◆ Human Brain Mapping 35:4827–4840 (2014) ◆

The Power of Neuroimaging Biomarkers for Screening Frontotemporal Dementia

Corey T. McMillan,[1]* Brian B. Avants,[2] Philip Cook,[2] Lyle Ungar,[3] John Q. Trojanowski,[4] and Murray Grossman[1]

[1]Department of Neurology, Penn Frontotemporal Degeneration Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania
[2]Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania
[3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, Pennsylvania
[4]Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania

Abstract: Frontotemporal dementia (FTD) is a clinically and pathologically heterogeneous neurodegenerative disease that can result from either frontotemporal lobar degeneration (FTLD) or Alzheimer's disease (AD) pathology. It is critical to establish statistically powerful biomarkers that can achieve substantial cost-savings and increase the feasibility of clinical trials. We assessed three broad categories of neuroimaging methods to screen underlying FTLD and AD pathology in a clinical FTD series: global measures (e.g., ventricular volume), anatomical volumes of interest (VOIs) (e.g., hippocampus) using a standard atlas, and data-driven VOIs using Eigenanatomy. We evaluated clinical FTD patients ($N = 93$) with cerebrospinal fluid, gray matter (GM) magnetic resonance imaging (MRI), and diffusion tensor imaging (DTI) to assess whether they had underlying FTLD or AD pathology. Linear regression was performed to identify the optimal VOIs for each method in a training dataset and then we evaluated classification sensitivity and specificity in an independent test cohort. Power was evaluated by calculating minimum sample sizes required in the test classification analyses for each model. The data-driven VOI analysis using a multimodal combination of GM MRI and DTI achieved the greatest classification accuracy (89% sensitive and 89% specific) and required a lower minimum sample size ($N = 26$) relative to anatomical VOI and global measures. We conclude that a data-driven VOI approach using Eigenanatomy provides more accurate classification, benefits from increased statistical power in unseen datasets, and therefore provides a robust method for screening underlying pathology in FTD patients for entry into clinical trials. *Hum Brain Mapp 35:4827–4840, 2014.*

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: National Institutes of Health, Contract grant number: AG043503, AG017586, NS044266, AG010124, AG015116; Contract grant sponsor: Wyncote Foundation.

*Correspondence to: Corey T. McMillan, Department of Neurology, University of Pennsylvania Perelman School of Medicine, 3400 Spruce Street, 3 West Gates, Philadelphia, PA 19104. E-mail: mcmillac@mail.med.upenn.edu

Received for publication 31 January 2014; Accepted 17 March 2014.

DOI 10.1002/hbm.22515

Published online 31 March 2014 in Wiley Online Library (wileyonlinelibrary.com).

Key words: frontotemporal degeneration; Alzheimer's disease; MRI; DTI; biomarkers; statistical power; classification

INTRODUCTION

Disease-modifying agents are emerging for clinical trials of frontotemporal lobar degeneration (FTLD) [Boxer et al., 2013a,b]. However, at autopsy approximately 20% of behavioral variant frontotemporal degeneration (bvFTD) cases actually have pathology consistent with Alzheimer's disease (AD) [Harris et al., 2013] and an even higher proportion of corticobasal syndrome (CBS) patients have AD pathology at autopsy [Hassan et al., 2011; Lee et al., 2011]. Previous power calculations suggest that neuroimaging measures outperform cognitive measures [Ard and Edland, 2011] and therefore a statistically robust and accurate neuroimaging screening tool would maximize statistical power and therefore yield substantial cost-savings and increase the feasibility of identifying optimal clinical trial entry criteria. Moreover, such a systematic analysis would address concerns that have been raised recently about power in neuroscience studies [Button et al., 2013]. Throughout this manuscript we adopt the standardized use of the term FTLD to refer to an autopsy-confirmed consensus diagnosis [Mackenzie et al., 2010] and we use FTD to refer to a spectrum of clinical syndromes commonly, but not necessarily, associated with FTLD pathology, including bvFTD [Rascovsky et al., 2011], primary progressive aphasia (PPA) [Gorno-Tempini et al., 2011], and CBS [Armstrong et al., 2013].

Neuroimaging methods are noninvasive and widely available and therefore may provide an ideal quantitative biomarker for screening candidate AD and FTLD patients for entry into clinical therapeutic trials. There is mounting evidence that a variety of neuroimaging modalities can reliably provide sensitive and specific classification of individual subjects with AD and FTLD [Davatzikos et al., 2008; Klöppel et al., 2008; McMillan et al., 2012; Zhang et al., 2009, 2011; Zhou et al., 2010]. The majority of these studies have suggested that a distribution of neuroanatomical changes that include frontal, temporal, and parietal regions is necessary to achieve accurate classification. However, each of these studies has used different methods of quantifying regional atrophy.

In this article, we focus on three broad methods of quantifying regional atrophy that have previously been suggested to perform magnetic resonance imaging (MRI)-based classification. These include global MRI measures such as ventricular and gray matter (GM) volume [Chou et al., 2010; Knopman et al., 2009] that can be measured relatively reliably and have been associated with clinical decline [Chou et al., 2010]. An alternative approach is the use of anatomically defined volumes of interest (VOI) such as hippocampal volume [Morra et al., 2009; Muñoz-Ruiz et al., 2012]. These anatomical approaches may benefit from increased regional specificity in comparison to global measures, but are potentially limited owing to user-defined boundaries or other sources of individual variability such as genetic modifiers [Goñi et al., 2013]. A third approach involves data-driven VOIs. These benefit from both regional specificity and user independence [Avants et al., 2012; Pelaez-Coca et al., 2011]. One example of this approach is Eigenanatomy, a dimensionality reduction approach that identifies VOIs accounting for the greatest statistical variance in the brain independent of a priori anatomic- or user-defined regions [Avants et al., 2012]. Instead, Eigenanatomy identifies a rank-ordered series of eigenvectors, each of which captures a cluster of voxels that explains a segment of the variance in the imaging dataset. By using a data-driven dimensionality reduction approach we can reduce imaging data that contain over a million voxels to a more computationally feasible set of predictors. This approach has been previously reported to be robust in MRI-based classification studies of AD [McMillan et al., 2013a] and FTLD [McMillan et al., 2013a,b].

In addition to different neuroimaging analysis approaches, there is increasing evidence that various neuroimaging modalities may improve discrimination between AD and FTLD. Beyond the MRI analyses of GM considered above, diffusion tensor imaging (DTI) of white matter (WM) is robust for discriminating between FTLD subtypes [McMillan et al., 2013b] and a multimodal combination of DTI and GM MRI achieves increased specificity for discriminating between FTLD and AD [McMillan et al., 2012].

We report a comparative study that assesses the classification accuracy and statistical power of global, anatomical, and data-driven VOI methods for screening FTLD and AD in the context of clinical FTD. We evaluate a representative sample of approximately 80% FTLD cases and 20% AD cases that is consistent with proportions of FTLD to AD in previously published young onset dementia autopsy series of clinical FTD. Each type of neuroimaging measure was evaluated using a multimodal combination of volumetric GM MRI and DTI of WM. Analyses were also performed using each of these modalities alone to allow comparison to the large number of previously published single modality neuroimaging studies. We use a full training and testing design to maximize generalizability and to minimize limitations associated with post hoc power estimates [Button et al., 2013]. After training a linear regression using each VOI approach and each neuroimaging modality we evaluate AD and FTLD classification accuracy and estimate minimum sample sizes for replication in an independent cohort. On the basis of our previous observations [McMillan et al., 2012], we hypothesized that a multimodal

TABLE I. Mean (SE) demographic profiles of clinical FTD patients with cerebrospinal fluid profiles consistent with Alzheimer's disease (AD) or frontotemporal lobar degeneration (FTLD)

Measure	AD	FTLD
N (female)	21 (12)	72 (31)
Age	67.81 (1.62)	62.61 (0.90)
Education	15.05 (0.66)	15.47 (0.41)
Disease duration	2.52 (0.31)	3.53 (0.32)
MMSE	25.19 (0.63)	26.29 (0.34)
Total-tau/Aβ	0.67 (0.08)	0.17 (0.01)
DTI sequence (12/30)	7/14	29/43

assessment using a data-driven approach would prove to be most robust at distinguishing between AD and FTLD. This study confirms that hypothesis.

