NeuroImage 82 (2013) 226–236

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Mapping registration sensitivity in MR mouse brain images

Matthijs C. van Eede [a,*], Jan Scholz [a], M. Mallar Chakravarty [b,c,d,e], R. Mark Henkelman [a,f], Jason P. Lerch [a,f]

[a] Mouse Imaging Centre, The Hospital for Sick Children, Toronto, Ontario, Canada
[b] Kimel Family Translational Imaging-Genetics Research Laboratory, the Centre for Addiction and Mental Health, Toronto, Canada
[c] Department of Psychiatry, University of Toronto, Toronto, Canada
[d] Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada
[e] Rotman Research Institute, Baycrest, Toronto, Canada
[f] Department of Medical Biophysics, University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Accepted 1 June 2013
Available online 10 June 2013

Keywords:
Deformation based morphometry
Evaluation
MRI
Neuroanatomy
Atrophy simulation
Mouse brain

ABSTRACT

Nonlinear registration algorithms provide a way to estimate structural (brain) differences based on magnetic resonance images. Their ability to align images of different individuals and across modalities has been well-researched, but the bounds of their sensitivity with respect to the recovery of salient morphological differences between groups are unclear. Here we develop a novel approach to simulate deformations on MR brain images to evaluate the ability of two registration algorithms to extract structural differences corresponding to biologically plausible atrophy and expansion. We show that at a neuroanatomical level registration accuracy is influenced by the size and compactness of structures, but do so differently depending on how much change is simulated. The size of structures has a small influence on the recovered accuracy. There is a trend for larger structures to be recovered more accurately, which becomes only significant as the amount of simulated change is large. More compact structures can be recovered more accurately regardless of the amount of simulated change. Both tested algorithms underestimate the full extent of the simulated atrophy and expansion. Finally we show that when multiple comparisons are corrected for at a voxelwise level, a very low rate of false positives is obtained. More interesting is that true positive rates average around 40%, indicating that the simulated changes are not fully recovered. Simulation experiments were run using two fundamentally different registration algorithms and we identified the same results, suggesting that our findings are generalizable across different classes of nonlinear registration algorithms.

Introduction

The structure and shape of the brain and its parts have been shown to be related to its function. Magnetic Resonance Imaging (MRI) allows the measurement of the intact brain's structural properties with high precision (Henkelman, 2010; Nieman et al., 2011). One MRI-based technique for probing shape differences in the brain, predominant in the study of animal models, is image registration (Badea et al., 2007; Chen et al., 2006; Lerch et al., 2008; Pitiot et al., 2007). Neuroanatomical differences between groups of subjects can then be estimated using deformation based morphometry (DBM) (Ashburner et al., 1998). Briefly, DBM works as follows. First, the MR images are aligned to a common registration target/template or one of the input images using affine transformations. Then a nonlinear registration is used for increasingly precise alignment to the target. This final registration yields the deformation fields which are then input into statistical analysis. The linear components of the deformation fields can be removed to account for overall differences in size and position. The Jacobian determinant of the deformation fields is often of interest as it summarizes the shape differences as expansions and contractions.

The lissencephalic mouse brain is homologous between individual mice, making DBM a particularly well suited method to capture differences in mouse brains. For example, Mercer et al. (2009) examined brain differences related to the *Magel2* gene, a gene that plays a role in cell differentiation and cell death. Several brain structures, such as the amygdala, the dentate gyrus and the nucleus accumbens were found to be 4–5% smaller in the mutants. The GSK-3α null mutants in Kaidanovich-Beilin et al. (2009) were shown to have the arbor vita enlarged by 9% and the pons by 5% as compared to their wild types litter mates. The gene products of GSK-3 are essential for cerebellar development and foliation. Mansouri et al. (2012) studied the effects of deficiencies in purine nucleoside phosphorylase (PNP) on the brain. Inherited defects in PNP in humans cause progressive neurological dysfunction. In their study regions in the cerebellum were shown to be about 15% smaller in PNP-knockout mice compared to wild type litter mates. In Spring et al. (2010) asymmetries in the mouse brain were investigated using DBM, which among others revealed an area in the cortex with a 15.5% difference between the left and right hemisphere.

* Corresponding author.
E-mail address: matthijsvaneede@gmail.com (M.C. van Eede).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.06.004

In the human brain with its complex and idiosyncratically folded cortex only subcortical structures may be homologous between subjects, however, DBM has been applied extensively on human MR data as well and is especially useful for longitudinal data (Brambati, 2009; Hyde et al., 2009; Kim, 2008; Tao, 2009).

Several studies have shown the superior effectiveness of MRI combined with automated image registration in comparison to stereology or manual segmentations in mice (Lau et al., 2008; Lerch et al., 2008; Spring et al., 2007). Other research, focused on human brain imaging, has compared the performance of multiple image registration packages based on a set of gold standard labels (Allen et al., 2008; Andreasen et al., 1996; Babalola et al., 2009; Bai et al., 2012; Christensen et al., 2006; Crum et al., 2004; Fischl et al., 2002; Heckemann et al., 2006; Hellier et al., 2003; Iosifescu et al., 1997; Klein et al., 2009; Ma et al., 2005; Quarantelli et al., 2002; Yassa and Stark, 2009). Calculating how accurately registration algorithms can reproduce gold standard segmentations, however, does not provide a direct estimation of their sensitivity to detecting changes in neuroanatomy, because overlap measures of labels cannot make inferences about the accuracy of registrations within those regions. To address that, Karacali and Davatzikos (2006) and Camara et al. (2006) have proposed methods to simulate atrophy on medical images. Their methods provide means to evaluating the sensitivity and accuracy of image registration techniques by simulating atrophy in a brain structure and subsequently trying to recover it.

Given the widespread use of DBM to investigate volumetric neuroanatomical differences in human and animal studies, we set out to assess its accuracy. Here we introduce a novel method to simulate atrophy and expansion in arbitrary areas in MR images and utilize this method to perform simulation experiments. The following seven points are addressed: 1) We evaluated how reliably image registration can recover change in the mouse brain by simulating atrophy in anatomical brain structures of a mouse brain atlas (Dorr et al., 2008). 2) Next we looked at how accurately the amount of simulated atrophy can be determined in the brain structures and how this depends on the size/shape/contrast at the boundaries of the structures. 3) We evaluated the influence that inherent variability in inbred mouse strains has on the recovery of simulated atrophy in brain structures, and 4) whether we systematically over- or underestimate simulated atrophy and expansion. 5) The simulation experiments were performed using two nonlinear registration methods to address the generality of our results. 6) Multiple comparison corrections were performed at a voxel level to determine the number of true and false positives. 7) Lastly, a small focal amount of atrophy was simulated in one half of a brain slice to investigate the impact of image features on the performance of image registration. The results demonstrate that three pieces of information influence the accuracy of registration on whole anatomical regions (size, compactness and inherent variability), and that while the accuracy of registration based measures of anatomy is impressive, not all simulated changes can be fully recovered.

Methods

Image acquisition

In this paper the simulation experiments used MRI scans acquired from 20 C57/bl6 male mice of young adult ages of 8–10 weeks [specimen preparation as in (Cahill et al., 2012)]. A multi-channel 7.0 T, 40 cm diameter bore magnet (Varian Inc. Palo Alto, CA) was used to acquire the anatomical images. A custom-built 16-coil solenoid array was used to image 16 samples concurrently (Lerch et al., 2011a). Parameters used in the scans were optimized for gray–white matter contrast: a T2-weighted 3D fast spin-echo sequence, with TR = 2000 ms, echo train length = 6, TE_{eff} = 42 ms, field-of-view (FOV) = 25 × 28 × 14 mm and matrix size = 450 × 504 × 250, giving an image with 56-micron isotropic voxels and an optimal signal to noise ratio (SNR) in the order of 40 (Kale et al., 2008). Total imaging time was 11.7 h. Scans were corrected for geometric distortions generated by the image acquisition process based on images of precision machined phantoms.

Creating an artificial deformation field

Atrophy and expansion were simulated by applying an artificial deformation field that produced the desired levels of volumetric loss and expansion (inspired by (Karacali and Davatzikos, 2006), see Fig. 1). First, a desired Jacobian determinant was generated for a region of interest (ROI). The Jacobian determinant of a deformation field specifies whether the change in each voxel is due to atrophy or expansion (Davatzikos et al., 1996). Voxels belonging to the ROI received either a reduced determinant (less than 1; atrophy), or an increased determinant (greater than 1; expansion). The remaining brain voxels received a determinant of 1 (no change). Next, a zero-vector deformation field was initialized. In an iterative procedure, the Jacobian determinant was calculated from the artificial deformation field and the vectors of each voxel's six nearest neighbors were adjusted to yield the approximate input determinant. To keep intracranial space constant, we created a tolerance map comprising the ventricles and the space between brain and skull (subarachnoid space) and allowed these areas to deform. To preserve topology, the volume change in these areas was limited to 50% of their original size.

