

MULTI-SITE META-ANALYSIS OF MORPHOMETRY

Neda Jahanshad¹, Joshua Faskowitz¹, Gennady Roshchupkin⁵, Derrek P. Hibar¹, Boris A. Gutman¹, Nicholas J. Tustison², Hieab H.H. Adams^{3,4}, Wiro J. Niessen⁵, Meike W. Vernooij^{3,4}, M. Arfan Ikram^{3,4,6}, Marcel P. Zwiers⁷, Alejandro Arias Vasquez⁸, Barbara Franke⁸, Jennifer L. Krolf^x, Benson Mwangi^x, Jair C. Soares^x, Alex Ing⁹, Sylvane Desrivieres⁹, Gunter Schumann⁹, Narelle K. Hansell^c, Greig I. de Zubicaray^a, Katie L. McMahon^b, Nicholas G. Martin^c, Margaret J. Wright^d, Paul M. Thompson¹
and the Alzheimer's Disease Neuroimaging Initiative*

¹Imaging Genetics Center, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, Marina del Rey, CA, USA

²University of Virginia, Charlottesville VA, USA

³Department of Epidemiology, ⁴Radiology, ⁵Medical Informatics, ⁶Neurology, Erasmus Medical Center, Rotterdam, NL

⁷Donders Institute, ⁸Department of Psychiatry, Human Genetics Radboud UMC, Nijmegen, NL

⁹King's College London, London UK

^xUniversity of Texas, Houston, Houston, TX, USA

^aFaculty of Health and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

^bCentre for Advanced Imaging, University of Queensland, Brisbane, QLD, Australia

^cQIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

^dQueensland Brain Institute, University of Queensland, Brisbane, QLD, Australia

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Submitted as a Regular Paper: IEEE/ACM Transactions on Computational Biology & Bioinformatics - Special Issue - BIODDD 2015

November 8, 2015; Revised May 27, 2016

Corresponding author: Neda Jahanshad
Email: neda.jahanshad@usc.edu

ABSTRACT

Genome-wide association studies (GWAS) link full genome data to a handful of traits. However, in neuroimaging studies, there is an almost unlimited number of traits that can be extracted for full image-wide big data analyses. Large populations are needed to achieve the necessary power to detect statistically significant effects, emphasizing the need to pool data across multiple studies. Neuroimaging consortia, e.g., ENIGMA and CHARGE, are now analyzing MRI data from over 30,000 individuals. Distributed processing protocols extract harmonized features at each site, and pool together only the cohort statistics using meta analysis to avoid data sharing. To date such MRI projects have focused on single measures

such as hippocampal volume, yet voxelwise analyses (e.g., tensor-based morphometry; TBM) may help better localize statistical effects. This can lead to 10^{13} tests for GWAS and become underpowered. We developed an analytical framework for multi-site TBM by performing multi-channel registration to cohort-specific templates. Our results highlight the reliability of the method and the added power over alternative options while preserving single site specificity and opening the doors for well-powered image-wide genome-wide discoveries.

Index Terms—ENIGMA, voxelwise, imaging genetics, multi-channel registration, multi-site, mass univariate, meta-analysis

1. INTRODUCTION

Imaging genetics is an emerging field in which variations in the human genome are related to brain differences. Genome-wide association studies (GWAS), test for statistical associations between brain measures and up to a million single nucleotide polymorphisms (SNPs), or base-pair variants, in the genome. To simplify the screening effort, studies often focus on a single measure extracted from brain scans, such as the overall volume of the hippocampus[1]. Despite success with simple summary measures, the image space contains many more features, e.g., at individual voxels, which can allow for a more complete understanding of the genetic influence on brain structure. Most meta-analyses in neuroimaging have been retrospective literature searches, pooling effect size estimates across published studies. However, when the original study authors are no longer involved, the information available from these studies is often limited to specific coordinates, which may be highly variable from study to study and dependent on imaging parameters. While methods have been developed to incorporate such spatial uncertainty into the meta analysis model [2], this does not specifically improve the power to detect effects across studies. More recently, large neuroimaging consortia have formed to systematically process MRI scans, leading to the testing of statistical effects on harmonized brain measures in a more prospective meta-analytical method, suggesting the need to optimize methods of improving correspondence between images across studies on an image-wide level.

The Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) Consortium (<http://enigma.ini.usc.edu>) in partnership with the CHARGE Consortia have shown that segmentation of brain MRI is reliable enough to detect consistent effects on intracranial and subcortical volumes of single-nucleotide changes in the genome across more than 30,000 individuals in over 48 cohorts worldwide [1, 3, 4] This type of meta-analysis is powerful enough to screen the genome and discover common genetic variants that account for as little as 0.5-1% of the variance in regional brain volume, yet they can be detected as they consistently affect MRI-derived brain structure in populations worldwide. This offers the

potential to study any factor that affects brain structure – from disease, to medication, to single-letter variants in the human genome. In genome-wide association studies, we are discovering that multi-site meta-analysis is crucial to accumulate sample sizes necessary for sufficient power in detecting subtle effects. Most notably, in imaging genetics, single sites rarely have a sample size large enough to pick up effects of single genetic variants.

However, single regional volumetric measures offer only a limited view of the information available in MRI; voxel-based analyses of MRI, which include voxel-based or tensor-based morphometry (VBM and TBM) are complementary and can yield statistical parametric maps of effects on the brain without the need for prior hypotheses or segmentation of brain regions. TBM, for example, computes regional volume differences between each individual brain and a template; voxel-based statistics of the resulting maps has revealed the profile of atrophy in Alzheimer’s disease [5], growth in childhood, and factors that affect them.

Before embarking on meta-analysis of genetic associations using voxelwise measures derived from images, the reliability and reproducibility of extracted measures must be established. Templates used for TBM may be biased if the population under study is not carefully considered. Cohorts also differ in age and the protocols used to scan them, so we developed a distributed processing method to compute site-specific templates and integrate their results for statistical analysis. The resulting method does not require the inter-site transfer of raw scans, which becomes impractical when $N >$ many thousands.

2. METHODS

2.1. Subject Demographics and Image Acquisitions

A random selection of MRI scans from seven cohorts were analyzed. **Table 1** summarizes basic demographic and image acquisition differences across the cohorts. For example, magnetic field strengths ranged from 1.5-4T. Cohorts included: the first and second phases of the Alzheimer’s Disease Neuroimaging Initiative: 1) ADNI1, 2) ADNI2 -although both ADNI, these are two

separate cohorts; both in terms of funding and because image and genetic acquisitions were also not identical across the initiatives- Further details of the ADNI project may be found at <http://www.adni-info.org> 3) the Brain Imaging Genetics study (BIG), 4) the Searching for Endophenotypes of Bipolar Disorders Study (BP), 5) the Queensland Twin Imaging Study (QTIM), 6) IMAGEN and 7) the Rotterdam Scan Study (RSS). Further details and cohort information may be found in **Table 1**.

As the goal of the overall analysis is to provide protocols that do not require centralized image analysis and vast amounts of inter-site data transfer, for this pilot protocol development project we asked that groups with data transfer clauses send only anonymized brain scans and parcellations of a small subset of the cohort. A representative group of 25-40 images was selected from each site to represent the healthy distribution of age and sex in the full cohort for creation of a

cohort-specific minimal deformation template (MDT).

The sharing of protected health information such as subject age is restricted at many institutions. To analyze this while respecting such considerations required sites to pre-select representative subsets of individuals to include in the template. Registrations were then performed on additional images, and final Jacobian maps transferred to eligible researchers for statistical analysis. As this is a pilot study, we limited transfer of images to approximately 60 scans, and used 50% to create a stable MDT. For a major portion of this work, we continued to use roughly equal number of images per site (ADNI1: 80, ADNI2: 52, BIG: 62, BP: 58, QTIM: 60, IMAGEN: 60; RSS: 64) to ensure effect sizes in the overall association were not driven by any one large site.

Table 1. Demographics and image acquisition information. *For some datasets, a larger sample was also processed for follow-up.

