NeuroImage 56 (2011) 788–796

Contents lists available at ScienceDirect

NeuroImage

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

Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques

Angela Rizk-Jackson [a], Diederick Stoffers [e], Sarah Sheldon [e], Josh Kuperman [b], Anders Dale [b], Jody Goldstein [b], Jody Corey-Bloom [b], Russell A. Poldrack [a,c,d], Adam R. Aron [e,*]

[a] Department of Neuroscience, University of California Los Angeles, Los Angeles, CA, USA
[b] Department of Neurosciences, University of California San Diego, San Diego, CA, USA
[c] Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA
[d] Department of Psychology, University of Texas at Austin, Austin, TX, USA
[e] Department of Psychology, University of California San Diego, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 16 December 2009
Revised 27 April 2010
Accepted 30 April 2010
Available online 6 May 2010

Keywords:
Huntington's disease
Biomarker
Neuroimaging
Machine learning
Support vector machine
Support vector regression

ABSTRACT

The development of MRI measures as biomarkers for neurodegenerative disease could prove extremely valuable for the assessment of neuroprotective therapies. Much current research is aimed at developing such biomarkers for use in people who are gene-positive for Huntington's disease yet exhibit few or no clinical symptoms of the disease (pre-HD). We acquired structural (T1), diffusion weighted and functional MRI (fMRI) data from 39 pre-HD volunteers and 25 age-matched controls. To determine whether it was possible to decode information about disease state from neuroimaging data, we applied multivariate pattern analysis techniques to several derived voxel-based and segmented region-based datasets. We found that different measures of structural, diffusion weighted, and functional MRI could successfully classify pre-HD and controls using support vector machines (SVM) and linear discriminant analysis (LDA) with up to 76% accuracy. The model producing the highest classification accuracy used LDA with a set of six volume measures from the basal ganglia. Furthermore, using support vector regression (SVR) and linear regression models, we were able to generate quantitative measures of disease progression that were significantly correlated with established measures of disease progression (estimated years to clinical onset, derived from age and genetic information) from several different neuroimaging measures. The best performing regression models used SVR with neuroimaging data from regions within the grey matter (caudate), white matter (corticospinal tract), and fMRI (insular cortex). These results highlight the utility of machine learning analyses in addition to conventional ones. We have shown that several neuroimaging measures contain multivariate patterns of information that are useful for the development of disease-state biomarkers for HD.

Introduction

In neurodegenerative disease, changes in the brain can precede overt symptoms by many years. For patients with these diseases it is imperative to develop in-vivo measures (biomarkers) that can track early disease-induced neural changes, especially before overt symptoms arise. Such biomarkers could provide metrics to evaluate neural change over time as well as the outcome of neuroprotective trials (reviewed by Bohanna et al., 2008; Hersch and Rosas, 2008).

One such neurodegenerative condition, Huntington's disease (HD), is aptly suited to the project of developing disease-state biomarkers. HD is a dominantly inherited disorder for which the genetic marker, an expanded CAG triplet on the huntingtin gene, is fully penetrant. This means that, given time, all individuals with the genetic marker will develop progressive neurodegeneration associated with HD. It is possible to estimate the degree of disease progression using an individual's age and the number of CAG repeats in the HD gene, this is referred to here as age/CAG-estimated years to onset, or YTO (Aylward et al., 1996; Langbehn et al., 2004). Currently, clinical diagnosis of the disease is based on motor symptoms, although it is clear that clinical motor symptoms are preceded by both cognitive and psychiatric changes (Beglinger et al., 2005, 2008; Duff et al., 2007; Marshall et al., 2007), as well as measurable changes in brain structure and function (reviewed below). Because of these characteristics, namely a fully penetrant genetic marker of disease, an estimable degree of disease progression (YTO), and evidence of disease manifestation prior to overt (motor) symptoms, HD is a prime target for the development of MRI-based disease-state biomarkers.

