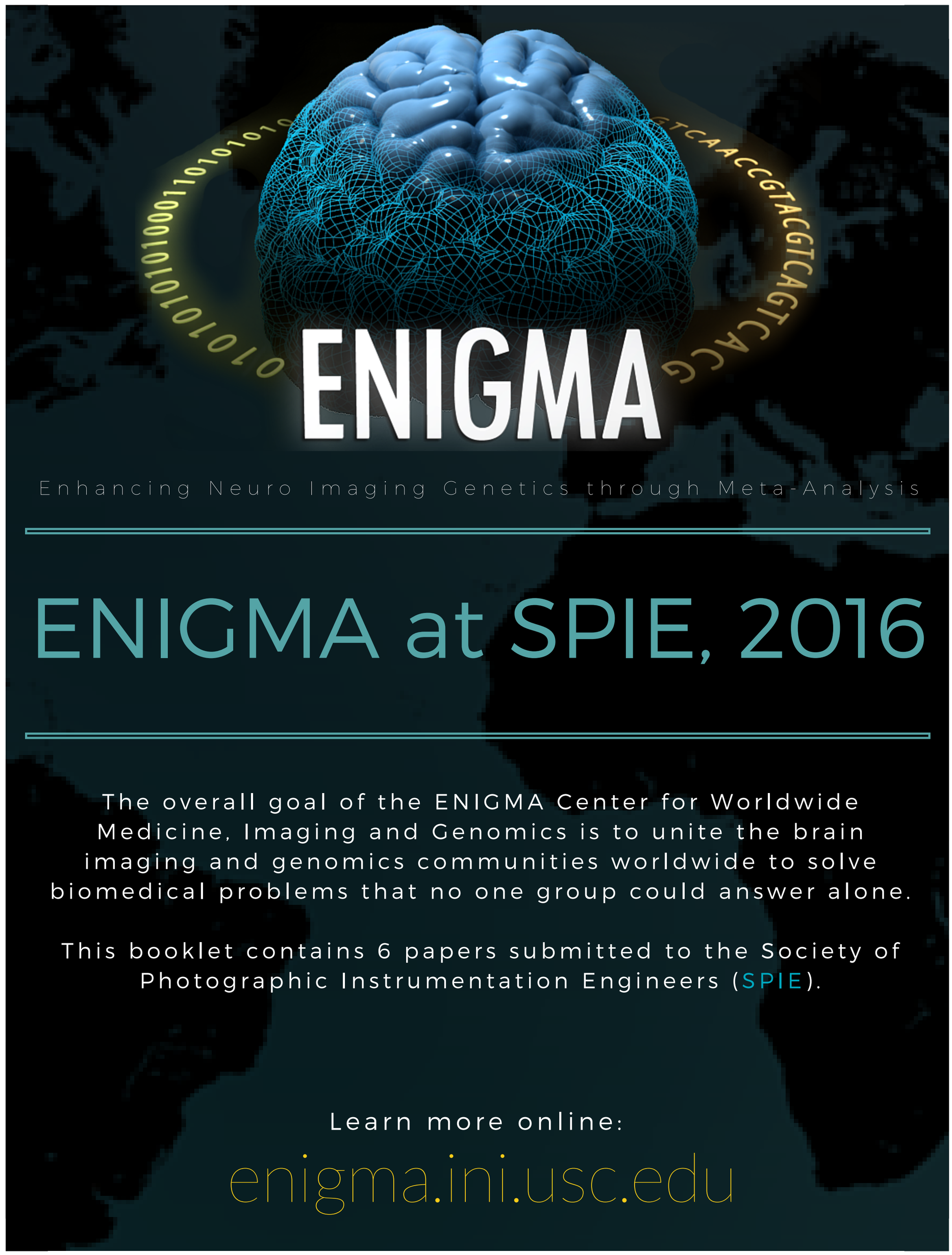


Enhancing Neuro Imaging Genetics through Meta-Analysis

ENIGMA at SPIE, 2016

The overall goal of the ENIGMA Center for Worldwide Medicine, Imaging and Genomics is to unite the brain imaging and genomics communities worldwide to solve biomedical problems that no one group could answer alone.

This booklet contains 6 papers submitted to the Society of Photographic Instrumentation Engineers (SPIE).

Learn more online:

enigma.ini.usc.edu

What is ENIGMA?

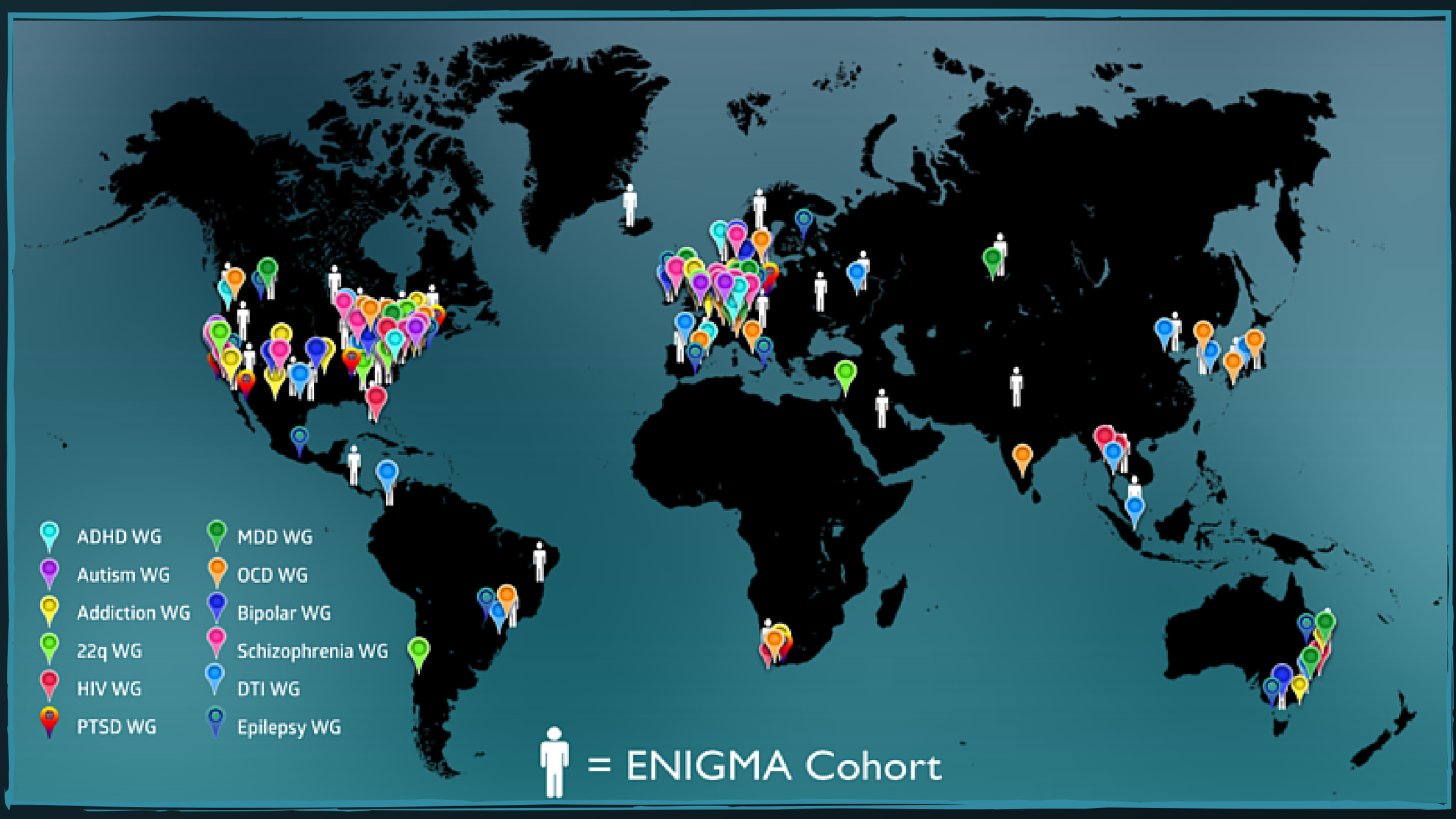
Paul Thompson

Principal Investigator, ENIGMA

Founded in 2009, ENIGMA is a worldwide network of over 500 scientists from 35 countries, working together to discover factors that help or harm the human brain. Pooling brain scans from over 50,000 people, ENIGMA has over 30 working groups, studying 12 major brain diseases. ENIGMA screens millions of letters of our genetic code to discover patterns in our DNA that affect brain function, integrity, and wiring. Awarded $11 million by the NIH in 2014, the ENIGMA “Big Data” Center of Excellence develops algorithms to crack the brain’s genetic code, developing new mathematics to discover what drives brain networks, and how our brain changes over the human lifespan, in 35+ countries worldwide.

Here we showcase 6 papers accepted for the Society of Photographic Instrumentation Engineers (SPIE) Medical Imaging conference, San Diego, California.

Each paper builds on the power of ENIGMA’s worldwide alliance to solve problems together in ways not previously dreamed of.



ENIGMA’s Worldwide Reach in Global Medical Graduate Training and Unprecedented Power. Experts from over 200 Institutions have come together to contribute to the ENIGMA Center for Worldwide Medicine, and achieve medical goals that each country could not achieve on its own.

CONTENTS

ENIGMA AT SPIE MEDICAL IMAGING, SAN DIEGO, CA

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.

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.

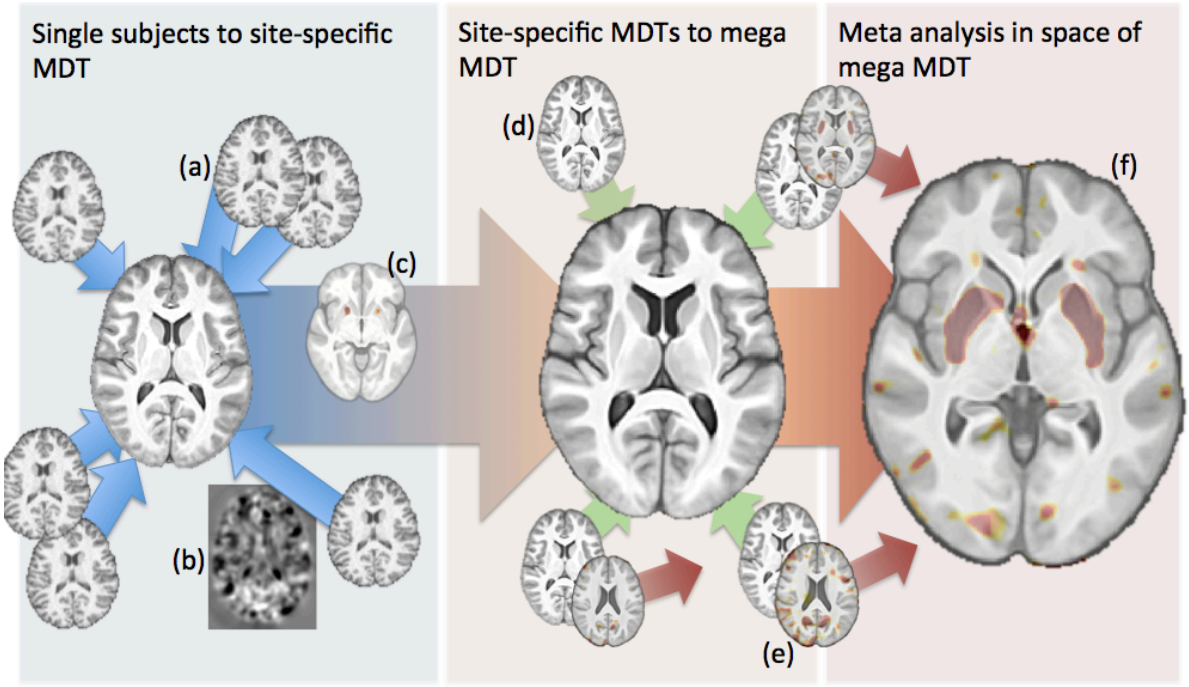


Figure 1. Diagram of our multi-site tensor-based morphometry framework. (a) Individual T1w images are non-linearly registered to a site-specific template. (b) Each subject to MDT registration yields a Jacobian map that quantifies volumetric expansion and contraction at each voxel from the subject to 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.

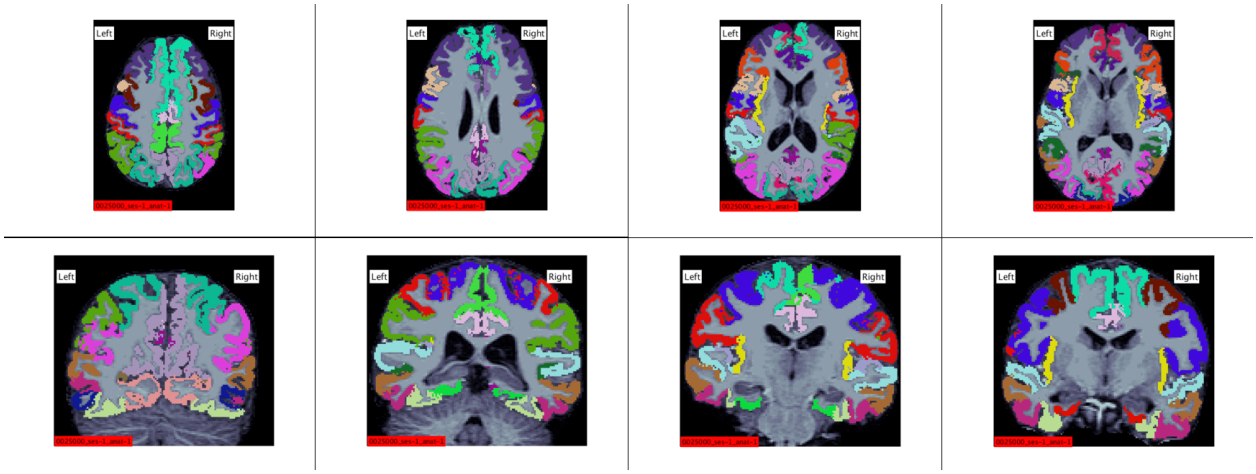


Figure 2. Shown is an example of the ENIGMA visual quality control for FreeSurfer parcellations from the NYU_2 cohort. Note that four slices in each dimension give a general idea of segmentation quality, but does not cover all parcellation labels of the *aparc+aseg.mgz*.

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.

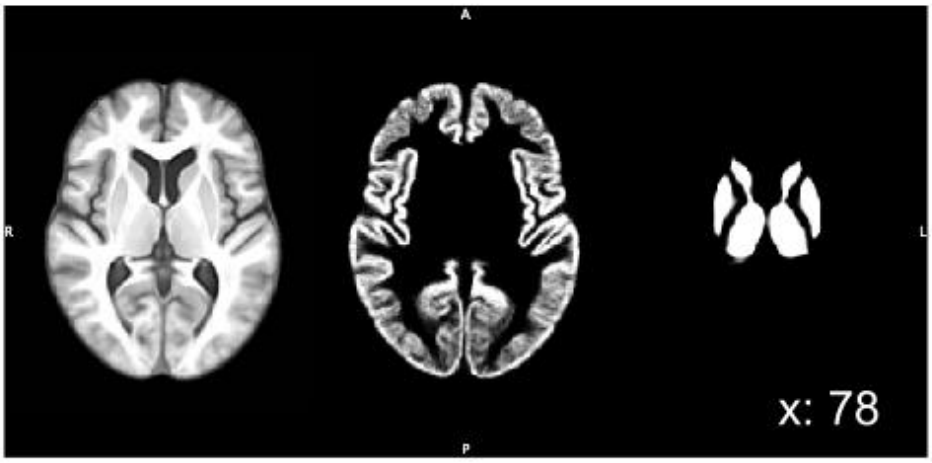


Figure 3. Example (from the ADNI_2 cohort) of the three channels of a multi-channel MDT.

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.

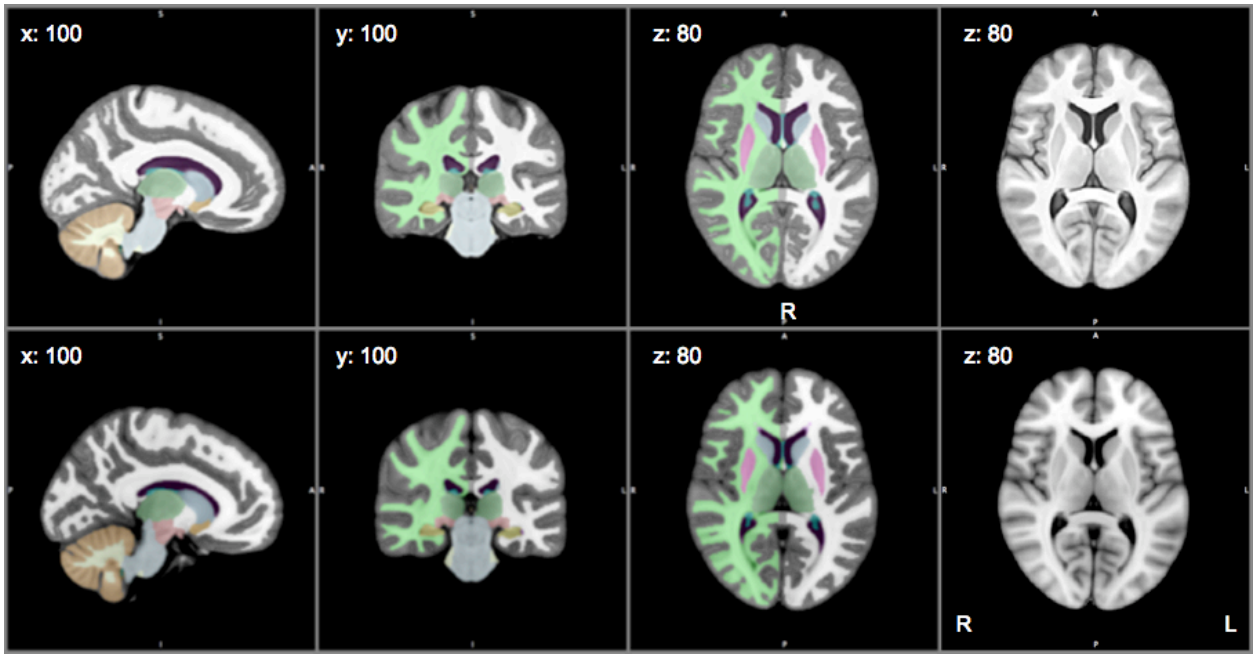


Figure 4. Visualization of the joint fusion labels made for the ENIGMA-Mega MDT (*top*) and MNI152 template (*bottom*). The mega MDT and MNI152 templates are shown without labels overlaid on far right. The ENIGMA-Mega MDT resembles the MNI152 as it is constructed from brain images initially linearly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