	ADNI-1	ADNI-2	BIG	BP	QTIM	IMAGEN	RSS
Scanner	GE, Siemens Philips,	GE	Philips	Siemens Allegra	Siemens Brunker	GE	GE
Field Strength	1.5T	3T	1.5, 3 T 50/50%	3T	4T	3T	1.5T
Location	US multi-site	US multi-site	Nijmegen, NL	Chapel Hill, US	Brisbane, AUS	EU, multisite	Rotterdam, NL
Voxel-size	1.25x1.25x1.2 mm ³	1.25x1.25x1.2 mm ³	1.0x1.0x1.0 mm ³	1.0x1.0x1.0 mm ³	0.9x0.9x0.9 mm ³	1.0x1.0x1.0 mm ³	1.0x1.0x1.6 mm ³
N (subset; h: healthy)	80	597 SNP, 197 for AD analysis (52h)	62	134 (58h)	870 (60h)	60	64
% female	49%	45%	50%	65%	61%	50%	50%
Age	75+/- 6.6 (60-89)	72.8+/-6.6 (48-90)	21.5+/- 1.7 (18-25)	31.9+/- 16.3 (19-72)	22.9+/- 2.8 (18-30)	14	67.49 +/- 11.40

2.2. Data Preprocessing

All images were processed according to previously validated protocols available online at <http://enigma.ini.usc.edu/ongoing/protocols/>. Subcortical and cortical segmentations were conducted at each site for each cohort. While the segmentation method is not strictly standardized

(many programs may be common for subcortical parcellations), all groups here used FreeSurfer <http://surfer.nmr.mgh.harvard.edu/> versions 5.0 or 5.3. Quality control protocols were implemented to flag any potential outliers, including visual inspection of slices in each plane, creating population distribution histograms of regional

measures (subcortical volume, cortical thickness and surface area) and flagging outliers for further inspection.

The brain was extracted from all images and visually inspected; adjustment of parameters and use of alternate software was allowed until skull stripping was successfully performed. An affine registration was applied to all skull-stripped T1-weighted images (and subsequently their corresponding labels) to the MNI space.

2.3. Site-specific and Multi-site template creation

The MDTs were constructed using the Advanced Normalization Tools (ANTs; <http://stnava.github.io/ANTs/> [6]) software package and accompanying scripts. (at commit: 88276f8). Approximately 30 scans per cohort were used in the template. Images were linearly aligned before SyN [7] non-linear registration. To create the multi-channel template, a weight was assigned to each channel, corresponding to the contribution of that channel to the total warp. We set the T1-weighted channel itself to 1, the cortical ribbon to 0.5 and the subcortical segmentations to 0.2. While countless combinations of weights may be applied, we assigned most of the weight to the T1-weighted scans while preserving to some degree thickness and volumetric properties that are being currently investigated in ENIGMA studies. Once the MDTs reached convergence, the MDTs were visually inspected to ensure adequate representation of the cohort; i.e., no extraneous non-brain tissue, defined cortical and subcortical regions.

Four of the maps (ADNI1, RSS, QTIM and BIG), represent two older adult and two younger adult cohorts, respectively. The age range in total covered early (18-30) to late adulthood (60-85 years of age). These were used to create a representative MDT of the cohorts, again using 3-channels for registration. We chose not to include all sites in the final map, to model the common situation where new sites join an ongoing study after the template has been created.

All cohort MDTs were then registered to the final MDT. These warps were maintained for later pooling of statistical maps to one space. Two alternate methods of template construction and registration were also evaluated:

1) *Single-channel template and registration.* To maintain parameters as similar as possible to the multi-channel method described, for each cohort the single channel output from the multi-channel MDT was used as the template and only T1-weighted images from the cohorts were mapped to it. Similarly, the overall group MDT used was the first channel from the multichannel output. These consistencies allow us to see differences in results that are only due to the added channels from single subject registrations to the cohort specific MDT, and not the creation of the MDTs.

2) *Registration to MNI.* To explore the effect of templates in meta-analysis, we registered all subjects, regardless of cohort directly to the MNI atlas. This has the advantage of staying consistent regardless of added subjects or added cohorts; however, use of a single template not drawn from the population may include many biases in the mapping and not allow subtle within site effects to be discovered.

2.4. Tensor Based Morphometry Meta Analysis

Once each subject-specific multi-channel template was created, each subject's individual 3 channels were registered to their corresponding template with a similar weighting scheme per channel as used in the template. The logarithm maps of the geometric Jacobian maps were carried forward for TBM analysis. Individual maps were smoothed with a 3D smoothing kernel with sigma = 2 mm.

Maps of Beta statistics and standard errors of association in the cohort-specific space were warped to the common space by applying the MDT-to-MDT warp.

