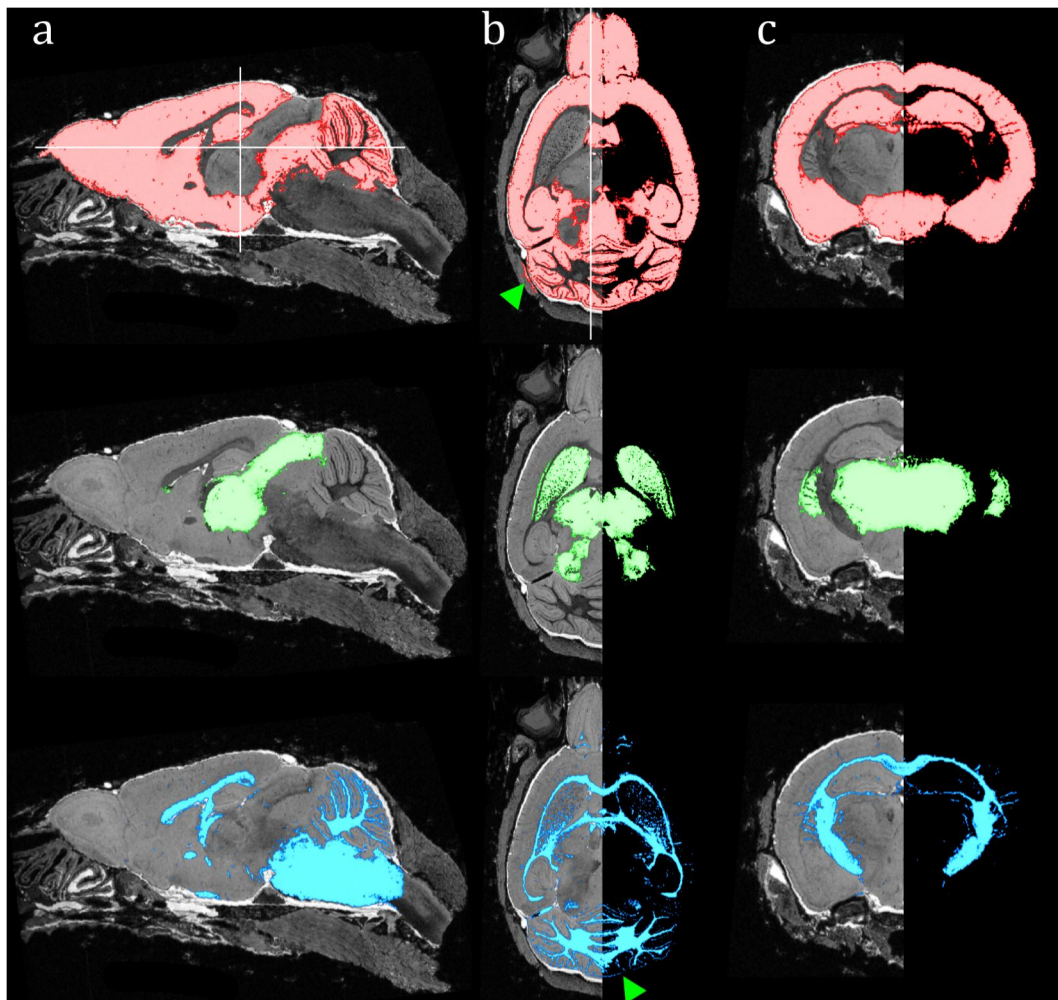
To account for variable locations of external CSF, I added an eCSF label to the outer rim by dilating the atlas masks – excluding the pituitary gland – by 1 voxel, and subtracting the original. I added a BG label in the same manner to account for skull tissue, dilating by 5 voxels and subtracting both the original mask and eCSF. These classes helped to avoid misclassification of the meninges and skull as brain tissue, as well as of GM PV voxels at the cortical surface as WM ([Good et al., 2001](#); [Teipel et al., 2004](#)). I again doubled the atlas size, by reflecting in the mid-sagittal plane ([§3.5](#)).

I registered the 10 structural NUS atlas images to each data image (12 DOF symmetric affine registration, then symmetric NRR to account for local morphological differences). I resampled the tissue labels to the data space following the resulting deformations, using linear interpolation. Average labels were then calculated in the data image space, probabilities were normalised at each voxel (to sum to unity) and the results smoothed (using a Gaussian kernel with a 4-voxel standard deviation), creating subject-specific TPMs of 5 classes {GM, WM, GWmix, eCSF, BG}.

I used these TPMs to initialise an iterative expectation-maximisation (EM) algorithm, implemented in *NiftySeg* ([Van Leemput et al., 1999](#); [Cardoso et al., 2009](#)), assuming approximately normally distributed tissue histograms. This was spatially constrained with a Markov Random Field (prior strength 0.25), giving robustness to noise and preventing the misclassification of brain surface PV voxels as WM, and proceeds until convergence. The priors were iteratively relaxed by a factor of 0.25 and regularised (Gaussian kernel standard deviation 0.5 voxels) to avoid overt bias towards the atlas and account for local anatomical variability ([Cardoso et al., 2013a](#)). The algorithm includes iterative NUC, assuming a smoothly-varying polynomial bias field. The dilated mask generated earlier

constrained classification to the brain and its immediate surrounds, limiting confounding external tissue.

To account for vCSF, which in most images was minimal, within the ventricular regions propagated from the atlas, I classified voxels within one standard deviation of the mean of the eCSF class as vCSF (binary), and subtracted these from other classes. Resulting TPMs for a single Tc1 mouse are shown in [Fig 3.15](#).



**Figure 3.15** Representative slices from a Tc1 brain illustrating probabilistic tissue segmentations.

Column a, sagittal; b, transverse; c, coronal. First row: GM (red); second: GM/WW mixture (green); third: WM (blue). Solid white lines indicate slice locations. Green markers indicate, from top, minor misclassification of external tissue as GM; thin dark GM Purkinje cell layer misclassified as WM.

In subsequent analyses, I replaced these TPMs with maps generated used the same scheme, omitting the GWmix class, which lacked direct biological meaning. The regions previously classified as GWmix ([Table 3.1](#)) were initialised prior to propagation to the data space with a probability of 50% GM and 50% WM. The updated segmentations ([Fig 3.16](#)) include PV as regions of lower GM or WM probability.

I calculated total intracranial volume (TIV), a measure used clinically to account for natural cross-sectional variability of head size ([Whitwell et al., 2001](#); [Ridgway et al., 2011](#)) as:

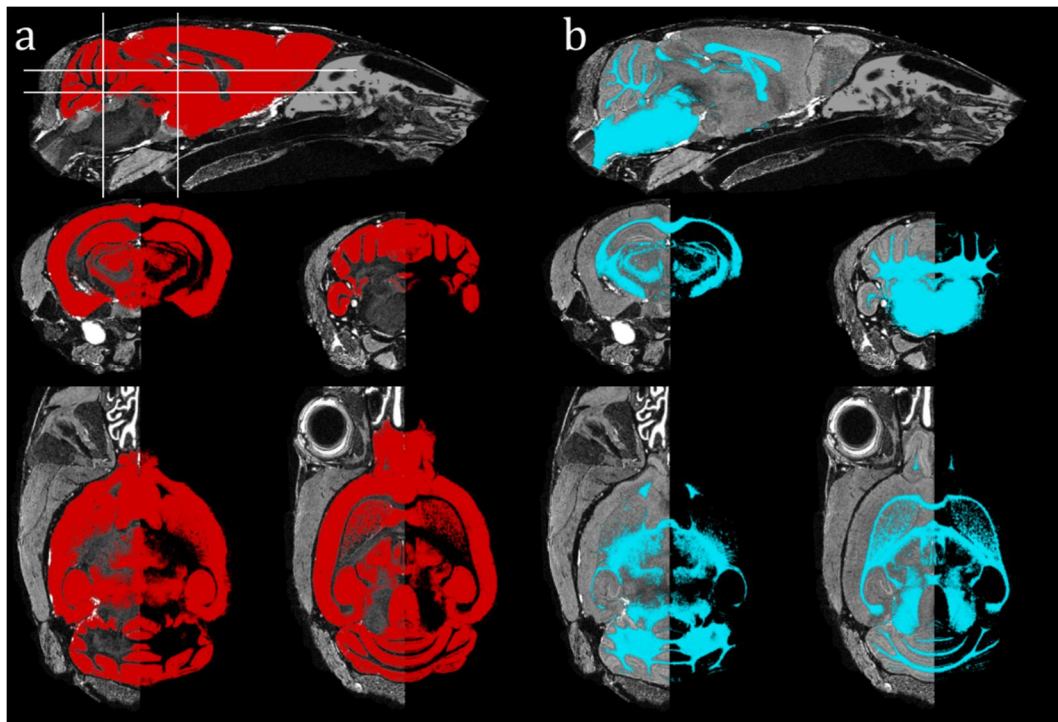
E3.13 
$$TIV = \sum_{all\ voxels} V \times (P_{GM} + P_{WM} + P_{eCSF} + vCSF)$$

Where  $V$  is the voxel volume and  $P_{GM}$  is the segmentation-derived probability of GM within the voxel. TIV may be used to normalise volumetric measurements of brain structures, and brain volume (BV) itself, which is calculated in the same way, excluding CSF. I generated a conservative brain mask by binarising the {GM, WM} probability images above 50%, adding vCSF, and binary-filling the resulting volume to include possible omitted internal regions. These refined masks excluded potentially misclassified external tissues.

CSF	cerebral aqueduct, lateral ventricles, third ventricle, fourth ventricle
GM	amygdala, auditory cortex, cerebellar cortex, cortex general, dentate gyrus, entorhinal cortex, frontal cortex, general basal ganglia, hippocampus CA1, hippocampus CA3, hippocampus general, hypothalamus, lateral olfactory tract, midbrain (remainder), motor cortex, olfactory system, periaqueductal grey, perirhinal cortex, septum, somatosensory cortex, substantia nigra, visual cortex
GWmix	caudate putamen (striatum), superior & inferior colliculi, thalamus
WM	anterior commissure, cerebellar lobules, cerebral peduncle, corpus callosum, fornix system, internal capsule, medulla, optic nerve, pons

**Table 3.1** Manual binary classification of NUS atlas labels used to initialise EM.

The NUS atlas ([Bai et al., 2012](#)) includes the pituitary gland, excluded here. In the second implementation ([Fig 3.16](#)), GM/WM mixture tissues were initialised as 50% GM, 50% WM.



**Figure 3.16** Improved tissue classifications.

The same sagittal (top), coronal (middle row) and transverse slices from a single representative WT brain in-skull, overlaid with probabilistic tissue classifications, including only GM (a; red) and WM (b; blue). Note the boundary between GM and WM at the brainstem is less definitive than in [Fig 3.15](#).

## 3.7 Intensity standardisation

The MRI intensity scale is inherently non-standardised and lacks diagnostic meaning: an identical tissue location in a repeat scan, using the same scanner and parameters under ideal conditions, may produce a different signal (*Nyúl & Udupa, 1999*). Subject movement, electronic and ambient noise, temperature and cellular tissue changes contribute to signal variation. These confounds are compounded between subjects and scanners. MR images are diagnostically valuable without standardisation, so procedures are surprisingly uncommon (*Bagci et al., 2010; Leung et al., 2010b*).

An intensity average image is created following each iteration of group-wise registration (§3.8). I standardised the MR intensity scale beforehand to the approximate [0 1] range,

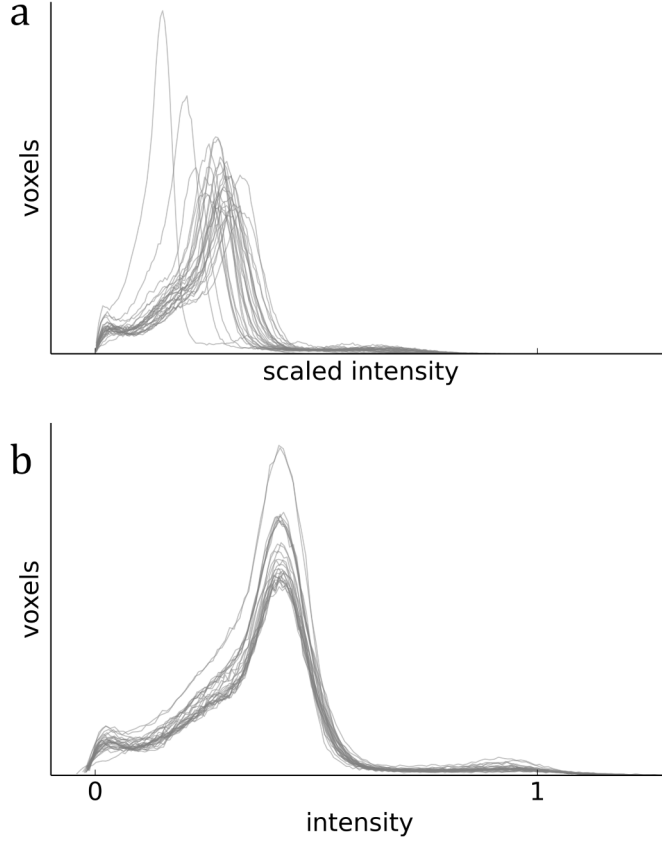
to prevent individual images' noise, intensity extremities and features from dominating, and to ensure similar intensity ranges represented equivalent anatomical regions between images. The best standardisation results are achieved after NUC (*Bagci et al., 2010*).

A 'robust' standardisation involves the trimming of the image histogram tails by e.g. 1%: voxels are sorted by intensity, the 1<sup>st</sup> and 99<sup>th</sup> percentiles  $pc$  of their total number found, and all voxels with values less than those at  $pc_1$  and greater than  $pc_{99}$  are set to these respective extremes. All intensity values are divided by the maximum, limiting the range to  $[0\ 1]$ . This beneficially broadens the histogram, increasing contrast, while maintaining relative intensity relationships between voxels. The intensity values themselves are not analysed in V/TBM studies, so their adjustment is not problematic, provided contrasts between structures, which drive the registration, are maintained. The arbitrary choice of 1% tails, however, fails to account for dominant intensities between  $[pc_1\ pc_{99}]$ : one image's GM peak may be at a higher intensity than another's – so will dominate the average, despite having the same maximum intensity. *Kovačević et al. (2005)* linearly scaled intensities to standardise them based upon the GM peak alone – which may have similar drawbacks, for non-GM structures.

I used the piecewise linear approach to standardisation described by *Nyúl et al. (2000)*. This method involves a training step, incorporating the whole dataset (or a subset), and an application step. During training, within the dilated brain masks, 11 histogram landmarks were found for each image: the 1<sup>st</sup> and 99<sup>th</sup> percentiles, and evenly-spaced at each 10<sup>th</sup> percentile. The landmark configuration as per *Nyúl et al. (2000)* was thus:

$$E3.14 \quad L = \{pc_1, pc_{99}, pc_{10}, pc_{20}, \dots, pc_{90}\}$$

The landmark means across images were found, and each image's intensities were scaled linearly between these (*Fig 3.17*). This piecewise-linear scaling was quick, so I trained using the entire dataset, avoiding bias to a sub-population.

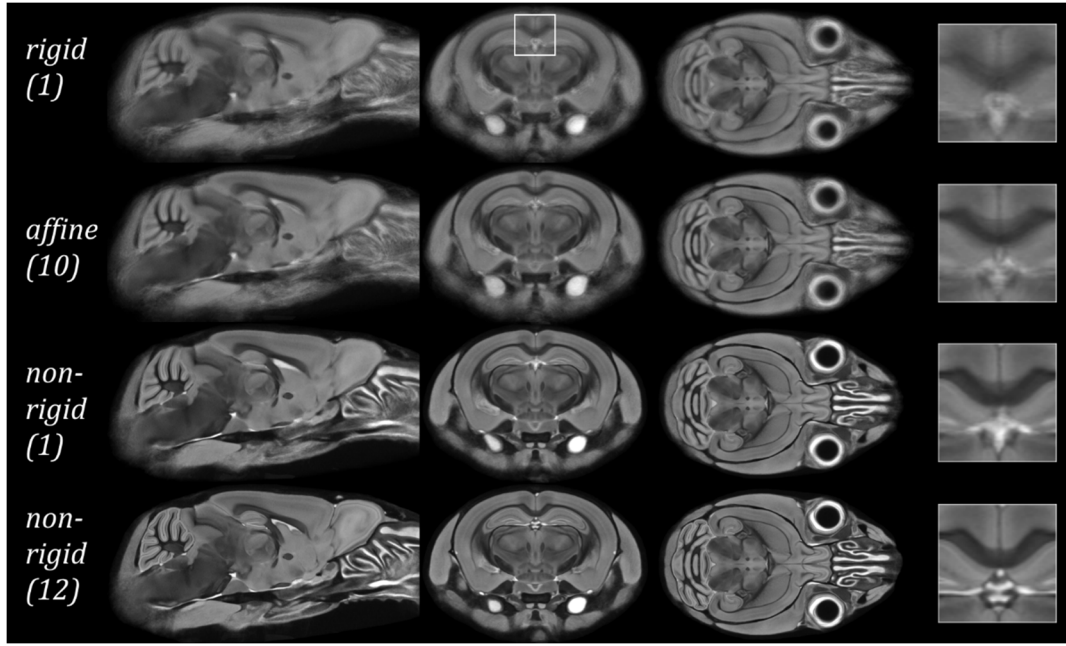


**Figure 3.17** Intensity standardisation results in the Tc1 (cohort 1, §4.1) dataset.  
 (a) original Tc1 images' histograms (within dilated masks), scaled to the [0 1] range for comparison with results of intensity standardisation (b).

## 3.8 Group-wise registration

As described in §2.2.6, registration acts to align corresponding voxels between images, to account for morphological differences arising from anatomical variation, by applying deformations and maximising a similarity measure. I spatially normalised all structural images into a common coordinate space (*Kovačević et al., 2004; Kovačević et al., 2005; Cleary et al., 2011a*) using *NiftyReg*, which performs global (rigid, affine) and local (non-rigid) registrations, with the group-wise registration (GWR) technique outlined in §2.2.6 and Fig 2.9. I employed 5-voxel dilated masks of each  $S_i$  to include a sufficient intensity gradient at the brain boundary for accurate alignment, while excluding irrelevant material. Results are shown in Fig 3.18.





**Figure 3.18** Group-wise registration average images

Results of the process illustrated by [Fig 2.9](#). The same sagittal, coronal and transverse slices of average images from the Tc1xJ20 cohort (N=49 brains, [§4.2](#)) showing progressive sharpening as the iterations (number at left) increase. By the final NRR iteration, cerebellar, hippocampal and cortical sublayers are clearly visible, at this high resolution. The right-hand column shows a zoom of the region indicated in the top coronal view. Voxel-wise standard deviation may also be used to assess the improvement in alignment, as illustrated in [Fig 3.20](#).

The first iteration was rigid-only, 6 DOF: 3D rotation and translation. Iterations 2—5 were affine (12 DOF, including 3D shears and scaling), employing a symmetric block-matching algorithm at progressively smaller scales ([Ourselin et al., 2001](#); [Modat et al., 2014](#)). ([Fig 3.20](#) shows 5 iterations of global registration, but I later adopted up to 10, as the standard deviation continued to reduce after 5 iterations.) Due to the small features of mouse brains and the global variability of images, I found that it was sometimes necessary to adjust the number of pyramidal coarse-to-fine-scale levels – for example, when performing affine registration between in-skull and ex-skull images. For the affine stages of GWR I used 4 levels. I performed 15—20 subsequent, non-rigid registration iterations using symmetric, inverse-consistent Free Form Deformation ([Rueckert et al., 1999](#); [Modat et al., 2012](#)): the image was warped using cubic B-splines deformed between

increasing numbers of control points placed at progressively finer scales, employing normalised mutual information as the similarity measure to be optimised. *NiftyReg* was parameterised with constraints on the final control point spacing: 5 voxels for *ex vivo* data; 5 levels of doubled CP spacing; and a penalty term for bending energy (0.005).

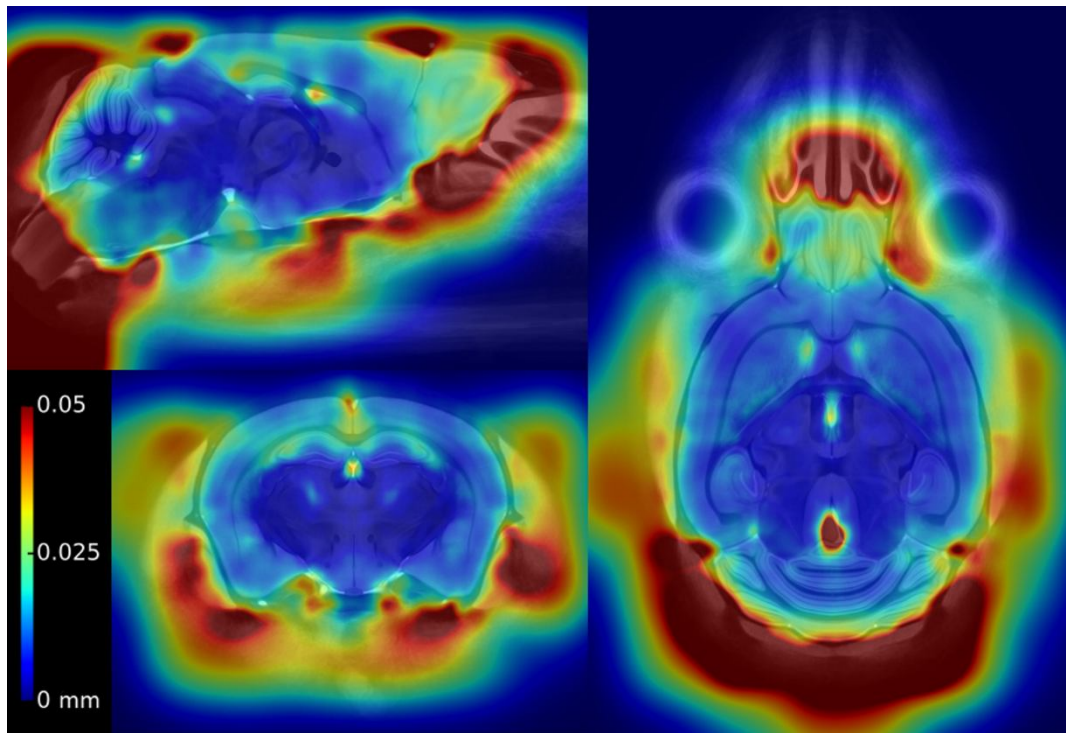
Deformations at a particular iteration of GWR may be summarised for the whole cohort with a mean positional distance image (MPD, [Fig 3.19](#)), wherein the magnitude of the 3D deformation vectors is calculated at each voxel and the mean found over all  $N$  image displacement fields,  $i$ :

$$\text{E3.15} \quad MPD(\text{voxel}) = \frac{1}{N} \sum_{i=1}^N \sqrt{x_i^2 + y_i^2 + z_i^2}$$

where  $x, y, z$  are the voxel's vector components from the non-rigid registration only (excluding rigid and affine).

Regions outside the brain, the brainstem and the ventricles show the greatest variability in [Fig 3.19](#). To avoid external tissues' deformations interfering with brain tissue, particularly at the cortical surface, I used conservative dilated masks (3—5 voxels *ex vivo*) when performing the GWR used for TBM analysis. The hippocampus, cortex and corpus callosum were more affected than the thalamus and midbrain in this cohort, although the relative intensity homogeneity within the latter structures also limits the ability of registration to recover morphological changes. The olfactory bulbs show more variability than the rest of the brain, but in these in-skull images the degree of deformations ( $<50\mu\text{m}$ ) is far less than that seen in the ex-skull images of [Kovačević et al. \(2005\)](#), which had deformations in this region  $>500\mu\text{m}$ , and damage or handling-related deformations in the cortical surface of around  $250\mu\text{m}$ . Similarly, in registering ex-skull *ex vivo* to in-skull *in vivo* images, (which I show in [§5.7.2](#)), [Ma et al. \(2008\)](#) showed deformations  $>500\mu\text{m}$  in the brainstem and olfactory bulbs, and changes around  $250\mu\text{m}$  in the cerebellum – all attributable to brain extraction (see [Fig 5.17](#)).





**Figure 3.19** *Ex vivo* mean positional distance image from a cohort scanned in skull. Sagittal, coronal and transverse slices, showing the mean magnitude of all images' NRR deformations overlaid on the final Tc1xJ20 group-wise average (§4.2). The scale, in mm, gives a summary of the amount each region deformed to meet the final average, and hence a quantified overview of local morphological differences within the cohort. The greater distances in extraneous material result from high variability, in part caused by damage and variable flesh and fat-stripping following skull excision.

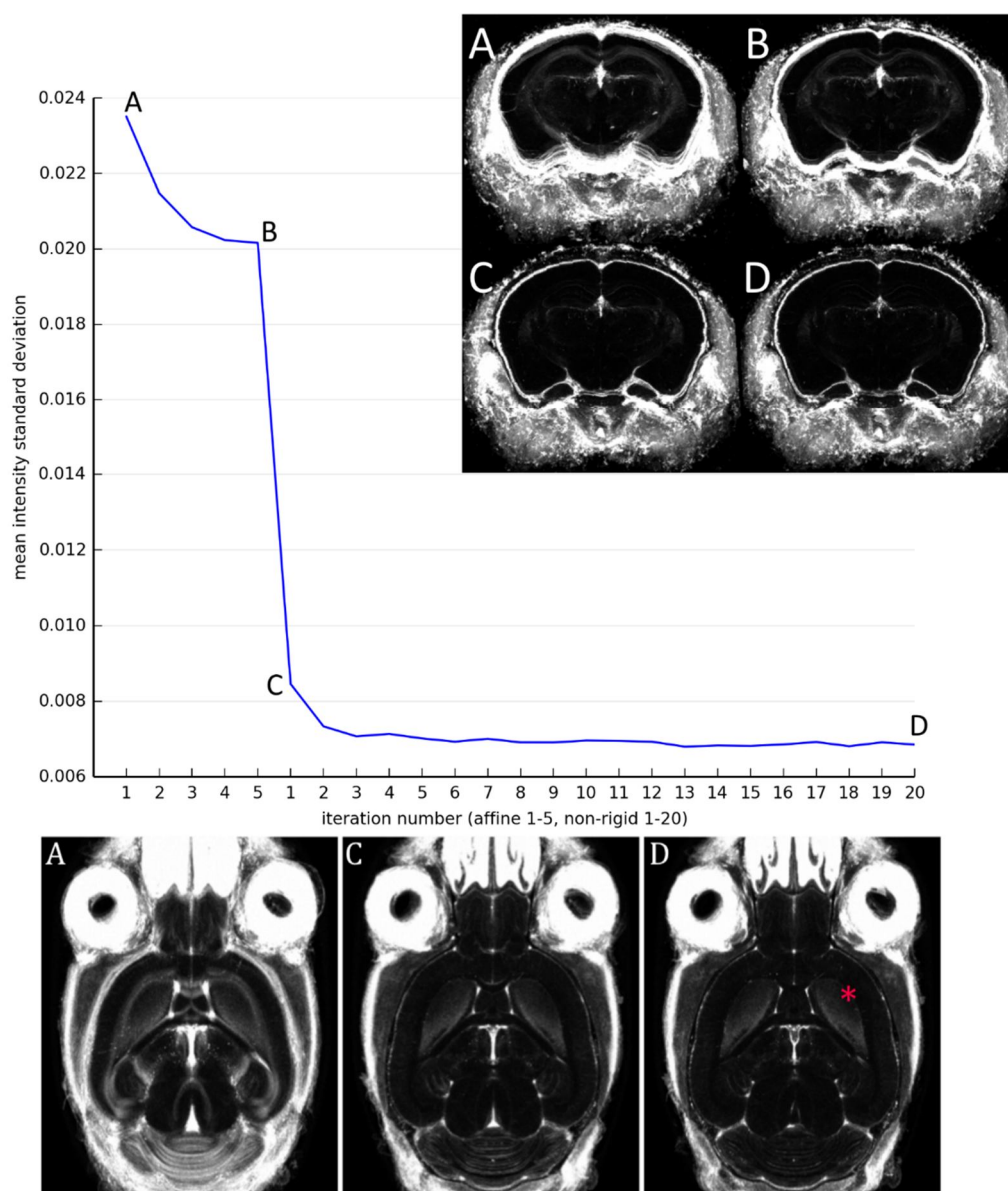
### 3.8.1 Assessing GWR quality

GWR is by far the most time-consuming aspect of the pipeline, taking several days to complete 20 NRR iterations of *ex vivo* data, even when run in parallel on a cluster (this was usually a combination of job queuing and the actual time taken per registration step). Each registration generally converged after less than eight hours, even at the highest resolutions (40 $\mu$ m). It was prudent to visually inspect each deformed image for misregistration, which may converge on local maximum similarities with false or non-physiological deformations which do not represent the overlapping of equivalent tissues. The performance of GWR can be quantitatively assessed, to determine the number of

iterations necessary to reasonably align images and sufficiently recover morphological variation.

A single registration's accuracy may be checked by manually segmenting both the target,  $T$ , and each source image,  $S_i$  (the whole brain or individual substructures), applying the transformations and deformations to the segmentation of  $S_i$ , and calculating a Dice score between the two resulting segmentations to quantify their overlap. Dice has disadvantages as it is sensitive to scale (as smaller structures are more sensitive to registration error, [Ma et al., 2014](#)), and this assumes each manual segmentation is equally accurate. [Lau et al. \(2008\)](#) assessed the performance of their GWR at each iteration by computing a voxel overlap metric: parcellating each individual image and propagating these to the average after registration. They showed acceptable convergence to 1 (perfect overlap) after about 6 iterations of NRR.

Registration errors may be revealed in the “difference image” between image pairs – the GWR target and each resampled floating image. The standard deviation (SD) of all resampled images at each voxel (requiring prior intensity standardisation) may be visualised ([Fig 3.20](#)) to summarise regions of high variability, where registration has more difficulty. Both highlight large intensity differences and are thus biased towards structures bounding CSF regions over subtler GM/WM boundaries.

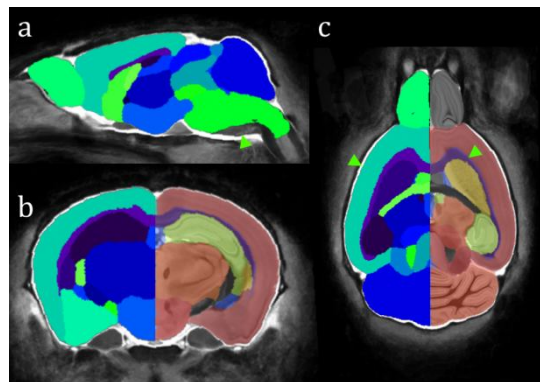


**Figure 3.20** Plot of the overall mean standard deviation (SD) of voxel intensities, within the original target brain mask, of the resampled structural Tc1 and WT images at each stage of GWR.

Inset images (top right) show the same coronal slice of the SD images at indicated iterations. Bottom row shows the same transverse slice. The maximum intensity of individual images, and hence the average, was approx. 1.2, after intensity standardisation; these images are thresholded so their maximum (white) is 0.05. The registration progressively improves alignment, reducing SD – dramatically so, after the first iteration (it.1) of NRR (C). Although improvement after NRR it.6 is not obvious and there is some minor fluctuation, in this case the minimum occurs at it.13. Note the variability of the striatum, even at it.20 (red asterisk). See also [Fig 5.5](#). This plot may still display low values despite misregistration; it was important to also visually inspect the final results.

### 3.9 Segmentation propagation for Jacobian integration

I segmented the final GWR average image using STEPS and the UFL atlas, as per [Ma et al. \(2014\)](#) and §2.2.3. The resulting parcellations (e.g. [Fig 3.21](#)) were used to localise statistical results, and, using the NRR deformations, may be back-propagated to individual subjects to measure volumes ([Cleary et al., 2011a](#)). This has the disadvantage of including an additional interpolation step, which may reduce accuracy. Instead, I measured the volume of each region in each mouse brain via integration of the determinants of the NRR deformations' Jacobian matrices (§2.2.7; [Boyes et al., 2006](#)).



**Figure 3.21** Final Tc1 NRR average image with UFL atlas parcellations overlaid. (a) Sagittal; (b) coronal; (c) transverse views. Right half semi-transparent in b, c. Arrows indicate, from top left, incomplete coverage of the brainstem owing to improper atlas registration; successful exclusion of eCSF; slightly over-generous coverage of the corpus callosum.

### 3.10 Transformation of Jacobian determinant maps

For TBM, I took the  $\log_e$  of the Jacobian determinant  $J_{det}$  at each voxel in the deformation fields generated from GWR (§2.2.7), having multiplied with the determinant of the affine matrices from the final affine iteration, to include global scaling

(discussed in more detail in §5.6). The log transformation gives a more intuitive (and symmetric) scale: values  $< 0$  represent contraction; values  $> 0$  represent expansion; values of 0 represent no change. I smoothed the result with a 3D Gaussian kernel (FWHM 0.16mm, both *ex vivo*, §4 and *in vivo*, §5.3), chosen to maintain sufficient resolution to identify small features, which helps to account for remaining registration imprecision and renders the values more normally distributed, an assumption of subsequent statistical tests (Ashburner & Friston, 2000).

### 3.11 Modulation of tissue class maps

For VBM, each image's tissue class maps were resampled to the final group-wise structural average space, thereby dilating or contracting (originally probabilistic) individual voxels, whose values remain constant. The deformed maps may then be modulated by multiplying with the  $J_{det}$  values, which represent the relative volume change of individual voxels, to preserve intra-subject tissue volumes, and smoothed (as above). The result is a comparison of voxel-wise volume of each tissue class between animals. The local concentration (proportional density) of the tissue class is compared if modulation is omitted, as it was for the Tc1 study (§4.1.9; Ashburner & Friston, 2000; Good *et al.*, 2001; Mechelli *et al.*, 2005).

### 3.12 Statistical tests

The final step is the generation of 3D statistical parametric maps (SPMs) consisting of False Discovery Rate (FDR)-corrected two-tailed t-statistics at every voxel of the final average image from the GWR. In most cases I used a conventional FDR threshold of  $q=0.05$  (Benjamini & Hochberg, 1995; Genovese *et al.*, 2002). I performed mass-univariate two-tailed t-tests using the General Linear Model (GLM) on the transformed deformation fields (TBM) and, for the Tc1 dataset (§4.1), unmodulated tissue class

images (VBM), with ANCOVA covariates for TIV (measured by the segmented tissue volumes).

The models used were (*ex vivo*) [E4.3](#), [E4.4](#); (*in vivo*) [E5.1](#). Some groups opted to threshold the resulting maps at a minimum cluster volume, to exclude very small regions of significance, which may be spurious or result from random noise (e.g. [Li et al., 2009](#) thresholded at 150 voxels; [Biedermann et al., 2012](#) at 200 voxels). I did not perform this additional thresholding step.

### 3.13 Discussion

μMRI statistical morphometry offers several advantages for phenotyping transgenic and disease model mice – by enabling computational approaches and, by covering large regions in an unbiased manner, reducing the need for time-consuming, destructive exploratory histology. Clinical image processing tools have been used to process preclinical data; however, this often requires adjustment to fit human-centric parameters. Manual image processing and preparation remains prohibitively time-consuming. Heretofore no approach has unified all the pre-processing steps necessary for fully automatic, large-cohort, high-throughput preclinical phenotyping with MRI.

This pipeline is applicable to *ex-* or *in vivo* mouse brain and *ex vivo* embryo studies employing μMRI analysis techniques including V/TBM and segmentation. Other preclinical models, such as rats, may also be analysed. The only steps requiring an applicable atlas are tissue segmentation and parcellation. The pipeline is designed to eliminate laborious steps, including multiple-subject extraction, orientation correction and skull-stripping, given appropriate data, with minimal intervention. I prioritised automation and robustness over speed; in these earlier steps, the process may be initialised with downsampled images to save time. Although the majority run in a few hours on a single modern processor, registration is better suited to parallel environments: I used a cluster running the Sun Grid Engine to queue jobs and run many registrations, and other tasks, in parallel.

A high resolution (40 $\mu$ m) imaging protocol was used for *ex vivo* structural phenotyping at 9.4T, giving greater SNR and spatial precision than is available at lower field strengths, and increasing contrast between anatomical structures and tissues, which mitigates PV effects (*Natt et al., 2002*). For a full investigation of the SNR and CNR achieved, see §4.1.5 (*ex vivo*) and §5.5 (*in vivo*). A subject separator eliminated complications arising from touching subjects in multi-brain scans. In these images, subjects were identified using brain or embryo volumes from training data; additional features such as intensity, texture or shape could also be included for this step.

The order of the steps is deliberate. Orientation correction is perhaps the most important, as it enables subsequent atlas-based approaches. However, some flexibility is beneficial; both multiple subject extraction and orientation correction may benefit from prior non-uniformity correction.

Orientation correction – to a standard space – enabled the multi-atlas techniques described to produce high quality brain masks, subject-specific tissue segmentations, and anatomical parcellations. The UFL atlas contains more brains and is thus suited to *ex vivo* and *in vivo* masking; the NUS atlas has more detailed parcellations, including cerebellar GM and WM, which were useful for initialising tissue segmentation. The earlier brain masks provide a good initial estimate of brain volume, which I employed in §5; the segmentation step may be skipped if VBM is not required.

Using registered priors from the NUS atlas resulted in significant qualitative improvements over prior-free segmentations based upon intensity alone. External non-brain voxels may be misclassified, for example, as GM, or cortical surface PV voxels as WM; separate background and eCSF classes mitigated this and improved GM classification at the brain surface (see *Fig 4.10*). As *Van Leemput et al. (1999a)* noted, external tissues may still be mis-segmented after affine registration with external priors; hence, to account for local variability, I used NRR. Extraneous tissues in the Tc1 and rTg4510 cohorts occasionally registered to the surface of the skull-stripped UFL atlas brains, particularly at the cerebellum and olfactory bulbs, where tissue separation is minimal. This required refinement of registration parameters, or substitution of a more appropriate atlas, for masking (§5.3.2).



The impetus for including GWmix as a separate class in segmentation ([Fig 3.14](#)) was that mouse brains exhibit a substantive proportion of PV, compared with human brains: GM and WM are not clearly delineated in some large subcortical regions, such as the thalamus, midbrain and pons. [Lee et al. \(2009\)](#) had a similar motivation. However, it was not clear what the implications of the GWmix class were, for VBM. Because the proportions of GM and WM in this region varied significantly between animals, statistical tests of GM or WM proportion did not appear to be meaningful. The class could be excluded from statistics, but this reduces the usefulness of morphometric tests, which ideally cover the entire brain. My subsequent segmentations, omitting GWmix, ([Fig 3.15](#)) appeared to be superior.

While not without faults (for example, misclassification of the dark Purkinje cell layer of the cerebellum as WM), these tissue segmentations ([Fig 3.15, 3.16](#)) represent significant improvements over those previously published in the literature ([§2.2.5](#)).

Atlas-based approaches should be used with caution given the extraordinary number of available mouse strains. A significant issue is the relevance of atlas-based TPMs to pathological data: TPMs based upon healthy subjects may be poor fits for those exhibiting gross brain changes. This concern should be reduced when expected morphological changes are likely to be subtle, accounted for by registration; and where V/TBM techniques are more relevant.

Most structural differences between the *in vivo*, skull-stripped NUS atlas the *ex vivo*, in-skull data discussed here were resolved with NRR. However, as the *ex vivo* brains' ventricles had mostly collapsed, segmentation improved after discarding ventricular CSF labels and classifying remaining internal CSF using eCSF intensities post-hoc (see [§5.7, §6.2.3](#)). Furthermore, the intra-class variance of GM and WM intensities was high, particularly between cortical and cerebellar GM, as well as between corpus callosum and deep cerebellar WM. Currently-available atlases are limited by relatively low contrast and resolution compared with those achievable at high field strengths. Both masking and segmentation could be improved with an in-skull atlas more suited to the datasets described in this thesis; many opportunities exist for extending the size and quality of currently available MRI mouse atlases.



Brain masks were refined for statistics using tissue segmentations, and group-wise registration resolved remaining spatial differences. The pipeline’s modularity enables flexibility, as well as the substitution of other analysis packages – such as SPMMouse – for later stages, if preferred. Permutation testing, e.g. using FSL’s Randomise, may be used to validate parametric statistics if desired ([Sawiak et al., 2013](#); [Winkler et al., 2014](#)).

Despite the difficulties posed to registration and segmentation techniques by the close proximity of brain and external tissue, in-skull imaging is vastly preferable to ex-skull, owing to near-inevitable damage incurred by the brain during extraction or manual handling ([Fig 3.19](#); [Kovačević et al., 2005](#); [Ma et al., 2008](#); [Scheenstra et al., 2009](#); [Sawiak et al., 2012](#)).

## 3.14 Conclusions

I have described an automated software package for the high-throughput phenotyping of large cohorts of mouse brains, from MRI to statistical parametric map, using novel approaches for separation from multiple-subject scans and orientation correction – previously performed manually – followed by skull-stripping and tissue class segmentation based upon external atlases. I applied freely available software, developed for use with clinical data, to this preclinical paradigm, using appropriate parameters and custom atlases, without requiring the data be adjusted to conform to human tools. In the following two chapters, I show the pipeline’s application to *ex vivo* and *in vivo* mouse brain datasets, including the Tc1 transgenic model of Down syndrome, and the rTg4510 model of tauopathy in neurodegenerative disease. In the latter I describe some adaptations for *in vivo* images.

As the pipeline is modular, the steps may be performed individually, enabling integration with other analysis techniques such as segmentation propagation or cortical thickness estimation based upon the deformation fields from GWR; or segmentation with label fusion techniques based upon prior parcellated atlases; as well as with other packages, such as SPM.

The pipeline may be of use to the phenotyping community. With the homogenisation of scan parameters, standardisation of analysis pipelines, and improved availability and accuracy of TPMs and atlases, multi-site and large cohort studies are possible, increasing the feasibility of  $\mu$ MRI and morphometry as important, powerful preclinical phenotyping tools. Many more mouse model MR images are available, and this software should greatly simplify their morphometric analysis.

## 4 Application of the pipeline to *ex vivo* mouse brains

The previous chapter described the development and testing of a software pipeline for automatically and robustly processing large cohorts of mouse brains and embryos for morphometric phenotyping. This chapter includes results, and discussion arising from two *ex vivo* mouse brain datasets processed with this pipeline:

§4.1 The Tc1 model of Down syndrome

§4.2 The Tc1xJ20 cross-bred mouse model of Down syndrome and AD

Each dataset is briefly introduced with background about the mouse model and the motivations behind the study. Study details are given, along with any necessary adjustments to the pipeline, followed by morphometry results output and discussion. Each dataset provided a unique opportunity for discussing various aspects of morphometry, including the statistical analysis. The Tc1 study was the most detailed investigation and comprises the bulk of this chapter. Some issues with the data, including a perfusion artefact, are discussed.

Research in this chapter was conducted with various collaborators and colleagues. Jon Cleary and Ben Sinclair acquired the first cohort of Tc1 data. Holly Holmes acquired the Tc1xJ20 dataset. Both of these originated with Professor Elizabeth Fisher and Dr. Frances Wiseman at the UCL Institute of Neurology. Da Ma provided the brain masking code for STEPS parcellations, and performed the thickness calculation on the hyperintense rim.

## 4.1 Morphometric analysis of the Tc1 model of Down syndrome

### 4.1.1 The Tc1 model

Down syndrome (DS) and mouse models were briefly introduced in §2.4.1. The Tc1 mouse model carries a freely segregating copy of human chromosome 21 (Hsa21) and is trisomic for 92% of Hsa21 genes (*O'Doherty et al. 2005; Gribble et al., 2013*). This mouse model of DS recapitulates many features including cardiac defects, short-term memory impairment, motor deficits and mandible malformation seen in humans and other DS mouse models, such as Ts65Dn and Ts1Cje (*Reeves 2006; Moore & Roper, 2007*). The brain, however, has yet to be fully characterised.

Using the pipeline described in §3, I performed a fully automated morphometric analysis of brains from the Tc1 mouse model of DS, using microscopic magnetic resonance imaging ( $\mu$ MRI) and Voxel- and Tensor-Based Morphometry. I compared 29 Tc1 brains with 28 wild-type (WT) littermate controls, imaged *ex vivo* at 9.4T. This first TBM analysis of a DS mouse model revealed global and local volumetric differences – both novel and previously described via histology. I show an unexpected increase in Tc1 TIV and, controlling for this, local volume and grey matter density reductions in the Tc1 brain compared to the WTs, most prominently in the cerebellum, in agreement with human DS and previous histological findings.

I here discuss the pipeline, results, custom considerations for this dataset, and further validate the Tc1 mouse as a preclinical model of human DS.

### 4.1.2 Mice and multiple brain imaging

Two cohorts  $C_1$ ,  $C_2$  of *ex vivo* adult Tc1 and WT brains were scanned at CABI. For reasons of statistical power (see discussion in §4.2.4), I analysed them together.  $C_1$  brains were scanned individually. For  $C_2$  brains, a multi-subject protocol was employed.

### ***Ethics statement***

This study was conducted following approval by the local Ethical Review Process of the MRC National Institute for Medical Research and authorisation by the UK Home Office, Animals (Scientific Procedures) Act 1986 under relevant Project Licence authority. The ERP approved and reported that all work reflects contemporary best practice. High standards in the design and conduct of work were applied and full implementation and consideration of the 3Rs (where appropriate) was made.

### ***Mice, genotyping and fixation***

C<sub>1</sub>: 28 male mice aged 4-5 months (14 Tc1, 14 WT littermate matched controls) were taken from a colony maintained by mating Tc1 females to F1(129S8×C57Bl/6JNimr) males and genotyped as per [Sheppard et al., 2012](#). C<sub>2</sub>: 29 male mice aged 15 months (15 Tc1, 14 WT littermates) were taken from a colony maintained by mating Tc1 females (from the same colony as those used to breed C<sub>1</sub>) to B6.Cg-Tg(PDGFB-APPSwInd)20Lms/2J males (see [Cohort 2 genotypes](#), below). Brains were perfuse-fixed using an optimised protocol for structural  $\mu$ MRI mouse brain phenotyping ([Cleary et al., 2011b](#)). The animals were terminally anaesthetised by administration of Euthanal (0.1mL) via intraperitoneal injection. Perfusion was performed through the left heart ventricle with 15-20mL 0.9% saline. Brains were then post-fixed (with 10% formal-buffered saline, doped with 8mM Gd-DTPA “Magnevist” contrast agent, flow rate 3mL per minute) and stored for 9 weeks at 4°C, then scanned in-skull to prevent damage.

### ***Cohort 2 genotypes***

J20 (B6.Cg-Tg(PDGFB-APPSwInd)20Lms/2J male mice (Jax stock code 006293) were used to breed C<sub>2</sub>, with Tc1 females (see [§4.2](#)). All mice included in this study were genotyped as being negative for the J20 transgene. These J20 male progenitors were maintained by mating J20 APP transgenic mice to C57Bl/6J for more than 12 generations ([Mucke et al., 2000](#)). Mice were genotyped using polymerase chain reaction for the human APP transgene (as per Jax 006293).

## Image acquisition

Brains were secured with surgical gauze inside a 20ml syringe ( $C_1$ ) or within a subject separator ( $C_2$ ). Syringes were filled with proton signal-free, non-viscous Fomblin perfluoropolyether (PFS-1, Solvay Solexis SpA., Bollate, Italy) to avoid air interface susceptibility artefacts, and scanned with an Agilent/Varian 9.4T VNMRS using a 60mm (inner diameter) gradient set. Single brain protocol ( $C_1$ ): 26mm quadrature volume coil (RAPID Biomedical GmbH, Würzburg, Germany), 3hr spoiled GE3D sequence, parameters: TE 4.03ms; TR 17ms; flip angle 52°; 6 averages; FOV 20.48x13.04x13.04mm (matrix 512x326x326, 40µm isotropic resolution). 3-brain protocol ( $C_2$ ): 33mm quadrature birdcage coil; 11hr 4min spoiled GE3D; TE 4.54ms; TR 17ms; flip angle 51°; 6 averages; FOV 32x25x25mm (matrix 800x625x625).

## 3-brain subject separator

For  $C_2$ , a custom-designed, 3D-printed plastic mouse brain holder (25.4mm diameter, 44mm length) ([Fig 3.5a](#)) secured three skulls inside a 50ml syringe. Its 1mm walls precluded touching or partial volume (PV) between neighbouring subjects. This significantly reduced overall preparation time, enabling unsupervised overnight scans (the additional time accounting for increased FOV). I measured SNR<sup>1</sup> ([E4.1](#)) and CNR ([E4.2](#)) in both cohorts using the tissue maps created as per [§3.6](#) (binarised at 50% probability).

$$\text{E4.1} \quad \text{SNR} = \frac{\text{mean signal}}{\text{standard deviation noise}}$$

$$\text{E4.2} \quad \text{CNR} = \frac{\text{signal}(GM) - \text{signal}(WM)}{\text{standard deviation noise}}$$

Brains were aligned in one z-direction layer to minimise signal drop-off and geometric distortion away from the bore isocentre. To ensure gradient accuracy, the system was

---

1: [McRobbie et al. \(2007\)](#) suggest multiplying the result of [E4.1](#) by 0.66, as the BG has an approximately Rician distribution, which has (because although noise can be negative, it is made positive when the magnitude image is calculated) a narrower standard deviation than Gaussian. The formula reported here is used for consistency with [Cleary et al. \(2011b\)](#), and is given by [www.mr-tip.com](http://www.mr-tip.com), which recommends taking the signal over a small homogeneous brain region and the BG over a larger external region.

calibrated prior to imaging (O’Callaghan et al., 2014). Gradient linearity was within manufacturer’s limits, as measured within a centred sphere (20mm radius) encompassing the 3-brain FOV. Scaling measurements were performed throughout data acquisition to measure temporal stability (see §4.1.5). To remove possible group bias caused by remaining distortion or gradient instability, genotypes were interleaved within scans, and positioning within the 3-brain holder was randomised. See Fig 3.5b for an example raw MR image containing 3 brains from this study. One WT brain from C<sub>2</sub> was excluded from GWR and analysis, owing to a collapsed skull unilaterally compressing the brain.

### 4.1.3 Voxel-wise statistical tests

After processing with the complete pipeline (§3.2—3.12), I performed voxel-wise two-tailed t-tests using a GLM. The model was:

$$E4.3 \quad Y = b_1P(Tc1) + b_2P(WT) + b_3Age + b_4Background + b_5TIV + \epsilon_i$$

where the vector  $Y$  represents, at a particular voxel, the response values from each animal:  $\log(J_{det})$  for TBM, or the proportional density of a particular tissue class for VBM.  $b$  are the regression coefficients;  $P(genotype)$  is the binary encoding (0 or 1) of each animal being Tc1 or WT; *Background* is a binary encoding for C<sub>1</sub> and C<sub>2</sub>; and  $\epsilon$  is the residual error vector. I controlled for TIV in TBM, to reveal differences in the Tc1 group independent of overall volume. The TIV covariate was excluded for VBM, as the intensity values were not thought to have a relationship with volume. Levels of the effect of interest (genotype) were compared using contrasts. Tests were constrained to the brain mask to limit the multiple testing problem (Nichols & Hayasaka 2003), and to exclude skull and external tissues, which exhibit high inter-individual variability. I also performed two-tailed t-tests on the probabilistic GM class and parcellation volumes, after normalising to TIV.

### 4.1.4 Group-wise registration assessment

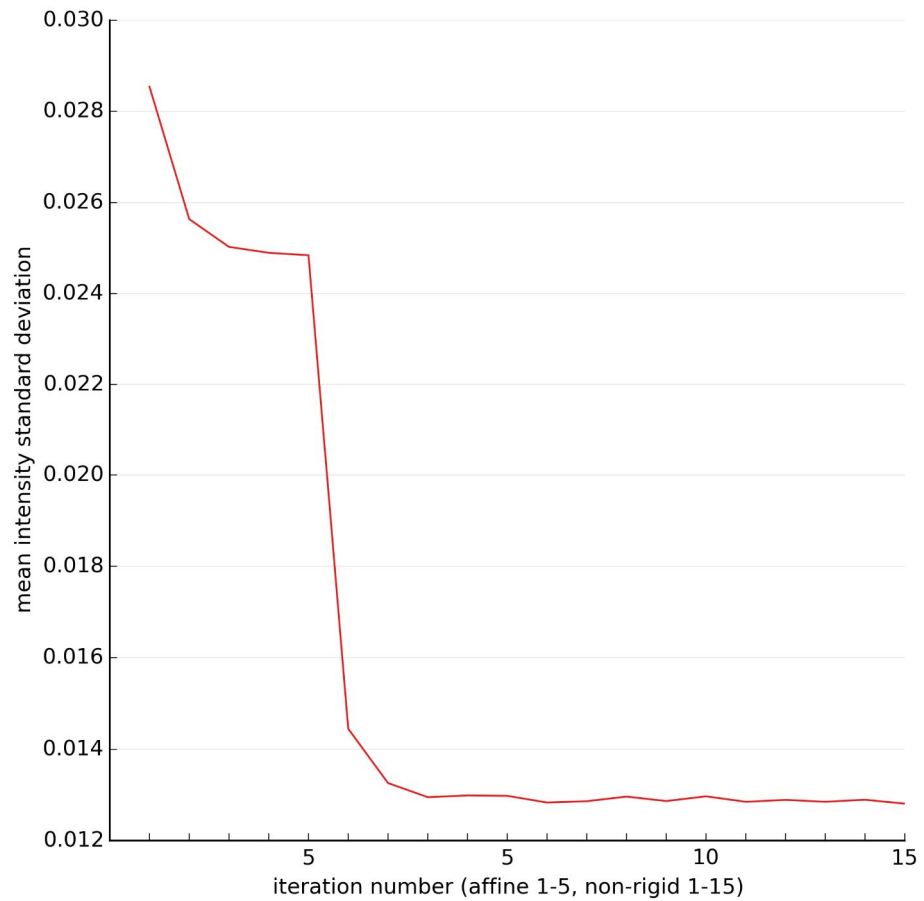
Morphometry is dependent upon intra-group structural registration accuracy. As registration is gradient-driven, contrast is crucial to success, and a driver of increasing

field strengths. Accuracy within homogeneous structures, such as the hypothalamus and brainstem, may be impeded: a local change in the centre of such a structure may be missed by TBM; a uniform change may be compounded and elicited only at its boundary. This may explain why the TIV covariate did not account for some local expansions. A similar issue faces parcellation-based studies when structural delineations are ambiguous, particularly as mice exhibit a high degree of PV.

The 3-brain protocol ( $C_2$ ) realised lower SNR and CNR than single-brain ( $C_1$ ) scans; however, contrast was sufficient to drive registration. [Van Eede et al. \(2013\)](#) found, using artificially simulated deformations, that while false positive rates were below 1%, FDR (at  $q=0.1$ ) recovered only 38% of 20% changes in volume, and that TBM (via integration of  $J_{det}$  values across deformation fields) generally underestimated true volume reductions. They attributed this underestimate to topological constraints put upon registration (such as bending energy, here). This could also arise from misregistration or high structural variance between groups; hence, it is important to check registration accuracy. The same group noted natural local volume variability within a wild-type cohort of up to 2.5%. If this was universal, it would be difficult to distinguish smaller inter-group differences with low animal numbers.

I quantified the progressive improvement in image alignment at each iteration of group-wise registration (GWR), and justified the number of iterations, by measuring the intensity standard deviation  $\sigma_I$  between resampled images at every voxel within the brain (as [§3.8.1](#)). Inter-image intensities in equivalent aligned regions should have low standard deviation, thanks to standardisation prior to GWR. [Fig 4.1](#) shows the mean  $\sigma_I$  over all voxels, at each iteration, within the brain mask. After a dramatic decrease with the first iteration of NRR,  $\sigma_I$  reaches a plateau after 5-10 iterations.





**Figure 4.1** Group-wise registration assessment for the Tc1 and WT groups. Mean standard deviation of all brain voxels in resampled images, within the brain mask. Although from a different group of images, this exhibits the same characteristic pattern as [Fig 3.20](#).

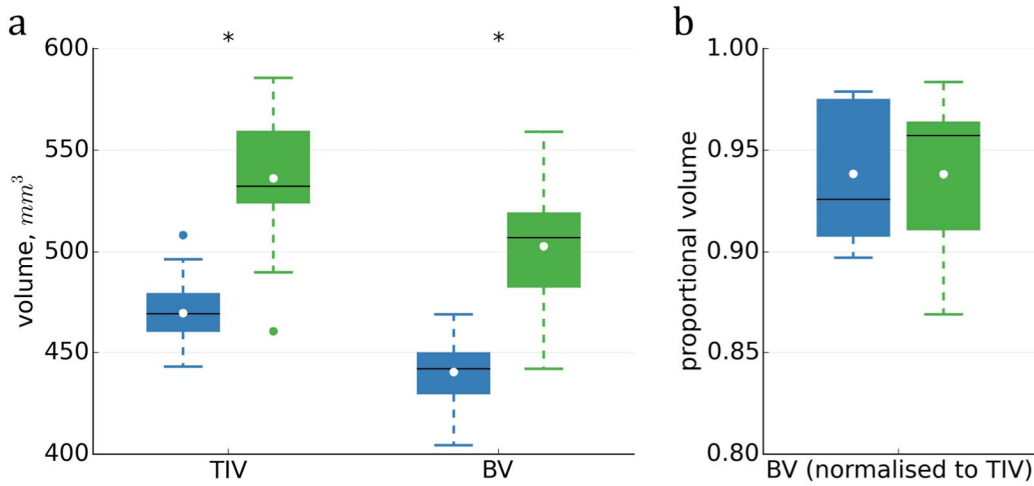
### 4.1.5 Global volume results

My analysis revealed several novel features of the Tc1 model, discussed below. Mean (SD) SNR in  $C_1$  was 29.3 (2.50);  $C_2$ : 13.6 (0.41). CNR in  $C_1$  was 12.4 (1.40);  $C_2$ : 7.06 (0.40). The sporadic hyperintense rim present in  $C_1$  was not found to have a deleterious effect upon results ([§4.1.14](#)).

The Tc1 mice exhibit greater brain and total intracranial volumes than WT littermates ([Fig 4.2](#); [Table 4.1a](#)), with little overlap. BV was on average 93.8% of both WT and Tc1 TIV (no significant difference,  $p=0.99$ ), indicating CSF did not play an appreciable role in separating groups.

Mean absolute volumes of the 20 anatomical parcellations are shown in [Table 1b](#) and [Fig 4.3](#). Most segmented tissues and parcellated regions had greater standard deviation and volume in the Tc1 mice. The brainstem, hippocampus, neocortex, thalamus and ventricles were all significantly larger in the Tc1s, both before and after normalisation. The cerebellum, internal capsule and olfactory bulb were all significantly smaller in the Tc1s after normalisation – but none were significantly different before.

The order of  $C_1$  scans was randomised, with genotypes interleaved, to avoid the possibility of scanner miscalibration affecting group volume differences. In  $C_2$ , genotypes were mixed randomly within 3-brain scans.  $C_{1,2}$  volumes were linearly scaled to account for phantom-based gradient calibration performed between acquisitions ([O'Callaghan et al., 2014](#)). Scaling factors were  $C_1$ : 1.0321;  $C_2$ : 0.9983 (5 s.f.). Although the scaling did act to separate cohort volumes, t-tests on intra-group mean TIVs did not discern a significant difference between cohorts before ( $p_{WT} > 0.45$ ;  $p_{Tc1} > 0.82$ ), or after ( $p_{WT} > 0.05$ ;  $p_{Tc1} > 0.14$ ) scaling.



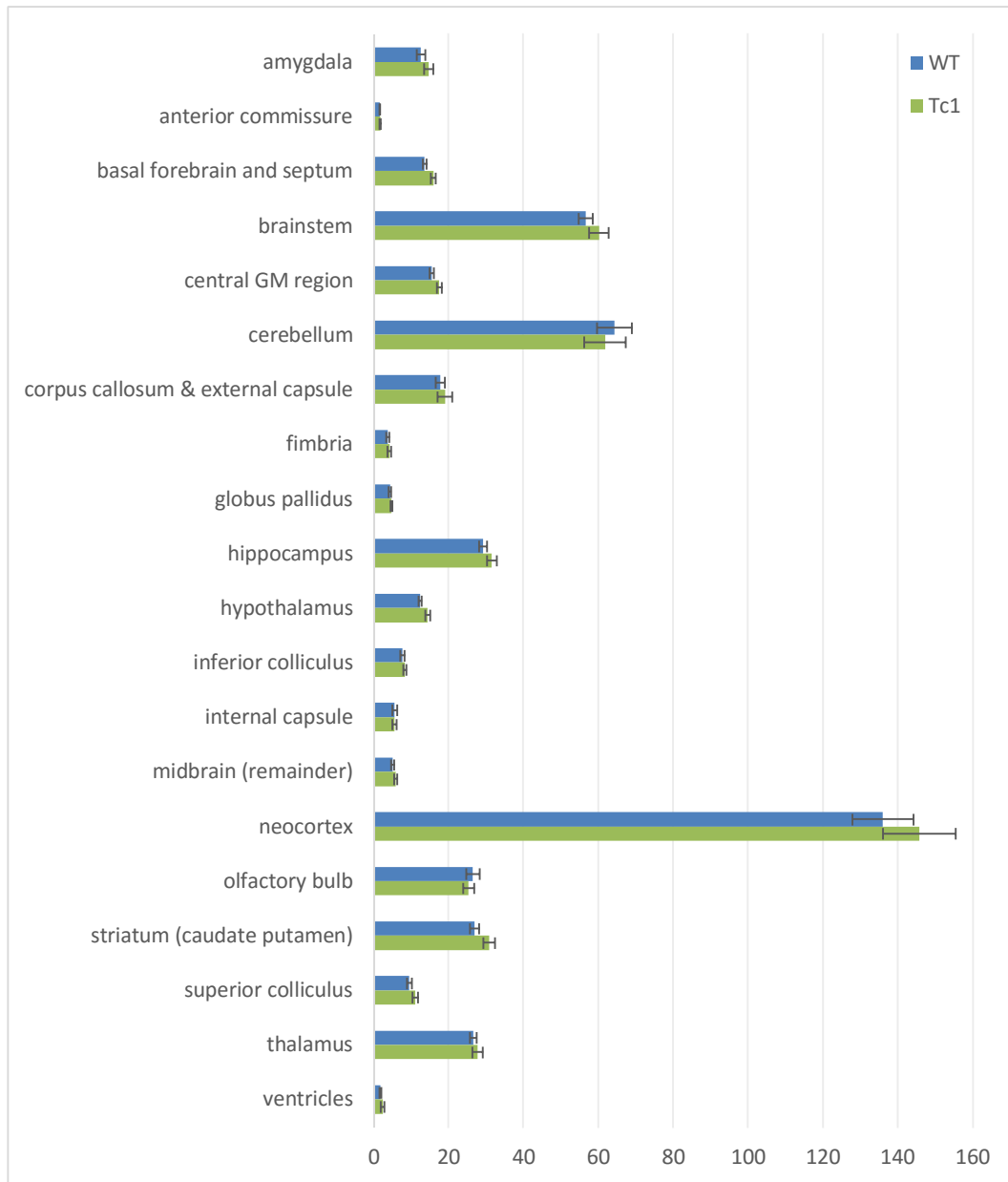
**Figure 4.2** Total intracranial and brain volumes (TIV, BV) of WT (blue) and Tc1 (green).

Means (white dots) for absolute volumes (a) were significantly different (\*), but after normalisation to TIV (b), BV means were almost identical. Outliers shown are  $>1.5$  inter-quartile range (IQR).