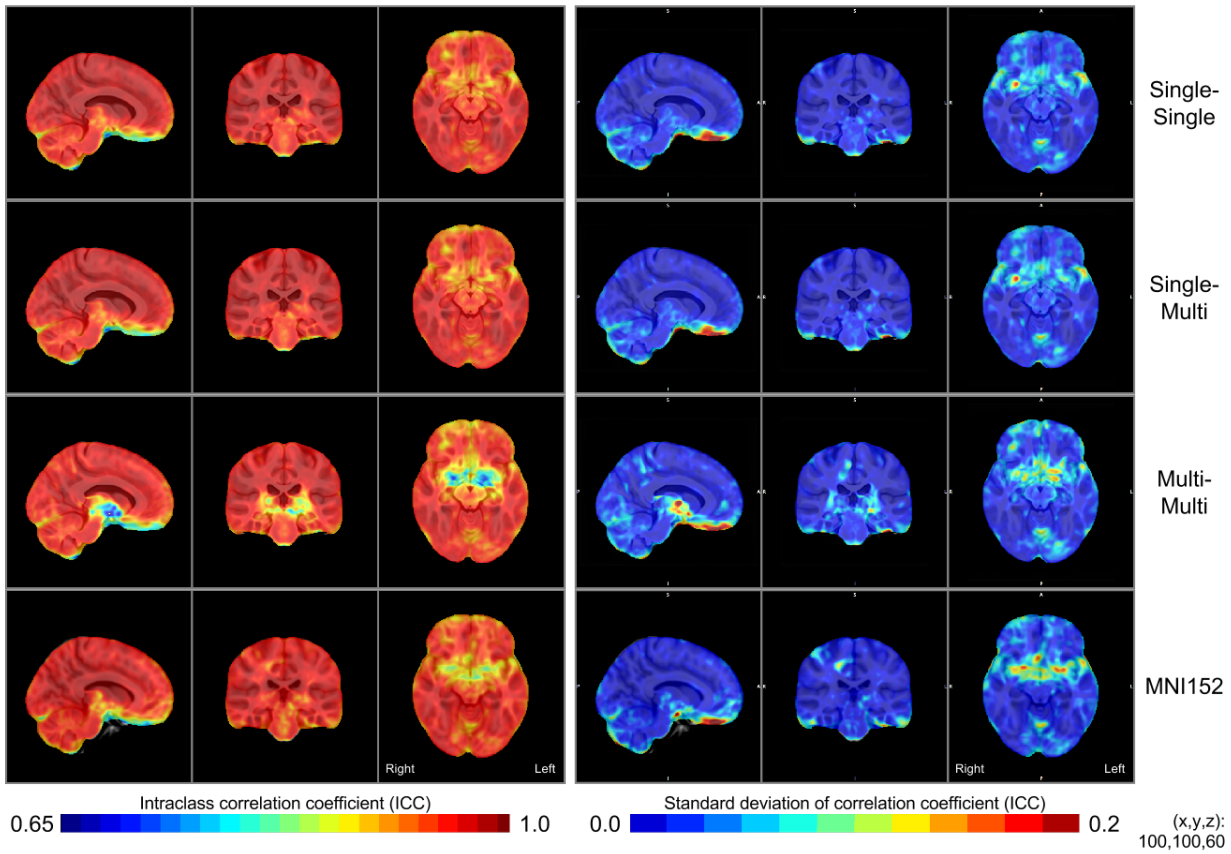


Figure 5. ICC maps were computed for nine cohorts, each comprised of 40 subjects. Average ICC and standard deviation of the ICC maps is visualized here. ICC standard deviation ranged from 0.0 to 0.32 across all registrations and was generally greatest in the medial orbitofronal region of the brain. As 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.

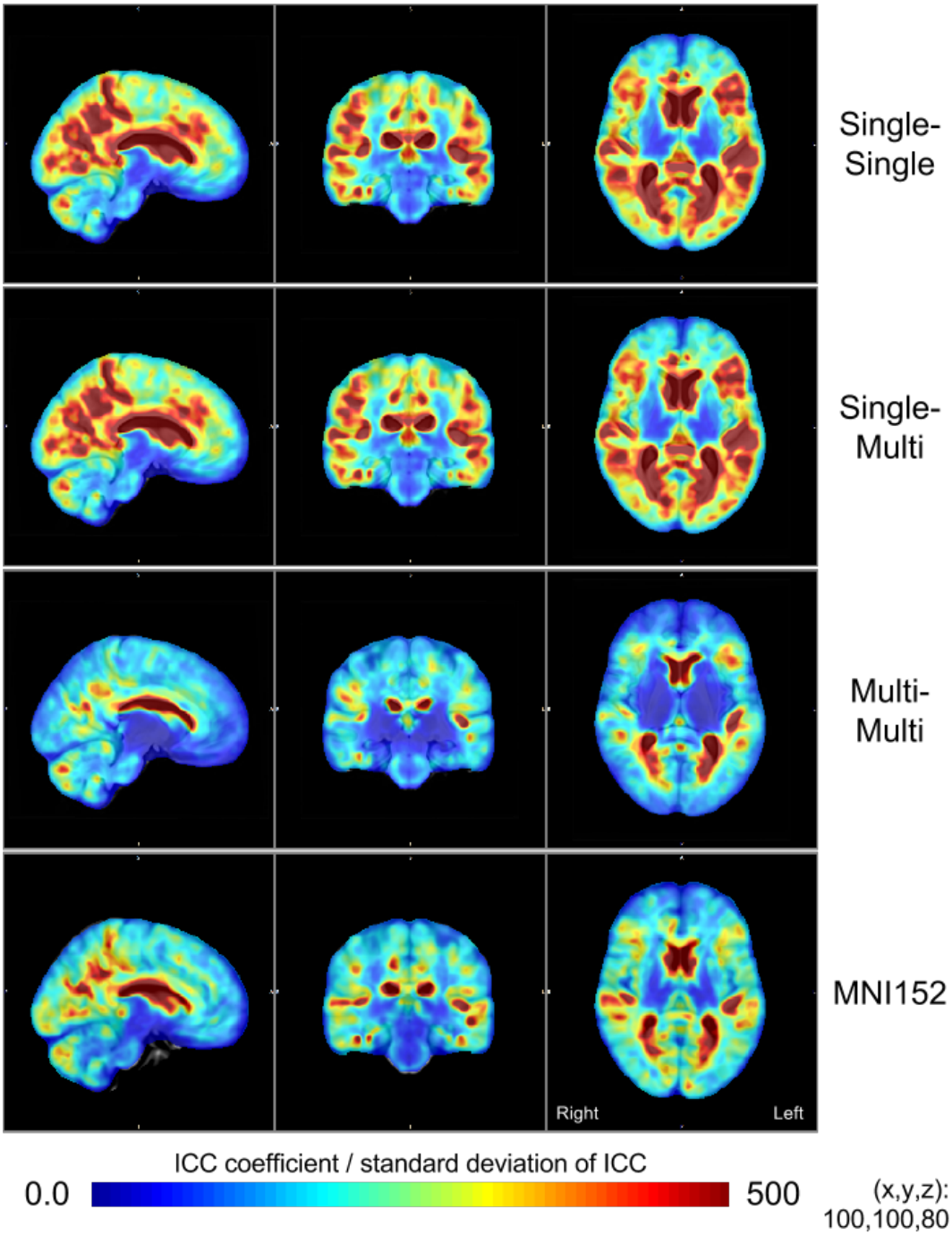


Figure 6. Shown here are the ICC signal-to-noise (SNR) maps, computed by dividing each cohort's ICC map by its respective standard deviation, and then averaged across the nine cohorts. SNR is generally the highest in the lateral ventricle areas to the target template. Registration to the ENIGMA-Mega 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

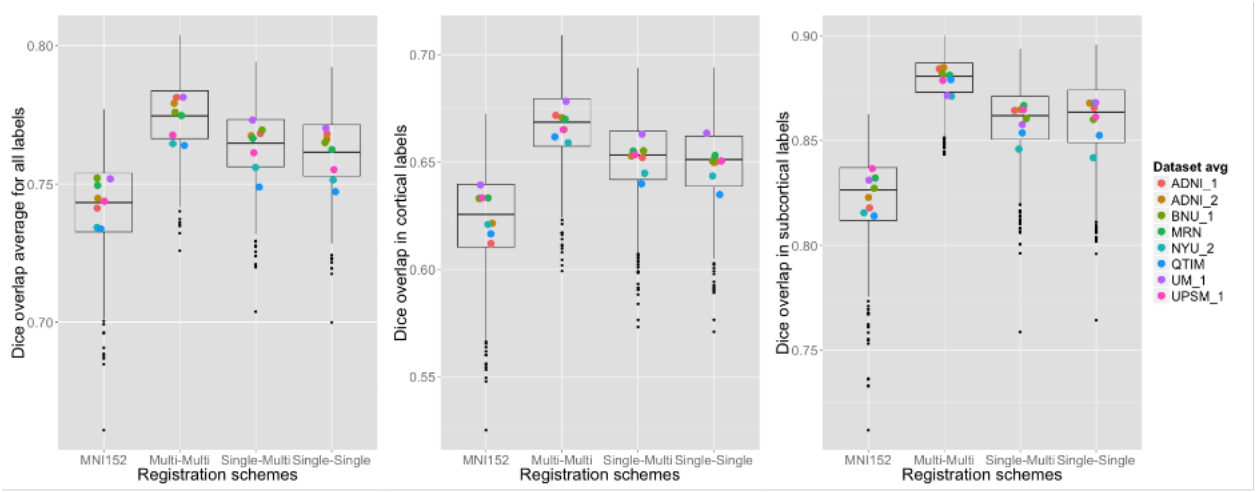


Figure 7. Boxplots showing the average label overlap between individual parcellations and parcellations in the ENIGMA-Mega MDT and MNI152 template space, for each registration scheme. Average cohort values are shown as colored dots overlaid on each boxplot.

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

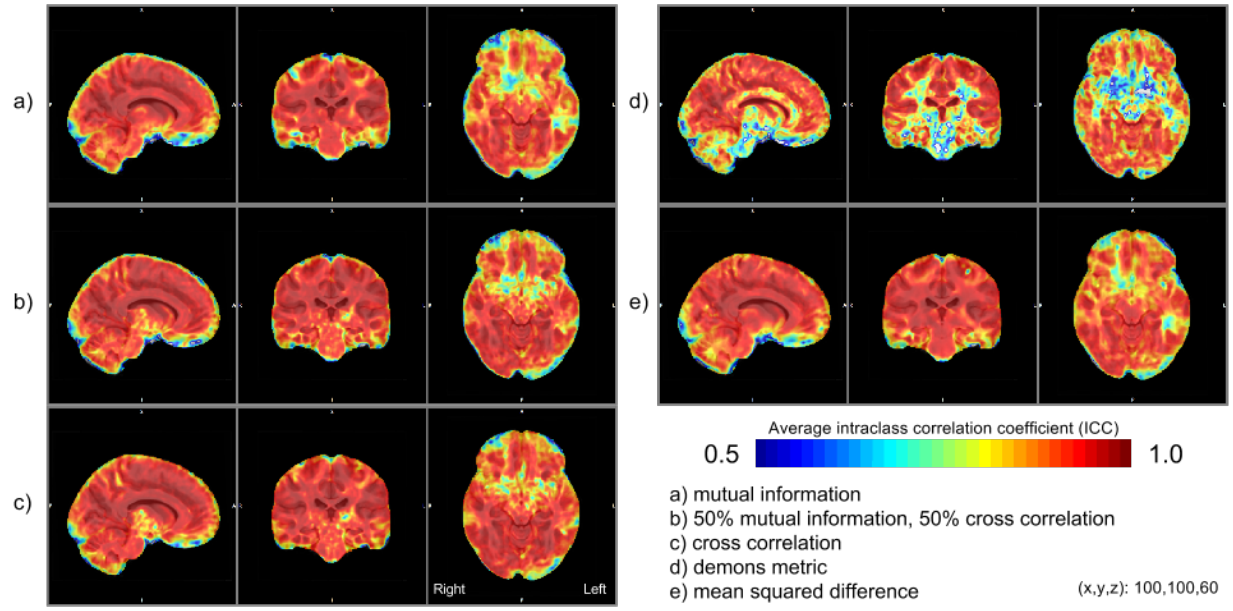


Figure 8. Voxelwise ICC of Jacobian maps in the QTIM cohort.

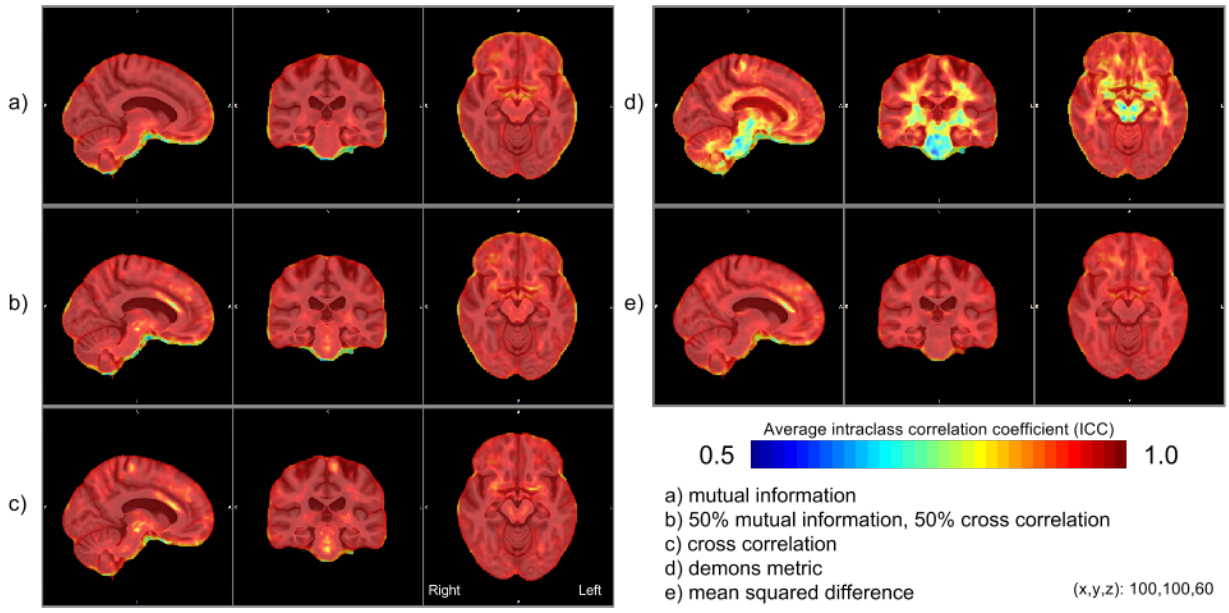


Figure 9. Voxelwise ICC of Jacobian maps in the ADNI_2 cohort.

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.

Relative Value of Diverse Brain MRI and Blood-Based Biomarkers for Predicting Cognitive Decline in the Elderly

Sarah K. Madsen[a], Greg Ver Steeg[c], Madelaine Daianu[a,b], Adam Mezher[a], Neda Jahanshad[a], Talia M. Nir[a], Xue Hua[a], Boris A. Gutman[a], Aram Galstyan[c], Paul M. Thompson[a,b,d]

[a]Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, University of Southern California, Marina del Rey, CA, USA
[b]Department of Neurology, UCLA School of Medicine, Los Angeles, CA, USA
[c]USC Information Sciences Institute, Marina del Rey, CA, USA
[d]Departments of Neurology, Psychiatry, Radiology, Engineering, Pediatrics, and Ophthalmology, University of Southern California, Los Angeles, CA, USA

ABSTRACT

Cognitive decline accompanies many debilitating illnesses, including Alzheimer's disease (AD). In old age, brain tissue loss also occurs along with cognitive decline. Although blood tests are easier to perform than brain MRI, few studies compare brain scans to standard blood tests to see which kinds of information best predict future decline. In 504 older adults from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we first used linear regression to assess the relative value of different types of data to predict cognitive decline, including 196 blood panel biomarkers, 249 MRI biomarkers obtained from the FreeSurfer software, demographics, and the AD-risk gene *APOE*. A subset of MRI biomarkers was the strongest predictor. There was no specific blood marker that increased predictive accuracy on its own, we found that a novel unsupervised learning method, CorEx, captured weak correlations among blood markers, and the resulting clusters offered unique predictive power.

Keywords: MRI, machine learning, brain, aging, cognitive decline

1. INTRODUCTION

Brain structure and the rate of brain tissue loss in old age can be assessed with a variety of neuroimaging tools [1-3]. An important question is which MRI measurements best predict cognitive decline, in elderly people with or without Alzheimer's disease (AD). We also need to understand how these brain MRI biomarkers perform in comparison to other biomarkers, such as those obtained from standard clinical tests of blood samples [4] and genetics [5], now that genotype information is readily available from saliva samples.

We considered a diverse set of over 400 potential biomarkers in 504 older adults from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain MRI measures included volume, gray matter thickness, and surface area measures for a standard set of anatomical regions of interest. We included potential predictors of cognitive decline covering a broad spectrum of health measures. We also included the AD-risk gene *APOE* – the strongest known genetic risk factor for late-onset AD [6], increasing lifetime risk around 3-fold for each copy of the *APOE4* allele carried [5].

We first used linear regression with cross-validation to compare the predictive value of including different types of variables (blood, MRI, genetic, demographic). To reduce dimensionality and avoid over-fitting, we used LASSO to select smaller subsets of variables within each type. We also used a novel unsupervised learning method, CorEx [7, 8], to learn low-dimensional representations that capture correlations among variables and to construct a hierarchical network that visually and quantitatively characterizes relationships among variables.

2. METHODOLOGY

2.1 Study Participants and Biomarkers

We analyzed data from ADNI participants for whom both MRI and blood test biomarkers were available, resulting in a sample of n=504 older adults (age: 75.0 ± 7.2 years; 198 women, 306 men; education: 15.6 ± 3.0 years; 57 cognitively normal, 351 with mild cognitive impairment, 97 with probable AD). For each participant, data was available for at least 152 of the biomarkers in our list. Cognitive decline was defined as one-year decline in Mini-Mental State Exam (MMSE) score [9]. Clearly, more refined measures of clinical decline are possible – including ADAS-Cog, clinical

dementia ratings, performance on specific neurocognitive tests – but here we used MMSE as it is a simple but widely used measure of decline.

Our model included a total of 451 predictor variables. This extensive list included 196 laboratory measures and 249 MRI measures obtained with the FreeSurfer analysis package. We also included standard demographic information (age, sex, years of education), height, weight, and *APOE* genotype (number of copies of *APOE4* carried by each participant).

Laboratory measures included blood tests relating to metabolism, kidney function, immune system function, growth factor levels, tumor biomarkers, inflammation levels, nutrition, oxidative stress, cardiovascular function, blood components (including red and white blood cell counts), hormone levels, liver function, and amyloid and tau pathology. A urine test of kidney function and a cerebrospinal fluid test (lumbar puncture/spinal tap) for amyloid and tau pathology were included. MRI measures included the volume of subcortical and cortical brain regions as well as mean gray matter thickness and total gray matter surface area for a standard set of cortical brain regions, as segmented by the 2006 Desikan-Killiany atlas (http://surfer.nmr.mgh.harvard.edu/).