The twenty mouse brain MRI scans were divided into two groups (mutants and controls) that minimized the structural differences between the groups. We simulated volume changes on two spatial scales; whole anatomical regions and focal spots.

First we induced three levels of atrophy (5, 10 and 20%) in 59 out of the 62 structures defined in Dorr et al. (2008). The 3 ventricle areas were left out and used as tolerance space. Two levels of expansion (5 and 10%) were induced in 14 randomly chosen structures with a size of at least 1 mm^3. Controls remain unchanged. Registrations at 5% atrophy were run using two registration algorithms, resulting in 236 registrations on simulated atrophy. 28 registrations were run on simulated expansion, totalling 264 registrations. A single registration took approximately 100 CPU hours, resulting in a total of about 26,400 h in single CPU time. This is equivalent to several weeks of processing on a mid-sized cluster of about 10 machines with an 8-core processor each.

Secondly, we induced focal atrophy in the mutants, the center of the simulated atrophy located at each voxel in one half of an axial brain slice (see Fig. 8). This change consisted of a 90% volume decrease in a discrete octahedron (radius of 3.5 voxels, volume of 63 voxels) translating into a 0.08 mm^3 volume loss. For computational efficiency, the data were resampled to 112-micron isotropic voxels. The entire axial brain slice contained 7640 brain voxels. We ran simulations in the left hemisphere of that slice, resulting in 3820 registrations. To ensure that our results displayed symmetry between the left and right hemisphere, we ran an additional 220 simulations in a rectangle in the right hemisphere. In total we ran 4040 registrations. Each registration at this coarser resolution took about 25 h of CPU time, totalling about 101,000 h of single CPU time for all registrations. This is equivalent to several months of processing on a mid-sized cluster. All processing was significantly sped up by using the GPC supercomputer at the SciNet HPC Consortium (Loken et al., 2010).

Image registration procedure

A multi-step image registration process was used to align the brains and create a consensus average (Kovačević et al., 2005). First, all brains were rigidly registered towards a pre-existing mouse brain average image. Then all possible pairwise 12-parameter registrations were carried out to create a linear average model of the entire data set. This linear average was the initial target for the final part of the registration. Here the scans were locally deformed through a mutli-scale nonlinear

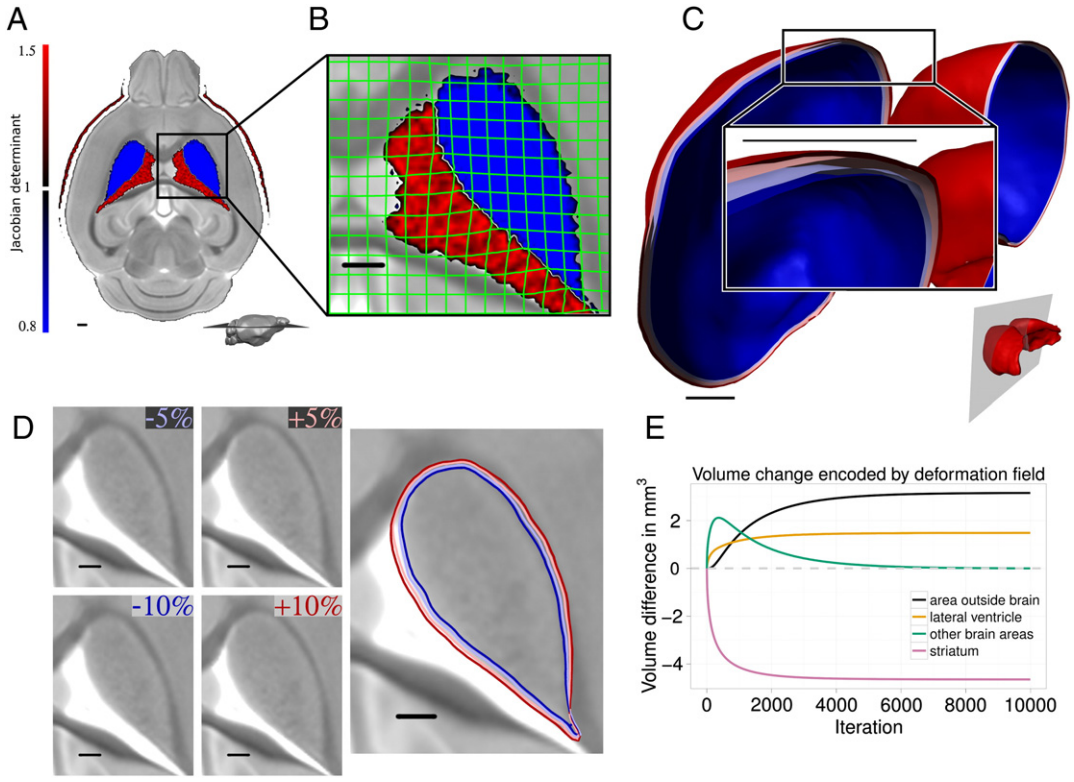

Fig. 1. Simulating atrophy in the striatum of the mouse. (A) A T2-weighted axial slice with the striatum highlighted in blue, neighboring ventricles and subarachnoidal volume are highlighted in red. The volume of the striatum is decreased (atrophy) by 20% and the volume of the ventricles and subarachnoidal volume are increased to minimize changes in other brain areas. Inset in the right lower corner indicates the position of the shown slice in the mouse brain. (B) A close-up shows the striatum and lateral ventricle area with the simulated deformation grid in green. (C) Surface rendering of the striatum. Change: 10% increase (red), 5% increase (pink), 5% decrease (light blue), and 10% decrease (dark blue). Note the very subtle shift in the surface of the striatum. (D) 4-panel image shows the striatum shrunk by 10%, 5% and increased by 10% and 5%. A close-up shows the original striatum with the contours of the 4 altered striatums overlaid. (E) Evolution of volume change in the entire brain during the iterative optimization of the simulated deformation field. The ventricular and subarachnoidal volumes increase to compensate for the decreasing striatal volume. This allows the rest of the brain to remain unchanged in volume. All scale bars in left lower corners represent 500 μm.

alignment procedure. After each iteration the images were averaged and the resulting atlas was used as the target for the next nonlinear iteration. This ensured that all scans were brought into exact correspondence without the preference for the shape of any particular input file. The deformation fields that were extracted for the scans hold the transformation that aligns them with the final atlas. We analyzed the Jacobian determinant of the extracted deformation fields to determine the difference that is recovered by the registration procedure between the two groups.

To demonstrate the generality of our results, we performed our simulation experiments using two fundamentally different registration algorithms. The first nonlinear registration procedure we used came from the Advanced Normalization Tools (ANTs) (Avants et al., 2008). This method came out as the best nonlinear registration method among 14 different algorithms (Klein et al., 2009). In particular, we used symmetric normalization, a diffeomorphic model, using as similarity metrics a combination of cross correlation on the regular image intensities and cross correlation on the 3D gradient magnitudes. A Gaussian regularizer was used that operates both on the similarity metric and on the deformation field. The second method, the Automatic Nonlinear Image Matching and Anatomical Labeling (ANIMAL) (Collins and Evans, 1997; Collins et al., 1995) algorithm, uses a multi-resolution, multi-scale methodology, and has been used in many studies to analyze morphometric variability (Chen et al., 2006; Dorr et al., 2008; Kovačević et al., 2005; Lerch et al., 2008; Mercer et al., 2009).

Volumetric data from extracted deformation fields

The volume of ROIs was determined by analysis of the extracted deformation fields. To reduce random noise and assure normality under the central limit theorem, the extracted deformation fields were blurred prior to analysis, and the logarithm of the Jacobian determinant was computed. Volumes were then determined by integrating over all voxels of the Jacobian determinant belonging to the ROI (brain structure or the octahedron comprising the focal area of change). To extract structure volumes a Gaussian filter with a full width at half maximum (fwhm) of 100 μm was used, and a fwhm of 200 μm was used for focal areas of atrophy. Overall brain size differences were accounted for by removing the linear component of the extracted deformation fields.

Measures on brain structures

The compactness of a brain structure was defined as $(\text{surface area})^{1.5}/(\text{volume})$. This shape measure has a minimum value of about 10.6 for spheres (see Fig. 4 for a visual comparison of the relationship between the size and the compactness for some brain structures). Two texture measures were examined: (1) the variance in image intensities and (2) the standard deviation of the image intensities within a brain structure. The investigated measures of boundary contrast were (1) the average gradient magnitude along the surface; (2) the percentage of brain structure surface with "high contrast", wherein MR images were normalized such that the median image intensity for all brain voxels was 1000, and high contrast was tested at two levels (an intensity difference of 300 or 500) and (3) the ratio of neighboring structures with high contrast versus low contrast.