	<i>WT</i> (N=27)		<i>Tc1</i> (N=29)		<i>p</i>	
<i>region</i>	<i>mean</i>	<i>SD</i>	<i>mean</i>	<i>SD</i>	<i>absolute</i>	<i>TIV-normalised</i>
<b><i>a</i></b>						
<i>GM</i>	316.51	9.63	363.33	21.03	1.82x10 <sup>-13</sup>	
<i>WM</i>	123.89	9.52	139.57	11.29	8.86x10 <sup>-6</sup>	
<b><i>BV</i></b>	440.40	15.48	502.91	28.67	1.10x10 <sup>-12</sup>	
<i>vCSF</i>	2.39	0.98	3.43	1.25	0.01	
<i>eCSF</i>	26.70	15.60	29.84	19.18		
<b><i>TIV</i></b>	469.49	15.34	536.19	28.33	7.30x10 <sup>-14</sup>	
<b><i>b</i></b>						
<i>amygdala</i>	12.51	1.19	14.55	1.23	1.60x10 <sup>-6</sup>	
<i>anterior commissure</i>	1.46	0.13	1.58	0.14		
<i>basal forebrain and septum</i>	13.42	0.49	15.78	0.59	1.15x10 <sup>-20</sup>	
<i>brainstem</i>	56.52	1.83	60.08	2.63	9.81x10 <sup>-6</sup>	3.39x10 <sup>-6</sup>
<i>central GM region</i>	15.32	0.54	17.39	0.65	2.09x10 <sup>-16</sup>	
<i>cerebellum</i>	64.21	4.59	61.71	5.60		1.40x10 <sup>-11</sup>
<i>corpus callosum &amp; external capsule</i>	17.70	1.19	18.88	1.91		
<i>fimbria</i>	3.55	0.44	4.05	0.49	5.12x10 <sup>-3</sup>	
<i>globus pallidus</i>	4.23	0.34	4.57	0.30	5.93x10 <sup>-3</sup>	
<i>hippocampus</i>	29.09	1.01	31.44	1.35	4.63x10 <sup>-8</sup>	1.06x10 <sup>-4</sup>
<i>hypothalamus</i>	12.26	0.49	14.30	0.61	1.24x10 <sup>-17</sup>	
<i>inferior colliculus</i>	7.50	0.57	8.12	0.41	4.31x10 <sup>-4</sup>	
<i>internal capsule</i>	5.48	0.60	5.40	0.60		1.21x10 <sup>-3</sup>
<i>midbrain (remainder)</i>	4.97	0.40	5.72	0.38	4.08x10 <sup>-8</sup>	
<i>neocortex</i>	136.02	8.17	145.78	9.68	4.36x10 <sup>-3</sup>	2.53x10 <sup>-3</sup>
<i>olfactory bulb</i>	26.37	1.85	25.19	1.46		1.26x10 <sup>-13</sup>
<i>striatum (caudate putamen)</i>	26.84	1.20	30.72	1.52	4.17x10 <sup>-13</sup>	
<i>superior colliculus</i>	9.36	0.62	10.97	0.80	9.52x10 <sup>-10</sup>	
<i>thalamus</i>	26.44	0.87	27.61	1.45	0.02	7.85x10 <sup>-8</sup>
<i>ventricles</i>	1.65	0.22	2.22	0.45	8.21x10 <sup>-6</sup>	4.56x10 <sup>-2</sup>

**Table 4.1** Parcellated volume results for Tc1 and WT groups.

Mean absolute volumes (mm<sup>3</sup>) of (a) probabilistic tissues: BV=GM+WM; TIV=BV+CSF and (b) parcellated regions via integration of Jacobian determinants, and their standard deviations (SD). (Bonferroni-adjusted two-tailed p-values shown, omitted where >>0.05).



**Figure 4.3** Parcellated WT and Tc1 region volumes from Jacobian integration.

Illustrating the results from [Table 4.1b](#). Error bars show  $\pm 1$  SD. The scale is  $\text{mm}^3$ .

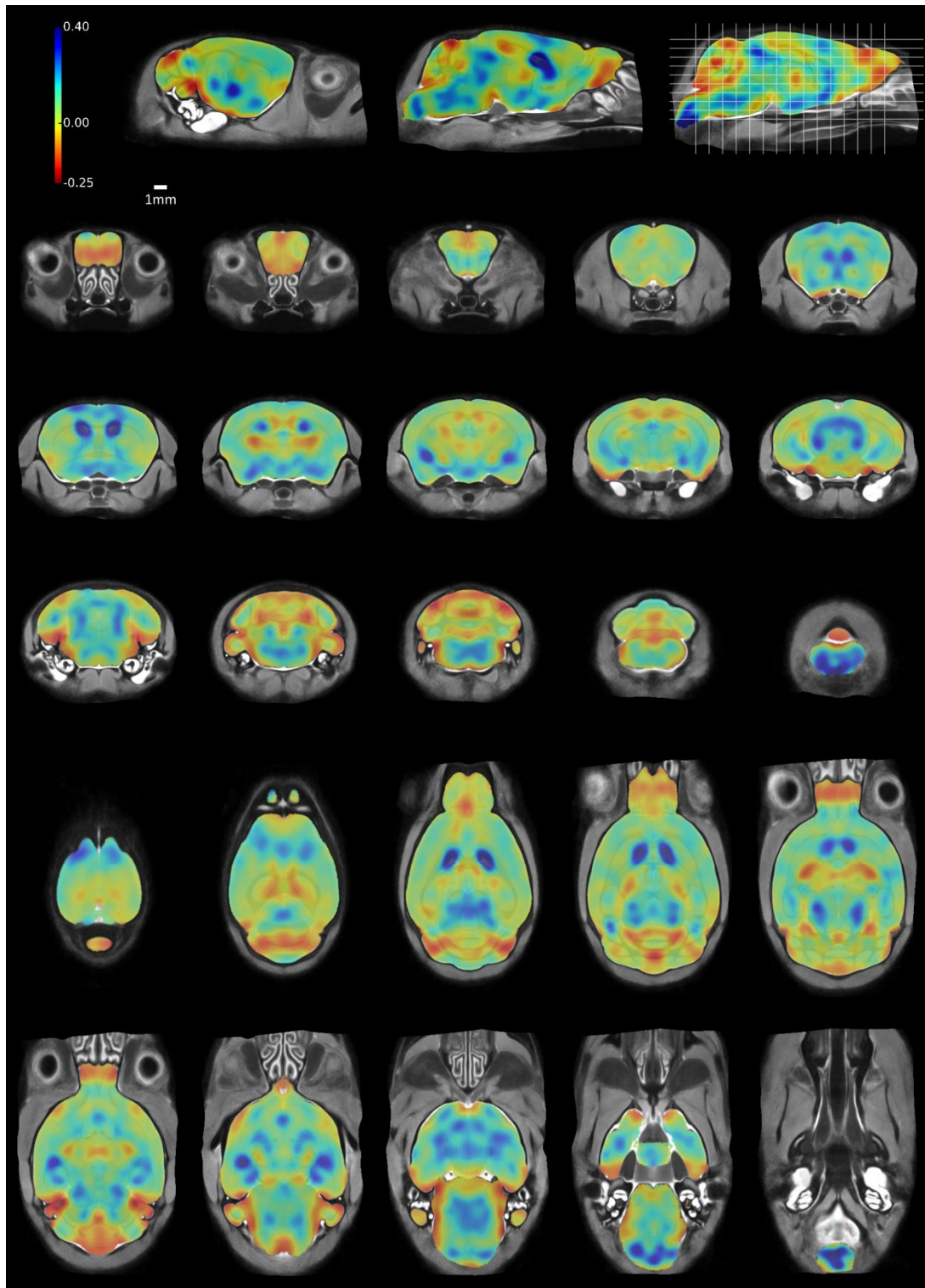
#### 4.1.6 Voxel-wise analysis

[Fig 4.4](#) shows, at each voxel, the mean proportional volume difference of the Tc1 group compared to the mean original WT volume, controlling crudely for TIV. To create this figure, I found the mean  $J_{det}$  values at each voxel (after smoothing), for both the WT and Tc1 groups, subtracted the mean WT values from the mean Tc1 values, and

calculated this difference as a proportion of the mean WT  $J_{det}$ . To remove the effects of global scaling, I divided this figure at each voxel by the relative proportion of Tc1 to WT mean TIV (E4.4). Each voxel  $v$  is given by:

$$E4.4 \quad v = \frac{mean(J_{det})_{Tc1} - mean(J_{det})_{WT}}{mean(J_{det})_{WT}} \bigg/ \frac{mean(TIV)_{Tc1}}{mean(TIV)_{WT}}$$

The illustration is comparable to TBM results, except without statistical tests or significance thresholding. It provides an overview of the volume differences between groups, using the WTs as a baseline control, and shows where statistically significant results may be expected, given sufficient power. Note the obvious bilateral expansion of the ventricles (blue) in the Tc1s, and the shrinkage of the olfactory bulbs, cerebellum, and hippocampus. The subsequent statistical tests of TBM help to determine the significance of these volume differences.



**Figure 4.4** Mean Tc1 voxel-wise volume difference from mean WT volume.

A semi-transparent per-voxel map after smoothing and correction for the mean relative difference in TIVs. Red: Tc1s originally smaller than WT; blue: larger. Sagittal (first row); coronal (rows 2-4) and transverse (rows 5 and 6) slices overlaid on the final group-wise structural average. 0 represents no difference; 0.1 where Tc1s are, on average, 10% larger

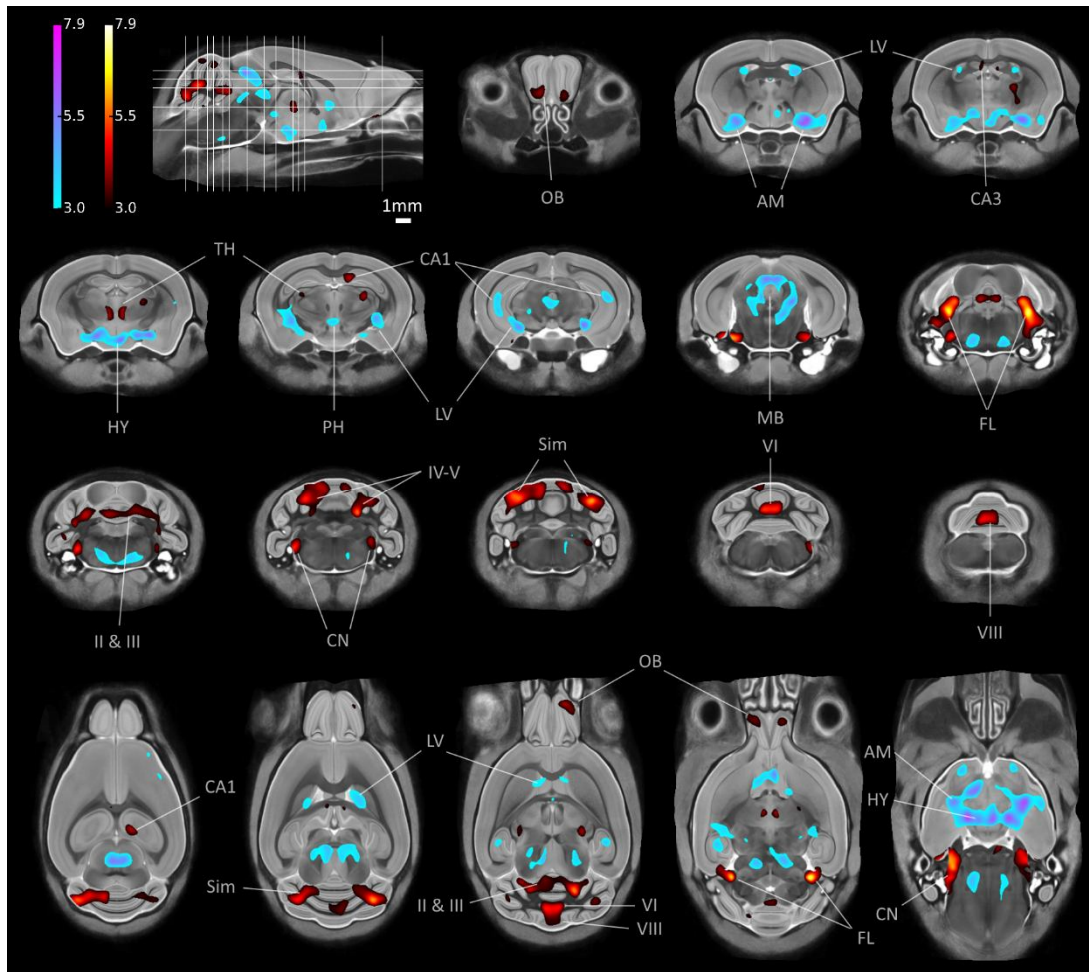
than WTs. Comparable to Fig 2 of *Hua et al. (2008)*, who for human AD showed “percentage reduction in volume relative to controls” prior to voxel-wise statistical significance testing.

#### 4.1.7 Tensor-Based Morphometry

I used TBM to highlight local volume differences between groups, by including TIV as a confounding factor in the GLM. *Fig 4.5* shows representative slices through the final average image after GWR, with significant voxels overlaid. To localise clusters, I referred to the parcellated labels and *Paxinos & Franklin (2012)*.

TBM revealed distinct local volume differences in the Tc1 brains compared with WTs. There were bilateral regions of localised expansion in the amygdala, lateral ventricles and hypothalamus. The reticular nucleus, superior colliculus, and periaqueductal grey regions of the midbrain also showed expansion, possibly secondary to that of the ventricular aqueduct and fourth ventricle. Unexpectedly, the hippocampus showed a degree of bilateral, localised enlargement in the Tc1 group, in CA1.

I observed significant bilateral reductions in local volume in the olfactory bulbs; the CA3 region of the hippocampus, rostrally; two distinct regions of the thalamus (the rhomboid nucleus and the dorsal sensory-motor region); and in the brainstem, the cochlear nuclei of the medulla. There was a unilateral reduction in the hippocampus medially in CA1. The most prominent volume reductions occurred throughout the Tc1 cerebellum, including bilaterally in the flocculi; the central lobules (II and III) of the cerebellar vermis; the simple lobule and culmen (lobules IV-V); and medially in declive VI and pyramus (VIII). See §4.1.8 for further investigation of significant regions.



**Figure 4.5** Tc1 TBM results

FDR-corrected ( $q=0.05$ ) t-statistics overlaid on slices (locations indicated top left) of the final structural GWR average. Blue: Tc1 group locally statistically significantly larger than WTs; red: Tc1s smaller.

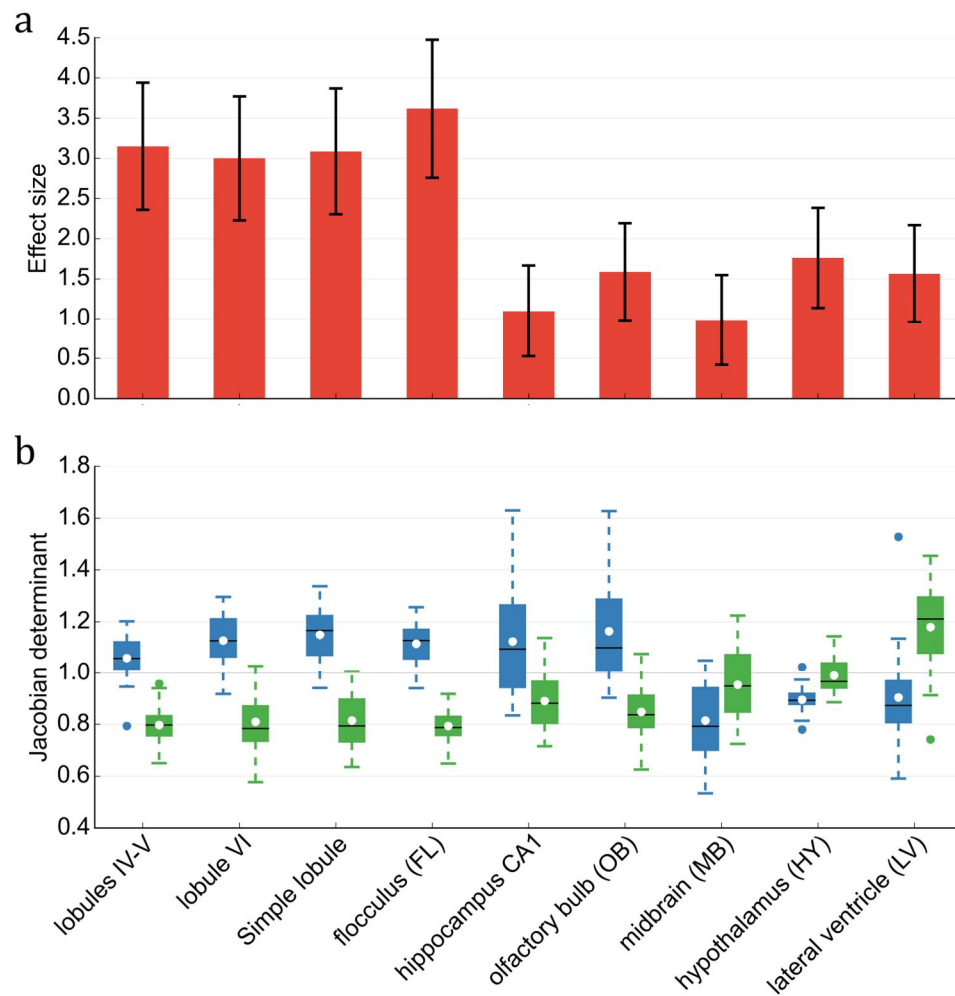
AM: amygdala; CA1, CA3: hippocampal sub-regions; CN: cochlear nucleus; FL: flocculus; HY: hypothalamus; LV: lateral ventricles; MB: midbrain; OB: olfactory bulb; PH: posterior hypothalamic nucleus; cerebellar lobules II & III; IV-V (culmen); declive VI and VIII (pyramus); Sim: simple lobule; TH: thalamus.

#### 4.1.8 Individual voxel analysis of TBM results

$J_{det}$  values at peak absolute t-statistic locations in selected regions (indicated in [Fig 4.5](#)) are shown in [Fig 4.6](#). These help to quantify the degree of group separation in different regions. They are also useful for directing potential follow-up histological investigations, which are likely to be constrained, by costs or time, to focus on fewer brain regions.



In the midbrain and hypothalamus, where the Tc1s were locally significantly larger than the WTs, the mean difference was relatively small compared with other regions. Of the regions shown, the midbrain had the lowest effect size (E5.3; discussed further in §5.7.3), thanks to the relatively large standard deviation of  $J_{det}$  values in this region, likely due to the structural variability of the fourth ventricle and aqueduct. Despite exhibiting the smallest difference between means, the peak t-statistic voxel within the hypothalamus had a moderate effect size, thanks to a low WT  $J_{det}$  standard deviation in this region.



**Figure 4.6** Peak t-statistic voxel values within selected significant clusters

Referring to significant clusters in Fig 4.5. (a) Effect sizes (Cohen's  $d$ ) with 95% confidence intervals; (b)  $J_{det}$  values for WT (blue) and Tc1 (green) groups. To show the effect independent of global volume, values were divided by TIV then multiplied by mean WT TIV.

### 4.1.9 Voxel-Based Morphometry

VBM highlights local differences in GM tissue proportion, as measured by the segmentation. This may be thought of as a representation of tissue density. Owing to the global volume difference between groups, modulated VBM (§3.11) would have required a TIV covariate in the GLM when performing voxel-wise statistical tests<sup>2</sup>; I show unmodulated VBM results here. The GM maps had no dependence upon  $J_{det}$  values and hence no volume component. I therefore excluded the TIV covariate.

[Fig 4.7](#) shows representative slices through the final GWR average image, with FDR-corrected t-statistics overlaid ( $q=0.05$ ). I smoothed with a Gaussian kernel, FWHM 0.16mm (4 voxels). To better anatomically localise some significant regions and improve spatial specificity, I also tested a 0.02mm smoothing kernel (0.5 voxels, [Fig 4.8](#)).

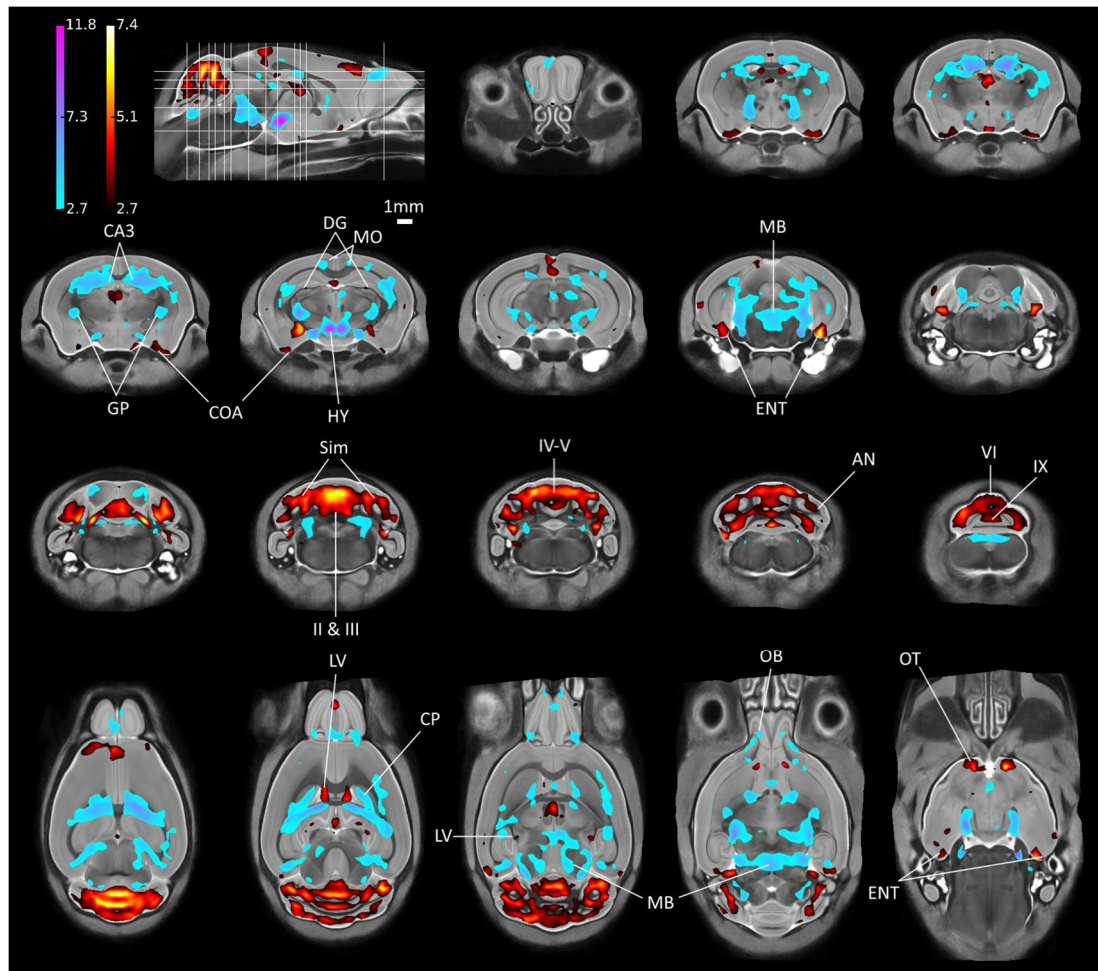
VBM detected reduced GM in the dentate gyrus (DG) hippocampal region, and bilateral increases in GM proportion in the olfactory bulbs, thalamus, hypothalamus, midbrain, globus pallidus, motor cortex and the CA3 region of the hippocampus. The mouse midbrain, especially adjoining the pons and brainstem, exhibits a high degree of GM/WM mixture: most voxels have some PV. Additionally, this region has few high-contrast features, which may have impeded internal registration accuracy, possibly contributing to the apparent increases in GM proportion seen here. There were few significant voxels in this region after application of the narrower smoothing kernel (see [Fig 4.8](#)), suggesting this regional apparent increase in GM proportion is sparse and nonspecific.

GM regions adjacent to the ventricles in the final average image, such as the septal nucleus, show reduced GM density in the Tc1s, likely due to ventricular expansion encroaching into GM tissue.

---

2: As discussed in §2.2.7, with high-dimensional warping in NRR, modulated VBM approximates TBM: the volume differences between groups dominate any intensity differences. Indeed, results of modulated VBM (not shown) highlighted similar locations as TBM results.

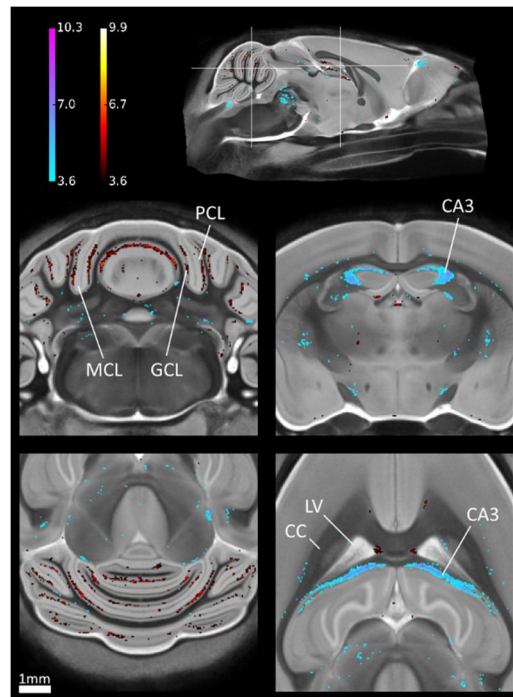
Bilateral regions of decreased GM density were detected in the dentate gyrus region of the hippocampus, the entorhinal cortex, and the olfactory tubercle. The cerebellum showed decreases in GM proportion throughout, particularly lobules II & III, IV-V, VI and IX, and the simple lobule. Upon inspection with a narrower smoothing kernel (*Fig 4.8*), these regions appeared predominately within the granule and Purkinje cell layers.



**Figure 4.7** Unmodulated VBM results

FDR-corrected ( $q=0.05$ ) t-statistics overlaid on coronal and transverse slices (locations indicated top left: same as *Fig 4.5*) of the final structural average. Blue: statistically greater local proportion of GM in Tc1 group; red: reduced.

CP: caudate putamen; ENT: entorhinal cortex; GP: globus pallidus; MO: motor cortex. Cerebellar regions: AN: ansiform lobule.



**Figure 4.8** VBM in the cerebellum and hippocampus, with a narrower smoothing kernel.

As [Fig 4.7](#). Sagittal, coronal and transverse views, employing a FWHM 0.02mm Gaussian smoothing kernel prior to statistical tests. Significant voxels are largely confined to the Purkinje and granule cell layers of the cerebellum and CA3 region of the hippocampus. Note few significant voxels in the midbrain.

CC: corpus callosum; GCL: granule cell layer; MCL: molecular cell layer; PCL: Purkinje cell layer.

#### 4.1.10 Discussion: morphometry of Tc1 brains

I identified several phenotypes in the Tc1 brain using the automated pipeline, the first such morphometric analysis of a DS brain model compared with WT's, and revealed local differences not distinguishable between samples by eye, histology, or segmentations alone.

In human DS, BV is reduced by around 18% ([Aylward et al., 1999](#); [Pinter et al., 2001a](#); [White et al., 2003](#)). Whereas the whole mandible is reduced in size in human DS, [O'Doherty et al. \(2005\)](#) measured only partial mandible reduction in Tc1s, and no craniofacial malformation or reduction in skull size. Distances between landmarks in

μCT images of Tc1 skulls also exhibited no significant size difference compared with WT<sub>s</sub>. I observed significantly increased Tc1 TIV, BV, GM and WM volumes ([Table 4.1a](#)).

This unexpected global finding was consistent in C<sub>1</sub> and C<sub>2</sub> – scanned independently and with considerations for gradient scaling – and indicates the utility of whole-brain MR and tissue segmentation over histology and landmark measurements, which are necessarily localised and limited to a few subjects and by rater variability.

All tissues and most parcellated regions displayed greater volume variance in the Tc1s. [O'Doherty et al. \(2005\)](#) reported that approximately 66% of Tc1 brain nuclei retain Hsa21. This mosaicism means Hsa21 levels are unpredictable and vary between organs, mice, and with background, leading to phenotypic variation ([Reeves, 2006](#)). [Olson et al. \(2004\)](#) noted: “most DS phenotypes are incompletely penetrant and variable in expressivity – the mechanism(s) by which increased gene dosage causes any specific DS feature is not established”. This could potentially be addressed using morphometry with more selective trisomic animal models.

The ventricles are enlarged in human DS ([White et al., 2003](#)) and both VBM and TBM detected their bilateral enlargement in the Tc1s. This is likely underestimated in *ex vivo* brains, which shrink slightly during fixation as tissues relax and ventricles partially collapse (§5.7; [Zhang et al., 2010](#)). Although mitigated in-skull ([Sawiak et al., 2013](#)), one would expect the same systematic effect across groups. Brain expansion due to hydrocephalus has been reported in a DS model ([Yu et al., 2010](#)); however, these mice died by 10 weeks of age, and the Tc1s did not exhibit gross ventricular enlargement of the same degree. (Also, only 6.5% of those mice exhibited hydrocephalus. They had rounded and enlarged skulls. Although I investigated volume, I did not quantify the shape or roundedness of the Tc1 brains. This would be possible, however, using their principal axes, as calculated in §3.3.) Overall vCSF volumes did differ significantly between groups before and after correction for TIV. However, this measurement is likely an underestimate, due to unpredictable ventricular collapse. Upon visual inspection, one brain exhibited uncollapsed 3<sup>rd</sup> and 4<sup>th</sup> ventricles absent of CSF (and therefore appearing

dark), however, this came from a WT littermate. As the ventricular spaces were present without bright vCSF, segmenting this brain required manual attention.

That the UFL atlas-segmented ventricles were significantly different in volume prior to TIV normalisation is likely a result of the atlas registering poorly in these regions, again owing to ventricular collapse in our data. The only tissue segmentation-derived region to exhibit a significant difference in volume surviving Bonferroni correction was GM/WM mixture (not shown). This may be due to increased PV in the Tc1 group, or imprecision in boundary delineation of this PV class (*Fig 3.15*); corresponding parcellations (broadly, the caudate putamen/striatum; midbrain; superior and inferior colliculi) showed differences. As noted in §3.6, for these reasons of imprecision, I later removed this measurement and replaced it with more smoothly-varying GM and WM segmentations, which better account for PV.

One may safely conclude that the Tc1 mouse exhibits an enlarged brain, in contrast to people with DS. Tc1 mice have some rearrangements of their copy of Hsa21 (*Gribble et al., 2013*), and it is possible that this contributes to the megaly phenotype; there may also be some non-specific interaction of the human chromosome in the mouse cells that gives rise to larger brains. Finding relatively greater subcortical GM volumes in DS patients compared to the total GM volume, and preservation of GM volume in the parietal cortex, *Pinter et al. (2001a)* suggested that greater basal ganglia and thalamus volumes in humans may result from insufficient apoptosis.

In humans and mice with deletion or truncation of the Hsa21 gene DYRK1A<sup>3</sup>, brain size and weight is reduced (*Sebrié et al., 2008*; *Guedj et al., 2012*). Through phosphorylation, this gene is thought to be tied to many DS phenotypes, and it is likely modulated by the presence of other genes (*Wiseman et al., 2009*). It is dose-dependent and hence, in humans and Tc1 mice, it is overexpressed. In two mouse models<sup>4</sup> of partial trisomy, overexpressing the DYRK1A gene, *Sebrié et al., (2008)* and *Guedj et al. (2012)* found increased brain size (measured via MRI, weight and histology). The thalamus, midbrain

---

3: dual-specificity tyrosine-phosphorylation-regulated kinase 1A.

4: hYACtgDyrk1a and mBACtgDyrk1a.

and colliculus were preferentially affected. In the thalamus, neuronal density and number increased, while both neuron size and extracellular space decreased. Conversely, cell density was negatively correlated with DYRK1A dosage in the hippocampus and somatosensory and entorhinal cortex (*Guedj et al., 2012*). This may underlie my TBM results, which showed expansion of the Tc1 midbrain and central thalamus. *Guedj et al.* noted that Ts65Dn mice, also with three copies of DYRK1A, do not exhibit elevated BVs, and that other genes may compensate.

In subsequent voxel-wise statistical tests I controlled for TIV to reveal differences in the Tc1 group independent of the total volume increase. TBM detected significant bilateral, local volume reductions in the olfactory bulbs. *Bianchi et al. (2014)* recently observed, via histology, impaired neurogenesis in the olfactory bulbs of 13-month-old Ts65Dn mice, and remarked that this may parallel the loss of smell in older human DS individuals. The reductions seen here suggest there may be a similar functional impairment in Tc1 mice.

I found significant local reductions in GM volume within the cerebellum, using both TBM and VBM, focussed medially in Declive VI, as well as unilaterally within the simple lobule and lobules 4/5, in both granular and molecular cell layers.

Cerebellar GM was reduced in a VBM study of non-demented people with DS, and exhibits reduced overall volume compared with TIV in humans (*Raz et al., 1995; White et al., 2003*) and the Ts65Dn, Ts1Cje and Tc1 mouse models (*Ma et al., 2014; Olson et al., 2004*). My TBM analysis reveals that rather than the cerebellum being uniformly reduced in volume, reductions have discrete local foci. There is evidence cerebellar lobules have distinct functional correlates (*Stoodley & Schmahmann, 2010*). It may be possible to map local volume reductions to functional topography and hence to behaviour in Tc1 mice.

The cerebellum is associated with fine motor control and cognitive processes. In children with DS, cerebellar hypoplasia is implicated in motor and speech difficulties (*Pinter et al., 2001a*). *Galante et al. (2009)* found motor learning and coordination deficits in Tc1 mice. Histological staining revealed reduced internal granule layer density in the Tc1



cerebellum compared with WT, mirroring observations of the Ts65Dn and Ts1Cje models (Baxter *et al.*, 2000; O'Doherty *et al.*, 2005; Olson *et al.*, 2004). I repeated these findings with VBM, showing reduced GM density in the granule cell layer of several lobules (Fig 4.7, 4.8). This supports the utility of VBM for informing histology. VBM also showed bilateral reductions in GM density in the entorhinal cortex, recapitulating the progressive atrophy of this region in human DS (Teipel *et al.*, 2004).

VBM also detected reduced GM proportion in the dentate gyrus (DG). Long-term potentiation in the DG – synaptic plasticity thought to be directly related to long-term memory – was found to be reduced in Tc1 mice (O'Doherty *et al.*, 2005), and behavioural observations demonstrated reduced spatial working memory (Morice *et al.*, 2008). I also observed bilaterally elevated GM proportion in the CA3 region of the Tc1 hippocampus. Insausti *et al.* (1998) measured elevated neuronal numbers in Ts65Dn CA3, and suggested this may compensate for reductions in DG, although Kurt *et al.* (2004) found normal neuron density, but reduced synapse density, in both structures. Witton *et al.* (2015) also recently showed decreased synapse density in the DG, and related this to the poorer performance of Tc1 mice in a radial arm maze, compared with WT. These cellular changes could underpin the differences in VBM GM signal observed here.

Mouse brain tissue classification is complicated by smaller structures and greater PV proportion than is found in humans. Both may be mitigated using higher field strengths, enabling greater SNR, spatial precision, and contrast (Natt *et al.*, 2002). Structural differences between the *in vivo*, skull-stripped NUS atlas used for tissue segmentation and this *ex vivo*, in-skull data were resolved with NRR. By employing this atlas, and explicitly modelling background, external tissues, PV and CSF, my tissue classifications included fine WM detail, including the deep WM of the cerebellum and PV regions such as the striatum and midbrain. I averted misclassifications which have befallen previously published TPMs, including the presence of a brain-enveloping 'rim' where GM PV is misinterpreted as WM (Fig 2.8).



#### 4.1.11 Comparison with earlier results

An earlier version of this pipeline, which excluded tissue segmentation and was thus reliant upon TIV measured from STEPS masks, produced less specific TBM results – similar to those of *Sinclair et al., (2011)*. As *Fig 3.14* shows, the masks exhibit good accuracy, but variability at the brainstem could substantially alter their volume. As TIV was thus imprecisely controlled in the GLM, some global voxel expansion (from rigid and affine stages of GWR) was included. Comparison with the results of *Sinclair et al., (2011)* indicates that these TBM maps are highly sensitive to the TIV covariate: greater accuracy (through measurement of TIV via probabilistic GM, WM and CSF volumes) improves the spatial specificity of morphometric results. For further discussion, see *§5.6*.

#### 4.1.12 Corroboration of V/TBM with cohort 2

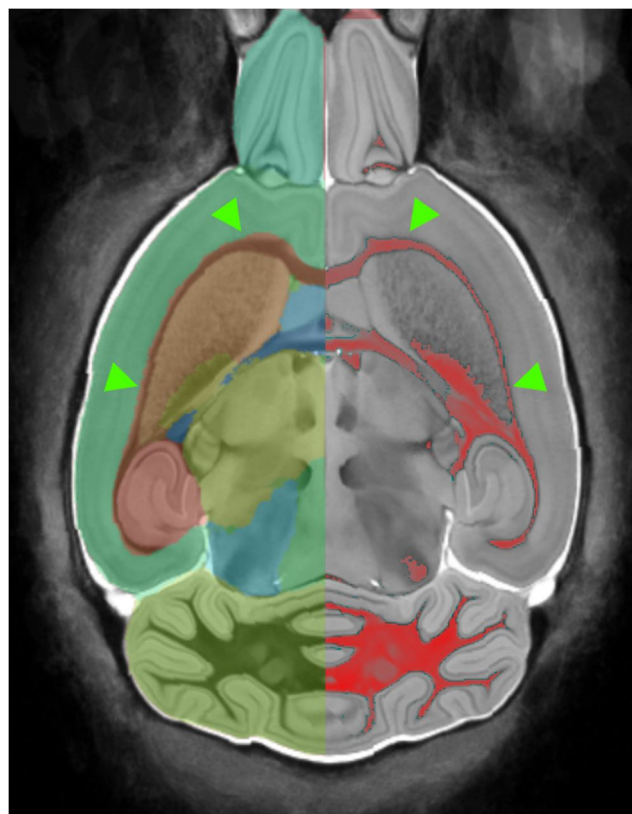
The ‘Tc1xJ20’ group (*§4.2*), from which C<sub>2</sub> brains were taken, allowed further validation of the results revealed above using V/TBM. Encouragingly, although not reported here, when the full pipeline and TBM was performed in C<sub>1</sub> and C<sub>2</sub> separately, very similar statistical maps were observed. However, C<sub>1</sub> results were mostly confined to the cerebellum and there were few significant voxels after FDR correction at  $q=0.05$ . A similar pattern as that reported here survived correction at  $q=0.1$ . C<sub>2</sub> was included to improve statistical power. C<sub>2</sub> results alone are shown in *§4.2 (Fig 4.12)*.

#### 4.1.13 Comparison with parcellation

Employing STEPS-based parcellations with the UFL atlas, *Ma et al. (2014)* reported significant shrinkage of the cerebellum and olfactory bulbs in Tc1 brains relative to BV, which was found to be significantly increased, but no other structures were significantly different in volume. Controlling for TIV, I repeated these findings using Jacobian integration, and additionally found the internal capsule to be reduced in size, possibly thanks to the increased contrast the group average image provides over individual scans; this thin structure is not easily segmented (*Fig 4.9*). Additional structures were also found to have volume differences here (*Table 4.1*). This increased sensitivity may be due to this

study's incorporation of  $C_2$ , in order to increase numbers and thereby boost statistical power. The earlier paper's use of STEPS to segment each image individually, rather than indirectly via a group average, may have improved its sensitivity.

Segmentation via parcellation enables volume- and shape-based analysis of substructures, but is limited in specificity by atlas detail. TBM here localised the contributory regions of difference *within* those structures. The noted reductions in the olfactory bulbs and cerebellum will have contributed to the earlier global volume findings. Furthermore, I observed local changes within the hippocampus, thalamus, hypothalamus, midbrain and ventricles undiscerned by segmentation alone. This surrogate biomarker is more useful for informing precise histological follow-up investigations.

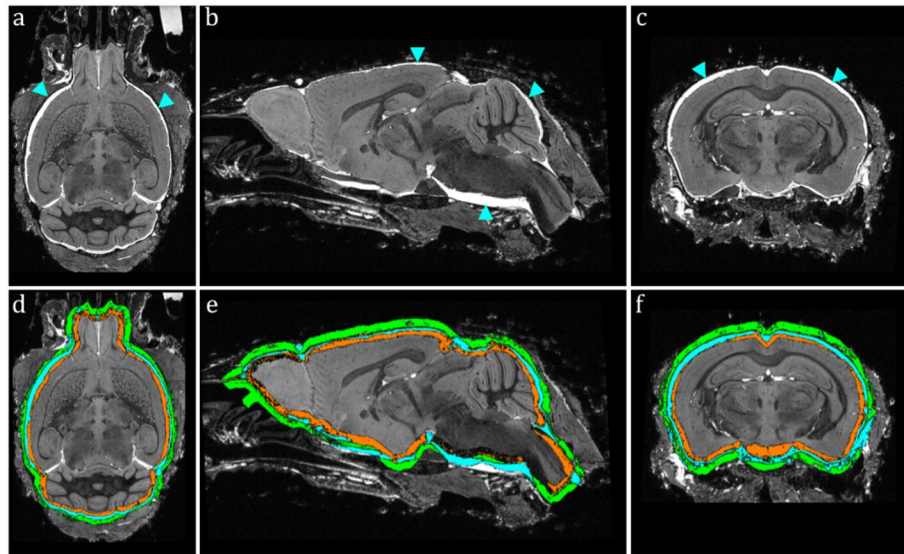


**Figure 4.9** Comparing label propagation with tissue segmentation.

Over-generous coverage of the corpus callosum (CC) by the propagated UFL atlas labels (semi-transparent, left) in the space of the final group-wise average of the Tc1 and WT brains (cropped transverse view). The WM segmentation (red, right, here binarised at 0.5) appears to cover the CC with greater accuracy.

Imprecise registration between the external atlas and the Tc1 data resulted in over-generous volumes for some regions, such as the corpus callosum (*Fig 4.9*). Unfortunately, few high-quality mouse atlases are presently available. Currently-available multi-subject atlases (UFL, NUS) are limited by relatively low contrast and resolution compared with those achievable at high field strengths; many opportunities exist for extending their size and quality. An atlas database with more subjects, parcellations (including non-brain) and finer detail would aid the specificity of segmentation-based approaches, and could complement these V/TBM results by allowing more precise anatomical localisation of significant voxels – for example, nodules of the cerebellum. I expect results to improve further with the increased availability of such atlases.

#### 4.1.14 Hyperintense rim



**Figure 4.10** Example of hyperintense rim, and its segmentation, in the Tc1 dataset. Transverse, sagittal and coronal views of a single Tc1 brain (same slices, both rows) exhibiting a region of high signal around some parts of the brain, indicated with light blue arrows (a,b,c). This was thought to be a perfusion artefact. I used an EM scheme to automatically classify the rim (light blue in d,e,f), in order to measure its thickness. Orange: brain GM; green: external material and BG.

I noted a hyperintense ‘rim’ partially enveloping most  $C_1$  brains (*Fig 4.10*). *Cabill et al. (2012)* suggested a similar artefact (“a blister of liquid beneath the dura mater”) measuring up to 500 $\mu$ m in thickness resulted from improperly dissolved perfusate, or a high perfusion rate, causing blockages and pressure build-up in the vasculature, leakage into CSF cavities, and resulting in ‘bubbles’ of high intensity perfusate forming between meninges and brain tissue. This could cause compression and hence, potentially, unpredictably affect registration, and render TBM unreliable. I therefore sought to characterise this artefact in the data.

$C_2$  brains were perfuse-fixed in CABI using a standardised protocol: commercial formal-buffered saline (VWR International Ltd., England), together with a low flow rate (2.5-3 ml/min) (*Cleary et al., 2011b*).  $C_2$  brains did not exhibit the artefact.  $C_1$  brains were perfuse-fixed at a different site, at 3ml/min, where powder PFA was mixed with the same commercial solution.

CSF is also hyperintense in  $T_2^*W$  images, and may become trapped during perfusion-fixation. I took the ‘rim’ volume as the eCSF volume from tissue segmentation. As noted in §3.6, this involved the addition of external labels, to incorporate eCSF, BG and external tissues into the EM scheme. A further refinement of this method is described in §6.2.3. Visual inspection ensured all parts of the rim were included. eCSF occurs naturally, is particularly prominent in models of brain atrophy (*Wells et al., 2015*), and pools in the cisterns. After its binary segmentation, my colleague Da Ma measured its 3D thickness using a Laplacian field-based algorithm typically applied to the cortex (*Ma et al., 2015*).

In contrast to *Cabill et al. (2012)*, I found a positive correlation between eCSF volume and BV ( $r=0.449$ ,  $p=0.017$ ), suggesting the rim in  $C_1$  did not compress the brain. It was most prominent within the interhemispheric fissure of the cortex, and never completely enveloped the brain. In  $C_1$ , mean (SD) eCSF volume was 44.13 $\mu$ L (8.38 $\mu$ L), 9.53% of average BV. In  $C_2$ : 12.0 $\mu$ L (4.08 $\mu$ L), 2.48% of mean BV, and no eCSF volume exceeded 23.7 $\mu$ L. There was no appreciable correlation between eCSF volume and BV in  $C_2$  ( $r=0.02$ ,  $p=0.92$ ). The mean thickness of the eCSF layer for all  $C_1$  brains was 162 $\mu$ m

(75 $\mu$ m). Thickness was only measured where it was greater than 1 voxel (40 $\mu$ m); as the rim was never fully enveloping, the mean over the entire brain surface will thus actually be much lower. As these measurements reveal the artefact to be less severe, and do not show the correlations reported by *Cahill et al. (2012)*, and additionally as I combined  $C_1$  and  $C_2$  during group registration, reducing the relative contribution of brains with the artefact to the structural average and to statistics, I do not believe it denigrated V/TBM or volumetric results, or made registration unreliable.

#### 4.1.15 Conclusions

My statistical morphometric analysis identified novel phenotypes in this first transchromosomic animal model of DS. Overall Tc1 brain volume was unexpectedly elevated and, controlling for this, I found local volume reductions in the cerebellum, olfactory bulbs, and other brain regions. I also found GM density reductions within the cerebellum, consistent with previous histological findings in this model, and human DS. Several newly identified regions, such as the dentate gyrus, specific cerebellar lobules, and the olfactory bulbs, may warrant additional histological follow-up, or suggest behavioural investigations.

## 4.2 Tc1xJ20 cross-breed study

### 4.2.1 Introduction

As introduced in §2.4, Down syndrome (DS), the leading genetic cause of intellectual disability, is caused by trisomy of human chromosome 21 (Hsa21). Individuals with DS have a greater predisposition to Alzheimer's disease (AD) in later life, thought to be due to increased dosage of the amyloid precursor protein (APP) gene that maps to Hsa21 – an established risk factor for AD (*Oliver & Holland, 1986*). Analysis of mouse models of DS improves our understanding of the significance of trisomy of Hsa21, and its relationship to AD (*Roizen & Patterson, 2003*).

The Tc1 mouse (*O'Doherty et al., 2005*) and the J20 mouse (*Mucke et al., 2000*) are, respectively, established models of DS and A $\beta$  deposition in AD. As previously discussed (§4.1), the Tc1 mouse carries an almost complete copy of Hsa21, but is not functionally trisomic for APP. It thus allows the study of the DS phenotype without the effects of APP trisomy. The J20 mouse expresses a mutant form of APP, as found in familial forms of AD (*Mucke et al., 2000*), and exhibits key features of human AD, including cognitive decline, synaptic loss and accumulation of amyloid plaques (§2.3). By cross-breeding Tc1 mice with J20 mice, progeny with four genotypes were generated: wild-type (WT); trisomic (Tc1); APP transgenic (J20); and double mutants: trisomic and APP transgenic (Tc1xJ20): a novel model of amyloidosis in AD, in the presence of DS. Exaggerated behavioural changes and increased plaque load have already been observed in the Tc1xJ20 double mutant mouse.

I here show the results of a comprehensive morphometric analysis of these mouse brains, to elucidate structural differences between Tc1, J20 and WT, and between the Tc1xJ20 and J20 littermates. This latter comparison enables the study of trisomy Hsa21 alongside mutant APP, in the same animal, and thus testing whether DS modulates or accentuates the AD phenotype.

## 4.2.2 Methodology

### ***Animals and preparation***

Tc1 females were bred with J20 males (*O'Doherty et al., 2005*; *Mucke et al., 2000*) by collaborators at the UCL Institute of Neurology/MRC Prion unit (London, UK). 50 adult male offspring mice (15 Tc1, 14 J20, 8 Tc1xJ20, 13 WT littermates) were aged to 15 months and culled. Their brains were perfuse-fixed (20mL 0.9% saline, then 50mL 10% buffered formal saline with 8mM Magnevist) – as §4.1.2.

### ***Image acquisition and quality control***

For scanning, the plastic brain holder (§3.2.3) was used to secure 3 brains at a time, in skull, within a 50ml syringe filled with proton signal-free, non-viscous Fomblin. To avoid any systematic effects from scanner distortion or gradient calibration affecting any one

group, genotypes were mixed randomly in each scan. Optimised protocols for high-resolution structural *ex vivo*  $\mu$ MRI mouse brain phenotyping were employed by my colleague Holly Holmes at CABI (Cleary *et al.*, 2011b) using the same parameters as for cohort 2 in §4.1.2.

Upon visual inspection, one brain, a WT, was found to have a partially collapsed skull, unilaterally, causing compression of the cortex by about 5mm. As this would likely have influenced morphometry results, this brain was excluded from further analysis. Images were adjusted for gradient scaling errors from the manufacturer calibration using the protocol described by O’Callaghan *et al.* (2014) and in §4.1.2.

### **Morphometric pipeline**

Images were all processed according to the TBM pipeline described in §3. All images in this study required multiple subject extraction. I performed tissue segmentation, for volume measurements, but not VBM. The statistical model used in the GLM was similar to E4.3, although I did not control for age or genetic background, as all animals were age-matched littermates:

$$E4.4 \quad Y = b_1P(WT) + b_2P(Tc1) + b_3P(J20) + b_4P(Tc1xJ20) + b_5TIV + \epsilon_i$$

where  $b$  are the regression coefficients;  $P(genotype)$  is the probability (0 or 1) of each animal being Tc1, WT, J20 or Tc1xJ20; and  $\epsilon$  is the residual error vector.

## **4.2.3 Results**

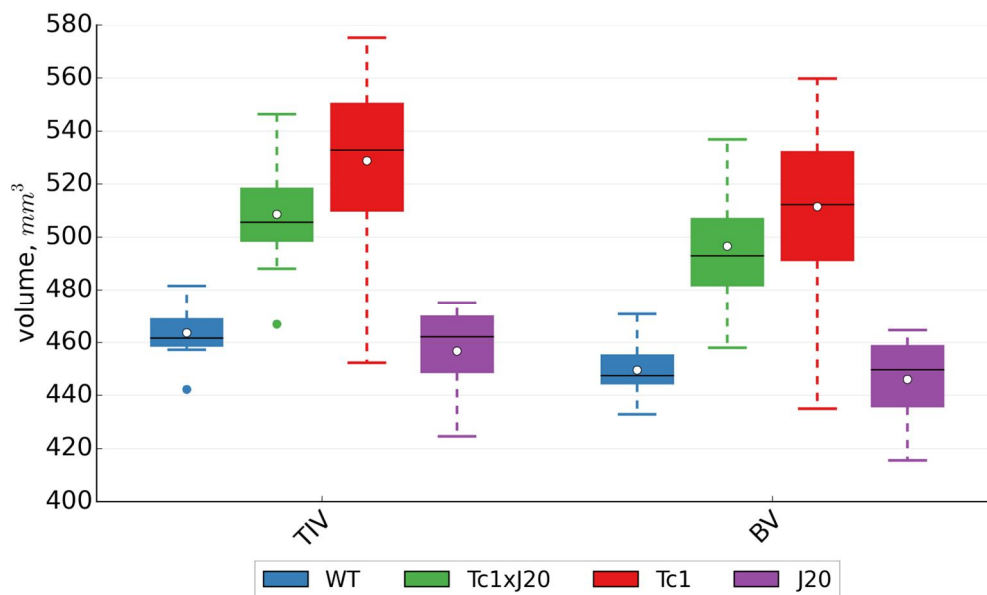
### **Segmentation**

Fig 4.11 and Table 4.2 show global intracranial volume results from the tissue segmentation performed across all brains in this study. (Note that the Tc1 and WT groups were included as part of §4.1 results and in Table 4.1, as cohort 2.)

	<i>WT</i> ( <i>n</i> =12)		<i>Tc1xJ20</i> ( <i>n</i> =8)		<i>Tc1</i> ( <i>n</i> =15)		<i>J20</i> ( <i>n</i> =14)	
	mean	SD	mean	SD	mean	SD	mean	SD
<i>GM</i>	312.28	6.61	346.58	24.67	360.35	26.95	304.60	11.37
<i>WM</i>	137.28	7.39	149.94	4.66	151.19	9.06	141.43	6.69
<i>BV</i>	449.56	10.16	496.52‡	27.29	511.54†	33.47	446.03◇	15.50
<i>eCSF</i>	11.97	4.54	9.72	3.02	13.66	3.65	9.13	2.22
<i>vCSF</i>	2.03	0.88	2.34	0.80	3.59	1.49	1.55	0.39
<i>TIV</i>	463.57	10.27	508.58‡	26.66	528.79†	33.06	456.71◇	15.84

**Table 4.2** Tissue volumes after segmentation in WT, Tc1, Tc1xJ20 and J20 mice.

For all groups, mean and standard deviation segmented tissue volumes ( $\text{mm}^3$ ) after EM segmentation, with BV and TIV (as [Table 4.1](#)). Two-tailed t-tests to compare BV and TIV means between group pairs gave Bonferroni-adjusted p-values  $< 0.01$ : † Tc1 vs. WT; ‡ Tc1xJ20 vs. WT; ◇ J20 vs. Tc1xJ20. (J20 vs. WT and Tc1xJ20 vs. Tc1: not significant, even at  $p < 0.1$ .)



**Figure 4.11** Box plots of TIV and BV for all four groups, from tissue segmentation. Values in [Table 4.2](#). Black bars represent medians; white dots means; outliers  $> 1.5$  IQR.

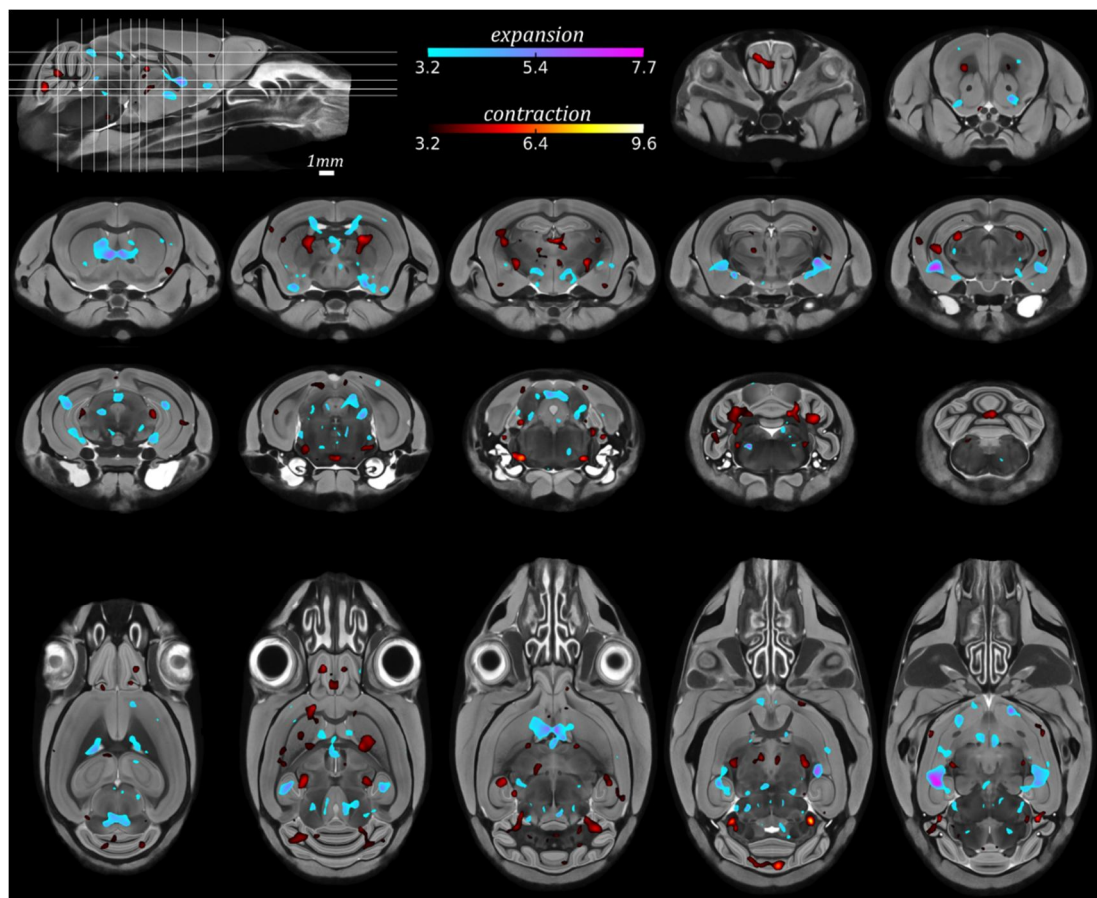
Tc1s were significantly greater in TIV than WT ( $p < 6.4 \times 10^{-6}$ ), a repeat of the unexpected finding from §4.1.5 in this data subset. J20s were significantly smaller than Tc1xJ20s ( $p < 1.1 \times 10^{-4}$ ) but not significantly different in volume, in any tissue, from WT. Tc1xJ20s were also not significantly different in volume from the Tc1 group. This may be due to



the considerable spread of the Tc1 group's volumes: apart from eCSF, which in this cohort was very low in volume thanks to the improved perfusion techniques discussed in §4.1.14, Tc1s had the greatest standard deviation in all tissue volumes.

### ***Tensor-based morphometry***

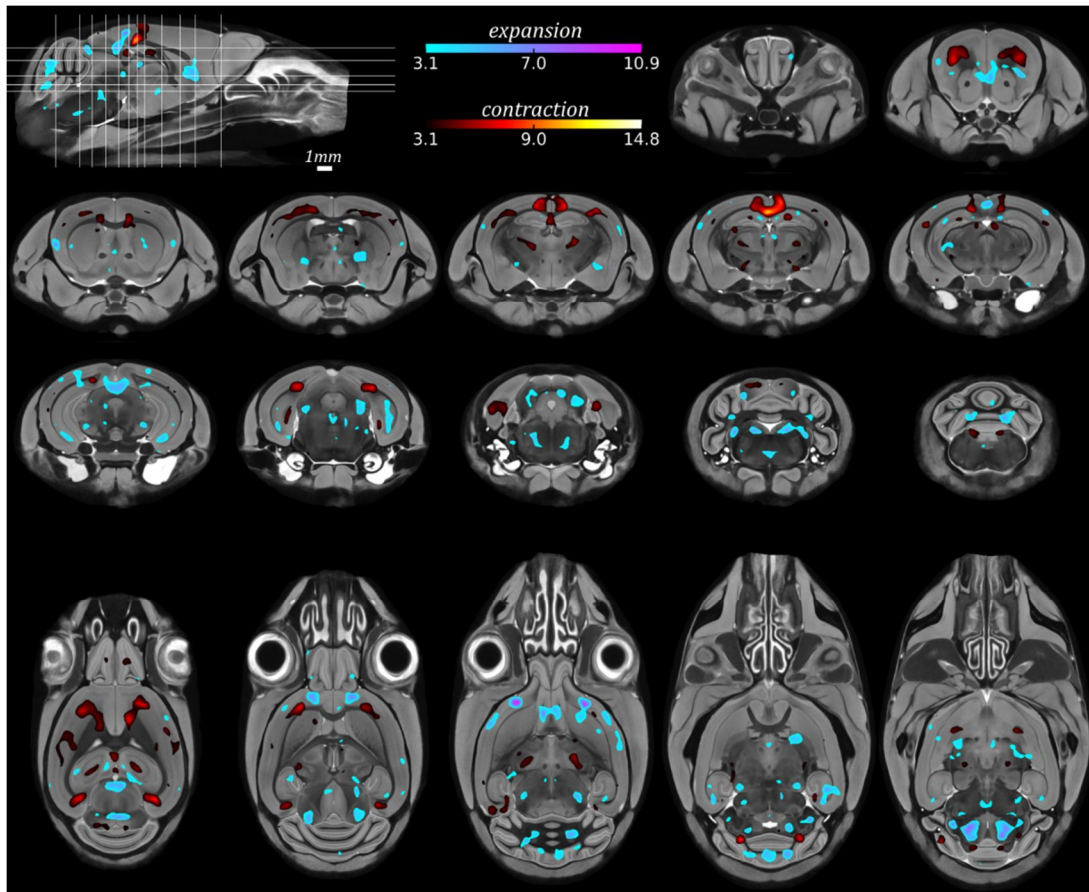
The TBM images show statistically significant, FDR-corrected ( $q=0.05$ ) t-statistics revealing local structural differences between the brains of each group, controlling for their respective TIVs.



**Figure 4.12** TBM results: Tc1 vs. WT (cohort 2)

Red: Tc1s locally smaller than WTs; blue: larger. Different slices are shown from [Fig 4.5](#). Many of the same regions of expansion and contraction are highlighted as reported earlier, including the lateral ventricles, CA1 region of the hippocampus, bilaterally (expansion); and the olfactory bulbs, bilateral and central lobules of the cerebellum, including the flocculus, declive VIII and the simple lobule (contraction). The same slices (indicated by solid lines in the sagittal view key, top left) are shown in [Figs 4.13](#), [4.14](#).

Tc1 vs. WT ([Fig 4.12](#)): The morphometric results indicate central and bilateral reductions in the 5/6 and simple lobules of the cerebellum of the Tc1 group, compared with WT. There were also significant bilateral reductions at foci within the thalamus, and the hypothalamus. The cortex appeared relatively preserved. That the regions implicated have fewer significant voxels in this image, compared with [Fig 4.5](#), despite the same degree of smoothing of the  $\log(J_{det})$  fields, may have two causes. First, this study incorporated fewer Tc1 and WT images (a subset of the earlier study), so at individual voxels, the group means may have had poorer separation, reducing significance. Secondly, the GWR average image in this study incorporates Tc1xJ20 and J20 animals; thus the ‘middle space’ defined by the deformation fields will have been slightly different.

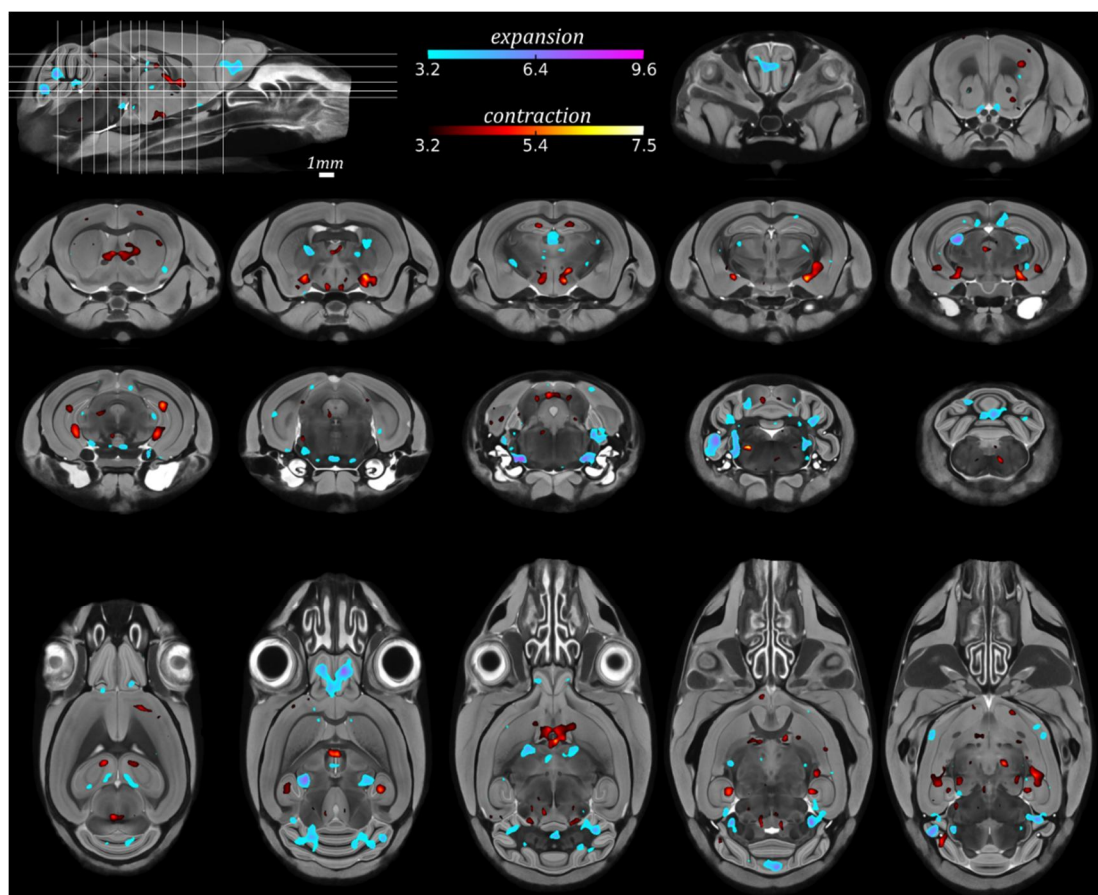


**Figure 4.13** TBM results: J20 vs. WT

Red: J20 animals locally smaller than WT; blue: larger. TBM detected extensive bilateral reductions in the J20 group in the cortex, hippocampus and corpus callosum. There were some scattered regions of reduction in the thalamus. There were small regions of expansion,

bilaterally, in the cortex. The cerebellum was locally larger, around declive VIII, than the group average.