ADNI was launched in 2004 by the National Institute of Health, the Food and Drug Administration, private pharmaceutical companies, and non-profit organizations to identify and evaluate biomarkers of AD for use in multisite studies. All ADNI data are publicly available online. The study was conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki, and the US 21 CFR Part 50–Protection of Human Subjects, and Part 56–Institutional Review Boards. Written informed consent was obtained from all participants in advance.

2.2 Scan Acquisition

Brain MRI data from 504 ADNI participants was included. High-resolution T1-weighted structural MRI brain scans were analyzed for all ADNI participants included in this analysis [10]. Scanning occurred on General Electric (Milwaukee, Wisconsin, USA), Siemens (Germany), or Philips (The Netherlands) scanners at 1.5T. Scan acquisition parameters were: 3D sagittal magnetization-prepared rapid gradient-echo (MP-RAGE) protocol; repetition time (2400 ms), flip angle (8°), inversion time (1000 ms), 24-cm field of view, a 192×192×166 acquisition matrix, a voxel size of 1.25×1.25×1.2 mm^3. Each scan was later reconstructed to 1 mm isotropic voxels.

2.3 Quantitative Predictive Power

We predicted cognitive decline for n=504 individuals for different sets of predictors using linear regression. Instead of predicting the raw change in MMSE score over one year, we predicted the percentile (based on a ranking) of a person's change with respect to all other ADNI study participants. Prediction using the raw score was extremely sensitive to outliers: while most MMSE score changes were less than 5 points, a few individuals appeared to lose more than 15 points, as shown in **Figure 1**. A person losing a single point is in the 46th percentile while anyone losing more than 10 points is below the 1st percentile. Of course, other definitions or reference cohorts would produce a somewhat different percentile ranking system.

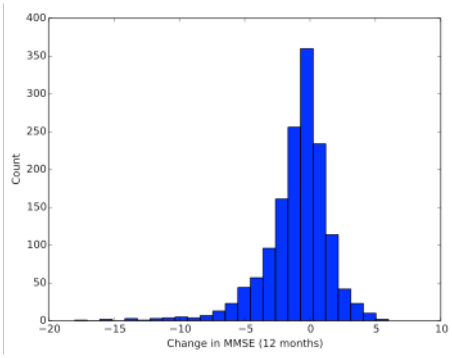


Figure 1. Histogram of changes in MMSE score over one year, in 504 ADNI participants.

For each set of predictors, we used 10-fold cross-validation and we report the mean across all ten folds of the mean squared error (MSE) on each test fold (using the same random set of folds for each experiment). This score estimates our prediction error on previously unseen examples. The accuracy of this estimate is reflected by the standard error of the mean across folds. The baseline MSE of 993.9 is obtained from always predicting the average percentile; the RMSE for this strategy is 31.5 percentile points. Regressing on, e.g., all 249 MRI measures leads to MSE worse than the

baseline, due to over-fitting. Therefore, for each group of variables, we first performed feature selection with LASSO [11].

The LASSO hyper-parameters were set using 10-fold cross-validation. LASSO selected a set of 47 MRI measures (*italicized* in the graph in **Fig. 2**), three blood tests (percent lymphocytes, CD5, APO C-I) and two cerebrospinal fluid biomarkers (amyloid beta I-42, ptau), along with age, weight, height, and *APOE* genotype.

Prediction results using different subsets of these variables are summarized in the *Results*. In addition to these variables, we also included a set of factors constructed to be maximally informative about correlations in the data using the novel machine learning technique, CorEx [7, 8].

2.4 Correlation Explanation (CorEx)

The recently introduced method of Correlation Explanation (CorEx) is an information-theoretic way to construct representations of high-dimensional data that are maximally informative about the data in some sense (http://www.github.com/gregversteeg/CorEx) [7, 8]. Let $X=(X_1,\ldots X_n)$ denote random variables that represent some observed measurements. We assume that we have iid samples of X drawn from some unknown distribution, $p_X(X=x)$, or p(x) for short. Information is measured using a multivariate mutual information measure historically called "total correlation" (TC), although in modern terms it would be better described as a measure of total dependence. TC is defined as:

$$TC(X) = D_{KL}(p(x)||\prod_{i=1}^{n} p(x_i)) = \sum_i H(X_i) - H(X)$$

where H denotes the Shannon entropy, and D_{KL} is the Kullback-Leibler divergence.

The total correlation among a group of variables, X, after conditioning on some other variable, Y, can be simply defined in terms of standard conditional entropies as:

$$TC(X|Y) = D_{KL}(p(x|y)||\prod_{i=1}^{n} p(x_i|y)) = \sum_i H(X_i|Y) - H(X|Y)$$

Then we can measure the extent to which some other variable, Y, explains the correlations in X by looking at how much the total correlation is reduced by conditioning on Y:

$$TC(X;Y) = TC(X) - TC(X|Y) = \sum_i I(X_i;Y) - I(X;Y)$$

TC(X|Y) is zero (and TC(X;Y) maximized) if and only if the distribution of X's conditioned on Y factorizes. This will be the case if Y contains full information about all of the common causes of the X_i's (in other words we say that Y explains all of the correlation in X).

The principle behind Correlation Explanation (CorEx) is to search for latent factors, $Y_1,\ldots,Y_m$, that maximize TC(X;Y).

$$\max_{\forall j, p(y_j|x)} TC(X;Y)$$

This optimization searches all possible functions of x for the m representatives that are most informative about the data. For a more detailed discussion of how this optimization is carried out see [7, 8]. This procedure can be carried out hierarchically so that we construct variables Y that explain correlations in X, then we can construct variables Z, which explain correlations in Y, etc. Applying this hierarchical procedure to MRI data leads to the structure in Figure 2. This technique has been used to out-perform standard learning methods for recovering structure in latent tree models, to perfectly reconstruct personality types from survey data, to automatically identify stock market sectors, and to automatically discover genetic sequences related to common ancestral origins [7, 8].

3. RESULTS

We computed the relative predictive accuracy of the biomarker groups alone and combined using both the original ("raw") data and using low-dimensional representations of the data learned with CorEx (**Table 1)**.

A graph of the structure of the latent factors for the brain MRI biomarkers in predicting cognitive decline is shown in **Figure 2**. Each raw predictive variable selected for inclusion by LASSO is italicized and slightly enlarged. Each selected CorEx factor (numbered circles) is italicized, enlarged, and highlighted in *yellow*. **Figure 2** shows brain MRI biomarkers only. The blood test biomarkers (not shown) were automatically split into a separate set of clusters by the unsupervised CorEx algorithm.

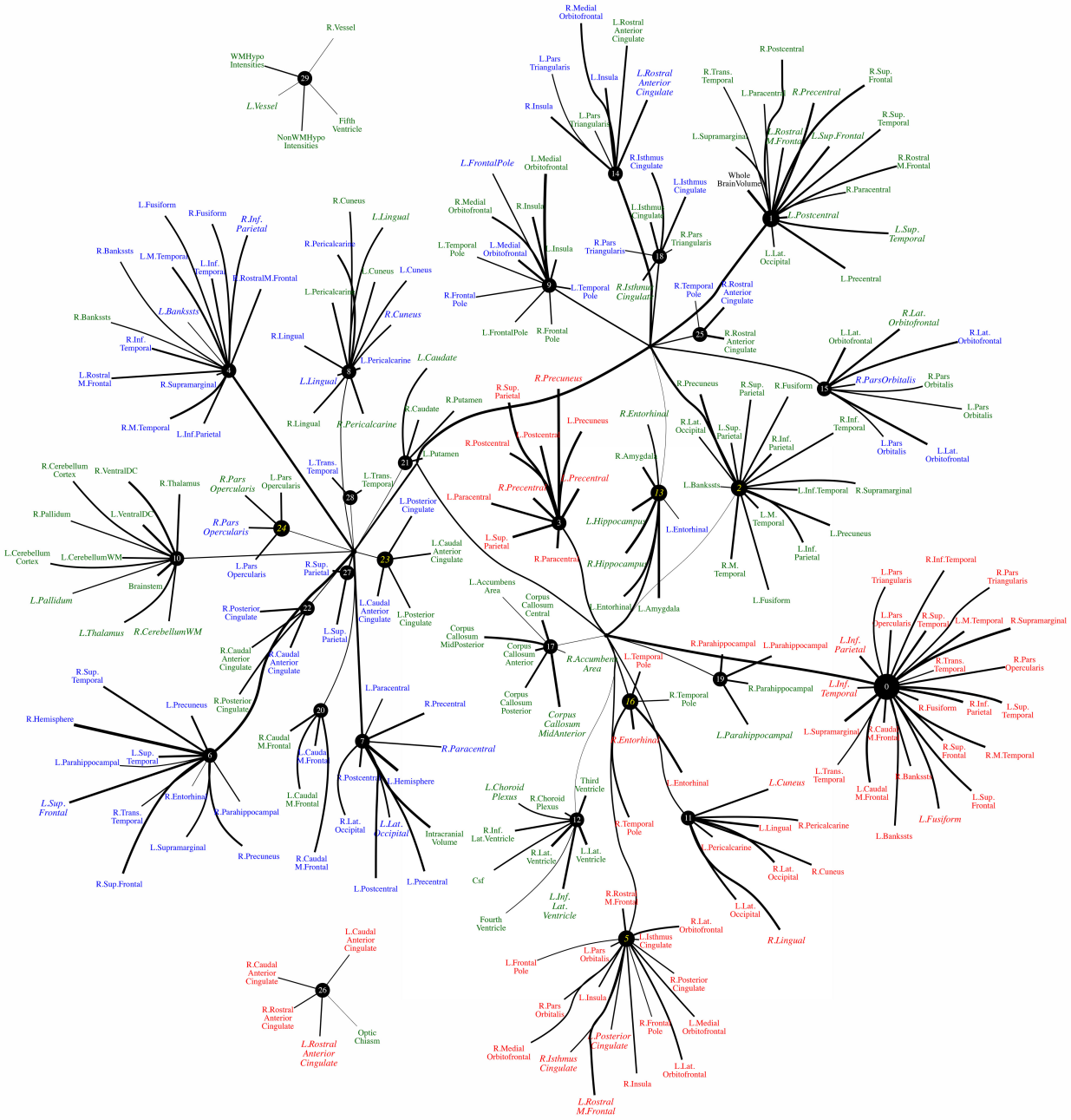


Figure 2. Graph of latent factors for brain MRI biomarkers of cognitive decline constructed by CorEx. Colors denote variable types (*green=volume, red=gray matter thickness, blue = gray matter surface area*). Edge width reflects mutual information, and node size reflects strength of correlation among connected variables. The subset of predictors of cognitive decline chosen by LASSO are in a larger, *italic* font. Note how the variables are grouped together in a biologically meaningful structure – for instance, measures of gray matter thickness are more closely grouped together. Importantly, measures that share high mutual information also tend to contribute redundant information, so we get diminishing returns from measuring all of them.

Separately:	Raw	CorEx	Combined:	Raw	Raw + CorEx Blood Test
Baseline	*993.9*	--	Brain MRI + Blood Test	805.3, 32.8	777.3, 25.5
Blood Test	938.3, 39.6	903.7, 40.5	Brain MRI + Other	790.3, 25.2	--
Brain MRI	800.8, 22.4	907.5, 22.5	Blood Test + Other	940.2, 42.3	--
Other (age, weight, height, APOE)	969.5, 31.9	--	Brain MRI + Other + Blood Test	800.6, 32.6	771.4, 28.2

Table 1. Relative predictive power of biomarker groups (mean squared error, standard error). Lower values indicate better performance (highlighted in *blue*).

4. NEW OR BREAKTHROUGH WORK BEING PRESENTED

Does combining different types of biomarkers improve accuracy in predicting cognitive decline? For the first time, we answer this question using a novel unsupervised machine learning technique, CorEx, and identified groups of measures with high multivariate mutual information. Combining the brain MRI and other biomarkers (demographics and genetics) achieved the greatest predictive accuracy compared to any individual biomarker group or combination of biomarker groups. Considering the high predictive power of the raw brain MRI measures [12], adding the blood test markers may not significantly improve prediction beyond what can be achieved with brain MRI biomarkers alone.

Even so, the CorEx representation of blood test biomarkers was a better predictor than the biomarkers individually. So we tested whether combining the CorEx factors with the raw brain MRI and other biomarker groups would further improve the prediction of cognitive decline. The performance did improve, compared to using all raw biomarker groups. In this combination, the best predictive accuracy was obtained when using all three biomarker groups together (CorEx blood test + raw brain MRI + raw other). This suggests that not only do correlations among blood biomarkers contain more valuable information than any individual biomarker on its own, but this information is not already present in MRI, *APOE*, or demographics.

5. CONCLUSION

In the raw data, the brain MRI biomarkers were the best predictors of cognitive decline, followed by the blood tests, and then the other categories of predictors (demographics and genetics).

Given a set of correlated predictors, LASSO tends to pick the best predictor and discards the rest. Instead of discarding correlations, CorEx will search for a latent factor that captures as much information as possible about all the correlated predictors. While the blood test biomarkers were relatively poor predictors individually, the CorEx representation of the blood data was a better predictor, suggesting that correlations among blood markers are more useful than the values of individual markers.

Interpreting the graphical representations of the data structure (**Fig. 2**) leads us to several key conclusions. The graph highlights how each individual biomarker variable relates to the entire set of biomarker variables. First, the unsupervised CorEx algorithm split the blood test and brain MRI biomarkers into two separate sets of clusters. This suggests that there is relatively little correlation between the variables in these two biomarker groups. This is somewhat unexpected, as these blood test biomarkers were included in ADNI based on evidence that they were associated with brain health, although the vast majority of this literature is based on studies analyzing a single blood test on its own. As expected, different brain MRI measures (volume, thickness, surface area) are largely clustered together, indicating a clear correlation in measures across brain regions. The clusters also correspond to known anatomical and functional subdivisions within the human brain, such as the clustering of temporal and frontal lobe substructures into distinct clusters.

Acknowledgements

Algorithm development and image analysis for this study was funded, in part, by grants to PT from the NIBIB (R01 EB008281, R01 EB008432) and by the NIA, NIBIB, NIMH, the National Library of Medicine, and the National Center for Research Resources (AG016570, AG040060, EB01651, MH097268, LM05639, RR019771 to PT). Algorithm development was also funded, in part, by ARO grant to AG (W911NF-15-1-0259), and AFOSR grant to GVS (FA9550-

12-1-0417). Data collection and sharing for this project was funded by ADNI (NIH Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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. 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. This research was also supported by NIH grants P30 AG010129 and K01 AG030514 from the National Institute of General Medical Sciences; and by a Consortium grant (U54 EB020403) from the NIH Institutes contributing to the Big Data to Knowledge (BD2K) Initiative, including the NIBIB and NCI.

References

[1] M. Daianu, A. Mezher, M. F. Mendez *et al.*, "Disrupted rich club network in behavioral variant frontotemporal dementia and early-onset Alzheimer's disease," Hum Brain Mapp, (2015).

[2] M. Daianu, M. F. Mendez, V. G. Baboyan *et al.*, "An advanced white matter tract analysis in frontotemporal dementia and early-onset Alzheimer's disease," Brain Imaging and Behavior, In Press (2015).

[3] M. Daianu, N. Jahanshad, T. M. Nir *et al.*, "Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network," Human Brain Mapping*,* 36(8), 3087–3103 (2015).

[4] M. Mapstone, A. K. Cheema, M. S. Fiandaca *et al.*, "Plasma phospholipids identify antecedent memory impairment in older adults," Nat Med*,* 20(4), 415-8 (2014).

[5] S. Sadigh-Eteghad, M. Talebi, and M. Farhoudi, "Association of apolipoprotein E epsilon 4 allele with sporadic late onset Alzheimer`s disease. A meta-analysis," Neurosciences (Riyadh)*,* 17(4), 321-6 (2012).

[6] M. Daianu, A. Mezher, N. Jahanshad *et al.*, "Spectral graph theory and graph energy metrics show evidence for the Alzheimer's disease disconnection syndrome in APOE-4 gene carriers," IEEE International Symposium of Biomedical Imaging (ISBI), 458-461 (2015).

[7] G. Ver Steeg, and A. Galstyan, "Discovering structure in high-dimensional data through correlation explanation," Advances in Neural Information Processing Systems, (2014).

[8] G. Ver Steeg, and A. Galstyan, "Maximally informative hierarchical representations of high-dimensional data," Artificial Intelligence and Statistics (AISTATS), (2015).

[9] V. C. Pangman, J. Sloan, and L. Guse, "An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice," Appl Nurs Res*,* 13(4), 209-13 (2000).

[10] C. R. Jack, Jr., M. A. Bernstein, N. C. Fox *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," J Magn Reson Imaging*,* 27(4), 685-91 (2008).

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso:a retrospective," Journal of the Royal Statistical Society*,* 73(3), 273-282 (2011).

[12] M. Daianu, G. ver Steeg, A. Mezher *et al.*, "Information-theoretic clustering of neuroimaging metrics related to cognitive decline in the elderly," Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science: Springer*,* In Press, (2015).

Axonal Diameter and Density Estimated with 7-Tesla Hybrid Diffusion Imaging in Transgenic Alzheimer Rats

Madelaine Daianu*[a,b], Russell E. Jacobs [c], Terrence Town[d], Paul M. Thompson[a,b,e]

[a]Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute,
University of Southern California, Marina del Rey, CA, USA
[b]Department of Neurology, UCLA School of Medicine, Los Angeles, CA, USA
[c]Biological Imaging Center, California Institute of Technology, Pasadena, CA
[d]Zilhka Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA
[e]Departments of Neurology, Psychiatry, Radiology, Engineering, Pediatrics, and Ophthalmology,
University of Southern California, Los Angeles, CA, USA

ABSTRACT

Diffusion-weighted MR imaging (DWI) is a powerful tool to study brain tissue microstructure. DWI is sensitive to subtle changes in the white matter (WM), and can provide insight into abnormal brain changes in diseases such as Alzheimer's disease (AD). In this study, we used 7-Tesla hybrid diffusion imaging (HYDI) to scan 3 transgenic rats (line TgF344-AD; that model the full clinico-pathological spectrum of the human disease) *ex vivo* at 10, 15 and 24 months. We acquired 300 DWI volumes across 5 *q*-sampling shells (b=1000, 3000, 4000, 8000, 12000 s/mm^2). From the top three *b*-value shells with highest signal-to-noise ratios, we reconstructed markers of WM disease, including indices of axon density and diameter in the corpus callosum (CC) – directly quantifying processes that occur in AD. As expected, apparent anisotropy progressively decreased with age; there were also decreases in the intra- and extra-axonal MR signal along axons. Axonal diameters were larger in segments of the CC (splenium and body, but not genu), possibly indicating neuritic dystrophy – characterized by enlarged axons and dendrites as previously observed at the ultrastructural level (see Cohen et al., J. Neurosci. 2013). This was further supported by increases in MR signals trapped in glial cells, CSF and possibly other small compartments in WM structures. Finally, tractography detected fewer fibers in the CC at 10 versus 24 months of age. These novel findings offer great potential to provide technical and scientific insight into the biology of brain disease.

