Comparison of template registration methods for multi-site meta-analysis of brain morphometry

Joshua Faskowitz[1], Greig I. de Zubicaray[2], Katie L. McMahon[3], Margaret J. Wright[4], Paul M. Thompson[1], Neda Jahanshad[1]

[1]Imaging Genetics Center, Keck School of Medicine of USC, Marina del Rey, CA, USA
[2]Queensland University of Technology, Brisbane, Australia
[3]Center for Advanced Imaging, University of Queensland, Brisbane, Australia
[4]Queensland Brain Institute, University of Queensland, Brisbane, Australia

ABSTRACT

Neuroimaging consortia such as ENIGMA can significantly improve power to discover factors that affect the human brain by pooling statistical inferences across cohorts to draw generalized conclusions from populations around the world. Voxelwise analyses such as tensor-based morphometry also allow an unbiased search for effects throughout the brain. Even so, such consortium-based analyses are limited by a lack of high-powered methods to harmonize voxelwise information across study populations and scanners. While the simplest approach may be to map all images to a single standard space, the benefits of cohort-specific templates have long been established. Here we studied methods to pool voxel-wise data across sites using templates customized for each cohort but providing a meaningful common space across all studies for voxelwise comparisons. As non-linear 3D MRI registrations represent mappings between images at millimeter resolution, we need to consider the reliability of these mappings. To evaluate these mappings, we calculated test-retest statistics on the volumetric maps of expansion and contraction. Further, we created study-specific brain templates for ten T1-weighted MRI datasets, and a common space from four study-specific templates. We evaluated the efficacy of using a two-step registration framework versus a single standard space. We found that the two-step framework more reliably mapped subjects to a common space.

Keywords: Multi-site, voxelwise, tensor-based morphometry, test-retest reliability

1. INTRODUCTION

Neuroimaging consortia such as ENIGMA (enigma.ini.usc.edu) [1] can boost power to discover subtle biological or clinical correlates of brain measures by pooling information across cohorts worldwide. Because of the scale of the datasets (tens of thousands of MRIs) [2], current efforts have successfully identified effects of single nucleotide polymorphisms (SNPs) that explain less than 0.5% of the variance in brain measures [3]. These studies use automatic segmentation tools to produce volume, thickness, and surface area measurements for regions of anatomical interest [4]. Measuring the brain in *a priori* regions can help in interpreting results while maintaining statistical power. However, this approach could limit searches of neuroimaging phenotypes to a handful of pre-defined features such as hippocampal or other subcortical volumes, and patterns of effects throughout the brain may be overlooked.

Voxelwise analyses such as tensor-based morphometry (TBM) allow for an unbiased search over the entire brain, and contrary to the handful of features often selected in region of interest (ROI) analyses, the number of features examined in TBM is on the order of millions. TBM is a sensitive approach to monitor longitudinal changes related to development and aging in health and disease [5], and cross sectional TBM studies can reveal the extent to which a certain trait or condition has an effect on brain volume and whether the effect is local or diffuse. Genetic studies are also common with TBM. Voxelwise genome-wide association studies have been proposed previously [6, 7], but were implemented for one or two specific dataset with limited statistical power for discovery. As with ROI analyses, for subtle genetic effects there is therefore a need to pool voxelwise TBM data from multiple sites to achieve power and adequate sample sizes. However, it is not immediately clear how to relate these features and spatially normalize data across multiple sites, imaging acquisitions and datasets [8-10]. Inaccurate alignment across sites may further limit the power to detect localized effects.

Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging,
edited by Barjor Gimi, Andrzej Krol, Proc. of SPIE Vol. 9788, 978822 · © 2016 SPIE
CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2217370

Here, we are developing a multistage framework to conduct voxel-wise meta-analysis of morphometry and other brain maps, to be distributed across sites in the ENIGMA consortium [11, 12]. In this framework, independent sites perform cross-sectional TBM with a site-specific MDT, yielding "Jacobian determinant" maps that localize volumetric variability within their cohort. Statistical hypotheses are tested image-wide across all voxels at each respective site.

Each site-specific template is then non-linearly registered to our ENIGMA-Mega template—a template made from multiple site-specific templates to normalize all site statistics into a single space [11]. The statistical maps from each site are aligned to this mega template common space by a second round of registrations, and are then in a common space in which meta-analysis of the statistical maps can occur (see **Figure 1**).

As each template may be derived from images taken with different acquisition parameters, field strengths, head coils, and image resolutions, ideally the parameters underlying the warps should be chosen to yield reliable registration, across multiple sites and cohorts. To better normalize T1-weighted (T1w) images across different sites, we proposed in earlier work [11, 12] to add additional target channels (gray matter ribbon mask and subcortical structure mask) to help further drive the nonlinear registration between each image and template. In line with our prior efforts in ENIGMA [2, 13, 14], we expect that sites would already have these parcellations quality controlled, and we have determined through these studies that these segmentations and parcellations are adequate in finding group-level meta-analyzed significance. Thus, our motivation for including additional channels is to enhance registration alignment with *a priori* high-quality anatomic information.

In this work, we perform one of the most comprehensive test-retest reliability estimates across 9 imaging studies, each with a unique population of individuals, different age ranges, and most importantly, different imaging acquisition parameters, to narrow down an optimal multisite registration scheme for harmonized multi-site voxelwise meta-analysis.