An inverse-weighted meta-analysis was then conducted at a voxelwise level.

We first note that the inverse-variance based aggregate p-value for the regression i at trait (voxel) v , is

$$p_{MA-SE}(i, v) = 2\Phi(|-Z_{MA-SE}(i, v)|), \text{ where } \Phi \text{ represents the normal transformation and } Z_{MA-SE}(i, v) = \frac{\beta(i, v)}{SE(i, v)} = \frac{\sum_j \beta_j(i, v) \times se_j^{-1}(i, v)}{\sqrt{1/\sum_j se_j^{-2}(i, v)}}$$

where, site j 's effect-size and standard error for the test of interest (sex effects, SNP, disease, etc.), i and trait v is $\beta_j(i, v)$ and $se_j(i, v)$.

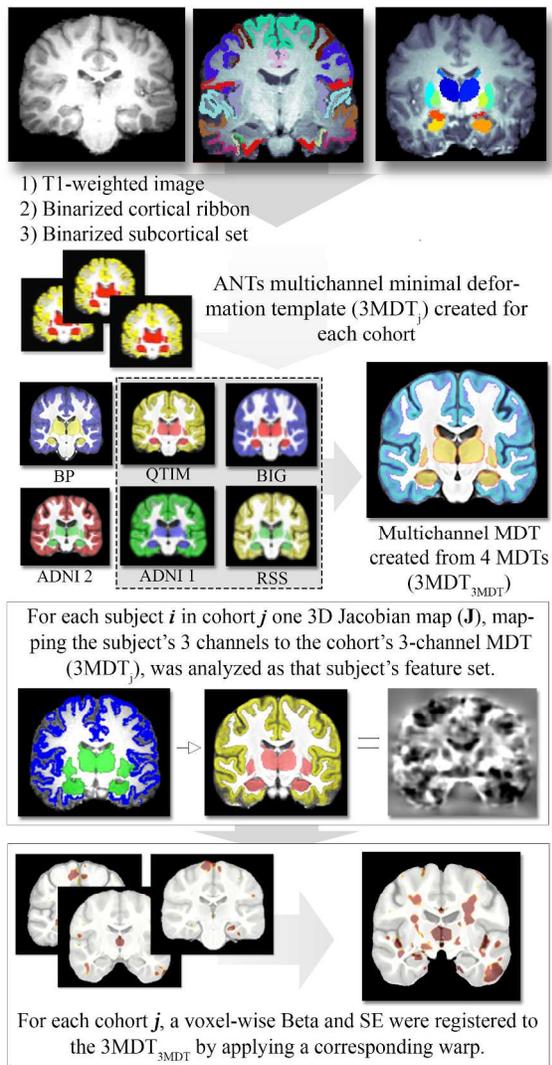


Figure 1 Flow diagram of template creation and registration.

2.5. Genetic Simulation and Effect Localization

As motivation for this work is distributed big data analytics on private full genome and image data, we generated 100,000 data-points per subject of each dataset to represent an additive genetic effect (0,1, or 2 at each “genetic locus”) using a 2D multinomial distribution with probabilities set to MAF and 1-MAF. The MAF was uniformly distributed (as approximated by public ADNI-2 data) and maintained greater than 0.1 to avoid rare variants. Of the generated SNPs, a handful were designed to meet certain criteria: **1)** a SNP with $MAF = 0.1$ was approximated to be marginally ($z=1.96$) associated with average bilateral thalamic volume in each cohort (after removing the

intracranial volume (ICV) effect). **2)** A SNP with $MAF= 0.3$ was generated to have $z=\min(N/10,5)$ when regressed with bilateral hippocampal volume (ICV effect removed) such that the significance was related to cohort size, yet was not excessive ($|z| < 5$). **3)** a SNP ($MAF=0.3$) was set to similarly associate with ICV, a feature intended to be removed from the voxelwise associations as TBM was performed on images that had been linearly aligned to include scaling after skull stripping. To approximately enforce the association, a correlation coefficient was determined from the set Z-statistic (1.96 for **1** above). Using the fact that vectors with mean =0 have $\text{corr}=\cos(\theta)$, where θ is the angle between them, we center and orthogonalize the response variable (HV) with a QR-decomposition and scale back; as this method does not lead to the integer values 0,1,2 needed, the values were rounded and correlation values were recomputed, the process iterated until the final correlation was ± 0.1 of the desired value.