Keywords: axonal diameter, multi-shell, HARDI, hybrid diffusion imaging (HYDI), rat, Alzheimer's disease

1. INTRODUCTION

Diffusion-weighted MR imaging (DWI) is sensitive to the organization and geometry of semipermeable barriers within living tissue microstructure [1, 2]. These barriers affect water diffusion in the white matter (WM) and can be picked up in DWI over millisecond timescales [1, 3]. In clinical research, these changes in the WM are most commonly characterized by measures of mean diffusivity and fractional anisotropy (FA) [4] – as reconstructed from the standard diffusion tensors. But although these metrics provide statistics of diffusion anisotropy at voxel level, they do not directly relate to features of tissue microstructure – such as cell size and packing density [1].

Axonal density and diameter are traditionally estimated with invasive histological procedures [e.g., electron microscopy (EM)] and therefore, are limited to *post mortem* tissue. Artifacts may arise from histological procedures (e.g., shrinkage, etc.) and from sampling of small regions of tissue [5]. To overcome some of these limitations, a few DWI-based methods have been proposed to map the axonal density and diameter of the WM matter structure, such as AxCaliber [6] [7], and more recently ActiveAx [1], that models orientation-invariant indices of axonal diameter and density. ActiveAx estimates the MR signal in four compartments of the WM, including intra-, extra-axonal populations of water, stationary water in glial cells and other small structures, and cerebrospinal fluid (CSF) [1].

Here, we implemented the ActiveAx framework and estimated indices of axonal density and diameter in three transgenic Alzheimer rats (line TgF344-AD) at different time points in the disease – 10, 15 and 24 months, using high-field multi-shell imaging, also known as hybrid diffusion imaging (HYDI) [8]. We collected 300 diffusion volumes across 5 distinct *q*-sampling shells and reconstructed the diffusion signal using the diffusion tensor imaging (DTI) model [2] in the corpus callosum (CC). Based on previous EM findings in these transgenic rats [9], we expected to find

indications of dystrophic neurites, or axonal swelling, characterized by enlarged axons and dendrites (possibly filled with vacuoles) as the disease progresses. We provide a summary of axonal density and diameter indices in addition to descriptors of anisotropy and demonstration of tractography based on HYDI, to show altered fiber integrity.

2. METHODS

2.1 Data and Analysis

Three transgenic Alzheimer rats (line TgF344-AD) were generated on a Fischer 344 background by co-injecting rat pronuclei with two human genes driven by the mouse prion promoter: "Swedish" mutant human APP (APPsw) and Δ exon 9 mutant human presenilin-1 (PS1ΔE9) [18]. Transgene integration was confirmed by genotyping and expression levels were evaluated by Western blot of brain homogenates. All experiments were conducted with protocols approved by the Institutional Animal Care and Use Committee (IACUC). The protocol called 'Peripheral TGF-beta Pathway Inhibitor Therapy in Alzheimer's Rats' was approved by the University of Southern California IACUC (Protocol Number: 20044). TgF344-AD rats were housed at the University of Southern California, Zilkha Neurogenetic Institute animal facility. Rats were maintained on normal lab chow and generally housed two per cage, in order to allow socialization. Nesting material was provided to all rats, and environmental enrichment in the form of plastic vertical barriers or tubes was added to all cages. Additionally, extraneous noise that may induce stress was minimized by keeping doors closed to housing rooms. A cage cleaning protocol was adopted that balanced hygiene with the need to retain some odor cues (e.g., scent-marked nesting material) to avoid stress and aggression. Finally, gentle and frequent handling of rats early in life was ensured. Before scanning, we anesthetized the animals with isoflurane, and then performed euthanasia. All efforts were made to minimize suffering.

We scanned the three rats *ex vivo* at 10, 15 and 24 months with a 7 Tesla Bruker BioSpin MRI scanner at the California Institute of Technology. After the three rats were sacrificed at the aforementioned ages, fixed brains (intact within the skull) were soaked in a gadolinium contrast agent (5mM ProHance) for 4 days prior to imaging to decrease the overall T1 of the tissue [10]. To ensure no leakage and that the signal would not change during acquisition, the samples were immersed in galden (perfluoropolyether with same magnetic susceptibility as water). During acquisition, the temperature was monitored via a fiber optic temperature sensor near the sample (and was 20°C for the whole scan) [4].

Using a 3D 8-segment spin echo EPI sequence with 1 average, we acquired 300 DWI volumes (133x233x60 matrix; voxel size: 0.15x0.15x0.25 mm^3; TE=34 ms; TR=500 ms; δ=11 ms; Δ=16 ms), yielding a 20-hour scan time. Specifically, 60 DWI volumes were acquired for each of the 5 *q*-sampling shells, b = 1000, 3000, 4000, 8000 and 12000 s/mm^2, with the same angular sampling, and 5 T2-weighted volumes with no diffusion sensitization (b_0 images). The relatively long δ and Δ values were required to achieve the largest *b*-values within the duty cycle constraints of our gradient coils. This long scan time precludes in vivo imaging, but this is not an issue for fixed samples [11].

During preprocessing, extra-cerebral tissue was removed using the "skull-stripping" Brain Extraction Tool from BrainSuite (http://brainsuite.org/) for both the anatomical images and the DWIs. We corrected for eddy current distortions using the "eddy correct FSL" tool (www.fmrib.ox.ac.uk/fsl) for which a gradient table was calculated to account for the distortions. DWIs were up-sampled to the resolution of the anatomical images (with isotropic voxels) using FSL's flirt function with 9 degrees of freedom; the gradient direction tables were rotated accordingly after each linear registration.

We calculated the signal-to-noise (SNR) ratio of each *q*-sampling shell by computing the average diffusion signal in the WM structure of each rat and estimated noise from the mean standard deviation of all 5 b=0 s/mm^2 images. Then, we ran a 2-tailed *t*-test to determine which *b*-value weighting provided the highest SNR.

2.2 Axonal Diameter Estimation

We modeled four tissue compartments in the WM defined based on the population of water molecules as outlined in [1]. Each compartment provided a separated normalized MR signal: (1) S_1 – signal trapped in the intra-axonal parallel cylinders of equal diameter, a, modeled with the Gaussian phase distribution approximation [12]; (2) S_2 – signal found in the extra-axonal water near but outside the cylinders modeled based on the diffusion tensor (DT) scheme, also assuming Gaussian distributed displacements [2]; (3) S_3 – signal found in the stationary water (S_4 =1 as signal is unattenuated by

diffusion weighting) that may be trapped in glial cells and other subcellular structures (4) S_4 – signal from the CSF; here, the cylinders did not influence the diffusion and the model is isotropic Gaussian displacements [1].

The equation modeled by ActiveAx assumes that no exchange takes place between the compartments of water molecules described above [1]:

$$S^* = S_0^* \sum_{i=1}^{4} f_i S_i \quad (\mathbf{1})$$

Here, S^* is the MR diffusion signal, S_0^* is the MR diffusion signal from the b_0 image, f_1 defines the proportion of water molecules in a population, i, where: $\sum_{i=1}^{4} f_i = 1$ and $0 \le f_i \le 1$.

The axonal diameter was estimated from the following model: $E(a) = \int p(a) a \, da$, where $E(a)$ is the mean axonal diameter and p is the true distribution of axon diameters [1]. For more specifics on the model, please refer to [1]. We assessed the axonal diameter indices at three different segments of the CC – the splenium, body and genu. In this work, we studied the CC as it has larger axonal radii, especially in the middle segment (the body), which can be feasibly reconstructed using the ActiveAx method. In addition, the CC does not have crossing fibers where models like DTI might fail to accurately reconstruct the intersecting fibers at voxel level.

3. RESULTS

To start, we show that the diffusion signal SNR is significantly higher in the lower b-value shells when comparing single- and multi-shell reconstructions between b=1000, 3000 and 4000 s/mm^2 and b=8000 and 12000 s/mm^2 (P=0.03, 2-tailed t-test) (**Fig. 1**). Therefore, we selected the lower 3-shell HYDI data to reconstruct the axonal density and diameter indices.

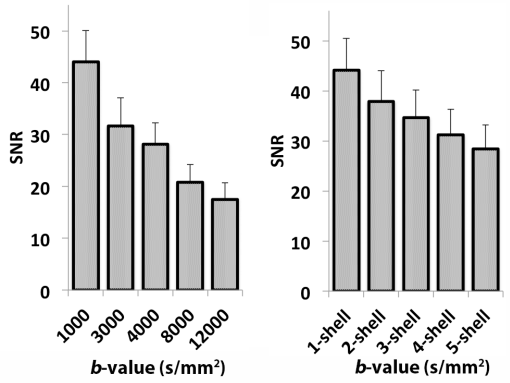


Figure 1. Signal-to-noise ratio (SNR) computed from the diffusion signal in the single b-value DWIs and the multi-shell HYDIs.

For each transgenic Alzheimer rat, we show patterns of alteration directly related to their tissue microstructure (summarized in **Figs. 2** and **3**). The MR signal quantified by the intra- and extra-axonal volume fraction (F1 and F2) decreased with disease progression (at the later stages of disease, 15 and 24 months), possibly indicating less directional water diffusion. This is closely related to decreases in FA; once again indicating altered anisotropic properties of the white matter tissue. The increase in the stationary water volume fraction (F3) and CSF volume fraction (F4) are further suggestive of the etiology of brain tissue disruptions noted on EM [9] in the transgenic Alzheimer rats (e.g., axonal swelling). Increases in F3 could indicate that more water is being trapped within glial cells and other small compartments that are modeled independently of the cylinder-like structures (i.e., modeling the axons).

In accord with our hypothesis, axonal radii were larger in the body of the CC than in the splenium and genu (**Fig. 3A**). Larger axons contain more water and contribute more to the MR signal. With disease progression, axonal radii increased in the splenium and body of the CC, but in the genu they showed opposing patterns; axonal radii in the genu

were also harder to map, so additional studies are needed to confirm these findings. Finally, from our fiber tract maps we show visible decreases in fiber density between the 10 and 24 month old transgenic Alzheimer rats (**Fig. 3B**).

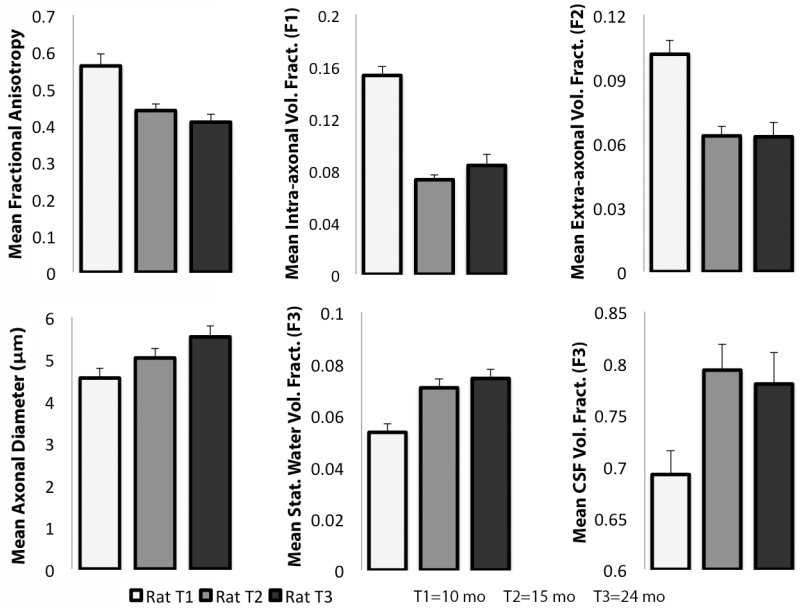


Figure 2. Plots of mean axonal density and anisotropy in the corpus callosum in each transgenic Alzheimer rat. Mean fractional anisotropy (FA) decreased in concordance with decreases in the intra- and extra-axonal volume fraction (F1 and F2). Stationary water and CSF volume fractions (F3 and F4) showed increasing patterns in the later stages of the disease.

4. NEW OR BREAKTHROUGH WORK BEING PRESENTED

In this study, we mapped and interpreted neuroimaging indices directly associated with WM microstructure in transgenic Alzheimer rats at three stages of the disease. To do this, we computed HYDI-based indices of axonal density and diameter and interpreted the observed patterns of disruption in relation to the previously investigated etiology of AD in a recent EM ultrastructural study using TgF344-AD rats [9]. For the first time, we show overlapping findings of altered WM structure going beyond the commonly reported apparent anisotropy changes in AD.

5. CONCLUSION

The involvement of the CC in AD is quite well understood, and the parietal, temporal and possibly occipital segments are among the first affected. Here, we show patterns of decreasing anisotropy in the CC and indices that may contribute to this phenomenon – decreasing intra- and extra-axonal volume fractions. Meanwhile, we also show patterns of increasing stationary water volume fraction that may influence the larger axonal radii observed at the later stages of the disease. Furthermore, increased CSF volume fractions could be explained by atrophy or tissue loss, and are also supported by the lower fiber density in the 24 month old transgenic Alzheimer rat. In this preliminary study we do not have the necessary number of subjects to detect statistically significant differences in the changes observed in the WM microstructure, but the landscape of disruption aligns with changes expected from EM in the cortex and hippocampus in this same transgenic Alzheimer rat line [9].

Understanding the network of WM pathways that supports communication in the brain is critically important to detecting and eventually preventing disease. Measures of axonal density and diameter are important morphological properties of WM, as their integrity is directly related to the rate of information transfer of a nerve bundle and can provide insight into the effects of neurological disease [5], otherwise not detectable with macroscopic neuroimaging metrics. The benefits of HYDI and detailed microscopic indices may be valuable for human [13-16] and animal connectome (**Fig. 3C**) projects and clinical research [4, 17, 18].

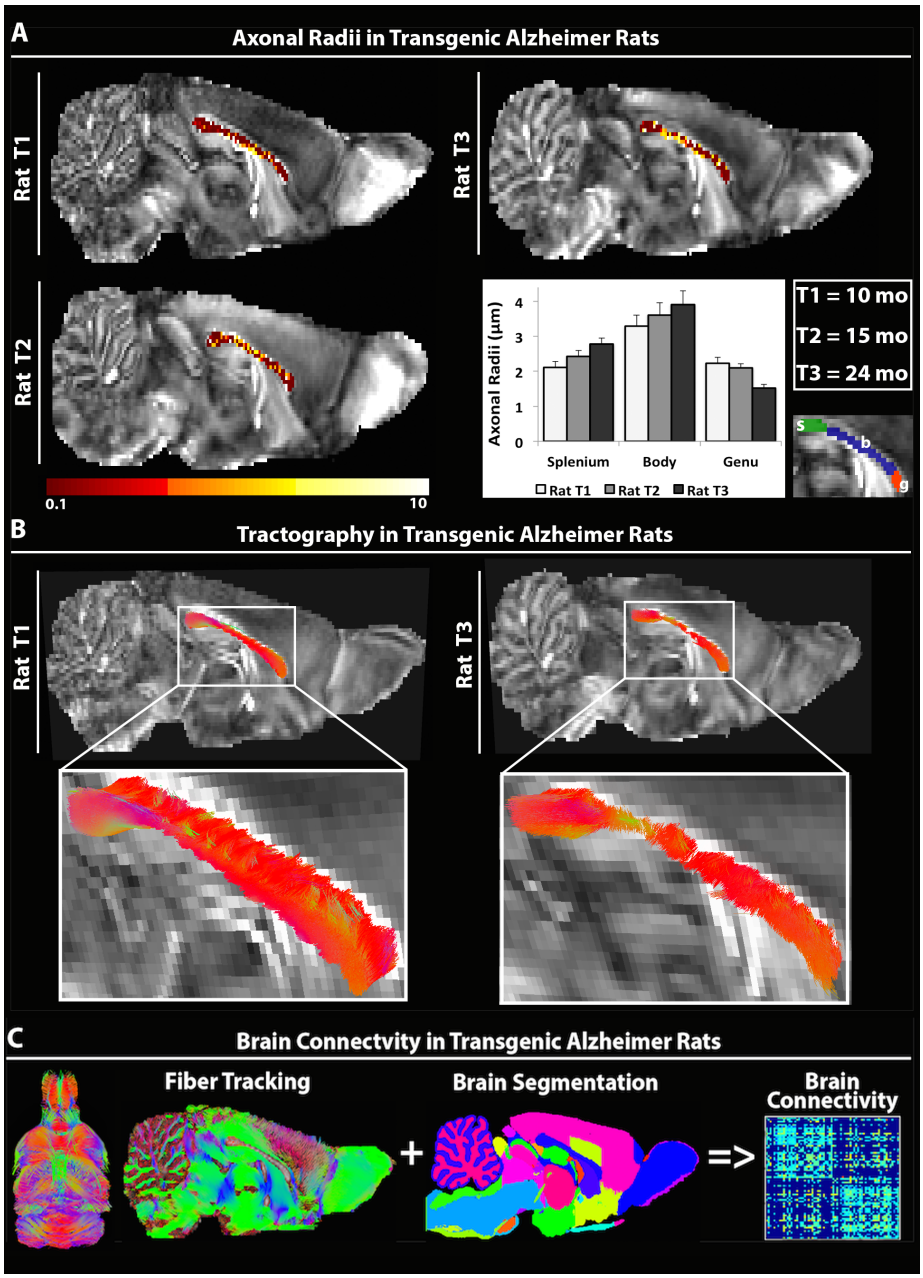


Figure 3. Maps of axonal diameter and fiber density in three transgenic Alzheimer rats. **A.** Axonal diameter indices were evaluated in the splenium, body and genu of the corpus callosum (CC) across the different stages of the disease. As expected, the axonal diameter was larger in the body of the CC than in the splenium and genu. With disease progression, axonal radii increased in the splenium and body of the CC (but not in the genu) - possibly signs of dystrophic neurites and axonal swelling found in electron microscopy studies in these TgF344-AD rats. **B.** Fiber density maps in the CC of the 10 vs. 24 month old rats indicated loss of fibers with disease progression. **C.** For our future work, we aim to reconstruct whole brain connectivity maps using the fiber density maps as depicted in **B**.

Acknowledgments

Algorithm development and image analysis for this study were funded, in part, by grants to PT from the NIBIB (R01 EB008281, R01 EB008432) and by the NIA, NIBIB, NIMH, and the National Library of Medicine (AG016570, AG040060, EB01651, MH097268, LM05639 to PT) and by an NINDS grant to TT (R01 NS076794). Data collection and sharing for this project was funded by NIH Grant R01 AG034499-05. NIBIB & the Beckman Institute at Caltech provided essential support for the scanner & personnel (NIBIB EB000993). TT and TMW were supported by an NIH/NINDS grant (5R01NS076794-05). This work was also supported in part by a Consortium grant (U54 EB020403) from the NIH Institutes contributing to the Big Data to Knowledge (BD2K) Initiative and R01 EB008432.