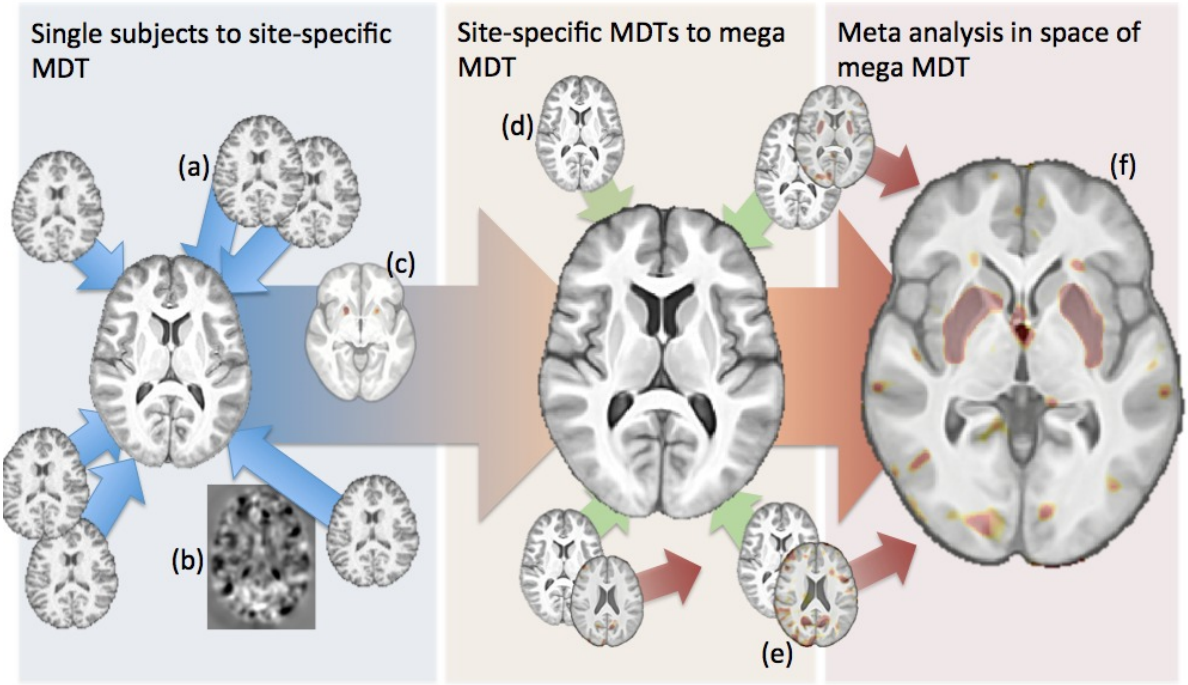

Figure 1. Diagram of our multi-site tensor-based morphometry framework. (a) Individual T1w images are non-linearly registered to a site-specific template. (b) Each subject to MDT registration yields a Jacobian map that quantifies volumetric expansion and contraction at each voxel from the subject to the MDT. (c) Each site participating in the framework runs voxelwise statistics on the Jacobian maps to render statistical maps in the space of the MDT. (d) Each site-specific MDT is non-linearly registered to a ENIGMA-Mega MDT. (e) Statistical maps in the space of each site-specific MDT are transformed by the non-linear mapping between site-specific MDT and ENIGMA-Mega MDT and are interpolated linearly. (f) Meta-analysis of statistics can occur in the ENIGMA-Mega MDT space.

2. METHODS

2.1. Cohorts and image acquisitions

MRI datasets from eleven cohorts were analyzed. Cohorts included: the first and second phases of the Alzheimer's Disease Neuroimaging Initiative (ADNI_1, ADNI_2) [15], the Queensland Twin Imaging Study (QTIM), Brain Genomics Superstruct Project (BrainGSP) [16], and multiple datasets from the Consortium for Reliability and Reproducibility (CoRR) [17] including: Beijing Normal University 1 and 2 (BNU_1, BNU_2), Mind Research Network (MRN), University of McGill (UM_1), New York University (NYU_2) and University of Pittsburgh School of Medicine (UPSM).

These cohorts cover a range of ages (18-30 years, to late adulthood, 60-85) and were collected on scanners around the world. These datasets provide us with a range of scan qualities (information in **Table 1**); image voxels range in size (0.9x0.9x0.9-1.25x1.25x2.0 mm) as do magnetic field strengths of the scanners used (1.5-4T). These datasets were also selected as they contained test-retest scanning sessions.

Table 1. Demographic information for datasets used. Note: if applicable, only healthy controls and non-related individuals in the dataset were analyzed. *Measures for subjects in test-retest subsample, if applicable.

Dataset	Image pairs for test-retest	Images for label overlap	Images pairs for metric comparison	Magnet strength	Voxel dimension	Age (mean; std dev; range)*	% Female*	Time between test-retest
ADNI_1	N/A	172	N/A	1.5/3.0 (multiple)	1.2x1.0x1.0mm^3	75.88; 5.21; 59.9-89.6	50%	N/A
ADNI_2	40	196	160	3T (multiple)	1.2x1.0x1.0mm^3	65.92; 2.68; 56.3-69.3	43%	3 months
BNU_1	40	57	N/A	3T (Siemens TrioTim)	1.33x1x1mm^3	23.1; 2.46; 19-30	53%	mean: 40.9
BNU_2	40	N/A	N/A	3T (Siemens TrioTim)	1.33x1x1mm^3	21.4; 0.8; 19.5-23.3	45%	mean: 157.6 days
BrainGSP	40	N/A	N/A	3T (Siemens TrioTim)	1.2x1.2x1.2mm^3	21.45; 2.7; 19-29	50%	mean: 77.3 days
MRN	40	48	N/A	3T (Siemens TrioTim)	1x1x1mm^3	25.00; 11.9; 10-53	43%	mean: 109 days
NYU_2	40	185	N/A	3T (Siemens Allegra)	1.33x1x1mm^3	22.10; 12.80; 7.5-53.0	45%	Same day
QTIM	40	40	40	4T (Bruker)	0.9x0.9x0.9mm^3	22.83; 2.48; 20.4-28.6	33%	3 months
UM_1	40	80	N/A	3T (Siemens TrioTim)	1x1x1mm^3	65.85; 6.86; 58-84	63%	mean: 111.4 days
UPSM_1	40	99	N/A	3T (Siemens TrioTim)	1x1x1mm^3	15.1; 2.87; 10.1-18.9	50%	mean: 624.4 days
Total:	**360**	**877**				**31.4; 19.72; 7.55-84**	**47%**	

2.2. Data preprocessing steps

All images were processed with previously tested protocols available online at http://enigma.ini.usc.edu/ongoing/protocols/. All T1-weighted (T1w) brain images were processed with FreeSurfer (version 5.3; freesurfer.net/fswiki) [4] to obtain subcortical and cortical parcellations. ENIGMA quality control protocols were implemented to flag outliers via visual inspection. The ENIGMA visual quality control protocol produced four coronal and four horizontal slices of FreeSurfer labels overlaid on a brain-extracted image (*brain.mgz*; Figure 2). The brain was extracted from the bias-field inhomogeneity corrected output (*nu.mgz*) of FreeSurfer using a cortical parcellation derived mask.