2.6. Comparisons with segmented volume at the site level

Protocols for the ENIGMA consortium were designed for multi-site meta-analysis, but they are also intended to be used for site-level analyses that do not use other consortium data. As ADNI2 was not used in creating the overall MDT, the cohort was used as a test case for assessing any differences between (1) testing a statistical effect of a covariate on subcortical volumetric extractions, and (2) testing the same covariate effect voxelwise, according to the various template and registration schemes. This was used to determine if indeed the voxelwise maps pick up statistical associations in the expected regions.

For ADNI2, we examined the voxelwise effect as well as the effect of sex differences across all 14 subcortical measures (both left and right thalamus, caudate, putamen, pallidum, hippocampus, and amygdala), controlling for sex, age, age^2 , age x sex interaction, and age^2 x sex interaction, as well as intra-cranial volume.

2.7. Pooling information on sex and disease-related differences

Within each cohort, sex differences in deformation maps were evaluated after adjusting for covariates including age, age^2 , age x sex interaction, and age^2 x sex interaction. The BIG cohort was also

adjusted for scanner field strength. Next, we compared the effect of Alzheimer's disease diagnosis on brain morphometry in both ADNI1 and ADNI2 subsamples used here, using methods described above, and pooled statistical results. While both cohorts are part of the same overall study, the imaging parameters are slightly different (1.5T vs 3.0T).

2.8. True genetic effects mapped voxelwise

In a final analysis, we extend the genetic discoveries from the ENIGMA consortium. We analyzed structural brain MRI data from the Queensland Twin Imaging Study (QTIM; N=870; mean age 23.1 +/- 2.9 years; 534 women; 501 families) and the Alzheimer's Disease Neuroimaging Initiative (ADNI2; N=597; mean age 73.0 +/- 7.2 years; 270 women). The recent publication from the ENIGMA Consortium performed meta-analysis of genome-wide association scans of the volumes of 7 subcortical structures [4] using the power of around 30,000 individuals from over 50 sites. The most significant association was found to be a SNP associated with the volume of the putamen (rs945270-C/G; $p=1.08 \times 10^{-33}$; N=28,275). In both QTIM and ADNI2 datasets, the number of minor alleles carried at that SNP was regressed against the 3-channel TBM maps. Age and sex (and their linear and nonlinear combinations) were used as covariates. In ADNI2, we also covaried for Alzheimer's disease diagnosis and cognitive impairment (for those with AD or MCI). Kinship was included in the statistical model for QTIM to account for relatedness between subjects and the fact that the monozygotic twins share their genomes. The Beta and standard error maps of the SNP-regressions were then warped to a common template space. An inverse variance-weighted meta-analysis was then performed at each voxel, to meta-analyze the aggregate effect size for the association of the genetic variant on regional brain volumes. While other methods are also feasible, (see cautionary note in the discussion), here the false discovery rate method was used to account for the multiple comparisons, across the entire image.

3. RESULTS AND DISCUSSION

3.1. Genetic Simulation

Figure 2a shows the effect of a single variant with set marginal effects on thalamic volume (SNP 1 from section 2.5 above) mapped across multiple possible voxelwise methods. The multichannel approach where the cortical and subcortical volume segmentations were used as added registration channels to help drive registration MDT creation and intersubject registration showed visibly greater specificity with thalamic variability. **Figure 2b** shows the effect of the multichannel registration approach when meta-analyzing a fixed SNP effect (SNP 4 above) on a voxelwise level. Voxel level analysis maintained regional specificity with FDR-significant voxels bilaterally in the hippocampus.

3.2. Single-site sex differences, mapped voxelwise compared to volume.

No site on its own (all with N=50-80, almost all healthy controls except for 13 ADNI1 patients) showed significant sex differences when controlling the FDR at $q=0.05$. However, when voxelwise maps were pooled across samples, all forms of pooling showed some regions with proportionately larger volumes in women; **Figure 3**. This is consistent with prior work showing sex differences in regional volumes after proportional scaling of overall brain size [8].

Using the MNI template for group registration led to the most dispersed regions of significance with smallest cluster sizes.