J20 vs. WT ([Fig 4.13](#)): This comparison exhibits the most significant difference between groups: the retrosplenial region of the cortex, centrally. The dorsal apex of the corpus callosum, caudally, was significantly reduced in volume, bilaterally. The most extensive region of reduction was the cortex, bilaterally. This is distinct from the retrosplenial area and appears predominately in the somatosensory cortex (multiple slices of [Fig 4.13](#)), extending rostrally into the primary motor cortex. Some regions of the midbrain and cortex, in the forebrain, appear bilaterally larger in the J20 animal, along with the third ventricle, and some small regions of the amygdala, also bilaterally.



**Figure 4.14** TBM results: J20 vs. Tc1xJ20

Red: J20s locally smaller than Tc1xJ20 group; blue: larger. The regions implicated in this comparison are less extensive. However, there were bilateral regions of expansion, or

preservation in the J20s relative to the Tc1xJ20s, in the cerebellum (including the paraflocculi) and olfactory bulbs. There were small, focussed, but bilateral regions of reduction, controlling for TIV, in the hippocampus (CA1 and CA3) and, in the forebrain, the lateral septal nucleus of the striatum. The ventricles in the J20 group also appear smaller.

J20 vs. Tc1xJ20 ([Fig 4.14](#)): The primary regions of comparative reduction in the J20 animals were the hippocampus (both CA1 and CA3), septal nucleus, amygdala and hypothalamus, all bilaterally. Regions of the cerebellum and paraflocculi appeared larger, reflecting the significant volume loss of the Tc1 group here ([§4.1](#), [Fig 4.12](#)).

In contrast with the J20 group in comparison to WT's ([Fig 4.13](#)), the corpus callosum here was not affected. That the paraflocculi were not affected in the J20s in that comparison suggests relative preservation of the J20 group here, compared with Tc1xJ20 reduction. The olfactory bulbs show more extensive and significant reduction in the double mutant cohort here than in the Tc1 comparison with WT's, and there were larger, bilateral reductions in the geniculate region of the thalamus.

#### 4.2.4 Discussion

An optimised  $\mu$ MRI protocol, previously described ([Cleary et al., 2011b](#)), and the fully automated software pipeline for TBM ([§3](#)) were here employed to investigate the global and local structural differences between Tc1, J20, Tc1xJ20 and WT mouse brains. The Tc1/WT comparison results were expected, as they formed a subset of animals in prior work ([§4.1](#)). This older group, with a lower total number (27), helped to corroborate those results, as I observed similar statistically significant regions of difference, and similar global volume measurements as in the younger group alone (data not shown). I concluded that the age difference between cohorts, controlled for in the GLM of [§4.1](#), did not significantly contribute to the regions of local reduction and expansion shown. The discrepancy in global volumes between Tc1 animals and WT's meant inclusion of a TIV covariate in the GLM was important, to discern local changes in both the Tc1 and

Tc1xJ20 groups. In contrast, the global volumes of the J20 and WT groups were not significantly different, despite significant local atrophy of the cortex.

Volume preservation in AD mouse models is not unprecedented: in APP/PS1 mouse models, *Teipel et al. (2011)* measured no reduction in GM volume, using VBM with SPMMouse. *Lau et al. (2008)* measured reduced volumes compared with WT, but no longitudinal atrophy, which they suggested indicated an earlier, developmental effect. *Delatour et al. (2006)* measured increased brain volumes from 2.5—24 months, as did *Maheswaran et al. (2009a)* from 6—14 months.

Within the J20 animals, I observed local atrophy in regions vulnerable to amyloid pathology in this model. These findings support the hypothesis linking amyloid plaques and morphometric changes in AD: *Mucke et al., 2000* noted that the cortex and hippocampus displayed the highest levels of APP, via immunoreactive histological staining. It should be noted, however, that correlations between atrophy and  $A\beta$  deposition are weaker than those with NFTs (*Giannakopoulos et al., 2003; Delatour et al., 2006; Chételat et al., 2010*). In the late stages of disease, after plaque aggregation, neurodegeneration has already begun and treatment may be less effective (*Hampel et al., 2010*).

In a study of younger J20 mice up to 9 months of age, *Wright et al. (2013)* observed localised, age-dependent neuron loss within the hippocampus (CA1 – but preservation of CA3), prior to the appearance of  $A\beta$  plaques. Here a small bilateral region of CA1 appeared to be locally reduced, while the CA3 region was also preserved compared with the WT group. *Hébert et al. (2013)* showed, using an atlas-based cortical thickness measurement, an increased rate of cortical thinning in this model, from 3.3—15 months, compared with WT. Initial thickness in transgenic mice was, however, greater – except in the retrosplenial, motor and somatosensory cortex regions, all of which were reduced here.

These findings contrast, in part, with a previous whole-brain, atlas-based and morphometric analysis of the J20 model alone (*Badhwar et al., 2013*), which unexpectedly found over 50% of structures were significantly enlarged in J20s compared



with WT, including the hippocampus, striatum and cortex. The corpus callosum and hippocampus exhibited localised regions of reduction. The greatest expansion was observed in the amygdala. I also observed regional expansion in the amygdala, focal volume increases in the cortex (forebrain), and losses bilaterally in the corpus callosum. However, cortical expansion was balanced by significant and larger regions of atrophy, both bilaterally in the somatosensory and primary motor regions, and centrally in the retrosplenial area, which *Badhwar et al. (2013)* found to significantly increase in volume, alongside cortical thickness increases. The authors suggested this expansion could be an inflammatory, or hypertrophic, response to A $\beta$ . The animals of that study were younger (6 months); the discrepancy with the results here could be due to the effects of age: the physical losses of tissue to atrophy over time may have superseded, or be downstream from, the effects of inflammation, which has itself been implicated in AD pathology (*Wright et al., 2013*).

The Tc1xJ20 vs. J20 comparison highlights regions where the presence of Hsa21 interacts with the J20 AD phenotype, resulting in morphological differences from the APP transgenic mutants alone. Here, the J20 group exhibited local hippocampal reductions (red in *Fig 4.14*) in both CA1 and CA3 (adjacent to the thalamus, bilaterally).

The Tc1xJ20 group included regions of local cerebellar reduction (blue in *Fig 4.14*) seen in the Tc1s, and displayed more significant reductions in the olfactory bulbs. Atrophy of the olfactory bulbs is known to be significant in the early stages of AD in humans, and has been detected in MCI patients (*Thomann et al., 2009*). The corpus callosum and cortex were relatively preserved. The geniculate region of the thalamus also appeared affected in the Tc1 vs. WT comparison, but here, compared with J20s, the region of significant difference appeared larger. These regions highlight the influence of APP trisomy on DS-related morphometric brain changes.

My findings highlight the usefulness of structural MRI, combined with TBM, for characterising novel phenotypes, localising morphological changes arising from genetic differences, and informing further studies. Further histological evaluation within these models, which may be targeted to specific regions thanks to TBM, may explore the biochemical mechanisms underpinning these volume changes, including within the

olfactory bulbs, and sub-regions of the thalami of the Tc1xJ20 animals. It would be interesting to compare these results with those at an earlier time-point, to investigate whether these changes are degenerative. Morphological tracking of disease progression over time is made possible *in vivo*. This is the focus of the following two chapters.





# 5 *In vivo* morphometry

This chapter begins the transition from *ex vivo* to *in vivo* MRI data first mentioned in §1.4. Much of the rest of this thesis depends upon one mouse model. That mouse, the rTg4510 model of tauopathy, is first introduced, and a large study, performed at CABI between 2012 and 2015, is described. This is followed by a translation of the morphometric image processing pipeline developed and applied in the previous two chapters to *in vivo* data. Appropriate parameters are investigated. Results from both cross-sectional (§5.5.1) and longitudinal (§5.5.2) analyses of the rTg4510 brains are then shown and discussed.

To complete the transition from *ex vivo* to *in vivo*, this chapter includes an investigation comparing the viability of TBM both *ex vivo* and *in vivo*, in the same dataset (§5.7). This work was conducted with Holly Holmes, who performed the *ex vivo* acquisitions and optimised the *in vivo* scan protocols; and Da Ma, who performed structural parcellation measurements. Once again, for this dataset, the images are first introduced and considerations made for the processing pipeline discussed.

All the rTg4510 work was conducted in collaboration with Eli Lilly. Scanning was performed at CABI by Niall Colgan, Ozama Ismail, Jack Wells, Ian Harrison, James O’Callaghan and Holly Holmes. The purpose was to assess the viability of various MRI measures – including arterial spin labelling (ASL), diffusion tensor imaging (DTI) and T<sub>2</sub>W structural – for the detection and quantification of brain atrophy in the rTg4510 mouse, with the eventual aim of assessing the effectiveness of experimental drug treatments. (ASL and DTI were collected and analysed by CABI collaborators.)

Parts of the single time-point section (§5.4) were published in *Wells et al. (2015)*. The cross-sectional analysis (§5.5) forms part of a publication by *Holmes et al. (2016)*.

## 5.1 The rTg4510 mouse model of tauopathy

Alzheimer's disease (AD) was introduced in §2.1, and I investigated the J20 APP mouse model in §4.2. Several potential AD therapies, targeting  $A\beta$  by reducing deposition or enhancing removal, have failed to elicit disease-slowng effects (§2.3.3). This suggests that hyperphosphorylated tau neurofibrillary tangles (NFTs), a second hallmark of AD and other neurodegenerative diseases, may be a better therapeutic target (*Giacobini & Gold, 2013*). NFT load predicts cognitive status better than amyloid load (*Giannakopoulos et al., 2003*), and better-correlates with neuron loss, which is intimately tied to cognitive decline (*Fox et al., 1999; Jack et al., 2013*). The deleterious effects of NFTs, and possible therapies, are investigated here.

The tetO-MAPT\*P301L (rTg4510)<sup>1</sup> mouse over-expresses a form of mutated human tau related to familial frontotemporal dementia with parkinsonism (FTDP-17) and AD (*Ramsden et al., 2005*). A significant advantage of the model is the rapid progression of gene dose-dependent AD-like features, including neuronal death, cognitive deficits, motor and behavioural problems. Pretangles form in the cortex by 2.5 months of age (*SantaCruz et al., 2005*). By 3–4 months, the cortex and forebrain begin to show signs of neurodegeneration. By 5 months, NFTs have infiltrated the hippocampus (*Ramsden et al., 2005*). The model is particularly suited to FTD, as these pathologies are limited to the frontal brain regions<sup>2</sup>. Their neurodegenerative effects are severe, coinciding with the exhibition of profound structural atrophy of the cortex, entorhinal cortex, hippocampus and forebrain (particularly the striatum), and ventricular enlargement, which progressively worsens throughout the animal's life when left untreated – mirroring AD

---

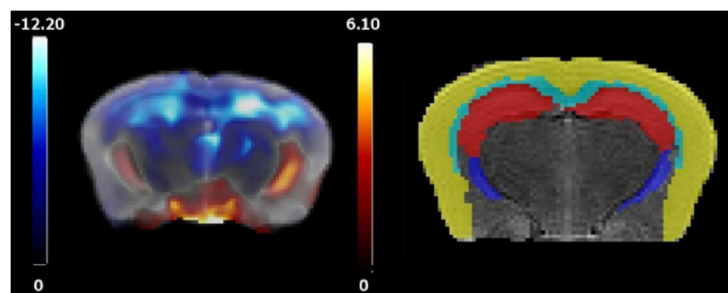
1: The tetracycline-responsive promoter (tetO) controls expression of the microtubule-associated protein tau (MAPT) encoding gene, whose mutation causes FTDP-17. The most common missense mutation is P301L. The model's tau is thus regulatable: (rTg(tau<sub>P301L</sub>)4510).

2: Also, there are no amyloid plaques.

and FTD in humans (*Braak & Braak, 1995; Janus, 2008; Jack et al., 2013*), which progress over many years. Adult rTg4510s are distinguishable from wild-type (WT) brains by eye in MR images (*Fig 5.7*).

To quantify neurodegeneration, *Ramsden et al. (2005)* measured brain weights in rTg4510s and WT littermates. Significant differences from WT emerged after 1 month of age. After 4 months, brains weighed 4-7% less. This difference widened in measurements up to 16 months. The forebrain was visibly atrophied in histological slices and neuron numbers were substantially reduced at 10 months.

*Yang et al. (2011)* used structural MRI with parcellations and, in another report, the same group used morphometry (*Xie et al., 2010*) to investigate the rTg4510 model. The group used a single iteration of GWR and a single atlas-based parcellation method. Their TBM detected enlarged ventricles and hippocampal reductions. Their parcellations, which confirmed these results, detected reduced cortical volume (*Fig 5.1*).



**Figure 5.1** Prior TBM results and parcellation in the rTg4510 mouse brain.

Adapted from *Xie et al. (2010)*, left, and *Yang et al. (2011)*, right, who used TBM and atlas-based parcellation to measure the differences between  $N = 7$  rTg4510 mice and  $N = 7$  WT littermates, aged 5 months. The scale bars show uncorrected two-tailed t-test results; blue here for local reductions (hippocampus and cortex); red for expansion (ventricles).

### ***Doxycycline***

The rTg4510 model's tauopathy is regulatable with doxycycline – an antibiotic, and tetracycline derivative – in the diet. Doxycycline's administration is known to suppress tau, halt atrophy and improve memory function (*SantaCruz et al., 2005*). Because of this

drug sensitivity, the rTg4510 mouse is a propitious test-bed for measuring brain morphometry and the sensitivity of TBM in an animal model *in vivo*.

### ***Epothilone D***

An experimental drug, Epothilone D (EpoD), was also administered, to determine its efficacy. EpoD is a microtubule stabilising agent, previously shown to reduce microtubule loss, axonal dystrophy, and hippocampal neuron loss, and improve cognitive performance in the PS19 mouse model<sup>3</sup> aged 9—12 months, even with established tauopathy ([Zhang et al., 2012](#)). [Barten et al. \(2012\)](#) showed EpoD in rTg4510 mice prevented cognitive deficits, reduced neuron loss and restored normal microtubule dynamics.

## 5.2 Study design and methodology

### 5.2.1 Animals

This study was spread over four years (2012—15), and involved 112 animals, 255 *in vivo* and 25 *ex vivo* structural MR scans. Mice were all female, and either rTg4510 or WT littermate controls. They were licensed from the Mayo Clinic (Jacksonville, FL, USA), bred by Taconic Biosciences (Germantown, NY, USA), and imported prior to imaging at CABI.

26 animals were scanned at a single time-point (single-TP, §5.4), aged 8 months, to ascertain whether various imaging protocols could reliably distinguish the tau pathology in 9 untreated rTg4510s (UT) at a late stage, compared with 17 WTs.

The remaining mice were scanned longitudinally at different ages, under different treatment regimes, to ascertain the effects of healthy ageing, tauopathy, and the ability of two drugs, doxycycline (DOX) and EpoD, to mitigate those detrimental effects. Those

---

3: MAPT\*P301S

images form the content of both this chapter (§5.5, 5.7), and §6. Fig 5.2 illustrates the time course of this study.

For the longitudinal study, 87 animals were scanned in three mixed, age- and litter-matched cohorts {C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>}, across up to three time-points {TP<sub>1</sub>, TP<sub>2</sub>, TP<sub>3</sub>}. Each cohort consisted of three groups. C<sub>1</sub> consisted of WT, UT and rTg4510 mice treated with doxycycline from age 4 months onward, TR<sub>DOX-4M</sub>. C<sub>2</sub> consisted of WT, UT and animals treated earlier, at 3 months of age, TR<sub>DOX-3M</sub>. C<sub>3</sub> consisted of UT and rTg4510 mice treated with EpoD in low and high doses, TR<sub>EpoD-LO</sub>, TR<sub>EpoD-HI</sub>. A few animals did not survive until TP<sub>3</sub>. There were also drop-outs due to poor image quality. For the full numbers, ages at each TP, and inter-TP periods, see Table 5.1 and Table 6.1. C<sub>1</sub> animals which survived until TP<sub>3</sub> were also scanned *ex vivo* (§5.7).

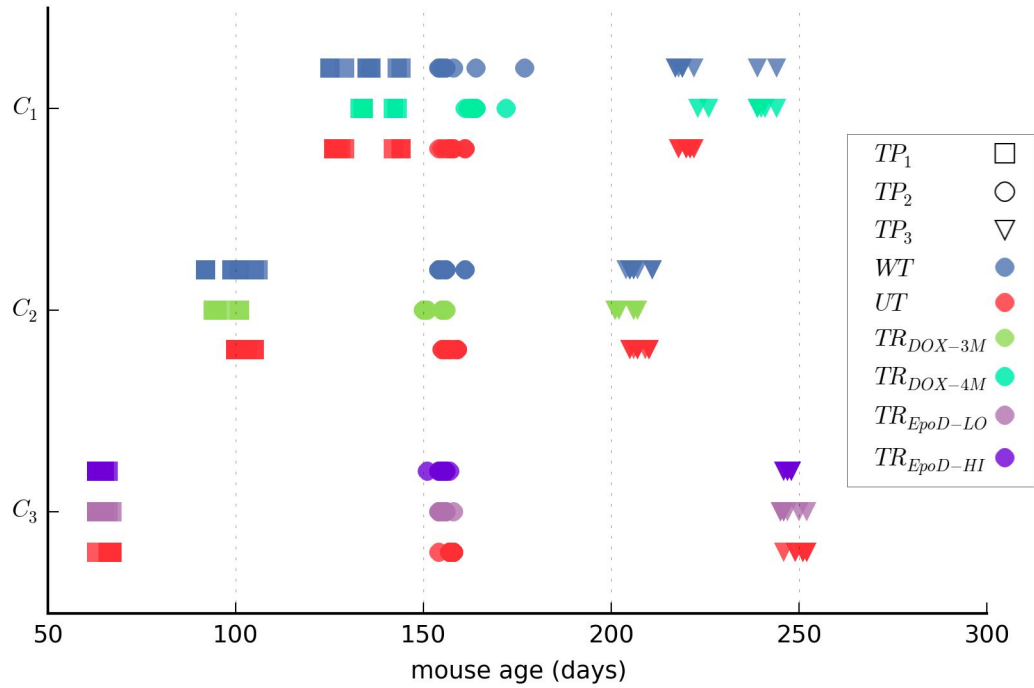
<i>group</i>	<b>N</b>	<b>TP<sub>1</sub></b>		<b>N</b>	<b>TP<sub>2</sub></b>		<b>N</b>	<b>TP<sub>3</sub></b>	
		<b>mean age</b>	<b>SD</b>		<b>mean age</b>	<b>SD</b>		<b>mean age</b>	<b>SD</b>
<b>WT</b>	19	115.6	19.1	19	157.4	5.5	19	214.7	11.3
<b>UT</b>	30	100	27.7	30	157.2	1.7	30	226.1	18.3
<b>TR<sub>DOX-3M</sub></b>	9	96.2	2.8	10	152.9	2.7	10	203.9	2.7
<b>TR<sub>DOX-4M</sub></b>	6	136.7	4.5	7	164.1	3.6	7	236	8.1
<b>TR<sub>EpoD-LO</sub></b>	10	64.5	1.4	10	155.3	1.3	9	246.9	2.5
<b>TR<sub>EpoD-HI</sub></b>	10	64.2	1	10	154.9	1.7	8	247.1	0.8

**Table 5.1** Number of rTg4510 mice, and their mean ages, for the longitudinal study.

For each TP, the N images included in this longitudinal investigation, and the mean and standard deviation age of the mice in days. For the time between TPs, see Table 6.1.

### ***Ethical statement***

This study (§5), including the following BSI investigation (§6) accorded with the United Kingdom Animals (Scientific Procedures) act, 1986, and was approved by UCL's ethical review panel.



**Figure 5.2** Longitudinal rTg4510 study design, showing mouse ages at each TP. Each point represents a 1.5-hour structural MR acquisition (many points overlap).

### 5.2.2 Treatment

TR<sub>DOX-3M</sub> animals began doxycycline treatment at ages 98 or 105 days; TR<sub>DOX-4M</sub> at 148d. Doxycycline treatment consisted of an initial dose of two 10mg/kg boluses via oral gavage, followed by 200mg/kg in mixed feed (by chow weight), accessed *ad libitum*. All TR<sub>EpoD</sub> animals began treatment at age 70d. EpoD, mixed with Captisol<sup>4</sup>, was injected twice-weekly intraperitoneally. TR<sub>EpoD-LO</sub> dosage was 0.3mg/kg. TR<sub>EpoD-HI</sub> dosage was 3mg/kg<sup>5</sup>. The C<sub>3</sub> UT group, as a control, was injected with Captisol vehicle without EpoD.

### 5.2.3 Image acquisition

Owing to its neurotoxicity ([Lerch et al., 2012](#)) and therefore potential to unpredictably interfere with, or corrupt, the effects of tauopathy, a manganese-based contrast agent was

<sup>4</sup>: Captisol, a cyclodextrin, is used as a medium for drug delivery.

<sup>5</sup>: For comparison, [Zhang et al. \(2012\)](#) injected 0.3mg/kg and 1.0mg/kg, once per week, for three months. [Barten et al. \(2012\)](#) administered 0.1mg/kg and 10mg/kg doses, and found the low dose was more beneficial.

not used for these *in vivo* studies<sup>6</sup>. The imaging protocol described below was optimised by my colleague Holly Holmes, who aimed to enhance the contrast between the hippocampus and adjacent structures.

Imaging was performed in the cohorts described above, as follows. Mice were anaesthetised (2% isoflurane) and secured in a custom, 3D-printed head holder with ear bars to ensure head stability, and maintained with 1.5% isoflurane (1L/min in 100% O<sub>2</sub>), as [Wells et al. \(2015\)](#). Structural images were generated using a T<sub>2</sub>W, 3D FSE sequence with parameters: FOV=19.2x16.8x12.0mm; isotropic resolution 150µm<sup>3</sup> (matrix 128x112x80); TR 2500ms, TE<sub>eff</sub> 43ms, echo train length: 4; number of signal averages: 1, imaging time approx. 1h30m per animal.

Scans were all performed with the same Agilent/Varian 9.4T horizontal bore VNMRS, with a 72mm birdcage transmission RF coil and 120mm (inner diameter) gradient set. Signal was received with either a quadrature mouse brain surface receiver coil (single-TP, C<sub>1</sub>{TP<sub>1</sub>, TP<sub>3</sub>} and C<sub>2</sub>), a 2-channel array mouse brain coil (C<sub>1</sub>TP<sub>2</sub>), or a 4-channel array rat head coil (C<sub>3</sub>) (all RAPID Biomedical GmbH, Würzburg, Germany). To test for differences in SNR and CNR between cohorts, I used [E4.1](#) and [E4.2](#). Data was scaled prior to analysis to account for gradient calibration, as measured by a custom MR phantom ([O'Callaghan et al., 2014](#)), to exclude the possibility of gradient scaling causing measured volume changes. Following imaging, histology of selected brains was performed by collaborators at Eli Lilly, with antibody staining for NFTs<sup>7</sup>.

Other MR measurements (not all reported here) shared the imaging session, including ASL (to measure CBF) and DTI (for probing structure and connectivity).

---

6: (Gadolinium, also toxic, was used in *ex vivo* scans.)

7: The PG-5 antibody was used to stain for hyperphosphorylated tau in the cortex, hippocampus and thalamus, and the NFT density reported.

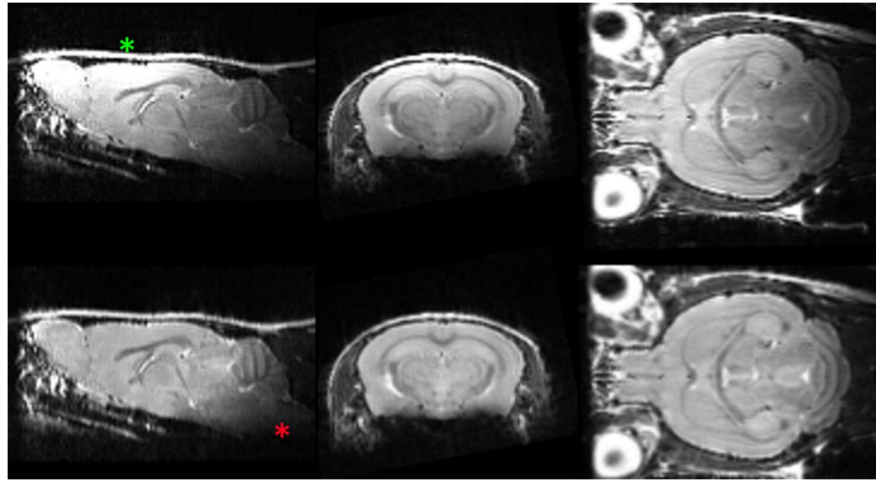
## 5.3 Adjustments to the pipeline for *in vivo* data

[Bock et al. \(2005\)](#) described the acquisition of up to 16 mouse brains *in vivo*, simultaneously, in a “mouse hive”. In addition to the difficulties already discussed ([§3.2](#)), it is practically difficult to simultaneously, stably maintain 16 mice *in vivo* for the time required for an MRI scan. Instead, for this study, the narrow bore of the Agilent 9.4T scanner only permitted scanning one mouse at a time. Multiple subject extraction was thus unnecessary. Orientation correction was able to proceed much faster, as (just as in the human *in vivo* paradigm) all images were in the same initial orientation, so after one image had been oriented ([§3.3](#)), the remainder could use the same initial affine matrix. Although a head-holder was used during structural scans, sometimes the head would slip out of place, resulting either in motion artefact or FOV misplacement (some images were discarded), or a brain ‘rotated’ from the usual frame of reference. Fine adjustment, using reflected images ([§3.3.2](#)), was required.

### 5.3.1 Non-uniformity correction

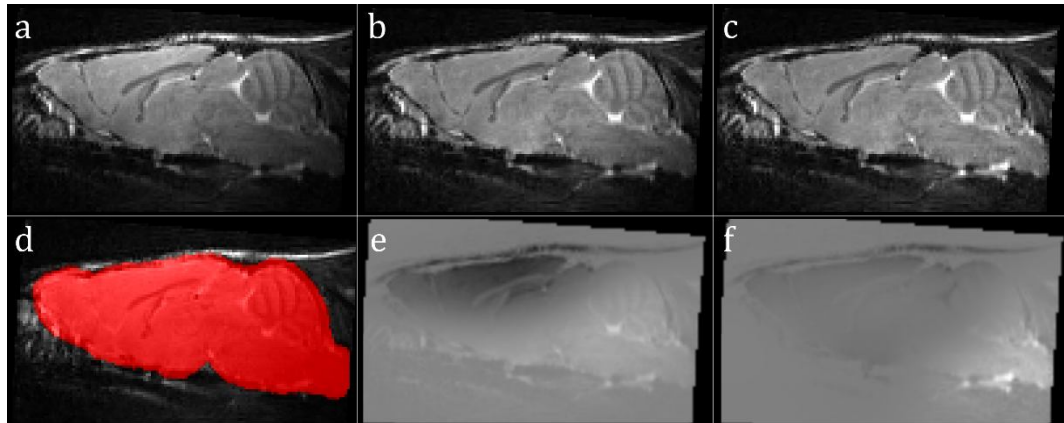
NUC was necessary at several stages, and is more important *in vivo* than *ex vivo*. Because, to better detect tauopathy in the forebrain, the surface receiver coils were placed above the rostral cortex and forebrain, signal drop-off usually manifested in the brainstem and posterior surface of the cerebellum, [Fig 5.3](#). In the four-coil setup, two coils were placed below the mouse head; however, their contributing signal was limited, owing to the distance between the inferior surface of the jaw and the brain. If the intensity inhomogeneity was severe (assessed visually) or non-symmetric, the intensity-weighted principal axis calculation for automated orientation could fail, and required prior NUC.





**Figure 5.3** Single-TP WT brain, *in vivo*, before and after initial NUC.

First row: typical uncorrected sagittal, coronal and transverse slices. Note the elevated GM intensities at the cortical surface, beneath the receiver coils (placed at the green asterisk). Second row: the same slices after N4, without a mask. The cerebellum and brainstem still exhibit signal drop-off and inhomogeneity (red asterisk), but contrast in the cerebellum is improved and its posterior surface is more easily distinguished.



**Figure 5.4** Intensity non-uniformity in an rTg4510 *in vivo* mouse brain image. Multiple iterations of NUC, with masking, confer additional benefit. The same slice of a WT brain at TP<sub>1</sub>, first row: (a) prior to NUC; (b) after initial application of N4ITK; (c) after additional NUC using a dilated brain mask. Second row: (d) semi-transparent dilated binary mask overlaid; subtraction images show intensity difference and hence improvement in brain homogeneity (e): b-a; (f): c-b. Compare with human data: [Prados et al. \(2014\)](#), figure 2.

NUC was usually appropriate prior to masking, in order that the low signal, poor-contrast caudal aspect of the cerebellum ([Fig 5.3](#)) properly registered with atlas data. This surface would often otherwise be underestimated in the rTg4510 brains, and required some manual intervention (see below). An additional application of NUC was beneficial after masking ([Fig 5.4](#)).

### 5.3.2 Brain masking with manual intervention and bootstrapping

I used the *in vivo* UFL atlas ([Ma et al., 2008](#)) to perform initial multi-atlas brain masking of the *in vivo* rTg4510 data. As the atlas' constituent images are in-skull, the region surrounding the brains matched the data better than the equivalent, skull-stripped *ex vivo* UFL atlas. Also, the atlas images' ventricles registered to most rTg4510 brains reasonably well (assessed via visual inspection). The NUS atlas, while based upon *in vivo* images and also exhibiting large ventricles, was not chosen, as it has fewer member images (5), lower contrast, and the brains are skull-stripped, so their surfaces are not surrounded by meninges, skull, or other flesh in the images ([Fig 2.4](#)) which might match the rTg4510 data and benefit initial registration.

The UFL atlas has 10 images and 20 parcellations each. I doubled the database size by reflecting images along the right/left axis (x, in RAS orientation – i.e., in the YZ mid-sagittal plane) – exploiting mouse brains' inherent approximate symmetry ([§3.3.2](#)). The multi-atlas label fusion algorithm STEPS was used (as per [§3.5.1](#)), with parameters  $X = 9$ ;  $\sigma = 3$  voxels (0.45mm). The output included binary and probabilistic brain masks.

One problem encountered with mask accuracy arose due to the significant volume of eCSF present in the rTg4510 images. Atlas brain surfaces were liable to erroneously register to the external surface of the eCSF, as the intensity gradient (in the atlas images) from GM to BG matched the gradient in the rTg4510 data from eCSF to BG better than that from GM to eCSF. This was a problem for *ex vivo* rTg4510 images (see [§5.7.4](#)), but *in vivo*, lower resolution and contrast compounded these difficulties. Additionally, the skull – separating the brain from external flesh above the cortex – often occupied less than a voxel's thickness (150 $\mu$ m), leading to PV effects in this area. Extra tissue above the cortex was also often included in the mask.

Consequently, my colleague Da Ma manually corrected a subset of the *in vivo* rTg4510 masks, removing the eCSF regions from the mask using ITK-Snap ([Yushkevich et al., 2006](#)). These images and manually-corrected masks formed a second, rTg4510-specific atlas database. I replaced the UFL atlas with this subset of the rTg4510 data, and employed the multi-atlas label fusion procedure to automatically, and accurately, mask the remainder of the dataset in a ‘bootstrapping’ procedure. Parameters with this larger database of  $N=25$  images were  $X = 12$ ;  $\sigma = 3$  voxels. Improved accuracy was assessed by visually inspecting all the results.

### 5.3.3 Group-wise registration

I investigated different combinations of GWR parameters to accommodate the lower resolution of the *in vivo* data. I used the C<sub>1</sub>TP<sub>3</sub> dataset, including WT and UT brains, to incorporate as much natural physical variability between brains as possible – thus testing registration’s performance when the images were dissimilar. For each test, I began with 10 iterations of globally-optimised registration using the same default parameters (1 iteration 6-DOF rigid; 9 iterations 12-DOF affine). For 15 subsequent NRR iterations, I varied two *NiftyReg* F<sub>3</sub>D parameters about their defaults,  $(sx, sy, sz) = 5$  voxels (the final grid spacing of control-point positions, all kept equal, to be isotropic) and  $be = 0.005$ , the bending energy penalty term:

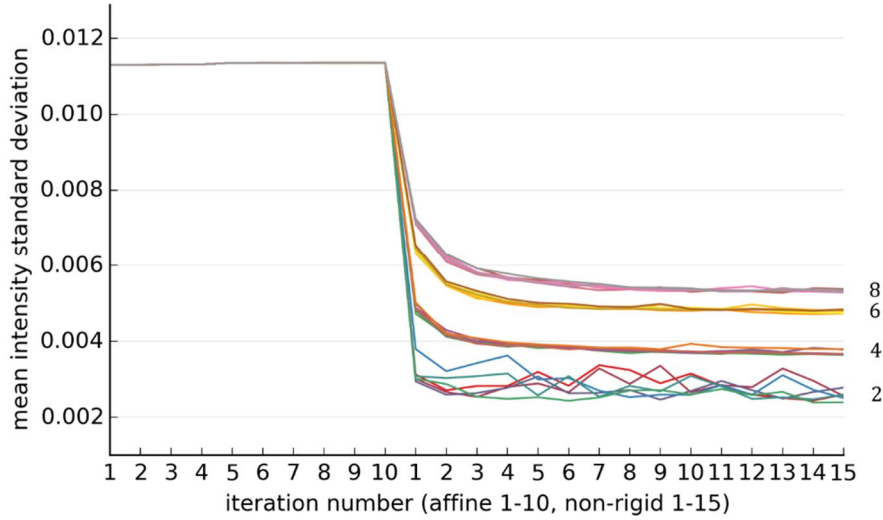
$$(sx, sy, sz) = \{2, 4, 6, 8\} \text{ voxels}$$

$$be = \{0.0005, 0.0025, 0.005, 0.0075, 0.01, 0.05\}$$

For *in vivo* data, initial CP spacing was doubled 4 times (4 levels). At each GWR iteration of the resulting 24 combinations, within a dilated brain mask, I calculated the mean of the voxel-wise standard deviation between registered, resampled images.

An example of the results is shown in [Fig 5.5](#). I chose  $(sx, sy, sz) = 4$  voxels;  $be = 0.005$  for *in vivo* brains. The lowest SD values in [Fig 5.5](#) are from  $(sx, sy, sz) = 2$ . However, upon visual inspection of the resampled images, there appeared to be some over-fitting: local structures were severely and uncharacteristically warped and distorted away from their natural physical appearance. The variability between iterations also

indicates some unreliability: unlike the other parameters, the SD results do not settle on a stable “inverted plateau” (or “valley floor”), even after 15 iterations.



**Figure 5.5** Sample group-wise registration parameter testing results.

The mean intensity standard deviation over all voxels within a dilated brain mask, across all resampled images, plotted at each GWR iteration, using different  $(sx, sy, sz)$  and  $be$  values (after the same affine parameters). The former are noted adjacent to clusters of results on the right (in voxels). The SD images of [Fig 3.20](#) were created in a similar manner.

In each *in vivo* investigation below, I performed statistical tests on the  $\log_e$ -transformed  $J_{det}$  values arising from the deformation fields generated by the final iteration of this GWR. Unless otherwise noted, I smoothed the resulting maps with a 3D Gaussian kernel (FWHM 0.16mm, 1.07 voxels), as per the *ex vivo* analysis ([§3.10](#)).

## 5.4 Single time-point morphometric investigation of the rTg4510 mouse brain

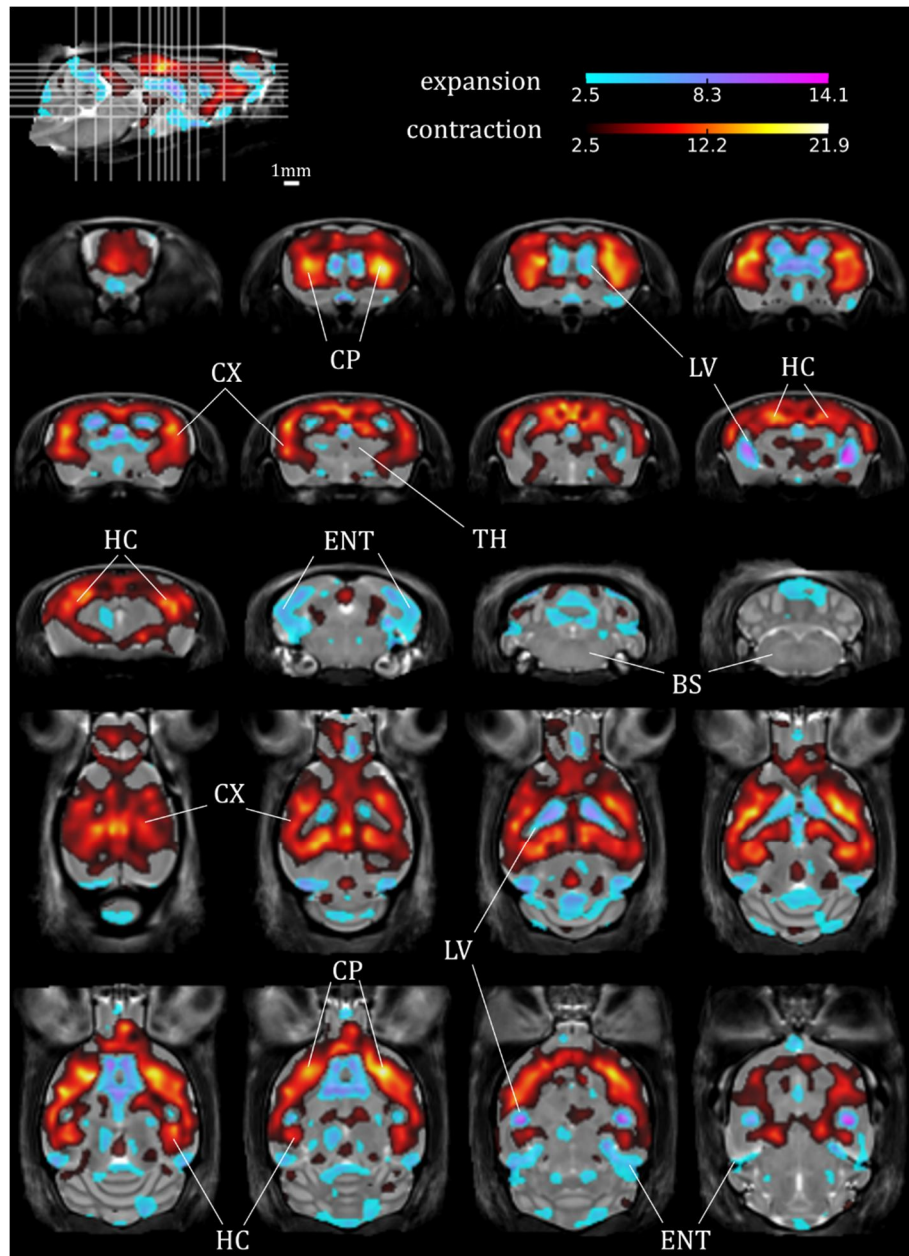
The differential structure of the 5-month-old rTg4510 mouse brain, compared with WT controls, was investigated previously by *Xie et al. (2010)* and *Yang et al. (2011)*. To ensure that my image analysis and processing techniques could sensitively detect these gross differences, prior to the larger, longitudinal study (§5.5), I performed a single, late TP investigation. This analysis, and the TBM results, were published in *Wells et al. (2015)*. Animal numbers are included above (§5.2.1), as are the processing pipeline steps. One brain, a WT, was excluded due to severe signal loss in the brainstem, which would lead to poor atlas and group-wise registration.

I used binary brain masks to measure TIV. The GLM covariates were for genotype (WT, UT) and TIV (E5.1). The resulting SPM (Fig 5.6) was FDR-corrected ( $q = 0.05$ ).

E5.1

$$Y = b_1P(WT) + b_2P(UT) + b_3TIV + \epsilon_i$$

TBM appears to readily detect extensive reductions in volume in the UT group compared with WTs at this late time-point in the animals' lives. As expected for this FTD-related tauopathy model, the forebrain, particularly the caudate putamen (striatum) appears most affected, bilaterally. Enlargement of the lateral, third and fourth ventricles is also clearly detected. TBM also detected enlargement of the eCSF spaces in the entorhinal cortex region. The cortex, hippocampus and olfactory bulbs are severely affected, beyond any group differences explicable by TIV alone. The results, particularly the coronal hippocampal slices, are in clear agreement with the single slice shown by *Xie et al. (2010)* (Fig 5.1), although it is not clear whether that group covaried with TIV; they also did not perform multiple testing correction. Histology of these brains revealed a clear negative correlation ( $r = -0.80$ ,  $p = 0.01$ ) between NFT density (assessed via cell staining) and rTg4510 hippocampal volume (assessed via atlas-based parcellation).



**Figure 5.6** Single time-point TBM results from 8-month-old rTg4510 mice. TBM revealed (red): local volume reductions in the UT group (N=9), compared with WT controls (N=16); (blue): local expansion. Results (FDR-corrected t-statistics) overlaid upon coronal and transverse slices of the structural average image after 15 iterations NRR. Solid lines in the key indicate slice locations. BS: brainstem; ENT: entorhinal cortex; CP: caudate putamen; CX: cortex; HC: hippocampus; LV: lateral ventricles; TH: thalamus.

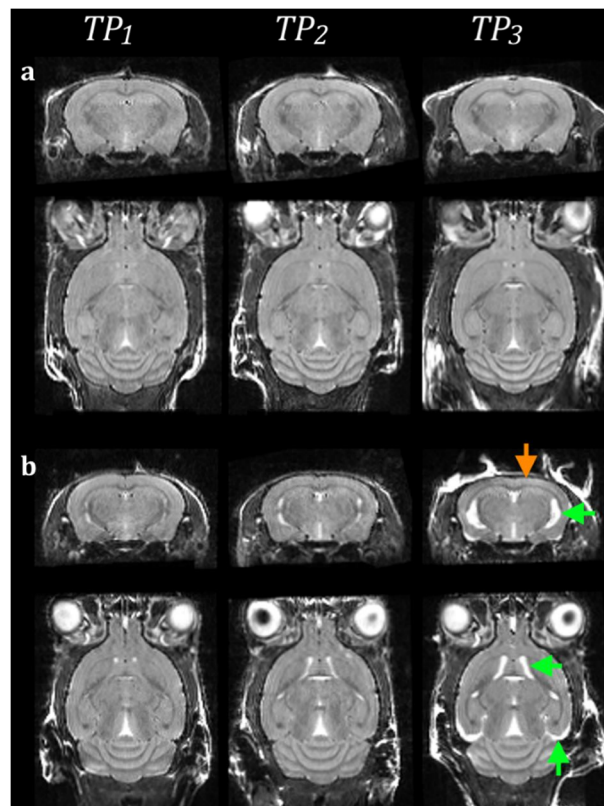
The centre of the thalamus, a low tau-burden region, appears relatively unaffected, along with the midbrain, brainstem and cerebellum, which exhibits some regional expansion

relative to the controls. *Yang et al. (2011)* recorded, via segmentation and manual landmark-based measurements, hippocampal and cortical volume reductions of 26.7% and 20.6%, respectively, in rTg4510s, and thinning of the entorhinal region (intra-rater measurement variability was up to 8.2%). They did not control for TIV, but found that the cerebellum exhibited no global difference in volume, which may explain the apparent local increases, proportional to TIV, detected here by TBM.

This first *in vivo* analysis of the rTg4510 brains indicated the viability of TBM for the detection of gross atrophy-related pathology at late time-points. The usefulness of the technique, however, is largely in sensitively detecting unforeseen localised changes which are not readily visible by eye alone; this is the purpose of including a large number of animals in each group. The subsequent multi-time-point study, detailed below, should therefore be of particular interest, as we should be able to find the developmental stage of pathology at which TBM becomes viable for detecting differences and separating groups.



## 5.5 Multiple time-point rTg4510 study



**Figure 5.7** Visible atrophy in the rTg4510 brain, compared with relative preservation of WTs.

Whereas the WT mouse brain is preserved over three imaging time-points (a), the UT rTg4510 brain (b) exhibits severe neuronal cell death, manifesting as cortical thinning (orange arrow), ventricular expansion, and external CSF (eCSF) accumulation at the entorhinal cortex (green arrows).

Progressive atrophy was clearly visible upon inspection of the scanned, reconstructed brains ([Fig 5.7](#)). The sequence was optimised for GM/WM contrast without a contrast agent, and this appeared successful: the corpus callosum and deep cerebellar WM were clearly visually distinguishable, as was the GM of the hippocampus. The cortex was visibly diminished in the later TP UT brains, alongside gross ventricular expansion.

I found comparable SNR and CNR across all images ([Table 5.2](#)). I performed both cross-sectional (within-TP) and longitudinal (between-TP) TBM analysis of each of these



cohorts, using the pipeline detailed above. These SNR measurements also compare favourably with those obtained from multiple-subject scans *ex vivo* (§4.1.5).

<i>cohort</i>	<i>SNR</i>	<i>SD</i>	<i>CNR</i>	<i>SD</i>
<b>C<sub>1</sub></b>	15.13	2.62	4.43	0.71
<b>C<sub>2</sub></b>	15.7	2.69	4.64	0.73
<b>C<sub>3</sub></b>	15.83	2.63	4.55	0.82

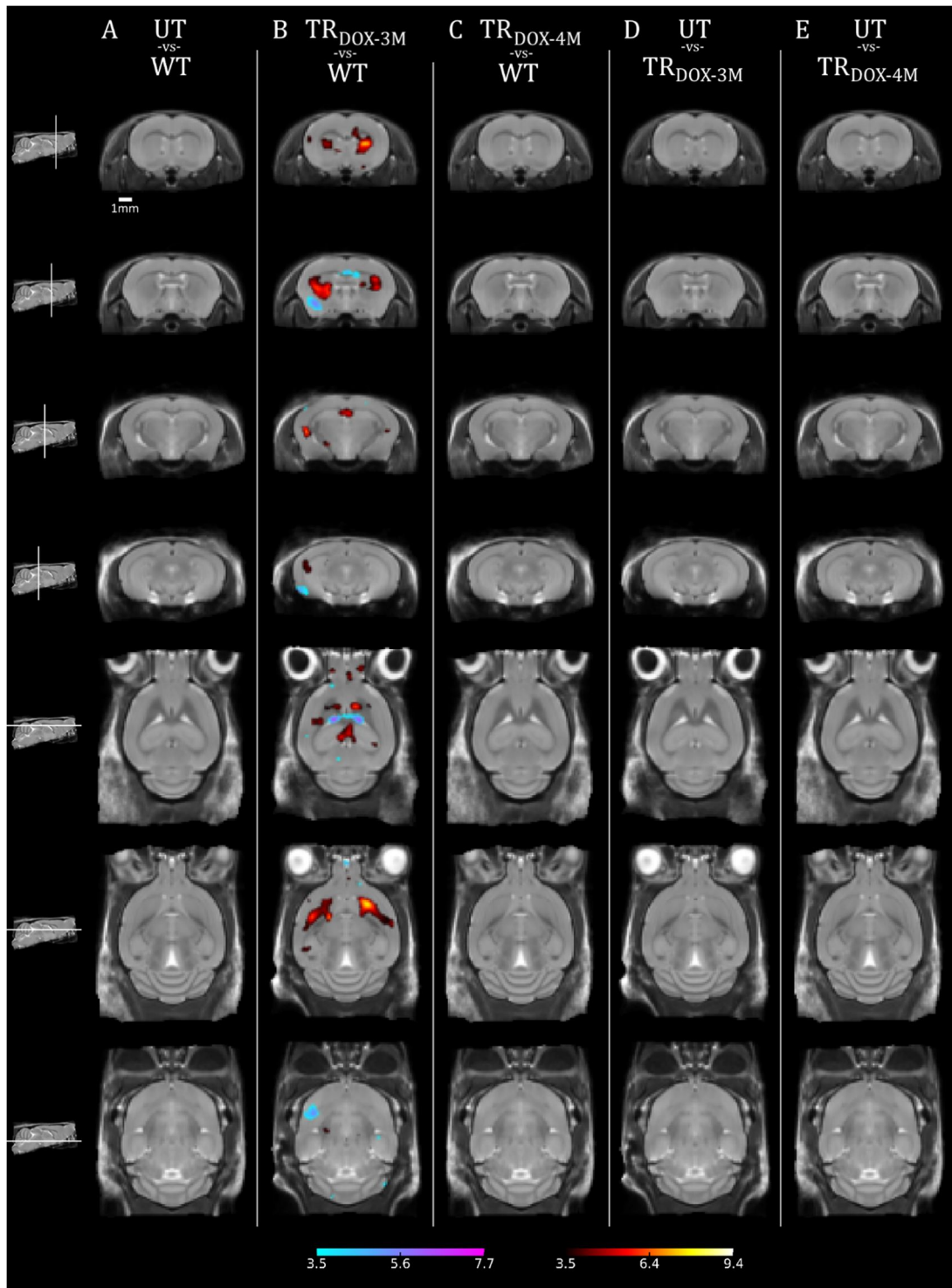
**Table 5.2** Mean SNR and CNR for all rTg4510 scans in the longitudinal study.

Means and standard deviations for each cohort. Although three different receiver coils were used, SNR and CNR remained stable throughout the study. For formulae, see E4.1, E4.2.

### 5.5.1 Cross-sectional analysis

The *in vivo* pipeline was used to perform TBM analyses to compare groups at each of the three TPs, within each cohort. Figs 5.8, 5.9, 5.10 show the GWR average images from C<sub>1</sub> and C<sub>2</sub>, overlaid with statistical results (C<sub>3</sub> omitted from figures owing to lack of significant voxels). Comparisons were within-cohort, and therefore between approximately age-matched mice. This analysis was published in Holmes *et al.* (2016).

For these results, I initially tested the GLM (E5.1, including a third  $P(\textit{genotype})$  term for the treated group) with a TIV covariate (as per the single-TP study, §5.4). However, in most cases, this removed the majority of significant voxels from the parametric maps, after FDR. This indicates a strong linear relationship between TIV and the local, voxel-wise volumetric differences between groups illustrated in the figures below. It also suggests that, at these earlier TPs, the reduced animal numbers (compared with the single-TP at 8 months of age) provided insufficient power to achieve statistical significance and detect the earlier, subtler effects of neurodegeneration. I investigated the effects of TIV in a group comparison separately (§5.6), and show effect size calculations at TP<sub>3</sub> in §5.7.3.



**Figure 5.8** TBM results at TP<sub>1</sub> of the cross-sectional multi-time-point study.

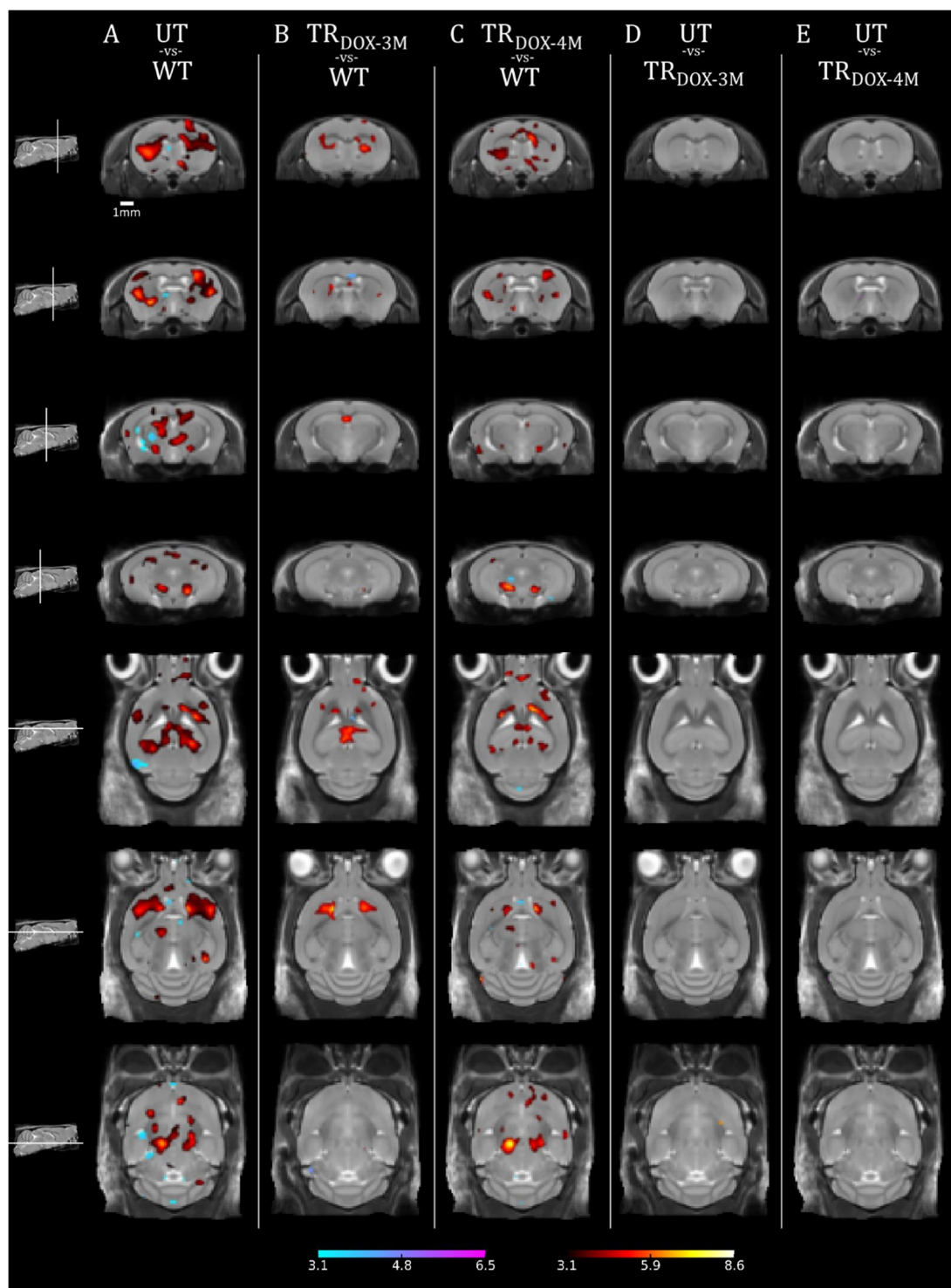
Each column in this figure, 5.9 and 5.10 represent a different comparison. FDR-corrected t-statistics are overlaid upon equivalent representative coronal and transverse slices of the final GWR average image, in each figure. Column A: cohort 1 (results for this comparison from cohort 2 not shown, but in all cases similar to those from cohort 1); cols B, D: cohort

2; cols C, E: cohort 1. In all three figures, (A,B,C), red: UT and TR groups locally smaller than WTs; blue: larger. (D,E), red: UT groups locally smaller than TR; blue: larger.

At TP<sub>1</sub> (*Fig 5.8*), there was a global difference in BVs between UT and WT groups. Although TBM could not reliably separate these groups, there were significant voxel-wise volume differences between the early treated group and WTs. This may be attributed to the greater animal number in this cohort (N=9 TR<sub>DOX-3MO</sub>; N=6 TR<sub>DOX-4M</sub>), improving the available statistical power: one would expect the latter group to show greater separation from WTs. Animals from the late treatment group were approximately one month older at this TP (*Table 5.1*). These early differences appear predominately bilaterally in the striatum. That no voxels survived FDR correction in the UT/WT group comparison may be because the global volume difference was evenly spread throughout the brain. No comparison of TR groups with UT animals showed significance at this TP. This is as expected, as the treatment start coincided with the TP<sub>1</sub> scan.

At TP<sub>2</sub> (*Fig 5.9*), the comparison between WT and TR<sub>DOX-3M</sub> brains revealed a very similar pattern of atrophy: the striatum appears bilaterally affected, and there is a central reduction in the treated group's hippocampus. Despite a similar difference in numbers as at TP<sub>1</sub>, the late-treated group exhibits a more extensive bilateral pattern of reduction compared with the control WTs, including apparent bilateral reductions in the thalamus. At this stage, the effects of doxycycline-linked tissue preservation are already apparent: the same regions are affected, more extensively, in UTs.

The GWR average images have larger ventricles than at TP<sub>1</sub>. Morphometry detects a difference between WT and UT ventricles, as well as subtle regions of expansion around the lateral ventricles in the TR groups, compared with WTs. No voxels survived FDR in the comparison between UT and TR groups, except a very small unilateral region of volume reduction in the forebrain of the UT group compared with TR<sub>DOX-3M</sub> (final row), indicating the earliest detection of structural preservation.



**Figure 5.9** TBM results for WT, UT and doxycycline-treated mice at TP<sub>2</sub>.  
 At this TP, reductions in the UT and TR groups, compared with WTs, are clearly visible.  
 The late-treated group appears preferentially reduced, over the early-treated group.

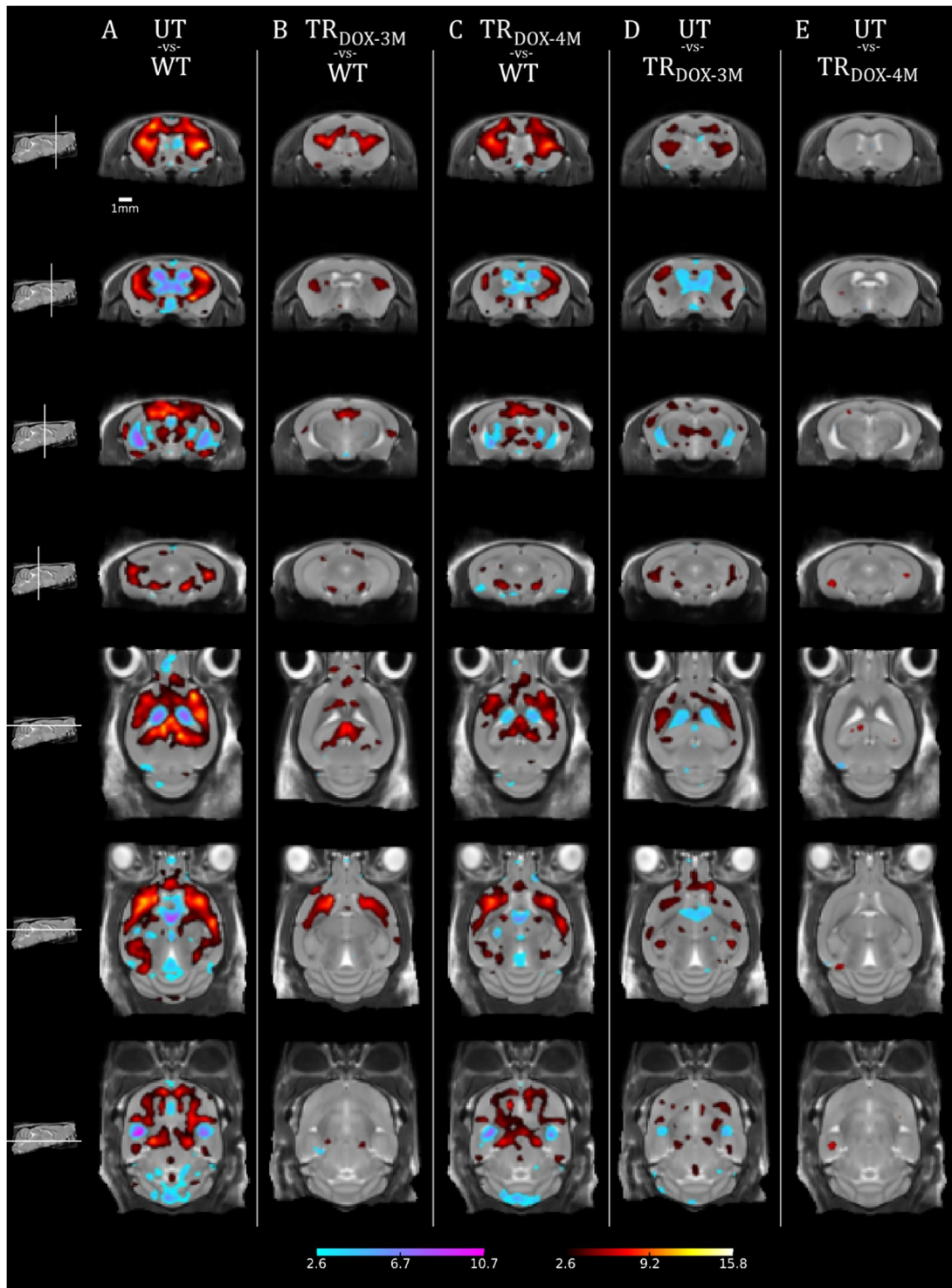
At TP<sub>2</sub> and TP<sub>3</sub>, the thalamus also appears affected in the UT and late-treatment groups, despite exhibiting low NFT counts in the post-imaging histology (data not shown). This may be an artefact of the global volume difference between groups.

At TP<sub>3</sub>, approximately one month prior to the single-TP investigation, very similar TBM results were observable in the comparison between WT and UT animals (*Fig 5.6, 5.10A*), although here without the TIV covariate. These consist of extensive bilateral reductions in the UT group c.f. WTs: the forebrain, striatum, hippocampus and cortex are all bilaterally reduced in volume. The UT cerebellum shows some increase in volume relative to WTs. This region was not assessed histologically for NFTs, but is not believed to be affected. Third and fourth ventricle expansion may have physically displaced the cerebellum, causing an apparent expansion (see §6 for further discussion). The lateral ventricles are also greatly enlarged, and TBM detects some expansion of CSF spaces in the entorhinal cortex region.

Early intervention appeared more successful at limiting atrophy and ventricular expansion. The early-treatment group is clearly differentiated from the UTs at TP<sub>3</sub>. Parts of the hippocampus, bilaterally; the thalamus, the striatum and the cortex all appear relatively preserved. The late-treatment group is just distinguishable from UTs, here, at bilateral foci in the hippocampus (coinciding with the early-treatment group), but TR<sub>DOX-4M</sub> mice exhibit a similar pattern of atrophy as the UTs, when compared with WTs. In both the early and late-treatment groups, the atrophy appears more severe than at TP<sub>2</sub>. As the WT brains do not increase in volume over time, this suggests the dosage of doxycycline did not completely halt progression of neurodegenerative tauopathy.

As well as the spread of statistically significant voxels increasing compared with earlier TPs, the t-statistics themselves were greater, indicating broader separation of group means. The differences between columns (*B,C*) and between (*D,E*) in *Figs 5.8-10* may be explained by the tau transgene suppression, and hence preservative effects of doxycycline intervention. However, at no time-point did TBM, even without a TIV covariate, distinguish UT from TR<sub>EpoD-LO</sub> or TR<sub>EpoD-HI</sub>, nor did it distinguish between those groups, after multiple-testing correction with FDR (q=0.05). Thus, slices from those GWR averages images are omitted.





**Figure 5.10** Cross-sectional TBM results at TP<sub>3</sub>.

The t-statistics are generally much greater than in [Fig 5.8](#) or [Fig 5.9](#), indicating better group separation. In both comparisons, the early-treated group exhibits better structural preservation than the late-treated group.

## 5.5.2 Longitudinal analysis

The previous, cross-sectional analysis sought to compare groups *at* each TP of the study, to determine the sensitivity of TBM to group differences in local brain structure, at different ages, stages of disease progression, and after different periods from the start of treatment. NRR may also be leveraged to compare the changes *between* TPs. In this section, registration and TBM are used to measure how each group deforms between TPs – the degree, extent, and location of atrophy may thus be compared.

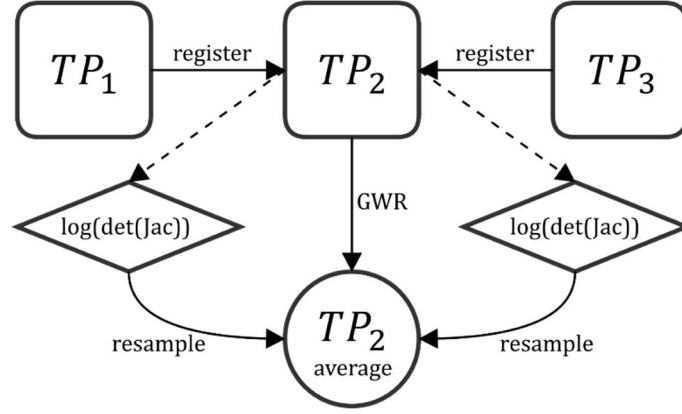
In a longitudinal investigation of an APP transgenic AD mouse model, [Lau et al. \(2008\)](#) registered all brains from all TPs into a single group average space, and controlled for age in a linear, mixed-effects model. The pair-wise registration technique I adopted was similar to that employed by [Maheswaran et al. \(2009a\)](#). After intensity standardisation from the *in vivo* pipeline described above, I individually registered each mouse brain from TP<sub>1</sub> and TP<sub>3</sub> to its corresponding image from TP<sub>2</sub> (from the same mouse – thus, mice with either member of the image pair missing were excluded from the group comparison). For each mouse brain image  $i$ :

$$iTP_1 \rightarrow iTP_2 \leftarrow iTP_3$$

Where the arrows represent a pair-wise registration (affine followed by non-rigid), with parameters as per the *in vivo* GWR (§5.3.3). Dilated masks from §5.3.2 were recycled to constrain the ROI for these registrations. The resulting pair-wise deformation field  $iDF_{pair}$  was in the space of  $iTP_2$ . One advantage of this technique is that, nominally,  $iTP_2$  will already approximately be in the mid-way average space from each set of  $\{iTP_1, iTP_2, iTP_3\}$ , thus requiring the least physical deformation from other TPs. I calculated the voxel-wise  $\log(J_{det})$  from each  $iDF_{pair}$  (again, in the space of  $iTP_2$ ).