Analysis on simulated focal atrophy

The experiments that examined the effect of focal atrophy on registration accuracy simulated a small amount of tissue loss (0.08 mm^3).

In order to examine the effect of this simulated volume loss in isolation, we eliminated the inherent variability present in the data. To this end, a registration was performed on the input files without any simulated changes. The Jacobian determinants extracted from this registration (inherent variability Jacobians) contain the inherent variability in the data set. Before examining the extracted Jacobian determinants from the simulation experiments (experiment Jacobians), the inherent variability Jacobians were subtracted from the experiment Jacobians.

To test the influence of proximal image features on the registration accuracy, average image intensities and average image gradient magnitudes were considered in all octahedrons with a radius of 3–15 voxels centered at the area of simulated atrophy.

Statistical comparison and multiple comparison correction

The anatomical structures of control and mutant groups with simulated atrophy or expansion were compared with a two-tailed *t*-test. P-values indicated are uncorrected. Linear regression was used to test for correlations. True and false positive voxels were determined for 14 randomly chosen structures larger than 1 mm^3, using a two-tailed *t*-test. To control for the number of false positives, we used the False Discovery Rate (FDR) (Genovese et al., 2002) and Threshold-Free Cluster Enhancement (TFCE) (Smith and Nichols, 2009). We adopted a 10% threshold for significance for FDR and $p < 0.1$ for TFCE.

Simulation experiments and results

The aim of this paper was to determine the sensitivity of image registration for MR mouse brain images. First we ran simulation experiments of atrophy and expansion over whole anatomical regions. We wanted to determine in which structures we can detect a change, and what factors influence how accurate we are at determining the simulated volumetric difference. We then investigated the ability of image registration to recover small local amounts of simulated atrophy.

Fractions of structures with simulated atrophy identified

Atrophy and expansion was simulated in the mouse brain structures defined in Dorr et al. (2008). After image registration, brain structure volumes were determined for all mutants and controls. Then, the volume of the brain structure with the simulated change in the mutant group was compared to the control group to establish whether DBM could detect this change.

At 5% simulated atrophy 20 out of 59 structures were recovered ($p < 0.05$, uncorrected; see Fig. 2A). One third of the brain structures defined in our atlas are smaller than 1 mm^3 and none of these 21 structures were recovered at $p < 0.05$. In contrast, 5 out of 13 structures that ranged in size between 1 and 3 mm^3 were recovered. Out of the remaining larger structures, 15 out of 25 were recovered. Given the lack of recoverability for structures smaller than 1 mm^3, we concluded that these structures span too few voxels (about 5700, 285 voxels lost at 5% atrophy) in a mouse brain MRI scan at 56 μm to be consistently recovered at this small amount of simulated atrophy.

After increasing the simulated change to 10% atrophy, 39 out of 59 structures were recovered ($p < 0.05$, uncorrected; see Fig. 2B). In this case 7 out of 21 structures smaller than 1 mm^3 were recovered, 8 out of 13 at 1–3 mm^3, and 24 out of 25 of the remaining structures. Finally at 20% simulated atrophy, 55 out of 59 structures were recovered ($p < 0.05$, uncorrected; not shown). Here, 17 out of the 21 structures smaller than 1 mm^3 were recovered as well as all the other remaining structures.

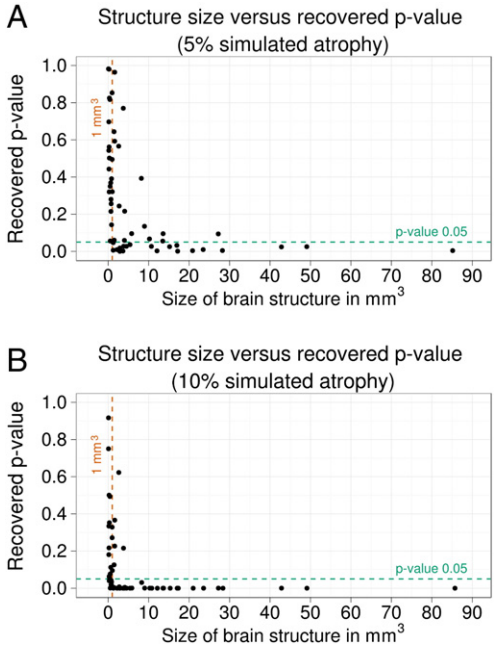

Fig. 2. The relationship between the size of a brain structure and the p-value at which it is recovered given (A) 5% and (B) 10% simulated atrophy.

Accuracy in recovered atrophy

First we tested what proportion of the simulated atrophy we could recover. For this test we excluded structures smaller than 1 mm^3 due to their poor general recoverability (see Fractions of structures with simulated atrophy identified section). There was a trend towards a positive correlation between structure size and the recovered atrophy ($p = 0.089$, adjusted $R^2 = 0.053$, see Fig. 3A). Compactness showed a negative correlation with the recovered atrophy in a structure ($p = 0.0023$, adjusted $R^2 = 0.21$, see Fig. 3B). Because the amount of simulated atrophy (5%) was small, we examined the effect that pre-existing differences within our data had on the results. To this end we calculated the inherent variability that existed between the structures in the control group and the mutant group without any simulated adjustment. We found a significant positive correlation between the amount of inherent variability in a structure and the recovered atrophy ($p = 3e\text{-}08$, adjusted $R^2 = 0.57$, see Fig. 3C). After increasing the simulated atrophy to 10% and 20% in the same structures, the effect of this inherent variability on the recovered atrophy dropped by ~80%. Conversely the effect of the size and the compactness of structures on the recovered atrophy increased by 100% and 75% respectively (see Fig. 3D). We further investigated two texture measures: the variance in image intensity and the standard deviation of image intensities within brain structures. Neither of these texture measures showed any relationship with the accuracy of the recovered atrophy. Finally, we examined the influence that the degree of contrast along the boundary of anatomical structures had on the accuracy in recovered atrophy. In particular, the average gradient magnitude along the surface was tested, the ratio of neighboring structures with high contrast versus low contrast as well as the percentage of structure surface with high contrast. Our MR images were normalized to have a median image intensity for all brain voxels of 1000. High contrast was tested at two levels: having an intensity difference of 300 and 500. These measures of boundary contrast did not show any correlations with the accuracy of recovered atrophy.

In the experiments where 5% atrophy was simulated we found that with the exception of a single structure the measured atrophy always underestimated the simulated atrophy. The exception was the globus pallidus for which the recovered atrophy was 5.3%, and

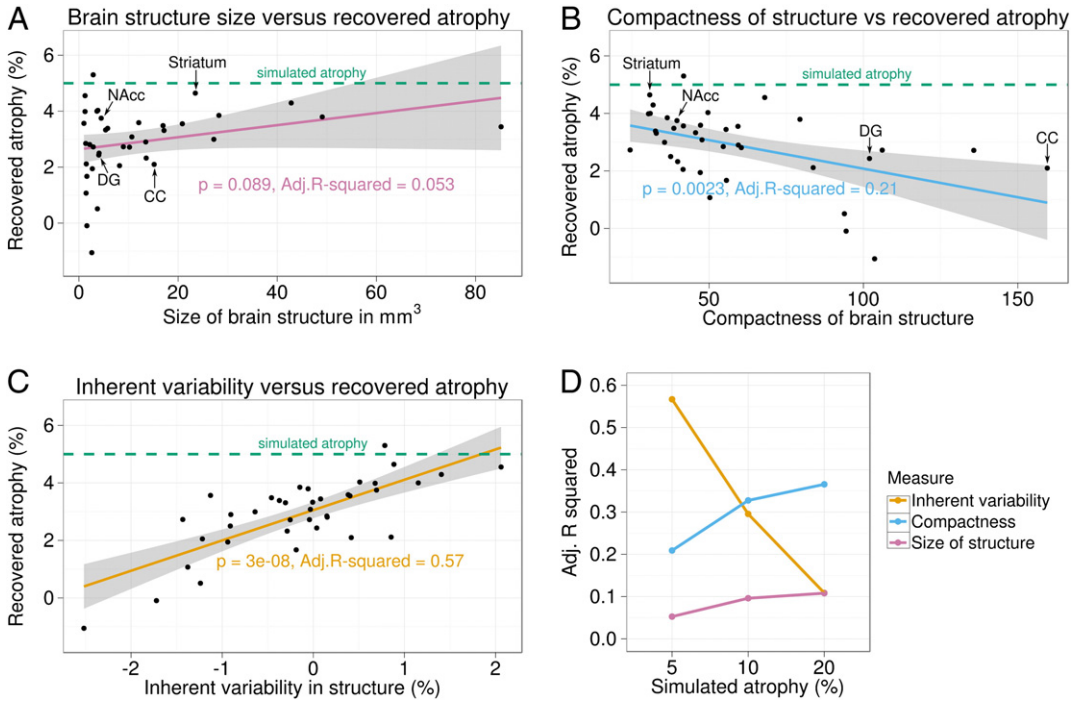

Fig. 3. What drives the accuracy of the recovered atrophy when 5% atrophy is simulated in structures at least 1 mm^3 in size. (A) There was a trend towards a positive correlation between the size of a structure and the recovered atrophy ($p = 0.089$). (B) The compactness of a structure had a negative correlation with the recovered atrophy ($p < 0.01$), and (C) a significant positive correlation was found at 5% simulated atrophy with the structural differences that exist without any simulated adjustment (represented as inherent variability where negative values indicate structures that were inherently larger in the mutant group compared to the controls in the data set) ($p < 0.001$). (D) The effect of the inherent variability decreased as the amount of simulated atrophy increased, whereas the effect of the size and compactness on the recovered atrophy increased. CC = corpus callosum, DG = dentate gyrus of the hippocampus, NAcc = nucleus accumbens.

the mutant group was 0.8% smaller than the controls due to the inherent variability in this structure. To save computation time, we then simulated both atrophy and expansion at 5% and 10% in 14 randomly selected regions. Regardless of the direction of change, the recovered changes consistently underestimated the simulated changes (see Fig. 5).