METHODS

Participants

We investigated 93 patients clinically diagnosed with a FTD-spectrum neurodegenerative disease from the Penn Frontotemporal Degeneration Center and Cognitive Neurology Clinic at the University of Pennsylvania. As therapeutic targets in a clinical trial would ideally be administered to patients with mild disease, we restricted our cohort to mild patients [Mini-Mental State Exam (MMSE) ≥ 20]. A board-certified neurologist who has extensive expertise in neurodegenerative diseases diagnosed all patients a FTD-spectrum disease using published criteria, including behavioral variant FTD [Rascovsky et al., 2011], PPA [Gorno-Tempini et al., 2011], CBS [Armstrong et al., 2013], and progressive supranuclear palsy (PSP) [Litvan et al., 1996] (see Supporting Information Table 1 for a summary of clinical syndromes). All patients participated in a high-resolution volumetric T1-weighted MRI scan and a diffusion-weighted imaging protocol. All patients also participated in a lumbar puncture, described below. Patient groups were comparable for all demographic features (all $P > 0.1$), including education, disease duration, and disease severity measured with the MMSE (see Table I for a summary of demographics). On average FTLD patients were younger than AD patients [$t(91) = 2.77$; $P = 0.007$] by ~5 years; therefore, we include age as a nuisance covariate in all statistical models. Written informed consent was obtained from all patients using a University of Pennsylvania Institutional Review Board approved protocol.

Cerebrospinal Fluid Analysis

Cerebrospinal fluid (CSF) analytes of total tau and beta-amyloid$_{1-42}$ were obtained using previously reported procedures and evaluated with either a sandwich ELISA 2 (INNOTEST, Innogenetics, Ghent, Belgium) or a LUMINEX xMAP platform (INNO-BIA AlzBio3, Innogenetics). A ratio of total tau to beta-amyloid (t-tau:Aβ) was generated across platforms using an autopsy-validated conversion factor that has been cross-validated across two independent series [Irwin et al., 2012]. Specifically, it has been demonstrated that a t-tau:Aβ ratio above threshold (>0.34) is 95.5% accurate across two autopsy series [Irwin et al., 2012]. In this study cohort, 11 patients had autopsy or a genetic mutation consistent with FTLD pathology and all of them were correctly classified with CSF as having FTLD pathology. Using this threshold we identified 72 patients with a CSF profile not consistent with AD, which we presume is FTLD, and 21 patients had a CSF profile consistent with AD. Our cohort that contains 22.5% AD cases provides a representative sample that is consistent with previous reports, suggesting that ~20–30% of clinical FTD cases have AD pathology [Harris et al., 2013; Lee et al., 2011].

Volumetric T1 MRI Acquisition and Preprocessing

All participants underwent a structural T1-weighted MPRAGE MRI acquired from a SIEMENS 3.0T Trio scanner with an eight-channel coil using the following parameters: repetition time (TR) = 1,620 ms; echo time (TE) = 3 ms; slice thickness = 1.0 mm; flip angle = 15°; matrix = 192 × 256; and in-plane resolution = 0.9 mm × 0.9 mm. MRI volumes were preprocessed using highly accurate [Klein et al., 2010] PipeDream (http://sourceforge.net/projects/neuropipedream/) and Advanced Normalization Tools (ANTs) [Avants et al., 2008], as previously reported [Avants et al., 2011]. Briefly, PipeDream deforms each individual dataset into a standard local template space in a canonical stereotactic coordinate system. A diffeomorphic deformation was used for registration that is symmetric to minimize bias toward the reference space for computing the mappings, and topology-preserving to capture the large deformation necessary to aggregate images in a common space. These algorithms allow template-based priors to guide GM. We then computed a registration-based measure of cortical thickness [Das et al., 2009] and smoothed the cortical thickness images using a 1.5-mm sigma kernel.

DTI Acquisition and Preprocessing

Diffusion-weighted images were acquired with either a 30- or 12-directional acquisition sequence. The 30-directional sequence included a single-shot, spin-echo, diffusion-weighted echo planar imaging sequence (FOV = 245 mm; matrix size = 128 × 128; number of slices = 57; voxel size = 2.2 mm isotropic; TR = 6,700 ms; TE = 85 ms; and fat saturation). In total, 31 volumes were

acquired per subject, one without diffusion weighting ($b = 0$ s/mm^2) and 30 with diffusion weighting ($b = 1{,}000$ s/mm^2) along 30 noncollinear directions. The 12-directional sequence included a single-shot, spin-echo, diffusion-weighted echo planar imaging sequence (matrix size = 128 × 128; number of slices = 40; voxel size = 3 mm; TR = 6,500 ms; and TE = 99 ms). In total, 12 noncollinear, noncoplanar, isotropic diffusion-encoding directions were acquired. An equal proportion of DTI data from each sequence was available per subject group ($X^2 = 1.42$, $P > 0.1$; see Table I for a summary) and across training (50%) and test (50%) cohorts ($X^2 < 0.1$; $P > 0.1$), thus reducing potential DTI-sequence bias within any one group. We additionally included a nuisance covariate for DTI sequence in all DTI analyses, as previously reported [McMillan et al., 2013b].

Diffusion-weighted images were preprocessed using ANTs [Avants et al., 2008] and Camino [Cook et al., 2006] within the associated PipeDream (http://sourceforge.net/projects/neuropipedream/) analysis framework. Motion and distortion artifacts were removed by affine co-registration of each diffusion-weighted image to the unweighted ($b = 0$) image. Diffusion tensors were computed using a linear least squares algorithm implemented in Camino [Salvador et al., 2005]. Each participant's T1 image was warped to the template via the symmetric diffeomorphic procedure in ANTS (as above). Distortion between participants' T1 and DT images was corrected by registering the FA image to the T1 image. The DT image was then warped to template space by applying both the intrasubject (FA to participant T1) and intersubject (participant T1 to template) warps. Tensors were reoriented using the preservation of principal directions algorithm [Alexander et al., 2001].

Calculation of Neuroimaging Classifiers

Once MRI and DTI images were preprocessed and normalized to standard stereotactic space we computed VOIs for classification using three methods. To calculate global measures we generated a probabilistically defined GM mask and we generated a lateral ventricle mask using an edge-based snake algorithm implemented in ITK-SNAP (http://www.itksnap.org). Each of these masks was generated in our local template. We then used the inverse warp from our ANTs registration routine to warp each of these masks into subject space where we computed the sum of all voxels to generate a single value for total GM volume and for total ventricular volume.

To calculate anatomically defined VOIs we used labels from previously published and widely used anatomical atlases of GM [Tzourio-Mazoyer et al., 2002] and WM [Oishi et al., 2008]. Each of these atlases was available for download in MNI space, which we used for all analyses. Atlases were masked using a probabilistically defined GM and WM mask to omit voxels that did not contain a cortical thickness or FA value. Within each atlas we generated a binary mask for each of the VOIs including 90 GM regions and 48 WM regions. We then computed the mean cortical thickness or mean FA within each region yielding a single value per anatomical VOI for each patient.

To calculate data-driven VOIs we used Eigenanatomy [Avants et al., 2012; McMillan et al., 2013b], a dimensionality reduction tool based on sparse singular value decomposition (SVD) and implemented in ANTs (http://stnava.github.io/ANTs/). To identify VOIs using Eigenanatomy, all GM thickness volumes or FA volumes are first transformed into a matrix. Then an anatomically constrained $L1$-penalized SVD is used to identify a set of 20 eigenvectors that account for 95% of the total variance in the matrix. The ANTs implementation of Eigenanatomy uses a sparseness penalty on the eigenvectors such that (1) the entries of the eigenvector are both sparse (i.e., have many zero entries) and non-negative and (2) the non-zero voxels are clustered and exceed a cluster extent threshold. The extent thresholds selected for this study were chosen to approximately match the average size of anatomical VOIs used for the comparative study: 1,000 adjacent voxels for GM and 500 adjacent voxels for FA. The sparseness and non-negativity allows the eigenvectors to be interpreted as weighted averages of the original data, resembling a distributed version of a traditional region of interest.

Classifier Training

Once GM MRI and WM DTI measures were computed for the data-driven, anatomical, and global approaches, as described above, we performed linear regression and cross-validation (see Fig. 1 for a schematic of the procedure) to identify the optimal neuroimaging classifiers. To perform cross-validation, our full patient cohort ($N = 93$) was randomly divided into a training ($N = 46$) and test ($N = 47$) cohort. For all linear regressions we included nuisance covariates that have previously been reported to contribute to the neuroanatomic distribution of disease. These included age at MRI, MMSE, disease duration, gender, and APOE status.