I next performed a GWR of all TP<sub>2</sub> images, from all groups (re-using dilated masks and parameters from §5.3.3). I used the resulting deformations to resample each  $\log(J_{det})$  image into the final TP<sub>2</sub> average space (using trilinear interpolation). This interpolation doubled as a “smoothing” step, which in the *ex vivo* pipeline was additional. Here, the comparatively large voxel sizes (and hence greater PV) renders smoothing less important.

These steps are illustrated in [Fig 5.11](#). The technique was thus similar to that of [Kipps et al. \(2005\)](#).



**Figure 5.11** Illustration of the registration steps for longitudinal TBM.

Each image from  $TP_1$  and  $TP_3$  was registered to its respective pair from  $TP_2$ , and  $\log(J_{det})$  values calculated from the deformations at each voxel.  $TP_2$  images were group-wise registered and the  $\log(J_{det})$  values resampled into the final average space, for voxel-wise statistics.

I performed voxel-wise statistics on the resampled  $\log(J_{det})$  images, which represent deformations from  $TP_{1-2}$  or from  $TP_{3-2}$ , using the GLM with the following model:

$$Y = b_1 P(TP_{X-2}(WT)) + b_2 P(TP_{X-2}(UT)) + b_3 P(TP_{X-2}(TR)) + b_4 Days + \epsilon_i$$

Where  $P(TP_{X-2}(genotype))$  is a dummy variable representing the binary group-and-time-point classifications.  $TP_{X-2}$  represents deformations *from*  $TP_1$  or  $TP_3$  to the same animal's  $TP_2$  image. The covariate *Days* represents, for each image, the number of days between the  $TP_1$  or  $TP_3$  scan and the corresponding  $TP_2$  scan. Thus I controlled for the variable time interval between scans ([Table 6.1](#) shows the mean interval per group). For these tests, I did not include a TIV covariate, as the purpose was not to compare volume across groups (the inclusion of TIV would mitigate any initial global scaling factor). Rather, all the volume changes statistically compared here were within-group (actually within-animal), between TPs, thus any global scaling is likely to be due to atrophy or natural volume change, and is of interest. [Whitwell et al. \(2001\)](#) used TIV to scale serial



MR measurements, but this was to account for possible variations in voxel size due to scanner calibration, which I corrected separately (§5.2.3).

There were no significant voxels after correction for FDR ( $q = 0.05$ ) when comparing:

$$TP_{1 \rightarrow 2}(UT_{C_1}) \text{ vs. } TP_{1 \rightarrow 2}(WT_{C_1})$$

This might be expected, as earlier in the animals' life and at earlier stages of the disease progression, changes are subtler, as indicated by the cross-sectional results above. Indeed, testing cohorts separately, the comparison of the later periods:

$$TP_{3 \rightarrow 2}(UT_{C_1}) \text{ vs. } TP_{3 \rightarrow 2}(WT_{C_1})$$

$$TP_{3 \rightarrow 2}(UT_{C_2}) \text{ vs. } TP_{3 \rightarrow 2}(WT_{C_2})$$

did show some significant voxels, at the same threshold, with similar N per group. When data from all cohorts was combined to compare the  $TP_{1 \rightarrow 2}$  deformations:

$$TP_{1 \rightarrow 2}(UT_{C_1}, UT_{C_2}, UT_{C_3}) \text{ vs. } TP_{1 \rightarrow 2}(WT_{C_1}, WT_{C_2})$$

there were still no significant results. This may be because there was significant variability in inter-time-point periods, because the  $C_1$   $TP_{1 \rightarrow 2}$  interval was very low ([Fig 5.2](#)), so any changes would be extremely subtle, or because the test was under-powered for such subtle changes, with this number of animals. I combined  $C_1$  and  $C_2$  to increase numbers and hence statistical power. [Fig 5.12](#) shows the results of comparing:

$$\text{Column 1: } TP_{1 \rightarrow 2}(UT_{C_1}, UT_{C_2}) \text{ vs. } TP_{1 \rightarrow 2}(WT_{C_1}, WT_{C_2})$$

$$\text{Column 2: } TP_{3 \rightarrow 2}(UT_{C_1}, UT_{C_2}) \text{ vs. } TP_{3 \rightarrow 2}(WT_{C_1}, WT_{C_2})$$

$$\text{Column 3: } TP_{1 \rightarrow 2}(UT_{C_1}, UT_{C_2}) \text{ vs. } TP_{1 \rightarrow 2}(TR_{C_1}, TR_{C_2})$$

$$\text{Column 4: } TP_{3 \rightarrow 2}(UT_{C_1}, UT_{C_2}) \text{ vs. } TP_{3 \rightarrow 2}(TR_{C_1}, TR_{C_2})$$

As doxycycline treatment began concurrently with the first scan, controlling for the inter-TP period in the GLM also controlled for the time from initial intervention (although not mouse age).

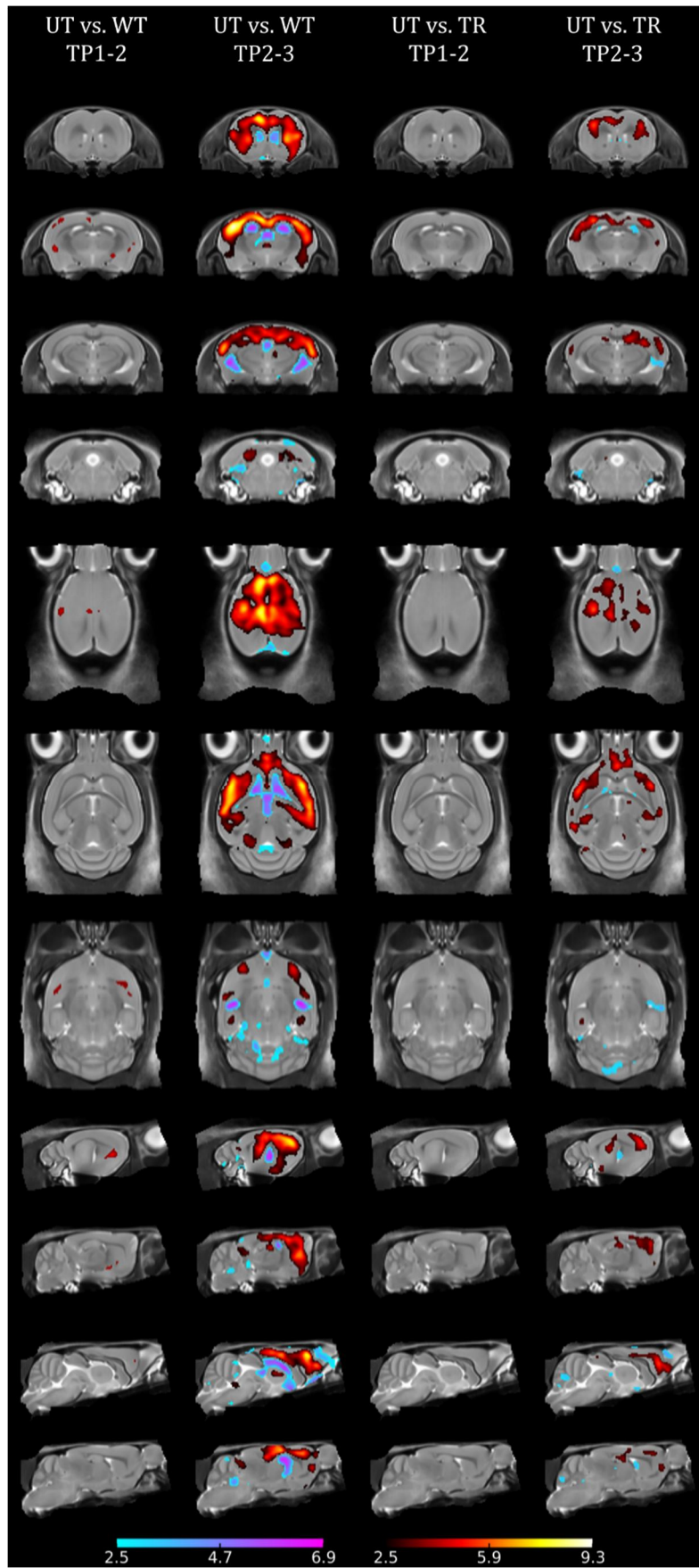


Figure 5.12 Longitudinal TBM results for UT vs. WT and UT vs. TR comparisons.

The same representative coronal, transverse and sagittal slices of the final TP<sub>2</sub> GWR structural average, overlaid with FDR-corrected t-statistics for each comparison. In each column, the TBM results are red: UT locally shrinking over time compared with either WT (columns 1 and 2) or TR (cols 3, 4); blue: expanding. C<sub>1</sub> and C<sub>2</sub> were combined in each case.

Very few, or no, voxels were significant between TP<sub>1-2</sub>, whether comparing UTs with WTs (a few bilateral reductions were detected in the cortex and forebrain) or TRs (none). The former comparison suggests that, rather than changes initially occurring in the entorhinal cortex or hippocampus, as would be consistent with human Braak staging (*Braak & Braak, 1995*), in this model of FTD-linked tauopathy, the earliest signs of degeneration occur in the cortex and forebrain. Although these are subtle effects, encouragingly, they are bilateral.

Between TP<sub>2-3</sub>, changes are more obvious: there is significant bilateral shrinkage of the cortex, striatum and hippocampus, and symmetric enlargement of the lateral, third and fourth ventricles, in UTs compared with WTs. The axial slices reveal expansion of eCSF pools in the entorhinal cortex region. These results suggest the most severe effects of tauopathy, manifesting as neurodegeneration, occur after approximately 150 days (5 months).

As expected, the differences between UT and TR<sub>DOX</sub> brains are less dramatic, but bilateral, and consistent with doxycycline's known ability to halt neurodegeneration in the model. The red regions of columns 3 and 4 show relative preservation of the TR group, compared with UTs, which shrink over time in comparison. The same regions are affected, but their extent is reduced, and the t-statistics, representing group separation, are lower. Doxycycline's preservative effects are, however, not detected with significant voxels until after TP<sub>2</sub>, 1—2 months after treatment began (*Table 6.1*).

## 5.6 The TIV covariate

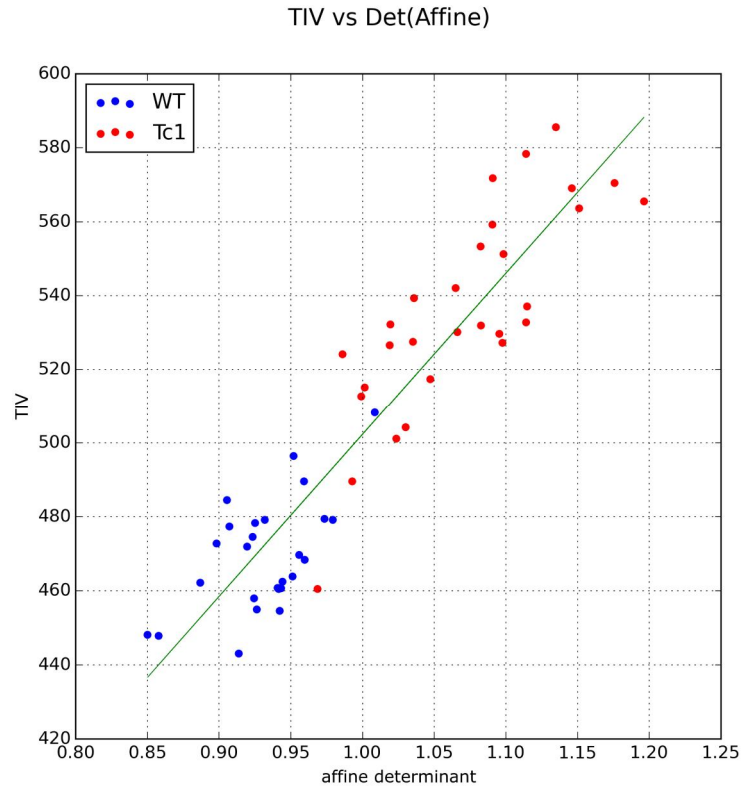
In both the Tc1 study (§4.1) and this rTg4510 chapter, consideration of the TIV covariate proved important for statistical analysis. The “TIV” measurement here is based upon binarised brain masks, including ventricular CSF and meninges (see Fig 5.19), but excluding some external CSF. Although it only approximates TIV, it is here likely closer to TIV than BV (GM and WM only), as in most cases there appeared to be a greater proportion of CSF in the ventricles than external spaces. The measurement’s accuracy could be improved by using the probabilistic output of STEPS, rather than a sum of the binary parcellations, ensuring all eCSF was included. (A more accurate, tissue-based estimation of each animal’s BV, which required further development, is included in §6.)

In healthy WT mice, the brain is tightly encased within the skull, and the ventricles are small. Especially in the UT rTg4510 group, however, BV decreases substantially with atrophy over time. As in human studies (*Whitwell et al., 2001; Scabill et al., 2003; Barnes et al., 2010*), I expected TIV to remain approximately constant<sup>8</sup>. This facilitates the simple comparison of cross-sectional and longitudinal measurements.

For morphometry, some studies remove global volume change from consideration in statistics by omitting the scaling factor applied by the affine matrix during GWR (e.g. *Xie et al., 2010*). This is the technique employed by Freesurfer (*Buckner et al., 2004*). Using the Tc1 and WT data from §4.1, I investigated whether the determinant of the affine matrices from the final iteration of affine registration during GWR correlated with the TIV measurement from tissue segmentation (§3.6; Table 4.1). Fig 5.13 shows the result. Pearson’s  $r^2 = 0.85$ ; the p-value on the slope was  $1.33 \times 10^{-23}$ .

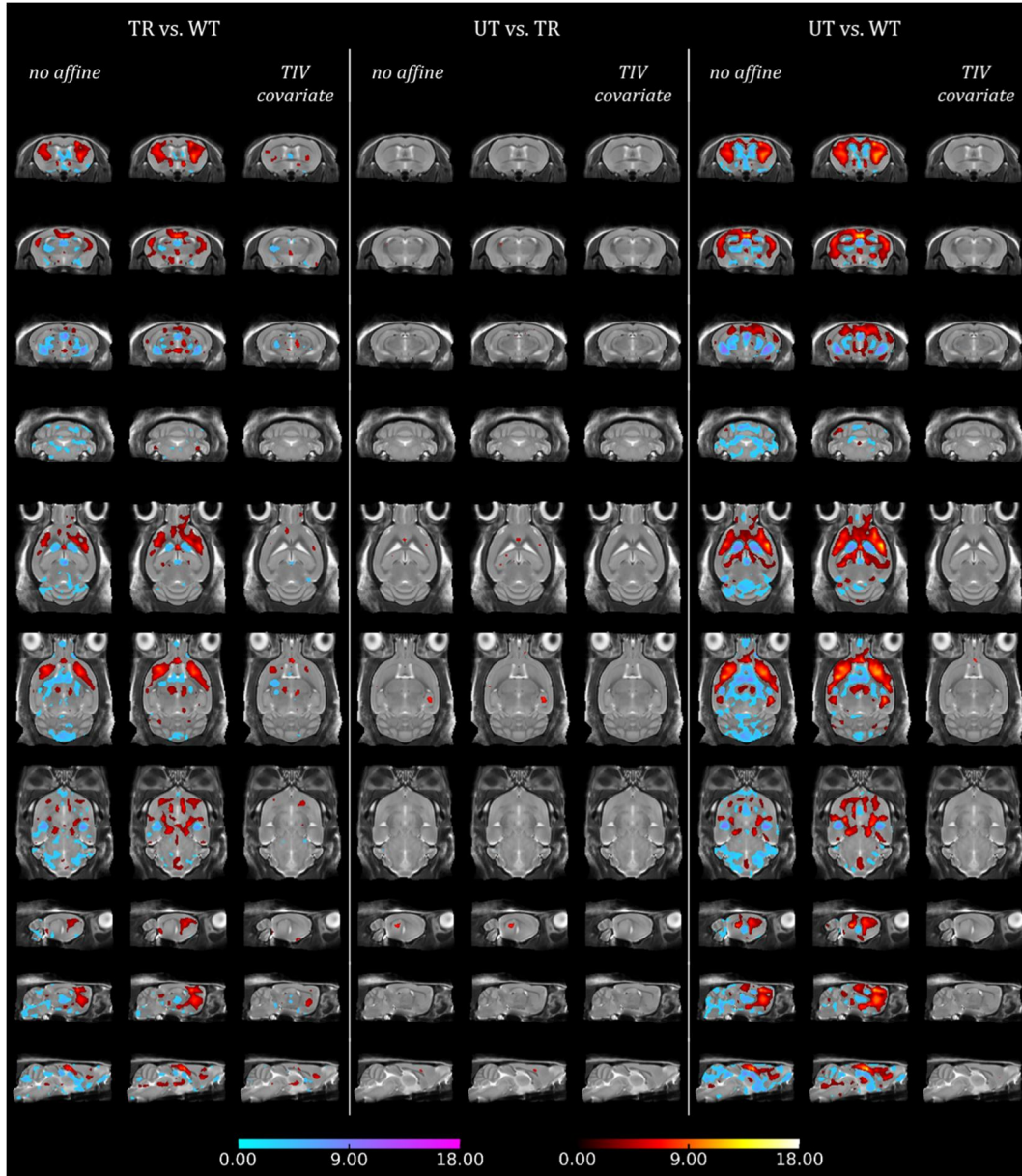
---

8: *Barnes et al., (2010)* reported a nonsignificant 0.1% TIV increase per annum in healthy human adults.



**Figure 5.13** Comparing the TIV measurement with the GWR affine determinant. WT and Tc1 TIV measurements and affine determinants from §4.1. The affine determinant scales each brain to the group average, so linearly correlates with tissue-derived TIV.

The linear relationship in [Fig 5.13](#) appears strong. However, the spread of points away from an exact linear relationship indicates that subtle effects differentiate the two measurements. The GWR is applied to voxels within a dilated brain mask. The registration will optimise the overall transformation using the skull, brain tissue, and any extraneous material included. It is therefore partially driven by variable tissues. The affine determinant will hence be an imperfect representation of consistent TIV, and may be biased by disease-related atrophy ([Ridgway et al., 2011](#)). [Malone et al. \(2015\)](#) found that segmentation-based estimates of TIV (e.g. in SPM12) outperformed affine registration-based methods.



**Figure 5.14** TBM results in rTg4510 brains, comparing affine and TIV covariates.

Three cross-sectional TBM comparisons (FDR corrected,  $q = 0.05$ ) overlaid on the same representative structural GWR average slices from  $C_1$  TP<sub>3</sub>: TR vs. WT; UT vs. TR and UT vs. WT. Within each, three versions of the statistics are shown: the first column omitting the affine from statistical tests (only NRR deformations tested); the second column including the affine; the third column, for each comparison, including the affine but also including the brains' TIV as a covariate in the GLM. Red:  $TR < WT$ ;  $UT < TR$ ;  $UT < WT$  locally; blue:  $TR > WT$ ;  $UT > TR$ ;  $UT > WT$  locally.

To investigate whether the two measurements had different effects on rTg4510 TBM results, I repeated the statistical tests of  $C_1$  TP<sub>3</sub> from the multi-TP study (§5.5.1), with the affine determinate excluded from the  $J_{det}$  values (to remove the global scaling from affine GWR); included, and included but with a TIV covariate to remove any global volume difference between animals' skulls. The results are shown in Fig 5.14. No additional smoothing was performed prior to statistics for these *in vivo* tests, as the combination of minor registration error, interpolation from resampling, and the larger voxel volume (compared with the *ex vivo* analysis) likely provided enough spatial smoothing.

In the comparison between UT and WT brains, removing the affine determinant had very little effect on TBM results, whereas including a TIV covariate in the GLM removed almost all voxels from significance. This suggests a strong linear relationship between the local volume reductions detected in the cortex, hippocampus and forebrain, as well as ventricular expansion. The relationship between affine determinant and TIV may be weaker in this dataset than in Fig 5.13. The TIV measurement also accounts for almost all the observed differences in the TR vs. WT and UT vs. TR comparisons. In the former, there remain some scattered, preferential local shrinkage in the forebrain not accounted for by global differences. At these locations, tissue atrophy may be particularly pronounced. In the latter, there are very few significant voxels which survive FDR correction in any comparison (as §5.5.1). The subtle differences between this TR<sub>DOX-4M</sub> group and the UTs are accounted for by TIV, but not by the affine determinant.

The strong correlation between all results and TIV justified my exclusion of the covariate when assessing cross-sectional results of the multi-TP study, above. Gross volume differences between groups, especially at this late TP, are visible by eye, particularly in the ventricles. However, these volume changes are not present globally throughout the brain: TBM remains useful to localise their concentration in the forebrain, hippocampus and cortex. The single-TP study (§5.4) included a TIV covariate, but many more statistically significant voxels survived FDR correction (Fig 5.6, c.f. column 3 of Fig 5.14). I attribute this improved sensitivity to the greater animal numbers of that study.



## 5.7 Comparing *ex vivo* morphometry to *in vivo*

Following their TP<sub>3</sub> scans, 25 C<sub>1</sub> mice were culled and perfuse-fixed using the protocol optimised for high-contrast *ex vivo* scans described in §4.1.2. After 9 weeks' fixation, the brains were secured in the 3-brain holder and MR scanned (acquisition parameters as per §4.1.2). 8 WT<sub>s</sub>, 10 UT<sub>s</sub> and 7 TR<sub>DOX-4M</sub> brains were included. This afforded the opportunity to directly compare the same brains with both *in vivo* and *ex vivo* techniques.

The purpose was to determine whether the compromises inherent to *in vivo*  $\mu$ MRI – including lower resolution (hence, greater partial volume and reduced contrast), potential movement artefacts, and lack of contrast agents – hamper TBM to such an extent that it is not as useful for detecting morphological differences between groups as *ex vivo* TBM. *Ex vivo* imaging has its own drawbacks, including the possibility of perfusion artefacts (§4.1.14; Cahill *et al.*, 2012), tissue dehydration and shrinkage (Zhang *et al.*, 2010; Lerch *et al.*, 2012) and possible damage (Ma *et al.*, 2008; Sawiak *et al.*, 2012). These may confound morphometric analysis. If equivalent sensitivity is achievable *in vivo*, a strong argument could be made for the exclusive use of *in vivo*  $\mu$ MRI, which allows other MR measurements, and consumes fewer scanner resources (including time), and, by enabling longitudinal studies, fewer mice. By comparing UT and WT, then TR and UT groups, I show the sensitivity of TBM to both large and small structural differences.

The first analysis undertaken was a repeat of the *ex vivo* TBM processing steps from §3, applied to this rTg4510 cohort. My colleague Da Ma also performed a structural parcellation comparison. Prior to processing, *in vivo* and *ex vivo* data were scaled according to a gradient-specific calibration (O'Callaghan *et al.*, 2014), to remove any global scaling effects of scanning each dataset with different coils.

### 5.7.1 TBM results compared, *in vivo* and *ex vivo*

For the *ex vivo* data, I applied the same complete set of image pre-processing steps for TBM as those described in §3 and §4, including multiple subject extraction and GWR. The GLM used for statistics was the same as that for *in vivo* C<sub>1</sub> TP<sub>3</sub>. Results from that

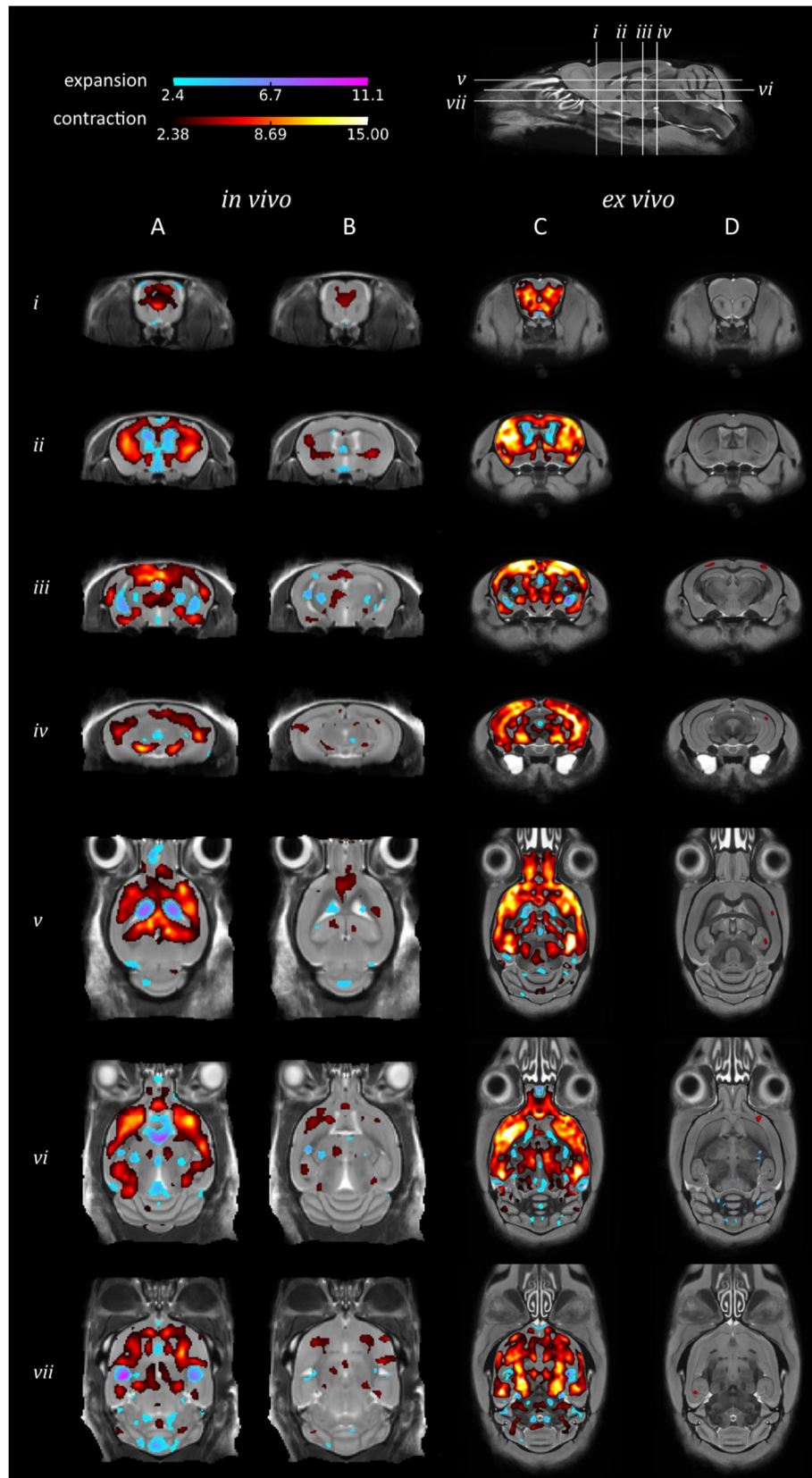


data (§5.5) are repeated here alongside *ex vivo* results. To discern its effect upon statistical maps in both datasets, and isolate local deformations, I tested both the inclusion and omission of a TIV covariate, where TIV was measured from binarised brain masks, including ventricles, as discussed above. Smoothing of the log-transformed  $J_{det}$  values was performed prior to statistics using a kernel with the same FWHM, 0.16mm, in both cases. All voxel-wise statistical results were corrected using FDR,  $q = 0.05$ .

In a comparison of UT and WT groups (Fig 5.15), both *in* and *ex vivo* TBM revealed substantial, extensive regions of significance when TIV was not covariated out (hence, in A & C, I included global scaling). Volume differences in the cortex, caudate putamen, hippocampus and thalamus were bilaterally significant in both, indicating widespread contraction of these tissues in the UT group. A greater proportion of voxels appeared significantly reduced in the caudate and cortex, and the t-score was generally higher, in the *ex vivo* comparison, despite slightly reduced subject numbers – indicating that, thanks to the increased local contrast available *ex vivo*, registration was better able to deform structures, and thus separate the two groups'  $J_{det}$  means. The *in vivo* voxels are limited by greater volume ( $150/40 \mu\text{m}^3 \approx 52.7$ ). More tissue types share a single sampled image region, increasing PV and homogeneity, and the available detail is thus lower.

Central and bilateral ventricular expansion was far more robustly recovered *in vivo*. Although the *ex vivo* images did reveal some ventricular (and entorhinal eCSF) change, t-statistic values were greater *in vivo*. This is likely due to the variable levels of ventricular collapse instigated by the perfusion-fixation process. Poor registration in these areas may contribute (see §5.7.4).

Using the TIV covariate (Fig 5.15 B & D), many of these regions disappear below the threshold of significance. *Ex vivo*, some small bilateral clusters of significant reductions survived FDR in the cortex (iii), and the hippocampus (v, vii). *In vivo*, more voxels retained significance, indicating local contractions (especially the cortex, hippocampus, and striatum) and expansion (the ventricles) in the UT group independent of the global volume difference. The process of tissue relaxation, dehydration and adjustment during fixation may have obscured true local volume differences – however, this would not explain the extensive significant regions seen without the TIV covariate.



**Figure 5.15** *In* and *ex vivo* TBM, comparing UT rTg4510s with WTs.

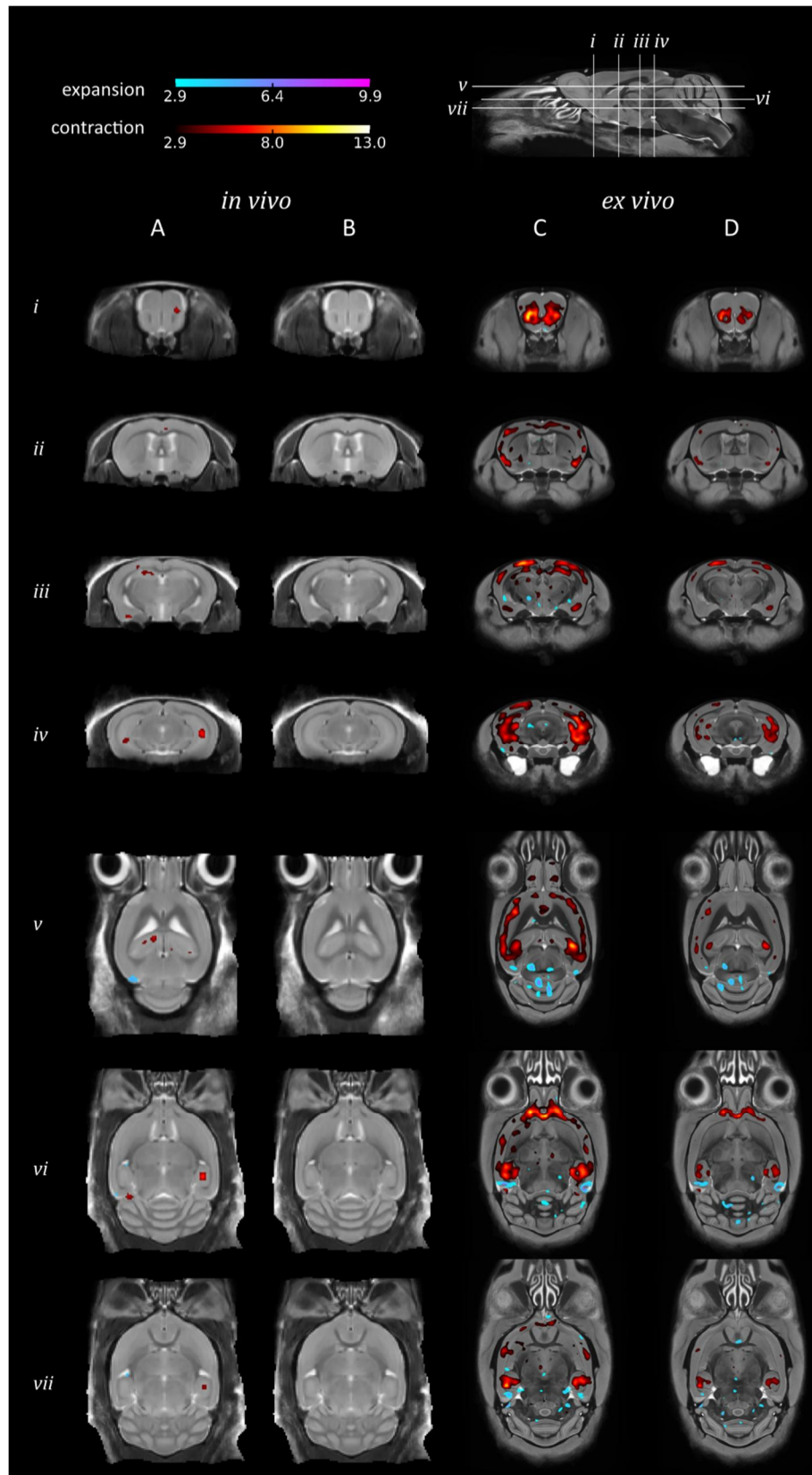
Results of *in vivo* (A, B) and *ex vivo* (C, D) morphometry (TBM two-tailed t-tests, after FDR correction  $q = 0.05$ ) overlaid on equivalent, representative coronal and transverse

slices of the final GWR structural average image, in each case. Without (A, C) and with (B, D) a TIV covariate in the GLM. Red: UT brains were relatively locally smaller than WTs; blue: larger. *In vivo* results from the same statistical tests on C<sub>1</sub> TP<sub>3</sub> as [Fig 5.10](#).

Instead, it is possible that the larger voxels *in vivo* allow small regions of volume change – homogeneous in intensity and thus poorly registering – to contribute collectively to larger structures with some elevated contrast, which do register well. These structures manifest the local changes which, *ex vivo*, are spread evenly across many voxels, and do not attain significance independently. This is similar to the improved power of parcellation-based volume measurements ([§4.1.5](#)).

In [Fig 5.16](#), I compare UT and TR<sub>DOX-4M</sub> groups. In this case, red regions represent relative preservation of the TR group in comparison with the UTs. Just as in [Fig 5.10](#), few local structural differences were detected in the late-treated group compared with UTs *in vivo*. There were some small indications of relative contraction bilaterally in the UT hippocampus ([A iv](#)). When TIV was used to standardise results ([B](#)), as above, these disappeared. The reduced contrast *in vivo*, in combination with the lower animal number (TR N=7), likely diminished the power of this test.

*Ex vivo* TBM encouragingly highlights more regions of significant difference, indicating bilateral conservation of the hippocampus, cortex, forebrain and olfactory bulbs in the TR animals. Again, this is likely due to the superior local contrast. Because the differences are subtler, the dispersion of significant voxels, and their t-scores, is reduced. Although expansion of the UT ventricles is highlighted by a small region of significance ([C v](#)), these suffered the same unpredictable collapse as described above. Thus, large significant changes are not discerned. As the *in vivo* brains' ventricles are not significant either, this may also be because intervention began too late to interrupt the period of ventricular expansion. Conversely, the entorhinal cortex eCSF region is highlighted by robust regions of significant bilateral expansion in the UTs ([C vi](#)). Because it is not surrounded by tissue, external pressure, and hence collapse, is not an issue during perfusion-fixation.



**Figure 5.16** *In* and *ex vivo* TBM, comparing TR rTg4510s with UTs.

*In vivo* (A, B) and *ex vivo* (C, D) TBM results, as [Fig 5.15](#) (two-tailed t-tests, with the same slices and FDR  $q = 0.05$ ). (A, C): without TIV covariate in the GLM; (B, D): with TIV.

Red: UTs locally significantly smaller than TRs; blue: larger. The most prominent changes *in vivo* and *ex vivo* are in the hippocampus. *Ex vivo*, substantial forebrain and cortical differences are also detected.

More of these regions survive with the TIV covariate (*D*) in this comparison than UT vs. WT, or indeed the *in vivo* set. This is possibly because TIV explains a greater proportion of the differences between UT and WT brains than TR and UT brains: in the latter comparison, significant changes are concentrated in the regions highlighted, and not attributed to global volume differences. The brains' initial volumes, at TP<sub>1</sub>, were indeed more similar to one another than to WTs.

The parcellation results (not shown in figures) quantified a global shrinkage, across groups, from *in vivo* to *ex vivo* (excluding the ventricles, WT: 7.5%; TR: 9.3%; UT: 9.6%, all significant within-group paired t-test results with  $p < 0.01$ ). While the overall correlation between equivalent structures, across groups, was strong ( $r^2 = 0.98$ ), shrinkage was non-uniformly distributed between structures. The cortex was relatively preserved; however, the ventricles almost completely collapsed (85.7% volume decrease) – confirming visual observations. This was expected: previously published work has implicated tissue fixation with dehydration and subsequent global shrinkage. *Zhang et al. (2010)* observed 4.4—7.8% whole-brain volume reductions, and concluded that, despite dramatic shrinkage of the ventricles, the structure of the rest of the brain was preserved sufficiently for morphometry. *Ma et al. (2008)* measured a 10.6% decrease in C57Bl/6J WT brain volumes, and similarly observed ventricular collapse.

### 5.7.2 Local structural differences between *in vivo* and *ex vivo*

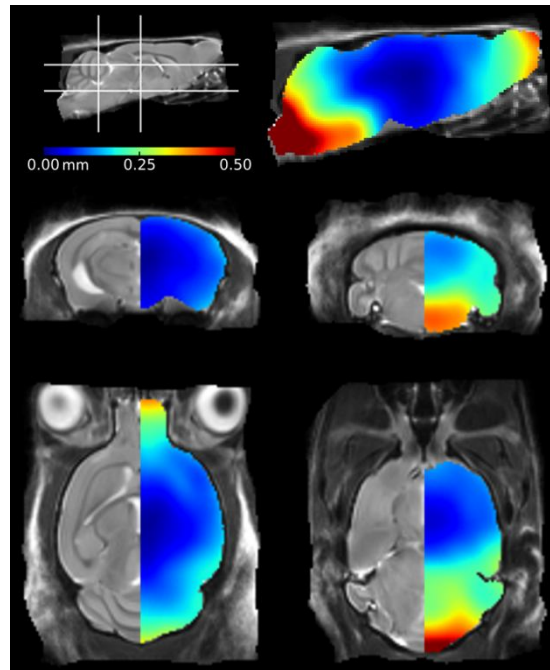
To measure, and visualise, the degree of local variability between *ex vivo* and *in vivo* images, I used a method similar to that employed by *Ma et al. (2008)*<sup>9</sup>. I registered each of the 25 *ex vivo* brains to their corresponding C<sub>1</sub> TP<sub>3</sub> *in vivo* counterpart, and calculated

---

<sup>9</sup>: Their Fig 8 (c & f) inspired this analysis, but is not exactly equivalent, as their registration was group-wise instead of the pair-wise registration here.

the mean positional distance (MPD) at each voxel (E3.15), after resampling the resulting deformation fields into the *in vivo* brains' average space, from the GWR of C<sub>1</sub> TP<sub>3</sub>.

Fig 5.17 shows the mean distance moved by each voxel during registration, when aligning *ex vivo* images pair-wise to in situ, *in vivo* images, from all three cohorts (NRR only; rigid and affine transformations were excluded). It thus shows physical differences between the brains, as scanned. This could tell us whether the *ex vivo* preparation of mouse brains added significant structural deformations. Coarsely, the figure could be considered a map of the degree of reliability of morphometry from *ex vivo* structural scans, compared with the *in vivo* paradigm, which is more likely to reflect in situ biological reality. As might be expected, the greatest displacements (between 0.25 and 0.5mm) occur within the brainstem, which was in different positions during *in vivo* and *ex vivo* scans – and was subject to different FOV placements or the position of the cut, when the skulls were removed from the mice.



**Figure 5.17** Mean positional distance image showing the mean distances moved by voxels during pair-wise registration from *ex vivo* to *in vivo* data.

Deformation maps showing the local distortions of *ex vivo* brains (the distance travelled, in mm) compared with *in vivo* counterparts. Voxel-wise deformations were calculated and resampled into the same average space for comparison.



The rostral aspect of the olfactory bulbs was also affected, although as the *ex vivo* scans were in skull, the displacements were not of the same degree as those seen by [Ma et al. \(2008\)](#). The ventricles do not show a great deal of variability between *in* and *ex vivo*. This may be because this group average, including WT, UT and TR groups, has such large variability that the vectors are averaged out. It may also be because the images are similar: as the registration was pair-wise, large deformations should not necessarily be expected between *in* and *ex vivo* brains of the same group. In the *ex vivo* scans, as in prior datasets, the ventricles usually partly or fully collapsed during perfusion and fixation (though not always; see [Fig 5.19](#)). Registration sometimes failed to capture this gross deformation (see [§5.7.4](#)). This could perhaps be corrected via the use of a more liberal bending energy penalty term in NRR – however, as the term is global, other regions of the brain, which do not require such freedom to register correctly, may then deform incorrectly.

The remainder of the olfactory bulbs and the cerebellum also showed displacements of between 0 and 0.25mm (under 2 voxels *in vivo*). The rest of the brain, including the midbrain, hippocampus, thalamus and cortex, was relatively unaffected *ex vivo*; I attribute this preservation to keeping the brains in skull for scanning, to prevent damage. It is notable that the cortical surface shows very few signs of systematic perturbation, which bodes well for the validity of *ex vivo* morphometric results in this structure.

Another possible cause of the apparent displacements seen in [Fig 5.17](#) is the different sequences used for *in vivo* (FSE) and *ex vivo* (GE) acquisition. The latter may have induced some geometric distortion away from the bore isocentre, giving a similar pattern. However, the *ex vivo* brains were always positioned in the scanner so that their centres were aligned with the isocentre. According to [O’Callaghan et al. \(2014\)](#), gradient warp distortion was below 0.1mm within  $\pm 20$ mm of the isocentre; mouse skulls easily fit within this window. Additionally, the displacements seen caudally appear predominately in the brainstem (rather than equally in the cerebellum, directly above), so it is more likely that physical distortion caused the dominant appearance here.

### 5.7.3 Power analysis

With enough subjects, any group difference between means can be declared “significant” (Cohen, 1994). TBM results themselves are often insufficient; what is important is a measure of the effect size; the magnitude of the physical difference between groups’ local volumes, and whether this magnitude is scientifically interesting (Coe, 2002).

As described, the *ex vivo* and *in vivo* groups underwent separate GWR and TBM analysis. Each dataset may be treated as a preliminary study from which to estimate the number of mice  $N$  (per group) required in a future *ex* or *in vivo* study to show meaningful differences between the WT and UT groups, with specified statistical power. I chose a conventional significance level of  $\alpha = 0.05$ , a false negative rate  $\beta = 0.2$ , and initially assumed the WT group provided a reasonable estimation of local structural variability within the rTg4510 mouse population. To detect a 25% local volume difference between WT and UT groups measured in  $J_{det}$  values, at each voxel, the required number of animals in each arm of the study is (Florey, 1993):

$$E5.2 \quad N_{arm} = \frac{(Z_{\alpha/2} + Z_{1-\beta})^2 \times 2 \times WT_{stdev}^2}{(0.25 \times WT_{mean})^2}$$

where  $Z_{\alpha/2} = 1.96$  is the approximate number of standard deviations from the mean of a standard normal distribution of  $1 - (\alpha/2) = 0.975$ ; likewise  $Z_{1-\beta} = 0.84$  for  $1 - \beta = 0.8$ .  $WT_{mean}$  and  $WT_{stdev}$  are the mean and SD of the WT  $J_{det}$  values, from GWR, at a particular voxel. These vary across the brain depending upon local WT variability.  $N_{arm}$  is rounded up to the nearest integer.

As the actual difference between mean UT and WT  $J_{det}$  values is known *post-hoc*, the effect sizes (Cohen’s  $d$ ), both *in vivo* and *ex vivo*, may be visualised and compared at each voxel. This helps to quantify which regions are most different, between groups, regardless of statistical significance:



E5.3

$$\text{Cohen's } d = \frac{UT_{mean} - WT_{mean}}{WT_{stdev}}$$

I calculated  $d$  using a pooled standard deviation, replacing  $WT_{stdev}$  in E5.3 with  $\sigma_{pooled}$ :

E5.4

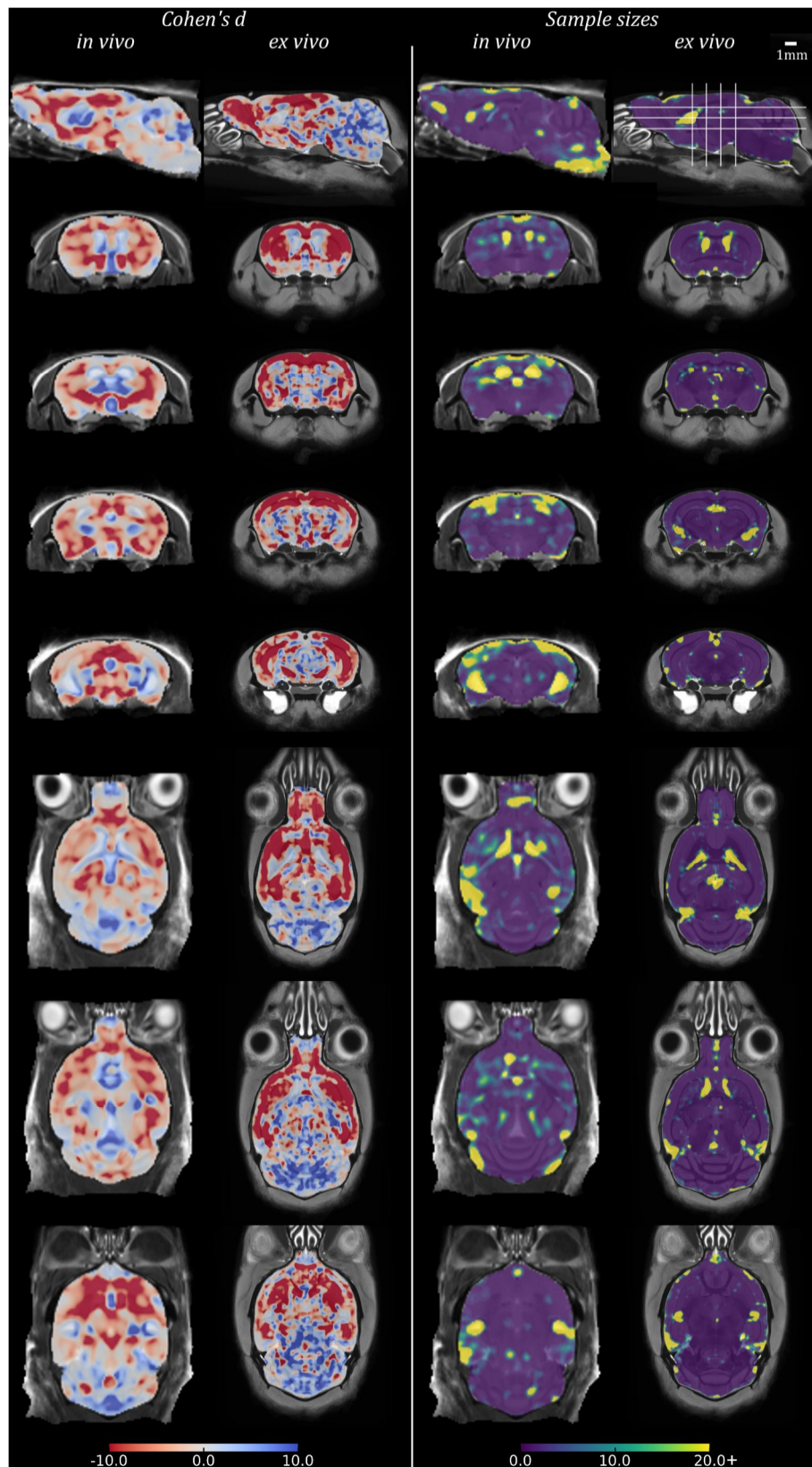
$$\sigma_{pooled} = \sqrt{\frac{(WT_{stdev}^2 \times (N_{WT} - 1) + UT_{stdev}^2 \times (N_{UT} - 1))}{N_{WT} + N_{UT} - 2}}$$

Because the  $WT_{stdev}$  was usually smaller than the overall standard deviation, and therefore not representative of the whole cohort, I made the same substitution in E5.2.

Fig 5.18 shows equivalent slices from the *in vivo* and *ex vivo* group-wise average images, overlaid with the  $d$  and  $N_{arm}$  required at each voxel to detect a 25% variation in  $J_{det}$  from the WT mean. We might wish for greater power (e.g.  $1 - \beta = 0.99$ ), or to detect a subtler effect (e.g.  $0.01 \times WT_{mean}$ ) – however, these estimates of  $N$  give a realistic lower bound for a future study.

In each case, *in vivo* and *ex vivo* exhibit similar patterns. The map of Cohen's  $d$ , in measuring group mean separation, is similar to the thresholded statistical maps of Fig 5.15. The separation of groups in the hippocampus and cortex is clearly greater *ex vivo*. This may be facilitated by higher contrast enabling better alignment during registration, reducing local standard deviation. The greatest group separation, *in vivo*, is achieved in the hippocampus and striatum.

The  $N_{arm}$  required to detect a significant effect, with the specified  $\alpha$  and  $\beta$ , is elevated around the ventricles and at the entorhinal cortex – likely both due to reduced WT volumes and increased variability around these structures. This may explain the greater numbers required in some regions of the *in vivo* group. That statistically significant results are still obtained in these regions, despite this study incorporating fewer animals, is due to the actual difference between groups being greater than  $0.25 \times WT_{mean}$ .



**Figure 5.18** Power analysis results in rTg4510 data: Cohen's  $d$  and sample sizes. Equivalent sagittal, coronal and transverse slices on *in vivo* and *ex vivo* GWR average images, overlaid with the Cohen's  $d$  (left) and sample sizes  $N$  (right) required to show a

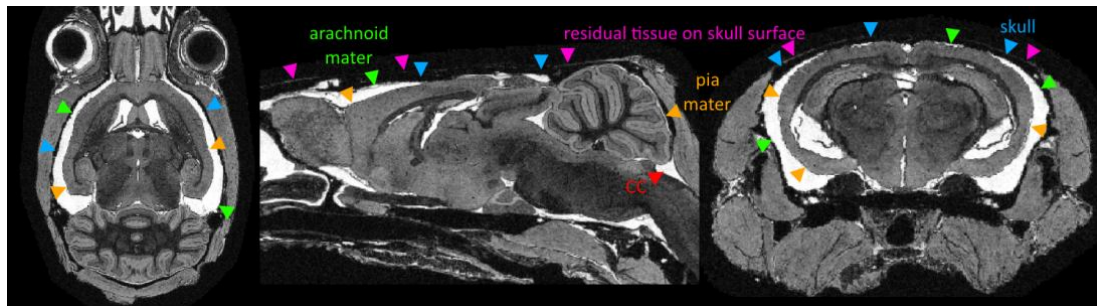
significant effect, with  $\alpha = 0.05$ ,  $\beta = 0.2$ , and an effect size 25% of the local WT mean volume. For Cohen's  $d$ : red, UT group locally smaller than WTs; blue: larger. Solid lines, top left, indicate slice locations.

From [E5.2](#), the power of TBM to detect a local volumetric expansion or contraction is dependent upon the standard deviation of  $J_{det}$  values in the groups at each voxel,  $WT_{stdev}$  (or  $\sigma_{pooled}$ ). This variability is itself dependent upon the ability of the registration algorithm to successfully align equivalent voxels, which depends upon constraint parameters (bending energy and control point spacing) as well as on the inherent image contrast of the brain structures. Low contrast regions are more likely to register poorly: structures with relatively homogeneous intensities have less information to inform registration. Here, misalignments may arise. This may result in greater local  $WT_{stdev}$  (or  $\sigma_{pooled}$ ), and hence a higher local effect size (or  $N_{arm}$ ) requirement in order to achieve the desired power. This appears to be the case in the midbrain and striatum of the *in vivo* group, where there is lower structural contrast than *ex vivo*, and registration consequently struggles to accurately align voxels.

The formula [E5.2](#) itself is well-known. However, voxel-wise implementations of this method in TBM studies are uncommon in the literature – even in humans. This may be because post-hoc tests are undesirable (although they may be interpreted to inform future studies); it may be an incidence of the frequently-noted observation ([Cohen, 1994](#)) that power analyses are uncommon; or it may be that their interpretation is not simple over millions of voxels. As the  $WT_{stdev}$  and  $WT_{mean}$  vary between brain regions, the power of each test varies and different sample size estimates are prescribed, throughout the brain. To obtain a useful sample size for a study focussing on a specific brain region (e.g. the hippocampus or cortex, in the rTg4510s), it would be sensible to use that region to set the  $N$  per study arm. [Maheswaran et al. \(2009a\)](#) reported that voxel-wise techniques required reduced sample sizes in several brain regions, compared with atlas-based techniques. Several other groups have reported the effect size of their TBM measurements, at specific voxels (as [§4.1.8](#), e.g. [Hua et al., 2008](#); [Lerch et al., 2008](#); [Badhwar et al., 2013](#)).

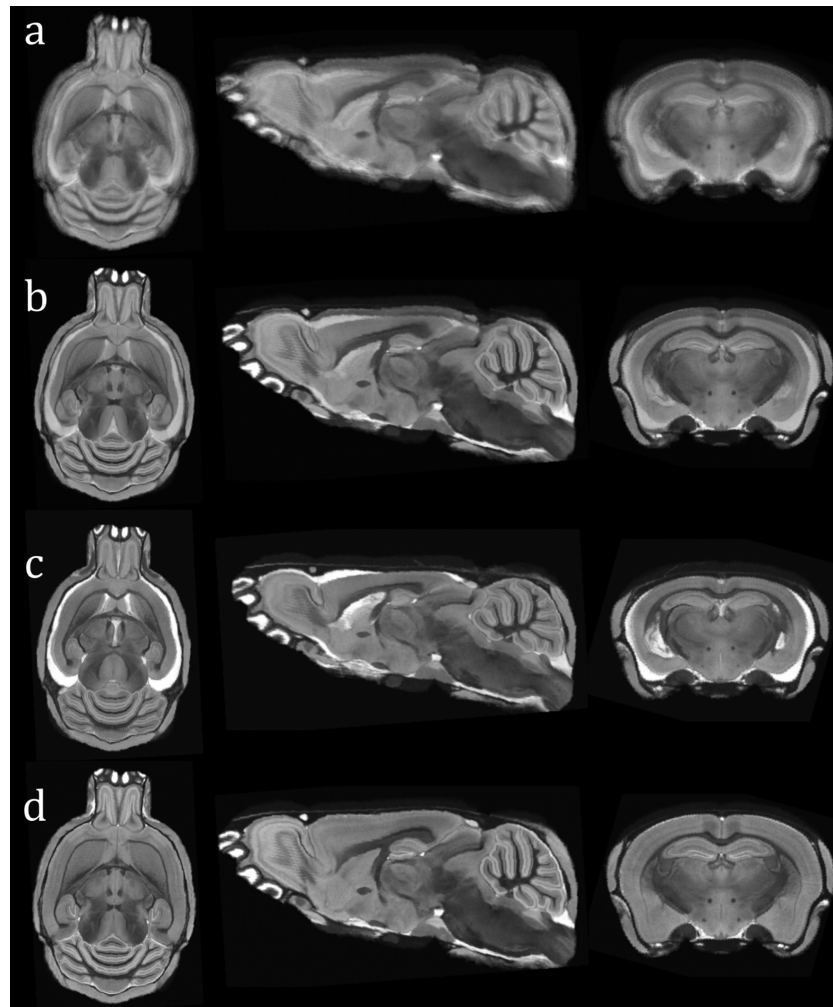
### 5.7.4 Fundamentally differing morphologies

Registration may fail between two images when one lacks features present in another: large ventricles in one brain image may incorrectly register to another whose ventricles have collapsed, leaving no visible CSF, for example. The presence and location of eCSF is especially variable in the rTg4510s. The large negative intensity gradients from eCSF to BG, or from cortical GM to BG, may be incorrectly deemed equivalent by the gradient-driven registration algorithm of *NiftyReg*, and eCSF and GM surfaces in different images may thus align. This problem afflicted the *ex vivo* images prior to registration parameter optimisation. During GWR, mid-grey intensities resulted where CSF and BG or brain tissue improperly overlap at the brain surface, producing artificial layers in the average image (later used as a registration target), which are not biologically meaningful or present in any one structural image (*Fig 5.20*).



**Figure 5.19** Selected anatomical features of the meninges and skull in an *ex vivo* rTg4510 brain (10 months of age), after severe atrophy due to tauopathy.

From the outside, layers include: surface flesh (a very thin residual layer indicated in pink, after the flesh has been removed); skull (blue), which tightly encases the brain and produces minimal signal; meninges; brain. The meninges include, again from the outer layer: dura mater (relatively hard and thick); arachnoid mater (green: thin, attached to the dura, like a loose-fitting sack); pia mater (orange: tightly seals the brain surface). CSF is contained within the “subarachnoid” space, between the arachnoid and pia mater. CC: cisterna cerebellomedullaris (*Drake & Lowrie, 2009; McGavern & Kang, 2011*).



**Figure 5.20** GWR averages of an *ex vivo* rTg4510 and WT littermate cohort (10 months of age), after one iteration of...

(a): ...rigid-only registration; (b): NRR; (c): NRR, rTg4510 only; (d): NRR, WT only.

While the cerebellum is sharply registered, the cortical surface between rTg4510 and WT images (a, b) did not correctly register: instead, mid-grey regions (average of GM and CSF, or GM and BG) form artificial layers in the average image. Within-group registrations (c, d) were more successful. In the WT, as is typical for *ex vivo* images, the ventricles almost completely collapsed upon fixation. The rTg4510s' large ventricles survive.

The rTg4510 mice undergo severe brain atrophy when the tauopathy is allowed to progress without suppressive drugs (such as doxycycline). This atrophy occurs predominately in regions analogous to the human brain: the ventricles, entorhinal cortex, and eventually the whole cortex. [Fig 5.19](#) shows the brain of a 10 month old rTg4510 mouse with large accumulations of CSF both in the expanded ventricles, and externally,

where tissue has atrophied. Steps were taken during fixation, such as using a low flow rate, to avoid pressure build-up and hence ventricular collapse. Here, the ventricles and eCSF spaces have been preserved and are likely filled with a mixture of CSF and fixative.

The eCSF in rTg4510s is thought to arise entirely from tauopathy, being consistently located within the untreated group and more severe ([Fig 5.19](#)) than that seen in the Tc1 cohort ([Fig 4.10](#)), with perfusion-linked external fluid. This variability could cause severe inter-group registration artefacts ([Fig 5.20](#)).

## 5.8 Discussion

### 5.8.1 Cross-sectional and longitudinal TBM

The rTg4510 dataset described in this chapter holds great promise for the monitoring of disease and drug intervention. Scanning WT and UT mice alongside treated groups is vital to provide controls exhibiting “normal healthy ageing” or natural disease progression. The effectiveness of therapeutics must be measured as divergence from this UT trajectory (see also [§6](#)).

The single-TP TBM investigation of rTg4510 brains encouragingly repeated previously published findings ([Xie et al., 2010](#); [Yang et al., 2011](#)) of localised volume reductions in rTg4510s, attributable to NFTs, compared with WT littermates. These volume differences were severe at the late time-point chosen, and spread throughout the forebrain, cortex, hippocampus and caudate putamen, supporting tau-driven neurodegeneration.

The purpose of the multi-time-point study was to assess the sensitivity of structural MRI, analysed using TBM, to these neurodegenerative effects at earlier TPs, and to the ability of two therapeutic interventions, doxycycline and EpoD, to modulate this detected atrophy. Other MRI measures shared each mouse’s total 3 hours of scan time (per TP), including CBF and DTI.



None of the cross-sectional TBM analyses was able to detect a significant difference between EpoD-treated animals and UTs, despite the application of comparable doses to a previous study (*Zhang et al., 2012*), in which EpoD treatment led to reductions in tau pathology and improvement in hippocampal neuronal integrity, measured via histology. Corroborating this negative finding, DTI measures (mean diffusivity, MD and fractional anisotropy, FA, in the high-NFT-burdened cortex) were also unable to distinguish a neuroprotective effect of EpoD, although both cortical CBF and its atlas parcellation-derived volume were significantly different from UTs in the high-dose group at the final TP, perhaps indicating preservation (data not shown).

TBM was, however, sensitive to the doxycycline treatment, and its ability to halt the progression of tauopathy-linked neurodegeneration was observed here, distinguishing UT, late, and early-treatment groups from WTs, as well as TR groups from UT. TBM therefore provided a sensitive downstream indicator of therapeutic effectiveness. In these mice, the measurement was not, however, reliably able to discern very early atrophy: cross sectional significant volume differences between groups at TP<sub>1</sub> were limited. The detection of such early differences, in humans, will be vital for intervention to successfully interrupt cognitive decline (*Weiner et al., 2013*).

Histology results (not shown) confirmed NFT deposition, with high densities in the cortex and hippocampus of UT animals, and comparatively low density in the thalamus. The TR<sub>DOX-3M</sub> (early) group had lower densities in these regions than the late-TR group, validating the correlation between NFTs and atrophy detected by TBM. The DTI measures, cortical FA and MD, were both able to distinguish TR groups from UT, and UT from WTs, but only at the final TP: structural MRI was a consistent, earlier surrogate marker of change.

It is important to note that doxycycline halts atrophy progression, saving tissues from future degeneration – but it does not promote re-growth. As expected, the earlier intervention appeared to substantially benefit brain preservation. The gradual increase in the sensitivity of TBM, at successive time-points, to gross progressive morphometric brain changes across groups would likely be improved by greater animal numbers.

Although groups of rTg4510 mice may be distinguished by eye at late time-points, the changes within each group over time are subtler and less easily discerned. My longitudinal inter-TP analysis demonstrates the versatility of registration and  $J_{det}$ -based techniques. It would be interesting, for drug studies, to observe WT changes over time, using a larger cohort and deformation magnitude maps, rather than thresholded statistics: only a few papers have reported longitudinal healthy ageing volumes of WT mouse brains (e.g. [Lau et al., 2008](#); [Maheswaran et al., 2009a](#); [Zhang et al., 2010](#)).

[Hua et al. \(2013\)](#) measured the power of TBM to detect atrophy rates in a longitudinal human study of the ADNI MR dataset. They noted an approximately linear rate of brain atrophy over two years. In the longitudinal TBM measurements here, the TP<sub>2-3</sub> pair showed a markedly greater degree of atrophy than TP<sub>1-2</sub>, controlling for the inter-TP period. This suggests that atrophy rates in rTg4510 mice are non-linear, and appear to accelerate in the time window studied here. A longer study would be necessary to ascertain the continued trajectory of this decline. It would be interesting, additionally to [Fig 5.12](#), to replicate [Hua et al. \(2013\)](#)'s measurements of within-group longitudinal atrophy, which use  $J_{det}$  values to show voxel-wise cumulative volume change over time.

A disadvantage of structural imaging for TBM is the relatively long (1.5-hour) acquisition time, which, while possible preclinically, could ideally be lowered to increase throughput. Reducing this may compromise resolution, SNR or CNR. Non-isotropic voxels could be used instead. However, an increase in slice thickness, maintaining in-plane resolution, would increase PV, and lower the performance of registration and TBM. This would likely have to be balanced with increased subject numbers to preserve statistical power. Nonetheless, it would be interesting to perform experiments in this direction.

### 5.8.2 *In vivo* versus *ex vivo*

Before undertaking a morphometric investigation of mouse brains – of which, thanks to large-scale phenotyping efforts and the need for accurate disease models and drug studies, there are an increasing number – it is important to consider whether *in vivo* or *ex vivo* imaging and analysis techniques would be superior, to detect differences between groups, or therapies' disease-modifying effects.



The *ex vivo* protocol employed here was optimised, as far as possible, for image quality and contrast (Cleary *et al.*, 2011b). Some distortion due to the perfusion and fixation process is inevitable. The ventricles collapse to an unpredictable degree, and the brainstem, which is physically severed, is moved. Had the brains been extracted from skulls, the olfactory bulbs would also have been free to move and the cortex highly susceptible to damage (Ma *et al.*, 2008). Despite the fixation-linked tissue shrinkage, as Zhang *et al.* (2010) observed, morphometry was here still able to resolve subtle and important local physical differences between the TR, UT and WT groups.