After multiple comparisons correction with the false discovery rate procedure at $q=0.05$, in the larger N=197 ADNI2 sample, sex differences in regional volume survived correction for the right thalamus, left and right caudates, right putamen, right amygdala and bilateral pallidum; **Figure 4**.

When mapping effects voxelwise (**Figure 5**, ADNI2), we detected slightly different patterns of association, depending on the registration scheme used. The right amygdalar region showed significance in all maps, with diffuse effects in other brain regions. All the same covariates were used for this comparison, but it was negligible whether or not ICV was used as a covariate.

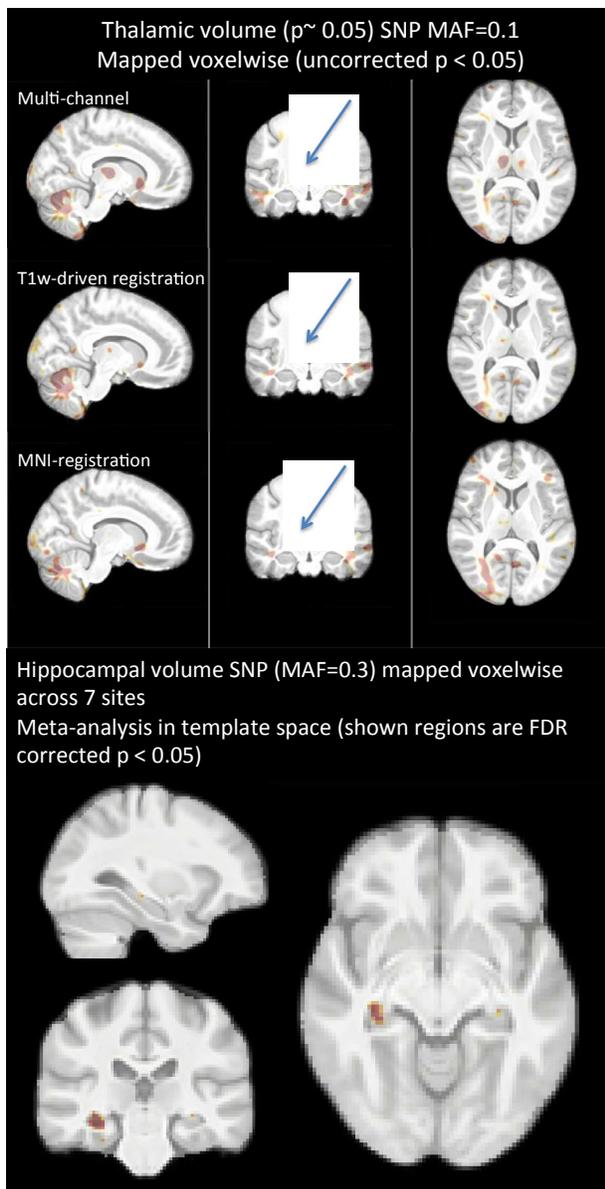


Figure 2. a) A SNP with MAF = 0.1 was simulated to be marginally ($z=1.96$) associated with average bilateral thalamic volume in a single cohort (after removing ICV). The effect of maintaining specificity to the thalami was compared between multiple templates. No method produced voxelwise significant maps, however, evaluating the uncorrected association results of the methods shows greater thalamic effects in the multi-channel method. **b)** A SNP with MAF=0.3 was generated for each of 7 cohorts, to have $z=\min(N/10,5)$ such that the significance was related to cohort size, yet was not excessive ($|z| < 5$). Beta and SE maps for all cohorts were mapped to template space, and voxelwise meta-analysis revealed associations localized to both hippocampi. As expected, no localized regions were detected with ICV simulated genetics.

3.3. Disease effects across sites. After multiple comparisons correction with FDR, both ADNI1 (13 AD vs 67 Controls) and ADNI2 (47 AD vs 150 Controls) cohorts showed strong volume reductions in the temporal lobes as well as ventricular expansion, as expected from the known anatomy of Alzheimer's disease. Meta-analyzing these maps through methods described above led to greatest effect sizes for differences detected using the multi-channel MDT, including well-defined cortical patterns, followed by the single channel cohort-specific template, and lastly the MNI standard template.