REFERENCES

[1] D. C. Alexander, P. L. Hubbard, M. G. Hall *et al.*, "Orientationally invariant indices of axon diameter and density from diffusion MRI," Neuroimage*,* 52(4), 1374-89 (2010).

[2] P. Basser, J. Mattiello, and D. LeBihan, "MR Diffusion Tensor Spectroscopy and Imaging," Biophysical Journal 66, 259-267 (1994).

[3] M. Daianu, M. F. Mendez, V. G. Baboyan *et al.*, "An advanced white matter tract analysis in frontotemporal dementia and early-onset Alzheimer's disease," Brain Imaging and Behavior, In Press (2015).

[4] M. Daianu, R. E. Jacobs, T. M. Weitz *et al.*, "Multi-Shell Hybrid Diffusion Imaging (HYDI) at 7 Tesla in TgF344-AD Transgenic Alzheimer Rats," PLoS One*,* 10(12), e0145205 (2015).

[5] D. Barazany, P. J. Basser, and Y. Assaf, "In vivo measurement of axon diameter distribution in the corpus callosum of rat brain," Brain*,* 132(Pt 5), 1210-20 (2009).

[6] Y. Assaf, T. Blumenfeld-Katzir, Y. Yovel *et al.*, "AxCaliber: a method for measuring axon diameter distribution from diffusion MRI," Magn Reson Med*,* 59(6), 1347-54 (2008).

[7] D. C. Alexander, "A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features," Magn Reson Med*,* 60(2), 439-48 (2008).

[8] M. Daianu, N. Jahanshad, J. E. Villalon-Reina *et al.*, "7T multi-shell Hybrid Diffusion Imaging (HYDI) for mapping brain connectivity in Mice," The International Society for Optics and Photonics (SPIE), Medical Imaging 2015: Image Processing*,* 9413, 941309-941309 (2015).

[9] R. M. Cohen, K. Rezai-Zadeh, T. M. Weitz *et al.*, "A transgenic Alzheimer rat with plaques, tau pathology, behavioral impairment, oligomeric AB, and frank neuronal loss," Journal of Neuroscience*,* 33(15), 6245-6256 (2013).

[10] H. Benveniste, K. Kim, L. Zhang *et al.*, "Magnetic resonance microscopy of the C57BL mouse brain," Neuroimage*,* 11(6 Pt 1), 601-11 (2000).

[11] O. I. Alomair, I. M. Brereton, M. T. Smith *et al.*, "In vivo high angular resolution diffusion-weighted imaging of mouse brain at 16.4 Tesla," PLoS One*,* 10(6), e0130133 (2015).

[12] J. S. Murday, and R. M. Cotts, "Self-Diffusion Coefficient of Liquid Lithium," The Journal of Chemical Physics*,* 48(11), 4938 (1968).

[13] M. Daianu, N. Jahanshad, T. M. Nir *et al.*, "Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network," Human Brain Mapping*,* 36(8), 3087–3103 (2015).

[14] M. Daianu, A. Mezher, M. F. Mendez *et al.*, "Disrupted rich club network in behavioral variant frontotemporal dementia and early-onset Alzheimer's disease," Hum Brain Mapp, (2015).

[15] M. Daianu, N. Jahanshad, T. M. Nir *et al.*, "Algebraic connectivity of brain networks shows patterns of segregation leading to reduced network robustness in Alzheimer's disease," Medical Image Computing and Computer Assisted Intervention (MICCAI), Computational Diffusion MRI: Springer, 55-64 (2014).

[16] M. Daianu, E. L. Dennis, N. Jahanshad *et al.*, "Alzheimer's disease disrupts rich club organization in brain connectivity networks," IEEE International Symposium of Biomedical Imaging (ISBI), 266-269 (2013).

[17] M. F. Mendez, A. Joshi, M. Daianu *et al.*, "White Matter Changes Associated with Resting Sympathetic Tone in Frontotemporal Dementia vs. Alzheimer's Disease," PLoS One*,* 10(11), e0142445 (2015).

[18] F. F. Roussotte, M. Daianu, N. Jahanshad *et al.*, "Neuroimaging and genetic risk for Alzheimer's disease and addiction-related degenerative brain disorders," Brain Imaging Behav*,* 8(2), 217-33 (2014).

The heritability of the functional connectome is robust to common nonlinear registration methods

George W. Hafzalla[a], Gautam Prasad[a], Vatche G. Baboyan[a], Joshua Faskowitz[a], Neda Jahanshad[a], Katie L. McMahon[b], Greig I. de Zubicaray[c], Margaret J. Wright[d], Meredith N. Braskie[a], and Paul M. Thompson[a]

[a]Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, California, USA, 90292
[b]Centre for Advanced Imaging, University of Queensland, Brisbane, QLD 4072, Australia
[c]Faculty of Health and Institute of Health Biomedical Innovation (IHBI), Queensland University of Technology, Brisbane, QLD, 4072
[d]QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia

ABSTRACT

Nonlinear registration algorithms are routinely used in brain imaging, to align data for inter-subject and group comparisons, and for voxelwise statistical analyses. To understand how the choice of registration method affects maps of functional brain connectivity in a sample of 611 twins, we evaluated three popular nonlinear registration methods: Advanced Normalization Tools (ANTs), Automatic Registration Toolbox (ART), and FMRIB's Nonlinear Image Registration Tool (FNIRT). Using both structural and functional MRI, we used each of the three methods to align the MNI152 brain template, and 80 regions of interest (ROIs), to each subject's T1-weighted (T1w) anatomical image. We then transformed each subject's ROIs onto the associated resting state functional MRI (rs-fMRI) scans and computed a connectivity network or functional connectome for each subject. Given the different degrees of genetic similarity between pairs of monozygotic (MZ) and same-sex dizygotic (DZ) twins, we used structural equation modeling to estimate the additive genetic influences on the elements of the function networks, or their heritability. The functional connectome and derived statistics were relatively robust to nonlinear registration effects.

Keywords: nonlinear registration, inter-subject alignment, structural tissue classification, resting state functional MRI, functional connectome, heritability, default mode network, multimodal evaluation

1. INTRODUCTION

This work assesses the effects of three different nonlinear deformation algorithms,[1] when computing maps of functional brain connectivity.[2] We evaluate their accuracy using overlap metrics for 80 cortical and subcortical regions of interest (ROIs) derived from structural MRI. We also assess whether functional connectivity networks derived from resting state data in the ROIs are robust to the choice of registration algorithm.

Further author information: (Send correspondence to George Hafzalla)
George W. Hafzalla: E-mail: hafzalla@usc.edu
Gautam Prasad: E-mail: gprasad@usc.edu
Paul M. Thompson: E-mail: pthomp@usc.edu

2. METHODS

2.1 Sub jects

We obtained 611 resting state functional MRI (rs-fMRI) and T1-weighted (T1w) structural MRI brain scans (232 men, 379 women; mean age 22.5 ± 2.8 years when scanned) as part of the Queensland Twin IMaging (QTIM) study, a neuroimaging and genetic study of young, healthy adult twins and their family members. From this sample, we used a subset of 276 participants (58% MZ, 42% DZ) for the functional connectome analysis, where we excluded a person if their twin was not also present in the sample. Opposite-sex DZ twins were also excluded from the functional connectome analysis.

2.2 Acquisition

rs-fMRI scans were 5 min. 19 sec. in duration, collected at 4T (Bruker Medspec), with parameters: TR/TE = 2100/30 ms; 230 mm FOV; 64 x 64 acquisition matrix; 3.6 x 3.6 mm voxels; 3 mm slices/0.6 mm gap. Participants were instructed to lay awake with their eyes closed and not to focus on particular thoughts. High resolution T1w structural scans were acquired with an inversion recovery rapid gradient echo sequence (TI/TR/TE = 700/1500/3.35 ms; 240 mm FOV; 256 x 256 acquisition matrix; 0.9 x 0.9 x 0.9 mm3 voxels/0 mm gap).

2.3 Data processing

We used a variety of neuroimaging tools from the FMRIB software library (FSL).[3] rs-fMRI data was analyzed using FSL–FEAT (FMRI Expert Analysis Tool, v5.0.7; http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). We performed motion correction using MCFLIRT, non-brain tissue removal using BET (Brain Extraction Tool), spatial smoothing using a Gaussian kernel with 6 mm full width at half maximum, and high pass temporal filtering of 100 s. Scans were registered to a standard MNI (Montreal Neurological Institute) atlas using a 12 degrees of freedom affine transform.

3D T1w structural MRI data were pre-processed with FreeSurfer[4] (v5.3; http://surfer.nmr.mgh.harvard.edu/) to correct intensities in the structural images and label cortical and subcortical regions as defined by the Deskian-Killiany atlas.[5] FreeSurfer outputs were returned to native space and the cortical parcellation was binarized, dilated six times and eroded five times to produce an accurate subject-specific brain mask of the gray and white matter. The purpose of this dilation and erosion process was to fill any holes in the mask while maintaining a tight fit to the brain. This mask was used for subsequent data processing.

Computing a highly accurate classification of gray matter for each participant required a series of steps. Because the FreeSurfer binary mask of the cortical ribbon is susceptible to gray matter exclusions, we performed tissue classification using FSL–FAST (FMRIB's Automated Segmentation Tool) on the skull-stripped brain in native space. This allowed us to create a binary CSF mask, which was then subtracted from our FreeSurfer-derived brain mask. This brain mask was further reduced by subtracting binary masks of the following FreeSurfer parcellations: CSF, choroid plexus, corpus callosum, optic chiasm, ventricles, and white matter. To correct for residual errors in classification by FSL–FAST, the FreeSurfer cortical and subcortical parcellations were then added to the brain mask. This mask was binarized and served as our gray matter mask for subsequent data analysis (see **Figure 1**).

2.4 Nonlinear registration

Three nonlinear registration methods, Advanced Normalization Tools (ANTs),[6] Automatic Registration Toolbox (ART),[7] and FMRIB's Nonlinear Image Registration Tool (FNIRT) from FSL, were used to coregister each participant's intensity non-uniformity corrected, intensity normalized, and skull-stripped brain image to the MNI152 brain template. For each program, we selected parameter options that optimized inter-subject alignment.

We created 80 regions of interest (ROIs) on the MNI152 brain template using the AFNI (Analysis of Functional Neuroimages)[8] program 3dUndump. ROIs were spherical in shape with 3-mm radii. We placed the center of each ROI in the centroid position of its respective region. Center of gravity (cog) coordinates for each ROI are shown in **Table 1**.

For all subjects, we applied each registration method to co-register their T1w image to the MNI152 brain template. These transformations were then used to back-project the 80 ROIs from the MNI152 brain template to the native T1w brain space (see **Figure 1**). Finally, across all methods, we applied to these ROIs the affine transformations produced by FSL–FEAT. This moved the native T1w image space ROIs to the functional space for further analysis.

2.5 Functional connectome

We computed the functional connectome[9] for each participant as the connectivity or edge strength between all pairs of the 80 ROIs or nodes that had been transformed into functional space. We computed the average time-series activation across all the voxels in an ROI and using the correlation of this time series between ROIs to quantify their connectivity strength. The resulting 3,160 connectivity values were stored as a symmetric connectivity matrix; the lower triangular matrix (without diagonal entries) was used for all subsequent analysis. In many graph theoretic analyses of networks, a threshold is applied to filter or remove spurious edges, but here we chose to retain the full network to understand how sensitive it was to the underlying registration method used.

Our data was derived for two types of twins: monozygotic (MZ) twins, who share 100% of their genetic material, and same-sex dizygotic (DZ) twins who share on average 50%. We computed the intra-class correlation (ICC) in both the MZ and DZ twins to quantify cross-twin correlations in the functional networks, where connections that have a higher MZ ICC and lower DZ ICC point to a larger genetic influence on the brain connectivity.[10] In addition, we used structural equation modeling from classical genetics to assess the degree of the variance due to additive genetics or heritability (A), shared environmental (C), and unique environmental (E) effects in each connection in the network.[11]

2.6 Evaluation of nonlinear registration methods

Nonlinear registration was assessed in three ways:

1. **Percent of ROI overlap with gray matter.** To assess differences in the registration algorithms based on the ROI overlap with gray matter we used a linear mixed effects model that factored in the family membership into the statistics comparing each method. In the model, family membership was encoded as the random effect, the percent of gray matter overlap in each ROI was used as the response, and a dummy variable was used to represent two methods being compared, either ANTs vs. ART, ANTs vs. FNIRT, or ART vs. FNIRT.
2. **Number of valid connections in the functional connectome across sub jects.** We computed the percentage of valid connections in the functional connectivity network with respect to the total number of possible connections (3,160 edges for the 80 node network). A valid connection required two ROIs in the functional space with > 0 voxels and located in a part of the image that included activation data. We averaged the percentage of valid connections in each subject across the entire set of subjects used in the heritability analysis.
3. **Heritability of the functional connectome.** We compared the MZ ICC, DZ ICC, and heritability values from the ACE model computed over the functional connectivity networks. We used paired-sample t-tests to compare the distribution of these values between methods and averaged across edges to summarize the results.

3. RESULTS

3.1 Structural tissue classification

We show the percentage of gray matter in 16 ROIs that are associated with the default mode network in **Figure 2** across three registration methods. All methods perform well for subcortical structures, but there is large variation for cortical regions. **Table 2** lists statistics comparing the percentage of gray matter in cortical ROIs.

3.2 Functional connectome and heritability

The percentage of connections that were valid across the functional connectome was 94.3% computed from the ANTs transformed ROIs, while the number dropped to 90.4% for ART, and 89.2% for FNIRT based approaches.

We show the results from the MZ ICC, DZ ICC, and heritability analysis in **Figure 3**; **Table 3** summarizes the average values across the connections in the network. As expected, ANTs, ART, and FNIRT all showed higher average intra-class correlations in MZ twins relative to DZ twins. ANTs had higher average heritability but not significantly different from ART, while ANTs and ART both had significantly higher average heritability compared to FNIRT. Hence, all registrations lead to a heritable functional connectome overall but on average the degree of heritability may differ in various connections dependent on the method.

4. CONCLUSIONS

We studied registration algorithms in the context of computing functional connectivity networks. Our assessment is more complex than prior studies that focus simply on the overlap of the two images being aligned in the nonlinear registration step. We were able to use both structural and functional MRI to understand the effects of ROIs during registration and their subsequent influence on functional connectivity networks.

We found distinct differences in nonlinear registration performance based on ROI placement in gray matter and the percentage of valid connections in functional connectomes derived using the ROIs as nodes. Despite these differences, nonlinear registration methods did not have a large impact on the functional connectome when we assessed the values using heritability analysis. The lower resolution of the functional images may make them more robust to errors in the higher resolution ROI registration.

Our findings could help in making decisions when designing a robust functional connectivity pipeline.

5. FIGURES AND TABLES

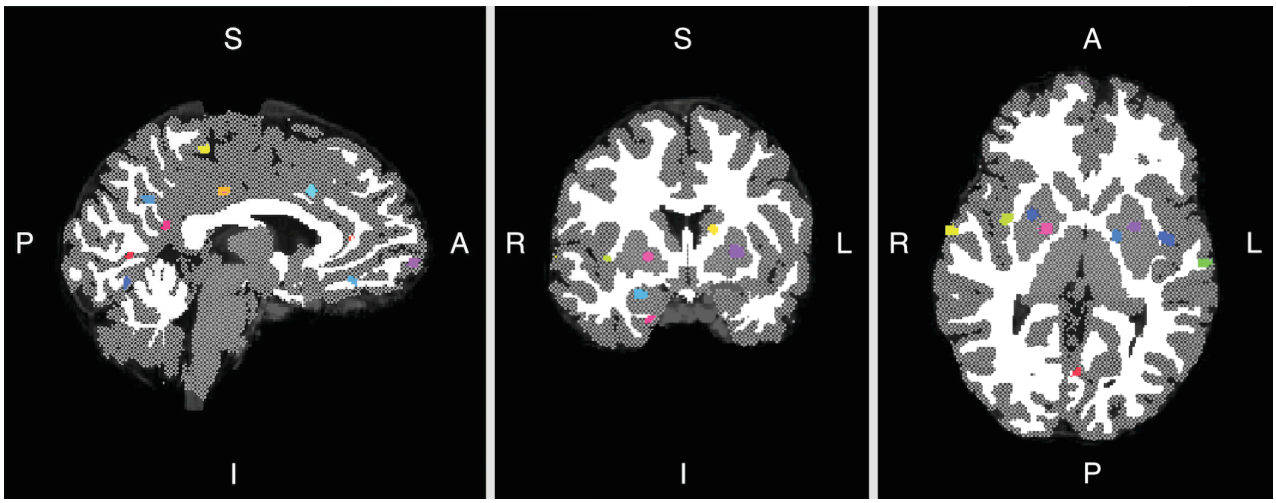


Figure 1: **Nonlinear registration applied to MNI152 ROIs.** A representative T1-weighted (T1w) structural MRI brain scan overlaid with a gray matter mask, white matter mask, and ROIs (regions of interest; random colors) that have been warped from the MNI152 brain template to a native scan using ANTs.

Table 1: **ROI cog Coordinates.** List of left and right hemisphere cog coordinates for each ROI (determined by FSL–fslstats), which were used to create a standard set of spherical ROIs on the MNI152 brain template. Because AFNI defaults to RAI orientation, the sign of the X coordinates listed were flipped to preserve orientation.