The spatial specificity of *in vivo* TBM is compromised by the necessity of larger voxels. Nonetheless, the voxel-wise statistical tests performed here were highly localised in comparison to atlas-based parcellation techniques, which are limited in resolution by the atlas labels. In comparing *in vivo* and *ex vivo* imaging, Zhang *et al.* (2010) noted that *in vivo* parcellation was unreliable owing to the reduced contrast. For parcellation and skull-stripping, it is important to select an appropriate atlas (an *in vivo* atlas would register better with *in vivo* data). Even though *in vivo* TBM performed well, and highlighted similar structural differences between groups, it should be noted that the imaging protocol described (§5.2.3) only used half the available three hours for which, under project license authority, animals could be anaesthetised. Other MR measurements shared the time. A longer scan might improve resolution or SNR, which was nevertheless comparable to the multi-brain *ex vivo* acquisitions (§4.1.5, Table 5.2), perhaps due to the larger voxel size *in vivo* (improving signal, Kale *et al.*, 2008) and dehydration effects inherent to *ex vivo* tissue fixation (decreasing proton density). A contrast agent, such as manganese (avoided here, as it could interact with the pharmacology of the therapeutics, or cause additional neuron loss), could also improve CNR *in vivo*. The comparison performed was not, therefore, between the best *possible ex* and *in vivo* images. Such a compromise is, however, realistic, and likely to be repeated in future preclinical imaging studies: one advantage of *in vivo* MR is the ability to collect other MR measures.

For increased sensitivity *in vivo*, it may yet be practical to use a manganese contrast agent. Allemang-Grand *et al.* (2015) investigated the effects of repeated manganese exposure on the *in vivo* mouse brain, in a longitudinal study of early-onset AD, and found no effect

on neuroanatomy or survival rate. The same group reported a similar direction and location of volumetric changes detected in lower-contrast images without manganese.

Given observed variability of WT mice, [Lerch \(2010\)](#) recommended using 10 mice per study arm, to detect at least 17% volume differences, in morphometry studies. Given [Fig 5.18](#) (showing numbers for 25% differences), this estimate appears reasonable, although it is important to note that WT variance is spatially-dependent, and more animals may be required if subtle differences are expected in regions of high variance, such as the ventricles.

[Lerch et al. \(2012\)](#) recommend the use of more mice, rather than more time-points, *in vivo*, to detect subtle effects, as they found structural variation across populations was greater than between measurements. A similar conclusion could be made here; the single-TP *in vivo* investigation, with more mice, yielded greater TBM sensitivity than the oldest TP of the multi-TP study, despite comparable mouse ages. However, this must be balanced by ethical considerations for the number of mice employed, and the time and cost of scanning.

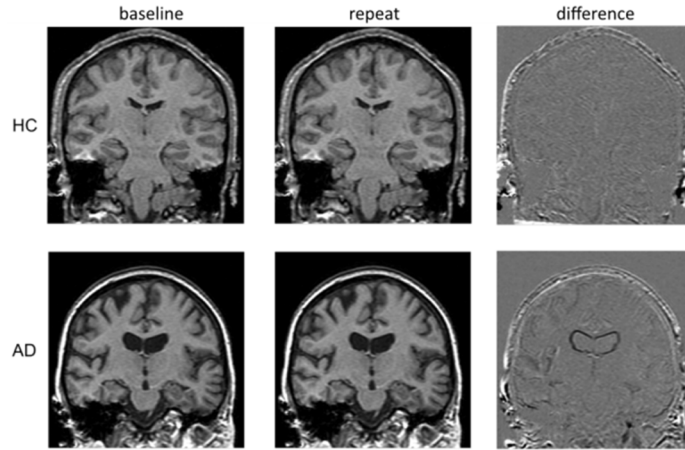
For this purpose, power analyses are invaluable. Likely thanks to increased contrast and improved registration, *ex vivo* TBM requires fewer animals to detect subtle differences between groups. It was also better able to localise those small differences, thanks to greater resolution. This does not preclude the usefulness of non-invasive, *in vivo* MRI, however. In the final chapter, I develop and investigate another longitudinal measure of structural volume change: the Boundary Shift Integral.

# 6 Longitudinal atrophy measurement using the Boundary Shift Integral in the mouse brain

The previous chapter focussed on the application of the V/TBM pipeline to perform both cross-sectional and longitudinal analysis of *in vivo* data using group-wise morphometric techniques and voxel-wise statistical tests. This chapter introduces a new technique for longitudinal analysis of individual animals' scans. The Boundary Shift Integral (BSI) is a sensitive, robust measure of tissue volume change in human MRI, and has become a popular method employed in clinical trials to assess neurodegenerative disease progression and, potentially, drug efficacy. The measurement uses voxel-wise intensity differences between aligned serial MRI scans to estimate the rate of volume change. This is constrained to structures' boundaries – the cortical surface; ventricles' edges – where atrophy is most discernible ([Fig 6.1](#)). Until now, it has never been applied outside of human studies – nor to T<sub>2</sub>-weighted (T<sub>2</sub>W) MR images.

In this final experimental chapter, I describe and demonstrate the adaptation of the BSI to T<sub>2</sub>W mouse brain images, using the rTg4510 mouse model of tauopathy introduced and investigated earlier. After testing different parameters to optimise the measurement, I used the BSI to record atrophy in wild-type (WT) and untreated (UT) mice, and groups treated (TR) with different drugs. The BSI distinguished reduced atrophy rates, with high

sensitivity, in doxycycline-treated groups. Epothilone D (EpoD) had a much smaller, but discernible, beneficial protective effect against neurodegeneration, in the short term.



**Figure 6.1** Progressive atrophy in the human brain revealed with difference images. T<sub>1</sub>W brain MRI of healthy control (HC) and Alzheimer’s disease (AD) patients, with baseline and registered repeat scans some time later (usually six months – one year). Equivalent coronal slices shown. The difference image highlights the change in position of the ventricle boundaries, as well as around the cortical surface and at the hippocampus, in AD. Note in the AD patient, the intensity difference at the cortical surface is the same as that at the ventricle boundary: a dark rim. Adapted from a presentation by Professor Daniel Rueckert (2013).

Ferran Prados developed the Generalised BSI (gBSI) implementation ([Prados et al., 2014](#)), and code for symmetric registration and multi-time-point differential bias correction (MTPDBC), at CMIC. Kelvin Leung coded the k-means (KN-BSI, [Leung et al., 2010b](#)) and original double window BSI (DW-BSI) implementations. I show modifications here for T<sub>2</sub>W mouse MRI.

## 6.1 Introduction

The early signs of neurodegeneration may fall within the normal range of tissue loss seen in healthy ageing or natural volume difference between subjects. Errors in manual and

segmentation-based methods of volume measurement may be of the same order as these natural differences (Fox & Freeborough, 1997). Expert manual segmentation is the current “gold standard” technique, but has high inter-animal and intra-rater variability. Such errors are likely to be compounded in the mouse brain, where structures are much smaller and partial volume more significant. An accurate, objective, quantitative and repeatable measure of tissue change over time within a single subject is therefore likely to be advantageous for the early detection of pathology, and to quantify the efficacy of disease-modifying drugs over short timescales, particularly in preclinical trials (Freeborough & Fox, 1997).

Imaging biomarkers are increasingly used to detect, diagnose and stage neurodegenerative disease (Rudin, 2007, Gustaw-Rothenberg et al., 2010). The BSI, a sensitive, registration-based measure of tissue volume change, has been shown to robustly quantify the rate of cerebral structure atrophy of in Alzheimer’s disease (AD, Freeborough & Fox, 1997; Leung et al., 2012), frontotemporal dementia (FTD, Chan et al., 2001), and Huntington’s disease (Hobbs et al., 2009; Tabrizi et al., 2013), in humans, by measuring the volume through which their boundaries shift in serial MRI.

In AD, atrophy occurs diffusely and inhomogeneously throughout the brain; not exclusively at the cortical and ventricular boundaries highlighted in Figs 6.1, 6.2. As the gross tissue structure adjusts to this cell death, volume change manifests at these surfaces (Freeborough & Fox, 1997). This may be measured directly, using changes in voxel intensities.

Atrophy is a macroscopic proximity marker of disease progression: it is dynamic at mild cognitive impairment (MCI) symptom onset (see Fig 2.13), and correlates well with cognitive test score decline and neuronal loss (Fox et al., 1999). Structural imaging biomarkers thus show potential as surrogate, objective measures of disease progression and drugs’ disease-modifying effects. The BSI has been used to reveal subtle atrophy rates distinguishing healthy ageing from MCI and AD over 6 month – 2 year periods (Freeborough & Fox, 1997; Schott et al., 2005; Leung et al., 2010a; Weiner et al., 2013; Gutman et al., 2014). It compares well with the accuracy of manual segmentation (Barnes

*et al.*, 2008), and can be fully automated (*Prados et al.*, 2014). However, the BSI has, as yet, never been applied outside of human studies, nor to my knowledge to T<sub>2</sub>W images.

Animal models are important for understanding disease onset and progression, and an essential component of the drug development pipeline. Preclinical trials are used to assess therapeutic effectiveness in living systems. As discussed in §1, mice are apt models: they are fast to breed and develop, cheap to house, and share 99% of their genes with humans (*Rosenthal & Brown*, 2007). For ethical reasons, as well as to reduce costs, it is desirable to use as few animals as possible<sup>1</sup>. Just as in humans, imaging biomarkers must offer the power and repeatability to detect subtle effects over short periods (*Gutman et al.*, 2014).

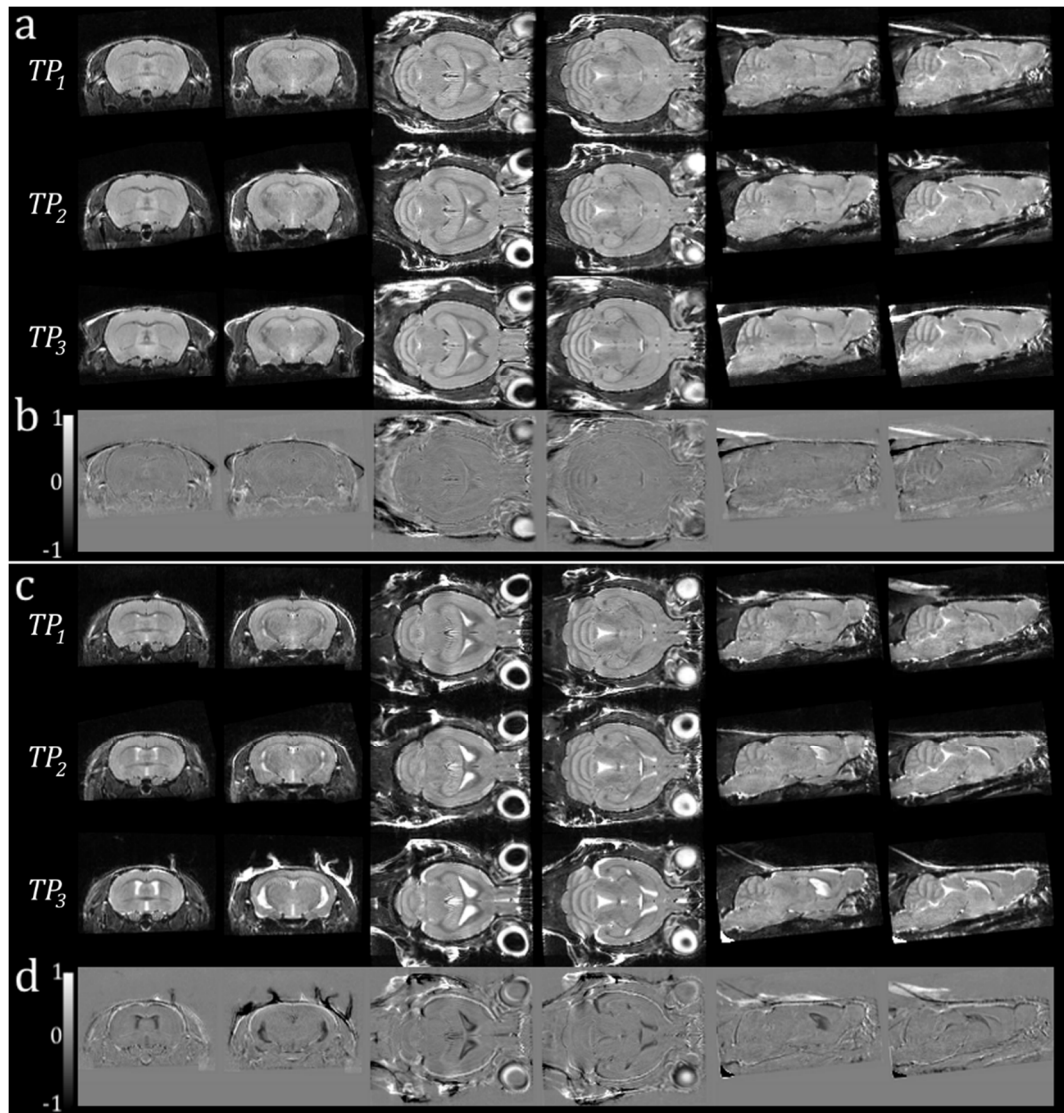
As I showed using morphometry in §5, the rTg4510 mouse exhibits profound structural atrophy of the entorhinal cortex, hippocampus and forebrain, and ventricular expansion, after 3 months of age – reflecting the progressive neurodegeneration seen in tauopathies such as human Alzheimer’s disease (AD, §2.3) and other dementias – over short timescales (c. one month) (*Fig 5.7, 6.2; Santacruz et al.*, 2005). These changes are interruptible via administration of doxycycline, a mutant-tau suppressor.

Existing drugs for AD seek to slow the onset of symptoms; there are as yet no cures or disease-modifying treatments (*Salomone et al.*, 2011). Early detection and intervention are vital to mitigate the effects of dementia, and accurate differential diagnosis, to distinguish diseases, is essential for treatments’ safe application (*Götz & Ittner*, 2008). Drug trials require clinically meaningful endpoints – measurable outcomes relevant to the disease and intervention – such as functional improvement, or survival. Atrophy is a surrogate biomarker which, being slowed or halted, may substitute (*Hampel et al.*, 2010).

The rTg4510 mouse is thus a good test-bed for the algorithmic development and measurement of the BSI in an animal model. If found to be a sensitive measure of the change in atrophy rate, the measurement could be applied in mouse trials of prospective therapies whose efficacy is unknown. Taking the brain BSI as an outcome measure, a change in atrophy rate is analogous to the drug’s disease-modifying effect.

---

1: See the NC3Rs’ “ARRIVE” guidelines: <https://www.nc3rs.org.uk>



**Figure 6.2** Progressive atrophy clearly visible in T<sub>2</sub>W rTg4510 mouse brain images. Equivalent coronal, transverse and sagittal slices from representative WT and UT brains at each time-point of the late intervention doxycycline study (C<sub>1</sub>), after rigid registration. Compare with [Fig 6.1](#) and [Freeborough & Fox \(1997\)](#) figure 1. (a) WT, showing relative brain preservation; (b) WT difference image  $TP_1 - TP_3$ ; (c) UT; (d) UT difference image  $TP_1 - TP_3$ . Scale units in difference images are arbitrary. In (d) atrophy is represented by both positive and negative differences: a bright rim around the cortex, and dark ventricles.

I describe below the algorithmic adaptations necessary for the BSI's application to T<sub>2</sub>W mouse brain images, including a modified “double window” approach, and assess ideal parameters. I based this upon the generalised BSI (gBSI, [Prados et al., 2014](#)), an automated pipeline, and show the necessary pre-processing steps. I measured the BSI



using *in vivo* structural MRI at three time-points in 87 rTg4510 mice from §5: a baseline (TP<sub>1</sub>) and two serial follow-up scans (TP<sub>2</sub>, TP<sub>3</sub>), and compared values between untreated (UT) animals, wild-type (WT) littermates, and those treated with doxycycline, and Epothilone D (EpoD), a microtubule stabilising agent (all introduced in §5.2).

In each group, I calculate the volume, percentage brain volume change (PBVC), and rate of atrophy between TPs. I evaluated my implementation via comparison with volume measurement from tissue segmentation; linear regression of atrophy rates; the transitive consistency over all TPs; and the ability of the BSI to distinguish groups and measure changes in atrophy rate – an assessment of the drugs’ effectiveness. I also calculate the sample sizes required in a future study for a drug to show a therapeutic effect. This is the first demonstration of the BSI’s viability in an animal study and the first objective longitudinal measure of atrophy in the rTg4510 model at short timescales (of around one month).

## 6.2 Methodology

### 6.2.1 Study design

This investigation was based upon the *in vivo* longitudinal rTg4510 mouse brain data, whose acquisition and study design is described fully in §5.2. Briefly, I applied the BSI to six groups of animals, each imaged serially *in vivo* at three scan time-points {TP<sub>1</sub>, TP<sub>2</sub>, TP<sub>3</sub>}, detailed in Table 5.1 and Fig 5.2. Groups were distinguished by genotype – either rTg4510 or wild-type (WT) littermates – and treatment regime. In untreated rTg4510s (UT), tauopathy was allowed to progress unhindered. Treatment of rTg4510 animals with doxycycline began ‘early’ (approx. 3 months of age, TR<sub>DOX-3M</sub>) or ‘late’ (4 months, TR<sub>DOX-4M</sub>), to elucidate the importance of intervention time. To distinguish the effects of the drug’s concentration, ‘low’ and ‘high’ doses of EpoD were administered (TR<sub>EpoD-LO</sub>; TR<sub>EpoD-HI</sub>), with the same treatment start time.



Up to three animals could practically be scanned per day. Consequently, inter-time-point periods differed between animals ([Fig 5.2](#)). I therefore report measurements of volume change standardised to 30 days (in humans, these are typically annualised). Mouse ages at each time-point are shown in [Table 5.1](#); periods between baseline and follow-up scans are shown in [Table 6.1](#).

<i>group</i>	<i>TP<sub>1-2</sub></i>			<i>TP<sub>2-3</sub></i>			<i>TP<sub>1-3</sub></i>		
	<i>N</i> pairs	mean interval	SD	<i>N</i> pairs	mean interval	SD	<i>N</i> pairs	mean interval	SD
<b>WT</b>	19	41.8	17.95	19	57.30	7.95	19	99.11	11.71
<b>UT</b>	30	57.2	27.64	30	68.90	18.31	30	126.07	42.71
<b>TR<sub>DOX-3M</sub></b>	9	56.4	2.35	10	51.00	0	9	107.44	2.35
<b>TR<sub>DOX-4M</sub></b>	6	27.8	6.11	7	71.86	6.28	6	101.00	7.13
<b>TR<sub>EpoD-LO</sub></b>	10	90.8	0.42	9	91.56	1.42	9	182.33	1.32
<b>TR<sub>EpoD-HI</sub></b>	10	90.7	0.95	8	91.75	0.46	8	182.75	0.46

**Table 6.1** Number of baseline and follow-up image pairs, and days between time-points, for the BSI investigation.

For each TP pair, the number (N) of complete image pairs included in this BSI investigation, and the mean and standard deviation (SD) scan interval, in days. For the number of animals scanned at each TP individually, see [Table 5.1](#).

I used the BSI to sensitively measure the rate of volume change between TP pairs  $\{TP_{1-2}, TP_{2-3}, TP_{1-3}\}$  under differing treatment regimes and at different stages of disease and treatment. For each TP pair, the earlier scan was denoted “baseline” and the later follow-up “repeat”. For the purpose of statistical tests, my null hypothesis  $H_0$  was the same atrophy measurement outcome for each group, although based upon earlier morphometry investigations in this model (§5; [Wells et al., 2015](#); [Holmes et al., 2016](#)), I expected WTs to display stable brain volumes (BV), or gradual, age-related, linear BV reduction over the study course, whereas UT animals should display severe atrophy. Similarly, one might expect doxycycline and EpoD treatments to improve brain volume preservation, and that early treatment with doxycycline and the higher dose of EpoD would more dramatically slow the rate of atrophy compared with UTs. I used two-tailed t-tests for statistics, to be agnostic as to the direction of any disease-modifying effect.

To evaluate results (§6.4), I measured the transitive error, regression intercept and ability of the BSI to differentiate groups. I correlated brain BSI with the difference between automatically segmented volumes, and calculated sample sizes required to measure 10, 25 and 50% reductions in the rate of atrophy. I compared volume loss in the UT, TR and WT groups and considered its rate as a percentage of total brain volume.

## **Animals**

All mice ( $N = 87$ ; numbers in [Tables 5.1](#) and [6.1](#)) were female rTg4510 or WT littermates, licensed, bred, imported and housed as previously described (§5.2.1; [SantaCruz et al., 2005](#); [Wells et al., 2015](#); [Holmes et al., 2016](#)). Animals were distinguished between scans using ear notches. Genotypes were re-confirmed following the study with PCR (polymerase chain reaction) and histology.

### **6.2.2 Data quality control, image selection and animal attrition**

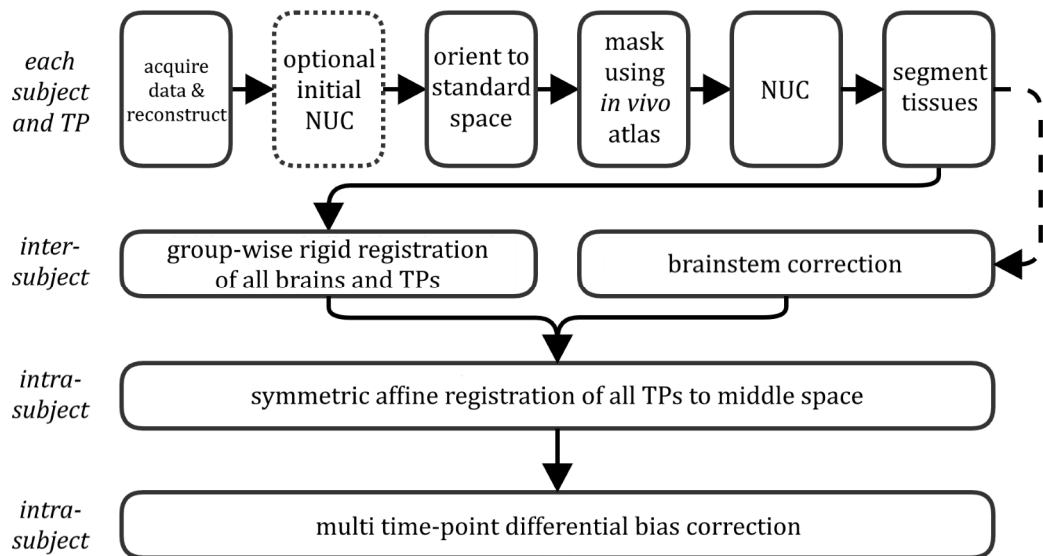
Fewer animals are reported here than in §5; this is due to attrition. In total, 255 scans were performed for cohorts C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>. To avoid bias from outliers and image artefacts, I quality controlled the data via visual inspection while blinded to the BSI results. 87 animals were imaged with more than one time-point. 81 completed all three ([Table 6.1](#)). Drop-outs were due to premature animal death ( $N=4$ ), poor image quality, assessed visually ( $N=8$ ), or indeterminate genotyping ( $N=1$ ). Poor image quality was due to large regions of unrecoverable signal loss at, or cropping of, the brainstem or cerebellum (due to FOV placement, or mouse motion during the scan, such as the head slipping out of the head holder); motion artefact; or severe noise arising from a loose cable.

I considered performing quality checks on the data using three more criteria: (1) SNR; (2) transitive error (see later); (3) volume discrepancy. Visual inspection could fairly reliably predict (3), which I measured using tissue segmentations at each TP. For (1) and (2), I calculated the Z-score  $\left(\frac{\text{measurement} - \text{mean}}{\text{standard deviation}}\right)$  to identify outliers, and chose an arbitrary cut-off. However, SNR (measured in the GM regions and BG) was reasonably consistent between images ([Table 5.2](#)) and a poor predictor of transitive error or signal drop-off. I did not use transitive error or volume discrepancy, as such post-hoc

measurements would artificially boost the power of the study, and images with consistent error or volume across all TPs (e.g. cerebellum cropped in all images) would not be excluded.

### 6.2.3 Image processing pipeline

I used the previously-described pipeline (§3, §5.3) to pre-process images, including image orientation; whole-brain masking based upon atlases (*Ma et al., 2014*); and non-uniformity correction (NUC) with N4. Brain tissue (GM, WM) and CSF were segmented using iterative expectation maximisation, beginning with prior tissue probability maps (TPMs) based upon the NUS atlas (*Bai et al., 2012*). I integrated aspects of the gBSI pipeline (*Prados et al., 2014*), including symmetric registration of all TPs to a middle space, and multi-time-point differential bias correction. I also replaced k-means tissue classification with the TPM-based expectation maximisation, changed intensity window parameters, and adjusted how different intensities were handled between those windows. The necessary pre-processing steps are outlined in *Fig 6.3*.



**Figure 6.3** Overview of pre-processing steps prior to BSI calculation

The segmented tissues were forward-propagated to the rigid GWR average space of all subjects and TPs. I then created a brainstem correction region, propagated this to each image, and used it to constrain segmentations. Segmented tissues were resampled into the intra-subject middle space, and used to construct the probabilistic XOR (pXOR) boundary

region (§6.2.4). “Intra-subject” here indicates between scans from the same animal. The GWR was used for brainstem correction only; images were resampled just once prior to BSI analysis, into the intra-subject middle space.

### ***Initial non-uniformity correction***

An accurate measurement of whole brain atrophy with the BSI requires the inclusion of all brain surfaces. Brain tissues must also occupy a similar intensity range, so the fixed intensity windows, calculated for each image pair, may be applied globally. Intensity non-uniformity correction (NUC) is thus vital. A smoothly-varying bias field arose from the use of both a high field strength magnet, and surface receiver coils placed over the cortex, and manifested predominately as signal drop-off caudally and ventrally (in particular at the brainstem). As discussed previously (§5.3.1), an initial application of NUC to the whole image benefits orientation correction. I used N4ITK, which provides robust NUC in high field-strength images (*Tustison et al., 2010*). Parameters were: 200 iterations; FWHM=0.15; subsampling factor 2; 4 levels; convergence threshold 0.001; 256 histogram bins.

### ***Probabilistic brain masks***

Brains were oriented to RAS space. Images from an external *in vivo* atlas database with accurate corresponding masks (UFL, *Ma et al., 2008*) were registered to each rTg4510 brain, using *NiftyReg*, with symmetric, inverse-consistent 12-DOF affine (*Modat et al. 2014*; 3 resolution levels and default parameters) followed by non-rigid registration (*Modat et al. 2012*; up to 150 iterations; bending energy penalty 0.01; control point spacing 5 voxels; 4 resolution levels; normalised mutual information used as the similarity measure).

I then performed multi-atlas label fusion using STEPS, a method shown to result in greater accuracy than single-atlas-based mask propagation (*Cardoso et al., 2013b; Ma et al., 2014; §2.2.3; §3.5.1*). During this procedure, registered atlas images were ranked in order of local similarity to the data (kernel standard deviation  $\sigma = 3$  voxels), and the  $X = 12$  most similar atlas images’ masks were propagated to the data space, following the

registration-derived deformations. A voting scheme then determined a particular voxel's inclusion in the mask. In cases where the UFL atlas failed to produce accurate whole-brain masks – such as by omitting the low-signal brainstem or cerebellum, or including extraneous material – (especially above the cortex, assessed visually), I substituted the *in vivo* rTg4510 database created by Da Ma and used a bootstrapping procedure to generate masks (as §5.3.2).

I binarised the resulting probabilistic masks, dilated them by 3 voxels, and used these to constrain a repeat NUC with N4. NUC with N3 is performed as part of the standard ADNI pipeline for human data<sup>2</sup>. *Prados et al. (2014)* showed that its re-application was beneficial in human brains. The example difference images in *Fig 5.4* reveal additional improvement in brain intensity homogeneity in mice, particularly at the brainstem.

The BSI requires accurate brain masks at each time-point, to initialise the intensity window selection and accurately identify surface regions of interest. The STEPS procedure provided probabilistic brain masks. In human images, this is sufficient. However, because the mouse brain is very tightly encased inside the skull, even at high resolutions, the separation between brain and external tissues (including the skull, meninges; fat on the skull surface; muscle tissue inferior to the brain and the olfactory epithelium) can be extremely small (less than one voxel – see *Fig 2.1*). Owing to their variability between groups, ventricle labels from the UFL brains were also sometimes inaccurate. Consequently, these masks required further improvement.

### ***Brain tissue, surface and ventricle classification***

In existing human BSI algorithms, brain masks are either dilated (the KN-BSI method) or are probabilistic (the gBSI method), and hence include some eCSF and vCSF, assumed to be dark. These low-signal voxels are included in subsequent k-means tissue classification. In human T<sub>1</sub>W images, the k-means scheme clusters CSF, GM and WM tissues. The mean intensity of each tissue underlying the resulting binary class masks is then used to standardise intensities between images (using linear regression) and calculate

---

<sup>2</sup>: The ADNI pre-processing pipeline is detailed at <http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing> (accessed November 2015).

intensity window limits (see below). This has been shown to reduce sample sizes required to detect a reduction in atrophy ([Leung et al., 2010b](#)) compared with the previous BSI implementation, which used only one value to normalise brain intensities (the mean intensity within the brain).

I found that the 3-class k-means algorithm performed poorly in the mouse brain, which, at minimum, exhibits four classes in T<sub>2</sub>W images (see [E6.5](#)): {BG, WM, GM, CSF}, with variable total volumes. The high degree of PV in mouse brains is also unsuitable for binary classification. Additionally, the tight-fitting skull and external material could be misclassified as brain tissue, without improved masks.

To mitigate these issues, I used TPMs propagated from the NUS atlas, as in [§3.6](#), to classify vCSF, GM and WM in the *in vivo* rTg4510 images with some spatial initialisation. I again tested a range of NRR parameters (bending energy and control point spacing in *NiftyReg*). To choose the optimum combination (bending energy penalty 0.001; control point spacing 4 voxels), I registered one NUS brain to all three TPs of a single UT rTg4510 and measured the resulting Pearson's  $r$  within a dilated brain mask.

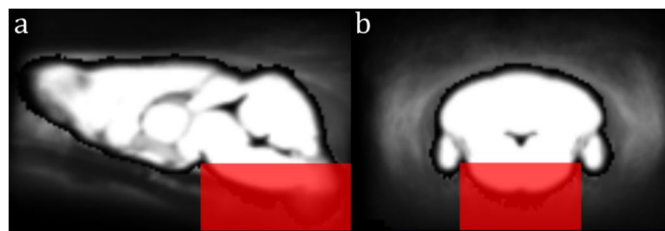
Ventricles, surface and external tissues required refinement. To classify external tissues, I binarised brain masks, dilated by 3 voxels, and subtracted the original mask, creating a 'rim region'. This constrained an intensity-based, prior-free expectation-maximisation (EM) classification using *NiftySeg*, with four classes, for {BG, GM, WM, eCSF}. External material (particularly adjacent to the cortex and caudal to the cerebellum) usually had similar intensity to brain tissue, and hence was misclassified as "GM" or "WM". To avoid classification of true brain tissue as external tissue, I weighted these classifications using a Euclidean distance transform up to three voxels from the brain surface, such that voxels adjacent to the existing binary brain mask were weighted 0.33; those two voxels from the brain were weighted 0.66; and voxels at least three voxels from the brain were weighted 1.0. I then used these weighted tissue classifications as priors in a second iteration of EM.

As the initial ventricular labels from the NUS atlas were generally over-generous, I binarised them and again used these masks to constrain a prior-free EM, with 3 classes, for {GM, WM, vCSF}.

Within the three-voxel-dilated masks, I used the results of the above EM schemes as TPMs (re-labelling external “GM” and “WM” as external material and combining eCSF and vCSF into a single class) to re-initialise EM using *NiftySeg*. I discarded the resulting external material and BG and combined the GM and WM segmentations to give a probabilistic brain mask ([Fig 6.6](#)), used for measuring whole brain volumes and atrophy rate. Upon visual inspection, the brain surface was considerably improved over the initial STEPS probabilistic masks, as were the ventricular regions.

### ***Mask brainstem consistency***

One of the greatest regions of variability in the brain masks was the brainstem, owing to the position of the mouse in the scanner, field-of-view selection, variable performance of the registration to atlas images, and signal drop-off (even accounting for NUC). This could potentially lead to substantial inconsistencies in the brain volume measurement (which is based upon binarised brain tissues), and hence mismeasurement of the atrophy rate (in terms of percentage brain volume change), as well as misrepresentations of the brain boundary in the brainstem region. In humans, a consistent cut-off point for BV segmentation is chosen; for example, in [Whitwell et al. \(2001\)](#) at “the lowest point of the cerebellum”.



**Figure 6.4** Binary brainstem region in average space.

Average of all probabilistic masks in the groupwise average space, with a painted binary brainstem region overlaid in red (a) sagittal; (b) coronal views.

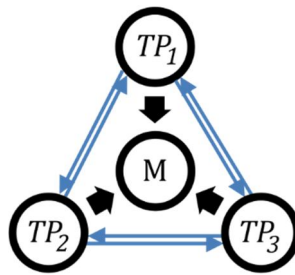
To improve accuracy, I defined a protocol to consistently crop the brainstem masks inferiorly and caudally. I performed a group-wise rigid registration of all brains to a randomly-chosen target member of the group, using *NiftyReg* (6 DOF: rotation and

translation only). Using the resulting transformations, I resampled the probabilistic masks to the target space using linear interpolation, and calculated their average. To reduce variability, I thresholded the average at 0.2, and binarised the result. I then manually painted a binary mask around the structural average image's brainstem using ITK-Snap ([Yushkevich et al., 2006](#)), ensuring the inferior brainstem surface was covered, from the cerebral peduncle to the cisterna magna ([Fig 6.4](#)).

I back-propagated the binarised, thresholded average probabilistic mask and the brainstem region mask to the space of the original images (using nearest neighbour interpolation). Then, within each brainstem region mask, I multiplied the original probabilistic masks by the binarised, thresholded average mask. This had the effect of constraining the probabilistic masks and substantially reducing variability.

### ***Symmetric, inverse-consistent registration***

The BSI measurement requires initial alignment of the baseline and repeat images. If the repeat is registered and resampled to the baseline (or vice versa), the resulting interpolation of only one image of the pair introduces directional bias ([Fox et al., 2011](#); [Leung et al., 2012](#)). Each animal's brain images were thus aligned to a “middle space” using symmetric, inverse-consistent registration of all 3 TPs, such that the resampling interpolation in all images was of the same order ([Fig 6.5](#)).



**Figure 6.5** Illustration of symmetric registration to a middle space, for three TPs.

Thin blue arrows represent inter-TP pair-wise 12-DOF symmetric affine registrations; black arrows represent equidistant resampling to the middle space (M), such that each TP image undergoes the same degree of interpolation.



I used *NiftyReg* with 12 DOF symmetric affine registration ([Modat et al., 2014](#); three resolution levels; normalised cross-correlation used as the similarity measure) to align each possible TP pair combination. After symmetric registration, the resulting affine transformation was inverted, to give the exact inverse. Thus the affine  $TP_2 \rightarrow TP_1$  inverted gave  $TP_1 \rightarrow TP_2$ ;  $TP_3 \rightarrow TP_1$  gave  $TP_1 \rightarrow TP_3$ ;  $TP_3 \rightarrow TP_2$  gave  $TP_2 \rightarrow TP_3$ .

The affine matrix  $A_i$  to resample each  $TP_i$  into the middle space is given by the log Euclidean average (similar to [E3.12](#), §3.3.2): the geometric mean of each pairwise affine transformation. For three TPs:

$$\begin{aligned} A_1 &= \expm\left(\frac{\logm(I) + \logm(TP_1 \rightarrow TP_2) + \logm(TP_1 \rightarrow TP_3)}{3}\right) \\ A_2 &= \expm\left(\frac{\logm(I) + \logm(TP_2 \rightarrow TP_1) + \logm(TP_2 \rightarrow TP_3)}{3}\right) \\ A_3 &= \expm\left(\frac{\logm(I) + \logm(TP_3 \rightarrow TP_2) + \logm(TP_3 \rightarrow TP_1)}{3}\right) \end{aligned} \tag{E6.1}$$

Where  $I$  is the 4x4 identity matrix which represents the affine transformation of an image to itself: the final pair,  $TP_i \rightarrow TP_i$ .

This technique has been shown to reduce bias in the BSI measurement ([Leung et al., 2012](#)). I resampled structural images using cubic interpolation and corresponding probabilistic masks with trilinear interpolation.

### **Differential bias correction**

In addition to the within-subject NUC already applied, after registration I also applied multi-time-point differential bias field correction (MTPDBC) between all three time-points, for each animal. This procedure, consistent with the KN-BSI and gBSI pipelines, helps to reduce local intensity variability between alike tissues, and has been shown to improve atrophy measurements in human data ([Lewis & Fox, 2004](#); [Leung et al., 2012](#)).

Each image  $v_i$  (where  $i$  is one of the TPs) is log-transformed. The difference image for each possible pair is found. A median filter is then applied at each voxel, over a given kernel size, to remove remaining small-scale structures such as registration errors, true

atrophy, and noise. This is inverse log-transformed (E6.3). The results are multiplied (voxel-wise), and the  $N^{\text{th}}$ -root found, where  $N$  is the number of TPs, to give the geometric mean. As the bias is considered to be multiplicative, each image is divided by the result. For the three TPs here, the corrected images are therefore:

$$\begin{aligned} v'_1 &= v_1 / \sqrt[3]{r_{11} \times r_{12} \times r_{13}} \\ v'_2 &= v_2 / \sqrt[3]{r_{21} \times r_{22} \times r_{23}} \\ v'_3 &= v_3 / \sqrt[3]{r_{31} \times r_{32} \times r_{33}} \end{aligned} \quad \text{E6.2}$$

Where the intensity inhomogeneity between a pair is for example:

$$r_{12} = \exp(\text{median filter}(\log(v_1)) - \log(v_2)) \quad \text{E6.3}$$

For humans, the median filter kernel has default radius 5 voxels (around 5-6mm)<sup>3</sup>. Thus the travelling kernel ‘cube’ has sides of 11 voxels. Structures less than half of this side-length are removed (Paranjape, 2009) and not considered part of the bias field. In these 150 $\mu\text{m}^3$  resolution mouse brain images, this kernel (equivalent 0.75mm) was of a similar order of scale as, or smaller than, fine anatomical variation between brains, including structures such as the corpus callosum and ventricles, and I found (via visual inspection) that it was thus too small. If a late-TP image had large ventricles, surrounding tissues in earlier TPs were artificially brightened. I therefore chose a broader kernel radius (10 voxels). This allowed removal of structural artefacts up to 10 voxels in size from the difference image. If atrophy, or misregistration, were expected to translate equivalent tissue boundaries greater than 10 voxels, a larger radius should be chosen.

## 6.2.4 Probabilistic XOR boundary shift region

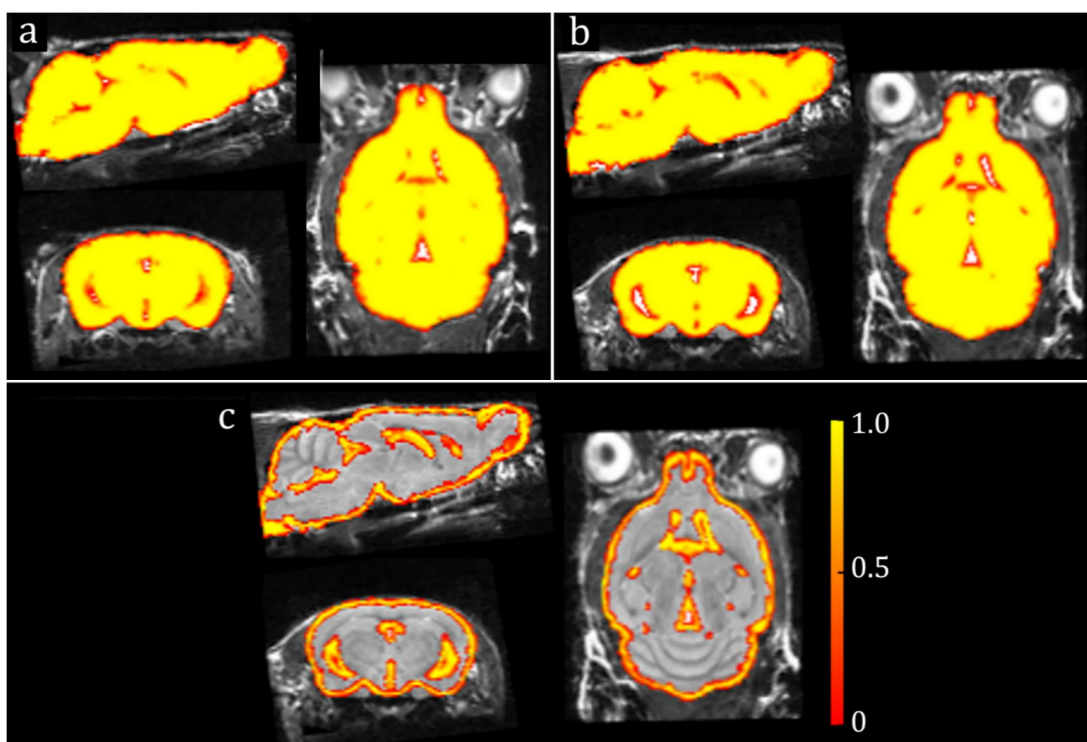
The boundary shift region confines the calculation of intensity differences between images to a ROI. This spatial window, adapted to each TP pair individually, is vital to

---

3: In Lewis & Fox (2004), resolution was approx.  $1 \times 1 \times 1.5\text{mm}$ ; in Leung et al. (2012) it was  $1.25 \times 1.25 \times 1.2\text{mm}$ .

exclude highly variable external tissues (revealed in the difference images of [Fig 6.1](#) and [6.2](#)) and include variable tissue interfaces, such as the ventricles.

For binary masks  $A$  and  $B$  at two TPs,  $A \cup B$  is the union, the set of voxels which are in  $A$ , or  $B$ , or both.  $A \cap B$  is the intersection, the set containing all those elements that  $A$  and  $B$  have in common.  $A \text{ XOR } B$  ( $A \oplus B$ ) is an exclusive OR, true when either  $A$  or  $B$ , but not both, are true. In earlier BSI implementations, the boundary region was identified using a binary XOR: the dilated union minus the eroded intersection of  $A$  and  $B$ .



**Figure 6.6** Probabilistic masks, and pXOR region.

Equivalent sagittal, coronal and transverse slices of a UT brain, showing probabilistic brain segmentations at  $TP_1$  (a) and  $TP_2$  (b), and the resulting pXOR region (c). The lateral, third and fourth ventricle boundaries are all included.

In gBSI, a probabilistic XOR (pXOR) region, generated using the STEPS masks, is used instead ([Fig 6.6c](#)). This beneficially weights the BSI calculation to the tissue boundaries: the inclusion of non-brain tissues and noise during binary erode/dilate operations may reduce sensitivity ([Ledig et al., 2012](#); [Prados et al., 2014](#)). This is especially relevant in the

mouse brain, wherein external tissues (dura; skull) are extremely close-fitting, compared with the human cranium.

The pXOR region is calculated from two probabilistic brain masks, resampled into the middle space, as follows:

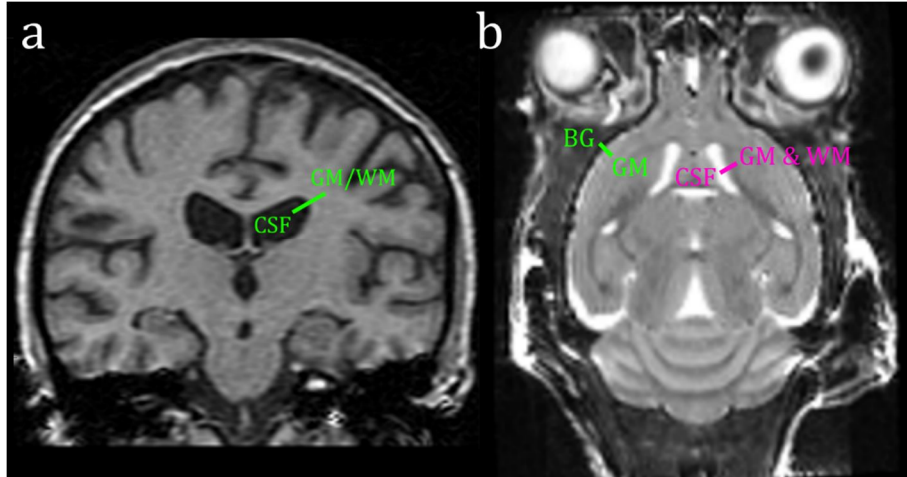
$$\text{E6.4} \quad pXOR(A, B) = (A \times \bar{B}) + (\bar{A} \times B) - ((A \times \bar{B}) \times (\bar{A} \times B))$$

where  $\bar{A}$  is the complement of  $A$ . It is greater when the segmentations disagree. Each voxel  $v$  (at 3D coordinate  $x, y, z$ ) in this region is then weighted by  $\kappa$ , the mean of all non-zero voxels in the pXOR region, such that:

$$\text{E6.5} \quad R_{pXOR} = \begin{cases} \frac{pXOR(v)}{\kappa} & \text{if } pXOR(v) < \kappa \\ 1 & \text{otherwise} \end{cases}$$

As  $\kappa < 1$ , this has the effect of broadening the ROI ([Prados et al., 2014](#)). An example of the resulting region ( $R_{pXOR}$ ) for a pair of TPs in a UT mouse is shown in [Fig 6.6c](#).

### 6.2.5 Double-window approach for T<sub>2</sub>W mouse brains



**Figure 6.7** Comparing T<sub>1</sub>W human and T<sub>2</sub>W mouse brain images

Human T<sub>1</sub>W MRI (a, coronal view) has one possible profile through either the ventricle or external cortical boundary region ([Fig 6.4a](#)), from CSF (low intensity) → GM/WM (higher). The mouse brain (b, transverse view) typically has little external CSF, as the skull more tightly encases the brain. In T<sub>2</sub>W MRI, this dark “background” region produces a

similar profile (green line; lower part of [Fig 6.4b](#)). The CSF (high intensity) → GM/WM (lower) boundary produces a different profile (pink line; upper part of [Fig 6.4b](#)). In late-stage rTg4510 mice, eCSF pools in the cisternae, above the cortex, and in the entorhinal region, so an equivalently aligned voxel may also transition between TPs from BG → CSF directly.

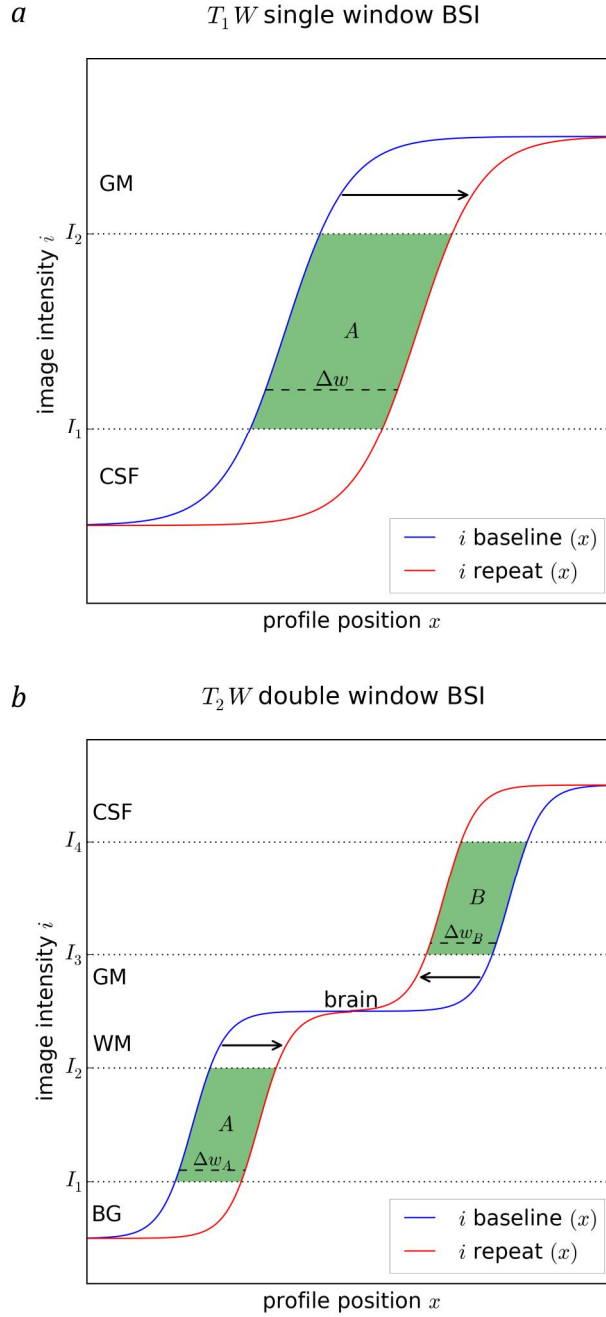
The BSI has so far been applied to, and designed exclusively for, T<sub>1</sub>W images, where CSF, air, skull, or background (BG) signal appear dark (low intensity), whereas GM and WM tissues have higher intensities. Tissue/CSF and tissue/BG interfaces thus share similar intensity gradients. The rTg4510 cohort was acquired with a T<sub>2</sub>W FSE sequence, in whose images {BG, skull, air} all appear dark, but CSF has higher intensity than tissue ([Fig 6.7](#)). The intensity scales are:

E6.6

$$\begin{aligned} T_1: \{CSF, BG, skull, air\} &< GM < WM \\ T_2: \{BG, skull, air\} &< WM < GM < CSF \end{aligned}$$

In most BSI implementations, only one intensity window is required to capture brain/non-brain boundaries ([Fig 6.8a](#)). [Hobbs et al. \(2009\)](#) showed how a double-window (DW) approach may be used to capture atrophy of the caudate putamen. [Leung et al. \(2010a\)](#) extended this to the hippocampus. Both are predominately GM structures bounded by CSF and WM.

I adapted the DW technique for the whole brain in T<sub>2</sub>W images ([Fig 6.8b](#)). I combined GM and WM tissue classes (see [Brain tissue, surface and ventricle classification](#) in §6.2.3 above) to give a single ‘brain’ class. The DW approach accommodates the BG-brain and brain-CSF intensity gradients, both of which may display the effects of atrophy in rTg4510s. [Figs 5.7, 6.2, 6.7](#) show cortical atrophy to both BG and eCSF (especially in the entorhinal region) and ventricular atrophy to vCSF.



**Figure 6.8** Idealised 1D boundary intensity profiles of (a)  $T_1W$  single window BSI and (b)  $T_2W$  double window BSI.

Sigmoidal representations of intensities along profiles drawn through brain boundaries in  $T_1W$  and  $T_2W$  images (as in Fig 6.7), at baseline and repeat time-points, in intensity-normalised scan pairs. The single window is appropriate for whole-brain BSI in  $T_1$ . In  $T_2$ , the double window BSI is necessary to capture the two boundary regions,  $BG \rightarrow$  brain and brain  $\rightarrow$  CSF. The black arrows indicate the direction of atrophy, on both fronts. The shaded areas A and B can provide estimates of the boundary shifts  $\Delta w_A$ ,  $\Delta w_B$ . These are modulated by both the actual intensity change and the probabilistic XOR region (see

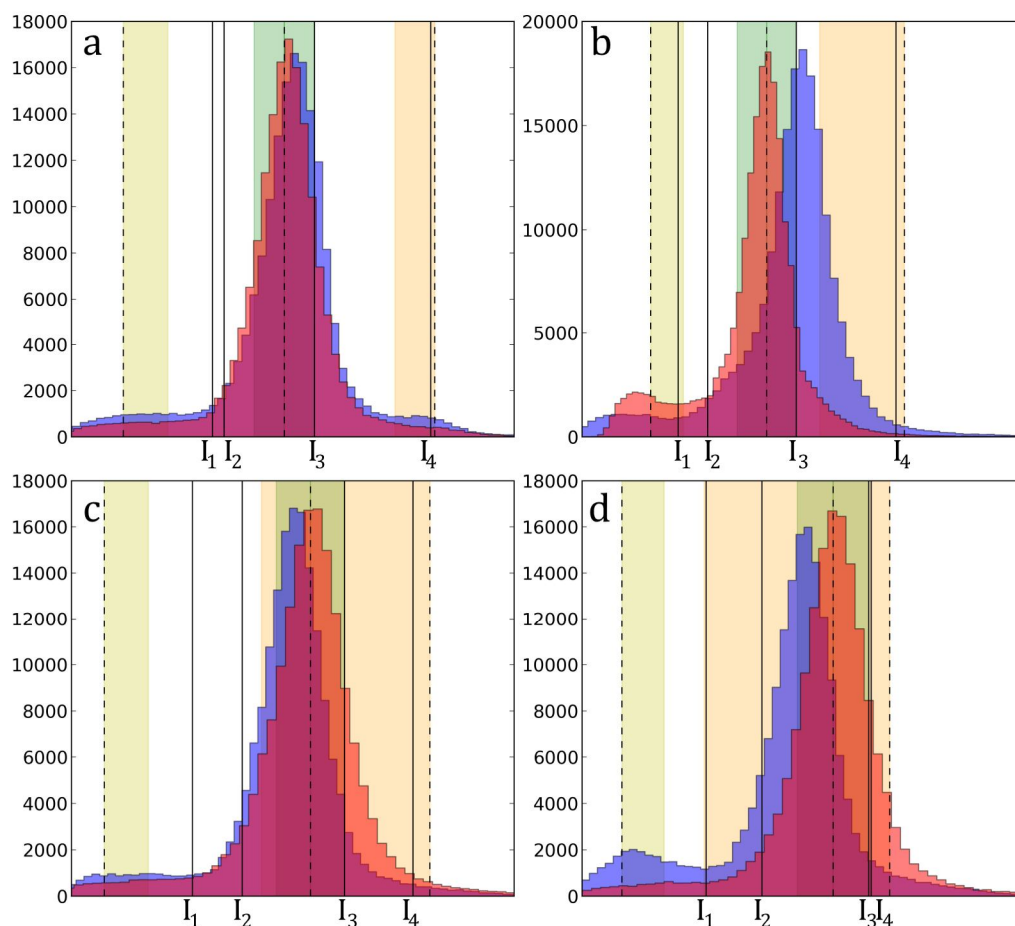
§6.2.7). Predominant tissue classes for each intensity range are shown at left. Note the brain consists of both GM and WM, so to accommodate intra-brain tissue intensity changes, the region between windows  $[I_2, I_3]$  should not be too narrow. Compare with Fig 2 of *Freeborough & Fox (1997)*, for the single-window T<sub>1</sub>W case, and Fig 2 of *Hobbs et al. (2009)* for the double-window T<sub>1</sub>W case in the human caudate.

## 6.2.6 Intensity standardisation and window selection

The intensities of each TP pair were standardised so that the boundary profiles of the baseline and repeat image pass through the same universally-applicable intensity window. In *Fig 6.8*, standardisation should result in the tissue means of each TP occupying approximately equal intensities.

I binarised the probabilistic tissue classifications (at 0.5) and found the mean and standard deviation (SD) of each class within these masks, as well as the mean intensity within the intersection region of the two binarised brain masks. I then used the coefficients of a linear regression of these intensities between TP pairs to normalise images' intensities (as per *Leung et al., 2010b; Prados et al., 2014*). Both images were then divided by the baseline CSF mean, to constrain the intensity range of most tissues between  $[0\ 1]$ , which helps for visualisation.

The DW BSI aims to capture tissue intensity changes between two windows (*Fig 6.8b*). To increase sensitivity, these windows should be as wide as possible. However, to increase specificity to intensity changes between relevant tissue types, intensity changes between the same or alike tissues (e.g. within the brain, between GM and WM) should be ignored. Thus in this figure, the range  $[I_3\ I_4]$  should be sufficiently broad. To achieve this, the SD of each tissue class was employed. This adaptive window selection method improves robustness over fixed intensity window limits (*Leung et al., 2010b*). *Fig 6.9* illustrates some image pairs' histograms and SDs, after intensity standardisation.



**Figure 6.9** Intensity windows overlaid upon baseline and repeat image histograms.

Example baseline (blue) and repeat (red-orange) histograms within a dilated mask, for four subjects' TP pairs. Dashed lines indicate intensity means of baseline image tissues, from left: BG; brain; CSF. Shaded regions cover respective standard deviations of each class, from left in (a), yellow-green: 1 BG SD above the BG mean; green: 1 SD  $\pm$  brain mean; yellow-orange: 1 SD below the CSF mean. Solid lines (and x-axis labels) indicate the lower and upper window boundaries:  $I_1$ ,  $I_2$ ,  $I_3$ ,  $I_4$ . (a) exhibits similar, overlapping histograms with aligned brain intensity peaks and a narrow CSF SD. However, in (a), the upper window is wide, decreasing sensitivity to small changes within the window.

As [Fig 6.9](#) shows, this technique had some limitations for these mouse brain images. Owing to variable intensity ranges, the linear regression was not always successful at aligning alike classes' means. The SD of tissue classes, especially CSF, was also highly variable. This may be due to poor initial classification performance, over-generous probabilistic masks, or the inclusion of (low intensity) blood vessels misclassified as CSF.



I chose the following window limits, with reference to [Fig 6.8](#):

$$\begin{aligned}
 I_1 &= BG_{mean} + 2 \times BG_{SD} \\
 I_2 &= brain_{mean} - 2 \times brain_{SD} \\
 I_3 &= brain_{mean} + 1 \times brain_{SD} \\
 I_4 &= CSF_{mean} - 0.1 \times CSF_{SD}
 \end{aligned}$$

To prevent negative window widths, if  $I_1 > I_2$ , I automatically set  $I_1 = I_2 - 1 \times brain_{SD}$ . Similarly, if  $I_4 < I_3$ ,  $I_4 = I_3 + 1 \times brain_{SD}$ .

The coefficients in [E6.7](#) were chosen following parameter tests about the DW BSI defaults, respectively {1, 1, 0.5, 0.5}. In mouse brains, these gave an over-wide lower window and, due to the large  $CSF_{SD}$ , a narrow upper window. As the tissues have different SDs, it is advantageous to vary each factor independently (the prior code used a single coefficient for each window). The  $BG_{SD}$  was multiplied by a factor of 2 in  $I_1$  to narrow the window, thereby increasing sensitivity to small changes (see later). The mean brain intensity was weighted in favour of GM (as there are more GM voxels than WM) – therefore, to allow some variation in WM intensity, the  $brain_{SD}$  in  $I_2$  was multiplied by 2, whereas in  $I_3$  it was multiplied by 1. In  $I_4$ , the factor 0.1 controlled the variability and breadth of the  $CSF_{SD}$  ([Fig 6.9](#)), sometimes at the expense of increasing the window width ([Fig 6.9a](#)).

## 6.2.7 Boundary Shift Integral calculation

For each animal, TP pairs were {TP<sub>1</sub>, TP<sub>2</sub>; TP<sub>2</sub>, TP<sub>3</sub>; TP<sub>1</sub>, TP<sub>3</sub>}. For each “baseline” and registered “repeat” scan pair ( $A, B$ ), the BSI was calculated between two intensity windows {[ $I_1, I_2$ ], [ $I_3, I_4$ ]} chosen (above) to incorporate the brain boundary. For the single window, T<sub>1</sub>W case ([Freeborough & Fox, 1997](#); [Prados et al., 2014](#)):

$$BSI = D \times \sum_{voxels \in V} R_{pXOR}(v) \times \frac{clip(I_A, I_{max}, I_{min}) - clip(I_B, I_{max}, I_{min})}{I_{max} - I_{min}}$$

where  $D$  is a voxel's volume;  $R_{pXOR}(v)$  is the weighting factor given by the pXOR region at the current voxel;  $I_A$  and  $I_B$  refer to individual voxel intensities;  $I_{max}$  and  $I_{min}$  are the

upper and lower window limits;  $V$  is the voxel space occupied by both registered scans, and clip is a function to include only intensities within the window, such that:

$$\text{clip}(a, I_{\max}, I_{\min}) = \begin{cases} I_{\max} & a > I_{\max} \\ a & \text{if } I_{\min} \leq a \leq I_{\max} \\ I_{\min} & a < I_{\min} \end{cases}$$

The effect of [E6.8](#) is to render intensity differences between voxels at baseline and repeat as a proportion of window width, before weighting by the pXOR. As tissue intensities are greater than BG or CSF in  $T_1$ , the BSI is given as the positive volume loss (atrophy).

For a particular voxel shared between  $(A, B)$ , the existing DW implementation calculated the BSI from [E6.8](#) using only one of the two intensity windows,  $\{[I_1, I_2], [I_3, I_4]\}$ :

$$\text{in E6.8, } \begin{cases} I_{\min} = I_3, I_{\max} = I_4 & \text{if } I_A > I_2 \text{ and } I_B > I_2 \\ I_{\min} = I_1, I_{\max} = I_2 & \text{otherwise} \end{cases}$$

Additionally, if  $I_A > I_2$  and  $I_B > I_2$ , the result was multiplied by -1, so atrophy involving the upper window was still given by a positive value. However, this does not accurately accommodate all possible  $T_2W$  image scenarios, as shown below.

For the  $T_2W$ , DW case I adjusted this to consider both windows for each voxel. This allows for inter-window PV and direct BG/CSF transitions:

$$\begin{aligned} \text{E6.9} \quad BSI \text{ (ml)} &= D \times \sum_{\text{voxels} \in V} R_{\text{pXOR}}(v) \\ &\times \left( \frac{\text{clip}(I_A, I_2, I_1) - \text{clip}(I_B, I_2, I_1)}{I_2 - I_1} \right. \\ &\quad \left. - \frac{\text{clip}(I_A, I_4, I_3) - \text{clip}(I_B, I_4, I_3)}{I_4 - I_3} \right) \end{aligned}$$

Here,  $D = (150\mu\text{m})^3 = 3.375 \times 10^{-6}\text{ml}$ . The first clipping term is as per [E6.8](#) and applies to the lower window, handling brain/BG boundaries. The second handles brain/CSF boundaries, but is subtracted (as CSF intensities are greater than brain tissues') so that a positive value still represents volume loss. To give some explicit examples of possible inter-TP voxel intensity transitions using [E6.9](#):

1. If  $I_A$  is CSF and  $I_B$  is BG,  $I_A > I_4 > I_1 > I_B$ , thus the first clipping term is +1, the second is +1 and the net BSI is  $D \times R_{pXOR}(v) \times (1 - 1) = 0$ .
2. If  $I_A$  is brain and  $I_B$  is CSF,  $I_B > I_4 > I_3 > I_A > I_2$ , thus the first clipping term is 0, the second is -1 and the net BSI is  $D \times R_{pXOR}(v) \times 1$ .
3. If  $I_A$  is PV brain/CSF and  $I_B$  is CSF,  $I_B > I_4 > I_A > I_3$ . Here, the first clipping term is determined by the distance of  $I_A$  from the window limits, and the second is 0. For example, if  $I_3 = 0.8$ ,  $I_4 = 1.0$ ,  $I_A = 0.95$ ,  $BSI = D \times R_{pXOR}(v) \times \frac{0.05}{0.2}$ .
4. If  $I_A$  is PV brain/BG and  $I_B$  is PV brain/CSF,  $I_4 > I_B > I_3 > I_2 > I_A > I_1$ . Each clipping term is determined by the relative position of  $I_A$  and  $I_B$  from respective window limits. For example, if  $I_1 = 0.1$ ,  $I_2 = 0.3$ ,  $I_3 = 0.8$ ,  $I_4 = 1.0$ ,  $I_A = 0.2$ ,  $I_B = 0.85$ ,  $BSI = D \times R_{pXOR} \times \left( \left( -\frac{0.1}{0.2} \right) - \left( -\frac{0.05}{0.2} \right) \right)$ . The BSI is negative, representing growth.

In examples (1) and (2), the intensities  $I_A$ ,  $I_B$  transition entire window widths; in (3) and (4) the windows are only partially crossed. For the same image pair, narrower windows increase the chance of whole-window transitions. This increased sensitivity is balanced by the corresponding increased chance of intra-class intensity changes which do not cross either window.