Within the training dataset we performed two stages of feature selection for each of the anatomical and data-driven approaches; we did not, however, perform feature selection for the global measures because these models, by definition, only include a single neuroimaging classifier. First, we performed an initial stage of feature selection using the Bayesian information criterion (BIC) to select VOIs that are potentially useful as classifiers. To achieve this we performed an exhaustive comparison of linear regression models using all possible combinations of features. These models were computed and ranked according to BIC using the Leaps package implemented in R (http://cran.r-project.org/web/packages/leaps/). We retained a total of 50 models with the lowest BIC and then hand-selected the features that appeared most frequently in

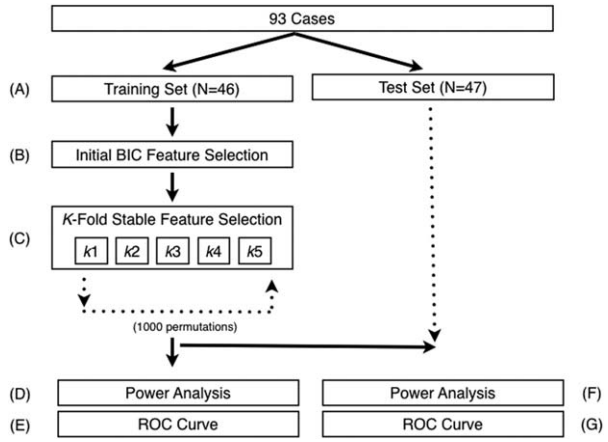

Figure 1.

Schematic overview of training and test prediction procedures. (**A**) Cohort is randomly divided into training and test datasets. (**B**) Initial feature selection is performed by determining which VOIs minimize Bayesian information criterion (BIC). (**C**) Fivefold cross-validation is performed within the training dataset by randomly dividing the cohort into five sets, calculating the features that achieve highest prediction accuracy, and permuting this process 1,000 times to identify the most stable VOIs for prediction. (**D**) Stable VOIs are entered into a power analysis in training cohort to confirm that there is a sufficient sample for test prediction. (**E**) ROC curve to calculate prediction accuracy in training cohort. (**F**) Power analysis in independent test cohort. (**G**) ROC curve to evaluate prediction accuracy in independent test cohort.

these models. Second, we refined our initial selection of features using fivefold cross-validation to select the most stable combination of features that most reliably achieved the highest prediction accuracy. This was accomplished by randomly dividing the dataset into five equally sized folds to identifying the optimal set of features. This fivefold procedure was then permuted 1,000 times in an effort to minimize bias that may be associated with unbalanced demographical features, unbalanced DTI parameters, or uneven assignment of pathological groups to each fold. This was done using the bestglm package in R (http://cran.rproject.org/web/packages/bestglm/). Using this approach, we generated a single, optimal linear regression model for our training dataset for each imaging measure (e.g., data-driven VOIs and anatomical VOIs) and modality (GM, FA, and multimodal combination of GM and FA). Each of these linear regression models contained the most stable VOIs identified during cross-validation.

Classification Accuracy

To evaluate classification accuracy of the global, anatomical, and data-driven approaches, we generated receiver operator characteristic (ROC) curves for each of the GM MRI, DTI, global, and multimodal analyses. In this study, we use sensitivity to refer to accuracy of diagnosing an individual as having AD when they do have AD. We report the area under the curve (AUC) and report sensitivity and specificity using the threshold that achieves the highest Youden-*J* index, a measure of the overall accuracy of a diagnostic test (sensitivity + specificity − 100). We report accuracy in the independent test cohort and include accuracy for our training dataset in Supporting Information.

Statistical Power Analysis

To quantify the statistical power associated with each neuroimaging analysis we calculated the minimum sample size required for classification of AD and FTLD in the independent test cohort. We first performed a regression analysis that only included the demographic nuisance covariates. We then compared the r^2 value of the demographic nuisance covariate model relative to each neuroimaging regression model that included the neuroimaging classifiers selected during training together with the demographic nuisance covariates. This resulted in a relative calculation of power $(1 - \beta)$ for values ranging from 0.1 to 1.0. These analyses were performed using the MBESS package implemented in R (http://cran.r-project.org/web/packages/mbess/).

TABLE II. Summary of test prediction results: Power (minimum sample size) and classification accuracy for MRI, DTI, global, and multimodal combination of measures

Modality	Method	Minimum *N*	AUC	*P*	Sensitivity	Specificity	Youden-*J*
Volumetric MRI	Data-driven	14	0.778	0.010	81	89	70
	Anatomical	87	0.802	0.005	54	100	54
DTI	Data-driven	14	0.808	0.005	46	100	54
	Anatomical	118	0.649	ns	78	56	34
Global	GM volume	82	0.820	0.003	65	100	65
	Ventricular volume	235	0.826	0.003	65	100	65
Multimodal	Data-driven	26	0.874	0.001	89	89	78
	Anatomical	66	0.742	0.026	70	78	48

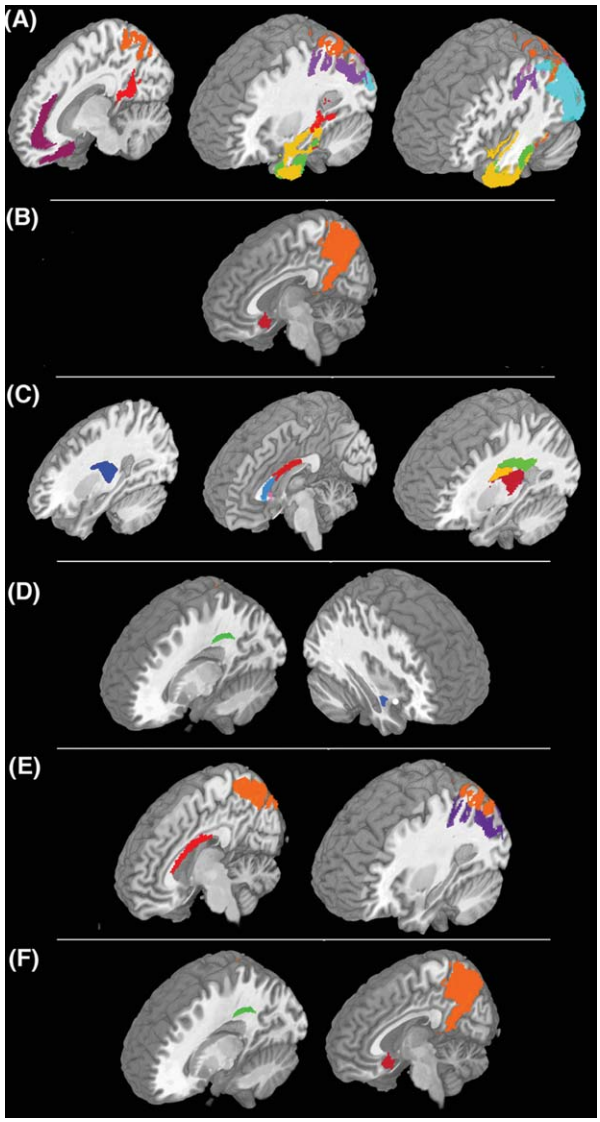

Figure 2.
Selected volumes of interest (VOIs) for volumetric MRI and DTI methods. (A) Data-driven MRI; (B) anatomical MRI; (C) data-driven DTI; (D) anatomical DTI; (E) data-driven multimodal (MRI + DTI); and (F) anatomical multimodal (MRI + DTI).

RESULTS

Classification Accuracy

All training analyses achieved statistically significant accuracies ($P < 0.05$) and are summarized in Supporting Information Table 2 and illustrated in Supporting Information Figure 1. In an independent evaluation of the test cohort ($N = 47$), we used VOIs generated in the training classification and their corresponding regression models to evaluate classification accuracy in an unseen group of patients. Classification accuracy is summarized in Table II and corresponding ROC curves are illustrated in Figure 3. Overall, these analyses revealed that the data-driven approach consistently outperformed anatomical and global measures in each comparison and that the date-driven multimodal analysis achieved the highest overall classification accuracy.

The data-driven multimodal approach included VOIs in left parietal cortex (purple), bilateral precuneus (orange), and the body of the corpus callosum (red) (Fig. 2E). As illustrated in Figure 3D, the data-driven multimodal approach thus achieved the highest overall classification accuracy with 89% sensitivity and 89% specificity (AUC = 0.874; $P < 0.001$). The anatomical multimodal approach, which selected the precuneus (orange), right ventral-medial prefrontal cortex (red), and the left superior longitudinal fasciculus (green) (Fig. 2F), achieved only modest classification accuracy (AUC = 0.742; $P = 0.026$).

Using MRI GM thickness alone, by comparison, Figure 3A shows that the data-driven approach achieved good sensitivity (81%) with high specificity (89%; AUC = 0.778; $P = 0.010$). The data-driven GM approach included seven VOIs as classifiers, illustrated in Figure 2A. These comprised a large bilateral VOI extending from angular gyrus to precuneus (orange), two adjacent VOIs in left anterior temporal cortex (green and yellow), left superior temporal extending into occipital cortex (blue), bilateral posterior cingulate (purple), left parahippocampal and fusiform gyri (red), and bilateral anterior cingulate extending into ventral-medial prefrontal cortex (magenta). In contrast, the anatomical MRI approach was poorly sensitive (54%), but was highly specific (100%; AUC = 0.802; $P < 0.005$). As illustrated in Figure 2B, the most stable VOIs for the anatomical GM MRI approach overlapped with some regions identified by the data-driven approach, including the right precuneus (orange) and right ventromedial prefrontal cortex (red).