Region of Interest (ROI)	Left Hemisphere cog Coordinates (X,Y,Z)	Right Hemisphere cog Coordinates (X,Y,Z)
Amygdala	-22, -4, -18	25, -3, -19
Bank of the Superior Temporal Sulcus	-63, -43, 8	48, -38, 8
Caudal Anterior Cingulate	-5, 17, 30	4, 22, 31
Caudal Middle Frontal	-42, 17, 49	45, 15, 51
Caudate	-14, 5, 13	16, 8, 13
Cuneus	-1, -81, 19	4, -78, 19
Entorhinal	-17, -7, -34	21, -3, -35
Frontal Pole	-5, 67, -11	5, 67, -11
Fusiform	-35, -42, -23	37, -39, -24
Hippocampus	-24, -23, -13	27, -22, -13
Inferior Parietal	-46, -70, 36	53, -62, 37
Inferior Temporal	-51, -37, -25	52, -31, -25
Insula	-40, -2, 0	39, 1, -1
Isthmus of the Cingulate	-4, -50, 19	6, -48, 18
Lateral Occipital	-37, -92, 0	41, -88, 0
Lateral Orbitofrontal	-24, 30, -17	24, 31, -18
Lingual	-8, -71, -8	14, -64, -8
Medial Orbitofrontal	-2, 37, -20	3, 37, -20
Middle Temporal	-67, -27, -12	68, -26, -13
Pallidum	-18, -4, 0	21, -3, 0
Paracentral	-3, -27, 60	7, -26, 57
Parahippocampal	-22, -31, -21	24, -31, -19
Pars Opercularis	-57, 19, 14	59, 16, 13
Pars Orbitalis	-48, 37, -15	45, 41, -18
Pars Triangularis	-54, 33, 2	54, 36, 7
Peri-calcarine	-10, -79, 6	13, -76, 8
Postcentral	-57, -17, 42	62, -11, 40
Posterior Cingulate	-2, -20, 36	3, -20, 35
Pre-central	-56, -1, 42	58, -3, 42
Precuneus	-3, -58, 36	4, -55, 36
Putamen	-26, 1, 1	28, 3, 0
Rostral Anterior Cingulate	-5, 41, 3	4, 39, 2
Rostral Middle Frontal	-37, 54, 19	41, 51, 21
Superior Frontal	-12, 27, 58	14, 29, 57
Superior Parietal	-21, -71, 60	18, -65, 63
Superior Temporal	-64, -13, 0	66, -7, 0
Supra-marginal	-64, -36, 33	64, -27, 33
Temporal Pole	-26, 17, -37	28, 20, -37
Thalamus	-10, -17, 8	13, -16, 8
Transverse Temporal	-47, -17, 8	44, -18, 9

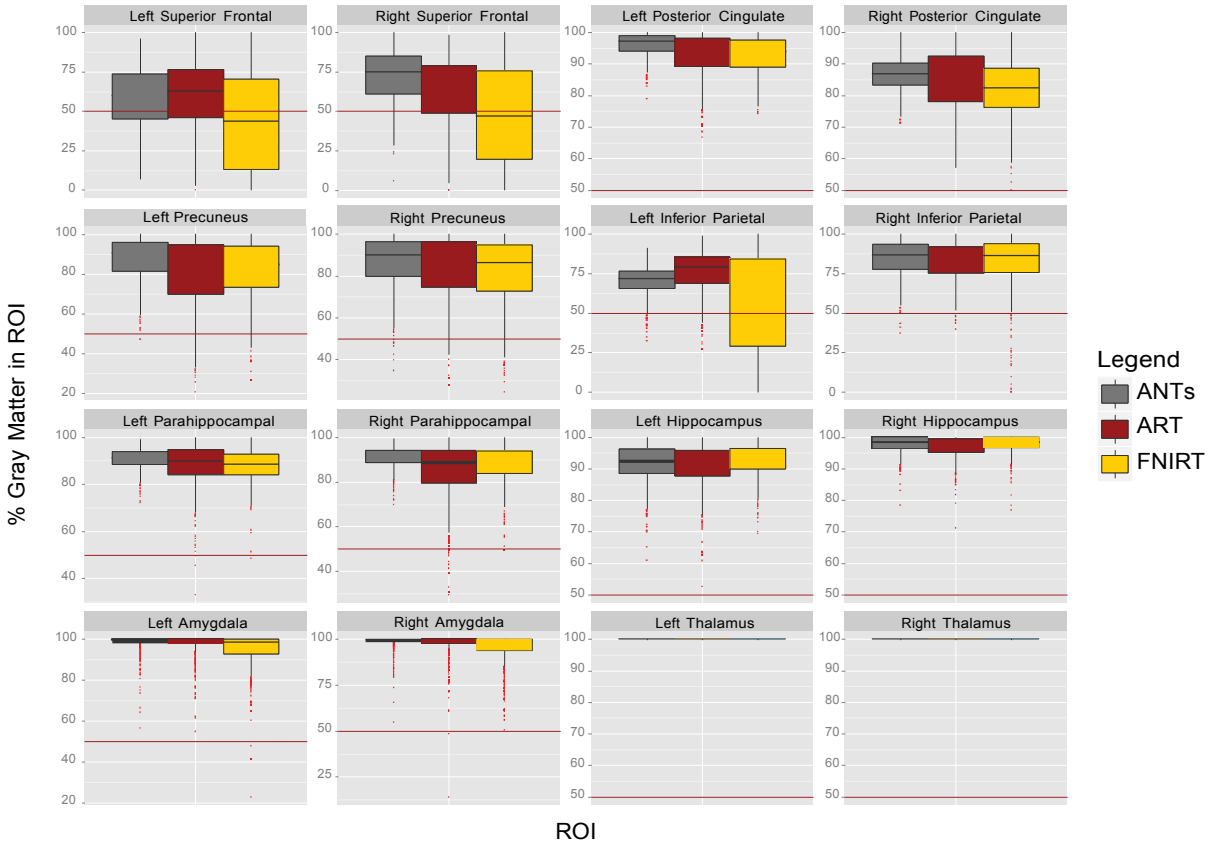


Figure 2: **Percentage of gray matter in transformed ROIs.** Boxplots for each of 16 ROIs associated with the default mode network show the relative performance of the three nonlinear registration methods, ANTs, ART, and FNIRT.

Table 2: **Comparison of nonlinear registration methods for default mode network relevant cortical ROIs.** P-values generated using a linear mixed effects model. **Bolded** p-values indicate significance while accounting for multiple comparisons using the false discovery rate (FDR) method.

Region of Interest (ROI)	ANTs vs. ART	ANTs vs. FNIRT	ART vs. FNIRT
Left Superior Frontal	0.85	**4.15E-33**	**4.63E-27**
Right Superior Frontal	**2.87E-21**	**9.57E-65**	**1.99E-23**
Left Posterior Cingulate	**8.93E-31**	**4.41E-38**	0.60
Right Posterior Cingulate	**3.14E-07**	**3.99E-31**	**4.60E-08**
Left Precuneus	**3.04E-19**	**4.27E-17**	**3.33E-02**
Right Precuneus	**2.25E-05**	**4.42E-11**	0.05
Left Inferior Parietal	**4.82E-23**	**1.34E-16**	**2.37E-32**
Right Inferior Parietal	**1.49E-04**	**7.20E-05**	0.21
Left Parahippocampal	**1.53E-11**	**6.40E-24**	0.43
Right Parahippocampal	**7.29E-25**	**1.64E-18**	**3.20E-07**

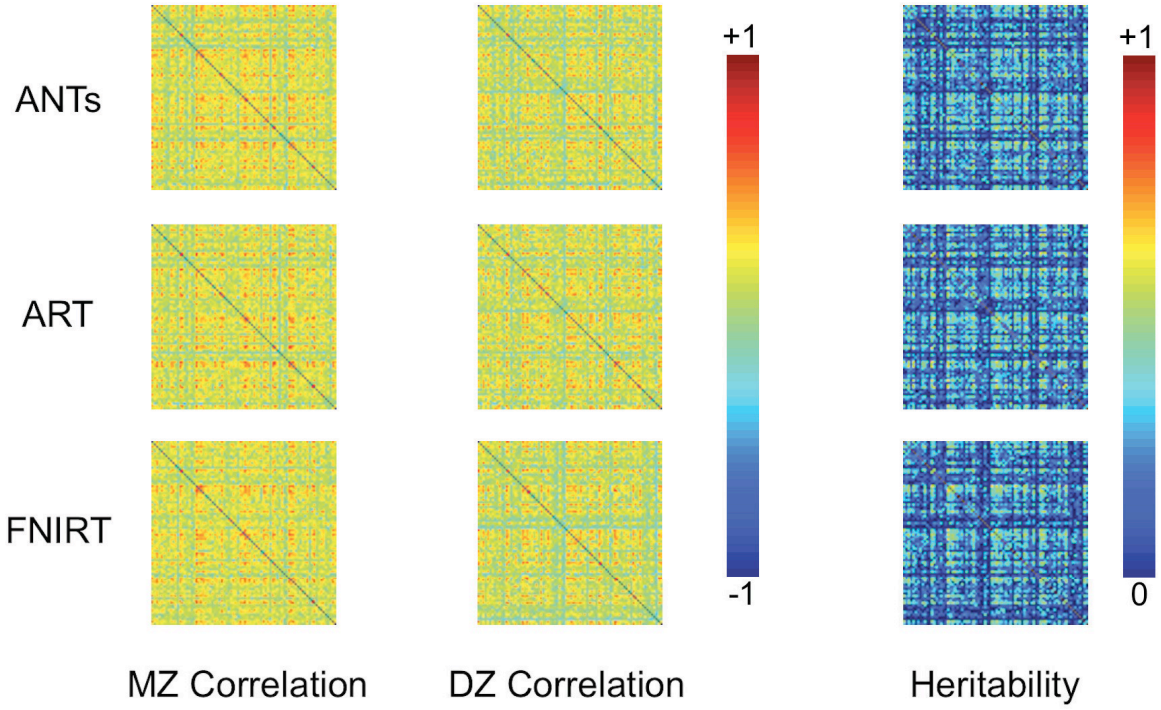


Figure 3: **Heritability of the Functional Connectome.** We present the results from the functional connectome across the three nonlinear registration techniques we used to spatially normalize the ROIs used as nodes in the network. MZ correlation represents the intra-class correlation (ICC) in the monozygotic twins, DZ correlation represents the ICC in the same-sex dizygotic twins, and the heritability column represents the additive genetic contribution to the variance in each connection in the network.

Table 3: **Comparison of Methods.** We summarize the mean across all connections±SD of the intra-class correlation (ICC) in the monozygotic (MZ) twins, ICC of the same-sex dizygotic (DZ) twins, and the additive genetic contribution to connectivity variance. ART and ANTs did not differ significantly in their assessment of heritability across the entire connectivity network.

Method	MZ Correlation	DZ Correlation	Heritability
ANTs	0.22±0.16	0.16±0.17	0.19±0.17
ART	0.21±0.16	0.18±0.17	0.19±0.17
FNIRT	0.22±0.15	0.15±0.18	0.17±0.17

ACKNOWLEDGEMENTS

Funded in part by NIH ENIGMA Center grant U54 EB020403, supported by the Big Data to Knowledge (BD2K) Centers of Excellence program, and R01 AG041915 and AG040060 (PI Thompson).

REFERENCES

[1] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., et al., "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *Neuroimage* **46**(3), 786–802 (2009).

[2] Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V., "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis," *Proceedings of the National Academy of Sciences* **100**(1), 253–258 (2003).

[3] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M., "FSL," *Neuroimage* **62**(2), 782–790 (2012).

[4] Fischl, B., "Freesurfer," *Neuroimage* **62**(2), 774–781 (2012).

[5] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage* **31**(3), 968–980 (2006).

[6] Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C., "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage* **54**(3), 2033–2044 (2011).

[7] Ardekani, B. A., Guckemus, S., Bachman, A., Hoptman, M. J., Wojtaszek, M., and Nierenberg, J., "Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans," *Journal of Neuroscience Methods* **142**(1), 67–76 (2005).

[8] Cox, R. W., "AFNI: what a long strange trip it's been," *Neuroimage* **62**(2), 743–747 (2012).

[9] Sporns, O., "The human connectome: a complex network," *Annals of the New York Academy of Sciences* **1224**(1), 109–125 (2011).

[10] Fornito, A., Zalesky, A., Bassett, D. S., Meunier, D., Ellison-Wright, I., Yücel, M., Wood, S. J., Shaw, K., O'Connor, J., Nertney, D., et al., "Genetic influences on cost-efficient organization of human cortical functional networks," *The Journal of Neuroscience* **31**(9), 3261–3270 (2011).

[11] Chiang, M.-C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., Klunder, A. D., Toga, A. W., McMahon, K. L., De Zubicaray, G. I., et al., "Genetics of brain fiber architecture and intellectual performance," *The Journal of Neuroscience* **29**(7), 2212–2224 (2009).

Embedded sparse representation of fMRI data via group-wise dictionary optimization

Dajiang Zhu[a], Binbin Lin[b], Joshua Faskowitz[a], Jieping Ye[b], Paul M. Thompson[a]
[a]Imaging Genetics Center, University of Southern California, CA, USA;
[b]University of Michigan, MI, USA

ABSTRACT

Sparse learning enables dimension reduction and efficient modeling of high dimensional signals and images, but it may need to be tailored to best suit specific applications and datasets. Here we used sparse learning to efficiently represent functional magnetic resonance imaging (fMRI) data from the human brain. We propose a novel embedded sparse representation (ESR), to identify the most consistent dictionary atoms across different brain datasets via an iterative group-wise dictionary optimization procedure. In this framework, we introduced additional criteria to make the learned dictionary atoms more consistent across different subjects. We successfully identified four common dictionary atoms that follow the external task stimuli with very high accuracy. After projecting the corresponding coefficient vectors back into the 3-D brain volume space, the spatial patterns are also consistent with traditional fMRI analysis results. Our framework reveals common features of brain activation in a population, as a new, efficient fMRI analysis method.

Keywords: sparse learning, fMRI

1. INTRODUCTION

Sparse learning is widely used in image and signal processing and in biomedical and computational neuroscience, as a general and effective solution to represent high-dimensional features in low-dimensional space [1]. Recently, there has been increasing interest in applying sparse representation methods to analyze functional magnetic resonance images of the brain [2-3]. Unlike other applications which pursue a compact or abstract representation of the original data, sparse learning of fMRI signals has additional justifications from a neuroscience point of view: though the exact mechanism is largely unknown, the human brain is widely considered to include a collection of specialized functional networks that flexibly interact when different brain functions are performed [4]. In such a scenario, activity in a specific brain region might recruit different neuroanatomical regions/networks in a temporal sequence, and the same brain region may also participate in multiple functional processes under different internal and/or external circumstances [5]. In such a situation, the same brain region may be involved in multiple functional activities and as a result, a single fMRI signal also tends to be composed of various components corresponding to multiple functional sources, called functional networks. Sparse representation of fMRI signals naturally accounts for the neuroscience principle that each single fMRI time-series can be approximated as a linear combination of common "blocks", which are so called *dictionary atoms*. And each dictionary atom also contributes to multiple fMRI signals simultaneously and the level of contribution is encoded in the coefficient matrix.

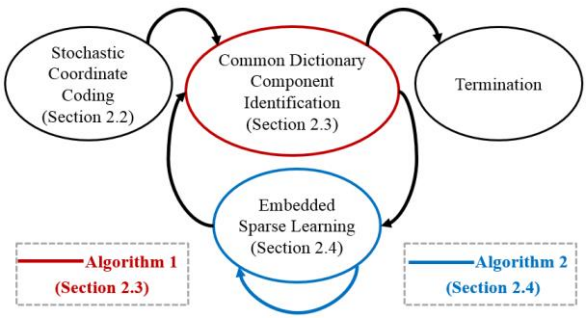


Figure 1. Illustration of the framework.

Here we propose a novel embedded sparse representation (ESR) framework, in which we explore the most consistent and common dictionary atoms, across different subjects, via an iteratively group-wise dictionary optimization procedure. In our proposed framework, we introduce additional criteria to make the learned dictionary atoms more consistent across different subjects; we embed this additional optimization procedure into the sparse learning problem. The framework is summarized in **Figure 1**. The most encouraging part of the results is that, given no prior expectation, we were able to successfully identify four optimized dictionary atoms that closely followed the external task stimulus. Moreover, after projecting the corresponding coefficient vectors back to the 3-D brain volume space, the spatial patterns are also consistent with the traditional fMRI analysis results. These results suggest that the proposed framework may reveal common features of brain activation, offering a new and efficient fMRI analysis method.

2. METHODS

2.1 Framework overview

We first define some notation to better illustrate the algorithms that follow:

1. D_j: the dictionary of subject *j*;
2. $d_{j,m}$: the m^{th} atom of the dictionary belonging to subject *j*;
3. $D_{j,F}$: the fixed dictionary for subject *j*, which only contains the fixed dictionary atoms;
4. s_j^i: similarity. *j* and *i* are the indices of the subjects and dictionary atoms, respectively.
5. *N and K*: the number of subjects and dictionary size.
6. *Corr* (. , .): Pearson Correlation Coefficient.
7. T^i: the i^{th} template signal.

In brief, the framework is composed of three sections. First, we perform stochastic coordinate coding [6] for each subject separately. The learned individual dictionary is then used for later exploration of the group-wise template signal. Then we identify a template signal so that each subject has at least one dictionary atom that is "similar" to this template signal. Lastly, we fix this atom within the individual dictionary training and tune it towards the template signal, during each iteration of the ESR process.

Overview of the framework

```
1. Dictionary initialization:
   for j=1 to N do
      D_j ← Stochastic Coordinate Coding (section 2.2)
   end for
2. Embedded sparse representation via group-wise dictionary optimization
   for i=1 to K do
      Explore template signal (T_i) and initialize D_F (Algorithm1 – section 2.3)
         if T_i is null do
            Terminate
         end if
      Optimize dictionary and update D_F (Algorithm 2 – section 2.4)
   end for
```

2.2 Stochastic coordinate coding

Given a data set $X = (x_1 \cdots x_n)$ of input signals, each signal is a p-dimensional vector, i.e., $x_i \in \mathbb{R}^p, i = 1, \dots, n$. Our aim is to learn a dictionary and sparse codes (the coefficient matrix) from these input signals so that each input signal may be approximated by a linear combination of a small number of basis elements from the dictionary:

$$\min \sum_{i=1}^{n} f_i(D, z_i) = \sum_{i=1}^{n} \frac{1}{2} \| Dz_i - x_i \|^2 + \lambda \| z_i \|_1 \quad (1)$$

λ is the regularization parameter. Each z_i is commonly called the sparse code. Here $D = (d_1 \cdots d_m) \in \mathbb{R}^{p \times m}$ is called the dictionary and each basis d_j satisfies $\| d_j \| \leq 1$. In this paper, X is the whole brain fMRI signals (n is the number of

voxels). D is the set of common "blocks" that can represent each voxel's signals by their linear combinations. Z is the coefficient matrix. An important characteristic of using sparse representation for fMRI signals is that the coefficient weight matrix (Z) naturally reveals the spatial patterns among those reconstructed brain networks [2, 3]. Here, we adopted stochastic coordinate coding (SCC) [6] to solve this ordinary sparse learning problem.

2.3 Common dictionary component identification (Algorithm 1)

Recent studies [3] suggest that some neuroscientifically meaningful dictionary atoms can always be achieved when the dictionary size lies within a reasonably wide range. In other words, some consistent dictionary atoms can be found across different subjects, even if each person's data is subjected to sparse learning separately. Inspired by this, we set out to explore and identify these common dictionary atoms. The core idea is that, for each subject, we calculate the consistency of its dictionary atoms with those belonging to other subjects. If we can find such a dictionary atom T, such that there exists at least one dictionary atom in all the other subjects that has a high correlation with T, then we call T a template signal candidate. After checking all the subjects, the dictionary atom possessing the maximal group-wise consistency (correlation) will be marked as a template signal. At the same time, those dictionary atoms that have the highest correlation with the template signal in other subjects (one atom for each subject) will be recorded for later dictionary initialization in algorithm 2.