All the simulation experiments described above were performed using ANTs. To ensure that our results generalize across registration algorithms we repeated all the experiments at 5% simulated atrophy using ANIMAL and found the same patterns. There was a positive correlation between structure size and the recovered atrophy ($p = 0.044$, adjusted $R^2 = 0.083$), a negative correlation between compactness and the recovered atrophy ($p = 0.00017$, adjusted $R^2 = 0.31$), and a positive correlation between the amount of inherent variability and the recovered atrophy ($p = 4.7e\text{-}08$, adjusted $R^2 = 0.56$) (see Fig. 6).

True positives and false positives

In 14 randomly chosen structures with a size of at least 1 mm^3 we simulated three levels of atrophy (5, 10 and 20%) and performed a voxel based analysis. The Jacobian determinants were blurred using a Gaussian filter with a fwhm of 200 μm. We then examined the amount of true positives and false positive voxels that we obtained

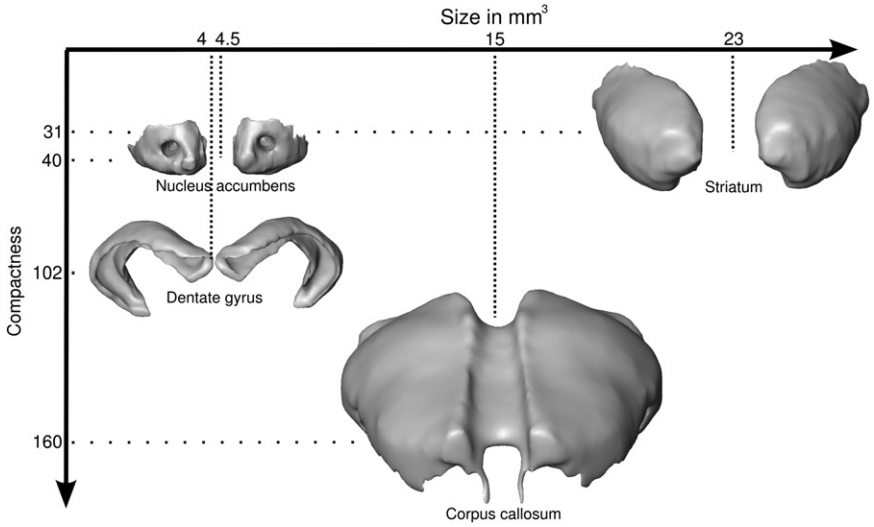

Fig. 4. The difference between the size of a structure and its compactness. The nucleus accumbens and the dentate gyrus of the hippocampus have about the same volume, but the shape of the former is a lot more compact. Due to its non-compact shape, the corpus callosum looks like a bigger structure than the striatum, but in volume it is about 35% smaller.

after correcting for multiple comparisons. At 20% simulated structural atrophy, the number of true positive voxels that we recovered on average was about 40% using TFCE and about 36% using FDR (see Fig. 7A). While in some of the smaller structures (cerebellar peduncle, dentate gyrus and fimbria, all smaller than 4 mm^3) we obtained very few true positives, about 80% of the thalamus (17 mm^3) was recovered. The true positives dwarf the number of false positives which, on average, was about 0.15% for the TFCE correction and 0.20% for the FDR correction (see Fig. 7B). Some of the voxels classified as false positives were located just outside the boundary of the structure under investigation (see Fig. 7C and D). After accounting for a minor spillover into adjacent structures by excluding a radius of 3 voxels around the structure of interest when determining the number of false positives, TFCE produced 0.03% false positives on average, and FDR 0.11%. This indicates that the overwhelming majority of the voxel based analysis produces true positives.

We then simulated 10% atrophy in these structures, and obtained on average 2.5% true positive voxels with TFCE and 1.1% true positives with FDR. The false positives at this level were 0.006% and 0.007% for TFCE and FDR respectively. We were able to detect 5% atrophy at a structure level using a matched filter (North, 1963), i.e. the structure definition itself. At a voxel based analysis using a 200-micron Gaussian filter however, no true or false positives voxels were recovered in any of the tested structures at 5% simulated atrophy.

Influence of image features on recovering simulated focal atrophy

To examine the influence of image features on the accuracy of image registration we ran simulation experiments with a small focal amount of atrophy at each voxel location within a region of interest. This change was identical in shape and size for each of the simulation experiments and consisted of a 90% volume decrease in a discrete octahedron with a radius of 3.5 voxels (63 voxels in total) translating into a 0.08 mm^3 volume loss. For computational efficiency, we simulated the atrophy in one half of a brain slice at 112-micron isotropic resolution, yielding 4040 registrations (see Fig. 8). As described in Accuracy in recovered atrophy section, the inherent variability in the data has a strong positive correlation on the accuracy of the recovered change obtained by image registration. To examine the effect of this small amount of simulated volume change in isolation, we eliminated the inherent variability in the data. Briefly, a registration was run on the data without any simulated changes. The Jacobian determinants from this registration encode the inherent variability in the data set. Prior to examining the extracted Jacobian determinants from the simulation experiments, the Jacobians that encode this inherent variability were subtracted.

We examined how image intensities and image gradient magnitudes influence the accuracy of the recovered atrophy as well as the t-statistic at which it was recovered. The average image intensities and the t-statistic at which the simulated atrophy was recovered showed a positive correlation (adjusted $R^2 = 0.2915$ at a radius of 13 voxels, see Fig. 8B). The average gradient magnitudes also showed a positive correlation with the t-statistics (adjusted $R^2 = 0.2466$ at a radius of 15 voxels, see Fig. 8C). Average intensity and average gradient magnitude accounted for 46% of the variance in the t-statistics (adjusted $R^2 = 0.4595$). We further found positive correlations between the image intensities and the accuracy in recovered atrophy (adjusted $R^2 = 0.1617$ at a radius of 3 voxels, see Fig. 8E) and a positive correlation between the image gradients and the accuracy in recovered atrophy (adjusted $R^2 = 0.2948$ at a radius of 13 voxels, see Fig. 8F). Here, average intensity and average gradient magnitude accounted for 34% of the variance (adjusted $R^2 = 0.3425$).

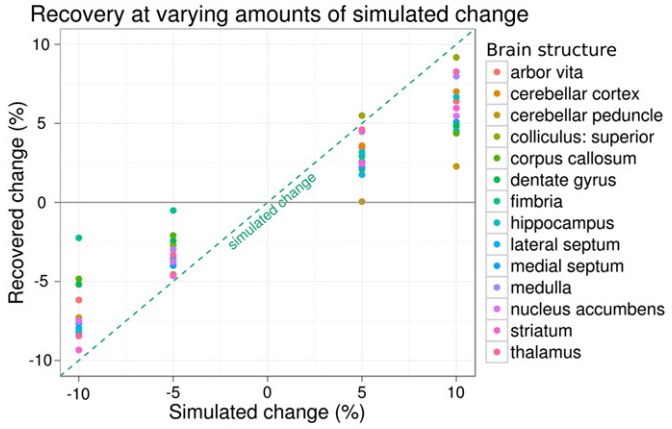

Fig. 5. Consistent underestimation of the recovered change versus the simulated change.

Discussion

In this study we investigated the effect that nonlinear registration algorithms have on extracting volumetric group differences from MRI mouse brain images and how these effects might vary across the brain. We simulated atrophy and expansion in whole anatomical brain regions as well as in small local areas and examined the ability of two registration algorithms to recover the simulated group differences.