Using DTI alone (Fig. 3B), the data-driven approach had poor sensitivity (46%) with high specificity (100%), and this was statistically robust (AUC = 0.808; $P = 0.005$). This analysis included three adjacent clusters in left superior longitudinal fasciculus (dark red, green, and yellow), right superior longitudinal fasciculus (blue), body of the corpus callosum (bright red), and genu of the corpus callosum (light blue) (Fig. 2C). However, the anatomical analysis of DTI did not achieve significance for classification accuracy (AUC = 0.649; ns). This anatomical analysis included left superior longitudinal fasciculus (green) and right uncinate fasciculus (blue) (Fig. 2D). The data-driven and anatomical approaches overlapped in the left superior longitudinal fasciculus. The global measure analyses achieved only modest sensitivity (65%) with high specificity (100%) for GM volume (AUC = 0.820; $P = 0.003$) and ventricular volume (AUC = 0.826; $P = 0.003$) and are summarized in Figure 3C.

In follow-up analyses we assessed the classification accuracy of each clinical syndrome included in our series and these results are summarized in Table III. Of particular interest is the classification accuracy of atypical syndromes that are more equally associated with AD or FTLD pathology, including CBS [Hu et al., 2009] and logopenic variant PPA

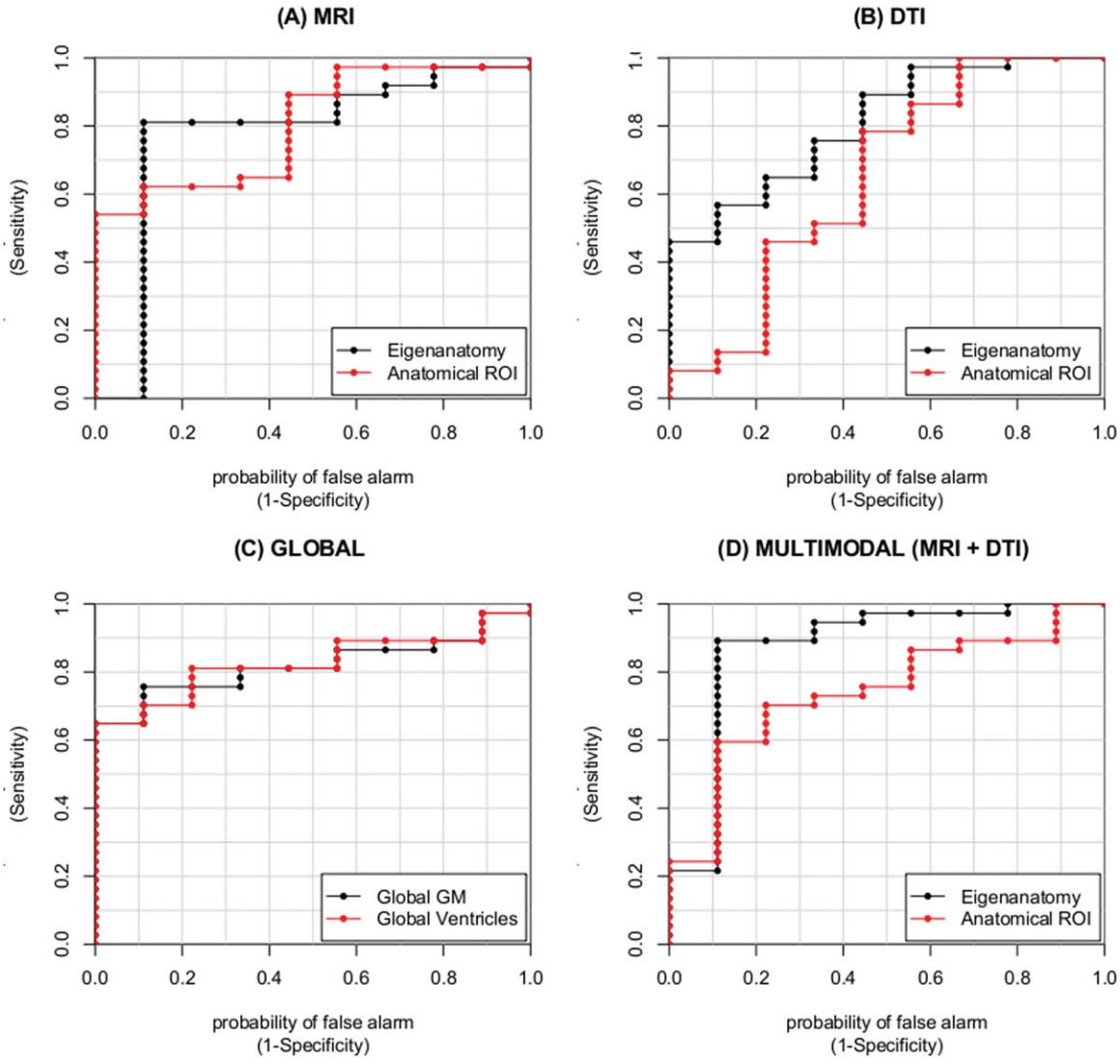

Figure 3.
Receiver operator characteristic (ROC) curves for volumetric, DTI, and multimodal combination of neuroimaging approaches in independent test dataset.

(lvPPA) [Mesulam et al., 2008]. Our results suggest that a multimodal approach achieves 100% accuracy of CBS and 78% accuracy of lvPPA. All other, more typical FTLD-associated syndromes achieve greater than 80% accuracy. The only exception includes PSP that only achieved 67% accuracy in which one of three cases was misclassified and this individual had AD-consistent CSF, suggesting that they may have comorbid underlying pathology [Toledo et al., 2012].

Statistical Power

To evaluate whether our independent test cohort ($N = 47$) was large enough to evaluate classification prediction accuracy, we first estimated the minimum sample size required for replication based on our training classification analyses. These results are summarized in Supporting Information Table 2 and illustrated in Supporting Information Figure 2. Overall, training power analyses suggested that our sample size for data-driven and anatomical approaches is indeed sufficiently large for VOI-based approaches. However, the global approaches required over 200 cases for both GM and ventricular volume analyses. As the performance of the global approach was much more modest than data-driven and anatomical approaches, we performed two post hoc quality control analyses reported in Supporting Information A.

The power analyses in the independent test cohort suggested that the data-driven approach requires approximately a third of the samples size to perform prediction

TABLE III. Classification accuracy for each data-driven neuroimaging biomarker approach across clinical syndromes

Clinical syndrome	*N*	Grey matter MRI	White matter DTI	Multimodal MRI + DTI
Corticobasal syndrome (CBS)	5	60	60	100
Logopenic variant PPA (lvPPA)	9	78	67	78
Nonfluent/agrammatic PPA (naPPA)	4	75	25	100
Progressive supranuclear palsy (PSP)	3	100	33	67
Semantic variant PPA (svPPA)	5	80	40	80
Behavioral variant FTD (bvFTD)	20	90	65	90

classification in comparison to the anatomical approaches. These are all summarized in Table II and illustrated in Figure 4. Specifically, the data-driven MRI GM thickness and DTI approaches have the smallest overall minimum sample size of 14 cases in comparison to a minimum sample of 87 cases for the GM anatomical approach and 118 cases for the DTI anatomical approach. The data-driven multimodal approach requires a slightly larger minimal sample size ($N = 26$) than unimodal methods, but was the most accurate approach (see above). Power analyses for the global measures suggested considerably larger sample sizes compared with the data-driven approach: 82 cases for GM volume and 235 cases for ventricular volume.

DISCUSSION

Neuroimaging has been suggested as a candidate biomarker to screen neurodegenerative patients for entry into clinical trials [McMillan et al., 2012, 2013a]. We evaluated the accuracy and statistical power of three published approaches for quantifying regional atrophy observed with neuroimaging using a multimodal combination of GM MRI and DTI. Because many prior publications use GM MRI or DTI alone, we also performed analyses using single modality datasets. Our findings suggested that a data-driven statistical approach using multimodal data provides the most accurate and powerful approach. In the sections below, we discuss the implications of data-driven neuroimaging methods for classification, statistical power, and broader research goals concerned with neurodegeneration and neuroimaging.