Algorithm 1 (Identify the *i th* common dictionary atom)

1. For each subject as a center, calculate the maximum group-wise consistency:
   for j=1 to N do
   $s_j^i \leftarrow \max(\ \sum_{j,m} \max(\ corr(\ d_{j,m}, d_{l,n}\ )\ )$, where $1 \le l \le N$, $j \neq l$, $i \le m$, $n \le K$, corr (. , .) $\ge Threshold_H$ and corr (. , T) $\le Threshold_L$
   end for
2. Find the center subject that has the maximum group-wise consistency:
   c = j, where $s_c^i = \max(s_j^i)$
3. Update $D_{j,F}$:
   for j=1 to N do
   $D_{j,F} \leftarrow d_{j,*}$, where $d_{j,*}$ contributes to s_c^i
   end for
4. Update T^i
   $T^i \leftarrow d_{c,*}$

Note that we have two predefined thresholds in the above algorithm. $Threshold_H$ is set as 0.5, which might be considered a fair indicator of moderately high correlation. $Threshold_L$ is used for filtering those template candidates that have high correlations with the template signals identified previously. In this paper, we set Threshold_L as 0; this means that the new template signal must be completely different from the previous ones. If we cannot find a template signal because of either the threshold constraint ($Threshold_H$) or a conflict with the previous templates ($Threshold_L$), then T^i will be null and the whole process will be terminated.

2.4 Dictionary optimization (Algorithm 2)

The main objective of algorithm 2 is to force the identified dictionary atoms to change towards the corresponding template signal (identified in Algorithm 1) and to maintain precision, as far as possible. Specifically, we first generate an initial dictionary using random data and replace the first *i* atoms with the fixed ones achieved in the previous iterations. Then we move the i^{th} atom with one step to make it more consistent with the template signal and deliver it to the SCC machine. The first *i* dictionary atoms would never be updated in the current sparse learning process. For example, when we optimize the second dictionary atom, we replace the first two atoms which were randomly generated with $D_{j,F}$. $D_{j,F}$ includes two atoms: the first one comes from the previous optimization (Algorithm 2) and the second one is obtained from Algorithm 1, in this round. During each iteration of algorithm 2, we keep the first atom unchanged, adjust the second atom with Δd, and update the other atoms as needed. Recall that before identifying common dictionary atoms, we have performed individual sparse learning with the same configuration (sparsity level and number of iterations) multiple times. Thus we acquire an estimate of the maximum residual (LASSO error) that we can tolerate, if some dictionary

atoms were fixed. In this paper, we perform SCC one hundred times for each subject and set the error Threshold_E as the mean plus the standard deviation derived from the multiple runs.

Algorithm 2 (Dictionary optimization for the i^{th} common atom)

```
for j=1 to N do
    Δd ← ( T^i - d_{j,i}) / StepNum
    do
        generate D_j with random data
        D_j ← D_{j,F}, d_{j,i} ← d_{j,i} + Δd
        Sparse learning with fixed dictionary atoms (d_{j,0}, …, d_{j,i})
    while residual ≤ Threshold_E and d_{j,i} ≠ T^i
end for
```

3. RESULTS

3.1 Dataset and sparse learning parameters

Our data source was the publicly available Human Connectome Project (HCP) Q1 release [7]. The HCP Q1 dataset has seven task-based fMRI datasets from 68 participants. In this paper, to illustrate the method on a specific problem, we randomly selected 10 subjects and focused on the motor task. More details of data acquisition and preprocessing may be found in [7].

3.2 Common dictionary atom

After applying the methods in section 2, we identified 4 template signals through Algorithm 1 and they were successfully optimized in algorithm 2. We listed the corresponding dictionary atoms before algorithm 1, after algorithm 1, and after algorithm 2 as "Before", "Initial" and "Optimized" in **Figure 2**. The white curves represent the external stimuli. Different subjects are represented by different colors.

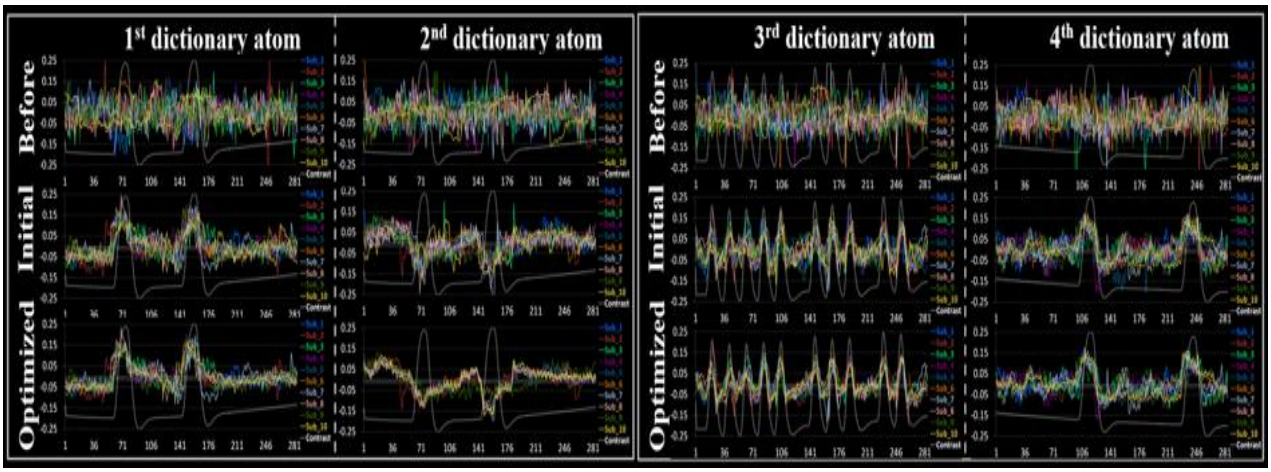


Figure 2. Four consistent dictionary atoms we identified. The first row ("Before") shows the signals at the same location among different subjects' dictionaries. The second ("Initial") and third row ("Optimized") show the results after applying algorithm 1 and algorithm 2, respectively. The white curves represent the external stimuli (*1st and 2nd dictionary atom - tongue movement*, *3rd* dictionary atom - *visual cue*, *4th* dictionary atom - *left hand*).

Before performing *algorithm 1*, there are no correspondences among the dictionaries computed from multiple subjects. Even though there are some consistent patterns in different subjects' dictionaries, we do not know what these patterns look like and where those atoms are located. Hence the signals in the same location display a random pattern (the rows marked with "Before"). In *algorithm 1*, the pattern with the maximum group-wise similarity was explored during each

iteration and the identified dictionary atom in each subject was assigned to the same location in the individual dictionary. The rows marked with "Initial" show the results after *algorithm 1*. As we can clearly see, there are still variations among different dictionary atoms, but the overall pattern is evident. Lastly, the *algorithm 2* iteratively optimized the above dictionary atoms and made them as similar as possible, until they are the same as the template signal or the residual exceeds the threshold. The optimized results are shown in the rows marked with "Optimized". The corresponding dictionary atoms among different subjects are more "similar", compared to the "Initial" ones. The changes in consistency (correlation) at each iteration during the dictionary optimization are shown in **Figure 3**. As the maximum iteration number is seven when optimizing the first dictionary atom, there are only seven statistic values in the first subfigure in **Figure 3**.

The most interesting finding is that, after exploration and optimization of the consistent dictionary atoms, the four consistent patterns identified can be perfectly matched with the external stimuli curves. In **Figure 2**, the first and second optimized dictionary atoms correspond to the tongue movement stimulus. Specifically, the first atom is positively correlated with the stimulus; however, the second atom is anti-correlated with the same stimulus. The third and fourth atoms are consistent with the stimuli of visual cue and left hand. Recall that we did not enforce or provide any prior knowledge of the task stimuli, and these stimuli-related dictionary atoms are automatically learned from the original fMRI data.

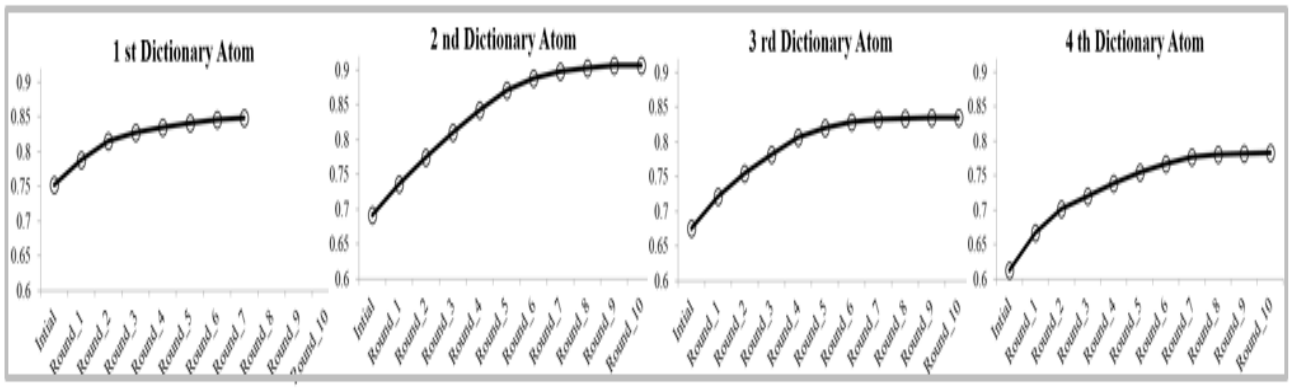


Figure 3. The improvement in consistency during the optimization in algorithm 2.

3.3 Spatial patterns in the identified common dictionary atoms

In the previous section, we displayed four consistent dictionary atoms after our optimization, which perfectly follow (or are anti-correlated with) the external stimuli. One natural question arises: Are their spatial patterns also similar to the activation map derived from traditional methods, such as the general linear model (GLM)? To answer this question, we show the spatial patterns (by projecting the corresponding coefficient vectors back to the 3D brain volume space) of these four dictionary atoms in **Figure 4**. We showed both results derived from our method (ESR) and GLM of ten subjects (*each row*). The GLM results of these four atoms correspond to tongue, anti-tongue (opposite of the tongue movement stimulus), visual cue and left hand. From visual examination, the spatial patterns from our method are visually similar to those activation maps detected by GLM. To quantitatively measure their similarity, we calculated the overlap rate ((ESR ∩ GLM)/(ESR U GLM)); results are shown in **Table 1**.

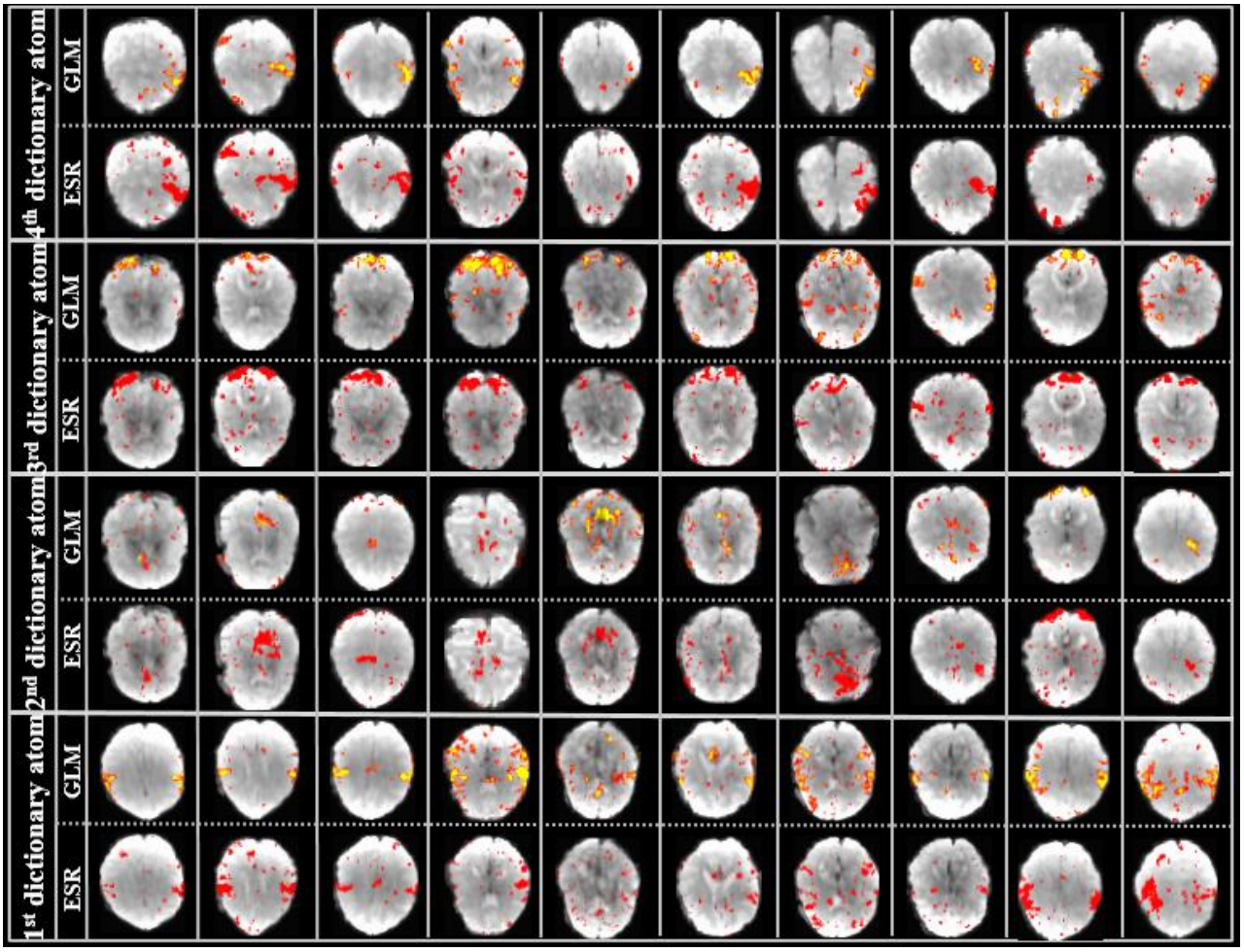


Figure 4. The corresponding spatial patterns of the identified four dictionary atoms.

Both results indicate two encouraging findings: 1) there are some common and stable dictionary atoms that exist across different subjects, even though the sparse learning algorithm is performed individually; 2) for those common dictionary atoms, their temporal patterns tend to follow some external stimuli, and most of their spatial patterns display an obvious similarity to the results derived from a traditional activation detection method (i.e., the GLM). However, we also noticed some inconsistent cases. For example, the 2nd atom displayed a more arbitrary pattern both in ESR and GLM. One possible explanation is that unlike the situations under external stimuli, the functional interactions under the "anti-task" are much more complicated. Another possible reason is that the dictionary size might be too large for this dataset. Some latent patterns, such as the Default Mode Network (DMN) which is known as an "anti-task" feature when performing task, could be over-decomposed into multiple components. This is worth investigating in future work.

Table 1. Overlap rate between ESR and GLM. "Sub" means subject, numbered 1 to 10.

%	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10
1st	18.2	16.7	22.6	25.0	23.9	25.9	18.1	28.9	28.7	23.4
2nd	16.5	13.4	14.0	14.1	25.4	16.4	16.0	24.2	14.0	21.8
3rd	22.0	20.9	31.9	20.5	28.2	29.0	22.1	21.9	37.6	23.0
4th	22.9	21.2	18.9	26.9	22.8	24.6	18.4	24.9	32.2	25.3

4. CONCLUSIONS

In this paper, we proposed a novel embedded sparse representation framework, in which we aimed to explore the most common and consistent dictionary atoms across different subjects. To do this, we used an iterative group-wise dictionary optimization procedure. **Figure 2** illustrates that, given no prior expectation or prior, we can successfully learn multiple common dictionary atoms based on ordinary sparse learning results and the learned template signals (dictionary atoms) perfectly match the external stimuli. In addition, the corresponding spatial patterns are also consistent with prevailing neuroscience knowledge.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Vinje, W.E., Gallant, J.L., "Sparse coding and decorrelation in primary visual cortex during natural vision," **Science** 287, 1273-1276 (2000).

[2] Varoquaux, G., Schwartz, Y., Pinel, P., Thirion, B., "Cohort-level brain mapping: learning cognitive atoms to single out specialized regions," IPMI - **Information Processing in Medical Imaging**. LNCS, 7917, 438-449 (2013).

[3] Lv, J., Jiang, X., Li, X., Zhu, D., Zhang, S., Zhao, S., Chen, H., Zhang, T., Hu, X., Han, J., Ye, J., Guo, L., Liu, T., "Holistic Atlases of Functional Networks and Interactions Reveal Reciprocal Organizational Architecture of Cortical Function," **IEEE Transactions on Biomedical Engineering**, 62(4), 1120-31 (2014).

[4] Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach N.U., church, J.A., Miezin, F.M., schlaggar, B.L., Petersen, S.E., "Functional Brain Networks Develop from a "Local to Distributed" Organization," **PLoS Computational Biology**, 5(5) (2009).

[5] Doucet, G., Naveau, M., Petit, L., Zago, L., Crivello, F., Jobard, G., Delcroix, N., Mellet, E., Tzourio-Mazoyer, N., Mazoyer, B., Joliot, M., "Patterns of hemodynamic low-frequency oscillations in the brain are modulated by the nature of free thought during rest," **NeuroImage**, 59 (4), 3194-200 (2012).

[6] Lin, B., Li, Q., Sun, Q., Lai, M., Davidson, L., Fan, W., Ye, J.," Stochastic Coordinate Coding and its Application for *Drosophila* Gene Expression Pattern Annotation," CoRR abs/1407.8147 (2014).

[7] Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Lelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, e., Uqurbil, K., Van Essen, D.C., Glasser, M.F., "Resting-state fMRI in the Human Connectome Project," **NeuroImage**, 80, 144-68 (2013).

Effects of EPI Distortion Correction Pipelines on the Connectome in Parkinson's Disease

Justin Galvis[1], Adam F. Mezher*[1], Anjanibhargavi Ragothaman[1], Julio E. Villalon-Reina[1], P. Thomas Fletcher[2], Paul M. Thompson[1], Gautam Prasad[1]*

**Both authors contributed equally*
1. Imaging Genetics Center, University of Southern California, Marina del Rey, CA, USA
2. Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT, USA

Abstract: Echo-planar imaging (EPI) is commonly used for diffusion-weighted imaging (DWI) but is susceptible to nonlinear geometric distortions arising from inhomogeneities in the static magnetic field. These inhomogeneities can be measured and corrected using a fieldmap image acquired during the scanning process. In studies where the fieldmap image is not collected, these distortions can be corrected, to some extent, by nonlinearly registering the diffusion image to a corresponding anatomical image, either a T1- or T2-weighted image. Here we compared two EPI distortion correction pipelines, both based on nonlinear registration, which were optimized for the particular weighting of the structural image registration target. The first pipeline used a 3D nonlinear registration to a T1-weighted target, while the second pipeline used a 1D nonlinear registration to a T2-weighted target. We assessed each pipeline in its ability to characterize high-level measures of brain connectivity in Parkinson's disease (PD) in 189 individuals (58 healthy controls, 131 people with PD) from the Parkinson's Progression Markers Initiative (PPMI) dataset. We computed a structural connectome (connectivity map) for each participant using regions of interest from a cortical parcellation combined with DWI-based whole-brain tractography. We evaluated test-retest reliability of the connectome for each EPI distortion correction pipeline using a second diffusion scan acquired directly after the participants' first. Finally, we used support vector machine (SVM) classification to assess how accurately each pipeline classified PD versus healthy controls using each participants' structural connectome.