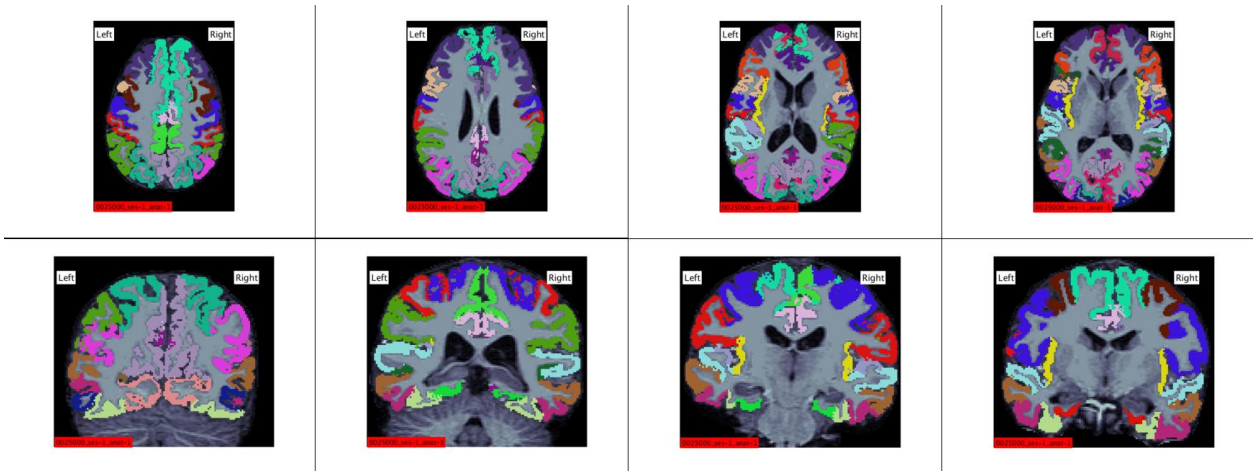

Figure 2. Shown is an example of the ENIGMA visual quality control for FreeSurfer parcellations from the NYU_2 cohort. Note that four slices in each dimension give a general idea of segmentation quality, but does not cover all parcellation labels of the *aparc+aseg.mgz*.

2.3 Template creation

For our multi-site tensor-based morphometry framework, we use the ANTs Symmetric Normalization (SyN) algorithm [18] for nonlinear registration, as it has been shown to be robust [19]. It is built on the widely used Insight ToolKit (ITK) [20], and is open source and freely available. Each of the cohort MDTs was constructed using the Advanced Normalization Tools (ANTs; stnava.github.io/ANTs/) software package and accompanying scripts (at commit: 88276f8). Approximately 24-30 scans per cohort were used to create each template. We used the three-channel (T1-weighed contrast, cortical ribbon, subcortical parcellations) registration approach described before [11]. The T1-weighted channel was the primary driving force of the registrations, with the highest weight. A meta-population template was created in our previous work from multiple datasets. Parameters used to create this template were identical to those used for site-specific templates. Not all sites were included in the meta-population template as in the case of the consortia, new groups are continually joining, and occasionally leaving, and it would be impractical to continually create new overall templates corresponding to groups currently in the consortium. All templates were created in the same coordinate system of the MNI 152 1mm template.

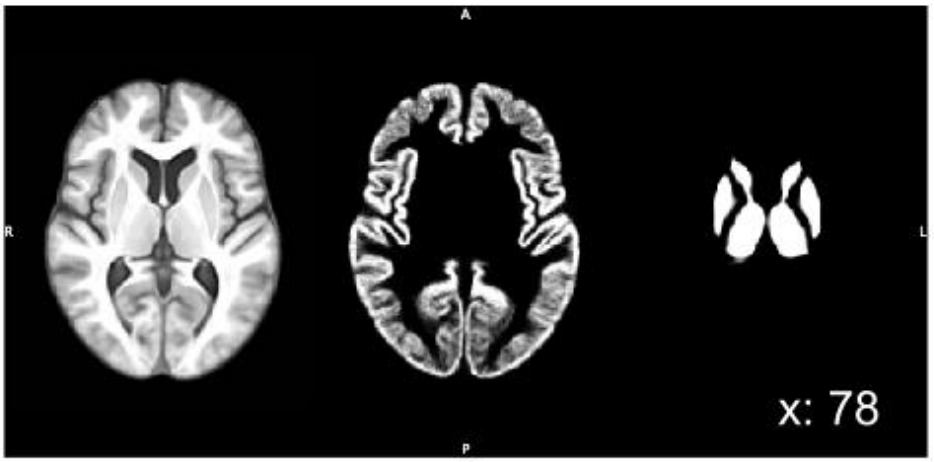

Figure 3. Example (from the ADNI_2 cohort) of the three channels of a multi-channel MDT.

2.4 Multi-site voxelwise Test-Retest Reliability

As the main contribution of this work, we seek quantitative measures of reliability and stability to aid in parameter selection across sites. We used antsRegistration.sh (github.com/stnava/ANTs/tree/master/Scripts) and observed the effects of modifying key aspects and parameters of the registration: the image similarity metric that drives the non-linear transformation, and the number of channels to use for the registration to the target.

- **Image similarity metric:** The image similarity metrics computed in the ANTs package are robust and fast [21], but template alignment can be based on a number of similarity metrics implemented in the ITK library. These options include: *mutual information*, based on the joint histogram entropy; *cross correlation*, based on the statistical similarity of image intensities; *mean squared difference* (MS), based on voxel intensity similarity; *Demons metric* (as implemented in ITK via ANTs), based on voxel intensity similarity. Additionally, these metrics can be combined and given weighting parameters.
- **Number and type of channels:** In our developing line of work, we have shown that using multiple channels for registration – adding the cortical ribbon and subcortical structure channels to the T1-weighted image – could improve correspondence of the TBM maps across cohorts; even so, it is unclear if multi-channel registration is also needed to warp cohort templates to an overall template, or whether this added computation does not significantly boost accuracy.

In the present analysis, we seek to evaluate two aspects of our multi-site morphometry framework: 1) the reliability of the "Jacobian-determinant" (Jacobian) maps for test-retest subjects in single channel versus multi channel registrations; 2) the label overlap in single channel versus multi channel registrations; 3) the reliability of Jacobian maps in registrations using different image similarity metrics.

2.4.1 Jacobian reliability evaluation

T1w images from each subject were run through an image-processing pipeline to produce Jacobian maps in the ENIGMA-Mega MDT space. Skull-stripped T1w images and corresponding FreeSurfer cortical parcellations were linearly aligned to a site-specific template using FSL's (fsl.fmrib.ox.ac.uk/fsl/) [22] *flirt* with 9 degrees of freedom. Then, T1w brains were nonlinearly registered to the site-specific template using ANT's *antsRegistration* using the Symmetric Normalization (SyN) transformation model [18] according to the parameters in **Table 2**. From this first registration we produced natural logged Jacobian maps characterizing the volumetric expansion and contraction at each voxel with ANT's *CreateJacobianDeterminantImage*. Each site-specific template was non-linearly registered to the ENIGMA-Mega MDT according to the parameters in **Table 2**. Each Jacobian map from the first registration was transformed into the ENIGMA-Mega MDT space according to the appropriate MDT to ENIGMA-Mega MDT transformation. Jacobian maps were interpolated linearly and smoothed with a Gaussian sigma of 2.