Classification Accuracy of Neuroimaging Biomarkers

In our comparative assessment of different neuroimaging approaches we observed that our data-driven VOI approach achieved the greatest accuracy using multimodal measures. This finding is consistent with a previous comparative study reporting that statistically defined measures generated with a principal components analysis outperformed anatomically defined measures for discriminating between AD and controls [Pelaez-Coca et al., 2011]. The observation that a multimodal dataset achieved the highest performance is also consistent with a previous report demonstrating that a multimodal combination of MRI, CSF, and FDG-PET was more powerful statistically than any of these modalities individually [Kohannim et al., 2010].

The observation that a multimodal neuroimaging approach achieves high classification accuracy converges with prior evidence in our laboratory that demonstrated a combination of GM MRI and WM DTI performs better than a single imaging modality [McMillan et al., 2012]. Both studies suggest that GM in parietal cortex and WM in the corpus callosum are the most accurate regions for classification, but this study contributes several additional advances. First, this study evaluated accuracy in an independent dataset using cross-validation, which is necessary to assess the generalizability of our method. Second, the DTI analyses in this study were identified using data-driven, and user-independent, WM VOIs rather than tract-specific analyses in our prior report, which required a priori assumptions about the loci of WM tracts and was limited to only 11 tracts that could be reliably parcellated. Third, this study additionally evaluated the statistical power of our classification methods, which we discuss in detail in a later section.

The selection of parietal regions and corpus callosum identified with our data-driven approach also overlapped substantially with VOIs identified by the anatomical approach. The major difference between these two approaches was the limited generalization of the regions identified by the anatomic approach from the training dataset to the independent cohort of test patients. In contrast, the data-driven approach proved to be nearly as robust in the independent test cohort as it was in the training cohort. One potential reason that anatomical VOI approaches may be suboptimal in this context is because of individual differences in anatomic structure and the anatomic distribution of disease. For example, FTLD patients may or may not have observable medial temporal lobe atrophy [Galton et al., 2001; Hornberger et al., 2012; Josephs et al., 2006] and atypical AD patients may have hippocampal sparing [Murray et al., 2011]. Findings such as these emphasize that strictly anatomical approaches may have limited value in comparative classification studies. Although the anatomical VOI approach used in this study used widely reported atlases of anatomically defined VOIs [Tzourio-Mazoyer et al., 2002], it is possible that alternative atlases may achieve better performance in future studies.

Global neuroimaging approaches have previously been suggested for use in clinical trials [Knopman et al., 2009]. However, although the global approaches reported in this

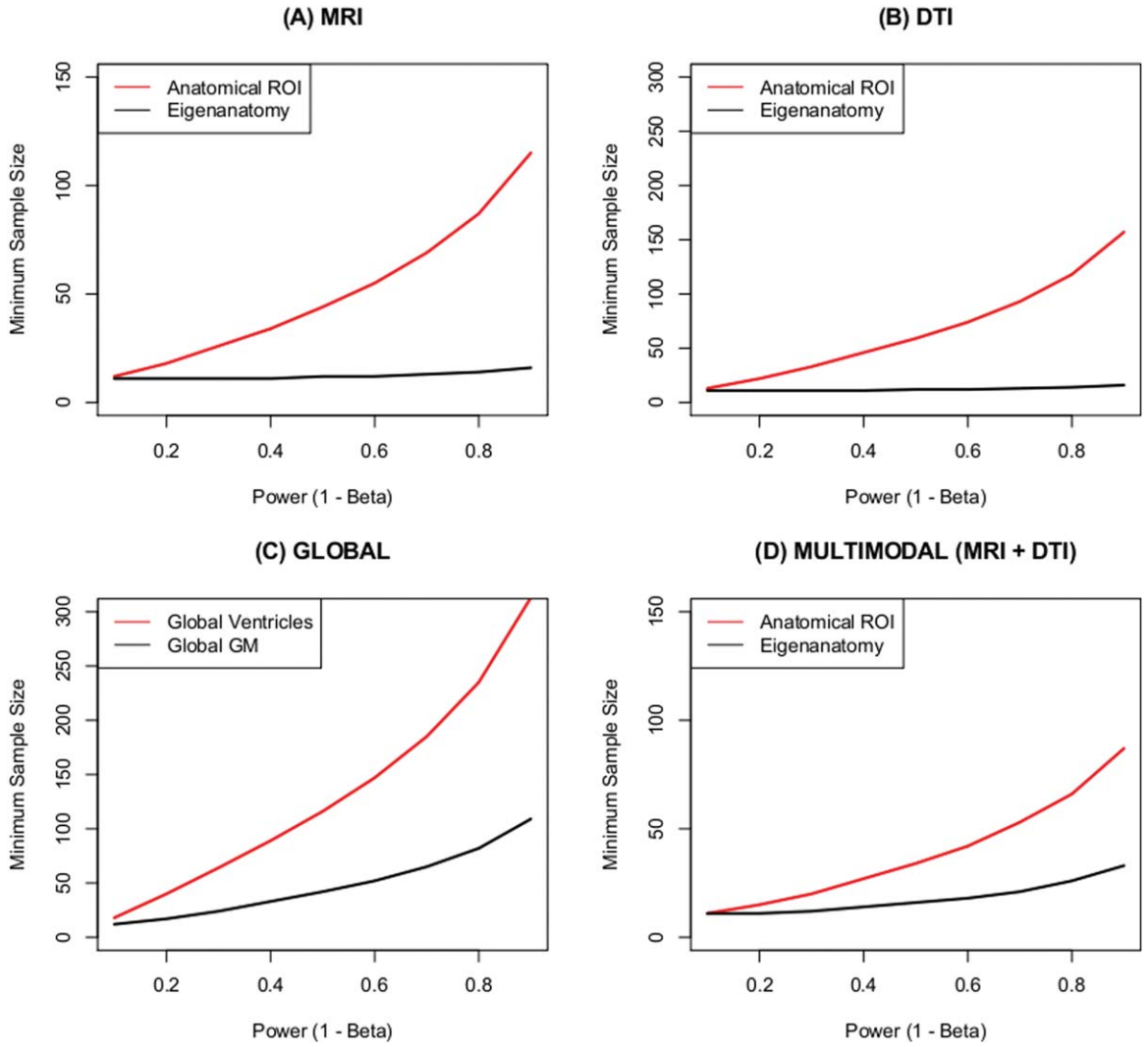

Figure 4.

Estimated minimum sample sizes for volumetric, DTI, and multimodal combination of neuroimaging approaches in independent test dataset.

study achieved very high specificity (100%), we only observed modest sensitivity (65%). Relative to anatomic approaches, global approaches are easy to implement and have the advantage that they are insensitive to individual differences in the anatomic distribution of disease. Although global approaches may be sensitive to detecting the presence of disease in patients compared with controls our findings suggest that global approaches are unlikely to be informative in comparative studies where two patient groups may have relatively equal amounts of two different diseases that are uniquely distributed throughout the brain. One possibility is that global measures may be more useful as clinical endpoints rather than as a screening tool in the context of a clinical trial. Indeed, global measures yield a single value reflecting overall disease that can easily demonstrate change over time, and this has been correlated with clinical and neuropsychological measures of disease progression [Chou et al., 2010; Knopman et al., 2009]. Although a single value can also be derived from data-driven and anatomic approaches [McMillan et al., 2013a], future research is necessary to comparatively evaluate the statistical power of each of the neuroimaging approaches for quantifying longitudinal decline. In a preliminary study, Eigenanatomy was demonstrated to have increased sensitivity to longitudinal decline relative to a standard voxel-based morphometry approach [Avants et al., 2012].

Although this study used linear regression to evaluate classification accuracy across three different VOI approaches, it is important to consider how our observations compared

to previous reports using alternative neuroimaging classification methods to discriminate between FTLD and AD [Avants et al., 2010; Davatzikos et al., 2008; Du et al., 2007; Klöppel et al., 2008; Lindberg et al., 2012; Lu et al., 2014; McMillan et al., 2012, 2013a; Rabinovici et al., 2007; Zhang et al., 2009, 2011; Zhou et al., 2010]. The vast majority of these studies converge by emphasizing the contribution of parietal regions [Avants et al., 2010; Du et al., 2007; Klöppel et al., 2008; McMillan et al., 2012; Zhang et al., 2011] and corpus callosum [Avants et al., 2010; Lu et al., 2014; McMillan et al., 2012; Zhang et al., 2009] for accurate classification. Another GM region commonly reported to accurately discriminate between FTLD and AD is ventromedial prefrontal cortex [Avants et al., 2010; Davatzikos et al., 2008; Klöppel et al., 2008; McMillan et al., 2013a], which contributed to our GM-only classification but did not contribute to classification accuracy when WM was added to the multimodal analysis. Other than the use of multiple imaging modalities, another major difference between our approach and other whole-brain approaches is related to the choice of statistical classifier such as support vector machine [Davatzikos et al., 2008; Klöppel et al., 2008] or canonical correlation analysis [Avants et al., 2010]. Our analyses focused on logistic regression in an effort to facilitate interpretation, but future work is required to directly compare different types of statistical classifiers. Other studies have also focused on sophisticated analyses of an anatomical structure such as shape-based analyses of the hippocampus [Lindberg et al., 2012] or putamen [Looi et al., 2012] and these detailed anatomical analyses may prove to be more sensitive than voxelwise anatomical analyses such as those reported in this study.