Keywords: diffusion-weighted imaging, echo planar imaging, non-linear registration

1. Introduction

Diffusion-weighted imaging (DWI) – a form of magnetic resonance imaging (MRI) – has grown in popularity as it enables detailed study of white matter microstructural abnormalities and connectivity patterns that are undetected in standard anatomical MRI images. Echo-planar imaging (EPI) is typically used for DWI, to generate high signal-to-noise images with a fast acquisition time. As EPI assumes a homogeneous static magnetic field that is hard to actually achieve in a typical brain MRI, it is susceptible to EPI distortions – nonlinear geometric distortions arising from inhomogeneities in the static magnetic field – primarily around air-tissue interfaces such as the sphenoid, ethmoid and frontal sinuses. Typically, the phase-encoding direction lies along 1 dimension of the image, usually the coronal axis (y-axis), and thus as do the distortions. Here we aimed to study how the choice of EPI distortion correction pipelines affects the resulting structural connectome in the study of Parkinson's disease. We build on prior work by using two statistical techniques to compare three image-based correction approaches: (1) a widely used 3D inverse-consistent mutual information elastic registration (3DMI) to a T1-weighted target; (2) a single-plane warp to nonlinearly register images to a T2-weighted image in the phase-encoding direction only (PDEC) (Tao, 2009) that was developed specifically for EPI correction; and (3) a "no correction" technique designed as a baseline group (nEC), for comparison. We hypothesized that PDEC would perform better than both the 3DMI and nEC groups because PDEC addresses to the physical sources of

EPI distortion. If the phase-encoding direction is the only dimension affected by EPI distortions, then warping only that direction should be sufficient to correct the image, avoiding unnecessary or incorrect warps of the diffusion image in other dimensions. We assessed connectivity networks from each approach (3DMI vs. PDEC vs. nEC) using intra-class correlation and support vector machines (SVMs) that classified healthy controls and Parkinson's disease (PD) patients in a dataset of 189 participants. Many prior studies have assessed the registration performance of EPI distortion correction techniques, but here we investigated whether effects of these different methods of correction might propagate to higher-level measures such as brain connectivity and classifications based on connectivity networks.

2. Methods

2.1 Participants & Brain Imaging

	Controls	Patients	Total
N	58	131	189
Age (mean ± SD in years)	60.6 ± 10.1	61.6 ± 9.4	61.3 ± 9.6
Sex	37M/21F	83M/48F	120M/69F

Table 1. Demographic information from PPMI including age and sex.

We analyzed data from 189 participants scanned as part of the Parkinson's Progression Markers Initiative (PPMI), including 58 healthy controls and 131 patients with Parkinson's disease (PD). **Table 1** above has a breakdown of our participant demographics. All participants received whole-brain 3-Tesla MRI image scans on Siemens scanners. For each subject, one high-resolution T1-weighted 3D magnetization prepared rapid gradient echo, or MP-RAGE scan (256x240x176 matrix; voxel size = 1.0x1.0x1.0 mm^3; TR = 2300 ms; TI = 900 ms; TE = 2.98 ms; flip angle = 9°), one non-contrast enhanced T2-weighted image, and two identical diffusion MRI sequences (116x116x72 matrix; voxel size = 2.0x2.0x2.0 mm^3; TR = 900 ms; TE = 88 ms) were acquired in the same session. For both diffusion MRI scans, 65 gradient directions were acquired: 1 T2-weighted image with no diffusion sensitization (b_0 image) and 64 diffusion-weighted images (b = 1000 s/mm^2).

2.2 DWI preprocessing & EPI Distortion Correction Techniques

Both DWIs acquired per subject, which we will refer to as DWI_1 and DWI_2, were pre-processed identically. To increase signal to noise ratio (SNR), images were first de-noised using local principal component analysis (Manjon, 2013). Each image was then corrected for eddy current distortions using the FMRIB Software Library (FSL) (Jenkinson, 2012). Using the b_0 as a reference volume, each of the gradient directions were then rotated, or affinely registered using FSL's Linear Image Registration Tool (FLIRT) algorithm to accommodate this correction. Brains were then "skull-stripped" using FSL's Brain Extraction Tool (BET) and underwent voxel-wise diffusion tensor modeling using FSL's DTIFIT algorithm. At this point, the DWI's for both DWI_1 and DWI_2 both underwent each EPI distortion correction pipeline separately. After EPI correction we had six total groupings of images; DWI_1: nEC, 3DMI, PDEC and DWI_2: nEC, 3DMI, PDEC.

2.2a General Elastic Registration for EPI Correction (The "3DMI" Technique)

For the first EPI correction technique, we used a 3D inverse-consistent mutual information elastic registration algorithm, referred to as 3DMI, to correct for the EPI susceptibility artifacts. T1-weighted images were first affinely transformed to the standard 1mm^3 MNI atlas using FSL's FLIRT (degrees of freedom [dof] = 6, cost = correlation ratio; interpolation = trilinear). We then affinely registered each participant's DWI_1 and DWI_2 to his or her T1-weighted image in MNI coordinate space (dof = 12; cost = correlation ratio; interpolation = trilinear). 3DMI was then applied to register and warp the affinely registered DWI_1 and DWI_2 to its corresponding T1-weighted image. Assuming the T1-weighted image is undistorted and that the registration is accurate, this alignment of the DWI to the T1 should correct for the EPI distortions.

2.2b Custom Variational Image-Based Elastic Registration for EPI Correction (The "PDEC" Technique)

Separately, all DWI_1 and DWI_2 scans that had been affinely alighned to their T1 were then corrected for EPI distortions using a previously described elastic registration method restricted to the *y*-phase direction of the scan, which we refer to as Phase Encoding Direction EPI Correction or PDEC. PDEC nonlinearly registers a diffusion image – the image with EPI distortions – to its corresponding T2-weighted image, to account for the EPI susceptibility artifacts. In theory, true EPI distortion exists solely in the phase-encoding direction, denoted by '*y*' below. Thus, PDEC ideally should correct the EPI distortions more accurately. The technique registers the b_0 image to its corresponding T2-weighted structural image using a squared-error penalty as detailed in the equation below, where I_E represents the EPI image and the second part of the equation is an elastic penalty on the displacement '*v*'. The method includes a correction for the conservation of the EPI signal and is solved using gradient descent.

$$\varepsilon_v = \frac{1}{2}\int_{\mathcal{D}} \left[\left(1+\frac{\partial v}{\partial y}\right) I_E(x,y+v,z) - I_S(x,y,z)\right]^2 dxdydz + \frac{\lambda}{2}\int_D |\nabla v|^2 dxdydz,$$

We skull-stripped and affinely registered the T2-weighted images to their corresponding DWI in native diffusion space with FLIRT (dof = 6; cost = correlation ratio; interpolation = trilinear). We then performed PDEC to nonlinearly warp each DWI_1 and DWI_2 to its corresponding T2-weighted image along the phase-encoding direction only (for these data, the anterior-posterior plane, also denoted as the '*y*' plane).

Images post-PDEC were affinely aligned to standard 1 mm^3 MNI coordinate space by first affinely aligning the T2-weighted image to the subject's corresponding T1-weighted image in MNI space with FLIRT (dof = 12; cost = correlation ratio; interpolation = trilinear) to normalize each subjects' images.

This transformation was then applied to each DWI_1 and DWI_2 image to upsample them to standard MNI coordinates.

2.2c Baseline Pipeline with No EPI Correction

A "no correction" group (nEC) was included as a control method as to set a baseline comparison for the two distinct EPI correction pipelines. Only a single affine transformation using FSL FLIRT was used on the preprocessed DWIs to the subject's corresponding T1-weighted image in the MNI coordinate space (dof = 12; cost = correlation ratio; interpolation = trilinear). **Figure 2** illustrates the three pipelines described above.

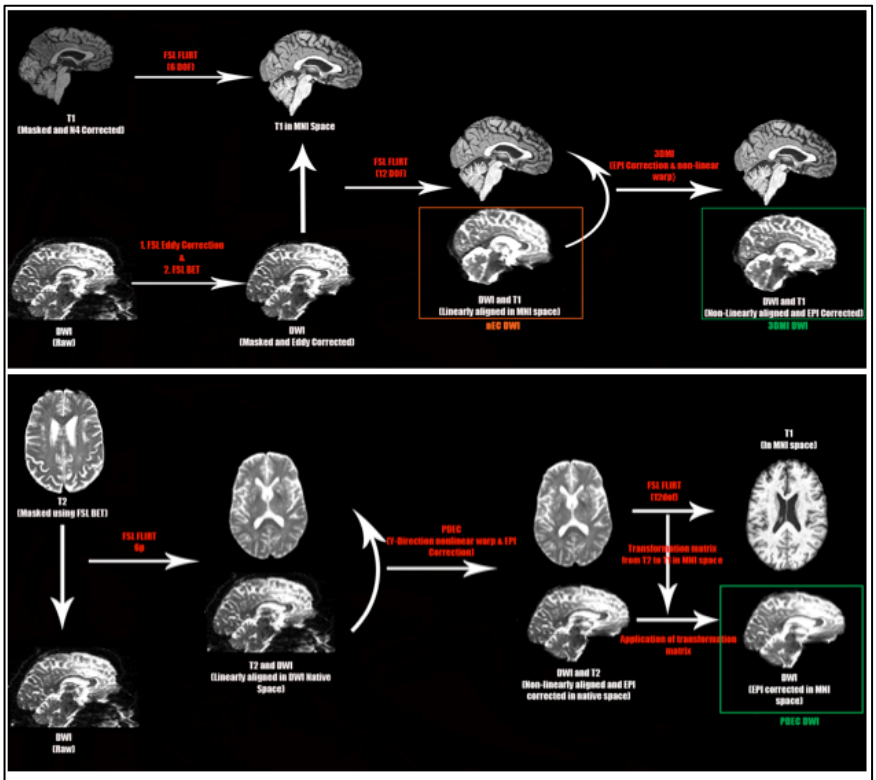


Figure 2. A visualization of each step involved in all three processing groups. 3DMI and nEC (top) are processed using only the diffusion and T1-weighted images. nEC performs no EPI correction, while the 3DMI group has EPI performed. In PDEC (bottom), the diffusion, T2, and T1-weighted images are all used in processing, with EPI distortions being corrected using only the DWI and T2.

2.3 Cortical Measurements

Each subject's T1-weighted MRI scan was processed using FreeSurfer version 5.3 (http://surfer.nmr.mgh.harvard.edu/). 34 cortical gray matter labels, defined by the Desikan-Killiany atlas (Desikan, 2006), were extracted for each hemisphere (68 labels in total). Using nearest neighbor interpolation, we affinely aligned the labels from the T1 native space to the standard 1-mm^3 MNI space by applying the same transformation used on the subject's T1-weighted image when it was aligned to MNI space. This allowed the diffusion images and the cortical parcellations to be aligned for the computation of the subjects' connectomes.

2.4 Tractography and Structural Connectivity Networks

For each comparison group, whole-brain tractography was combined with the FreeSurfer-generated cortical labels to map each subject's cortical connectivity. First, we computed tractography using the EPI corrected, MNI aligned DWI_1 and DWI_2 images using Constant Solid Angle Orientation Density Function (CSA-ODF's) (Aganj, 2010) generated from an optimized global tractography algorithm (Aganj, 2011; Prasad, 2013) with 10,000 fibers. From this, we constructed connectivity networks by computing the number of fibers intersecting pairs of the 68 cortical regions of interest (ROIs) from the FreeSurfer parcellation. The number of fibers between a pair of ROIs quantified the connectivity between each of the 68 regions. We normalized this value by the total number of fibers extracted per brain and arranged them into a symmetric 68x68 connectivity matrix with 34 ROIs in both the right and left hemispheres.

2.5 Intraclass Correlation Coefficient

The intraclass correlation coefficient (ICC) (McGraw, 1996) is a reliability index that can quantify the consistency or reproducibility of measurements. We used ICC (3, 1)=(BMS-EMS)/(BMS+EMS), where BMS is the between-targets mean square and EMS is the residual mean squares. This ICC variant assumes each target (participant) is rated by a fixed set of 2 judges (in this case, DWI_1 and DWI_2). We computed ICC (3, 1) at every edge that was assigned a positive weight across the dataset in the 68x68 connectivity networks (described above). We then used paired-sample Student's *t*-tests to compare this distribution of ICC values at each edge, across the three EPI correction techniques.

2.6 Cross-Validated Support Vector Machine Classification

We used linear kernel support vector machines (SVMs) (Cortes, 1995) to classify participants as either a healthy control or a patient with PD. SVM learns a hyperplane in the d=(68/2)=2278 dimensional connectivity network space. This hyperplane is optimized to separate the two classes that exist in the data. We set the penalty parameter C, that controls the number of support vectors or observations used in the hyperplane, to a default value of 1.0. This has been shown to be robust across a wide variety of datasets (Pedregosa, 2011). Because the class sizes in our dataset differed, we used balanced accuracy (Brodersen, 2010) to quantify classification performance instead of typical accuracy that may not correctly assess the performance on the smaller class. Balanced accuracy is the mean of sensitivity and specificity and equally weights the accuracy on each class. For all three EPI correction techniques and for each subject, we used 10-times repeated 10-fold cross-validation (Kohavi, 1995) to assess classification accuracy across correction methods.

3. Results

The ICC analysis in **Table 2** contrasts the mean of the test-retest statistic for all three processing pipelines. The PDEC method consistently held the highest ICC in all groups and was significantly higher (p-value < 0.05) than 3DMI in all cases. A higher ICC in the PDEC group shows higher consistency in the EPI correction and aligning of the DWI to the T2-weighted image. Even so, PDEC was marginally different (p~0.05) from nEC in only the control group.

In **Table 3**, we list the balanced accuracy results from the cross-validated SVM classifier of the control group vs. PD in both DWI_1 and DWI_2. 3DMI consistently showed significantly higher accuracy (comparing accuracy across the ten repeated measures) compared to both nEC and PDEC. In the DWI_1 dataset, PDEC ranked second in performance and performed better (significant when correcting for multiple comparisons) than nEC.

EPI Correction Technique	Controls	PD	Combined
nEC	0.825	0.808	0.822
3DMI	0.802	0.797	0.808
PDEC	0.832	0.810	0.824

Table 2. Mean intraclass correlation coefficient (ICC) are listed for each EPI distortion correction pipeline and broken into controls, patients, and a combination of both groups. An ICC value was computed for each connection in the connectivity network across all subjects in their respective group. The PDEC approach was marginally significantly different from no EPI correction (nEC) in the controls and significantly different in performance from 3DMI in all cases.

EPI Correction Technique	DWI_1 Balanced Accuracy	DWI_2 Balanced Accuracy
nEC	51.38%	53.61%
3DMI	60.10%	59.50%
PDEC	55.48%	52.06%

Table 3. We present the linear SVM 10-times repeated 10-fold cross-validation results using the denoised cortical connectivity networks. The balanced accuracy for 3DMI was significantly higher ($p << .05$) compared to the cases of no EPI distortion correction and correction using PDEC based on the T2 image.

4. Conclusions

Though EPI distortion algorithms have been studied through standard registration metrics, here we study how EPI distortion correction affects brain connectivity measures in the analysis of Parkinson's disease, specifically connectomics. Errors induced from pre-processing issues can be propagated to higher-level derived measures of brain connectivity, and classifiers based on those measures, which we assess here for the first time.

The DWI_1 and DWI_2 images in the PDEC group had higher ICC values, compared to those in the 3DMI and nEC groups. PDEC may provide a more consistent, but not necessarily more accurate, EPI correction based on these test-retest metrics.

In terms of disease classification, images that were EPI corrected using 3DMI had better classification accuracy than those corrected using PDEC and nEC. 3DMI uses nonlinear warps in three planes rather than only in the phase-encoding direction, and this may influence classification accuracy. It is possible, that the 3D nonlinear alignment of the DWI to the T1 provides a higher accuracy in terms of matching the intricate gyral patterns of the cortex, which is of utmost importance when computing the structural connectomes. The nature of the structural connectomes is multimodal, where discrete cortical regions are segmented from T1-weighted images and tractographies are reconstructed from DWI images. Hence, a better alignment of both image modalities is critical to resolve the intersection of each cortical region with the right reconstructed fiber/streamline. Our results also suggest that the anatomical correctness of the alignment is likely more important in terms of computing structural connectomes suitable for classification purposes, than an accurate correction of the diffusion gradient directions accounting for 3D non-linear deformation fields.

8. Acknowledgements

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

5. References

[1] Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, "Reconstruction of the orientation distribution function in single‐and multiple‐shell q‐ball imaging within constant solid angle", Magnetic Resonance in Medicine, 64(2), 554-566, 2010.

[2] I. Aganj, C. Lenglet, N. Jahanshad, E. Yacoub, N. Harel, P.M. Thompson, and G. Sapiro, "A Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography" Medical Image Analysis, 15(4), 414-425, 2011.

[3] K.H. Brodersen, C.S. Ong, K.E. Stephan, and J.M. Buhmann, "The balanced accuracy and its posterior distribution", IEEE International Conference on Pattern Recognition (ICPR), 20, 3121-3124, 2010.

[4] C. Cortes and V. Vapnik, "Support-vector networks", Machine learning, 20(3), 273-297, 1995.

[5] R.S. Desikan, F. Ségonne, B. Fischl, B.T. Quinn, B.C. Dickerson, D. Blacker et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest", NeuroImage, 31(3), 968-980, 2006.

[6] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, and S.M. Smith, "FSL", NeuroImage, 62(2), 782-790, 2012.

[7] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", IJCAI, 14(2), 1137-1145, 1995.

[8] K.O. McGraw and S.P. Wong, "Forming inferences about some intraclass correlation coefficients", Psychological methods, 1(1), 30-46, 1996

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., "Scikit-learn: Machine learning in Python", The Journal of Machine Learning Research, 12, 2825-2830, 2011.

[10] G. Prasad, T.M. Nir, A.W. Toga, and P.M. Thompson, "Tractography density and network measures in Alzheimer's disease", IEEE International Symposium on Biomedical Imaging (ISBI), 10, 692-695, 2013.

[11] J. V. Manjon, P. Coupé, L. Concha, A. Buades, D. L. Collins, M. Robles. "Diffusion Weighted Image Denoising using overcomplete Local PCA." PLoS ONE 8(9): e73021. doi:10.1371/journal.pone.0073021

[12] Tao R, Fletcher PT, Gerber S, Whitaker RT. "A Variational Image-Based Approach to the Correction of Susceptibility Artifacts in the Alignment of Diffusion Weighted and Structural MRI," Information processing in medical imaging : proceedings of the conference. 2009; 21:664-675.