Table 2. Parameters used in the non-linear ANTs registrations.

antsRegistration parameter	Selected value	Annotation
-c, --convergence	[100x70x50x15,1e-6,5]	Four level registration with 100 iterations at the top (most smoothed) and 15 iterations at the last level (full-resolution); iterations at current level if the slope of the last 5 gradient descent values dips below convergence value of $1.0x10^{-6}$
-f, --shrink-factors	8x4x2x1	Image representation subsampled via ITK's ShrinkImageFilter by N value at each level
-g, --restrict-deformation	1x1x1	Non-linear deformation is not restricted in any dimension
-m, --metric	CC, demons, demons	Multi-channel: Cross correlation image similarity metric used for T1w channel, demons image similarity metric used for overlap of cortical grey matter and subcortical structure masks (masks are smoothed prior to *antsRegistration* call); weights for each channel were set at: 1.0, 0.5, 0.2 Single-channel: Cross correlation image similarity metric used for T1w channel
-s, --smoothing-sigmas	3x2x1x0vox	Image representation smoothed by a sigma of N voxels at each level; at last (full-resolution) level there is no smoothing
-t, --transform	SyN[0.1,3,0]	Symmetric normalization with a gradient step size of 0.1
-u, --use-histogram-matching	1	Histogram match images before registration

Four different registration schemes were compared. The first three registrations follow our two-step registration scheme, with individual level templates and registration, followed by group level mapping, and we compare this to the more simplistic approach of mapping all subjects from all cohorts directly to a known template. The four schemes are:

1. ***single-single***: both registration steps used only the T1w image to drive the registration;
2. ***single-multi:*** T1w image registration to site-specific MDT, and then a multi-channel registration from site MDT to ENIGMA-Mega MDT
3. ***multi-multi:*** the three channel registration was used for both steps.
4. ***MNI152:*** a single registration to the MNI152 template brain (skipping the site-site specific MDT step)

Test-retest Jacobian maps in ENIGMA-Mega MDT space were compared voxelwise between time point one and time point two Jacobian maps. Intraclass correlation (ICC; one-way ANOVA fixed effects model) was computed at each voxel using R's *psych* package (personality-project.org/r/psych/) for each cohort (40 subjects).

Further, signal-to-noise ratio was computed at a voxelwise level. Standard deviation of the ICC fit was calculated for each cohort rendering an ICC standard deviation map. Each cohort's ICC map was divided by its standard deviation map, to render an SNR map. SNR maps were averaged across cohorts to render an average SNR map for each of the four registration schemes.

2.4.2 Label overlap computation

To evaluate registration label overlap, we measured label overlap of individual parcellations with ENIGMA-Mega MDT parcellations. Individual FreeSurfer parcellations (*aparc+aseg.mgz*) were transformed into the *ENIGMA-Mega -MDT* space according to the four registration schemes described previously: 1) multi-multi; 2) single-multi; 3) single-single; 4) MNI152. Parcellation maps on the ENIGMA-Mega MDT and MNI152 template were created using joint label fusion [23]. Joint label fusion involves using multiple representative atlases to label a target image. We employed an ANTs script (*antsJointLabelFusion.sh*) to nonlinearly register 136 brain images to the ENIGMA-Mega MDT and MNI152 template. The 136 subjects represent the individuals from the four datasets used to create

the site-specific MDTs that constructed the ENIGMA-Mega MDT [11]. Parcellations were mapped to the target space and a single parcellation map in the target space was derived using a label fusion approach that automatically determines optimal weights [24].

Label overlap between transformed individual labels and the template labels were quantified by the Dice overlap coefficient [25]. Dice coefficient overlap for two labels X and Y is defined by the following equation:

$$Dice = 2\frac{|X \cap Y|}{|X| + |Y|}$$

Dice coefficients were measured between each subject parcellation label and the corresponding label of the target image. Dice coefficients were then averaged across all labels to render an average Dice overlap measure. Additionally, average Dice overlap was computed for cortical labels and subcortical labels, respectively.

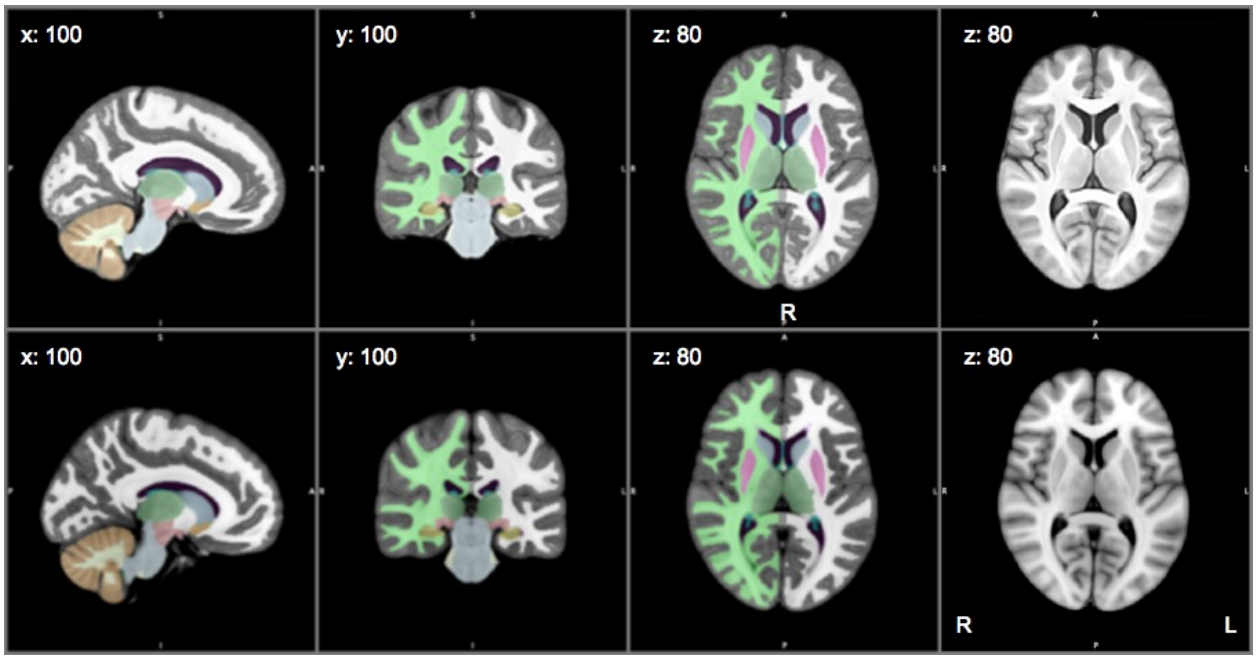

Figure 4. Visualization of the joint fusion labels made for the ENIGMA-Mega MDT (*top*) and MNI152 template (*bottom*). The mega MDT and MNI152 templates are shown without labels overlaid on far right. The ENIGMA-Mega MDT resembles the MNI152 as it is constructed from brain images initially linearly aligned to the MNI152 space.