Statistical Power for Neuroimaging Biomarkers

The results of this study suggest that a data-driven approach requires fewer participants than anatomical or global approaches to perform accurate classification. We observed a small trade-off between accuracy and sample size because our data-driven sample size estimate is slightly larger for the multimodal approach compared to single modalities. However, despite this trade-off, the data-driven approach using a multimodal dataset can be used for clinical trial screening with as few as 26 cases in comparison to the multimodal anatomical assessment that estimated as many as 66 cases would be required and over 80 cases would be required using a global measure. Sample size is an important consideration in the context of designing a cost-effective, statistically robust, and feasible method to screen patients for an etiologically driven treatment trial. The ability to classify with smaller sample sizes is especially important in the context of less common diseases like FTLD in which there is a relatively limited pool of individuals to recruit for clinical trials.

Critically, our power analyses were based on an independent test cohort and thus address recent objections concerning adequate power in neuroscience studies. If post hoc power analyses are performed in the absence of an independent sample, studies may have based their observed power on overfitted data and therefore may have underestimated the sample size required for replication [Button et al., 2013]. A comparison of estimated minimum sample sizes in our training and testing analyses emphasizes this potential pitfall.

Additional Advantages of Data-Driven Statistical Approaches for Neuroimaging

There are several additional advantages to data-driven statistical methods that extend beyond high classification accuracy and increased statistical power. Although anatomical approaches are constrained to regions that have previously been clearly defined, data-driven approaches may facilitate the detection of novel regions that are important for improving our understanding of FTLD and AD. This is of particular concern for DTI analyses in which anatomical protocols are not well defined owing to challenges associated with defining regions that contain crossing fibers or U-shaped subcortical fibers.

Another potential advantage of data-driven approaches is that multivariate statistics can be used to identify latent signals that may not be captured by more simplistic univariate statistical procedures [Moeller and Habeck, 2006]. In a recent study, we used a similar statistical procedure as reported here to identify cortical networks of neurodegeneration in AD and FTLD that were associated with distinct domains of cognition, including social, linguistic, executive, and memory neuropsychological measures [Avants et al., 2014]. These findings suggest that distributed networks of correlated voxels, rather than a single cluster of adjacent voxels, best account for cognitive deficits observed in neurodegenerative diseases. A similar study in our laboratory suggests that distributed networks of GM and WM are associated with genetic risk factors in FTLD [McMillan et al., 2014]. The advantages of multivariate approaches have also been highlighted in other imaging modalities such as BOLD measures of functional connectivity [Seeley et al., 2009; Zhou et al., 2010] and FDG-PET [Seo et al., 2013] studies that emphasize the importance of network-level variance in FTLD and AD.

As clinical trials emerge it will be necessary to develop statistical methods that are protected from potential challenges associated with multicenter studies such as variance associated with different MRI scanners. We speculate that the data-driven VOI approach may be ideal for multicenter studies because any scanner-to-scanner variance may potentially be captured by an additional nuisance eigenvector that can be residualized during analysis, though future research is required to evaluate this.

Although data-driven approaches appear to have several advantages, detailed anatomical analyses are likely to continue to contribute to an improved understanding of neurodegeneration in FTLD and AD. In particular there has

been an emergence of promising high-field imaging methods [Kerchner, 2011] in AD, some of which include ex vivo histopathological validation [Adler et al., 2014; Augustinack et al., 2013]. However, there is still substantial work required to translate these sophisticated techniques into clinical practice. Some studies suggest that data-driven methods may lack the level of detail required to make subtle and detailed neuroanatomical distinctions required for challenging diagnoses [Klauschen et al., 2009]. Preliminary evidence suggests that data-driven methods can be leveraged to refine anatomically defined cortical regions in order to provide an optimal "hybrid" approach for neuroimaging analysis [Dhillon et al., 2013].

A more practical potential benefit of neuroimaging biomarkers, in comparison to CSF or PET, is that it is widely available, noninvasive, and cost-effective. Nearly every major hospital has access to an MRI scanner while CSF analysis and PET imaging may only be available at specialized tertiary medical centers. Although lumbar punctures in practice have minimal risks associated with them, patients and physicians often vary in their attitudes and perceived risks of this more invasive procedure. In comparison to PET studies, MRI is relatively less expensive and the acquisition of multiple modalities such as GM MRI and DTI in a single session is feasible.

Potential Limitations and Future Considerations

Several caveats should be kept in mind when considering our findings. Our results were based on screening FTLD and AD in a clinical FTD cohort, but FTLD itself is a histopathologically heterogeneous condition that results from either tau, TDP-43, or rarely FUS inclusions [Mackenzie et al., 2010]. It is nonetheless important to screen out AD patients in the context of a clinical trial. For example, a clinical trial (NCT01626378) is currently recruiting bvFTD patients independent of tau or TDP-43 pathology, but lists AD pathology as an exclusion criterion. Prior evidence also suggests that Eigenanatomy, the data-driven method used here, provides highly accurate classification of tau and TDP-43 subtypes of FTLD when using DTI [McMillan et al., 2013b], and unpublished work demonstrates the utility of Eigenanatomy analyses of GM and WM in distinguishing the underlying causes of CBS [Goldmann Gross et al., 2012]. Thus, these findings suggest that Eigenanatomy may be a useful tool for resolving diagnostic challenges when clinical phenotype is less informative. Although our classification analyses were based on a CSF t-tau:Aβ values rather than pathologically confirmed cases, the t-tau:Aβ cutoff that we used has been cross-validated across two autopsy series and has been demonstrated to have greater than 90% sensitivity and specificity across both series [Irwin et al., 2012; Toledo et al., 2012]. Moreover, the brain regions selected by an Eigenanatomy analysis have previously been validated by direct sampling of histopathology in the identified regions [McMillan et al., 2013b].

When considering the histopathological heterogeneity of our cohort it is important to acknowledge that the DTI analyses reported in this study were constrained to FA measures of WM. However, it is possible that different DTI metrics may provide more specific measures of distinct pathological processes in AD or FTLD. For example, some studies have suggested that radial diffusivity provides a more sensitive measure of FTD pathology [Zhang et al., 2009, 2013], but future research with animal models is required to evaluate the controversial relationship between DTI metrics and specific pathological processes [Jones et al., 2013].

In addition to histopathological heterogeneity, the clinical syndromes associated with AD or FTLD pathology can be heterogeneous. In this study, patients were initially screened for a clinical diagnosis of an FTD-spectrum syndrome and we evaluated the accuracy of neuroimaging approaches to identify FTLD or AD underlying pathology. Importantly, clinical evaluation with current criteria [Armstrong et al., 2013; Gorno-Tempini et al., 2011; Litvan et al., 1996; Rascovsky et al., 2011] is a critical first step for identifying individuals with an FTD syndrome and is important for identifying the appropriate treatment and management of clinical symptoms. However, we argue that clinical diagnosis alone must be complemented with other biomarkers to identify the underlying pathological source of disease. For example, visuospatial difficulties are observed in CBS and posterior cortical atrophy, behavioral difficulties are observed in bvFTD and frontal-variant AD, and language difficulties are observed in PPA, which have been associated with AD and FTLD. In a post hoc analysis we observed that our classification procedure performed well in more heterogeneous clinical syndromes like CBS, though these analyses were based on small numbers of cases and require additional validation in a larger series. Accounting for differences in disease distribution across clinical syndromes is important given prior evidence that suggests that underlying pathology mediates the distribution of disease observed within a given clinical syndrome [Hu et al., 2010]. For example, within a single clinical syndrome such as lvPPA and nonfluent/agrammatic variant of PPA (naPPA) patients exhibit a more anterior distribution of disease with FTLD pathology and more posterior distribution with AD pathology [Hu et al., 2010]. Similarly, MRI and DTI demonstrate distinct patterns of atrophy in CBS patients with AD or FTLD histopathology [Goldmann Gross et al., 2012]. Thus, it appears that clinical syndrome alone may not be a reliable predictor of regional atrophy in AD or FTLD.