First we evaluated how reliably image registration can recover simulated atrophy in mouse brain structures using the atlas from Dorr et al. (2008). We found that when 5% atrophy was simulated, image registration reliably recovered this change in structures that are 1 mm^3 or larger in size. We expect however, that in general it is not the structure size that determines this threshold, but that it is a combination of the imaging resolution and the number of voxels that is lost or gained with a simulated change. Indeed we found that at 10% simulated atrophy, where the loss in voxels doubles, we could recover one third of the structures that are smaller than 1 mm^3.

Next we looked at how accurately we could determine the amount of simulated atrophy in the brain structures. The structure size showed a trend towards a positive correlation with the accuracy in recovered atrophy. Simulated atrophy tended to be recovered more accurately in larger structures. This positive correlation was weak at small amounts of change, and only became significant as the amount of change was large. We assume that this trend ties in with the increasing number of voxels that was lost in larger structures. Our results showed that 21–37% of the variance in accuracy could be explained by the compactness of brain structures. This shape measure showed a significant negative correlation on the recovered atrophy regardless of the how much atrophy was simulated. A simulated change in more spherical (compact) shapes was recovered more accurately than in elongated or sheet like (non-compact) structures. This might be due to the difference in boundary movement resulting from an equal volume change in compact and non-compact shapes. The boundaries of compact shapes have a larger displacement than those in more elongated shapes at a similar change in volume. This larger displacement should be easier to detect, and thus the accuracy in the recovered change should be increased as well.

Registration algorithms operate on the image intensities and gradients. We expected to find a correlation between sensitivity in recovery and the strength of the gradients on the boundaries of brain structures. We tested this but found no relationship. This is likely due to an inherent need for a visible boundary in order to be able to define and include a structure in a segmented atlas. When we looked at local changes, we did find a correlation between gradient magnitudes and registration sensitivity.

Brain structure volumes among inbred mice of the same strain vary in size on the order of 2–15% (Chen et al., 2006). These inherent structure size differences in our data ranged up to 2.5%. We evaluated what the impact of this natural inter-individual variability in structure size was on the accuracy of the recovered atrophy. When only

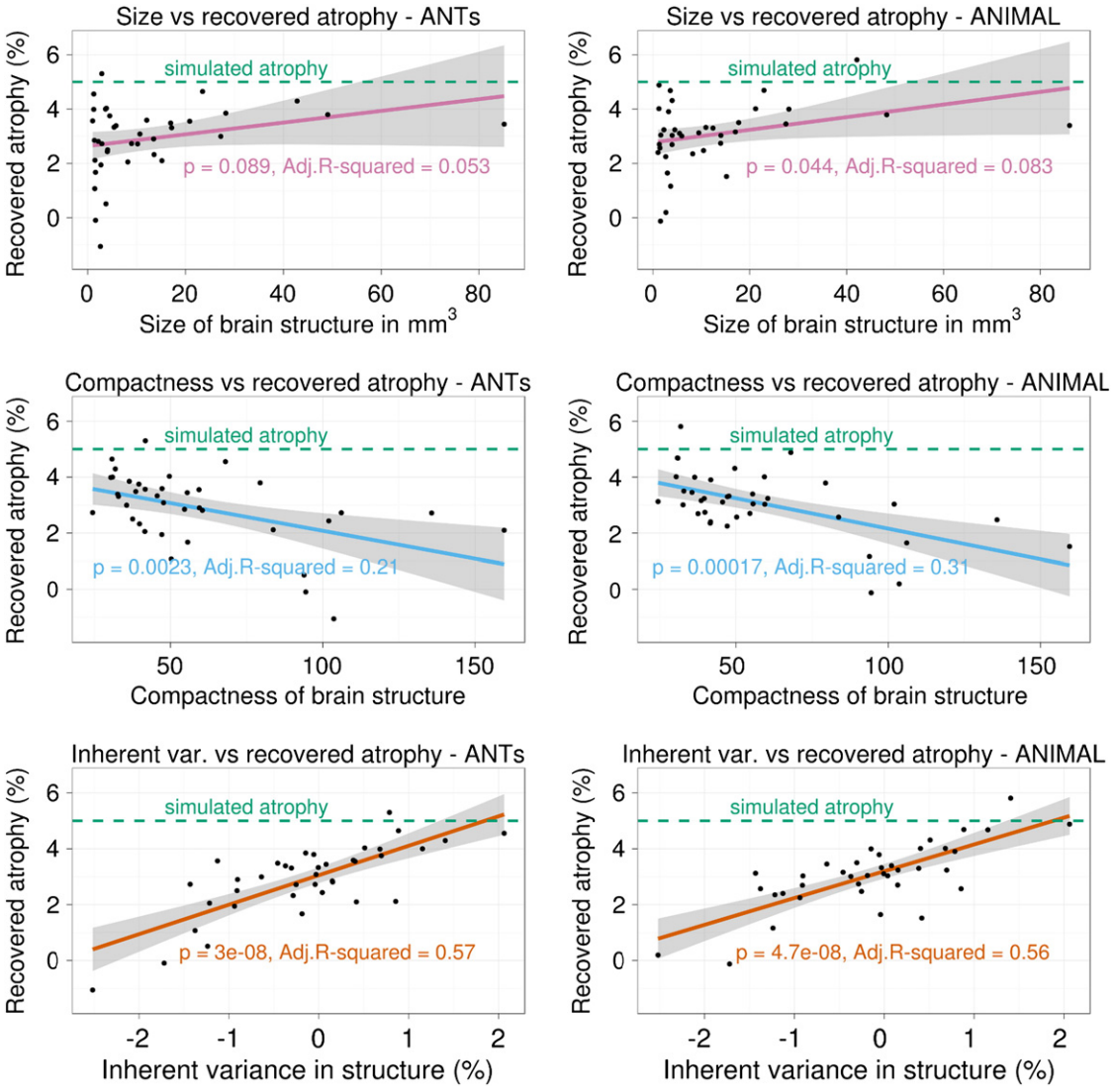

Fig. 6. The same correlation patterns were found using both the Advanced Normalization Tools (ANTs) (results in the left column; same as in Fig. 3) and the Automatic Nonlinear Image Matching and Anatomical Labeling (ANIMAL) algorithm (results in the right column).

5% atrophy was simulated, inherent variability showed a strong positive correlation with the accuracy in recovered atrophy and accounted for as much as 57% of the variance. However, as the amount of simulated atrophy increased and thus the inherent inter-individual variability was reduced to a smaller portion of the simulated change, the strength and impact of this positive correlation decreased.

We also found that, in general, the recovered amount of atrophy/expansion underestimated the amount simulated in the data. There were two exceptions, and in both cases the overestimation is likely due to the inherent variability that exists in the data (the globus pallidus: 5% atrophy simulated, 5.3% atrophy recovered, 0.8% inherent variability and the superior colliculus: 5% expansion simulated, 5.5% expansion recovered, 1.4% inherent variability). On average, 61 ± 21% of the simulated changes was recovered. In Pieperhoff et al. (2008), a processing pipeline is presented based on DBM to analyze the progression of neurodegenerative disease in humans. To validate their methods, atrophy is simulated in 6 areas of the brain and expansion was simulated in the lateral ventricles. They report a mean recognition rate (recovered atrophy/simulated atrophy) of 64 ± 13%, which agrees well with the underestimation we found. Image registration algorithms use regularization on for example the similarity metrics or the produced deformation fields in order to preserve topology in the resulting vector fields and avoid self-intersecting deformations. We hypothesize that it is these regularizations that result in the consistent underestimation of the volume changes we simulated.

Next we investigated whether our results applied to image registration procedures used to extract groupwise differences in general. Several studies have shown that the smoothness of Jacobian determinants varies for different registration algorithms (Han et al., 2012; Yanovsky et al., 2009). These differences stem from the fact that registration methods use different transformation models and regularization techniques. To ensure that our findings are not biased towards a single registration algorithm, we ran all our experiments at 5% simulated atrophy using two fundamentally different registration algorithms. ANTs which use a symmetric diffeomorphic transformation model with cross correlation and ANIMAL, which uses a multi-resolution, multi-scale methodology to produce a global nonlinear transformation based on local translations. We recovered the same correlations with both algorithms, suggesting that our findings are not limited to one registration approach.