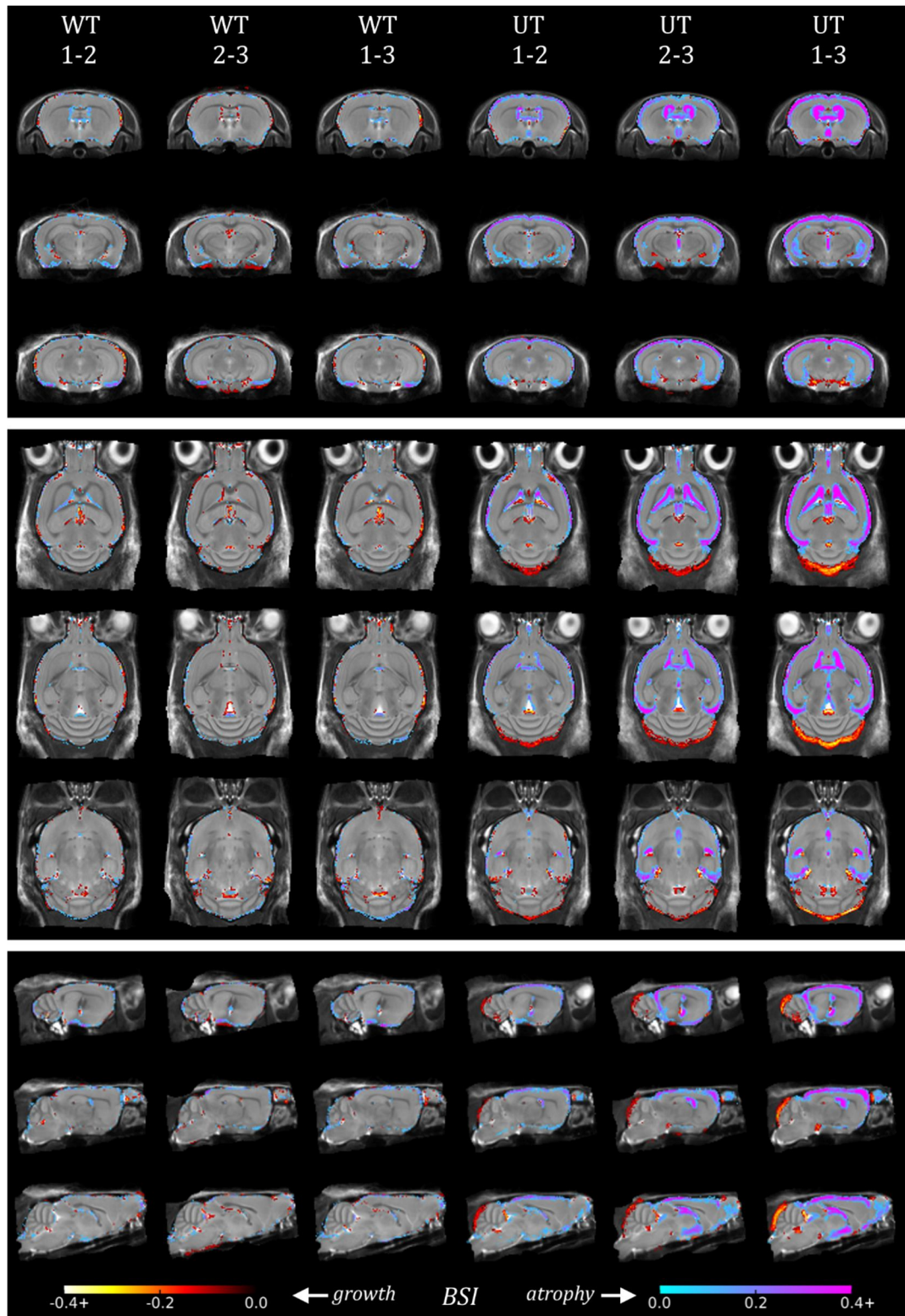
To illustrate the improvement offered by [E6.9](#), consider examples (1) and (4). In the earlier DW implementation, only the lower window would be considered in each case. In (1), the clipping term would be +1 and the BSI is incorrectly positive (atrophy). In (4), the clipping term becomes  $(-1/2)$ : the tissue increase is overestimated.

To increase robustness, because the linear regression is asymmetric ([Leung et al., 2010b](#)), I calculated the “forward” and “backward” BSI for each TP pair (i.e. baseline and repeat images swapped; linear regression and window limits re-calculated), and took the mean.

## 6.3 Results

[Fig 6.10](#) illustrates the primary regions of atrophy in the two extreme cases, WT and UT rTg4510 groups, by showing the voxel-wise average brain BSI results for each TP pair (achieved using GWRs of each group and resampling BSI results into the final average space). The ventricles, entorhinal cortex and cortical surface are all well-delineated by tissue volume loss in the UT group, and these clearly differentiate the UTs from WTs. The WT group displays some scattered regions of growth and minor atrophy, but this appears stable across TPs. This is likely due to the sensitivity of the intensity windows chosen ([§6.2.6](#)); however, image artefacts and imprecise registration may also contribute.

Apparent expansion of the cerebellum is particularly noticeable in the UT group. To investigate this unexpected feature, my colleague Da Ma used the multi-atlas parcellation protocol described earlier to measure WT and UT cerebellums in  $C_1$  and  $C_2$  ([Table 6.2](#)). The group means were not significantly different at any TP, but overall mean cerebellar volumes increased in both cases (also non-significantly,  $p > 0.05$ ). Despite this, the mean BSI image highlights much greater regions of apparent expansion in the UT group. This was detected by cross-sectional TBM ([Fig 5.10](#)). [Fig 6.15](#) shows that overall UT brain volumes decreased over time, whereas WT volumes remained relatively stable.



**Figure 6.10** WT and UT brain BSI result visualisations

Voxel-wise average brain BSI results overlaid on  $TP_1$  and  $TP_2$  structural averages, after GWR of the respective WT and UT images. WT (columns 1—3) and UT (columns 4—6) groups, with TP pairs indicated. Three equivalent coronal, transverse and sagittal views of each. The scale is mean BSI (voxel proportion), hence atrophy (volume loss) is positive.

							% change in mean		
	<i>TP</i> <sub>1</sub>	SD	<i>TP</i> <sub>2</sub>	SD	<i>TP</i> <sub>3</sub>	SD	<i>TP</i> <sub>1-2</sub>	<i>TP</i> <sub>2-3</sub>	<i>TP</i> <sub>1-3</sub>
<b>WT</b>	93.32	5.81	93.11	5.75	96.06	4.04	-0.24	3.18	2.93
<b>UT</b>	91.33	4.83	92.26	4.52	93.92	3.77	1.03	1.79	2.84

**Table 6.2** WT and UT cerebellar volumes for *C*<sub>1</sub>, *C*<sub>2</sub>.

Mean and standard deviation (SD) cerebellar volumes (mm<sup>3</sup>), from automatic parcellation with STEPS, and inter-TP % volume changes of the means (not standardised for age or to inter-TP period). Volumes were not significantly different (*p*>0.1 at all TPs).

I used the BSI to assess rates of whole brain volume change in the WT, treated (doxycycline and EpoD) and UT rTg4510 groups. The BSI measurements provide the absolute volume loss (in ml) over time (standardised, here, to 30 days, as the inter-TP periods varied), while the percentage brain volume change (PBVC) allows direct comparison between animals with different initial BVs.

[Tables 6.3](#), [6.4](#), [6.5](#) show the mean and SD measures of BSI, segmentation and PBVC, for each TP pair, as well as overall means. The PBVC was calculated, for a TP pair *TP*<sub>*i-j*</sub>:

$$\text{E6.10} \quad PBVC_{i-j} (\%) = -100 \times \frac{BSI_{i-j}}{TP_0 \text{ volume}}$$

where *TP*<sub>0</sub> *volume* is the segmentation-derived brain volume (probabilistic mask thresholded at 50%) at the earliest available TP. The segmentation volume change is the difference between thresholded, segmented volumes between each TP.

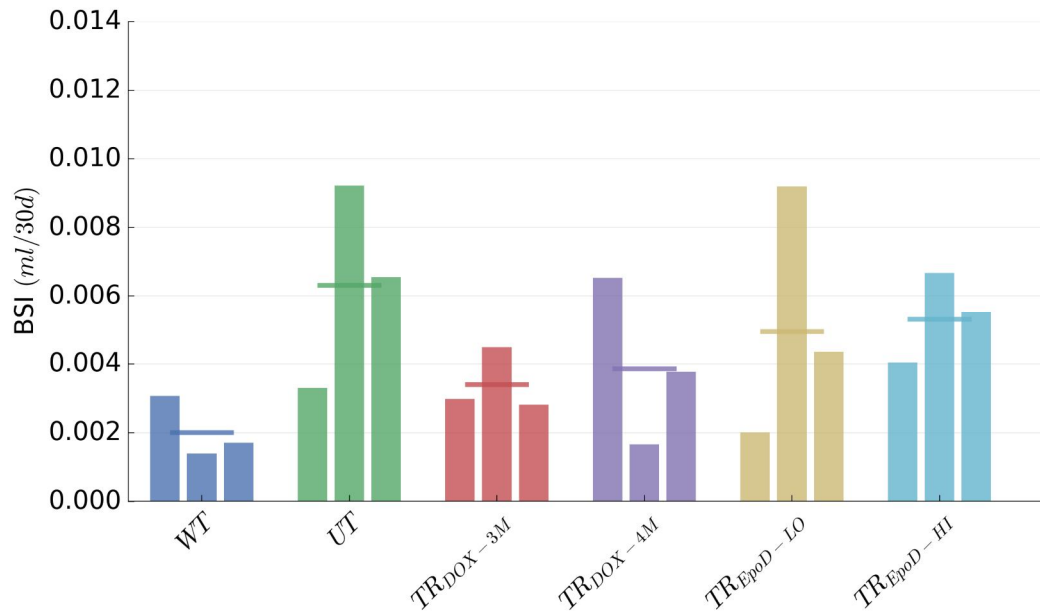
The standardised BSI measurements illustrated in [Fig 6.11](#) suggest the rate is different between TPs. To elucidate whether this was a factor of mouse age, I performed a linear regression between the standardised BSI (ml/30d) and age at the mid-TP point, for *TP*<sub>1-2</sub> and *TP*<sub>2-3</sub>. Only the UT and TR<sub>EPOD-LO</sub> groups had slopes significantly different from zero (*p*<0.05), suggesting atrophy accelerated slightly<sup>4</sup> with mouse age during the period studied, in rTg4510s, but not when treated with doxycycline or the high dose of EpoD.

<sup>4</sup>: For both:  $y(\text{ml}/30d) \approx 7 \times 10^{-5}x - 5.6 \times 10^{-3}$ , where *x* is in days. Approximately 0.69 voxels/day, per day of age, -1659 voxels/day (intercept).

Both the WT and TR<sub>DOX-4M</sub> groups exhibited greater rates of volume loss between the earlier TPs – however, these groups had by far the greatest SDs in all measures. This may be an artefact of measurement error, or the very short TP<sub>1-2</sub> period for some mice in C<sub>1</sub> (minimum 10 days, [Fig 5.2](#)). Discounting error, the TP<sub>1-3</sub> measurement should be the average of the other two; this was true in most cases.

The variance of all measures was relatively large, compared to respective means, despite my exclusion of 8 poor-quality images. Evaluation ([§6.4](#)) showed that, overall, BSI measurements were reasonable across groups and TPs. The measurements' means (TP<sub>1-2</sub>, TP<sub>2-3</sub>, TP<sub>1-3</sub>) provide an estimate of overall BSI and PBVC ([Fig 6.13](#)). These values are broadly as expected for each group (see [§6.2.1](#)). The differences in segmented volumes – from the same masks used to generate the pXOR region – are shown for comparison ([Fig 6.12](#), values in [Table 6.4](#)). These follow a similar pattern to the BSI, but the recovered volumes are larger – notably, the WT group between TP<sub>1-2</sub>, which may suggest the segmentation was more prone to measurement bias. This is explored further in [§6.4.1](#).

As brains had different initial volumes ([Fig 6.15](#)), the PBVC provides a useful comparison of proportional volume loss between groups. Early treatment with doxycycline resulted in the lowest rate of proportional volume loss, followed by late doxycycline treatment, the low dose of EpoD, and the high dose of EpoD (although the actual rate of volume loss was slightly greater in the TR<sub>EpoD-HI</sub> group).



**Figure 6.11** Mean absolute brain volume loss (ml), from the BSI, per 30 days.

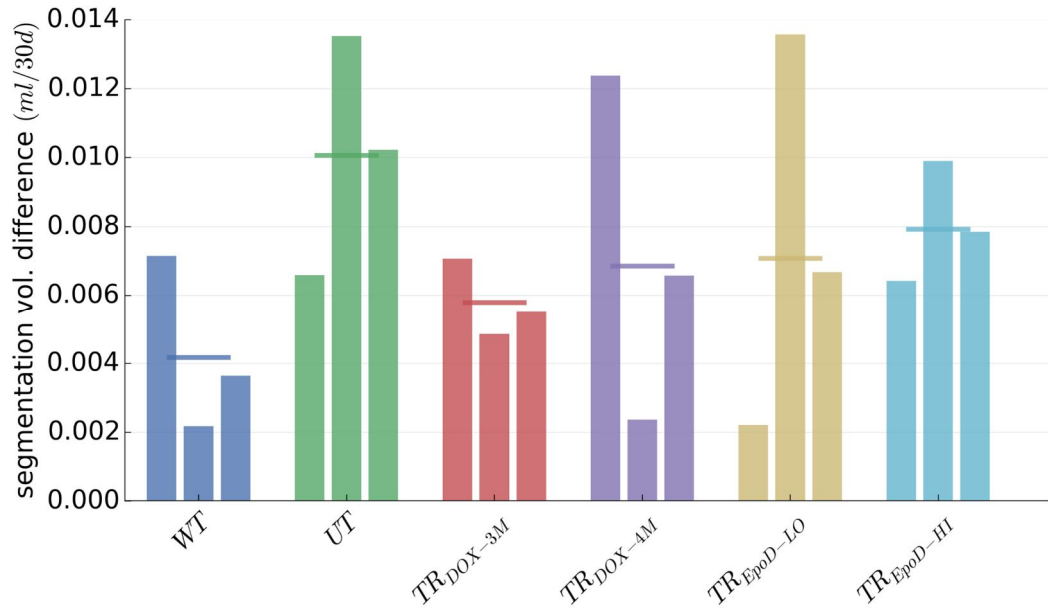
BSI measurements from each TP pair (from left for each group: TP<sub>1-2</sub>, TP<sub>2-3</sub>, TP<sub>1-3</sub>). Values are shown in [Table 6.3](#). Horizontal lines represent overall mean (all TPs). In this plot and for [Figs 6.12](#), [6.13](#), SDs are shown in accompanying tables, as they were comparatively large.

group	TP <sub>1-2</sub> mean	SD	TP <sub>2-3</sub> mean	SD	TP <sub>1-3</sub> mean	SD	overall mean	SD
WT	0.0031	0.0106	0.0014	0.0064	0.0017	0.0025	0.0020	0.0070
UT	0.0033	0.0058	0.0092	0.0036	0.0065	0.0023	0.0063	0.0047
TR <sub>DOX-3M</sub>	0.0030	0.0061	0.0045	0.0021	0.0028	0.0037	0.0034	0.0040
TR <sub>DOX-4M</sub>	0.0065	0.0106	0.0017	0.0050	0.0038	0.0054	0.0039	0.0070
TR <sub>EPOD-LO</sub>	0.0020	0.0029	0.0092	0.0016	0.0043	0.0012	0.0049	0.0036
TR <sub>EPOD-HI</sub>	0.0040	0.0032	0.0067	0.0034	0.0055	0.0030	0.0053	0.0032

**Table 6.3** Standardised BSI results (ml/30 days).

Mean and standard deviation BSI measurements (ml), all standardised to 30d. The overall mean for each group includes all three TP pairs. Values greater than  $\pm 2$  SD removed in this table and in [Tables 6.4](#), [6.5](#).





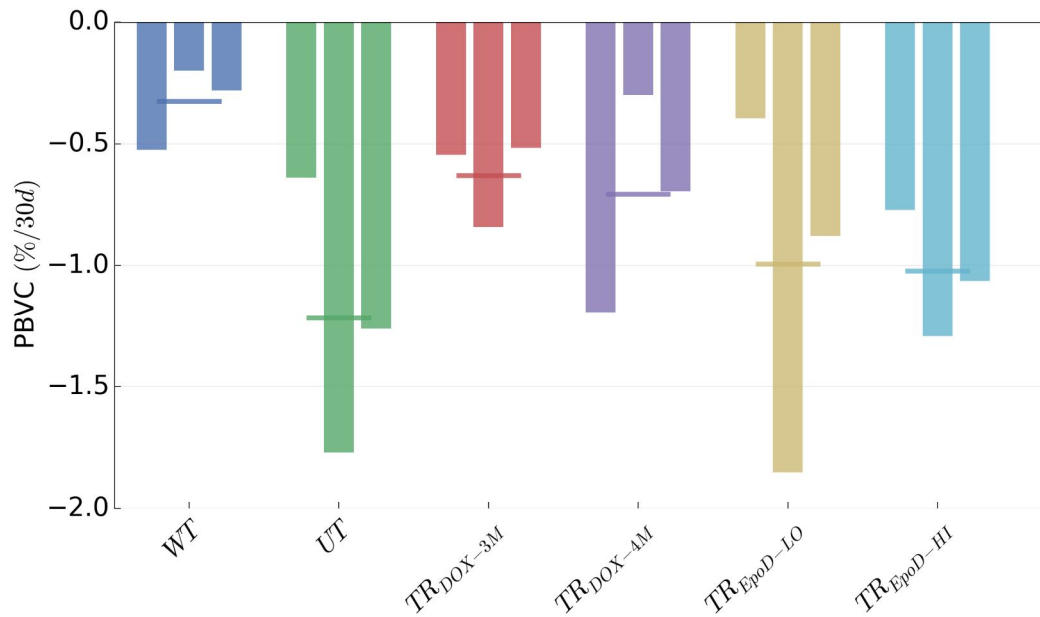
**Figure 6.12** Mean segmentation volume change measurements.

Mean measurements from all groups and TP pairs (TP<sub>1-2</sub>, TP<sub>2-3</sub>, TP<sub>1-3</sub>), standardised to 30d. Horizontal lines represent mean of all TPs. Values in [Table 6.4](#).

group	TP <sub>1-2</sub> mean	SD	TP <sub>2-3</sub> mean	SD	TP <sub>1-3</sub> mean	SD	overall mean	SD
<b>WT</b>	0.0071	0.0131	0.0022	0.0099	0.0036	0.0032	0.0042	0.0095
<b>UT</b>	0.0066	0.0078	0.0135	0.0051	0.0102	0.0038	0.0101	0.0063
<b>TR<sub>DOX-3M</sub></b>	0.0071	<u>0.0048</u>	0.0049	0.0091	0.0055	0.0053	0.0058	0.0063
<b>TR<sub>DOX-4M</sub></b>	0.0124	0.0114	0.0024	0.0059	0.0066	<u>0.0042</u>	0.0069	0.0082
<b>TR<sub>EpoD-LO</sub></b>	0.0022	0.0039	0.0136	0.0027	0.0067	0.0016	0.0071	0.0055
<b>TR<sub>EpoD-HI</sub></b>	0.0064	0.0037	0.0099	0.0040	0.0078	0.0034	0.0079	0.0038

**Table 6.4** Standardised segmentation-based volume change (ml/30 days).

Mean and standard deviation segmentation volume differences between TPs, (ml) all standardised to 30d. The overall mean for each group includes all three TP pairs. The means and standard deviations are greater than those from the BSI in almost all cases (exceptions underlined).



**Figure 6.13** Overall standardised PBVC per group, per 30 days.

Mean PBVC (relative to earliest available BV) calculated from the BSI, for all three TP pairs: TP<sub>1-2</sub>, TP<sub>2-3</sub>, TP<sub>1-3</sub> and overall mean (horizontal lines), standardised to 30d. Values in [Table 6.5](#).

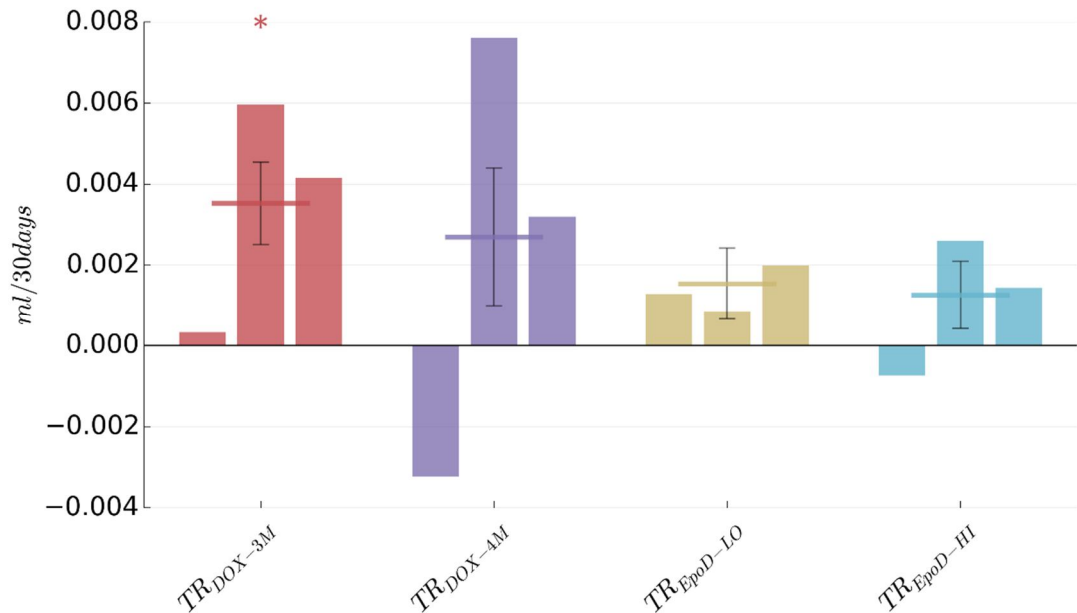
group	TP <sub>1-2</sub> mean	SD	TP <sub>2-3</sub> mean	SD	TP <sub>1-3</sub> mean	SD	overall mean	SD
WT	-0.52	1.75	-0.20	1.10	-0.28	0.42	-0.33	1.17
UT	-0.64	1.12	-1.77	0.68	-1.26	0.45	-1.22	0.91
TR <sub>DOX-3M</sub>	-0.54	1.12	-0.84	0.40	-0.52	0.68	-0.63	0.75
TR <sub>DOX-4M</sub>	-1.19	1.98	-0.30	0.97	-0.70	1.02	-0.71	1.32
TR <sub>EpoD-LO</sub>	-0.40	0.57	-1.85	0.30	-0.88	0.24	-0.99	0.72
TR <sub>EpoD-HI</sub>	-0.77	0.63	-1.29	0.67	-1.06	0.60	-1.02	0.64

**Table 6.5** Percentage brain volume change (PBVC) per group, per 30d.

Mean and standard deviation PBVC, standardised to 30d and relative to TP<sub>1</sub> volume ([E6.10](#)). Overall mean includes all three TP pairs. Excludes values where the corresponding BSI ([Table 6.3](#)) was  $>\pm 2$  SD.

To assess effectiveness of doxycycline and EpoD, I calculated each TR group's effect size from BSI measurements, between each TP pair, by finding the difference in means from the UTs ( $UT_{mean} - TR_{mean}$ , [Fig 6.14](#)). I did not standardise this (e.g. with Cohen's d,

E5.3), as the units of difference (millilitres per 30 days) are inherently meaningful. Only the early treatment with doxycycline produced an effect significantly different from the UT animals.



**Figure 6.14** Effect sizes (difference of means from UT) for each treatment group.

The difference in each TR group’s inter-TP mean from the corresponding UT mean atrophy rate (ml/30d), using BSI. From left, TP<sub>1-2</sub>, TP<sub>2-3</sub>, TP<sub>1-3</sub>. Horizontal lines represent overall mean difference (all TPs). The y-axis corresponds to the tissue volume “preserved”, ml per 30 days, by the treatment. Values are negative when the group exhibits a greater atrophy rate than UTs. The error bars represent  $\pm 1$  SD of the difference between overall means. Only the TR<sub>DOX-3M</sub> group was significantly different, overall (\*,  $p < 0.05$ , [Table 6.6](#)). Outliers ( $\pm 2$  SDs) excluded from [Tables 6.3](#), [6.4](#), [6.5](#) were included in this calculation.

## 6.4 Evaluation

A quandary for clinical atrophy assessment is the lack of “ground truth” tissue loss, which must always be indirectly measured (until death). In mice, this can be serially, histologically investigated *ex vivo* (with neuron counts in proxy animals at different time-points), but in this longitudinal study, the true rate of atrophy was unknown. [Fox et al.](#)

(2011) discussed some “bronze standard” assessments to evaluate whether measurements are consistent and reasonable, including comparison with known disease biology and other measures; transitivity; and symmetry. *Leung et al. (2012)* included bias assessment via linear regression.

- Transitivity: given three TPs, the atrophy  $TP_{1-2} + TP_{2-3}$  should equal  $TP_{1-3}$ .
- Symmetry: the atrophy  $TP_{1-2}$  should equal the absolute of the reverse,  $TP_{2-1}$ .
- Linear regression: atrophy should approach zero as the period between TPs decreases; the regression intercept should be zero if there is no measurement bias.

#### 6.4.1 Comparison with segmentation

*Fig 6.15* illustrates the progression of each group’s brain volume, with time, using the binarised tissue class masks. WT and UT groups are clearly differentiated, but there is overlap between each of the treated groups. To compare the sensitivity of the BSI with the differences in segmented volumes, I performed a t-test between WT and UT atrophy rates (ml/30d), which should be different (*Smith et al., 2007*). *Freeborough & Fox (1997)* showed the BSI better-able to separate human AD and healthy control groups than segmentation. To identify any significant disease-modifying effect on atrophy rates, I performed t-tests between the UT and each TR group, with both BSI and segmentation (*Table 6.6*).

I also compared the difference in overall atrophy rates between BSI and segmentation measures. In the WT, TR<sub>DOX</sub> and TR<sub>EpoD-LO</sub> groups, atrophy rates did not differ significantly ( $p > 0.05$ ). In the UT and TR<sub>EpoD-HI</sub> groups, the segmentation value was significantly greater (respectively,  $p < 0.001$ ;  $p < 0.05$ ).

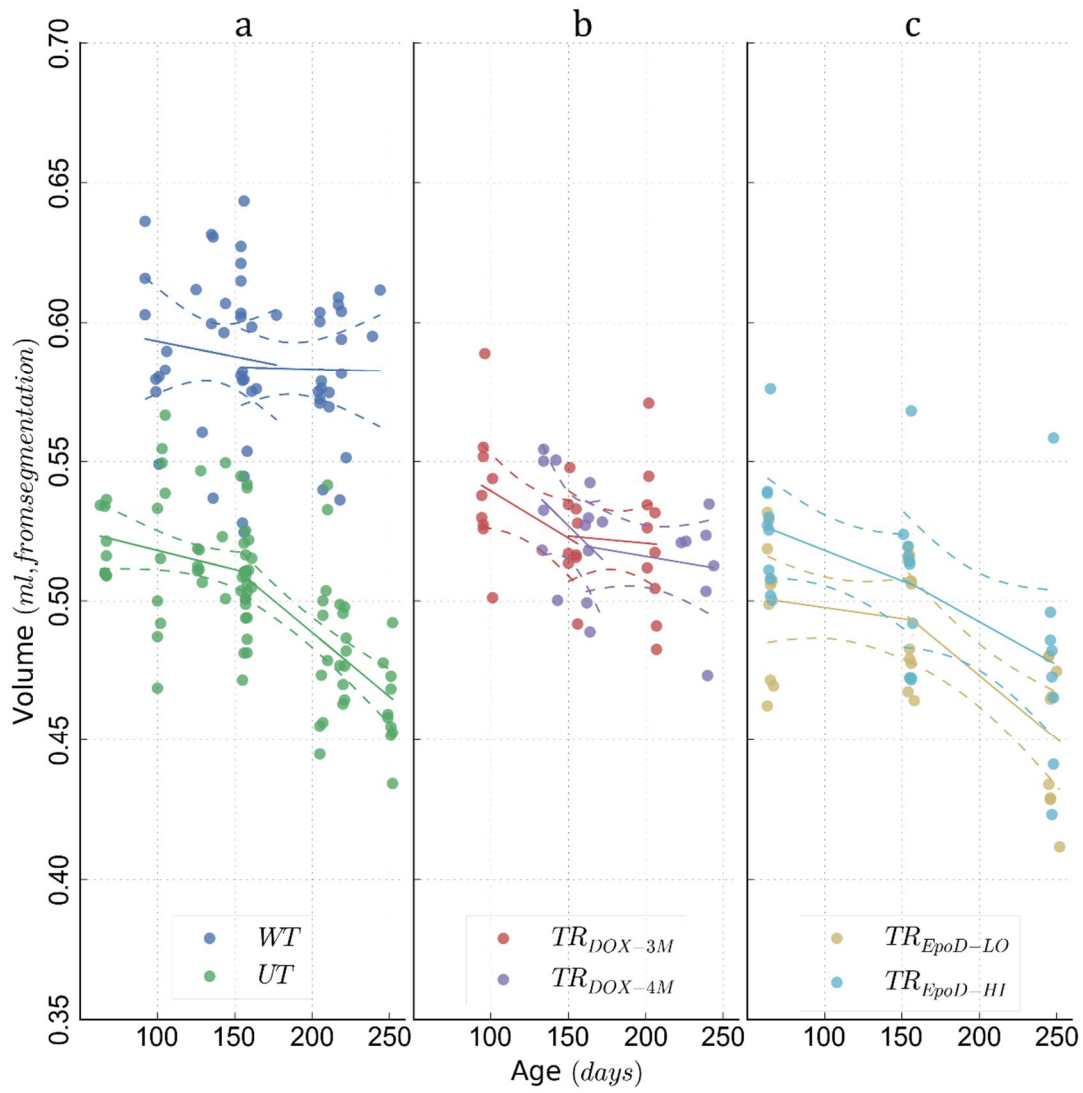
		<i>TP</i> <sub>1-2</sub>	<i>p</i>	<i>TP</i> <sub>2-3</sub>	<i>p</i>	<i>TP</i> <sub>1-3</sub>	<i>p</i>	<i>all TP pairs</i>	<i>p</i>
<b>WT vs. UT</b>	<b>BSI</b>	0.68	0.507	-4.53	<b>1.0x10<sup>-4</sup></b>	-5.49	<b>5.1x10<sup>-6</sup></b>	-2.32	<b>0.023</b>
	<b>seg.</b>	0.84	0.414	-3.96	<b>4.5x10<sup>-4</sup></b>	-6.08	<b>4.1x10<sup>-7</sup></b>	-2.36	<b>0.021</b>
<b>UT vs. TR<sub>DOX-3M</sub></b>	<b>BSI</b>	0.14	0.894	3.76	<b>0.002</b>	3.14	<b>0.010</b>	3.42	<b>0.001</b>
	<b>seg.</b>	-0.21	0.838	2.75	<b>0.018</b>	2.69	<b>0.021</b>	3.15	<b>0.003</b>
<b>UT vs. TR<sub>DOX-4M</sub></b>	<b>BSI</b>	-0.72	0.499	3.68	<b>0.005</b>	1.41	0.211	1.55	0.135
	<b>seg.</b>	-1.18	0.280	4.05	<b>0.002</b>	2.16	0.067	1.57	0.129
<b>UT vs. TR<sub>EPOD-LO</sub></b>	<b>BSI</b>	0.89	0.380	0.69	0.499	2.24	<b>0.039</b>	1.73	0.089
	<b>seg.</b>	2.22	<b>0.033</b>	0.12	0.908	2.83	<b>0.010</b>	2.21	<b>0.031</b>
<b>UT vs. TR<sub>EPOD-HI</sub></b>	<b>BSI</b>	-0.48	0.634	1.77	0.096	1.22	0.249	1.49	0.142
	<b>seg.</b>	0.09	0.928	1.54	0.138	1.99	0.068	2.04	<b>0.045</b>

**Table 6.6** BSI and segmentation's ability to separate WT and UT groups

Two-tailed t-test (unequal variances) t scores comparing 30d-standardised BSI and segmentation-based atrophy rates between WT and UT animals, and between UT and each treatment group, over each TP pair, and overall. Significant p-values bold (<0.05).

Original values in [Tables 6.3, 6.4](#).

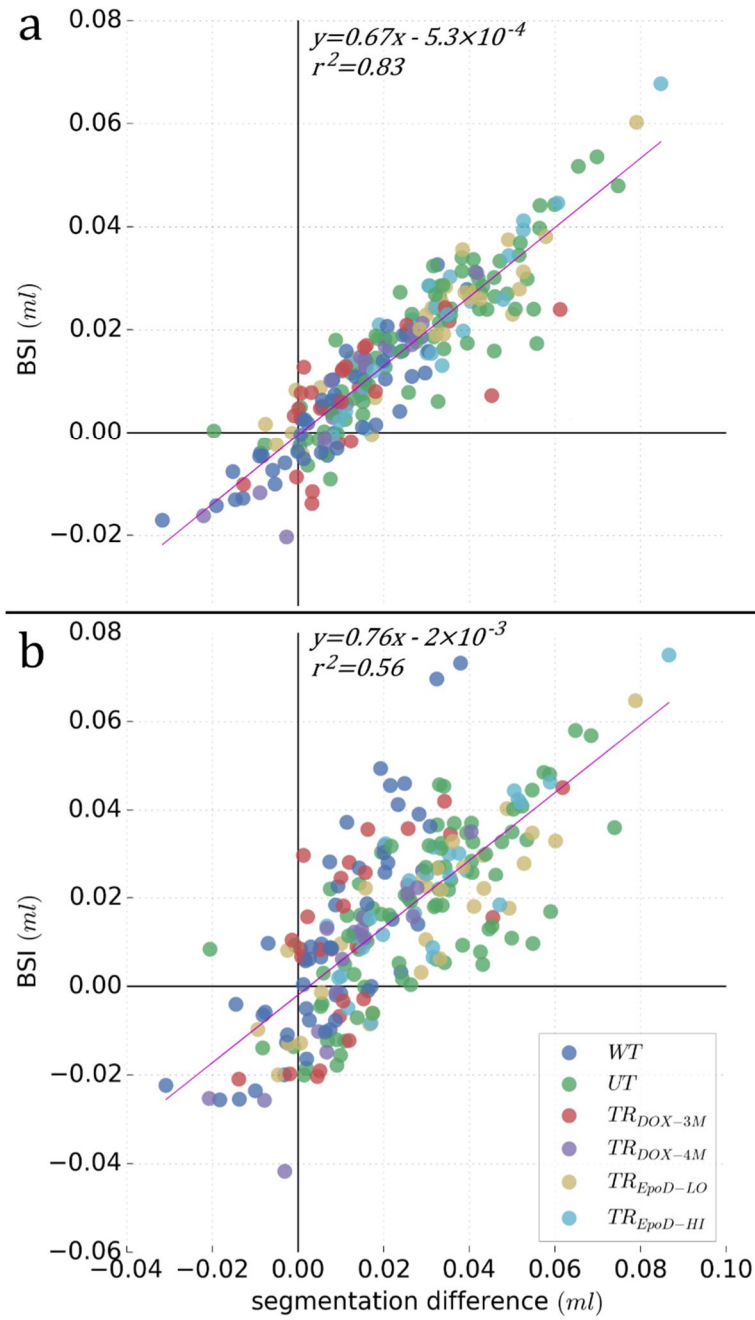
Segmentation was slightly better-able than BSI to separate WT and UT groups, except between TP<sub>2-3</sub>, where BSI was superior. This may be thanks to the increased late atrophy rate in UTs ([Fig 6.11](#)). Neither method separated groups between TP<sub>1-2</sub>, except BSI for UT and TR<sub>EPOD-LO</sub>. Segmentation performed slightly better overall, except distinguishing UT from TR<sub>DOX-3M</sub>. Both measures reveal a disease-modifying effect in TR<sub>DOX</sub> groups between TP<sub>2-3</sub>, and overall in TR<sub>DOX-3M</sub>. The EpoD treatment's preservative effect was smaller ([Fig 6.14](#)) and only detected as significant via segmentation.



**Figure 6.15** Segmented volumes vs. mouse age.

Brain tissue masks' volumes, for each group, at each scan TP. (a): WT, UT; (b):  $TR_{DOX-3M}$ ,  $TR_{DOX-4M}$ ; (c):  $TR_{EpoD-LO}$ ,  $TR_{EpoD-HI}$ . The linear regression fits between  $TP_{1-2}$  and  $TP_{2-3}$  (solid lines) and their 95% confidence intervals (dashed lines) are shown.

[Fig 6.16](#) shows the correlation between brain BSI and the difference between semi-automatically segmented brain volumes (the probabilistic brain segmentation thresholded at 50% and binarised). In the first plot [a](#), the relationship was  $y = 0.67x - 5.3 \times 10^{-4}$ ,  $r^2 = 0.83$ . [Freeborough & Fox \(1997\)](#) showed (in their equivalent figure 4 including human AD and healthy controls) a gradient of  $y = 0.899x + 4.014$ .



**Figure 6.16** BSI vs. segmentation-measured volume differences (in ml).

Illustrating two approaches with original probabilistic brain masks (a) and smoothed by one voxel (b). The proportion of segmented difference recovered by BSI is greater in the latter, at the expense of consistency.

To test whether the comparative under-representation of mouse BV loss measured by BSI was due to a conservative pXOR region, I smoothed the probabilistic masks (Gaussian kernel with FWHM 1 voxel) – thus broadening the high-probability region of

the pXOR. In the second plot [Fig 6.16b](#),  $y = 0.76x - 2 \times 10^{-3}$ ,  $r^2 = 0.56$ . More external materials (dura, skull, non-brain tissues) were included, increasing the variability – notably in the WTs. The BSI recovered a still greater proportion of segmented differences using a binarised XOR region (not shown) – however, this foregoes the advantages of the pXOR, and decreased the  $r^2$  measure further. For this work, I used the unsmoothed pXOR ([Fig 6.16a](#)). An alternative could be to increase the  $\kappa$  parameter in [E6.5](#), which would narrow the high-probability peak in the pXOR and decrease the influence of variable nearby tissues.

## 6.4.2 Transitivity

I evaluated the consistency of BSI measurements using the transitive error (TE) for each animal, defined as:

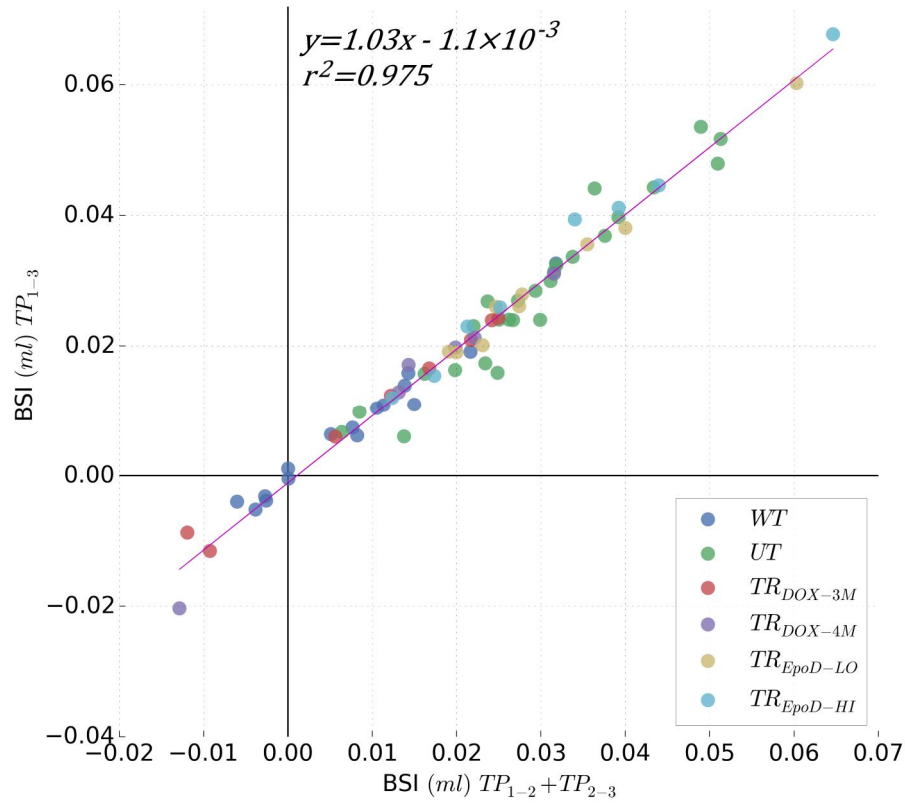
$$\text{E6.11} \quad TE \text{ (ml)} = BSI_{1-2} + BSI_{2-3} - BSI_{1-3}$$

If there is no inherent measurement bias, TE should be zero. To compare TE between animals, I calculated the transitive error rate, defined as this error divided by the mean of absolute BSI measurements for the total period studied:

$$\text{E6.12} \quad TE_{rate} \text{ (\%)} = abs\left(100 \times \frac{TE}{mean(abs(BSI_{1-2} + BSI_{2-3}), abs(BSI_{1-3}))}\right)$$

This was more sensitive in the WTs, as the denominator tended to be smaller. [Fig 6.17](#) shows the linear relationship between  $TP_{1-2} + TP_{2-3}$  and  $TP_{1-3}$ . [Table 6.7](#) shows the mean TE (in ml) for each group, and [Fig 6.18](#) illustrates the rate.





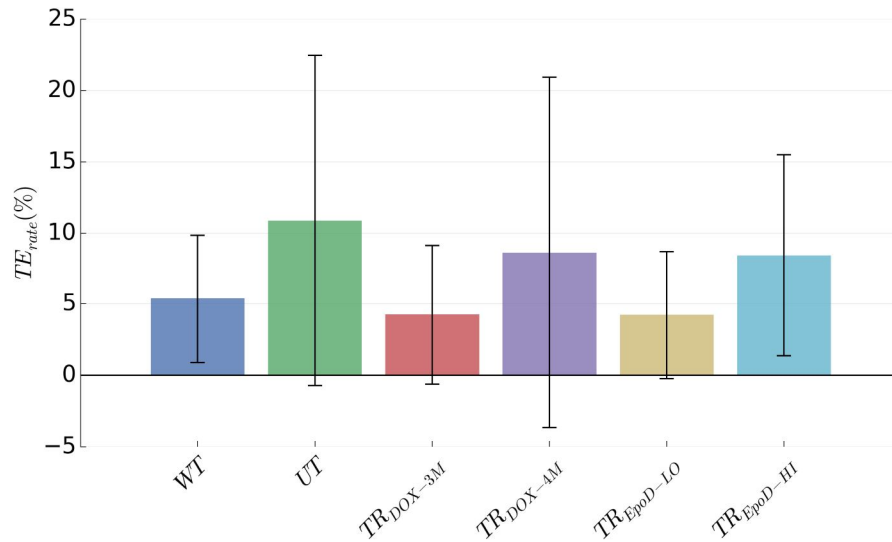
**Figure 6.17** Scatter plot illustrating BSI transitive error.

The technique achieved a high transitive consistency with low error.  $BSI_{1-3}$  was generally slightly greater than  $BSI_{1-2} + BSI_{2-3}$ :  $y = 1.03x - 1.1 \times 10^{-3}$ ;  $r^2 = 0.975$ . Using the smoothed brain masks (§6.4.1),  $y = 0.99x - 1.3 \times 10^{-3}$ ;  $r^2 = 0.911$  (not shown). Equivalent to [Fox & Freeborough \(1997\)](#), figure 7, in human brains.

<i>group</i>	<i>N</i>	<i>mean TE (ml)</i>	<i>SD</i>	<i>CI (95%)</i>	<i>p</i>
<b>WT</b>	16	0.000407	0.001603	[-0.0066, 0.0074]	0.911264
<b>UT</b>	26	0.000956	0.003592	[-0.0058, 0.0077]	0.782980
<b><math>TR_{DOX-3M}</math></b>	8	$6.82 \times 10^{-5}$	0.001548	[-0.0133, 0.0135]	0.992545
<b><math>TR_{DOX-4M}</math></b>	6	0.001078	0.003400	[-0.0162, 0.0184]	0.911695
<b><math>TR_{EpoD-LO}</math></b>	9	0.000642	0.001302	[-0.0110, 0.0123]	0.918278
<b><math>TR_{EpoD-HI}</math></b>	8	-0.001380	0.002209	[-0.0179, 0.0151]	0.878062

**Table 6.7** Mean transitive error measurements, for each group.

The mean and standard deviation TE (ml), with confidence interval and p-value for the t-test comparing  $TP_{1-2} + TP_{2-3}$  with  $TP_{1-3}$ . The difference never approached significance, and there was no appreciable group bias toward greater TE.

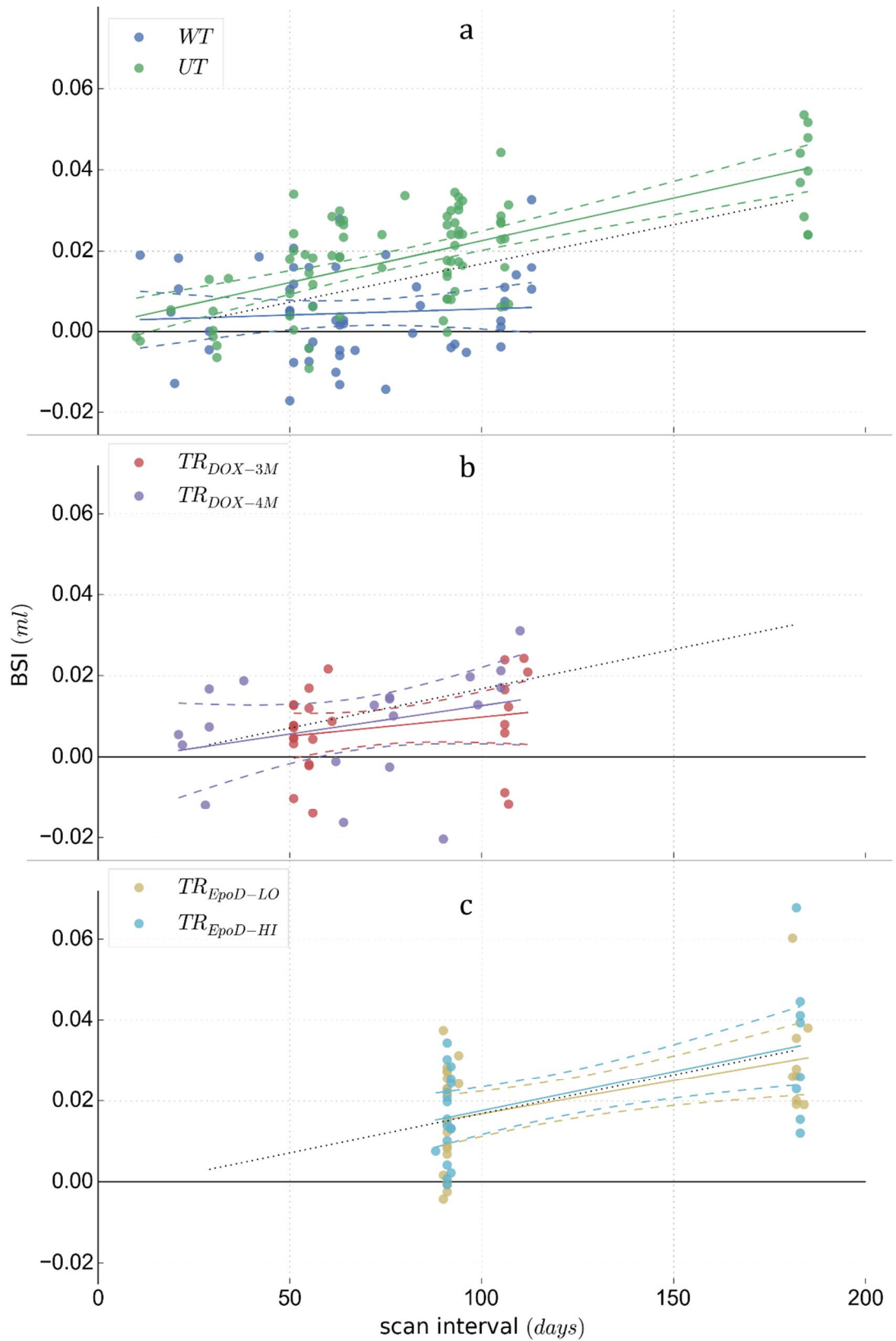


**Figure 6.18** Transitive error rate, per group.

The mean transitive error standardised between groups using [E6.12](#). Error bars:  $\pm 1$  SD.

### 6.4.3 Linear regression

To provide an estimate of systematic bias in the BSI and segmentation measurements, I performed a linear regression of all independent measurements ( $TP_{1-2}$ ,  $TP_{2-3}$ ,  $TP_{1-3}$ ) with the number of days between scans. If there is no bias, the y-intercept should be zero in [Fig 6.19](#) and [Tables 6.8, 6.9](#).



**Figure 6.19** Linear regression of BSI measurements with scan interval between TPs. Illustrating a test for bias in the BSI. If there is no bias, the y-intercept should not be significantly different from zero. The black dotted line is the same in (a,b,c) and represents

the linear fit for all data. Split into three plots for clarity, (a): WT, UT; (b):  $TR_{DOX-3M}$ ,  $TR_{DOX-4M}$ ; (c):  $TR_{EpoD-LO}$ ,  $TR_{EpoD-HI}$ . As per [Leung et al. \(2012\)](#) Fig 4. The slope reveals any linear relationship between time and BSI.  $TP_{1-2}$ ,  $TP_{2-3}$  and  $TP_{1-3}$  are all plotted here, as each is independently measured. See [Table 6.8](#) for values.

<i>group</i>	<i>slope</i>	<i>r<sup>2</sup></i>	<i>intercept</i>	<i>CI (95%)</i>	<i>p</i>
<b>WT</b>	$3.0 \times 10^{-5}$	0.006	$2.53 \times 10^{-3}$	[-0.006 0.011]	0.534
<b>UT</b>	$2.1 \times 10^{-4}$	0.438	$1.51 \times 10^{-3}$	[-0.004 0.007]	0.552
<b><math>TR_{DOX-3M}</math></b>	$9.4 \times 10^{-5}$	0.054	$4.90 \times 10^{-4}$	[-0.012 0.013]	0.937
<b><math>TR_{DOX-4M}</math></b>	$1.4 \times 10^{-4}$	0.101	$-1.29 \times 10^{-3}$	[-0.017 0.014]	0.864
<b><math>TR_{EpoD-LO}</math></b>	$1.6 \times 10^{-4}$	0.236	$4.17 \times 10^{-4}$	[-0.015 0.016]	0.955
<b><math>TR_{EpoD-HI}</math></b>	$1.9 \times 10^{-4}$	0.293	$-1.87 \times 10^{-3}$	[-0.018 0.014]	0.813
<b>all</b>	$1.9 \times 10^{-4}$	0.307	$-2.51 \times 10^{-3}$	[-0.006 0.001]	0.174

**Table 6.8** Linear regression of BSI (ml) with scan interval (days), for each group.

Slope and intercept values corresponding to [Fig 6.19](#). The confidence interval and p-value are shown for the intercept term. No group's p-value approached (arbitrary) significance  $p < 0.05$ , separately or combined. As per [Leung et al. \(2012\)](#) table 5.

<i>group</i>	<i>slope</i>	<i>r<sup>2</sup></i>	<i>intercept</i>	<i>CI (95%)</i>	<i>p</i>
<b>WT</b>	$4.8 \times 10^{-5}$	0.008	$5.17 \times 10^{-3}$	[-0.006 0.016]	0.347
<b>UT</b>	$3.1 \times 10^{-4}$	0.495	$3.27 \times 10^{-3}$	[-0.003 0.010]	0.323
<b><math>TR_{DOX-3M}</math></b>	$2.1 \times 10^{-4}$	0.118	$-2.10 \times 10^{-3}$	[-0.021 0.016]	0.817
<b><math>TR_{DOX-4M}</math></b>	$1.8 \times 10^{-4}$	0.141	$1.35 \times 10^{-3}$	[-0.015 0.018]	0.865
<b><math>TR_{EpoD-LO}</math></b>	$2.5 \times 10^{-4}$	0.266	$-1.32 \times 10^{-3}$	[-0.023 0.020]	0.902
<b><math>TR_{EpoD-HI}</math></b>	$2.6 \times 10^{-4}$	0.355	$7.42 \times 10^{-4}$	[-0.018 0.019]	0.935
<b>all</b>	$2.8 \times 10^{-4}$	0.342	$-1.93 \times 10^{-3}$	[-0.007 0.003]	0.425

**Table 6.9** Linear regression of segmentation difference (ml) with scan interval (days).

For comparison with [Table 6.8](#) (plots not shown). CI and p-values again shown for intercept term. The slope and  $r^2$  values are greater in every case. The intercept term is also greater, and its p-value lower, in the majority of cases (and overall), indicating greater bias in the segmentation measurements.

The results of linear regression with segmentation-based measures of volume change ([Table 6.9](#)) indicate that BSI had a smaller measurement bias than segmentation in the

WT, UT and doxycycline-treated groups, although when all groups were considered together – assuming methodological bias is equal across groups (*Holland et al., 2012*) – segmentation performed better. This may be because of the greater scan interval available for the EpoD measurements, reducing the importance of measurement bias. To better-estimate true volume change, the overall linear regression intercept could be subtracted from all measurements (*Hua et al., 2013* – however, I did not do this).

#### 6.4.4 Power calculation

A good quantification of a measurement’s sensitivity and effectiveness, compared with other measures, is an estimate of the minimum sample sizes required for a hypothetical future study to detect a “significant” effect, i.e., a meaningful reduction in the rate of atrophy, under different treatment regimes. This is also import for ethical reasons, to reduce the number of mice used as far as possible (for the 3Rs); as well as to reduce costs and increase efficiency. I used the group atrophy rates and measurement standard deviations recorded above to compare the BSI with segmentation-derived volume differences.

When considering a reduction in atrophy rate, a more realistic alternative than reducing the rate to 0% is a restoration to WT levels (healthy ageing; *Fox et al., 2011*). This likely requires additional sensitivity, hence more animals. I therefore calculated reductions in the atrophy rate as proportions of the difference between UT and WT rates between each TP pair (*Fox et al., 2000; Schott et al., 2010; Leung et al., 2010b; Prados et al., 2014*). To calculate sample sizes, I used *E6.13* (as *E5.2*):

$$E6.13 \quad N_{arm} = \frac{(Z_{\alpha/2} + Z_{1-\beta})^2 \times 2 \times WT_{stdev}^2}{\Delta\mu^2}$$

Here,  $Z_{\alpha/2} = 1.96$  for two-tailed, 5% significance, and  $Z_{1-\beta} = \{0.84, 1.28\}$  for 80% and 90% power, respectively.  $\Delta\mu$  is 10%, 25% or 50% of the difference between mean 30-day WT and UT atrophy rates for each TP pair. Results are shown in *Table 6.10*.

In keeping with [Fox et al. \(2000\)](#), I calculated increased sample sizes to account for subject attrition ( $\times 110\%$ ) and insufficient quality in either image of the scan pair ( $\times 110\%$ ).

	<i>power</i>	<i>effect</i>	<i>TP<sub>1-2</sub></i>	<i>TP<sub>2-3</sub></i>	<i>TP<sub>1-3</sub></i>	<i>all TP pairs</i>	<i>all, with increase</i>
<b>BSI</b>	80%	50%	62922	16	3	14	17
		25%	251687	61	12	54	66
		10%	1573040	378	71	332	402
	90%	50%	72810	18	4	16	20
		25%	291238	70	14	62	76
		10%	1820232	437	82	384	465
<b>seg.</b>	80%	50%	15377	18	6	6	8
		25%	61506	69	22	21	26
		10%	384407	430	132	130	158
	90%	50%	17793	20	7	7	9
		25%	71171	80	25	25	31
		10%	444814	498	153	151	183

**Table 6.10** Sample sizes for a significant effect, using BSI or segmentation

The number of animals,  $N$ , per arm, to detect 10% or 25% changes in the rate of atrophy at 80 and 90% power. The maximum possible effect (100%) was taken as a reduction to WT levels rather than to no atrophy, which would produce lower sample size estimates. The increase represents 121% of the value calculated from all TP pairs. All calculated values rounded up to the nearest integer. Note the very large  $N$  for  $TP_{1-2}$ : error in some measurements was likely exaggerated when scaling a short inter-TP period to 30 days.

As can be seen from [Table 6.10](#), the  $N_{arm}$  for the first TP pair is extremely unrealistic, due to both high WT variability and the very small difference between early WT and UT rates. Note from [Fig 5.2](#) and [Table 6.1](#) that the  $TP_{1-2}$  periods were short in  $C_1$ ; this likely increased the importance of measurement error when standardising to 30-day rates. The sample sizes were lowest over the longest inter-TP period ( $TP_{1-3}$ ).

Between  $TP_{2-3}$  and  $TP_{1-3}$ , the BSI required lower  $N_{arm}$  than segmentation to detect a significant effect ( $\alpha = 0.05$ ). To detect a 25% reduction in atrophy, these BSI sample sizes were respectively 11.5% and 45.5% lower. When  $TP_{1-2}$  was included ('all TP pairs'), the sample size increased. This is likely peculiar to this study and it may be beneficial to

remove more WT scan pairs from this measurement before calculating the mean and standard deviation atrophy.

## 6.5 Discussion

I have for the first time demonstrated the use of the Boundary Shift Integral, a sensitive, quantitative and objective measure of longitudinal tissue volume change – which meets or exceeds the accuracy of manual segmentation in humans – in a mouse model of disease. To better classify mouse brain tissues, I substituted the k-means algorithm of KN-BSI for a prior-based expectation maximisation method. I adapted a double window approach to accommodate T<sub>2</sub>W images, which are commonly employed in the preclinical paradigm. I used this to capture brain boundary shifts from GM/WM to CSF (in the ventricles, or at various intracranial locations such as the cisternae and entorhinal cortex) and background (including skull, meninges and other low-signal external material). My new DW approach more accurately accounts for direct BG—CSF and BG/brain—brain/CSF PV transitions.

The rTg4510 mouse model of tauopathy exhibits gross whole-brain atrophy after approximately 3 months of age. I measured the BSI at three time-points in the brains of untreated (UT) rTg4510s, WT littermate controls, and rTg4510s treated with doxycycline (a known tau suppressor in this model) or Epothilone D (EpoD, an experimental microtubule stabiliser), to demonstrate the usefulness of the BSI both as a measure of progressive volume change and as a preclinical outcome measure for the effectiveness of disease-modifying drug treatments.

I monitored PBVC with respect to a baseline BV at the earliest time-point. While WT brains were relatively preserved, exhibiting around 0.33% volume reduction between the TPs observed (comparable to the annual rate in humans), as expected, the UT group exhibited the greatest tissue loss of all groups between most TPs (*Figs 6.11, 12, 13* – the TR<sub>EPOD-HI</sub> group had a higher rate between TP<sub>1-2</sub>). This is in line with UT brain weight loss previously observed up to 16 months (*Ramsden et al., 2005*): atrophy continues

throughout the animal's life when left untreated. The rate was most severe, at 1.77% BV loss per 30 days, between TP<sub>2-3</sub>, and steadily increased (age had a small, significant effect on the rate of UT atrophy), although a longer study would be required to thoroughly investigate this trajectory and its linearity. Taking a combination of all BSI measurements, all treatments preserved tissue relative to UT progression (*Fig 6.14; Table 6.6*), but only the early doxycycline treatment had a statistically significant effect on the rate of atrophy.

The BSI returns a whole-brain summary measure, combining the actual atrophy volume with growth, which should balance misregistration. The apparent cerebellar expansion in UTs highlighted by *Fig 6.10* may thus have masked some atrophy, which was restricted to the forebrain. Such an artefact – growth caudally, balanced by atrophy rostrally – is consistent with a rigid registration misalignment. This may partially account for the observed pattern. However, visual inspection of both BSI and registration results reveals a more complex picture. Unlike the human brain, wherein structural changes are relatively subtle over the course of years, the rTg4510 brain undergoes severe morphometric changes – both locally, to internal structures (ventricles; hippocampus), and to the global shape. The forebrain, including ventricles, and the entorhinal cortex region all registered well between TPs; the misalignment of the caudal cerebellar surface appears to result from a shape change to this structure not accounted for by affine registration. Cerebellar expansion could be compensatory to the deleterious forebrain degeneration, but was also observed in WTs (*Table 6.2*).

### 6.5.1 Brain volumes in human and healthy mouse ageing

Healthy human adults exhibit slight progressive deviation of BV below TIV (*Scabhill et al., 2003*, found that the atrophy rate was constant, but accelerated after age 70); AD symptoms manifest as a greater deviation. In human AD, the PBVC has been estimated using the BSI at around 1.3—2.4%; the ventricles grow by around 5.4—13.8%, and the hippocampus atrophies by around 4.6%, all per year. In normal healthy ageing, the brain atrophies by around 0.2—0.5%; the ventricles grow by 1.7—4.1%; and the hippocampus shrinks by 0.3—1.1% per year (*Preboske et al., 2006; Barnes et al., 2008*;



*Leung et al., 2010a; Prados et al., 2014*). Human studies must space scans over long periods to assess these subtle volume changes. Additionally, large numbers of subjects are often needed in study arms, in order to confidently show a drug's disease-modifying effect: with the latest gBSI pipeline,  $N_{arm} \approx 180$  subjects for whole brain BSI, controlling for normal ageing (as here)<sup>5</sup>. Both these factors contribute to clinical drug trials' high costs. In the rTg4510 model, I was able to accurately assess whole-brain atrophy, with good sensitivity, over short (1-3 month) timescales. As mice breed, mature and age quickly, this could be highly advantageous in potential drug therapy treatment trials.

As mentioned earlier (§5.8.1), relatively few imaging studies have investigated mice serially *in vivo*. *Delatour et al. (2006)*, *Lau et al. (2008)*, and *Maheswaran et al. (2009b)* all measured longitudinal BV increases in APP/PS1 mice. The latter group noted that TIV unexpectedly also increased, up to 14 months of age in adulthood, allowing both the ventricles and brain to expand, but that “there appears to be no literature on neurocranial change in the mature normal mouse”. This should be investigated. *Lau et al. (2008)* and *Maheswaran et al. (2009b)* also measured parallel WT BV increases.

As brain growth patterns differ with mouse strain, comparison with age-matched controls – WT and UT animals here – is vital to monitor the natural process of ageing, and hence evaluate the effects of disease and treatments. WT volumes in this study mirrored the slow volume loss of healthy human adults, and UT rTg4510s the atrophy of AD.

## 6.5.2 Treatments and power

To elucidate the sensitivity of the BSI to different treatments, and their effectiveness at reducing disease progression, I calculated changes in atrophy rates in rTg4510s under treatment with doxycycline (from either an early, 3-month or late, 4-month initial TP) and Epothilone D, a microtubule stabiliser.

---

5: 25% effect, 80% power,  $\alpha = 0.05$ . Without controlling for normal aging,  $N_{arm}$ : 66 (1.5T) – 82 (3T).

The BSI revealed a disease-modifying effect – a reduction in the rate of atrophy from UT levels – in both doxycycline and EpoD-treated groups ([Fig 6.14](#)). This was only significant in the early TR<sub>DOX-3M</sub> group, however. Segmentation was able to show a significant, but smaller, beneficial effect of EpoD treatment at both high and low doses. However, the effect size of EpoD was much lower. It could be that a longer treatment regime, or higher doses, are necessary before substantial slowing of the atrophy rate occurs.

Compared with segmented whole-brain volume differences, the BSI underestimated atrophy. However, BSI SD was almost always lower ([Table 6.4](#)), and in all but the TR<sub>EpoD</sub> groups, its measurement bias was lower ([Tables 6.8, 6.9](#)). Between time-points after 50 days of age, and at equivalent power, the BSI also required fewer animals than segmentation to detect a significant disease-modifying effect (a change in the rate of atrophy, [Table 6.10](#)). This bodes well for its usefulness at short timescales, which is apt for pre-clinical trials. (Between TP<sub>1-2</sub>, both the BSI and segmentation had high SD: the shorter inter-TP period may have magnified errors.) Segmentation was better able to separate TR groups from UT, but the BSI performed better in separating UT from WT ([Table 6.6](#)). As noted in §6.4.4, selectively removing some image pairs with high TP<sub>1-2</sub> variance could have produced more achievable sample size estimates for this period. However, [Hua et al. \(2013\)](#), in a TBM study, noted how selective data exclusion introduced an implicit bias and “undue optimism” into sample size estimates. It is not considered good statistical practice to remove data post-hoc, and this would not be done in a clinical trial, so here I did not remove the problematic image pairs.

Mouse attrition was around 5.4%<sup>6</sup>. This is a likely corollary of practical investigations; for this reason, sample numbers are usually inflated above the minimum necessary to detect a significant effect. For example, [Fox et al. \(2000\)](#) assumed both a 10% patient dropout rate and that 10% of scan pairs would be unusable.

---

6: From 261 possible scans (87 mice, 3 TPs), 4 were lost to premature death (4.6% of animals); 2 were lost to indeterminate genotyping; 8 (3.1% of the remaining 255 scans) were discarded for image quality reasons.

Given the relatively small treatment group sample sizes ([Table 6.1](#)), some of the measurements from this study (retrospectively) were underpowered to detect the desired effect sizes, at the given significance. However, even with the increases to account for image problems and animal loss, the numbers are reasonable for a future study to detect 25—50% reductions in atrophy across all three TPs, at 80—90% power ([Table 6.10](#)). As [Lerch et al. \(2012\)](#) noted, increased animal numbers in longitudinal studies are beneficial for statistical power. Here, the inter-TP period also played an important role in separating measurement bias and variance from the true effect of interest. Whereas random errors, such as image noise and inhomogeneity, have a stable magnitude, atrophy accumulates over time, thus SNR increases ([Hua et al., 2013](#)). Required sample sizes could be decreased by increasing the separation of TPs, thereby increasing the intervening volume losses. Although it is better to detect a drug’s disease modifying effect quickly, with limited animal numbers and allowing for attrition, this should be balanced with sufficient inter-TP time, to ensure a study’s robustness. Around two months appears reasonable in this case. A caveat is that the rTg4510 model exhibits severe atrophy, which manifests early in life and is visible by eye. Future work in a model of subtler changes may require greater numbers or longer inter-TP periods.