3.4. True genetic effects mapped voxelwise. Neither ADNI2 nor QTIM on its own showed significant voxelwise associations with the ENIGMA2 top hit[4]. However, after mapping the results to the same space, a meta-analysis of the two cohorts showed remarkable significance -- specifically localized to the putamen only. The lowest p -value obtained across the full map was 4.9×10^{-8} and the FDR-corrected threshold was $p=1.3 \times 10^{-5}$.

As an extension of the current work presented here, recent follow-up evaluations have confirmed the robustness of the proposed approach by examining the test-retest reliability of the two-step meta-analysis approaches compared to the single step, single MNI template approach and found significant improvements in reliability across ten datasets individually and when combined [9].

In this work, our aim is to present an analytical framework for harmonizing tensor based morphometry across cohorts based on previously quality-controlled segmentation assessments. Our framework allows cohort-level specificity to be maintained, while also allowing knowledge driven registrations to help improve correspondence between sites for meta-analysis. The overall meta-analysis results in a voxelwise statistical map, with which multiple methods may be used for correction across voxels. Our results are highlighted using FDR, the most widely used correction scheme for neuroimaging to date, despite limitations induced by insufficient account of the spatial smoothness of the images[10]. We note that careful consideration must be taken into account when attempting cluster-base statistical

inferences on the voxels as meta-analytical maps derive their smoothness from the individual study statistical maps, and are likely not uniform. Previous work, including Eickhoff et al., 2012[11], have looked into comparing various multiple comparisons correction methods including FDR and RFs for voxelwise meta analysis. While non-parametric means of permutation (sign-flipping of cohort statistical maps) are needed to derive clusterwise smoothness estimates, it should be noted that cluster size alone may reflect a counter-intuitive notion of inaccurately registered maps, and should be approached with caution.

4. CONCLUSION

We introduced a method for multi-site voxel-based statistical analysis of brain morphometry. It does not require centralized pooling of all scans, and benefits from distributed data processing at multiple sites using harmonized analysis protocols. By applying multi-channel registration to cohort-specific templates, analyses can proceed largely independently while avoiding re-computation when new sites join. In examples given here, genetic, sex and disease effects on structure were detected in a meta-analytic setting, and cross-validated against effects of the same covariates on volumes. Regardless of the registration scheme used, the pattern of effects was similar in the voxelwise images. Collaborative neuroimaging consortia (eg, ENIGMA, CHARGE), now analyze 30,000+ scans, so harmonized distributed computation is more efficient and scalable than mass transfer of data to a common site.

In future, greater effect sizes may be obtainable by using quality-controlled labels of cortical and subcortical extractions. Even so, the current method illustrates the potential of pooling data from multiple sites into a common space for large-scale meta-analysis of genetic and clinical data.

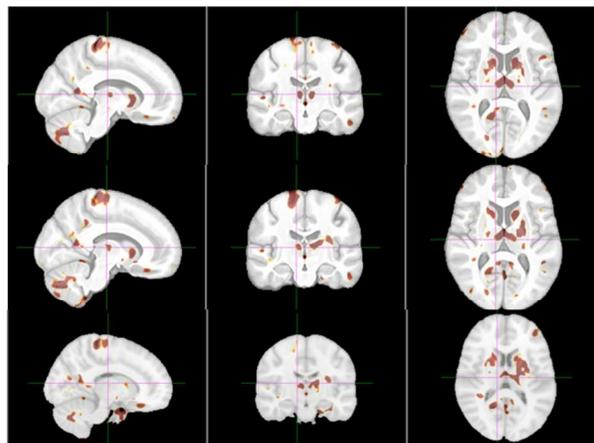


Figure 3. Multi-site meta-analysis of sex differences in regional brain volumes. Subcortical volumes do not scale linearly with the overall size of the brain[8, 12], leading to a pattern of sex differences in scaled data.

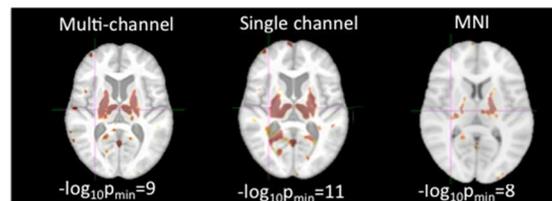


Figure 4. Single-site (ADNI2) sex associations are shown for (*top*: multichannel, *middle row*: single channel, *bottom row*: standard MNI template approach).