Lastly, to implement a data-driven and multimodal screening procedure for clinical trials it may be necessary to validate our observations in an independent dataset. This study used independent training and testing cohorts from our center and demonstrated that Eigenanatomy performed well in both datasets. However, unlike large AD neuroimaging repositories such as ADNI or OASIS, there are not publically available datasets that contain a large

FTD neuroimaging series. Eigenanatomy is available for free open-source download and we encourage investigators to independently evaluate the reliability and replicability of our tool in their own FTD datasets.

With these caveats in mind, our results suggest that a data-driven approach to analyze multimodal neuroimaging provides an optimal strategy for screening patients for entry into clinical trials. The proposed methods are user-independent, provide reliable VOIs, and demonstrate adequate power for use in therapeutic trials of rare neurodegenerative diseases such as FTLD.

ACKNOWLEDGMENTS

All authors have no relevant disclosures. Eigenanatomy is provided through open-source licensing in the SCCAN package of Advanced Normalization Tools (ANTs). See https://github.com/stnava/sccan for download and address software support questions to Brian Avants (stnava@gmail.com).

REFERENCES

Adler DH, Pluta J, Kadivar S, Craige C, Gee JC, Avants BB, Yushkevich PA (2014): Histology-derived volumetric annotation of the human hippocampal subfields in postmortem MRI. Neuroimage 84:505–523.

Alexander DC, Pierpaoli C, Basser PJ, Gee JC (2001): Spatial transformations of diffusion tensor magnetic resonance images. IEEE Trans Med Imaging 20:1131–1139.

Ard MC, Edland SD (2011): Power calculations for clinical trials in Alzheimer's disease. J Alzheimers Dis 26 (Suppl 3):369–377.

Armstrong MJ, Litvan I, Lang AE, Bak TH, Bhatia KP, Borroni B, Boxer AL, Dickson DW, Grossman M, Hallett M, Josephs KA, Kertesz A, Lee SE, Miller BL, Reich SG, Riley DE, Tolosa E, Tröster AI, Vidailhet M, Weiner WJ (2013): Criteria for the diagnosis of corticobasal degeneration. Neurology 80:496–503.

Augustinack JC, van der Kouwe AJW, Fischl B (2013): Medial temporal cortices in ex vivo magnetic resonance imaging. J Comp Neurol 521:4177–4188.

Avants BB, Epstein CL, Grossman M, Gee JC (2008): Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12:26–41.

Avants BB, Cook PA, Ungar L, Gee JC, Grossman M (2010): Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis. Neuroimage 50: 1004–1016.

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011): A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54:2033–2044.

Avants B, Dhillon P, Kandel BM, Cook PA, McMillan CT, Grossman M, Gee JC (2012): Eigenanatomy improves detection power for longitudinal cortical change. Med Image Comput Comput Assist Interv 15:206–213.

Avants BB, Libon DJ, Rascovsky K, Boller A, McMillan CT, Massimo L, Coslett HB, Chatterjee A, Gross RG, Grossman M (2014): Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. Neuroimage 84:698–711.

Boxer AL, Gold M, Huey E, Gao F-B, Burton EA, Chow T, Kao A, Leavitt BR, Lamb B, Grether M, Knopman D, Cairns NJ, Mackenzie IR, Mitic L, Roberson ED, Van Kammen D, Cantillon M, Zahs K, Salloway S, Morris J, Tong G, Feldman H, Fillit H, Dickinson S, Khachaturian Z, Sutherland M, Farese R, Miller BL, Cummings J (2013a): Frontotemporal degeneration, the next therapeutic frontier: Molecules and animal models for frontotemporal degeneration drug development. Alzheimers Dement 9:176–188.

Boxer AL, Knopman DS, Kaufer DI, Grossman M, Onyike C, Graf-Radford N, Mendez M, Kerwin D, Lerner A, Wu C-K, Koestler M, Shapira J, Sullivan K, Klepac K, Lipowski K, Ullah J, Fields S, Kramer JH, Merrilees J, Neuhaus J, Mesulam MM, Miller BL (2013b): Memantine in patients with frontotemporal lobar degeneration: A multicentre, randomised, double-blind, placebo-controlled trial. Lancet Neurol 12:149–156.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013): Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365–376.

Chou Y-Y, Leporé N, Saharan P, Madsen SK, Hua X, Jack CR, Shaw LM, Trojanowski JQ, Weiner MW, Toga AW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative (2010): Ventricular maps in 804 ADNI subjects: Correlations with CSF biomarkers and clinical decline. Neurobiol Aging 31:1386–1400.

Cook P, Bai Y, Nedjati-Gilani S (2006): Camino: Open-source diffusion-MRI reconstruction and processing. Presented at the 14th Scientific Meeting of the International Society for Magnetic Resonance Imaging in Medicine, Seattle, WA, p 2759.

Das SR, Avants BB, Grossman M, Gee JC (2009): Registration based cortical thickness measurement. Neuroimage 45:867–879.

Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM (2008): Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 41: 1220–1227.

Dhillon P, Gee JC, Ungar LH, Avants B (2013): Anatomically-constrained PCA for image parcellation. Presented at the International Workshop on Pattern Recognition in Neuroimaging (PRNI), Philadelphia, PA, pp 25–28.

Du A-T, Schuff N, Kramer JH, Rosen HJ, Gorno-Tempini ML, Rankin K, Miller BL, Weiner MW (2007): Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. Brain 130:1159–1166.

Galton CJ, Patterson K, Graham K, Lambon-Ralph MA, Williams G, Antoun N, Sahakian BJ, Hodges JR (2001): Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. Neurology 57:216–225.

Goldmann Gross R, McMillan C, Cook P, Grossman M (2012): Volumetric imaging of gray and white matter in corticobasal syndrome (P05.035). Neurology 78:P05.041.

Goñi J, Cervantes S, Arrondo G, Lamet I, Pastor P, Pastor MA (2013): Selective brain gray matter atrophy associated with APOE ϵ4 and MAPT H1 in subjects with mild cognitive impairment. J Alzheimers Dis 33:1009–1019.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M (2011): Classification of primary progressive aphasia and its variants. Neurology 76:1006–1014.

Harris JM, Gall C, Thompson JC, Richardson AMT, Neary D, Plessis, Du D, Pal P, Mann DMA, Snowden JS, Jones M (2013): Sensitivity and specificity of FTDC criteria for behavioral variant frontotemporal dementia. Neurology 80:1881–1887.

Hassan A, Whitwell JL, Josephs KA (2011): The corticobasal syndrome-Alzheimer's disease conundrum. Expert Rev Neurother 11:1569–1578.

Hornberger M, Wong S, Tan R, Irish M, Piguet O, Kril J, Hodges JR, Halliday G (2012): In vivo and post-mortem memory circuit integrity in frontotemporal dementia and Alzheimer's disease. Brain 135:3015–3025.

Hu WT, Rippon GW, Boeve BF, Knopman DS, Petersen RC, Parisi JE, Josephs KA (2009): Alzheimer's disease and corticobasal degeneration presenting as corticobasal syndrome. Mov Disord 24:1375–1379.

Hu WT, McMillan C, Libon D, Leight S, Forman M, Lee VM-Y, Trojanowski JQ, Grossman M (2010): Multimodal predictors for Alzheimer disease in nonfluent primary progressive aphasia. Neurology 75:595–602.

Irwin DJ, McMillan CT, Toledo JB, Arnold SE, Shaw LM, Wang L-S, Van Deerlin V, Lee VM-Y, Trojanowski JQ, Grossman M (2012): Comparison of cerebrospinal fluid levels of tau and Aβ 1–42 in Alzheimer disease and frontotemporal degeneration using 2 analytical platforms. Arch Neurol 69:1018–1025.

Jones DK, Knösche TR, Turner R (2013): White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. Neuroimage 73:239–254.

Josephs KA, Whitwell JL, Jack CR, Parisi JE, Dickson DW (2006): Frontotemporal lobar degeneration without lobar atrophy. Arch Neurol 63:1632–1638.

Kerchner GA (2011): Ultra-high field 7T MRI: A new tool for studying Alzheimer's disease. J Alzheimers Dis 26 (Suppl 3): 91–95.

Klauschen F, Goldman A, Barra V, Meyer-Lindenberg A, Lundervold A (2009): Evaluation of automated brain MR image segmentation and volumetry methods. Hum Brain Mapp 30:1310–1327.

Klein A, Ghosh SS, Avants B, Yeo BTT, Fischl B, Ardekani B, Gee JC, Mann JJ, Parsey RV (2010): Evaluation of volume-based and surface-based brain image registration methods. Neuroimage 51:214–220.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR, Ashburner J, Frackowiak RSJ (2008): Automatic classification of MR scans in Alzheimer's disease. Brain 131:681–689.