There has been recent progress toward this end in HD (Georgiou-Karistianis, 2009; Paulsen, 2009; Paulsen et al., 2006a). Potential biomarkers include structural MRI measures of grey and white matter,

* Corresponding author. Department of Psychology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.
E-mail address: adamaron@ucsd.edu (A.R. Aron).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.04.273

as well as functional MRI (fMRI). The characteristic neuropathology in HD is degeneration of subcortical structures within the basal ganglia, mainly the striatum (Vonsattel and DiFiglia, 1998). Consistent with this, studies of pre-HD have shown that striatal atrophy begins a decade or more before estimated clinical diagnosis and becomes more severe as clinical symptom onset approaches (Aylward et al., 1996, 2000, 2004). Other studies of pre-HD have also shown abnormalities in 1) grey matter regions outside of the basal ganglia (Gómez-Ansón et al., 2009; Kipps et al., 2005; Paulsen et al., 2006b; Rosas et al., 2005; Thieben et al., 2002), 2) white matter (Klöppel et al., 2008; Reading et al., 2005; Rosas et al., 2006) and 3) fMRI signal (Paulsen et al., 2004; Reading et al., 2004; Wolf et al., 2007; Zimbelman et al., 2007). Importantly, some of these effects are only seen in individuals who are close to predicted onset, thus indicating the presence of progressive longitudinal changes in the brain imaging data of pre-HD individuals.

The abovementioned studies have all used conventional analyses of imaging data to provide a strong foundation upon which to build. Here we use novel analysis methods based on machine learning algorithms that can examine imaging datasets in new ways (Pereira et al., 2009). These methods make use of the entire multivariate pattern present in a dataset in order to create models that allow predictions to be made about new data. This approach was recently demonstrated by Klöppel et al. using segmented grey matter data (Klöppel et al., 2009), as well as white matter data (Klöppel et al., 2008) to discriminate pre-HD from controls.

Here we sought to replicate and extend the findings of Klöppel et al. by using similar analysis methods on multiple forms of imaging data including segmented grey matter images, segmented region-based morphometric data, diffusion weighted (white matter) images, and fMRI images. In addition to testing classification accuracy using these various measures, we also used regression models to examine whether it was possible to make quantitative predictions about a person's disease progression (i.e. predicted years to clinical onset) based on imaging data. If so, this would suggest that these methods may be useful not merely for discriminating pre-HD individuals from controls, but also for indexing the level of disease progression. This is an important requirement for biomarkers that may be used for measuring longitudinal neurodegeneration and neuroprotection.

Methods

Participants

Neuroimaging data were obtained from 39 pre-HD individuals and 25 age- and sex-matched controls. Each pre-HD individual had more than 36 CAG repeats and a Unified Huntington's disease Rating Scale confidence rating below 2. The data were collected at UC San Diego and consent was obtained in accordance with the UCSD Institutional Review Board. Neuroimaging data used in the following analyses were derived from T1-weighted MRI structural scans, diffusion weighted imaging scans, and fMRI scans. Due to technical problems, one pre-HD participant was excluded from the voxel-based grey matter analysis (due to severe motion-induced artifacts), one control and one pre-HD participant were excluded from the voxel-based white matter analysis (due to imaging acquisition failures), five participants (1 control and 4 pre-HD) were excluded from the subcortical volume segmented region-based dataset, and one pre-HD from the brain volume segmented region-based dataset (due to above mentioned issues in addition to low grey/white contrast). The evaluation of grey/white contrast was a quality-control step carried out by research assistants performing the FreeSurfer segmentation. They were blind to pre-HD vs. control classification of the data (Table 1).

For pre-HD individuals, we obtained two different estimates of years to onset (YTO): YTO-L using the Langbehn method (Langbehn et al., 2004), and YTO-A using the Aylward method (Aylward et al., 1996). Each of these methods makes use of the individual's age and number of CAG repeats to generate YTO estimates. YTO-A was simply calculated using the formula

$$YTO = (-0.81(CAG) + 0.51(PO) + 54.87) - AGE$$

where CAG is the number of CAG repeats, PO is the age of parental symptom onset, and AGE is the current age of the individual. YTO-L was derived from tables associated with the survival analysis formula of Langbehn et al. (2004). Both of these estimation techniques have been widely used in the field, though the Aylward formula reflects a simple linear relationship and the more recent Langbehn method employs a more sophisticated probabilistic modeling technique using a non-linear parametric equation to generate estimates. The YTO-L and YTO-A estimates are correlated with $r = 0.55$ in our subject sample. These continuous markers of disease progression were used in the regression analyses described below.