### 6.5.3 BSI of substructures

In humans, the hippocampus is a prime target for the BSI measurement, owing to its early, preferential sensitivity to neurodegeneration in AD ([Braak & Braak, 1991](#); [Leung et al., 2010a](#)). It is predominantly surrounded by CSF, providing good contrast: [Barnes et al., 2004](#) and [2008](#) used the single-window BSI. As some WM also bounds the structure, [Leung et al. \(2010a\)](#) applied a DW technique. I have shown hippocampal shrinkage via TBM in rTg4510s ([§5](#)). The mouse hippocampus is comparatively large, and bounded by WM of the corpus callosum and fornix, ventricular CSF, and GM of the thalamus and midbrain ([Fig 2.1](#)). A DW approach would almost certainly be required – but especially at these GM boundaries, it may be very difficult to evoke sufficient contrast to accurately measure mouse hippocampal BSI, even with highly accurate parcellation. A contrast agent may aide the differentiation of hippocampal tissues; this

would require prior investigation. In this study, scan parameters were optimised for high GM/WM contrast, to aide inter-subject group registration. A contrast agent was not administered, because manganese has neurotoxic effects, which would have been difficult to disentangle from the longitudinal effects of tauopathy ([Lerch et al., 2012](#)).

Similar impediments might be expected in the mouse caudate putamen (striatum) – a simple symmetric structure in the mouse bounded by the cortex (GM); corpus callosum and fornix (WM); and ventricles. TBM detected severe, early volumetric reductions here, in rTg4510s (e.g. [Fig 5.8](#)). [Hobbs et al. \(2009\)](#) nevertheless successfully measured caudate BSI in humans, using the DW approach.

[Freeborough & Fox \(1997\)](#) found that human ventricular BSI recovered around 97% of segmentation-based volume difference (whole brain BSI: 90%). This performance is attributable to the structures' high contrast. As discussed earlier ([§3.6](#), [§5.3.2](#)), ventricular registration with the available multi-subject mouse atlases (UFL; NUS) can be problematic: ventricle shape and volume between subjects and groups was highly variable ([§5.7.4](#)). The ventricles were here removed from brain masks via expectation-maximisation-based segmentation. They could also be removed via thresholding, which would require prior intensity standardisation. Once segmented, ventricular BSI may be calculated. This would only require a single intensity window in  $T_2W$  images, as both brain tissue and BG have lower intensities. It may be more sensitive than whole-brain BSI, as the ventricles appear to grow dramatically in UT brains at all surfaces. However, brain-ventricle contrast was here lower than brain-BG ([§6.2.6](#)), and the fourth ventricle was especially variable, even in WTs ([Fig 6.10](#)). It would be interesting to include eCSF in this measurement. In humans, modelling eCSF explicitly can be impractical as its segmentation in  $T_1W$  images is easily confused with skull and BG ([Smith et al., 2002](#); [Malone et al., 2015](#)).

#### 6.5.4 Limitations and potential future improvements

Several features of mouse brains make the BSI's application non-trivial. The thin, close-fitting skull may allow brain and external tissues to share PV voxels, complicating tissue classification, surface detection, and likely compounding measurement errors. The

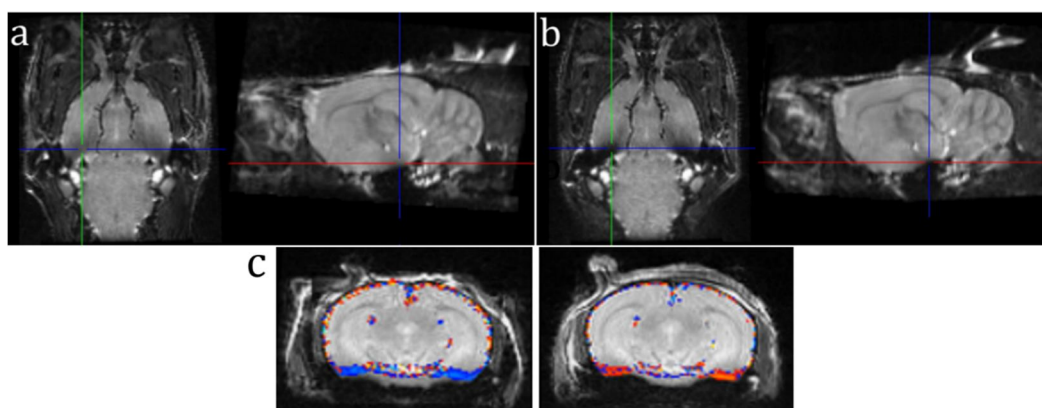
pXOR region, inherited here from the gBSI pipeline, is thus well-suited to the mouse brain. Finer resolution may further improve results; however, the increased time requirements may be unrealistic for high-throughput or multi-parametric studies (§5.8.2).

The many gyri and sulci of the folded human cortex provide a large surface area, proportional to BV, over which to detect atrophy and calculate the BSI. The smooth, lissencephalous mouse cortex has a significantly lower surface area compared with BV, which is also several orders of magnitude smaller (compare the available boundary regions of Figs 2.16, 6.6c). This magnifies the significance of small measurement and parcellation (or tissue classification) errors.

To lend confidence to outcome measures of experimental therapies, the repeatability of structural imaging analysis is vital. This is especially important because human treatment dosages can be selected from preclinical trials (Reagan-Shaw *et al.*, 2008; Jucker, 2010). All scans for this study were performed on the same scanner. In large-scale investigations, robustness must extend between centres and scanner manufacturers. Previous BSI-driven studies have reported standardised analysis parameters (Prados *et al.*, 2014). One of the BSI's chief advantages and reasons for its popularity is its repeatability at different sites. Future work should target refinement of the adaptive parameters selected here, such as the intensity window limits. In particular, the upper window's dependence upon the highly variable CSF intensity SDs (§6.2.6) could be limited – for example by standardising tissue intensities and using a fixed window, as in earliest BSI implementations. Narrowing the window would increase sensitivity – the BSI may recover a greater proportion of the atrophy measured by segmentation – but decrease specificity to subtle intensity changes.

Common imaging artefacts, such as noise, motion and differing contrast severely affect BSI measurements (Freeborough & Fox, 1997; Preboske *et al.*, 2006). I excluded 8 images with severe, subjective quality problems (including motion artefact and noise). However, I did not exclude images with smaller or localised problems. A more conservative approach, such as by ranking images with different raters and excluding subsets of consistently low-ranked images, may have reduced measurement variability, particularly

that unexpectedly observed between TP<sub>1-2</sub> in WTs, although would likely require a boost in animal numbers. Both signal drop-off and the animals' head position contributed to variability and error around the base of the brainstem. I constrained the brain mask to account for this, resulting in fewer detected intensity changes in this region ([Fig 6.10](#)). Susceptibility artefacts also caused some variability ([Fig 6.20](#)).



**Figure 6.20** Susceptibility artefact from the aural cavity causes false BSI results in two registered TP images of a WT brain.

Cross-hairs indicate equivalent positions in the images, and slice locations. There is significant local intensity loss, compared with (a), in (b). BSI results (c) show artificial atrophy (left, from a—b) and growth (right, from b to the next TP). This region, near the mouse ear canals, contains both air and bone ([Goodburn et al., 2015](#)). A similar feature is noticeable in the group mean images ([Fig 6.10](#)), suggesting this was a common artefact.

It is disappointing that the variance of measures was so high, early in the study ([Figs 6.11, 6.12; Tables 6.3, 6.4](#)) – in particular, the WT group. It is likely that subtle early separation of WT and UT atrophy rates was missed.

In addition to the evaluation methods employed here, as a further test of robustness, in a future study it would be informative to perform “back-to-back” scans of the same animal, in the same scanner, at one TP. (The total scan time would need to remain under three hours per session.) The BSI could then be calculated between TPs from each scan to assess the scan-rescan reproducibility (same-day error): no atrophy would be expected between the scan-rescan pair, so any difference in the longitudinal measurement should

be due to noise, scanner artefacts, movement or animal physiology (*Smith et al., 2007*; *Leung et al., 2012*). This would have the benefit of providing a “back-up” scan, in case there were problems with one acquisition. *Lewis & Fox (2004)* suggested that a very small amount of true atrophy may have been obscured by differential bias correction. Although I made parameter adjustments to account for this, the kernel size could be optimised further using back-to-back scans.

A potential extension of this work would be to compare the BSI measurements with Jacobian integration, using the longitudinal TBM technique described in §5.5.2 to generate  $J_{det}$  maps, and integrating the volume changes within brain masks. *Boyes et al. (2006)* performed this comparison in humans using the original implementation of the BSI technique (binary XOR region; asymmetric registration), and found that the BSI non-significantly underestimated atrophy.

SIENA<sup>7</sup>, an alternative technique also only requiring affine registration, has to my knowledge not been applied to mice, but this could also be attempted. *Smith et al. (2007)* showed high correlation between SIENA and BSI results. The method relies upon skull delineation (for registering brains), EM-based tissue classification for accurately identifying brain surfaces, and a normal vector projected from each point of the brain surface, along which intensities are recorded to a given distance. The gradient intensity change along this vector is found at each point, and the surface shift between TPs estimated from the distance required to move vectors for maximum correspondence.

## 6.6 Conclusions

I used the Boundary Shift Integral – an objective and accurate measure of atrophy used to detect structural changes associated with neurodegenerative disease – in a mouse model of tauopathy, and showed early brain tissue losses at short timescales (around one month). The measure showed the ability of doxycycline to limit atrophy in rTg4510 mice, and

---

7: Structural Image Evaluation, using Normalisation, of Atrophy (part of FSL).

differentiated early and late interventions. Epothilone D, a microtubule-stabilising trial drug, did not appear from the BSI to have a significant beneficial effect on brain preservation at the doses tested, although segmentations of the whole brain did detect a comparatively small effect.

This study indicates the viability of BSI in the mouse brain. Unlike TBM, which relies upon non-rigid registration to align voxels and which is hence best-suited to subtle changes, the BSI can be readily deployed to measure gross longitudinal volume differences. Because it does not involve non-rigid registration, the calculation itself is faster. The technique shows promise as a sensitive measure of brain atrophy in mouse disease models – it may be used as an outcome measure in future preclinical drug trials, and to detect early signs of neurodegeneration. The rTg4510 model is well-suited to this paradigm.

While the BSI has been well-optimised in the human brain, the mouse brain posed some challenges, including an increased proportion of PV; a narrower separation of brain and extraneous material; and the reduced surface area offered by the cortex. Despite these difficulties, the BSI showed a comparable ability to segmentation to separate groups in terms of atrophy rate, good consistency (with low transitive error) and lower bias than segmentation. The  $T_2W$  double window technique could also be applied to  $T_2W$  human brain images, and the pipeline could be applied to other preclinical animal images, such as rat brains.

The BSI provides a sensitive, quantified measure for assessing early structural changes in mouse brains. It is a promising technique for future longitudinal tests of preclinical therapeutic efficacy in slowing neurodegenerative disease.



## 7 General discussion, conclusions and future work

This thesis followed the development of software pipelines for phenotyping mouse brains and assessing pathology using voxel-wise techniques; their application to several datasets, and the advance from the ideal case for image quality, *ex vivo*, to *in vivo* scans – enabling longitudinal analyses, including the Boundary Shift Integral, which are useful for monitoring disease progression and for drug studies. The main contributions are:

- The development of a complete, fully automated software pipeline for phenotyping mouse brains with Voxel- and Tensor-Based Morphometry, integrating open-source tools (*NiftyReg*; *NiftySeg*) and best-practice clinical analysis techniques (multi-atlas based parcellation and subject-specific tissue segmentation; symmetric registration). With appropriate atlases (e.g. [Johnson et al., 2010](#); [Cleary et al., 2011a](#); [Calabrese et al., 2013](#)), it would be simple to extend this to other animals' brains, or embryos, for high-throughput studies.
- The application of this pipeline to several datasets and models, both *ex vivo* and *in vivo*, and investigation of the results, including the first morphometric analysis of a mouse model of Down syndrome, as well as a DS model cross-bred with an APP model, and the first longitudinal morphometric investigation of a tauopathy mouse model. The comprehensive whole-brain results ([§4](#), [§5](#)), including novel morphological findings, demonstrate the utility of these techniques for the unbiased structural characterisation of brains.

- An assessment of cross-sectional and longitudinal TBM, in the same dataset, for detecting atrophy, and of *ex vivo* and *in vivo* morphometry for sensitively distinguishing groups.
- The adaptation and application of the Boundary Shift Integral for the first time to non-human animal brains, and also its first application to T<sub>2</sub>W MRI. I showed its sensitive measurement of atrophy over short time periods of 1—2 months and the assessment of a novel drug for reducing atrophy. This technique, too, could be simply translated to other preclinical subjects or T<sub>2</sub>W scans, with an appropriate atlas for prior tissue segmentation.

Although widely used with clinical data, software customised for high-throughput preclinical analysis with these techniques was hitherto unavailable. Labour-intensive pre-processing was required. Consequently, few mouse models have been analysed with the powerful morphometric techniques common to clinical studies. Longitudinal *in vivo* studies are also uncommon, but confer several advantages for monitoring disease progression, including the reduction of inter-individual variability and the number of animals employed. Furthermore, the use of up-to-date, bias-reducing techniques, such as symmetric registration and multi-atlas tissue segmentation and parcellation, should be considered especially important in preclinical studies, whose sensitivity and repeatability are vital for drug trials. However, they have been slower to catch on, and their use is not yet standardised (*McConville et al. 2005*).

### **Limitations**

In this thesis, I focussed on voxel-wise statistical tests of local structural volume or intensity difference between groups. Statistical parametric maps are valuable for exploratory, unbiased or naïve investigations. However, interpretation is still required; a disadvantage of null hypothesis testing is the p-value's disconnection from values of scientific interest. For this reason, I calculated effect sizes and in the BSI investigation the sample sizes required to detect a suitable effect, which should be more useful for therapeutic trials (*Cohen, 1994*). Additionally, statistical tests can only reject H<sub>0</sub>, rather than prove an alternative hypothesis or give an explanatory cause for a difference.

Artefacts not due to volume change or tissue density may arise systematically, for example if one group is more likely to move during the scan.

Significance thresholding is quite arbitrary. Some factors may lend credence to a significant result: bilaterally significant voxels some distance apart likely did not interact during registration and are fully independent. Certain results may also be expected, for example based upon previous studies. Finally, images can be visually inspected for regions of misregistration. An extension of this could involve the integration of nonparametric permutation tests, such as with FSL's *Randomise*, which allows integration of a GLM. These are more robust to data which are not normally distributed (*Ashburner & Friston, 2000; Nichols & Holmes, 2001*), and are increasingly employed in morphometric investigations (e.g. *Chen et al., 2005; Sawiak et al., 2013*).

Like atrophy, which is quantified via the BSI, V/TBM changes are downstream effects of a presumed upstream cause (such as genetic difference, or disease). As *Davatzikos (2004)* noted, and unlike atrophy, these may have many underlying causes and expressions, which are only partially detected using morphometric techniques and not fully explained. With V/TBM, the effect size is only of the concentration or volume difference; with BSI, it is more easily interpretable, and can be directly related to drug treatment. A disadvantage of using atrophy as a disease marker is that many neurochemical processes are upstream in pathology: although treatments preventing atrophy may be beneficial, they may not address the underlying causes.

In §5.7.3 and §6.4.4 I discussed statistical power. Studies which hope to detect a subtle effect without enough subjects are a waste of resources. Power calculations should be made before all morphometric studies, considering the expected effect size, and numbers increased as per *Fox et al. (2000)*: +10% to allow for longitudinal attrition and +10% for image quality issues. These unexpected deleterious outcomes will inevitably occur, no matter how careful and rigorous the study.

Voxel-wise morphometry can complement parcellation-based segmentation, allowing the detection and quantification of small changes within atlas-defined structures. As *Zhang et al. (2010)* noted, subtle changes which are diffuse throughout a particular structure

may be better-detected by parcellation techniques, by combining the contribution of all voxels, as individual voxel-wise tests may lack the power to reach significance. Indeed, the finding of *Van Eede et al. (2013)* that with TBM only 38% of true positive voxels were detected from a 20% volume change is disconcerting. The authors found that a “matched filter” – Jacobian integration over the structure – was more sensitive. With a prior theorised region of difference, one might limit scans to a few ROIs. However, this limits the usefulness of morphometry; whole-brain, unbiased coverage allows the discovery of unexpected changes.

## 7.1 Can *in vivo* replace *ex vivo*?

This thesis followed a progression from *ex vivo* to *in vivo* imaging. There remain some advantages to the former: at high field strengths, long (overnight) multi-subject scans can improve throughput, reduce operator workload and, with optimised fixation, scanning protocols and contrast agents (*Cleary et al., 2011b*), allow fine structural detail at high contrast, improving sensitivity to subtle effects. For high-throughput image processing, these necessitate automated techniques such as multi-subject extraction and orientation (§3.2; §3.3). Steps were employed to reliably fix brains without artefacts, such as by using a low flow rate during perfusion (§4.1.14). Unfortunately, ventricular collapse is common, and occurs to an unpredictable degree. External CSF spaces, such as the entorhinal region, do not appear as susceptible, perhaps due to the lack of surrounding tissues causing compression. Tissues also take a long time to fix (around 9 weeks); however, *von Bohlen und Halbach et al. (2014)* recently showed high-resolution *ex vivo* scans, with good contrast, without fixative, which substantially reduced the necessary preparation time and also precluded tissue shrinkage. *Ex vivo* scans are also currently essential for assessing structural development in embryos (*Nieman et al., 2011*).

*In vivo* opens a wide range of analysis possibilities, and the paradigm additionally allows many other quantitative physiological MR measurements to be made in parallel with structural imaging, such as ASL, DTI, and functional response. These measures may be

more powerful than individual biomarkers, when combined (*Kohannim et al., 2010; Teipel et al., 2013; Wells et al., 2015*). UK regulations stipulate a maximum time a live animal may be anaesthetised to ensure recovery post-scan (the project license for §5 allowed 3 hours, *Wolfensohn & Lloyd, 2007*), and hence image resolution, SNR and CNR are compromised. Additionally, scans risk physiological or motion artefacts. These issues may be overcome using custom equipment. *Baltes et al. (2009)* and *Biedermann et al. (2012)* reported *in vivo* SNR improvements of  $\times 2.4 - 3.5$  by using cryogenically cooled surface coils, allowing high-resolution isotropic voxels ( $60\mu\text{m}^3$  in the former, in under one hour) suitable for VBM, visualisation of the Purkinje cell layer, and angiography.

The structural data acquired for §5 and §6 was complemented by other MR techniques (including ASL and DTI: see *Wells et al., 2015; Holmes et al., 2016*). The structural scan time was necessarily shortened to include these methods; thus, the *in vivo/ex vivo* comparison did not represent a like-for-like comparison of the best “possible” scenarios. However, this is a realistic reflection of “real-world” usage, and the volumetric results from structural data nevertheless provided the earliest marker of tauopathy.

## 7.2 Future work

I hope that the tools developed and described for this thesis may be of use to future investigators. The morphometry pipeline uses Python 2.7 and widely-available, free libraries, including: NumPy, SciPy, NiBabel and Pandas<sup>1</sup>, alongside the open-source, cross-platform tools developed at CMIC (*NiftyReg*, *NiftySeg* and *NiftyView* for visualisation). It was developed for ease-of-use, requiring only command-line input. For example, to orient brains, this is simply:

```
run mm_orient.py -i [input directory] -o [output directory]
```

---

1: Respectively: <http://www.numpy.org>; <http://scipy.org>; <http://nipy.org/nibabel>; <http://pandas.pydata.org>.

where `input_directory` contains NIfTI images. The majority runs in both Windows and Linux environments. It should be straightforward to perform future phenotyping studies using this pipeline, without prior image processing knowledge. *Ashburner & Friston (2005)* noted that unifying segmentation, registration and bias field correction into a single framework can improve results, but it also increases complexity and reduces the ability to customise parameters easily. This is still necessary for the preclinical paradigm, where fewer atlases are available, acquired images can be more variable, and techniques remain non-standardised. It is an advantage to be able to use pipeline steps modularly, for example performing non-uniformity correction before and after masking, and to better-elucidate the source of errors.

The task of identifying phenotypes for each of the 20—25,000 mouse genes is too great for one centre to perform alone. Standardisation and validation of mouse phenotyping screens is recognised as an important step for collaboratively building a complete picture of mammalian gene function (*de Angelis et al., 2015*). MR image analysis techniques for phenotyping should also be standardised, with methodological transparency and the open release of data for validation and duplication between groups, as is increasingly performed in the clinical paradigm (e.g. ADNI; *Cash et al., 2015*).

The BSI could be extended to the mouse ventricles and external CSF spaces. These have been shown in humans to be highly sensitive to disease (*Schott et al., 2005*), could be earlier predictors than whole brain volume, and are less likely to be diluted by corresponding regions of apparent “growth”, as was observed in the cerebellum in rTg4510 mice (*Fig 6.10*). Owing to variability of shape and size, their segmentation in mice is difficult. The hippocampus is another obvious target, but would likely require administration of a contrast agent to achieve sufficient delineation of its boundaries *in vivo*. It should also be possible to translate the T<sub>2</sub>W BSI technique I developed directly back to clinical, human data.

Another approach to the assessment of the BSI could be to artificially simulate atrophy, giving a known “ground truth”, which is lacking from clinical data and would be helpful for comparing techniques. Atrophy simulations are still being developed (e.g. *Camara et*

*al.*, 2008), and may however introduce their own biases, such as the types of deformations employed (*Fox et al.*, 2011).

### 7.2.1 Further phenotyping investigations

Future  $\mu$ MRI phenotyping studies should ideally use *in vivo* subjects; multi-time-point TBM and the BSI are powerful and sensitive measures of both volume difference between groups, and change over time, in the mouse brain, even without contrast agent. However, at the end of a study, *ex vivo* MRI may be warranted. The optimised perfusion and acquisition protocols described by *Cleary et al.* (2011b) should be combined with a low fixative flow rate during perfusion, to prevent artefacts such as the hyperintense rim discussed in §4.1.14. The SNR and CNR achieved in multi-subject scans (which enable increased throughput) were lower than single-subject, but sufficient for morphometry. High resolution, isotropic voxels should ideally be used, but the lower-resolution *in vivo* protocols (100 $\mu$ m voxels) were sufficient to detect early morphometric changes in rTg4510 mice. As MR field strengths continue to increase, higher resolutions may be possible. However, given the success of morphometry *in vivo*, phenotyping may be possible using much smaller, cheaper, lower field strength scanners, such as the Bruker “benchtop” 1T. This would increase the accessibility of structural investigations.

For the development of the pipeline in §3, I focused upon robustness to different  $\mu$ MR mouse imaging paradigms: brain (*in* and *ex vivo*) and embryo; single and multi-subject scans. Thanks to the wealth and variety of data made available through collaboration with CABI, pitfalls arising from variations in image quality and brain shape were encountered and overcome: for example, touching subjects in multi-subject scans (§3.2) and subtle, systematic changes which may have influenced morphometry, such as the “hyperintense rim” perfusion artefact (§4.1.14).

The pipeline has been successfully applied to other *ex vivo* mouse brain datasets, including (but not limited to) a study investigating the effects of sciatic nerve injury on mouse brain morphology, an *ex vivo* single time-point analysis of another FTD mouse model, and a morphological investigation of young transgenic knockout mice modelling Adenosine deaminase (ADA) deficiency, a serious immunodeficiency in humans. I also

optimised the pipeline for high-throughput mouse embryo data (as briefly alluded to, using a model of spina bifida) and rat brains, investigating the neuroprotective effect of nicotinamide in a model of Parkinson's disease.

The early pipeline steps cover necessary pre-processing for atlas-based parcellation, for volume measurement, regional investigation (*Ma et al., 2014*), or localisation of fMRI responses. Steps may be repeated automatically, using their initial outputs, on data corresponding to the structural images. For example, phase contrast images and MR angiographs can be automatically oriented, skull-stripped and aligned into the final GWR average space using the orientation matrices, masks, and deformations generated from structural data. This allows comparison in the average space. It may also be interesting to complement voxel-wise morphometric tests with shape analysis of larger structures: *Badea et al. (2007b)* examined the distance maps at structures' surfaces to compare a mutant mouse with WT's. *Costafreda et al. (2011)* used deformations from the human hippocampal surface to predict AD onset in MCI patients. The unexpected cerebellum change in untreated rTg4510 brains (§6) hints at global shape changes over the whole brain. These could be captured by shape analysis.

Behavioural studies could be undertaken to follow-up the morphometry findings of localised reductions in the Tc1 cerebellum. They could also be used to corroborate a negative result, such as a finding of no structural differences with morphometry or atlas parcellation in the abovementioned nerve injury dataset. However, behavioural or cognitive differences may be further downstream from pathology than structural changes in the brain.

It is valuable to follow-up  $\mu$ MRI investigations with histology, to confirm unexpected volumetric findings and investigate the cytoarchitectural changes (cell size; number; density; type) underlying V/TBM results.  $\mu$ MRI does not reach cellular resolution (*Bock et al., 2006*). Histology was performed in rTg4510 brains (*Wells et al., 2015*; *Holmes et al., 2016*): TBM findings of local volume loss correlated with elevated immunohistochemically stained neurons containing hyperphosphorylated tau, as well as visibly-reduced tissue volumes. Histology was attempted for the Tc1 dataset (§4.1); however, unfortunately, the tissues had deteriorated from prolonged fixation, rendering



histology impossible. Promisingly, [Keifer et al. \(2015\)](#) recently showed increases in dendritic spine density in the auditory cortex, after fear conditioning, correlated with VBM findings of increased GM voxel intensity. Other imaging methods to investigate microstructure, such as high-resolution episcopic microscopy, may enable computational techniques on more detailed 3D images ([Norris et al., 2013](#)).

This thesis was limited to structural MRI. Morphological phenotypes may be examined using many other imaging modalities. For example, [Verma et al. \(2005\)](#) and [Zhang et al. \(2005\)](#) used DTI to measure mouse brain morphological development and phenotypes. [Wong et al. \(2014\)](#) recently showed an atlas-based morphometry pipeline in  $\mu$ CT images of embryos, including a technique for the assessment of gross structural differences not resolvable via registration. Aligned images were instead subtracted to find regions of large intensity difference, and hence morphological variation. This is similar in concept to precursors of the BSI ([Fox et al., 1996](#)), but was quantified in terms of significant differences in intensity. The authors suggested  $\mu$ CT was more appropriate than MRI; excellent contrast and detail was possible; radiation exposure is not an issue *ex vivo*, and the modality is cheaper and faster.

There is also scope to improve understanding of inter-WT mouse brain variability (which would aid sample size calculations) and volume changes in adulthood, especially given the number of in-bred strains ([Beck et al., 2000](#)) and that different WT backgrounds are often used as controls in experimental studies. In following APP models of AD, [Lau et al. \(2008\)](#) and [Maheswaran et al. \(2009b\)](#) both unexpectedly showed BV increase in WT brains throughout adulthood (a finding not repeated in §5 or §6). Skull volume could be monitored longitudinally using  $\mu$ CT for a variety of different strains. This highlights the importance of using WT controls in longitudinal therapeutic investigations. Given the overall TIV finding in Tc1s, controlling for global volume is extremely important in TBM to highlight local differences.

Machine learning methods, such as support vector machines, could be employed to investigate phenotypes and stage disease. [Davatzikos \(2004\)](#) advocated the use of support vector machines over whole images, to separate groups, rather than voxel-wise tests on specific features, (volume; intensity). Advantageously for phenotyping, these might not

require images be segmented, parcellated, or registered to perform comparisons, thus negating the issues of atlas applicability or error, or quality loss and artefacts from interpolation. Machine learning techniques may also combine multiple biomarkers to increase statistical power (*Kohannim et al., 2010; Teipel et al., 2013*).

### 7.2.2 Mouse brain atlasing

Much of the work conducted for this thesis was dependent upon the availability of applicable mouse brain atlases. There are very few multi-subject mouse atlases presently publicly available. *Ma et al. (2014)* produced a summary. The UFL and NUS atlas databases have several drawbacks: the former contains just 10 brains, compared with hundreds, for humans. Although its *ex vivo* images are high resolution (47µm isotropic voxels), they were skull-stripped prior to scanning (with consequent damage), so registration with in-skull images (preferred for phenotyping) is sometimes problematic. Additionally, tissue contrast is low and only 20 structures are delineated. The NUS atlas has more (39), but fewer brains (5), and images are lower resolution (100µm isotropic), *in vivo* and skull-stripped. GM/WM contrast is also low. That, for example, the thin, dark Purkinje cell layer is not delineated in the NUS atlas led to its misclassification as WM in *Fig 3.15*. This was consistent across all Tc1 cohort scans and was not therefore thought to affect VBM results.

Several other public mouse brain atlases exist or are in development, but effort is focussed on the construction of single-subject, probabilistic or average atlases rather than multi-subject databases (*Dorr et al., 2008; Johnson et al., 2010; Ullmann et al., 2012*), rendering the advantages of multi-atlas approaches impossible. *Sawiak et al. (2012)* released a database of 399 WT and transgenic mouse brain scans online, with GM, WM tissue classes from SPMMouse. The segmentations have problems, such as a WM rim (*§2.2.5*), and some of the brains are damaged from skull stripping.

Given the quantity of high-quality structural MRI data acquired by my collaborators during this study, an opportunity exists for the development of a large, high-resolution mouse brain atlas database which could be used for multi-atlas approaches, cover a wide range of ages and strains and fit a diverse array of morphologies. The utility of such

databases has been demonstrated in humans ([Leung et al., 2011](#)). Accurate brain masks (for skull-stripping) and TPMs (for tissue segmentation) would be extremely useful for other studies, and could be generated semi-automatically. A method for doing so might follow these steps:

1. Gather prospective wild-type *in* or *ex vivo* mouse brain images, and associated metadata (genetic background; age; sex; scan parameters, etc.)
2. Filter for quality (exclude damaged brains and images with substantial artefacts)
3. Orient to standard (RAS) space ([§3.3](#))
4. Create brain masks using the UFL atlas, with STEPS ([§3.5](#))
5. Correct intensity inhomogeneities and standardise the intensity scale across images ([§3.4](#), [§3.7](#))
6. Create probabilistic TPMs using the NUS atlas ([§3.6](#)), including GM, WM, ventricular and external CSF, background and non-brain tissue classes.

Steps 4 and 6 could employ a “bootstrapping” strategy ([§5.3.2](#)):

- a. Rank or classify results by quality, and select the  $N$  best
- b. Manually correct these  $N$  if necessary
- c. Include these with the external atlas database, or replace the external atlas entirely
- d. Repeat masking or tissue classification on the remaining brains using the augmented atlas
- e. Repeat a—d until all brains are masked or segmented to a high quality.

A more ambitious extension might be the parcellation of anatomical structures. Manually delineating a large number would be time-consuming and laborious, and require meticulous expert assessment. This could be performed with increasing granularity (larger structures first), but would not be necessary for the atlas to be useful. Other groups (above) have released high-resolution, detailed parcellations using documented protocols. It would be sensible to extend these, using semi-automated techniques, to create large databases. Manual correction would be important, so as not to propagate errors.

The benefits of propagating atlas labels to experimental data are not limited to structural investigations: functional MRI experiments rely upon judgment of the locations of the signal: for this, anatomical reference images could be registered to a sufficiently detailed

atlas. Other investigations requiring ROI measurements (e.g. calculation of  $T_1$  or  $T_2$ ; or of FA from DTI) may also benefit from automatic parcellation, which is more repeatable than manual delineation.

This large database of wild-type scans could also be used in future to reduce the number of WT animals it is necessary to scan to compare with a transgenic mouse or disease model of interest. However, this may be impractical; a large number of variables must be recorded and kept constant for proper comparison, including background strain; age; sex; feeding and light/dark regime (*Zeiss et al., 2012*).

### 7.2.3 The future of animal studies

Common ancestors likely underlie homologous genetic sequences shared between mice and humans. However, diverging evolution of genes' function between species, and the relative heterogeneity of the human genome, mean mouse models must address specific research questions to avoid the confounding effects of interacting genes and environment (*Crum et al., 2003; Zeiss et al., 2012*). The failure of A $\beta$  immunisation in human AD patients (after successfully restoring function in mice) threw the validity of mouse models into question (*Giacobini & Gold, 2013*). Tau may be a better therapeutic target. It is also clearly important to report apparent negative findings – such as the EpoD results reported in §5 and §6 – so studies are not needlessly repeated and a balanced picture can be formed of where mouse models are appropriate (*Jucker, 2010*). EpoD did show some effect in §6, and there is likely an ideal time to begin treatment (*Kipps et al., 2005*). Longitudinal MRI, with the pipelines and computational techniques developed here, could be used to further investigate when this is.

As discussed in §1.2, in many cases mice are excellent models for helping to understand disease processes and potential interventions. However, the eventual aim should be to reduce the use of animals in research as far as possible. Computational simulations of neuronal interactions and the entire brain, such as the *OpenWorm* project<sup>2</sup>, may eventually realise this goal. To this end, more immediately, data should be shared;

---

2: <http://www.openworm.org>

techniques standardised for reproducibility; and highly sensitive, high-powered analysis techniques such as morphometry and the BSI should be used where appropriate. I have shown comparable performance of these techniques *ex vivo* and *in vivo*. *In vivo* scans offer comparable sensitivity and permit longitudinal studies, other acquisition parameters, and a reduction of mice used overall.



# References

- Aggarwal, M., Zhang, J., Miller, M.L., Sidman, R.L., Mori, S. (2009). Magnetic resonance imaging and micro-computed tomography combined atlas of developing and adult mouse brains for stereotaxic surgery. *Neuroscience*, 162(4), 1339–50. doi:10.1016/j.neuroscience.2009.05.070
- Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Phelps, C.H., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 270–279. doi:10.1016/j.jalz.2011.03.008
- Ali, A.A., Dale, A.M., Badea, A., Johnson, G.A. (2005). Automated segmentation of neuroanatomical structures in multispectral MR microscopy of the mouse brain. *NeuroImage*, 27(2), 425–35. doi:10.1016/j.neuroimage.2005.04.017
- Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D. (2009). Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*, 46(3), 726–38. doi:10.1016/j.neuroimage.2009.02.018
- Allemang-Grand, R., Scholz, J., Ellegood, J., Cahill, L.S., Laliberté, C., Fraser, P.E., Josselyn, S.A., Sled, J.G., Lerch, J.P. (2015). Altered brain development in an early-onset murine model of Alzheimer's disease. *Neurobiology of aging*, 36(2), 638–47. doi:10.1016/j.neurobiolaging.2014.08.032
- Alpert, N.M., Bradshaw, J.F., Kennedy, D., Correia, J.A. (1990). The principal axes transformation—a method for image registration. *J Nucl Med.*, 31(10), 1717–22.
- Arata, L.K., Dhawan, A.P., Broderick, J.P., Gaskil-Shipley, M.F., Levy, A.V., Volkow, N.D. (1995). Three-dimensional anatomical model-based segmentation of MR brain images through principal axes registration. *IEEE transactions on bio-medical engineering*, 42(11), 1069–78. doi:10.1109/10.469373
- Arnold, J.B., Liow, J.-S., Schaper, K.A., Stern, J.J., Sled, J.G., Shattuck, D.W., Worth, A.J., Cohen, M.S., Leahy, R.M., Mazziotta, J.C., Rottenberg, D.A. (2001). Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*, 13(5), 931–43. doi:10.1006/nimg.2001.0756
- Ashburner, J., Friston, K.J. (2000). Voxel-based morphometry—the methods. *NeuroImage*, 11(6 Pt 1), 805–21. doi:10.1006/nimg.2000.0582
- Ashburner, J., Friston, K.J. (2001). Why voxel-based morphometry should be used. *NeuroImage*, 14(6), 1238–43. doi:10.1006/nimg.2001.0961
- Ashburner, J., Friston, K.J. (2003) Morphometry. (Ed. Penny, W.D.), in *Human Brain Function* (2nd ed., pp.707-722). Academic Press. ISBN:978-0-12-264841-0
- Ashburner, J., Friston, K.J. (2005). Unified segmentation. *NeuroImage*, 26(3), 839–51. doi:10.1016/j.neuroimage.2005.02.018
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage* 38 95–113. doi:10.1016/j.neuroimage.2007.07.007
- Ashburner, J., Friston, K.J. (2007). Rigid body registration. In: *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press, pp.49-62.
- Ashe, K.H., Zahs, K.R. (2010). Probing the Biology of Alzheimer's Disease in Mice. *Neuron*, 66(5), 631–645. doi:10.1016/j.neuron.2010.04.031

- Avants, B., Epstein, CL., Grossman, M., Gee, JC. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. doi:10.1016/j.media.2007.06.004
- Aylward, EH., Li, Q., Honeycutt, NA., Warren, AC., Pulsifer, MB., Barta, PE., Chan, MD., Smith, PD., Jerram, M., Pearlson, GD. (1999). MRI volumes of the hippocampus and amygdala in adults with Down's syndrome with and without dementia. *Am J Psychiatry* 156(4), 564–568. doi:10.1176/ajp.156.4.564
- Badea, A., Ali-Sharief, A., Johnson, G. (2007a). Morphometric analysis of the C57BL/6J mouse brain. *Neuroimage*, 37(3), 683–93. doi:10.1016/j.neuroimage.2007.05.046
- Badea, A., Nicholls, PJ., Johnson, GA., Wetsel, WC. (2007b). Neuroanatomical phenotypes in the Reeler mouse. *NeuroImage* 34 1363–1374, doi:10.1016/j.neuroimage.2006.09.053
- Badea, A., Johnson, GA., Jankowsky, JL. (2010). Remote sites of structural atrophy predict later amyloid formation in a mouse model of Alzheimer's disease. *NeuroImage*, 50(2), 416–427. doi:10.1016/j.neuroimage.2009.12.070
- Badea, A., Gewalt, S., Avants, BB., Cook, JJ., Johnson, GA. (2012). Quantitative mouse brain phenotyping based on single and multispectral MR protocols. *NeuroImage*, 63(3), 1633–45. doi:10.1016/j.neuroimage.2012.07.021
- Badhwar, A., Lerch, JP., Hamel, E., Sled, JG. (2013). Impaired structural correlates of memory in Alzheimer's disease mice. *NeuroImage: Clinical*, 3, 290–300. doi:10.1016/j.nicl.2013.08.017
- Bae, MH., Pan, R., Wu, T., Badea, A. (2009). Automated segmentation of mouse brain images using extended MRF. *NeuroImage*, 46(3), 717–25. doi:10.1016/j.neuroimage.2009.02.012
- Bagci, U., Udupa, JK., Bai, L. (2010). The influence of intensity standardization on medical image registration. *Proc. of SPIE Vol*, 1–15. doi:10.1117/12.843969
- Baghdadi, L., Zamyadi, M., Sled, JG., Schneider, JE., Bhattacharya, S., Henkelman, RM., Lerch, JP. (2011). Semi-automatic segmentation of multiple mouse embryos in MR images. *BMC Bioinformatics*, 12(1), 237. doi:10.1186/1471-2105-12-237
- Bai, J., Trinh, TLH., Chuang, KH., Qiu, A. (2012). Atlas-based automatic mouse brain image segmentation revisited: model complexity vs. image registration. *Magnetic Resonance Imaging*, 30(6), 789–98. doi:10.1016/j.mri.2012.02.010. Atlas: [http://www.bioeng.nus.edu.sg/cfa/mouse\\_atlas.html](http://www.bioeng.nus.edu.sg/cfa/mouse_atlas.html)
- Balaban, RS., Hampshire, VA. (2001). Challenges in small animal noninvasive imaging. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources*, 42(3), 248–62. doi:10.1093/ilar.42.3.248
- Ballatore, C., Lee, VM-Y., Trojanowski, JQ. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nature reviews Neuroscience*, 8(9), 663–672. doi:10.1038/nrn2194
- Baltes, C., Radzwill, N., Bosshard, S., Marek, D., Rudin, M. (2009). Micro MRI of the mouse brain using a novel 400 MHz cryogenic quadrature RF probe. *NMR in biomedicine*, 22(8), 834–42. doi:10.1002/nbm.1396
- Barber, CB., Dobkin, DP., Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4), 469–483. doi:10.1145/235815.235821
- Barnes, J., Scahill, RI., Boyes, RG., Frost, C., Lewis, EB., Rossor, CL., Rossor, MN., Fox, NC. (2004). Differentiating AD from aging using semiautomated measurement of hippocampal atrophy rates. *NeuroImage*, 23(2), 574–81. doi:10.1016/j.neuroimage.2004.06.028
- Barnes, J., Foster, J., Boyes, RG., Pepple, T., Moore, EK., Schott, JM., Frost, C., Scahill, RI., Fox, NC. (2008). A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4), 1655–71. doi:10.1016/j.neuroimage.2008.01.012
- Barnes, J., Ridgway, GR., Bartlett, J., Henley, SMD., Lehmann, M., Hobbs, N., Clarkson, MJ., MacManus, DG., Ourselin, S., Fox, NC. (2010). Head size, age and gender adjustment in MRI studies: A necessary nuisance? *NeuroImage*, 53(4), 1244–1255. doi:10.1016/j.neuroimage.2010.06.025
- Barten, DM., Fanara, P., Andorfer, C., Hoque, N., Wong, PYA., Husted, KH., Albright, CF., et al. (2012). Hyperdynamic microtubules, cognitive deficits, and pathology are improved in tau transgenic mice with low doses of the microtubule-stabilizing agent BMS-241027. *The Journal of neuroscience*, 32(21), 7137–45. doi:10.1523/JNEUROSCI.0188-12.2012



- Bateman, R.J., Aisen, P.S., de Strooper, B., Fox, N.C., Lemere, C.A., Ringman, J.M., Salloway, S., Sperling, R.A., Windisch, M., Xiong, C. (2011). Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease. *Alzheimer's Research & Therapy*, 3:1. doi:10.1186/alzrt59
- Baxter, L.L., Moran, T.H., Richtsmeier, J.T., Troncoso, J., Reeves, R.H. (2000). Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse. *Human Molecular Genetics*, 9(2), 195–202. doi:10.1093/hmg/9.2.195
- Beacher, F., Daly, E., Simmons, A., Prasher, V., Morris, R., Robinson, C., Lovestone, S., Murphy, K., Murphy, DGM. (2009). Alzheimer's disease and Down's syndrome: an in vivo MRI study. *Psychological medicine*, 39(4), 675–84. doi:10.1017/S0033291708004054
- Beck, J.A., Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J.T., Festing, MFW., Fisher, EMC. (2000). Genealogies of mouse inbred strains. *Nature genetics*, 24(1), 23–25. doi:10.1038/71641
- Beckett, L.A., Donohue, M.C., Wang, C., Aisen, P., Harvey, D.J., Saito, N. (2015). The Alzheimer's Disease Neuroimaging Initiative phase 2: Increasing the length, breadth, and depth of our understanding. *Alzheimer's & Dementia* 11, 823–831. doi:10.1016/j.jalz.2015.05.004
- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1 (1995): 289–300.
- Benveniste, H., Blackband, S. (2002). *Progress in Neurobiology*. 67:5, pp.393–420. doi:10.1016/S0301-0082(02)00020-5
- Bertram, L., Tanzi, R.E. (2008). Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature Reviews Neuroscience*, 9(10), 768–778. doi:10.1038/nrn2494
- Bianchi, P., Bettini, S., Guidi, S., Ciani, E., Trazzi, S., Stagni, F., Ragazzi, E., Franceschini, V., Bartesaghi, R. (2014). Age-related impairment of olfactory bulb neurogenesis in the Ts65Dn mouse model of Down syndrome. *Experimental Neurology*, 251, 1–11. doi:10.1016/j.expneurol.2013.10.018
- Biedermann, S., Fuss, J., Zheng, L., Sartorius, A., Falfán-Melgoza, C., Demirakca, T., Gass, P., Ende, G., Weber-Fahr, W. (2012). In vivo voxel based morphometry: detection of increased hippocampal volume and decreased glutamate levels in exercising mice. *NeuroImage*, 61(4), 1206–12. doi:10.1016/j.neuroimage.2012.04.010
- Bock, N.A., Nieman, B.J., Bishop, J.B., Henkelman, R.M. (2005). In vivo multiple-mouse MRI at 7 Tesla. *Magnetic Resonance in Medicine*, 54(5), 1311–1316. doi:10.1002/mrm.20683
- Bock, N.A., Kovacevic, N., Lipina, T.V., Roder, J.C., Ackerman, S.L., Henkelman, R.M. (2006). In vivo magnetic resonance imaging and semiautomated image analysis extend the brain phenotype for cdf/cdf mice. *The Journal of neuroscience*, 26(17), 4455–9. doi:10.1523/JNEUROSCI.5438-05.2006
- Bookstein, F.L. (2001). "Voxel-based morphometry" should not be used with imperfectly registered images. *NeuroImage*, 14(6), 1454–62. doi:10.1006/nimg.2001.0770
- Bowden, D.M., Johnson, G.A., Zaborsky, L., Green, W.D.K., Moore, E., Badea, A., Dubach, M.F., Bookstein, F.L. (2011). A symmetrical Waxholm canonical mouse brain for NeuroMaps. *Journal of neuroscience methods*, 195(2), 170–5. doi:10.1016/j.jneumeth.2010.11.028
- Boyes, R.G., Rueckert, D., Aljabar, P., Whitwell, J., Schott, J.M., Hill, D.L.G., Fox, N.C. (2006). Cerebral atrophy measurements using Jacobian integration: comparison with the boundary shift integral. *NeuroImage*, 32(1), 159–69. doi:10.1016/j.neuroimage.2006.02.052
- Boyes, R.G., Gunter, J.L., Frost, C., Janke, A.L., Fox, N.C., et al. (2008). Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *Neuroimage*, 39(4), 1752–1762. doi:10.1016/j.neuroimage.2007.10.026.Intensity
- Braak, H., Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3), 271–278. doi:10.1016/0197-4580(95)00021-6
- Brambati, S.M., Renda, N.C., Rankin, K.P., Rosen, H.J., Seeley, W.W., Ashburner, J., Weiner, M.W., Miller, B.L., Gorno-Tempini, M.L. (2007). A tensor based morphometry study of longitudinal gray matter contraction in FTD. *NeuroImage*, 35(3), 998–1003. doi:10.1016/j.neuroimage.2007.01.028
- Brett, M., Penny, W., Kiebel, S. (2003). An Introduction to Random Field Theory. *Human Brain Function*, 1–23. doi:10.1049/sqj.1969.0076

- Brettschneider, J., del Tredici, K., Lee, VM-Y., Trojanowski, JQ. (2015). Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nature reviews. Neuroscience*, 16(2), 109–20. doi:10.1038/nrn3887
- Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, HM. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*, 3(3), 186–191. doi:10.1016/j.jalz.2007.04.381
- Brown, S., Moore, M. (2012). Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Disease models & mechanisms*, 5(3), 289–92. doi:10.1242/dmm.009878
- Buchhave, P., Minthon, L., Zetterberg, H., Wallin, ÅK., Blennow, K., Hansson, O. (2012). Cerebrospinal Fluid Levels of  $\beta$ -Amyloid 1-42, but Not of Tau, Are Fully Changed Already 5 to 10 Years Before the Onset of Alzheimer Dementia. *Arch Gen Psychiatry*. 69(1):98-106. doi:10.1001/archgenpsychiatry.2011.155
- Buckner, RL., Head, D., Parker, J., Fotenos, AF., Marcus, D., Morris, JC., Snyder, AZ. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, 23(2), 724–38. doi:10.1016/j.neuroimage.2004.06.018
- Budin, F., Hoogstoel, M., Reynolds, P., Grauer, M., O'Leary-Moore, SK., Oguz, I. (2013). Fully automated rodent brain MR image processing pipeline on a Midas server: from acquired images to region-based statistics. *Front Neuroinform*, 7(August), 15. doi:10.3389/fninf.2013.00015
- Cahill, LS., Laliberté, CL., Ellegood, J., Spring, S., Gleave, JA., Eede, MC. van, Lerch, JP., Henkelman, RM. (2012). Preparation of fixed mouse brains for MRI. *NeuroImage*, 60(2), 933–9. doi:10.1016/j.neuroimage.2012.01.100
- Calabrese, E., Badea, A., Watson, C., Johnson, GA. (2013). A quantitative magnetic resonance histology atlas of postnatal rat brain development with regional estimates of growth and variability. *NeuroImage*, 71C, 196–206. doi:10.1016/j.neuroimage.2013.01.017
- Camara, O., Schnabel, JA., Ridgway, GR., Crum, WR., Douiri, A., Scahill, RI., Hill, DLG., Fox, NC. (2008). Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal Alzheimer's disease images. *NeuroImage*, 42(2), 696–709. doi:10.1016/j.neuroimage.2008.04.259
- Cardoso, MJ., Clarkson, MJ, Ridgway, GR., Modat, M., Fox, NC., Ourselin, S. (2009). Improved maximum a posteriori cortical segmentation by iterative relaxation of priors. *Med Image Comput Comput Assist Interv.*, 12(Pt 2), 441–9. doi:10.1007/978-3-642-04271-3\_54
- Cardoso, MJ., Clarkson, MJ., Ridgway, GR., Modat, M., Fox, NC., Ourselin, S. (2011). LoAd: a locally adaptive cortical segmentation algorithm. *NeuroImage*, 56(3), 1386–97. doi:10.1016/j.neuroimage.2011.02.013
- Cardoso, MJ., Melbourne, A., Kendall, GS., Modat, M., Robertson, NJ., Marlow, N., Ourselin, S. (2013a). AdaPT: an adaptive preterm segmentation algorithm for neonatal brain MRI. *NeuroImage*, 65, 97–108. doi:10.1016/j.neuroimage.2012.08.009
- Cardoso, MJ., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, NC., Ourselin, S. (2013b). STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Medical Image Analysis*, 17(6), 671–84. doi:10.1016/j.media.2013.02.006
- Carducci, F., Onorati, P., Condoluci, C., Di Gennaro, G., Quarato, PP., Pierallini, A., Sarà, M., Miano, S., Cornia, R., Albertini, G. (2013). Whole-brain voxel-based morphometry study of children and adolescents with Down syndrome. *Functional Neurology*, 28(1), 19–28.
- Carey, AN., Liu, X., Mintzopoulos, D., Paris, JJ., Muschamp, JW., McLaughlin, JP., Kaufman, MJ. (2013). Conditional Tat protein expression in the GT-tg bigenic mouse brain induces gray matter density reductions. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 43, 49–54. doi:10.1016/j.pnpbp.2012.12.018
- Cash, DM., Frost, C., Iheme, LO., Ünay, D., Kandemir, M., Fripp, J., Salvado, O., Bourgeat, P., Reuter, M., Ourselin, S., et al. (2015). Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *NeuroImage*, 123, 149–164. doi:10.1016/j.neuroimage.2015.07.087
- Chakravarty, MM., Steadman, P., van Eede, MC., Calcott, RD., Gu, V., Shaw, P., Raznahan, A., Collins, DL., Lerch, JP. (2013). Performing label-fusion-based segmentation using multiple automatically generated templates. *Human Brain Mapping*, 34(10), 2635–54. doi:10.1002/hbm.22092

- Chan, D., Fox, NC., Jenkins, R., Schill, RI., Crum, WR., Rossor, MN. (2001). Rates of global and regional cerebral atrophy in AD and frontotemporal dementia. *Neurology*, 57(10): 1756-1763. doi:10.1212/WNL.57.10.1756
- Chen, XJ., Kovacevic, N., Lobaugh, NJ., Sled, JG., Henkelman, RM., Henderson, JT. (2005). Neuroanatomical differences between mouse strains as shown by high-resolution 3D MRI. *NeuroImage*, 29(1), 99–105. doi:10.1016/j.neuroimage.2005.07.008
- Chen, Y., Dyakin, VV, Branch, CA., Ardekani, B., Yang, D., Guilfoyle, DN., Peterson, J., Peterhoff, C., Ginsberg, SD., Cataldo, AM., Nixon, RA. (2009). In vivo MRI identifies cholinergic circuitry deficits in a Down syndrome model. *Neurobiology of Aging*, 30(9), 1453–65. doi:10.1016/j.neurobiolaging.2007.11.026
- Cheng, Y., Peng, Q., Hou, Z., Aggarwal, M., Zhang, J., Mori, S., Ross, CA., Duan, W. (2011). Structural MRI detects progressive regional brain atrophy and neuroprotective effects in N171-82Q Huntington's disease mouse model. *Neuroimage*, 56(3), 1027–1034. doi:10.1016/j.neuroimage.2011.02.022.Structural
- Chételat, G., Landeau, B., Eustache, F., Mézenge, F., Viader, F., de la Sayette, V., Desgranges, B., Baron, J-C. (2005). Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *NeuroImage*, 27(4), 934–946. doi:10.1016/j.neuroimage.2005.05.015
- Chételat, G., Villemagne, VL., Bourgeat, P., Pike, KE., Jones, G., Ames, D., Ellis, KA., Szeoke, C., Martins, RN., O'Keefe, GJ., Salvado, O., Masters, CL., Rowe, CC. (2010). Relationship between atrophy and  $\beta$ -amyloid deposition in Alzheimer disease. *Annals of Neurology*, 67(3), 317–324. doi:10.1002/ana.21955
- Chou, N., Wu, J., Bingren, JB., Qiu, A., Chuang, K-H. (2011). Robust Automatic Rodent Brain Extraction Using 3-D Pulse-Coupled Neural Networks (PCNN). *IEEE Transactions on Image Processing*, 20(9), 2554–2564. doi:10.1109/TIP.2011.2126587
- Chuang, N., Mori, S., Yamamoto, A., Jiang, H., Ye, X., Xu, X., Richards, LJ., Nathans, J., Miller, MI., Toga, AW., Sidman, RL., Zhang, J. (2011). An MRI-based atlas and database of the developing mouse brain. *NeuroImage*, 54(1), 80–9. doi:10.1016/j.neuroimage.2010.07.043
- Chung, MK., Worsley, KJ., Paus, T., Cherif, C., Collins, DL., Giedd, JN., Rapoport, JL., Evans, AC. (2001). A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14(3), 595–606. doi:10.1006/nimg.2001.0862
- Chung, MK. (2012). *Computational Neuroanatomy: The Methods*. World Scientific. ISBN:978-981-4335-43-0
- Church, DM., Goodstadt, L., Hillier, LW., Zody, MC., Goldstein, S., She, X., Ponting, CP., The Mouse Genome Sequencing Consortium, et al. (2009). Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS biology*, 7(5), e1000112. doi:10.1371/journal.pbio.1000112
- Clapcote, SJ., Lipina, TV., Millar, JK., Mackie, S., Christie, S., Ogawa, F., Lerch, JP., Trimble, K., Uchiyama, M., Sakuraba, Y., Kaneda, H., Shiroishi, T., Houslay, MD., Henkelman, RM., Sled, JG., Gondo, Y., Porteous, DJ., Roder, JC. (2007). Behavioral Phenotypes of Disc1 Missense Mutations in Mice. *Neuron*, 54(3), 387–402. doi:10.1016/j.neuron.2007.04.015
- Clark, CM., Davatzikos, C., Borthakur, A., Newberg, A., Leight, S., Lee, VM-Y., Trojanowski, JQ. (2008). Biomarkers for Early Detection of Alzheimer Pathology. *Neurosignals*, 16(1), 11–18. doi:10.1159/000109754
- Cleary, JO., Price, AN., Thomas, DL., Scambler, PJ., Kyriakopoulou, V., McCue, K., Schneider, JE., Ordridge, RJ., Lythgoe, M. F. (2009). Cardiac phenotyping in ex vivo murine embryos using microMRI. *NMR in biomedicine*, 22(8), 857–66. doi:10.1002/nbm.1400
- Cleary, JO., Modat, M., Norris, FC., Price, AN., Jayakody, SA., Martinez-Barbera, JP., Greene, NDE., Hawkes, DJ., Ordridge, RJ., Scambler, PJ., Ourselin, S., Lythgoe, MF. (2011a). Magnetic resonance virtual histology for embryos: 3D atlases for automated high-throughput phenotyping. *NeuroImage*, 54(2), 769–78. doi:10.1016/j.neuroimage.2010.07.039
- Cleary, JO., Wiseman, FK., Norris, FC., Price, AN., Choy, M., Tybulewicz, VLJ., Ordridge, RJ., Brandner, S., Fisher, EMC., Lythgoe, MF. (2011b). Structural correlates of active-staining following magnetic resonance microscopy in the mouse brain. *NeuroImage*, 56(3), 974–83. doi:10.1016/j.neuroimage.2011.01.082
- Coe, R. (2002). It's the effect size, stupid. *Annual Conf. British Educational Research Association* (pp.1–15).
- Cohen, J. (1994). The Earth Is Round ( $p < .05$ ). *American Psychologist*, 49(12), 997–1003.