Knopman DS, Jack CR, Kramer JH, Boeve BF, Caselli RJ, Graff-Radford NR, Mendez MF, Miller BL, Mercaldo ND (2009): Brain and ventricular volumetric changes in frontotemporal lobar degeneration over 1 year. Neurology 72:1843–1849.

Kohannim O, Hua X, Hibar DP, Lee S, Chou Y-Y, Toga AW, Jack CR, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative (2010): Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging 31:1429–1442.

Lee SE, Rabinovici GD, Mayo MC, Wilson SM, Seeley WW, DeArmond SJ, Huang EJ, Trojanowski JQ, Growdon ME, Jang JY, Sidhu M, See TM, Karydas AM, Gorno-Tempini ML, Boxer AL, Weiner MW, Geschwind MD, Rankin KP, Miller BL (2011): Clinicopathological correlations in corticobasal degeneration. Ann Neurol 70:327–340.

Lindberg O, Walterfang M, Looi JCL, Malykhin N, Ostberg P, Zandbelt B, Styner M, Paniagua B, Velakoulis D, Orndahl E, Wahlund L-O (2012): Hippocampal shape analysis in Alzheimer's disease and frontotemporal lobar degeneration subtypes. J Alzheimers Dis 30:355–365.

Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, Goetz CG, Golbe LI, Grafman J, Growdon JH, Hallett M, Jankovic J, Quinn NP, Tolosa E, Zee DS (1996): Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP international workshop. Neurology 47:1–9.

Looi JCL, Rajagopalan P, Walterfang M, Madsen SK, Thompson PM, Macfarlane MD, Ching C, Chua P, Velakoulis D (2012): Differential putaminal morphology in Huntington"s disease, frontotemporal dementia and Alzheimer"s disease. Aust N Z J Psychiatry 46:1145–1158.

Lu PH, Lee GJ, Shapira J, Jimenez E, Mather MJ, Thompson PM, Bartzokis G, Mendez MF (2014): Regional differences in white matter breakdown between frontotemporal dementia and early-onset Alzheimer's disease. J Alzheimers Dis 39:261–269.

Mackenzie IRA, Neumann M, Bigio EH, Cairns NJ, Alafuzoff I, Kril J, Kovacs GG, Ghetti B, Halliday G, Holm IE, Ince PG, Kamphorst W, Revesz T, Rozemuller AJM, Kumar-Singh S, Akiyama H, Baborie A, Spina S, Dickson DW, Trojanowski JQ, Mann DMA (2010): Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: An update. Acta Neuropathol 119:1–4.

McMillan CT, Brun C, Siddiqui S, Churgin M, Libon D, Yushkevich P, Zhang H, Boller A, Gee J, Grossman M (2012): White matter imaging contributes to the multimodal diagnosis of frontotemporal lobar degeneration. Neurology 78:1761–1768.

McMillan CT, Avants B, Irwin DJ, Toledo JB, Wolk DA, Van Deerlin VM, Shaw LM, Trojanoswki JQ, Grossman M (2013a): Can MRI screen for CSF biomarkers in neurodegenerative disease? Neurology 80:132–138.

McMillan CT, Irwin DJ, Avants BB, Powers J, Cook PA, Toledo JB, McCarty Wood E, Van Deerlin VM, Lee VM-Y, Trojanowski JQ, Grossman M (2013b): White matter imaging helps dissociate tau from TDP-43 in frontotemporal lobar degeneration. J Neurol Neurosurg Psychiatry 84:949–955.

McMillan CT, Toledo JB, Avants BB, Cook PA, Wood EM, Suh E, Irwin DJ, Powers J, Olm C, Elman L, McCluskey L, Schellenberg GD, Lee VM-Y, Trojanowski JQ, Van Deerlin VM, Grossman M (2014): Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. Neurobiol Aging 35:1473–1482.

Mesulam M, Wicklund A, Johnson N, Rogalski E, Léger GC, Rademaker A, Weintraub S, Bigio EH (2008): Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. Ann Neurol 63:709–719.

Moeller JR, Habeck CG (2006): Reciprocal benefits of mass-univariate and multivariate modeling in brain mapping: Applications to event-related functional MRI, H(2) (15)O-, and FDG-PET. Int J Biomed Imaging 2006:79862.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR, Schuff N, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative (2009): Automated 3D mapping of hippocampal atrophy and its clinical correlates in 400 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Hum Brain Mapp 30:2766–2788.

Muñoz-Ruiz MÁ, Hartikainen P, Koikkalainen J, Wolz R, Julkunen V, Niskanen E, Herukka S-K, Kivipelto M, Vanninen R, Rueckert D, Liu Y, Lötjönen J, Soininen H (2012): Structural

MRI in frontotemporal dementia: Comparisons between hippocampal volumetry, tensor-based morphometry and voxel-based morphometry. PLoS One 7:e52531.

Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW (2011): Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: A retrospective study. Lancet Neurol 10:785–796.

Oishi K, Zilles K, Amunts K, Faria A, Jiang H, Li X, Akhter K, Hua K, Woods R, Toga AW, Pike GB, Rosa-Neto P, Evans A, Zhang J, Huang H, Miller MI, van Zijl PCM, Mazziotta J, Mori S (2008): Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. Neuroimage 43:447–457.

Pelaez-Coca M, Bossa M, Olm C; Alzheimer's Disease Neuroimaging Initiative (ADNI) (2011): Discrimination of AD and normal subjects from MRI: Anatomical versus statistical regions. Neurosci Lett 487:113–117.

Rabinovici GD, Seeley WW, Kim EJ, Gorno-Tempini ML, Rascovsky K, Pagliaro TA, Allison SC, Halabi C, Kramer JH, Johnson JK, Weiner MW, Forman MS, Trojanowski JQ, Dearmond SJ, Miller BL, Rosen HJ (2007): Distinct MRI atrophy patterns in autopsy-proven Alzheimer's disease and frontotemporal lobar degeneration. Am J Alzheimers Dis Other Demen 22:474–488.

Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, van Swieten JC, Seelaar H, Dopper EGP, Onyike CU, Hillis AE, Josephs KA, Boeve BF, Kertesz A, Seeley WW, Rankin KP, Johnson JK, Gorno-Tempini ML, Rosen H, Prioleau-Latham CE, Lee A, Kipps CM, Lillo P, Piguet O, Rohrer JD, Rossor MN, Warren JD, Fox NC, Galasko D, Salmon DP, Black SE, Mesulam M, Weintraub S, Dickerson BC, Diehl-Schmid J, Pasquier F, Deramecourt V, Lebert F, Pijnenburg Y, Chow TW, Manes F, Grafman J, Cappa SF, Freedman M, Grossman M, Miller BL (2011): Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134:2456–2477.

Salvador R, Peña A, Menon DK, Carpenter TA, Pickard JD, Bullmore ET (2005): Formal characterization and extension of the linearized diffusion tensor model. Hum Brain Mapp 24: 144–155.

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD (2009): Neurodegenerative diseases target large-scale human brain networks. Neuron 62:42–52.

Seo EH, Lee DY, Lee J-M, Park J-S, Sohn BK, Lee DS, Choe YM, Woo JI (2013): Whole-brain functional networks in cognitively normal, mild cognitive impairment, and Alzheimer's disease. PLoS One 8:e53922.

Toledo JB, Brettschneider J, Grossman M, Arnold SE, Hu WT, Xie SX, Lee VM-Y, Shaw LM, Trojanowski JQ (2012): CSF biomarkers cutoffs: The importance of coincident neuropathological diseases. Acta Neuropathol 124:23–35.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

Zhang Y, Schuff N, Du A-T, Rosen HJ, Kramer JH, Gorno-Tempini ML, Miller BL, Weiner MW (2009): White matter damage in frontotemporal dementia and Alzheimer's disease measured by diffusion MRI. Brain 132:2579–2592.

Zhang Y, Schuff N, Ching C, Tosun D, Zhan W, Nezamzadeh M, Rosen HJ, Kramer JH, Gorno-Tempini ML, Miller BL, Weiner MW (2011): Joint assessment of structural, perfusion, and diffusion MRI in Alzheimer's disease and frontotemporal dementia. Int J Alzheimers Dis 2011:546871.

Zhang Y, Tartaglia MC, Schuff N, Chiang GC, Ching C, Rosen HJ, Gorno-Tempini ML, Miller BL, Weiner MW (2013): MRI signatures of brain macrostructural atrophy and microstructural degradation in frontotemporal lobar degeneration subtypes. J Alzheimers Dis 33:431–444.

Zhou J, Greicius MD, Gennatas ED, Growdon ME, Jang JY, Rabinovici GD, Kramer JH, Weiner M, Miller BL, Seeley WW (2010): Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. Brain 133:1352–1367.