2.4.3 Similarity metric comparison

Test-retest reliability of Jacobian maps was evaluated while modulating the image similarity cost computation of the non-linear registration. Each T1w skull-stripped brain was registered with the single T1w contrast channel to its appropriate site-specific MDT. This registration was performed using various image similarity cost metrics including: (1) mutual information; (2) cross correlation; (3) mean squared difference; (4) 50% mutual information, 50% cross correlation; (5) demons metric as implemented in ITK via ANTs [21, 26]. We tested the ICC on the unsmoothed test-retest Jacobian maps of each registration, as described in section 2.4.1.

3. RESULTS

3.1 Jacobian reliability evaluation

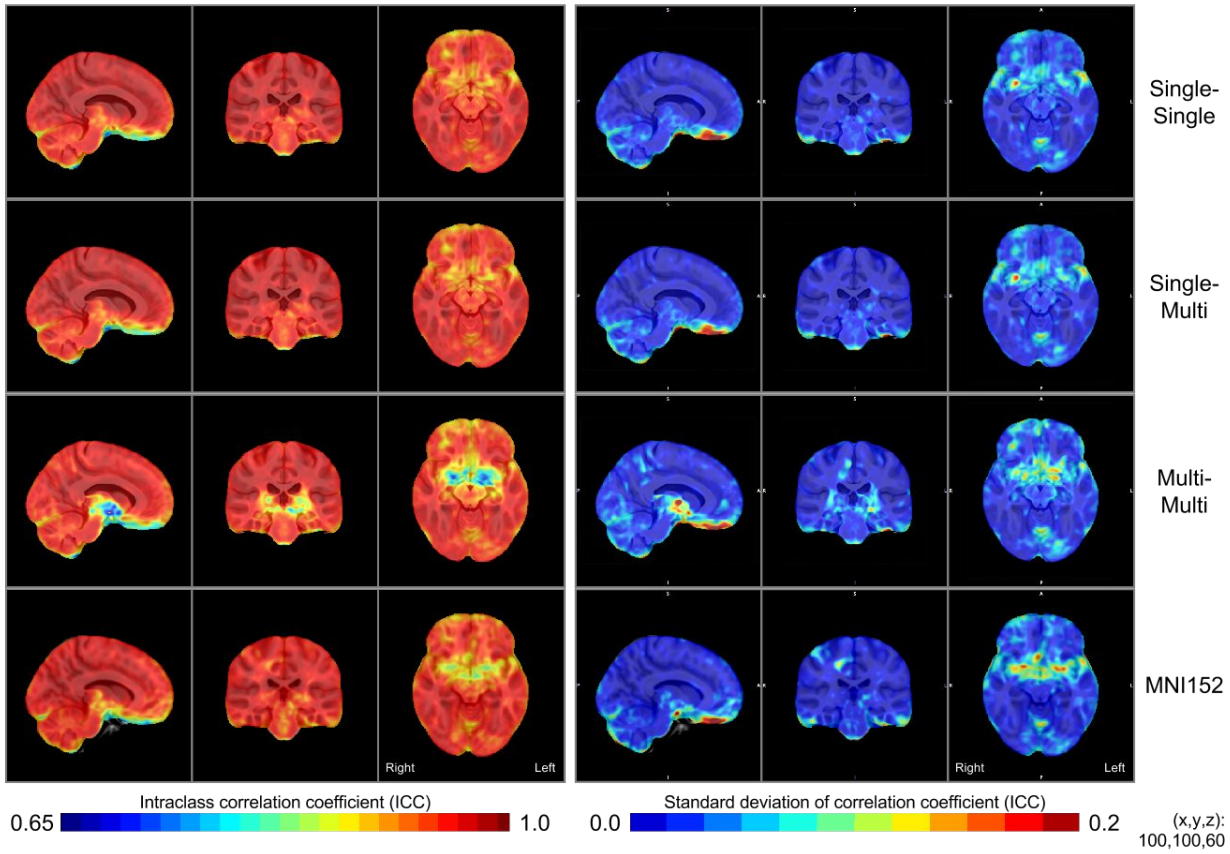

Figure 5. ICC maps were computed for nine cohorts, each comprised of 40 subjects. Average ICC and standard deviation of the ICC maps is visualized here. ICC standard deviation ranged from 0.0 to 0.32 across all registrations and was generally greatest in the medial orbitofronal region of the brain. As the ICC of both single-single and single-multi for each cohort were calculated after the registration to the single-channel cohort template, the maps appear similar (though not identical) after being warped with either the single-channel or the multi-channel registration scheme to the ENGIMA-Mega MDT, respectively.