In addition to assessing the accuracy in recovery of entire structures, we assessed the number of true positive and false positive voxels recovered using a voxelwise analysis. A *t*-test was performed at each voxel in the brain, and to control for the number of false positives, FDR (Genovese et al., 2002) and TFCE (Smith and Nichols,

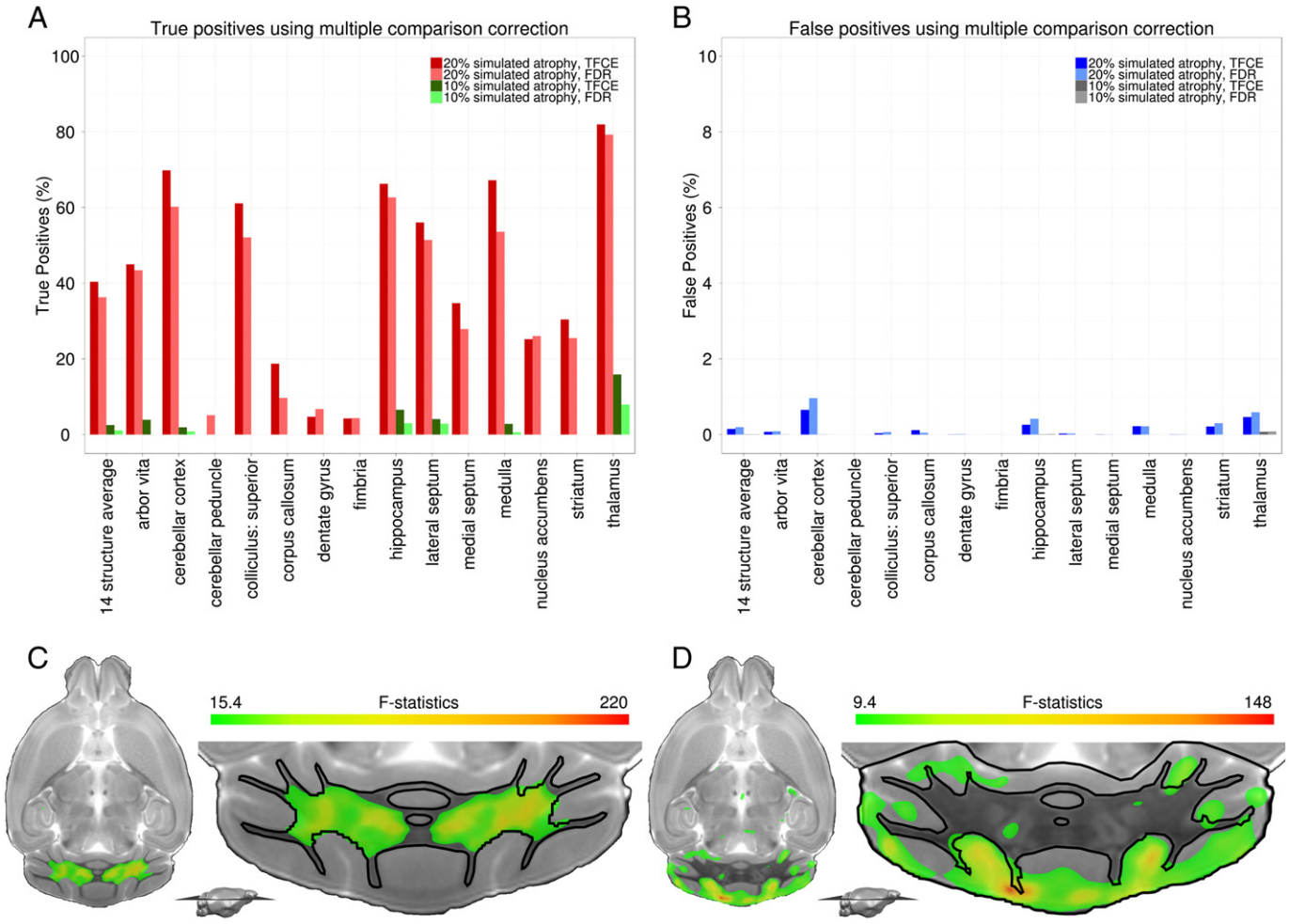

Fig. 7. True and false positives obtained through a voxel based analysis. (A) The percentage of true positive voxels and (B) false positive voxels captured in 14 structures using multiple comparison correction methods TFCE and FDR. The structures were reduced in size either by 20% or 10%. Note the different scales of the y-axes. (C) An axial slice through the mouse brain with a close-up of the cerebellum area with the corresponding f-statistics map for significant voxels at 10% FDR. The arbor vita of the cerebellum (outlined in black) was reduced in size by 20%. (D) Similarly for the analysis of the cortex of the cerebellum reduced in size by 20%.

2009) were used. At 20% simulated atrophy, the average amount of true positive voxels was about 38% and the average amount of false positives was about 0.17%. At smaller amounts of simulated atrophy, very few voxels reached statistical significance after correcting for multiple comparisons. In this experiment atrophy was simulated over an entire anatomical structure. The optimal way to extract a change in brain structure size from the Jacobian determinants is by using a matched filter (North, 1963), i.e. the label of a structure. However, when a voxelwise analysis is performed one generally does not have a priori knowledge of the size or shape of the signal to be extracted, and a Gaussian filter is usually used instead. In our simulations to detect this brain structure change we used a Gaussian filter with a fwhm of 200 μm, similar to our MRI phenotyping studies (Ellegood et al., 2011; Lerch et al., 2011b; Wheeler et al., 2013). Indeed, when we used a matched filter (i.e. integrated the Jacobians across a structure label as done in Fig. 2), simulated atrophy could be detected down to 5% in these same brain regions.

Our experiments indicate that when performing a voxelwise analysis and subsequently correcting for multiple comparisons the majority of the findings are indeed true positives. However, we should keep in mind that at 20% simulated atrophy the true positive voxel rates across structures ranged from 0 to 80%, indicating that false negatives ranged from 20 to 100% and we thus do not recover all the changes present in the data. In contrast, when instead of a voxelwise analysis, an analysis using structure labels is performed at 20% atrophy all changes are recovered (data not shown).

Lastly, we investigated the impact of image features on the performance of image registration. We simulated a small local amount of atrophy in one half of a brain slice and found that image intensities and gradient magnitudes show positive correlations with the recovered t-statistic (accounting for 46% of the variance). They also showed positive correlations with the accuracy of the recovered atrophy (accounting for 34% of the variance). In areas with higher average image intensities and higher average gradient magnitudes the simulated atrophy was recovered more accurately and with a higher t-statistic. At first glance, the positive correlations with respect to the image intensities were surprising. In T2 weighted MR images, ventricles have high image intensities as well as high gradients with the surrounding tissue. It is these areas in which the highest t-statistics were recovered and where the simulated atrophy was most accurately recovered. When we excluded the ventricles from the analysis, the adjusted R^2 dropped from 0.1617 to 0.0307 (data not shown). This suggests that the ventricles alone predominantly contributed to the positive correlation we found between accuracy in recovered atrophy and image intensities. We think that because of the presence of high intensities and high gradients in certain areas, image intensities work as a proxy for certain tissues. The striatum has low intensity values and low gradient magnitudes. In those areas, the sensitivity of image registration is low. Ventricles show the opposite pattern.

The experiments described in this paper assessed the ability of DBM to detect atrophy and expansion using simulated data. In order