- Collins, F., Rossant, J., Wurst, W. (2007). A mouse for all reasons. *Cell*, 128(1), 9–13. doi:10.1016/j.cell.2006.12.018
- Costafreda, S.G., Dinov, I.D., Tu, Z., Shi, Y., Liu, C.-Y., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Simmons, A., et al. (2011). Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *NeuroImage*, 56(1), 212–219. doi:10.1016/j.neuroimage.2011.01.050
- Crum, W., Griffin, L., Hill, D.L., Hawkes, D.J. (2003). Zen and the art of medical image registration: correspondence, homology, and quality. *NeuroImage*, 20(3), 1425–1437. doi:10.1016/j.neuroimage.2003.07.014
- Cryan, J.F., Holmes, A. (2005). The ascent of mouse: advances in modelling human depression and anxiety. *Nature reviews. Drug discovery*, 4(9), 775–90. doi:10.1038/nrd1825
- Davatzikos, C. (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, 23(1), 17–20. doi:10.1016/j.neuroimage.2004.05.010
- de Angelis, M.H., Nicholson, G., Selloum, M., White, J.K., Morgan, H., Ramirez-Solis, R., Brown, S.D.M., et al. (2015). Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nature genetics* (9). doi:10.1038/ng.3360
- Delatour, B., Guégan, M., Volk, A., Dhenain, M. (2006). In vivo MRI and histological evaluation of brain atrophy in APP/PS1 transgenic mice. *Neurobiology of aging*, 27(6), 835–47. doi:10.1016/j.neurobiolaging.2005.04.011
- Do, C.B., Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature biotechnology*, 26(8), 897–899. doi:10.1038/nbt1406
- Dorr, A.E., Lerch, J.P., Spring, S., Kabani, N., Henkelman, R.M. (2008). High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult C57Bl/6J mice. *NeuroImage* 42 60–69. doi:10.1016/j.neuroimage.2008.03.037
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., May, A. (2004). Changes in grey matter induced by training. *Nature*, 427, 311–312. doi:10.1038/427311a
- Drake, R., Lowrie, D.J. (2009). *Gray's Anatomy for Students*. W.B.Saunders Co. ISBN:9781416065920
- Duff, K., Suleman, F. (2004). Transgenic mouse models of Alzheimer's disease: how useful have they been for therapeutic development? *Briefings in Functional Genomics and Proteomics*, 3(1), 47–59. doi:10.1093/bfpgp/3.1.47
- Ellegood, J., Pacey, L.K., Hampson, D.R., Lerch, J.P., Henkelman, R.M. (2010). Anatomical phenotyping in a mouse model of fragile X syndrome with magnetic resonance imaging. *NeuroImage*, 53(3), 1023–1029. doi:10.1016/j.neuroimage.2010.03.038
- Ellegood, J., Lerch, J.P., Henkelman, R.M. (2011). Brain abnormalities in a Neuroligin3 R451C knockin mouse model associated with autism. *Autism Research*, 4(5), 368–376. doi:10.1002/aur.215
- Ellegood, J., Babineau, B.A., Henkelman, R.M., Lerch, J.P., Crawley, J.N. (2013). Neuroanatomical analysis of the BTBR mouse model of autism using magnetic resonance imaging and diffusion tensor imaging. *NeuroImage*, 70, 288–300. doi:10.1016/j.neuroimage.2012.12.029
- Evans, A.C., Janke, A.L., Collins, D.L., Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, 62(2), 911–22. doi:10.1016/j.neuroimage.2012.01.024
- Ewers, M., Teipel, S.J., Dietrich, O., Schönberg, S.O., Jessen, F., Heun, R., Scheltens, P., van de Pol, L., Freymann, N.R., Moeller, H.-J., Hampel, H. (2006). Multicenter assessment of reliability of cranial MRI. *Neurobiology of Aging*, 27(8), 1051–1059. doi:10.1016/j.neurobiolaging.2005.05.032
- Ewers, M., Sperling, R.A., Klunk, W.E., Weiner, M.W., Hampel, H. (2011). Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends in neurosciences*, 34(8), 430–42. doi:10.1016/j.tins.2011.05.005
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J.V., Pieper, S., Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging*, 30(9), 1323–41. doi:10.1016/j.mri.2012.05.001

- Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Scazufca, M., et al. (2005). Global prevalence of dementia: A Delphi consensus study. *Lancet*, 366, 2112–2117. doi:10.1016/S0140-6736(05)67889-0
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Dale, A.M., et al. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355. doi:10.1016/S0896-6273(02)00569-X
- Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B. (2014). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in Neurobiology*, 117, 20–40. doi:10.1016/j.pneurobio.2014.02.004
- Florey, C.D.V. (1993). Sample size for beginners. *BMJ (Clinical research ed.)*, 306(6886), 1181–1184. doi:10.1136/bmj.306.6886.1181
- Fox, N.C., Freeborough, P.A., Rossor, M.N. (1996). Visualisation and quantification of rates of atrophy in Alzheimer's disease. *The Lancet*, 348(9020), 94–97. doi:10.1016/S0140-6736(96)05228-2
- Fox, N.C., Freeborough, P.A. (1997). Brain atrophy progression measured from registered serial MRI: Validation and application to Alzheimer's disease. *Journal of Magnetic Resonance Imaging*, 7(6), 1069–1075. doi:10.1002/jmri.1880070620
- Fox, N.C., Scahill, R.I., Crum, W.R., Rossor, M.N. (1999). Correlation between rates of brain atrophy and cognitive decline in AD. *Neurology*, 52(8), 1687–1689. doi:10.1212/WNL.52.8.1687
- Fox, N.C., Cousens S., Scahill R., Harvey R.J., Rossor M.N. (2000). Using Serial Registered Brain Magnetic Resonance Imaging to Measure Disease Progression in Alzheimer Disease: Power Calculations and Estimates of Sample Size to Detect Treatment Effects. *Arch Neurol*. 57(3):339-344. doi:10.1001/archneur.57.3.339
- Fox, N.C., Ridgway, G.R., Schott, J.M. (2011). Algorithms, atrophy and Alzheimer's disease: cautionary tales for clinical trials. *NeuroImage*, 57(1), 15–8. doi:10.1016/j.neuroimage.2011.01.077
- Freeborough, P.A., Fox, N.C. (1997). The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans Med Imag*, 16(5), 623–629. doi:10.1109/42.640753
- Frisoni, G.B., Scheltens, P., Galluzzi, S., Nobili, F.M., Fox, N.C., Robert, P.H., Soininen, H., Wahlund, L.-O., Waldemar, G., Salmon, E. (2003). Neuroimaging tools to rate regional atrophy, subcortical cerebrovascular disease, and regional cerebral blood flow and metabolism: consensus paper of the EADC, 1371–1381. doi:10.1136/jnnp.74.10.1371
- Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6(2), 67–77. doi:10.1038/nrneurol.2009.215
- Frisoni, G.B., Bocchetta, M., Chételat, G., Rabinovici, G.D., de Leon, M.J., Kaye, J., Reiman, E.M., Scheltens, P., Barkhof, F., Black, S.E., Brooks, D.J., Carrillo, M.C., Decarli, C., et al. (2013). Imaging markers for Alzheimer disease: which vs how. *Neurology*, 81(5), 487–500. doi:10.1212/WNL.0b013e31829d86e8
- Fuchs, H., Gailus-Durner, V., Adler, T., de Angelis, M.H., et al. (2011). Mouse phenotyping. *Methods*, 53(2), 120–35. doi:10.1016/j.ymeth.2010.08.006
- Galante, M., Jani, H., Vanes, L., Daniel, H., Fisher, E.M.C., Tybulewicz, V.L.J., Bliss, T.V.P., Morice, E. (2009). Impairments in motor coordination without major changes in cerebellar plasticity in the Tc1 mouse model of Down syndrome. *Human Molecular Genetics*, 18(8), 1449–63. doi:10.1093/hmg/ddp055
- Gauthier, S. (ed) *Clinical diagnosis and management of Alzheimer's disease*, 3<sup>rd</sup> ed. Informa. ISBN:9780415372992
- Genovese, C.R., Lazar, N.A., Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15(4), 870–8. doi:10.1006/nimg.2001.1037
- Gerig, G., Oguz, I., Gouttard, S., Lee, J., An, H., Lin, W., McMurray, M., Grewen, K., Johns, J., Styner, M.A. (2011). Synergy of image analysis for animal and human neuroimaging supports translational research on drug abuse. *Frontiers in psychiatry*, 2(October), 53. doi:10.3389/fpsy.2011.00053
- Giacobini, E., Gold, G. (2013). Alzheimer disease therapy--moving from amyloid- $\beta$  to tau. *Nature reviews. Neurology*, 9(12), 677–86. doi:10.1038/nrneurol.2013.223

- Giannakopoulos, P., Herrmann, FR., Bussière, T., Bouras, C., Kövari, E., Perl, DP., Morrison, JH., Gold, G., Hof, PR. (2003). Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology*, 60(9), 1495–1500. doi:10.1212/01.WNL.0000063311.58879.01
- Goedert, M., Spillantini, MG. (2006). A century of Alzheimer's disease. *Science*, 314:5800, pp.777-781 doi:10.1126/science.1132814
- Good, C., Johnsrude, IS., Ashburner, J., Henson, RNA., Friston, KJ., Frackowiak, SJ. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, 14(1 Pt 1), 21–36. doi:10.1006/nimg.2001.0786
- Goodburn, R., Powell, N., O'Callaghan, J., Walker-Samuel, S., Shmueli, K. (2015). Characterising and Modelling Susceptibility Artifacts in the Mouse Brain at 9.4T. *Proc. British Chapter ISMRM*.
- Götz, J., Ittner, LM. (2008). Animal models of Alzheimer's disease and frontotemporal dementia. *Nature reviews Neuroscience*, 9(7), 532–544. doi:10.1038/nrn2420
- Gribble, SM., Wiseman, FK., Clayton, S., Prigmore, E., Langley, E., Yang, F., Maguire, S., Fu, B., Rajan, D., Sheppard, O., Scott, C., Hauser, H., Stephens, PJ., Stebbings, LA., Ng, BL., Fitzgerald, T., Quail, MA., Banerjee, R., Rothkamm, K., Tybulewicz, VLJ., Fisher, EMC., Carter, NP. (2013). 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*, 8(4), e60482. doi:10.1371/journal.pone.0060482
- Groppe, DM., Urbach, TP., Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology*, 48(12), 1711–25. doi:10.1111/j.1469-8986.2011.01273.x
- Guedj, F., Pereira, PL., Najas, S., Barallobre, M.-J., Chabert, C., Souchet, B., Sebrie, C., Verney, C., Herault, Y., Arbones, M., Delabar, JM. (2012). DYRK1A: a master regulatory protein controlling brain growth. *Neurobiology of disease*, 46(1), 190–203. doi:10.1016/j.nbd.2012.01.007
- Gunter, C., Dhand, R. (2002). Human biology by proxy. *Nature*, 420. doi:10.1038/420509a
- Gustaw-Rothenberg, K., Lerner, A., Bonda, DJ., Lee, H., Zhu, X., Perry, G., Smith, MA. (2010). Biomarkers in Alzheimer's disease: past, present and future. *Biomarkers in Medicine*, 4(1), 15–26.
- Gutman, BA., Wang, Y., Yanovsky, I., Hua, X., Toga, AW., Jack Jr, CR., Weiner, MW., Thompson, PM. (2014). Empowering imaging biomarkers of Alzheimer's disease. *Neurobiology of aging*. doi:10.1016/j.neurobiolaging.2014.05.038
- Hall, AM., Roberson, ED. (2012). Mouse models of Alzheimer's disease. *Brain Research Bulletin*, 88(1), 3–12. doi:10.1016/j.brainresbull.2011.11.017
- Hammers, A., Allom, R., Koeppe, MJ., Free, SL., Myers, R., Lemieux, L., Mitchell, TN., Brooks, DJ., Duncan, JS. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human brain mapping*, 19(4), 224–47. doi:10.1002/hbm.10123
- Hempel, H., Bürger, K., Teipel, SJ., Bokde, ALW., Zetterberg, H., Blennow, K. (2008). Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's & Dementia*, 4(1), 38–48. doi:10.1016/j.jalz.2007.08.006
- Hempel, H., Frank, R., Broich, K., Teipel, SJ., Katz, RG., Hardy, J., Herholz, K., Bokde, ALW., Jessen, F., Hoessler, YC., Sanhai, WR., Zetterberg, H., Woodcock, J., Blennow, K. (2010). Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. *Nature Reviews Drug Discovery*, 9(7), 560–574. doi:10.1038/nrd3115
- Hardy, J., Selkoe, DJ. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 297(5580), 353–356. doi:10.1126/science.1072994
- Hébert, F., Grand'Maison, M., Ho, M.-K., Lerch, JP., Hamel, E., Bedell, BJ. (2013). Cortical atrophy and hypoperfusion in a transgenic mouse model of Alzheimer's disease. *Neurobiology of Aging*, 34(6), 1644–1652. doi:10.1016/j.neurobiolaging.2012.11.022
- Henkelman, R. (2010). Systems biology through mouse imaging centers: experience and new directions. *Annual review of biomedical engineering*. doi:10.1146/annurev-bioeng-070909
- Hernandez, D., Mee, PJ., Martin, JE., Tybulewicz, VLJ., Fisher, EMC. (1999). Transchromosomal mouse embryonic stem cell lines and chimeric mice that contain freely segregating segments of human chromosome 21. *Human molecular genetics*, 8(5), 923–33. doi:ddc099

- Herrup, K. (2015). The case for rejecting the amyloid cascade hypothesis. *Nat Neurosci*, 18(6), 794–799. doi:10.1038/nn.4017
- Hobbs, NZ., Henley, SMD., Wild, EJ., Leung, KK., Frost, C., Barker, RA., Scahill, RI., Barnes, J., Tabrizi, SJ., Fox, NC. (2009). Automated quantification of caudate atrophy by local registration of serial MRI: evaluation and application in Huntington's disease. *NeuroImage*, 47(4), 1659–65. doi:10.1016/j.neuroimage.2009.06.003
- Holland, D., McEvoy, LK., Dale, AM. (2012). Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI. *Human brain mapping*, 33(11), 2586–602. doi:10.1002/hbm.21386
- Holmes, CJ., Hoge, R., Collins, L., Woods, R., Toga, AW., Evans, AC. (1998). Enhancement of MR Images Using Registration for Signal Averaging. *Jrnl. Computer Assisted Tomography*. 22(2):324-333. ISSN:0363-8715
- Holmes, HE., Colgan, N., Ismail, O., Ma, D., Powell, NM., O'Callaghan, JM., Harrison, IF., Johnson, RA., Murray, TK., Ahmed, Z., Heggenes, M., Fisher, A., Cardoso, MJ., Modat, M., Walker-Samuel, S., Fisher, EMC., Ourselin, S., O'Neill, MJ., Wells, JA., Collins, EC., Lythgoe, MF. (2016). Imaging the accumulation and suppression of tau pathology using multi-parametric MRI. *Neurobiology of Aging*. doi:10.1016/j.neurobiolaging.2015.12.001
- Hua, X., Leow, AD., Lee, S., Klunder, AD., Toga, AW., Lepore, N., et al. (2008). 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. *NeuroImage*, 41(1), 19–34. doi:10.1016/j.neuroimage.2008.02.010
- Hua, X., Hibar, DP., Ching, CRK., Boyle, CP., Rajagopalan, P., Gutman, BA., Leow, AD., Toga, AW., Jack Jr, CR., Harvey, D., Weiner, MW., Thompson, PM. (2013). Unbiased tensor-based morphometry: improved robustness and sample size estimates for Alzheimer's disease clinical trials. *NeuroImage*, 66, 648–61. doi:10.1016/j.neuroimage.2012.10.086
- Iglesias, JE., Sabuncu, MR., Van Leemput, K. (2013). A unified framework for cross-modality multi-atlas segmentation of brain MRI. *Medical Image Analysis*, 17(8), 1181–1191. doi:10.1016/j.media.2013.08.001
- IHGSC (International Human Genome Sequencing Consortium), et al. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 50(2), 162–8. doi:10.1038/nature03001
- Insausti, AM., Megías, M., Crespo, D., Cruz-Orive, LM., Dierssen, M., Vallina, TF., Insausti, R., Flórez, J. (1998). Hippocampal volume and neuronal number in Ts65Dn mice: a murine model of Down syndrome. *Neuroscience Letters*, 253(3), 175–178. doi:10.1016/S0304-3940(98)00641-7
- Ittner, LM., Götz, J. (2011). Amyloid- $\beta$  and tau--a toxic pas de deux in Alzheimer's disease. *Nature reviews. Neuroscience*, 12(2), 65–72. doi:10.1038/nrn2967
- Jack Jr, CR., Knopman, DS., Jagust, WJ., Petersen, RC., Weiner, MW., Aisen, PS., Shaw, LM., Vemuri, P., Wiste, HJ., Weigand, SD., Lesnick, TG., Pankratz, VS., Donohue, MC., Trojanowski, JQ. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet neurology*, 12(2), 207–16. doi:10.1016/S1474-4422(12)70291-0
- Janke, AL., Ullmann, JFP. (2015). Robust methods to create ex vivo minimum deformation atlases for brain mapping. *Methods*, 73, 18–26. doi:10.1016/j.ymeth.2015.01.005
- Janus, C. (2008). Conditionally inducible tau mice--designing a better mouse model of neurodegenerative diseases. *Genes, Brain and Behavior*, 7 Suppl 1, 12–27. doi:10.1111/j.1601-183X.2007.00375.x
- Jenkinson, M., Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156. doi:10.1016/S1361-8415(01)00036-6
- Jessen, F., Wolfgruber, S., Wiese, B., Bickel, H., Mösch, E., Kaduszkiewicz, H., Wagner, M., et al. (2014). AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimer's & Dementia*, 10(1), 76–83. doi:10.1016/j.jalz.2012.09.017
- Johnson, GA., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., Nissanov, J. (2010). Waxholm space: an image-based reference for coordinating mouse brain research. *NeuroImage*, 53(2), 365–72. doi:10.1016/j.neuroimage.2010.06.067
- Jucker, M. (2010). The benefits and limitations of animal models for translational research in neurodegenerative diseases. *Nature Medicine*, 16(11), 1210–1214. doi:10.1038/nm.2224

- Kainz, B., Keraudren, K., Kyriakopoulou, V., Rutherford, M., Hajnal, JV., Rueckert, D. (2014). Fast fully automatic brain detection in fetal MRI using dense rotation invariant image descriptors. In IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014 (pp. 1230-1233).
- Kale, SC., Lerch, JP., Henkelman, RM., Chen, XJ. (2008). Optimization of the SNR-resolution tradeoff for registration of magnetic resonance images. *Hum. Brain Mapp.*, 29, 1147–1158. doi:10.1002/hbm.20453
- Karas, GB., Burton, EJ., Rombouts, SARB., van Schijndel, RA., O'Brien, JT., Scheltens, Ph., McKeith, IG., Williams, D., Ballard, C., Barkhof, F. (2003). A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. *NeuroImage*, 18(4), 895–907. doi:10.1016/S1053-8119(03)00041-7
- Keifer Jr, OP., Hurt, RC., Gutman, DA., Keilholz, SD., Gourley, SL., Ressler, KJ. (2015). Voxel-based morphometry predicts shifts in dendritic spine density and morphology with auditory fear conditioning. *Nature Communications*, 6(May), 7582. doi:10.1038/ncomms8582
- Keller, SS., Wilke, M., Wiesmann, UC., Sluming, VA., Roberts, N. (2004). Comparison of standard and optimized voxel-based morphometry for analysis of brain changes associated with temporal lobe epilepsy. *NeuroImage*, 23(3), 860–8. doi:10.1016/j.neuroimage.2004.07.030
- Kiebel, SJ., Holmes, AP. (2006). The General Linear Model. In: *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, ed. Friston, KJ., Ashburner, JT., Kiebel, SJ., Nichols, TE., Penny, William D. ISBN-13:978-0-12-372560-8
- Kielar, C., Sawiak, SJ., Negredo, PN., Tse, DHY., Morton, AJ. (2012). Tensor-based morphometry and stereology reveal brain pathology in the complexin1 knockout mouse. *PloS one*, 7(2), e32636. doi:10.1371/journal.pone.0032636
- Kim, J., Fessler, JA. (2004). Intensity-based image registration using robust correlation coefficients. *IEEE transactions on medical imaging*, 23(11), 1430–44. doi:10.1109/TMI.2004.835313
- Kipps, CM., Duggins, AJ., Mahant, N., Gomes, L., Ashburner, J., McCusker, EA. (2005). Progression of structural neuropathology in preclinical Huntington's disease: a tensor based morphometry study. *Journal of neurology, neurosurgery, and psychiatry*, 76, 650–655. doi:10.1136/jnnp.2004.047993
- Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A. (2009). Evaluation of automated brain MR image segmentation and volumetry methods. *Hum Brain Mapp*, 30(4), 1310–1327. doi:10.1002/hbm.20599
- Kohannim, O., Hua, X., Hibar, DP., Lee, S., Chou, Y-Y., Toga, AW., Jack, CR., Weiner, MW., Thompson, PM. (2010). Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiology of aging*, 31(8), 1429–42. doi:10.1016/j.neurobiolaging.2010.04.022
- Kovačević, N., Chen, J., Sled, J., Henderson, J., Henkelman, M. (2004). Deformation based representation of groupwise average and variability. In: *Medical Image Computing and Computer-Assisted Intervention – (MICCAI) 2004*. Vol.3216, Lecture Notes in Computer Science, pp615-622. doi:10.1007/978-3-540-30135-6\_75
- Kovačević, N., Henderson, JT., Chan, E., Lifshitz, N., Bishop, J., Evans, AC., Henkelman, RM., Chen, XJ. (2005). A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability. *Cerebral cortex*, 15(5), 639–45. doi:10.1093/cercor/bhh165
- Kurt, AM., Kafa, MI., Dierssen, M., Davies, DC. (2004). Deficits of neuronal density in CA1 and synaptic density in the dentate gyrus, CA3 and CA1, in a mouse model of Down syndrome. *Brain Research*, 1022(1-2), 101–109. doi:10.1016/j.brainres.2004.06.075
- Lang, K., Huang, H., Lee, DW., Federico, V., Menzin, J. (2013). National trends in advanced outpatient diagnostic imaging utilization: an analysis of the medical expenditure panel survey, 2000-2009. *BMC medical imaging*, 13, 40. doi:10.1186/1471-2342-13-40
- Lau, JC., Lerch, JP., Sled, JG., Henkelman, RM., Evans, AC., Bedell, BJ. (2008). Longitudinal neuroanatomical changes determined by deformation-based morphometry in a mouse model of Alzheimer's disease. *NeuroImage* 42 19–27. doi:10.1016/j.neuroimage.2008.04.252
- Lebenberg, J., Hérard, A-S., Dubois, A., Dhenain, M., Hantraye, P., Delzescaux, T. (2011). A combination of atlas-based and voxel-wise approaches to analyze metabolic changes in autoradiographic data from Alzheimer's mice. *NeuroImage*, 57(4), 1447–57. doi:10.1016/j.neuroimage.2011.04.059



- Ledig, C., Wolz, R., Aljabar, P., Jyrki, L., Rueckert, D. (2012). PBSI: A symmetric probabilistic extension of the Boundary Shift Integral. *Medical Image Computing and Computer-Assisted Intervention–MICCAI Workshop on Novel Imaging Biomarkers for Alzheimer’s Disease and Related Disorders*, 117–124.
- Lee, VM., Goedert, M., Trojanowski, JQ. (2001). Neurodegenerative tauopathies. *Annual review of neuroscience*, 24, 1121–59. doi:10.1146/annurev.neuro.24.1.1121
- Lee, J., Jomier, J., Aylward, S., Tyska, M., Moy, S., Lauder, J., Styner, M. (2009). Evaluation of atlas based mouse brain segmentation. *Proc. SPIE, Medical Imaging: Image Processing*, 725943–725949. doi:10.1117/12.812762
- Leow, AD., Yanovsky, I., Chiang, M-C., Lee, AD., Klunder, AD., Lu, A., Becker, JT., Davis, SW., Toga, AW., Thompson, PM. (2007). Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration. *IEEE transactions on medical imaging*, 26(6), 822–32. doi:10.1109/TMI.2007.892646
- Lerch, JP., Carroll, JB., Spring, S., Bertram, LN., Schwab, C., Hayden, MR., Henkelman, RM. (2008). Automated deformation analysis in the YAC128 Huntington disease mouse model. *NeuroImage*, 39(1), 32–9. doi:10.1016/j.neuroimage.2007.08.033
- Lerch, JP. (2010). Tips for Advanced MRI Screening of Mice. *Proc. ISMRM* 18.
- Lerch, JP., Yiu, AP., Martinez-Canabal, A., Pekar, T., Bohbot, VD., Frankland, PW., Henkelman, RM., Josselyn, SA., Sled, JG. (2011). Maze training in mice induces MRI-detectable brain shape changes specific to the type of learning. *NeuroImage*, 54(3), 2086–95. doi:10.1016/j.neuroimage.2010.09.086
- Lerch, JP., Gazdzinski, L., Germann, J., Sled, JG., Henkelman, RM., Nieman, BJ. (2012). Wanted dead or alive? The tradeoff between in-vivo versus ex-vivo MR brain imaging in the mouse. *Frontiers in neuroinformatics*, 6(March), 6. doi:10.3389/fninf.2012.00006
- Leung, KK., Barnes, J., Ridgway, GR., Bartlett, JW., Clarkson, MJ., Macdonald, K., Schuff, N., Fox, NC., Ourselin, S. (2010a). Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s disease. *NeuroImage*, 51(4), 1345–59. doi:10.1016/j.neuroimage.2010.03.018
- Leung, KK., Clarkson, MJ., Bartlett, JW., Clegg, S., Jack Jr, CR., Weiner, MW., Fox, NC., Ourselin, S. (2010b). Robust atrophy rate measurement in Alzheimer’s disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *NeuroImage*, 50(2), 516–23. doi:10.1016/j.neuroimage.2009.12.059
- Leung, KK., Barnes, J., Modat, M., Ridgway, GR., Bartlett, JW., Fox, NC., Ourselin, S. (2011). Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. *NeuroImage*, 55(3), 1091–108. doi:10.1016/j.neuroimage.2010.12.067
- Leung, KK., Ridgway, GR., Ourselin, S., Fox, NC. (2012). Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *NeuroImage*, 59(4), 3995–4005. doi:10.1016/j.neuroimage.2011.10.068
- Lewis, EB., Fox, NC. (2004). Correction of differential intensity inhomogeneity in longitudinal MR images. *NeuroImage*, 23(1), 75–83. doi:10.1016/j.neuroimage.2004.04.030
- Li, Q., Cheung, C., Wei, R., Hui, E., Hui, ES., Feldon, J., Meyer, U., Chung, S., Chua, SE., Sham, PC., Wu, EX., McAlonan, GM. (2009). Prenatal immune challenge is an environmental risk factor for brain and behavior change relevant to schizophrenia: evidence from MRI in a mouse model. *PLOS ONE*, 4(7), e6354. doi:10.1371/journal.pone.0006354
- Liu, Y., Collins, R., Rothfus, W. (1998). Automatic bilateral symmetry (midsagittal) plane extraction from pathological 3D neuroradiological images. *Proc. SPIE 3338, Medical Imaging 1998: Image Processing*, 1528–1539. doi:10.1117/12.310886
- Lott, IT., Head, E. (2001). Down syndrome and Alzheimer’s disease: A link between development and aging. *Mental Retardation and Developmental Disabilities Research Reviews*, 7, 172–178. doi:10.1002/mrdd.1025
- Ma, Y., Hof, PR., Grant, SC., Blackband, SJ., Bennett, R., Slatest, L., McGuigan, MD., Benveniste, H. (2005). A three-dimensional digital atlas database of the adult C57Bl/6J mouse brain by magnetic resonance microscopy. *Neuroscience*, 135(4), 1203–15. doi:10.1016/j.neuroscience.2005.07.014

- Ma, Y., Smith, D., Hof, P.R., Foerster, B., Hamilton, S., Blackband, S.J., Yu, M., Benveniste, H. (2008). In vivo 3D digital atlas database of the adult C57Bl/6J mouse brain by magnetic resonance microscopy. *Front Neuroanat.* 17;2:1. doi:10.3389/neuro.05.001.2008
- Ma, D., Cardoso, M.J., Modat, M., Powell, N., Wells, J., Holmes, H., Wiseman, F., Tybulewicz, V., Fisher, E., Lythgoe, M.F., Ourselin, S. (2014). Automatic structural parcellation of mouse brain MRI using multi-atlas label fusion. *PLOS ONE*, 9(1), e86576. doi:10.1371/journal.pone.0086576
- Ma, D., Cardoso, M.J., Zuluaga, M.A., Modat, M., Powell, N., Wiseman, F., Tybulewicz, V., Fisher, E., Lythgoe, M.F., Ourselin, S. (2015). Grey matter sublayer thickness estimation in the mouse cerebellum. *Proc. Medical Image Computing and Computer-Assisted Intervention–MICCAI* (648).
- MacKenzie-Graham, A., Lee, E.-F., Dinov, I.D., Bota, M., Shattuck, D.W., Ruffins, S., Toga, A.W., et al. (2004). A multimodal, multidimensional atlas of the C57BL/6J mouse brain. *Journal of anatomy*, 204(2), 93–102. doi:10.1111/j.1469-7580.2004.00264.x
- MacKenzie-Graham, A., Tinsley, M.R., Shah, K.P., Aguilar, C., Strickland, L.V., Boline, J., Martin, M., Morales, L., Shattuck, D.W., Jacobs, R.E., Voskuhl, R.R., Toga, A.W. (2006). Cerebellar cortical atrophy in experimental autoimmune encephalomyelitis. *NeuroImage*, 32(3), 1016–1023. doi:10.1016/j.neuroimage.2006.05.006
- Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S.J., Frith, C.D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences USA*, 97(8), 4398–403. doi:10.1073/pnas.070039597
- Maheswaran, S., Barjat, H., Bate, S.T., Aljabar, P., Hill, D.L.G., Tilling, L., Upton, N., James, M.F., Hajnal, J.V., Rueckert, D. (2009a). Analysis of serial magnetic resonance images of mouse brains using image registration. *NeuroImage*, 44(3), 692–700. doi:10.1016/j.neuroimage.2008.10.016
- Maheswaran, S., Barjat, H., Rueckert, D., Bate, S.T., Howlett, D.R., Tilling, L., James, M.F., et al. (2009b). Longitudinal regional brain volume changes quantified in normal aging and Alzheimer's APP x PS1 mice using MRI. *Brain research*, 1270, 19–32. doi:10.1016/j.brainres.2009.02.045
- Malone, I.B., Leung, K.K., Clegg, S., Barnes, J., Whitwell, J.L., Ashburner, J., Fox, N.C., Ridgway, G.R. (2015). Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. *NeuroImage*, 104, 366–372. doi:10.1016/j.neuroimage.2014.09.034
- Mangialasche, F., Solomon, A., Winblad, B., Mecocci, P., Kivipelto, M. (2010). Alzheimer's disease: clinical trials and drug development. *Lancet neurology*, 9(7), 702–16. doi:10.1016/S1474-4422(10)70119-8
- Marion J.B., Thornton, S.T. (2004). *Classical Dynamics of Particles and Systems*, 5th ed. Brooks/Cole, pp.415–418. ISBN:978-0534408961.
- McConville, P., Moody, J.B., Moffat, B.A. (2005). High-throughput magnetic resonance imaging in mice for phenotyping and therapeutic evaluation. *Current Opinion in Chemical Biology*, 9(4), 413–20. doi:10.1016/j.cbpa.2005.06.004
- McGavern, D.B., Kang, S.S. (2011). Illuminating viral infections in the nervous system. *Nature reviews. Immunology*, 11(5), 318–29. doi:10.1038/nri2971
- McGowan, E., Eriksen, J., Hutton, M. (2006). A decade of modeling Alzheimer's disease in transgenic mice. *Trends in Genetics*, 22(5), 281–289. doi:10.1016/j.tig.2006.03.007
- McRobbie, D.W., Moore, E.A., Graves, M.J., Prince, M.R. (2007). *MRI from Picture to Proton*. 2<sup>nd</sup> Ed. Cambridge University Press. ISBN:9780521683845
- Mechelli, A., Price, C.J., Friston, K.J., Ashburner, J. (2005). Voxel-based morphometry of the human brain: methods and applications. *Current Medical*, 1–9. doi:10.2174/1573405054038726
- Mercer, R.E., Kwolek, E.M., Bischof, J.M., van Eede, M., Henkelman, R.M., Wevrick, R. (2009). Regionally reduced brain volume, altered serotonin neurochemistry, and abnormal behavior in mice null for the circadian rhythm output gene *Magel2*. *American Journal of Medical Genetics, B: Neuropsychiatric Genetics*, 150(8), 1085–1099. doi:10.1002/ajmg.b.30934
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S. (2010). Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3), 278–84. doi:10.1016/j.cmpb.2009.09.002

- Modat, M., Cardoso, M.J., Daga, P., Cash, D., Fox, N.C., Ourselin, S. (2012). Inverse-consistent symmetric free form deformation. *Biomedical Image Registration: Lecture Notes in Computer Science Volume 7359*, pp79–88. doi:10.1007/978-3-642-31340-0\_9
- Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S. (2014). Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging*, 1(2), 024003. doi:10.1117/1.JMI.1.2.024003
- Modersitzki, J. (2004). *Numerical Methods for Image Registration*. Oxford University Press. ISBN:9780198528418
- Moore, C.S., Roper, R.J. (2007). The power of comparative and developmental studies for mouse models of Down syndrome. *Mammalian genome*, 18(6-7), 431–43. doi:10.1007/s00335-007-9030-8
- Morice, E., Andreae, L.C., Cooke, S.F., Vanes, L., Fisher, E.M.C., Tybulewicz, V.L.J., Bliss, T.V.P. (2008). Preservation of long-term memory and synaptic plasticity despite short-term impairments in the Tc1 mouse model of Down syndrome. *Learning & Memory (Cold Spring Harbor, NY.)*, 15(7), 492–500. doi:10.1101/lm.969608
- Morris, J., Alberman, E. (2009). Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register. *British Medical Journal*, 1–5. 339:b3794, doi:10.1136/bmj.b3794
- Mucke, L., Masliah, E., Yu, G.Q., Mallory, M., Rockenstein, E.M., Tatsuno, G., Hu, K., Kholodenko, D., Johnson-Wood, K., McConlogue, L. (2000). High-level neuronal expression of abeta 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *The Journal of Neuroscience*, 20(11), 4050–8.
- Natt, O., Watanabe, T., Boretius, S., Radulovic, J., Frahm, J., Michaelis, T. (2002). High-resolution 3D MRI of mouse brain reveals small cerebral structures in vivo. *Journal of Neuroscience Methods*, 120(2), 203–9. doi:10.1016/S0165-0270(02)00211-X
- Nichols, T.E., Holmes, A.P. (2001). Nonparametric Permutation Tests For Functional Neuroimaging: A Primer with Examples, *Human Brain Mapping* 15:1–25. doi:10.1002/hbm.1058
- Nichols, T., Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*, 12(5), 419–46. doi:10.1191/0962280203sm341ra
- Nie, J., Shen, D. (2013). Automated segmentation of mouse brain images using multi-atlas multi-ROI deformation and label fusion. *Neuroinformatics*, 11(1), 35–45. doi:10.1007/s12021-012-9163-0
- Nieman, B.J., Bock, N.A., Bishop, J., Chen, X.J., Sled, J.G., Rossant, J., Henkelman, R.M. (2005). Magnetic resonance imaging for detection and analysis of mouse phenotypes. *NMR in biomedicine*, 18(7), 447–68. doi:10.1002/nbm.981
- Nieman, B.J., Flenniken, A.M., Adamson, S.L., Henkelman, R.M., Sled, J.G. (2006). Anatomical phenotyping in the brain and skull of a mutant mouse by magnetic resonance imaging and computed tomography. *Physiological genomics*, 24(2), 154–62. doi:10.1152/physiolgenomics.00217.2005
- Nieman, B., Lerch, J., Bock, N., Chen, X.J., Sled, J.G., Henkelman, R.M. (2007). Mouse behavioral mutants have neuroimaging abnormalities. *Human brain mapping*, 28(6), 567–75. doi:10.1002/hbm.20408
- Nieman, B.J., Wong, M.D., Henkelman, R.M. (2011). Genes into geometry: imaging for mouse development in 3D. *Current opinion in genetics & development*, 21(5), 638–646. doi:10.1016/j.gde.2011.08.009
- Norris, F.C., Wong, M.D., Greene, N.D.E., Scambler, P.J., Weaver, T., Weninger, W.J., Mohun, T.J., Henkelman, R.M., Lythgoe, M.F. (2013). A coming of age: advanced imaging technologies for characterising the developing mouse. *Trends in Genetics*, 29(12), 700–11. doi:10.1016/j.tig.2013.08.004
- Novak, G., Fox, N., Clegg, S., Nielsen, C., Einstein, S., Lu, Y., Tudor, I.C., Gregg, K., Di, J., Collins, P., Wyman, B.T., Yuen, E., Grundman, M., Brashear, H.R., Liu, E. (2016). Changes in Brain Volume with Bapineuzumab in Mild to Moderate Alzheimer's Disease. *Journal of Alzheimer's Disease*, 49(45), 709–719. doi:10.3233/JFS-120647
- Nyúl, L.G., Udupa, J.K. (1999). On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*, 42(6), 1072–81. doi:10.1002/(SICI)1522-2594(199912)42:6<1072::AID-MRM11>3.0.CO;2-M
- Nyúl, L.G., Udupa, J.K., Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2), 143–150. doi:10.1109/42.836373

- O'Callaghan, J., Wells, J., Richardson, S., Holmes, H., Yu, Y., Walker-Samuel, S., Siow, B., Lythgoe, MF. (2014). Is your system calibrated? MRI gradient system calibration for pre-clinical, high-resolution imaging. *PLOS ONE*, 9(5), e96568. doi:10.1371/journal.pone.0096568
- O'Doherty, A., Ruf, S., Mulligan, C., Hildreth, V., Errington, ML., Cooke, S., Sesay, A., Modino, S., Vanes, L., Hernandez, D., Linehan, JM., Sharpe, PT., Brandner, S., Bliss, TVP., Henderson, DJ., Nizetic, D., Tybulewicz, VLJ., Fisher, EMC. (2005). An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science*, 309(5743), 2033–7. doi:10.1126/science.1114535
- Oguz, I., Lee, J., Budin, F., Rumple, A. (2011). Automatic skull-stripping of rat MRI/DTI scans and atlas building. *Proc Soc Photo Opt Instrum Eng*. doi:10.1117/12.878405.Automatic
- Oguz, I., Zhang, H., Rumple, A., Sonka, M. (2014). RATS: Rapid Automatic Tissue Segmentation in rodent brain MRI. *Journal of Neuroscience Methods*, 221, 175–182. doi:10.1016/j.jneumeth.2013.09.021
- Oliver, C., Holland, AJ. (1986). Down's syndrome and Alzheimer's disease: a review. *Psychological Medicine*, 16(02), 307–322. doi:10.1017/S0033291700009120
- Olson, LE., Roper, RJ., Baxter, LL., Carlson, EJ., Epstein, CJ., Reeves, RH. (2004). Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes. *Developmental dynamics*, 230(3), 581–9. doi:10.1002/dvdy.20079
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.*, 9, pp.62–66.
- Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N. (2001). Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing*, 19, 25–31. doi:10.1016/S0262-8856(00)00052-4
- Paranjape, RB., (2009). Fundamental Enhancement Techniques. In: Bankman, IN., *Handbook of Medical Image Processing and Analysis*. Academic Press. p10. ISBN:978-0-12-373904-9.
- Parker, SE., Mai, CT., Canfield, MA., Rickard, R., Wang, Y., Meyer, RE., Anderson, P., Mason, CA., Collins, JS., Kirby, RS., Correa, A. (2010). Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004–2006. *Birth defects research. Part A, Clinical and molecular teratology*, 88(12), 1008–16. doi:10.1002/bdra.20735
- Paxinos, G., Franklin, K. (2012). *The Mouse Brain in Stereotaxic Coordinates*, 4<sup>th</sup> Ed. Academic Press, ISBN:9780123910578.
- Pinter, JD., Eliez, S., Schmitt, JE., Capone, GT., Reiss, AL. (2001a). Neuroanatomy of Down's syndrome: a high-resolution MRI study. *Am J Psychiatry*, 158(10), 1659–1665. doi:10.1176/appi.ajp.158.10.1659
- Pinter, JD., Brown, WE., Eliez, S., Schmitt, JE., Capone, GT., Reiss, AL. (2001b). Amygdala and hippocampal volumes in children with Down syndrome: a high-resolution MRI study. *Neurology*, 56(7), 972–974. doi:10.1212/WNL.56.7.972
- Piras, F., Piras, F., Chiapponi, C., Girardi, P., Caltagirone, C., Spalletta, G. (2015). Widespread structural brain changes in OCD: A systematic review of voxel-based morphometry studies. *Cortex*, 62, 89–108. doi:10.1016/j.cortex.2013.01.016
- Prados, F., Cardoso, MJ., Leung, KK., Cash, DM., Modat, M., Fox, NC., Wheeler-Kingshott, CAM., Ourselin, S. (2014). Measuring brain atrophy with a generalised formulation of the boundary shift integral. *Neurobiology of Aging*, 1–10. doi:10.1016/j.neurobiolaging.2014.04.035
- Prastawa, M., Gilmore, JH., Lin, W., Gerig, G. (2005). Automatic segmentation of MR images of the developing newborn brain. *Medical image analysis*, 9(5), 457–66. doi:10.1016/j.media.2005.05.007
- Preboske, GM., Gunter, JL., Ward, CP., Jack, CR. (2006). Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. *NeuroImage*, 30(4), 1196–202. doi:10.1016/j.neuroimage.2005.10.049
- Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., Ferri, CP. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's & Dementia*, 9(1), 63–75.e2. doi:10.1016/j.jalz.2012.11.007
- Prince, M., Knapp, M., Guerchet, M., McCrone, P. Alzheimer's Society, et al. (2014). *Dementia UK*, 2nd edition. ISBN:978-1-906647-31-5

- Ramsden, M., Kotilinek, L., Forster, C., Paulson, J., McGowan, E., SantaCruz, K., Guimaraes, A., Yue, M., Lewis, J., Carlson, G., Hutton, M., Ashe, KH. (2005). Age-dependent neurofibrillary tangle formation, neuron loss, and memory impairment in a mouse model of human tauopathy (P301L). *Journal of neuroscience*, 25(46), 10637–47. doi:10.1523/JNEUROSCI.3279-05.2005
- Raz, N., Torres, IJ., Briggs, SD., Spencer, WD., Thornton, AE., Loken, WJ., Gunning, FM., McQuain, JD., Driesen, NR., Acker, JD. (1995). Selective neuroanatomic abnormalities in Down's syndrome and their cognitive correlates: evidence from MRI morphometry. *Neurology*, 45(2), 356–366. doi:10.1212/WNL.45.2.356
- Reagan-Shaw, S., Nihal, M., Ahmad, N. (2008). Dose translation from animal to human studies revisited. *FASEB journal: pub. of the Fed. of Amer. Soc. Exp. Biology*, 22(3), 659–61. doi:10.1096/fj.07-9574LSF
- Redwine, JM., Kosofsky, B., Jacobs, RE., Games, D., Reilly, JF., Morrison, JH., Young, WG., Bloom, FE. (2003). Dentate gyrus volume is reduced before onset of plaque formation in PDAPP mice: a magnetic resonance microscopy and stereologic analysis. *Proc. Nat. Academy of Sciences USA*, 100(3), 1381–6. doi:10.1073/pnas.242746599
- Reeves, RH. (2006). Down syndrome mouse models are looking up. *Trends in molecular medicine*, 12(6), 237–40. doi:10.1016/j.molmed.2006.04.001
- Reitz, C., Mayeux, R. (2014). Alzheimer disease: Epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical Pharmacology*, 88(4), 640–651. doi:10.1016/j.bcp.2013.12.024
- Rengachary, SS., Ellenbogen, RG. (Eds.). (2005). *Principles of neurosurgery*. Elsevier Mosby. ISBN:0723432228
- Riegler, J., Cheung, KK., Man, YF., Cleary, JO., Price, AN., Lythgoe, MF. (2010). Comparison of segmentation methods for MRI measurement of cardiac function in rats. *Journal of MRI*, 32, 869–877. doi:10.1002/jmri.22305
- Ridgway, G., Barnes, J., Pepple, T., Fox, N. (2011). Estimation of total intracranial volume; a comparison of methods. *Alzheimer's & Dementia*, 7(4), S62–S63. doi:10.1016/j.jalz.2011.05.099
- Rodgers, JL., Nicewander, WA. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1), 59–66. doi:10.1080/00031305.1988.10475524
- Rogers, DC., Jones, DNC., Nelson, PR., Jones, CM., Quilter, CA., Robinson, TL., Hagan, JJ. (1999). Use of SHIRPA and discriminant analysis to characterise marked differences in the behavioural phenotype of six inbred mouse strains. *Behavioural Brain Research*, 105(2), 207–217. doi:10.1016/S0166-4328(99)00072-8
- Rohlfing, T., Brandt, R., Menzel, R., Maurer, CR. (2004). Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4), 1428–1442. doi:10.1016/j.neuroimage.2003.11.010
- Roizen, NJ., Patterson, D. (2003). Down's syndrome. *The Lancet*, 361(9365), 1281–1289. doi:10.1016/S0140-6736(03)12987-X
- Rosenthal, N., Brown, S. (2007). The mouse ascending: perspectives for human-disease models. *Nature cell biology*, 9(9), 993–999. doi:10.1038/ncb437
- Rowland, DJ., Garbow, JR., Laforest, R., Snyder, AZ. (2005). Registration of [18F]FDG microPET and small-animal MRI. *Nuclear Medicine and Biology*, 32(6), 567–572. doi:10.1016/j.nucmedbio.2005.05.002
- Rudin, M. (2007). Imaging readouts as biomarkers or surrogate parameters for the assessment of therapeutic interventions. *European Radiology*, 17(10), 2441–2457. doi:10.1007/s00330-007-0619-9
- Rueckert, D., Sonoda, LI., Hayes, C., Hill, DLG., Leach, MO., Hawkes, DJ. (1999). Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging, IEEE Transactions on*, 18(8), 712–721. doi:10.1109/42.796284
- Ruppert, GCS., Teverovskiy, L., Yu, C-P., Falcão, AX., Liu, Y. (2011). A new symmetry-based method for mid-sagittal plane extraction in neuroimages. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp.285–288. doi:10.1109/ISBI.2011.5872407
- Salomone, S., Caraci, F., Leggio, GM., Fedotova, J., Drago, F. (2011). New pharmacological strategies for treatment of Alzheimer's disease: focus on disease modifying drugs. *British Journal of Clinical Pharmacology*, 73(4), 504–517. doi:10.1111/j.1365-2125.2011.04134.x

- Sansare, K., Khanna, V., Karjodkar, F. (2011). Early victims of X-rays: A tribute and current perception. *Dentomaxillofacial Radiology*, 40(2), 123–125. doi:10.1259/dmfr/73488299
- SantaCruz, K., Lewis, J., Spires, T., Paulson, J., Kotilinek, L., Ingelsson, M., Guimaraes, A., DeTure, M., Ramsden, M., McGowan, E., Forster, C., Yue, M., Orne, J., Janus, C., Mariash, A., Kuskowski, M., Hyman, B., Hutton, M., Ashe, KH. (2005). Tau suppression in a neurodegenerative mouse model improves memory function. *Science*, 309(5733), 476–81. doi:10.1126/science.1113694
- Sawiak, SJ., Wood, NI., Williams, GB., Morton, AJ., Carpenter, TA. (2009a). SPMMouse: A new toolbox for SPM in the animal brain. In *Proc. ISMRM 17th Scientific Meeting & Exhibition*.
- Sawiak, SJ., Wood, NI., Williams, GB., Morton, AJ., Carpenter, TA. (2009b). Use of magnetic resonance imaging for anatomical phenotyping of the R6/2 mouse model of Huntington's disease. *Neurobiology of disease*, 33(1), 12–9. doi:10.1016/j.nbd.2008.09.017
- Sawiak, SJ., Wood, NI., Williams, GB., Morton, AJ., Carpenter, TA. (2009c). Voxel-based morphometry in the R6/2 transgenic mouse reveals differences between genotypes not seen with manual 2D morphometry. *Neurobiology of disease*, 33(1), 20–7. doi:10.1016/j.nbd.2008.09.016
- Sawiak, SJ., Wood, NI., Carpenter, TA., Morton, AJ. (2012). Huntington's Disease mouse models online: high-resolution MRI images with stereotaxic templates for computational neuroanatomy. *PLOS ONE*, 7(12), e53361. doi:10.1371/journal.pone.0053361
- Sawiak, SJ., Wood, NI., Williams, GB., Morton, AJ., Carpenter, TA. (2013). Voxel-based morphometry with templates and validation in a mouse model of Huntington's disease. *Magnetic resonance imaging*, 31(9), 1522–1531. doi:10.1016/j.mri.2013.06.001
- Scahill, RI., Frost, C., Jenkins, R., Whitwell, JL., Rossor, MN., Fox, NC. (2003). A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Archives of neurology*, 60(7), 989–994. doi:10.1001/archneur.60.7.989
- Scheenstra, AEH., Van de Ven, RCG., Weerd, L., Maagdenberg, AMJM., Dijkstra, J., Reiber, JHC. (2009). Automated Segmentation of In Vivo and Ex Vivo Mouse Brain Magnetic Resonance Images. *Methods*, 8(1), 35–44. doi:10.2310/7290.2009.00004
- Scheenstra, AEH. (2011). Automated morphometry of transgenic mouse brains in MR images. (Thesis.) Retrieved from <https://openaccess.leidenuniv.nl/handle/1887/16649> ISBN-13:978-90-9026004-4
- Schenk, D., Barbour, R., Dunn, W., Gordon, G., Grajeda, H., Guido, T., Hu, K., Huang, J., Johnson-Wood, K., Khan, K., Seubert, P., et al. (1999). Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature*, 400(6740), 173–177. doi:10.1038/22124
- Schott, JM., Price, SL., Frost, C., Whitwell, JL., Rossor, MN., Fox, NC. (2005). Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. *Neurology*, 65(1), 119–24. doi:10.1212/01.wnl.0000167542.89697.0f
- Schott, JM., Bartlett, JW., Barnes, J., Leung, KK., Ourselin, S., Fox, NC. (2010). Reduced sample sizes for atrophy outcomes in Alzheimer's disease trials: baseline adjustment. *Neurobiology of aging*, 31(8), 1452–62, 1462.e1–2. doi:10.1016/j.neurobiolaging.2010.04.011
- Sebrié, C., Chabert, C., Ledru, A., Guedj, F., Po, C., Smith, DJ., Delabar, J-M., et al. (2008). Increased dosage of DYRK1A and brain volumetric alterations in a YAC model of partial trisomy 21. *Anatomical record (Hoboken, NJ:2007)*, 291(3), 254–62. doi:10.1002/ar.20640
- Serra, L., Cercignani, M., Lenzi, D., Perri, R., Fadda, L., Caltagirone, C., Macaluso, E., Bozzali, M. (2010). Grey and white matter changes at different stages of Alzheimer's disease. *Journal of Alzheimer's disease*, 19(1), 147–59. doi:10.3233/JAD-2010-1223
- Sheppard, O., Plattner, F., Rubin, A., Slender, A., Linehan, JM., Brandner, S., Tybulewicz, VLJ., Fisher, EMC., Wiseman, FK. (2012). Altered regulation of tau phosphorylation in a mouse model of down syndrome aging. *Neurobiology of Aging*, 33(4), 828.e31–44. doi:10.1016/j.neurobiolaging.2011.06.025
- Sinclair, B., Cleary, JO., Modat, M., Norris, F., Wiseman, F., Tybulewicz, V., Fisher, E., Lythgoe, M., Ourselin, S. (2011). Tensor-based morphometry of the Tc1 mouse model of Down syndrome highlights previously undetected phenotypes. *Proc. Intl. Soc. Mag.*, 309(1), 2037.

- Sled, J., Pike, G. (1998). Understanding intensity non-uniformity in MRI. *Medical Image Computing and Computer-Assisted Intervention*. Vol 1496 Lecture Notes in Computer Science, pp 614–622. doi:10.1007/BFb0056247
- Sled, JG., Zijdenbos, AP., Evans, AC. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1), 87–97. doi:10.1109/42.668698
- Small, SA., Duff, K. (2008). Perspective Linking A $\beta$  and Tau in Late-Onset Alzheimer's Disease: A Dual Pathway Hypothesis. *Neuron*, 60(4), 534–542. doi:10.1016/j.neuron.2008.11.007
- Smith, SM. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3), 143–55. doi:10.1002/hbm.10062
- Smith, SM., Zhang, Y., Jenkinson, M., Chen, J., Matthews, PM., Federico, A., de Stefano, N. (2002). Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis. *NeuroImage*, 17(1), 479–489. doi:10.1006/nimg.2002.1040
- Smith, SM., Rao, A., De Stefano, N., Jenkinson, M., Schott, JM., Matthews, PM., Fox, NC. (2007). Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: cross-validation of BSI, SIENA and SIENAX. *NeuroImage*, 36(4), 1200–6. doi:10.1016/j.neuroimage.2007.04.035
- Smith, SM., Nichols, TE. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83–98. doi:10.1016/j.neuroimage.2008.03.061
- Sperling, RA., Aisen, PS., Beckett, LA., Bennett, DA., Craft, S., Fagan, AM., Phelps, CH., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 280–292. doi:10.1016/j.jalz.2011.03.003
- Spillantini, MG., Goedert, M. (2013). Tau pathology and neurodegeneration. *Lancet neurology*, 12(6), 609–622. doi:10.1016/S1474-4422(13)70090-5
- Spring, S., Lerch, JP., Henkelman, RM. (2007). Sexual dimorphism revealed in the structure of the mouse brain using three-dimensional magnetic resonance imaging. *NeuroImage*, 35(4), 1424–33. doi:10.1016/j.neuroimage.2007.02.023
- Spring, S., Lerch, JP., Wetzel, MK., Evans, AC., Henkelman, RM. (2010). Cerebral asymmetries in 12-week-old C57Bl/6J mice measured by magnetic resonance imaging. *NeuroImage*, 50(2), 409–415. doi:10.1016/j.neuroimage.2009.12.043
- Stoodley, CJ., Schmahmann, JD. (2010). Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex*, 46(7), 831–844. doi:10.1016/j.cortex.2009.11.008
- Studholme, C., Hill, DL. G., Hawkes, DJ. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1), 71–86. doi:10.1016/S0031-3203(98)00091-0
- Sussman, D., Ellegood, J., Henkelman, M. (2013). A gestational ketogenic diet alters maternal metabolic status as well as offspring physiological growth and brain structure in the neonatal mouse. *BMC Pregnancy and Childbirth*, 13(1), 1–10. doi:10.1186/1471-2393-13-198
- Tabrizi, SJ., Scahill, RI., Owen, G., Durr, A., Leavitt, BR., Roos, RA., Langbehn, DR., et al. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *The Lancet Neurology*, 12(7), 637–649. doi:10.1016/S1474-4422(13)70088-7
- Tao, G., He, R., Datta, S., Narayana, PA. (2009). Symmetric inverse consistent nonlinear registration driven by mutual information. *Computer Methods and Programs in Biomedicine*, 95(2), 105–115. doi:10.1016/j.cmpb.2009.01.011
- Tarasoff-Conway, JM., Carare, RO., Osorio, RS., Glodzik, L., Butler, T., Fieremans, E., de Leon, MJ., et al. (2015). Clearance systems in the brain—implications for Alzheimer disease. *Nature Reviews Neurology*, 11(8), 457–470. doi:10.1038/nrneurol.2015.119
- Teipel, SJ., Alexander, GE., Schapiro, MB., Möller, H-JM., Rapoport, SI., Hampel, H. (2004). Age-related cortical grey matter reductions in non-demented Down's syndrome adults determined by MRI with voxel-based morphometry. *Brain*, 127(Pt 4), 811–24. doi:10.1093/brain/awh101

- Teipel, S.J., Kaza, E., Hadlich, S., Bauer, A., Plath, A-S., Krohn, M., Scheffler, K., Walker, L.C., Lotze, M., Pahnke, J. (2011). Automated MRI detection of Beta-Amyloid in a transgenic model of AD, 23(2), 221–237. doi:10.3233/JAD-2010-101035.Automated
- Teipel, S.J., Grothe, M., Lista, S., Toschi, N., Garaci, F.G., Hampel, H. (2013). Relevance of Magnetic Resonance Imaging for Early Detection and Diagnosis of Alzheimer Disease. *Medical Clinics of North America*, 97(3), 399–424. doi:10.1016/j.mcna.2012.12.013
- Terry, R.D., Masliah, E., Salmon, D.P., Butters, N., DeTeresa, R., Hill, R., Hansen, L.A., Katzman, R. (1991). Physical basis of cognitive alterations in Alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. *Annals of Neurology*, 30(4), 572–580. doi:10.1002/ana.410300410
- Thacker, N.A. (2003). Tutorial: A Critical Analysis of Voxel Based Morphometry (VBM). Retrieved from: <http://www.tina-vision.net/docs/memos/2003-011.pdf>
- Thomann, P.A., Dos Santos, V., Seidl, U., Toro, P., Essig, M., Schröder, J. (2009). MRI-derived atrophy of the olfactory bulb and tract in mild cognitive impairment and Alzheimer's disease. *Journal of Alzheimer's disease*, 17(1), 213–221. doi:10.3233/JAD-2009-1036
- Thompson, P.M., Rapoport, J.L., Cannon, T.D., Toga, A.W. (2003). Automated analysis of structural MRI data. In: *Brain Imaging in Schizophrenia*, ed. Lawrie, S., Johnstone, E.C., Weinberger, D. Oxford University Press.
- Thompson, W.K., Holland, D. (2011). Bias in tensor based morphometry Stat-ROI measures may result in unrealistic power estimates. *NeuroImage*, 57(1), 1–4. doi:10.1016/j.neuroimage.2010.11.092
- Tohka, J., Krestyannikov, E., Dinov, I.D., Graham, A.M., Shattuck, D.W., Ruotsalainen, U., Toga, A.W. (2007). Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE transactions on medical imaging*, 26(5), 696–711. doi:10.1109/TMI.2007.895453
- Tondelli, M., Wilcock, G.K., Nichelli, P., de Jager, C.A., Jenkinson, M., Zamboni, G. (2012). Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. *Neurobiology of aging*, 33(4), 825.e25–36. doi:10.1016/j.neurobiolaging.2011.05.018
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6), 1310–20. doi:10.1109/TMI.2010.2046908
- Uberti, M.G., Boska, M.D., Liu, Y. (2009). A semi-automatic image segmentation method for extraction of brain volume from in vivo mouse head magnetic resonance imaging using Constraint Level Sets. *Journal of neuroscience methods*, 179(2), 338–344. doi:10.1016/j.jneumeth.2009.02.007.A
- Ullmann, J.F.P., Keller, M.D., Watson, C., Janke, A.L., Kurniawan, N.D., Yang, Z., Richards, K., Paxinos, G., Egan, G.F., Petrou, S., Bartlett, P., Galloway, G.J., Reutens, D.C. (2012). Segmentation of the C57BL/6J mouse cerebellum in magnetic resonance images. *NeuroImage* 62 1408–1414. doi:10.1016/j.neuroimage.2012.05.061
- Ullmann, J.F.P., Watson, C., Janke, A.L., Kurniawan, N.D., Reutens, D.C. (2013). A segmentation protocol and MRI atlas of the C57BL/6J mouse neocortex. *NeuroImage*, 78, 196–203. doi:10.1016/j.neuroimage.2013.04.008
- Ullmann, J.F.P., Janke, A.L., Reutens, D., Watson, C. (2015). Development of MRI-based atlases of non-human brains. *Journal of Comparative Neurology*, 523(3), 391–405. doi:10.1002/cne.23678
- Van Eede, M.C., Scholz, J., Chakravarty, M.M., Henkelman, R.M., Lerch, J.P. (2013). Mapping registration sensitivity in MR mouse brain images. *NeuroImage*. doi:10.1016/j.neuroimage.2013.06.004
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P. (1999a). Automated Model-Based Bias Field Correction of MR Images of the Brain. 18(10), 885–896. doi:10.1109/42.811268
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P. (1999b). Automated Model-Based Tissue Classification of MR Images of the Brain. 18(10), 897–908. doi:10.1109/42.811270
- Vemuri, P., Whitwell, J.L., Kantarci, K., Josephs, K.A., Parisi, J.E., Shiung, M.S., Knopman, D.S., Boeve, B.F., Petersen, R.C., Dickson, D.W., Jack, C.R. (2008). Antemortem MRI based STructural Abnormality iNDEX (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. *NeuroImage*, 42(2), 559–567. doi:10.1016/j.neuroimage.2008.05.012



- Vemuri, P., Wiste, HJ., Weigand, SD., Shaw, LM., Trojanowski, JQ., Weiner, MW., Knopman, DS., Petersen, RC., Jack, CR. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects: Predicting future clinical change. *Neurology*, 73, 294–301. doi:10.1212/WNL.0b013e3181af79fb
- Vemuri, P., Jack Jr, CR. (2010). Role of structural MRI in Alzheimer's disease. *Alzheimer's Research & Therapy*. doi:10.1186/alzrt47
- Vemuri, P., Weigand, SD., Przybelski, SA., Knopman, DS., Smith, GE., Trojanowski, JQ., Jack, CR., et al. (2011). Cognitive reserve and Alzheimer's disease biomarkers are independent determinants of cognition. *Brain*, 134(5), 1479–1492. doi:10.1093/brain/awr049
- Verma, R., Mori, S., Shen, D., Yarowsky, P., Zhang, J., Davatzikos, C. (2005). Spatiotemporal maturation patterns of murine brain quantified by diffusion tensor MRI and deformation-based morphometry. *Proc. National Academy of Sciences USA*, 102(19), 6978–83. doi:10.1073/pnas.0407828102
- von Bohlen und Halbach, O., Lotze, M., Pfannmoller, JP. (2014). 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. *Front Neuroanat*, 8(June), 47. doi:10.3389/fnana.2014.00047
- Wang, X., Li, W., Wang, X., Qian, Z. (2010). Segmentation of Scalp, Skull, CSF, Grey Matter and White Matter in MRI of Mouse Brain. *Computer, (Bmei)*, 550–554. doi:10.1109/BMEI.2010.5639992
- Wang, H., Suh, JW., Das, SR., Pluta, JB., Craige, C., Yushkevich, PA. (2013). Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Pattern Analysis and Machine Intelligence* 35:3, pp611–623. doi:10.1109/TPAMI.2012.143
- Warfield, SK., Zou, KH., Wells, WM. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging*, 23(7), 903–21. doi:10.1109/TMI.2004.828354
- Weiner, MW., Veitch, DP., Aisen, PS., Beckett, LA., Trojanowski, JQ., et al. (2013). The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & dementia*. doi:10.1016/j.jalz.2013.05.1769
- Wells, JA., O'Callaghan, JM., Holmes, HE., Powell, NM., Johnson, RA., Siow, B., Torrealdea, F., Ismail, O., Walker-Samuel, S., Golay, X., Rega, M., Richardson, S., Modat, M., Cardoso, MJ., Ourselin, S., Schwarz, AJ., Ahmed, Z., Murray, TK., O'Neill, MJ., Collins, EC., Colgan, N., Lythgoe, MF. (2015). In vivo imaging of tau pathology using multi-parametric quantitative MRI. *NeuroImage*, 111, 369–378. doi:10.1016/j.neuroimage.2015.02.023
- White, NS., Alkire, MT., Haier, RJ. (2003). A voxel-based morphometric study of nondemented adults with Down Syndrome. *NeuroImage*, 20(1), 393–403. doi:10.1016/S1053-8119(03)00273-8
- Whitwell, JL., Crum, WR., Watt, HC., Fox, NC. (2001). Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJNR. American journal of neuroradiology*, 22(8), 1483–9.
- Whitwell, JL., Josephs, KA., Murray, ME., Kantarci, K., Przybelski, SA., Weigand, SD., Vemuri, P., Jack, CR., et al. (2008). MRI correlates of neurofibrillary tangle pathology at autopsy: A voxel-based morphometry study. *Neurology*, 71(10), 743–749. doi:10.1212/01.wnl.0000324924.91351.7d
- Winkler AM., Ridgway GR., Webster MA., Smith SM., Nichols TE. (2014). Permutation inference for the general linear model. *NeuroImage*, 2014;92:381-397. doi:10.1016/j.neuroimage.2014.01.060
- Wiseman, FK., Alford, KA., Tybulewicz, VLJ., Fisher, EMC. (2009). Down syndrome--recent progress and future prospects. *Human molecular genetics*, 18(R1), R75–83. doi:10.1093/hmg/ddp010
- Wiseman, FK., Al-Janabi, T., Hardy, J., Karmiloff-Smith, A., Nizetic, D., Tybulewicz, VLJ., Fisher, EMC., Strydom, A. (2015). A genetic cause of Alzheimer disease: mechanistic insights from Down syndrome. *Nature Reviews Neuroscience*, (August), 1–11. doi:10.1038/nrn3983
- Witton, J., Padmashri, R., Zinyuk, LE., Popov, VI., Kraev, I., Line, SJ., Jones, MW., et al. (2015). Hippocampal circuit dysfunction in the Tc1 mouse model of Down syndrome. *Nature Neuroscience*, 18(9), 1291–1298. doi:10.1038/nn.4072
- Wolfensohn, S., Lloyd, M. (2007). *Handbook of Laboratory Animal Management and Welfare* (3<sup>rd</sup> Ed.) Blackwell. doi:10.1002/9780470751077; ISBN:9781405111591

- Wong, MD., Dorr, AE., Walls, JR., Lerch, JP., Henkelman, RM. (2012). A novel 3D mouse embryo atlas based on micro-CT. *Development*, 139(17), 3248–3256. doi:10.1242/dev.082016
- Wong, MD., Maezawa, Y., Lerch, JP., Henkelman, RM. (2014). Automated pipeline for anatomical phenotyping of mouse embryos using micro-CT. *Development*, 141(12), 2533–41. doi:10.1242/dev.107722
- Wood, TC., Lythgoe, DJ., Williams, SCR. (2013). rBET: making BET work for rodent brains. *Proc. Intl. Soc. Mag. Reson. Med.*, 21, 2706.
- World Bank (2014), UK GDP: <http://data.worldbank.org/country/united-kingdom>
- Wright, I., McGuire, P., Poline, JB., Travers, J., Murray, R., Frith, C., Frackowiak, RSJ., Friston, K. (1995). A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *Neuroimage*, 2(4), 244–252. doi:10.1006/nimg.1995.1032
- Wright, AL., Zinn, R., Hohensinn, B., Konen, LM., Beynon, SB., Tan, RP., Clark, IA., Abdipranoto, A., Vissel, B. (2013). Neuroinflammation and neuronal loss precede A $\beta$  plaque deposition in the hAPP-J20 mouse model of Alzheimer's disease. *PloS one*, 8(4), e59586. doi:10.1371/journal.pone.0059586
- Wu, M., Rosano, C., Lopez-Garcia, P., Carter, CS., Aizenstein, HJ. (2007). Optimum template selection for atlas-based segmentation. *NeuroImage*, 34(4), 1612–1618. doi:10.1016/j.neuroimage.2006.07.050
- Wu, T., Bae, MH., Zhang, M., Pan, R., Badea, A. (2012). A prior feature SVM-MRF based method for mouse brain segmentation. *NeuroImage*, 59(3), 2298–306. doi:10.1016/j.neuroimage.2011.09.053
- Xie, Z., Yang, D., Stephenson, D., Morton, D., Hicks, C., Brown, T., Bocan, T. (2010). Characterizing the regional structural difference of the brain between tau transgenic (rTg4510) and wild-type mice using MRI. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI* (Vol. 13, pp.308–315). doi:10.1007/978-3-642-15705-9\_38
- Yang, D., Xie, Z., Stephenson, D., Morton, D., Hicks, CD., Brown, TM., Sriram, R., O'Neill, S., Raunig, D., Bocan, T. (2011). Volumetric MRI and MRS provide sensitive measures of Alzheimer's disease neuropathology in inducible Tau transgenic mice (rTg4510). *NeuroImage*, 54(4), 2652–8. doi:10.1016/j.neuroimage.2010.10.067
- Yu, T., Li, Z., Jia, Z., Clapcote, SJ., Liu, C., Li, S., Asrar, S., Pao, A., Chen, R., Fan, N., Carattini-Rivera, S., Bechard, AR., Spring, S., Henklman, RM., Stoica, G., Matsui, SI., Nowak, NJ., Roder, JC., Chen, C., Bradley, A., Yu, YE. (2010). A mouse model of Down syndrome trisomic for all human chromosome syntenic regions. *Human Molecular Genetics*, 19(14) 2780-2791. doi:10.1093/hmg/ddq179
- Yu, X., Nieman, BJ., Sudarov, A., Szulc, KU., Abdollahian, DJ., Bhatia, N., Lalwani, AK., Joyner, AL., Turnbull, DH. (2011). Morphological and functional midbrain phenotypes in Fibroblast Growth Factor 17 mutant mice detected by Mn-enhanced MRI. *NeuroImage*, 56(3), 1251–1258. doi:10.1016/j.neuroimage.2011.02.068
- Yushkevich, PA., Piven, J., Hazlett, HC., Smith, RG., Ho, S., Gee, JC., Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*, 31(3), 1116–28. doi:10.1016/j.neuroimage.2006.01.015
- Yushkevich, PA., Avants, BB., Das, SR., Pluta, J., Altinay, M., Craige, C. (2010). Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 T MRI data. *NeuroImage*, 50(2), 434–45. doi:10.1016/j.neuroimage.2009.12.007
- Zamyadi, M., Baghdadi, L., Lerch, JP., Bhattacharya, S., Schneider, JE., Henkelman, RM., Sled, JG. (2010). Mouse embryonic phenotyping by morphometric analysis of MR images. *Physiological genomics*, 42A(2), 89–95. doi:10.1152/physiolgenomics.00091.2010
- Zatorre, RJ., Fields, RD., Johansen-berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature Neuroscience*, 15(4), 528–536. doi:10.1038/nn.3045
- Zeiss, CJ., Ward, JM., Allore, HG. (2012). Designing phenotyping studies for genetically engineered mice. *Vet Pathol*; 49(1): 24–31. doi:10.1177/0300985811417247
- Zhang, Y., Brady, M., Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1), 45–57. doi:10.1109/42.906424

- Zhang, J., Chen, Y., Hardwick, JM., Miller, MI., Plachez, C., Richards, LJ., Yarowsky, P., van Zijil, P., Mori, S. (2005). Magnetic resonance diffusion tensor microimaging reveals a role for Bcl-x in brain development and homeostasis. *Journal of Neuroscience*, 25(8), 1881–8. doi:10.1523/JNEUROSCI.4129-04.2005
- Zhang, J., Peng, Q., Li, Q., Jahanshad, N., Hou, Z., Jiang, M., Masuda, N., Langbehn, DR., Miller, MI., Mori, S., Ross, CA., Duan, W. (2010). Longitudinal characterization of brain atrophy of a Huntington's disease mouse model by automated morphological analyses of magnetic resonance images. *NeuroImage*, 49(3), 2340–51. doi:10.1016/j.neuroimage.2009.10.027
- Zhang, B., Carroll, J., Trojanowski, JQ., Yao, Y., Iba, M., Potuzak, JS., Hogan, AML., Xie, SX., Ballatore, C., Smith III, AB., Lee, VM-Y., Brunden, KR. (2012). The microtubule-stabilizing agent, epothilone D, reduces axonal dysfunction, neurotoxicity, cognitive deficits, and Alzheimer-like pathology in an interventional study with aged tau transgenic mice. *Journal of neuroscience*, 32(11), 3601–11. doi:10.1523/JNEUROSCI.4922-11.2012
- Zigman, WB., Lott, IT. (2007). Alzheimers disease in down syndrome: Neurobiology and risk. *Mental Retardation and Developmental Disabilities Research Reviews*. doi:10.1002/mrdd.20163



Unless noted, URLs were checked c. November 2015.

The C57Bl/6J mouse image on the final page is from The Jackson Laboratory, Bar Harbor, Maine, USA  
(used with permission).



*This page intentionally left completely blank.*

*W*ee, sleekit, cowran, tim'rous beastie,

*O, what a panic's in thy breastie!  
Thou need na start awa sae hasty,  
Wi' bickering brattle!  
I wad be laith to rin an' chase thee,  
Wi' murd'ring pattle!*

*I'm truly sorry Man's dominion  
Has broken Nature's social union,  
An' justifies that ill opinion,  
Which makes thee startle,  
At me, thy poor, earth-born companion,  
An' fellow-mortal!*

*I doubt na, whyles, but thou may thieve;  
What then? poor beastie, thou maun live!  
A daimen-icker in a thrave 'S a sma' request:  
I'll get a blessin wi' the lave,  
An' never miss't!*

(...)

Robert Burns, 1785