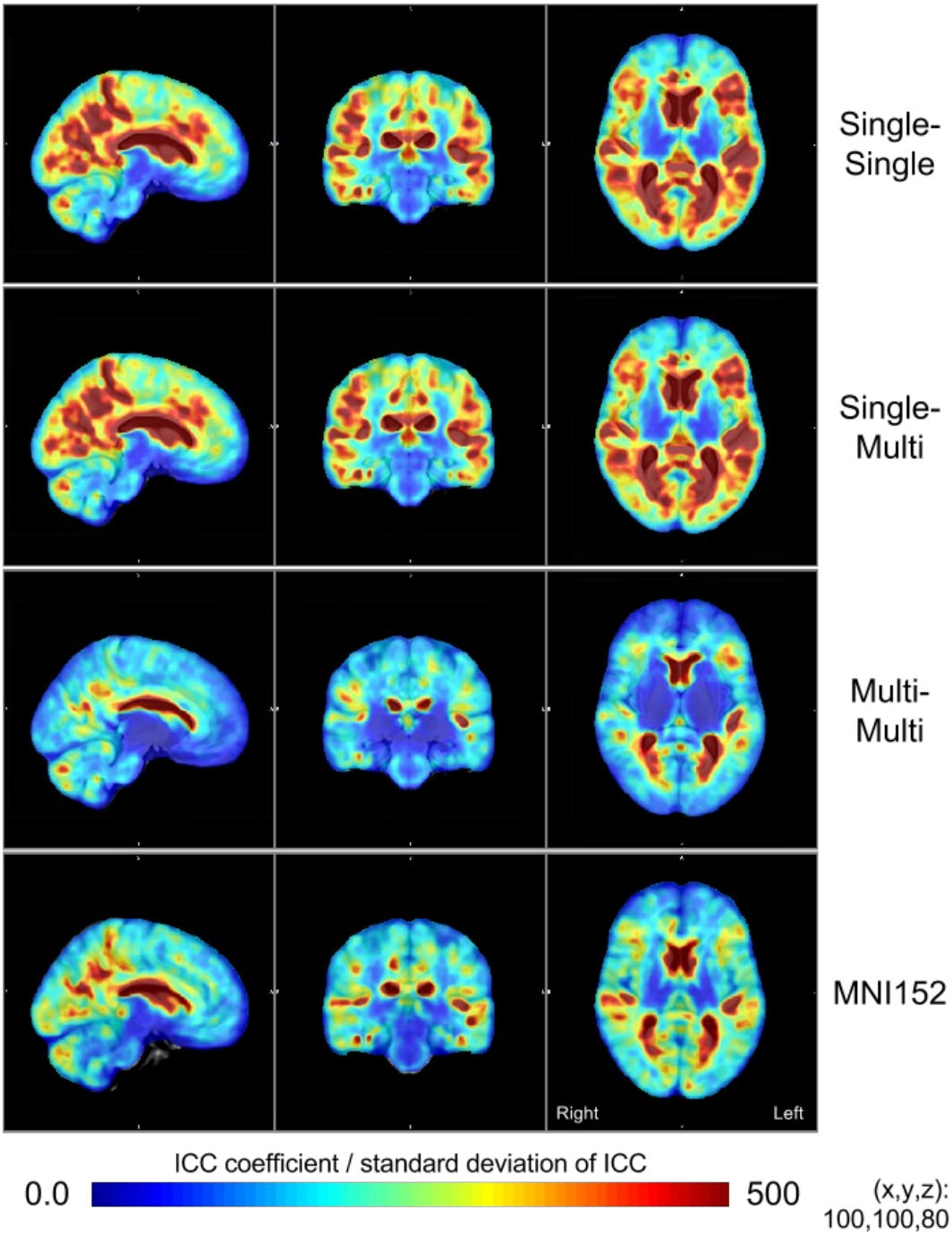

Figure 6. Shown here are the ICC signal-to-noise (SNR) maps, computed by dividing each cohort's ICC map by its respective standard deviation, and then averaged across the nine cohorts. SNR is generally the highest in the lateral ventricle areas to the target template. Registration to the ENIGMA-Mega MDT showed notably higher SNR that registration to the MNI152 template. SNR generally decreased ventrally. The single-multi and single-single registrations show high SNR along the cortical gray matter ribbon, while the multi-multi registration does not.

3.2 Label overlap evaluation

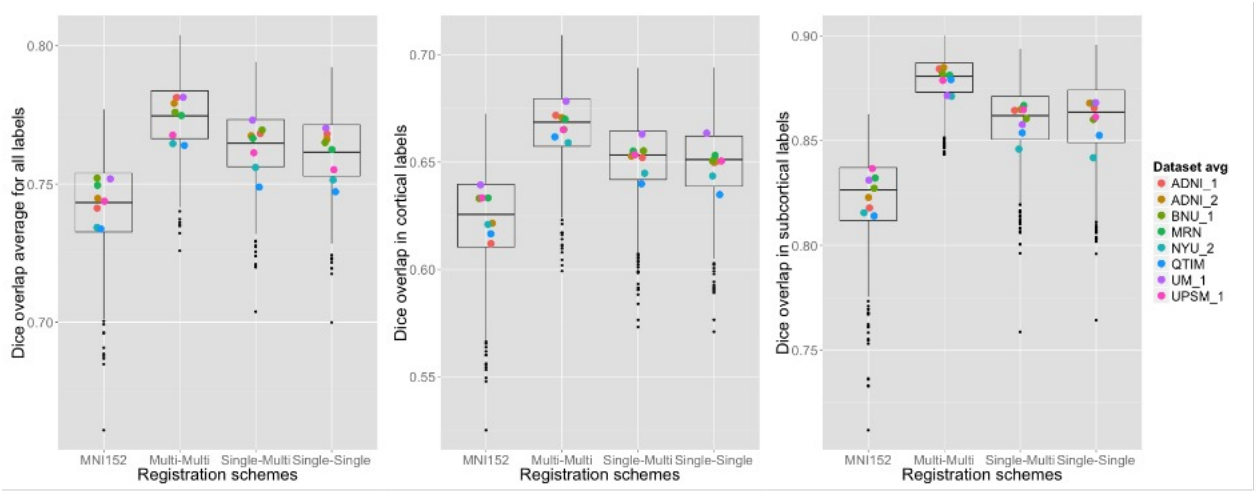

Figure 7. Boxplots showing the average label overlap between individual parcellations and parcellations in the ENIGMA-Mega MDT and MNI152 template space, for each registration scheme. Average cohort values are shown as colored dots overlaid on each boxplot.

The multi-multi registration scheme achieved the greatest Dice overlap on average in three separate comparisons: 1) average dice across all Freesurfer parcellation labels; 2) average Dice across all parcellation labels of the cortical grey matter mask; 3) average Dice coefficient across all parcellation labels of the subcortical structures.

3.3 Similarity metric evaluation

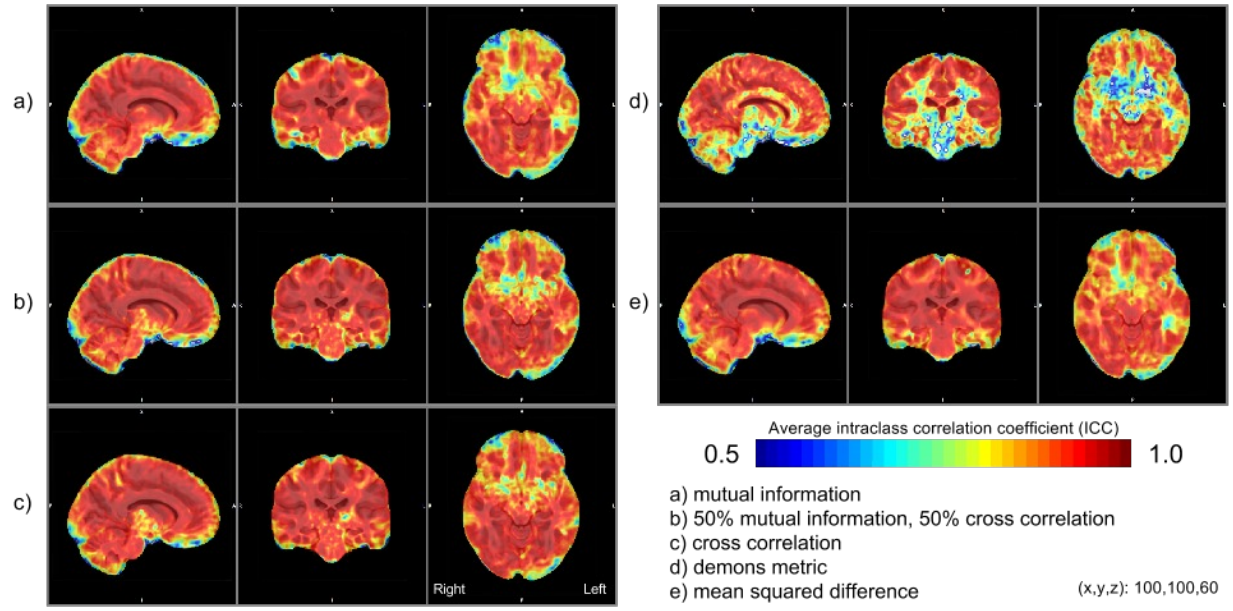

Figure 8. Voxelwise ICC of Jacobian maps in the QTIM cohort.