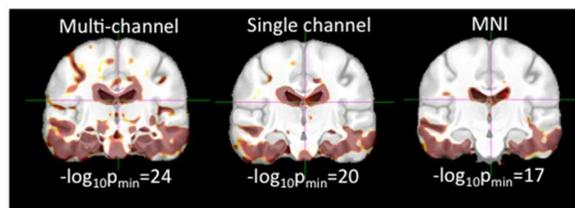


Figure 5. Meta-analysis of the effect of Alzheimer's disease on the brain as compared to healthy age- and sex-matched controls in ADNI1 and ADNI2. All results are overlaid on the MNI template for comparison.

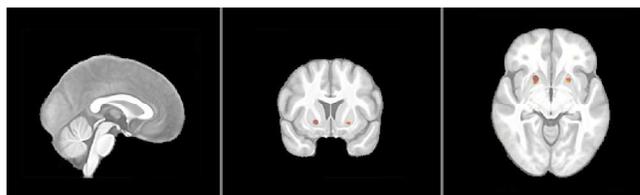


Figure 6. A meta-analysis of the ADNI2 and QTIM, two cohorts with an average age difference of approximately 50 years, was conducted. The analysis showed remarkable significance -- specifically localized to the putamen only with the voxelwise meta analysis. The lowest p -value obtained across the full map was 4.9×10^{-8} and the FDR-corrected threshold was $p=1.3 \times 10^{-5}$.

5. REFERENCES

- [1] J. L. Stein, S. E. Medland, A. A. Vasquez, D. P. Hibar, R. E. Senstad, A. M. Winkler, *et al.*, "Identification of common variants associated with human hippocampal and intracranial volumes," *Nat Genet*, vol. 44, pp. 552-61, May 2012.
- [2] G. Salimi-Khorshidi, S. M. Smith, J. R. Keltner, T. D. Wager, and T. E. Nichols, "Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies," *Neuroimage*, vol. 45, pp. 810-23, Apr 15 2009.
- [3] M. A. Ikram, M. Fornage, A. V. Smith, S. Seshadri, R. Schmidt, S. Debette, *et al.*, "Common variants at 6q22 and 17q21 are associated with intracranial volume," *Nat Genet*, vol. 44, pp. 539-44, 2012.
- [4] D. P. Hibar, J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivieres, N. Jahanshad, *et al.*, "Common genetic variants influence human subcortical brain structures," *Nature*, vol. 520, pp. 224-9, Apr 9 2015.
- [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. 648-61, Feb 15 2013.
- [6] 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," *Med Image Anal*, vol. 12, pp. 26-41, Feb 2008.
- [7] 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. 2033-44, Feb 1 2011.
- [8] C. C. Brun, N. Lepore, E. Luders, Y. Y. Chou, S. K. Madsen, A. W. Toga, *et al.*, "Sex differences in brain structure in auditory and cingulate regions," *Neuroreport*, vol. 20, pp. 930-5, Jul 1 2009.
- [9] J. Faskowitz, G. I. de Zubicaray, K. L. McMahon, M. J. Wright, P. M. Thompson, and N. Jahanshad, "Comparison of template registration methods for multi-site meta-analysis of brain morphometry," pp. 978822-978822.
- [10] J. R. Chumbley and K. J. Friston, "False discovery rate revisited: FDR and topological inference using Gaussian random fields," *Neuroimage*, vol. 44, pp. 62-70, Jan 1 2009.
- [11] S. B. Eickhoff, D. Bzdok, A. R. Laird, F. Kurth, and P. T. Fox, "Activation likelihood estimation meta-analysis revisited," *Neuroimage*, vol. 59, pp. 2349-61, Feb 1 2012.
- [12] A. N. Ruigrok, G. Salimi-Khorshidi, M. C. Lai, S. Baron-Cohen, M. V. Lombardo, R. J. Tait, *et al.*, "A meta-analysis of sex differences in human brain structure," *Neurosci Biobehav Rev*, vol. 39, pp. 34-50, Feb 2014.

6. ACKNOWLEDGMENTS

Funding for the ENIGMA Center for Worldwide Medicine Imaging and Genomics is provided as part of the 2014 NIH Big Data to Knowledge (BD2K) Initiative under U54EB020403 (PI: Thompson) to support big data analytics, management and distribution of programs.

* **ADNI**: 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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; 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.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; 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 (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.