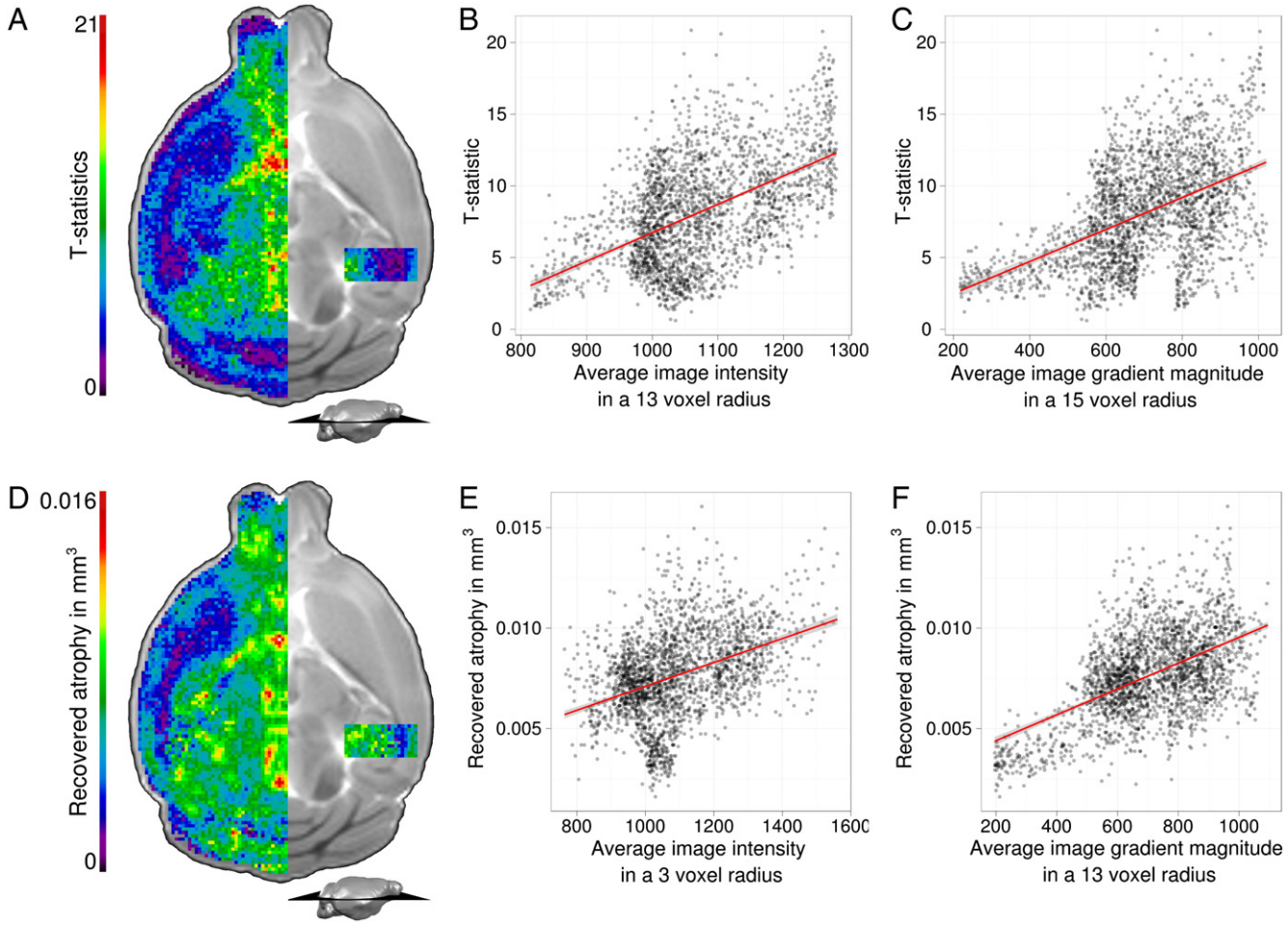

Fig. 8. Results of simulated focal atrophy. (A) A T2-weighted axial slice with t-statistics for each voxel at which focal atrophy was simulated overlaid. (B) There was a positive correlation (adjusted $R^2 = 0.2915$) between the t-statistic and the average image intensity in a radius of 13 voxels around the center of the simulated atrophy, and (C) a positive correlation (adjusted $R^2 = 0.2466$) between the t-statistic and the average image gradient magnitude in a radius of 15 voxels. (D) A T2-weighted transversal slice with the recovered atrophy in mm^3 overlaid. The recovered atrophy was adjusted for pre-existing differences in the data. (E) There was a positive correlation (adjusted $R^2 = 0.1617$) between the recovered atrophy and the average image intensity in a radius of 3 voxels, and (F) a positive correlation (adjusted $R^2 = 0.2948$) between the recovered atrophy and the average image gradient magnitude in a radius of 13 voxels.

to simulate these changes, topology-preserving deformation fields were created and applied to MR images of mouse brains. The resampling of the data that is involved when applying these deformations blurs the images, potentially leading to a reduction in contrast in the areas with simulated change. In turn, this could lead to an increased amount of false negatives. In an ideal world, our experiments would have been performed on data sets with known atrophy/expansion. In the absence of data with precisely known biological differences, we consider the simulated data a very reasonable substitute.

In general the differences we encounter in mouse brain studies are small; in the order of 5–20%. And we find that the hierarchical model building approach described in the manuscript suits these problems well with a stable set of parameters and even across registration algorithms. Both the diffeomorphic mapping strategy, as well as the strategy based on local affine transformations perform well. Our results indicate that recovering very subtle atrophy in small structures is difficult. To increase statistical power when these changes are expected, increased subject numbers could be used. When large amounts of inherent variability are expected within the study population, a longitudinal in-vivo approach could be beneficial. For an examination of the tradeoff between using in-vivo and ex-vivo experiment setups to recover differences in the brain, we refer the reader to Lerch et al. (2012).

Conclusion

In this paper we apply a novel method to simulate deformation fields with known structural tissue atrophy or expansion to simulate changes in mouse brain MRIs. We then tested how our ability to recover these changes depended on the amount and location of these perturbations as well as the properties of the affected anatomical structures and underlying image features. Our findings can be summarized as follows: (1) We can reliably pick up 5% atrophy in structures that are at least 1 mm^3 in size at a 56-micron isotropic resolution (a loss of 285 voxels). (2) The accuracy with which simulated atrophy of whole mouse brain structures can be recovered depends on the size (adjusted R^2 0.05–0.11) and compactness (adjusted R^2 0.21–0.37) of the structure. Larger structures tended to be recovered more accurately in the presence of small amounts of simulated change. This relationship was significant only at large amounts of change. More compact structures can be recovered more accurately regardless of how much change was simulated. (3) The presence of natural inter-individual variability in the size of brain structures has a considerable impact (57%) on the accuracy of recovering small changes (at 5% simulated atrophy). (4) Simulated atrophy and expansion recovered by spatial normalization underestimates the volume difference simulated in the data. (5) Similar patterns identified between accuracy in recovered atrophy and shape measures

and the existence of prior differences using two fundamentally different nonlinear registration algorithms suggest that these findings apply to image registration in general. (6) Voxelwise results corrected for multiple comparisons show fairly low rates of true positives. We simulated 20% atrophy in 14 brain structures, and on average about 40% of the affected voxels were recovered as true positives. In the extreme case we do not recover any true positive voxels: the cerebellar peduncle when using TFCE correction (see Fig. 7A). (7) Finally, the presence of strong image gradient magnitudes increases the accuracy of image registration at a voxel level. In T2 weighted images, so do the image intensities, but we think that here they act as a proxy for areas with high contrast (ventricles). Ongoing development of registration algorithms will improve their ability to capture small changes accurately. On the statistical side of the analysis, new methods could be developed that account for the varying sensitivity across the brain.

The code we used to generate the simulated deformation fields as well as a link to the documentation for the code can be found on GitHub (https://github.com/mcvaneede/generate_deformation_fields).

Acknowledgments

We gratefully acknowledge the financial support from the Canadian Institutes of Health Research (CIHR). Computations were performed on the GPC supercomputer at the SciNet HPC Consortium (Loken et al., 2010). SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund – Research Excellence; and the University of Toronto. In excess of 125,000 CPU hours were used to create the simulated deformation fields and perform the registrations.

Conflict of interest

The authors have no conflict of interest to declare.

References

Allen, J.S., Bruss, J., Mehta, S., Grabowski, T., Brown, C.K., Damasio, H., 2008. Effects of spatial transformation on regional brain volume estimates. NeuroImage 42, 535–547.

Andreasen, N.C., Rajarethinam, R., Cizadlo, T., Arndt, S., Swayze, V.W., Flashman, L.A., OLeary, D.S., Ehrhardt, J.C., Yuh, W.T.C., 1996. Automatic atlas-based volume estimation of human brain regions from MR images. J. Comput. Assist. Tomogr. 20, 98–106.

Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., 1998. Identifying global anatomical differences: deformation-based morphometry. Hum. Brain Mapp. 6, 348–357.

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

Babalola, K.O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., Smith, S., Cootes, T., Jenkinson, M., Rueckert, D., 2009. An evaluation of four automatic methods of segmenting the subcortical structures in the brain. NeuroImage 47, 1435–1447.

Badea, A., Ali-Sharief, A.A., Johnson, G.A., 2007. Morphometric analysis of the C57BL/6J mouse brain. NeuroImage 37, 683–693.

Bai, J., Trinh, T.L.H., Chuang, K.-H., Qiu, A., 2012. Atlas-based automatic mouse brain image segmentation revisited: model complexity vs. image registration. Magn. Reson. Imaging 30, 789–798.

Brambati, S.M., 2009. Atrophy progression in semantic dementia with asymmetric temporal involvement: a tensor-based morphometry study. Neurobiol. Aging 30, 103–111.

Cahill, L.S., Laliberté, C.L., Ellegood, J., Spring, S., Gleave, J.A., Van Eede, M.C., Lerch, J.P., Henkelman, R.M., 2012. Preparation of fixed mouse brains for MRI. NeuroImage 60, 933–939.

Camara, O., Schweiger, M., Scahill, R.I., Crum, W.R., Sneller, B.I., Schnabel, J.A., Ridgway, G.R., Cash, D.M., Hill, D.L.G., Fox, N.C., 2006. Phenomenological model of diffuse global and regional atrophy using finite-element methods. IEEE Trans. Med. Imaging 25, 1417–1430.

Chen, X.J., Kovacevic, N., Lobaugh, N.J., Sled, J.G., Henkelman, R.M., Henderson, J.T., 2006. Neuroanatomical differences between mouse strains as shown by high-resolution 3D MRI. NeuroImage 29, 99–105.

Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H., 2006. Introduction to the Non-rigid Image Registration Evaluation Project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (Eds.), Biomedical Image Registration, Lecture Notes in Computer Science. Springer Berlin, Heidelberg, pp. 128–135.

Collins, D.L., Evans, A.C., 1997. Animal: Validation and Applications of Nonlinear Registration-Based Segmentation. Int. J. Pattern Recognit. Artif. Intell. 11, 1271–1294.

Collins, D.L., Holmes, C.J., Peters, T.M., Evans, A.C., 1995. Automatic 3-D model-based neuroanatomical segmentation. Hum. Brain Mapp. 3, 190–208.

Crum, W.R., Rueckert, D., Jenkinson, M., Kennedy, D., Smith, S.M., 2004. A framework for detailed objective comparison of non-rigid registration algorithms in neuroimaging. In: Barillot, C., Haynor, D.R., Hellier, P. (Eds.), Medical Image Computing and Computer-Assisted Intervention — MICCAI 2004, Lecture Notes in Computer Science. Springer Berlin, Heidelberg, pp. 679–686.

Davatzikos, C., Vaillant, M., Resnick, S.M., Prince, J.L., Letovsky, S., Bryan, R.N., 1996. A computerized approach for morphological analysis of the corpus callosum. J. Comput. Assist. Tomogr. 20, 88–97.

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.

Ellegood, J., Lerch, J.P., Henkelman, R.M., 2011. Brain abnormalities in a Neuroligin3 R451C knockin mouse model associated with autism. Autism Res. 4, 368–376.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.

Han, Z., Thornton-Wells, T.A., Dykens, E.M., Gore, J.C., Dawant, B.M., 2012. Effect of nonrigid registration algorithms on deformation-based morphometry: a comparative study with control and Williams syndrome subjects. Magn. Reson. Imaging 30, 774–788.

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

Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Le Goualher, G., Collins, D.L., Evans, A., Malandain, G., Ayache, N., Christensen, G.E., Johnson, H.J., 2003. Retrospective evaluation of intersubject brain registration. IEEE Trans. Med. Imaging 22, 1120–1130.

Henkelman, R.M., 2010. Systems biology through mouse imaging centers: experience and new directions. Annu. Rev. Biomed. Eng. 12, 143–166.

Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C., Schlaug, G., 2009. Musical training shapes structural brain development. J. Neurosci. 29, 3019–3025.

Iosifescu, D.V., Shenton, M.E., Warfield, S.K., Kikinis, R., Dengler, J., Jolesz, F.A., McCarley, R.W., 1997. An automated registration algorithm for measuring MRI subcortical brain structures. NeuroImage 6, 13–25.

Kaidanovich-Beilin, O., Lipina, T.V., Takao, K., Van Eede, M., Hattori, S., Laliberté, C., Khan, M., Okamoto, K., Chambers, J.W., Fletcher, P.J., MacAulay, K., Doble, B.W., Henkelman, M., Miyakawa, T., Roder, J., Woodgett, J.R., 2009. Abnormalities in brain structure and behavior in GSK-3alpha mutant mice. Mol. Brain 2, 35.

Kale, S.C., Lerch, J.P., Henkelman, R.M., Chen, X.J., 2008. Optimization of the SNR-resolution tradeoff for registration of magnetic resonance images. Hum. Brain Mapp. 29, 1147–1158.

Karacali, B., Davatzikos, C., 2006. Simulation of tissue atrophy using a topology preserving transformation model. IEEE Trans. Med. Imaging 25, 649–652.

Kim, J., 2008. Structural consequences of diffuse traumatic brain injury: a large deformation tensor-based morphometry study. NeuroImage 39, 1014–1026.

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786–802.

Kovačević, N., Henderson, J.T., Chan, E., Lifshitz, N., Bishop, J., Evans, A.C., Henkelman, R.M., Chen, X.J., 2005. A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability. Cereb. Cortex 15, 639–645.

Lau, J.C., Lerch, J.P., Sled, J.G., Henkelman, R.M., Evans, A.C., Bedell, B.J., 2008. Longitudinal neuroanatomical changes determined by deformation-based morphometry in a mouse model of Alzheimer's disease. NeuroImage 42, 19–27.

Lerch, J.P., Carroll, J.B., Spring, S., Bertram, L.N., Schwab, C., Hayden, M.R., Mark Henkelman, R., 2008. Automated deformation analysis in the YAC128 Huntington disease mouse model. NeuroImage 39, 32–39.

Lerch, J.P., Yiu, A.P., Martinez-Canabal, A., Pekar, T., Bohbot, V.D., Frankland, P.W., Henkelman, R.M., Josselyn, S.A., Sled, J.G., 2011a. Maze training in mice induces MRI-detectable brain shape changes specific to the type of learning. NeuroImage 54, 2086–2095.

Lerch, J.P., Sled, J.G., Henkelman, R.M., 2011b. MRI phenotyping of genetically altered mice. In: Modo, M., Bulte, J.W.M. (Eds.), Magnetic Resonance Neuroimaging, Methods in Molecular Biology. Humana Press, pp. 349–361.

Lerch, J.P., Gazdzinski, L., Germann, J., Sled, J.G., Henkelman, R.M., Nieman, B.J., 2012. Wanted dead or alive? The tradeoff between in-vivo versus ex-vivo MR brain imaging in the mouse. Front. Neuroinform. 6.

Loken, C., Gruner, D., Groer, L., Peltier, R., Bunn, N., Craig, M., Henriques, T., Dempsey, J., Yu, C.-H., Chen, J., Dursi, L.J., Chong, J., Northrup, S., Pinto, J., Knecht, N., Zon, R.V., 2010. SciNet: Lessons Learned from Building a Power-efficient Top-20 System and Data Centre. J. Phys. Conf. Ser. 256, 012026.

Ma, Y., Hof, P.R., Grant, S.C., Blackband, S.J., Bennett, R., Slatest, L., McGuigan, M.D., Benveniste, H., 2005. A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. Neuroscience 135, 1203–1215.

Mansouri, A., Min, W., Cole, C.J., Josselyn, S.A., Henderson, J.T., Van Eede, M., Henkelman, R.M., Ackerley, C., Grunebaum, E., Roifman, C.M., 2012. Cerebellar abnormalities in purine nucleoside phosphorylase deficient mice. Neurobiol. Dis. 47, 201–209.

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. Am. J. Med. Genet. B Neuropsychiatr. Genet. 150B, 1085–1099.

Nieman, B.J., Wong, M.D., Henkelman, R.M., 2011. Genes into geometry: imaging for mouse development in 3D. Curr. Opin. Genet. Dev. 21, 638–646.

North, D.O., 1963. An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems. Proc. IEEE 51, 1016–1027.

Pieperhoff, P., Südmeyer, M., Hömke, L., Zilles, K., Schnitzler, A., Amunts, K., 2008. Detection of structural changes of the human brain in longitudinally acquired MR images by deformation field morphometry: methodological analysis, validation and application. NeuroImage 43, 269–287.

Pitiot, A., Pausova, Z., Prior, M., Perrin, J., Loyse, N., Paus, T., 2007. Magnetic resonance imaging as a tool for in vivo and ex vivo anatomical phenotyping in experimental genetic models. Hum. Brain Mapp. 28, 555–566.

Quarantelli, M., Larobina, M., Volpe, U., Amati, G., Tedeschi, E., Ciarmiello, A., Brunetti, A., Galderisi, S., Alfano, B., 2002. Stereotaxy-based regional brain volumetry applied to segmented MRI: validation and results in deficit and nondeficit schizophrenia. NeuroImage 17, 373–384.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44, 83–98.

Spring, S., Lerch, J.P., Henkelman, R.M., 2007. Sexual dimorphism revealed in the structure of the mouse brain using three-dimensional magnetic resonance imaging. NeuroImage 35, 1424–1433.

Spring, S., Lerch, J.P., Wetzel, M.K., Evans, A.C., Henkelman, R.M., 2010. Cerebral asymmetries in 12-week-old C57Bl/6J mice measured by magnetic resonance imaging. NeuroImage 50, 409–415.

Tao, G., 2009. Deep gray matter atrophy in multiple sclerosis: a tensor based morphometry. J. Neurol. Sci. 282, 39–46.

Wheeler, A.L., Lerch, J.P., Chakravarty, M.M., Friedel, M., Sled, J.G., Fletcher, P.J., Josselyn, S.A., Frankland, P.W., 2013. Adolescent cocaine exposure causes enduring macroscale changes in mouse brain structure. J. Neurosci. 33, 1797–1803.

Yanovsky, I., Leow, A.D., Lee, S., Osher, S.J., Thompson, P.M., 2009. Comparing registration methods for mapping brain change using tensor-based morphometry. Med. Image Anal. 13, 679–700.

Yassa, M.A., Stark, C.E.L., 2009. A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. NeuroImage 44, 319–327.