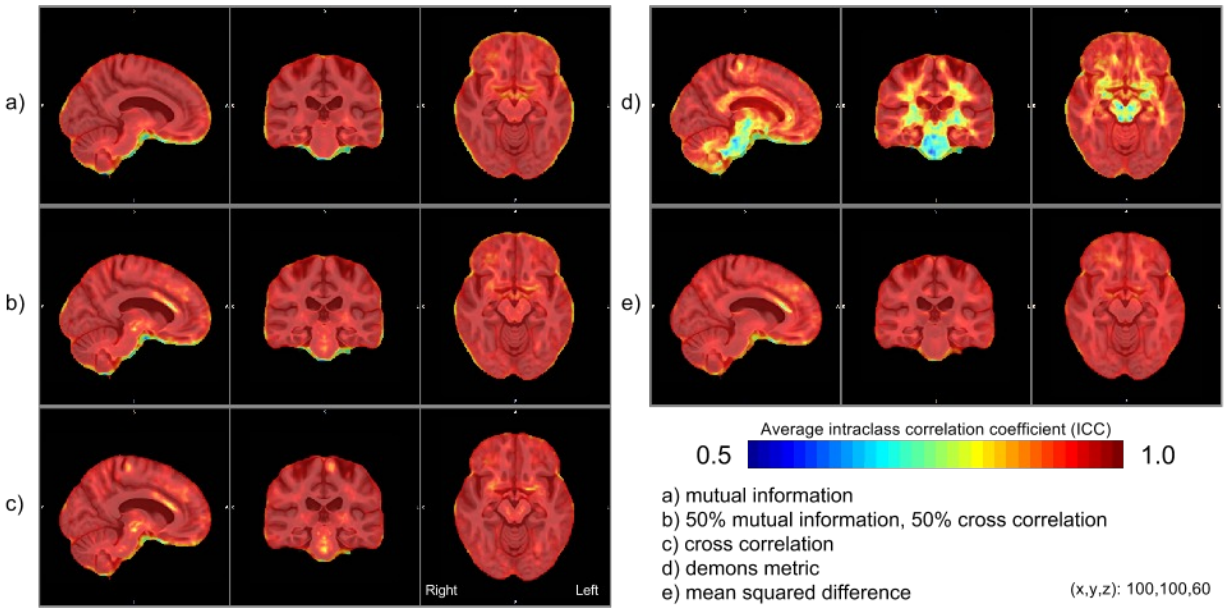

Figure 9. Voxelwise ICC of Jacobian maps in the ADNI_2 cohort.

In **Figures 7 and 8** we show the results of ICC after changing the image similarity cost metric for the Jacobian test-retest reliability. As expected, performance varies by dataset and by metric. Registration guided by the Demons metric performs comparatively worse; specifically, it results in lower ICC in white matter areas of the target template. Registrations guided by mutual information exhibit lower reliability near the edge of the template brain -- an area susceptible to errors in brain extraction. Registrations guided by cross correlation show slightly lower ICC around the subcortical structures and peripheral white matter regions.

4. DISCUSSION

To harmonize voxelwise processing protocols across cohorts for joint statistical analyses, we must ensure reliable estimates across cohorts in relation to the many parameters that can be adjusted and tuned for registration. Here we evaluate the performance of the non-linear registration tools used in our multi-site tensor-based morphometry (TBM) framework, which to the best of our knowledge is the first cross-cohort evaluation aimed to pooling multisite statistics in a stable manner. We find that a two step approach to normalizing TBM measures is more reliable than the alternative of using a general stereotaxic target such as the MNI152 template. The two-step approach also results in higher Dice coefficients of overlap when individual labels are transformed into the final space, where meta-analysis occurs. The benefit of using site-specific MDTs as we see here is corroborated by prior work highlighting the benefits of age-specific templates [27, 28].

We find that multi-multi registrations achieve the highest label overlap; this is generally expected, as the additional channels of the multi-channel registration are comprised of the same labels used in the Dice overlap analysis. While the Dice overlap measure was calculated by averaging each of the subcortical and cortical regions separately, and the registration channels were driven by 1) all subcortical regions merged and smoothed under one label and 2) all cortical regions together, there is still an inherent bias in the estimates. Although the multi-channel registration is only partially weighted by these channels (the T1w channel is weighted the most), they guide the overall registration enough to make Dice measures significantly higher. In future work, we will use a different software to calculate the parcellations for Dice overlap than the one used for the parcellations driving the registration (such as FSL's fast) to remove biases or circularity in estimating Dice overlap coefficients [29].

We also see that in these multi-multi registrations, the reliability of the voxelwise Jacobian maps near the subcortical structures is lower than in the registrations where a single channel is used within cohorts. The decreased voxelwise ICC could indicate inconsistent segmentation between test and retest scans, or that perhaps quality-control that was

not rigorous enough to include manual edits. Due to the large sample sizes in many cohorts, manual edits of automated segmentations may be time consuming, so we cannot assume all cohorts would perform rigorous correction of segmentations. Here, through our multi-multi analysis, we see the consequences of carrying forward these imperfections in an added channel of registration. Another reason for reduced SNR in the multi-multi scheme may be inconsistencies in automated segmentations from one time point to the next; FreeSurfer now has implemented a longitudinal scheme, which may provide more stable estimates for within subject segmentations [30]. This may in turn improve reliability estimates for voxelwise registration schemes using this prior information.

When evaluating the performance of various image similarity cost metrics, there was a general pattern of decreased reliability in the subcortical regions of the target template across all metrics. As expected, performance of these metrics can change considerably for different cohorts.

5. CONCLUSIONS

Reliability of the Jacobian maps differs across metrics and cohorts. As most of the TBM estimates across the parameters we tested are extremely reliable across 9 cohorts of various populations and imaging parameters, it is acceptable to keep one set of chosen metrics in our framework for multi-site voxelwise TBM meta-analysis. As expected, the two-step multi-site registration greatly outperforms that of a single step registration to a single MNI152 template. Future work will further evaluate the added benefit of incorporating additional anatomical information into the registration between the cohort specific MDT and the ENIGMA-Mega MDT.

ACKNOWLEDGMENTS

This study was funded in part by NIH ENIGMA Center grant U54 EB020403, supported by the Big Data to Knowledge (BD2K) Centers of Excellence program.

ADNI_1 & ADNI_2: Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

BrainGSP: Data were provided [in part] by the Brain Genomics Superstruct Project of Harvard University and the Massachusetts General Hospital, (Principal Investigators: Randy Buckner, Joshua Roffman, and Jordan Smoller), with support from the Center for Brain Science Neuroinformatics Research Group, the Athinoula A. Martinos Center for Biomedical Imaging, and the Center for Human Genetic Research. 20 individual investigators at Harvard and MGH generously contributed data to GSP Open Access Data Use Terms Version: 2014-Apr-22 the overall project.

CoRR: The National Institute on Drug Abuse (NIDA) and the National Natural Science Foundation of China (NSFC) have been instrumental in the CoRR collaboration providing the necessary funding and manpower to build the foundation of the project along with the Child Mind Institute, the Institute of Psychology, Chinese Academy of Sciences and the Nathan Kline Institute.

QTIM: This study was supported by grant T15 LM07356 from the NIH/National Library of Medicine, and Project Grant 496682 from the National Health and Medical Research Council, Australia. We are extremely grateful to the twins for their participation, to the radiographer, Matt Meredith, Centre for Magnetic Resonance, University of Queensland, for image acquisition, and research nurses, Marlene Grace and Ann Eldridge, Queensland Institute of Medical Research, for twin recruitment.

REFERENCES

[1] P. M. Thompson, J. L. Stein, S. E. Medland, D. P. Hibar, A. A. Vasquez, M. E. Renteria, *et al.*, "The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data," *Brain Imaging and Behavior,* vol. 8, pp. 153-182, 2014.

[2] D. P. Hibar, J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivières, N. Jahanshad, *et al.*, "Common genetic variants influence human subcortical brain structures," *Nature,* vol. 520, pp. 224-229, 2015.

[3] S. E. Medland, N. Jahanshad, B. M. Neale, and P. M. Thompson, "Whole-genome analyses of whole-brain data: working within an expanded search space," *Nature Neuroscience,* vol. 17, pp. 791-800, 2014.

[4] B. Fischl, "FreeSurfer," *Neuroimage,* vol. 62, pp. 774-781, 2012.

[5] X. Hua, D. P. Hibar, C. R. Ching, C. P. Boyle, P. Rajagopalan, B. A. Gutman, *et al.*, "Unbiased tensor-based morphometry: improved robustness and sample size estimates for Alzheimer's disease clinical trials," *Neuroimage,* vol. 66, pp. 648-661, 2013.

[6] J. L. Stein, X. Hua, S. Lee, A. J. Ho, A. D. Leow, A. W. Toga, *et al.*, "Voxelwise genome-wide association study (vGWAS)," *Neuroimage,* vol. 53, pp. 1160-1174, 2010.

[7] D. P. Hibar, J. L. Stein, O. Kohannim, N. Jahanshad, A. J. Saykin, L. Shen, *et al.*, "Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects," *Neuroimage,* vol. 56, pp. 1875-1891, 2011.

[8] N. K. Focke, G. Helms, S. Kaspar, C. Diederich, V. Tóth, P. Dechent, *et al.*, "Multi-site voxel-based morphometry—not quite there yet," *Neuroimage,* vol. 56, pp. 1164-1170, 2011.

[9] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, *et al.*, "Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data," *Neuroimage,* vol. 30, pp. 436-443, 2006.

[10] A. D. Leow, A. D. Klunder, C. R. Jack, A. W. Toga, A. M. Dale, M. A. Bernstein, *et al.*, "Longitudinal stability of MRI for mapping brain change using tensor-based morphometry," *Neuroimage,* vol. 31, pp. 627-640, 2006.

[11] N. Jahanshad, G. Roshchupkin, J. Faskowitz, D. P. Hibar, B. A. Gutman, H. H. Adams, *et al.*, "Multi-site meta-analysis of image-wide genome-wide associations of morphometry," in *MICCAI Imaging Genetics Workshop, 2015*, 2015.

[12] P. M. Thompson, O. A. Andreassen, A. Arias-Vasquez, C. E. Bearden, P. S. Boedhoe, R. M. Brouwer, *et al.*, "ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide," *Neuroimage,* 2015.

[13] L. Schmaal, D. J. Veltman, T. G. van Erp, P. Sämann, T. Frodl, N. Jahanshad, *et al.*, "Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group," *Molecular Psychiatry,* 2015.

[14] T. van Erp, D. Hibar, J. Rasmussen, D. Glahn, G. Pearlson, O. Andreassen, *et al.*, "Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium," *Molecular Psychiatry,* 2015.

[15] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging,* vol. 27, pp. 685-691, 2008.

[16] A. J. Holmes, M. O. Hollinshead, T. M. O'Keefe, V. I. Petrov, G. R. Fariello, L. L. Wald, *et al.*, "Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures," *Scientific data,* vol. 2, 2015.

[17] X.-N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautzik, *et al.*, "An open science resource for establishing reliability and reproducibility in functional connectomics," *Scientific Data,* vol. 1, 2014.

[18] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis,* vol. 12, pp. 26-41, 2008.

[19] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang*, et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *Neuroimage,* vol. 46, pp. 786-802, 2009.

[20] B. B. Avants, N. J. Tustison, M. Stauffer, G. Song, B. Wu, and J. C. Gee, "The Insight ToolKit image registration framework," *Frontiers in Neuroinformatics,* 2014.

[21] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *Neuroimage,* vol. 54, pp. 2033-2044, 2011.

[22] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage,* vol. 62, pp. 782-790, 2012.

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

[24] H. Wang, J. W. Suh, J. Pluta, M. Altinay, and P. Yushkevich, "Optimal weights for multi-atlas label fusion," in *Information Processing in Medical Imaging*, 2011, pp. 73-84.

[25] N. Tustison and J. Gee, "Introducing Dice, Jaccard, and other label overlap measures to ITK," *Insight J,* pp. 1-4, 2009.

[26] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis,* vol. 2, pp. 243-260, 1998.

[27] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins*, et al.*, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage,* vol. 54, pp. 313-327, 2011.

[28] U. Yoon, V. S. Fonov, D. Perusse, A. C. Evans, and B. D. C. Group, "The effect of template choice on morphometric analysis of pediatric brain data," *Neuroimage,* vol. 45, pp. 769-777, 2009.

[29] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, A. Hammers*, et al.*, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *Neuroimage,* vol. 51, pp. 221-227, 2010.

[30] M. Reuter and B. Fischl, "Avoiding asymmetry-induced bias in longitudinal image processing," *Neuroimage,* vol. 57, pp. 19-21, Jul 1 2011.