Table 1
Characteristics of the full cohort of our subject population. Though a small number of different subjects were excluded from each particular dataset (see Methods for details), these exclusions present no significant changes in the distributions of the characteristics displayed in this table. WP accuracy and reaction time are two performance measures collected during the fMRI weather prediction task. MMD (mean movement distance) is a calculated metric describing head motion during fMRI. No significant group differences are seen with these measures (see *p*-values).

	Control ($N = 25$)	Pre-HD ($N = 39$)	*p*-value
Age (years, mean ± SD [range])	39.1 ± 12.1 [21,64]	40.5 ± 10.4 [22,64]	0.63
Sex (M/F)	9/16	17/22	
WP performance (% accuracy ± SD)	0.64 ± 0.1	0.67 ± 0.1	0.28
WP reaction time (ms ± SD)	1.11 ± 0.23	1.14 ± 0.26	0.50
MMD (mean ± SD)	0.080 ± 0.044	0.081 ± 0.048	0.95
Number of CAG repeats (mean ± SD [range])	–	42.2 ± 2.4 [38,48]	
YTO-L (mean ± SD [range])	–	14.9 ± 7.8 [5,37]	
YTO-A (mean ± SD [range])	–	6.3 ± 7.8 [−5,23]	

MRI data and preprocessing

Voxel-based grey matter (GM)

The T1-weighted MRI structural scans were collected on a General Electric 1.5 T EXCITE HD scanner with an 8-channel phased-array head coil (TE = 2.796 ms, TR = 6.496 ms, TI = 600 ms, flip angle = 12°, FOV = 24 cm, matrix = 256 × 192, and slice thickness = 1.2 mm). The images were corrected for non-linear warping caused by gradient coil non-linearities using in-house software and the image intensities were normalized with the ratio of a body coil to a head coil scan in order to correct for spatial sensitivity inhomogeneities in the head coil.

Voxel-based measures of grey matter distribution were obtained from the structural images using a VBM-style analysis (Ashburner and Friston, 2000) implemented in FSL (4.0) (http://www.fmrib.ox.ac.uk/fsl/). Following brain extraction, FAST4 was used to carry out tissue-type segmentation. The resulting grey matter images were aligned to MNI standard space using linear registration (FLIRT) followed by non-linear registration (FNIRT). A study-specific template was then created by averaging these images of the 25 control subjects along with those of 25 age- and sex-matched pre-HDs. All 63 of the grey matter images in native subject-space were then non-linearly registered to this study-specific template. The grey matter partial volume images were then divided by the Jacobian of the warp field and smoothed with a 3 mm Gaussian kernel to produce the final grey matter (GM) dataset. For full details, see Stoffers et al. (2010).

Voxel-based white matter (WM)

Voxel-based measures of white matter were obtained from DTI scans acquired using single-shot EPI with isotropic 2.5 mm voxels

(TE = 80.4 m, TR = 13.2 s, FOV = 24 cm, matrix = 96 × 96, 47 axial slices, slice thickness = 2.5 mm, 51 directions, and $b = 1000$ mm^2/s). Images were preprocessed using in-house software that included unwarping to remove geometric distortion, and also eddy current correction. These data were then spatially normalized by projecting to an invariant white matter skeleton using the tract-based spatial statistics approach in FSL (Smith et al., 2006). Fractional anisotropy (FA) values were extracted and skeletonized white matter tracts were used to mask the FA images in order to produce the final white matter (WM) dataset. For full details, see Stoffers et al. (2010).

Voxel-based fMRI

Voxel-based measures of brain activation were extracted from the blood oxygenation level dependent (BOLD) signal collected during fMRI scans. These were acquired on a General Electric 3 T scanner. The imaging parameters were the same as reported in Aron et al. (2006); i.e. we acquired T2*-weighted echoplanar images (EPI) (slice thickness = 4 mm, 33 slices, TR = 2 s, TE = 30 ms, flip angle = 90°, matrix 64 × 64, FOV = 200 mm). For these data, subjects performed a feedback-based probabilistic learning task known as "weather prediction" (Aron et al., 2006) while BOLD signal was collected. The weather prediction paradigm is extremely well-motivated as a task for activating the fronto-striatal systems that are probably most affected in pre-HD (Shohamy et al., 2004; Aron et al., 2004, 2006; Knowlton et al., 1996). A standard GLM analysis was performed in order to extract contrast images representing the BOLD signal associated with task minus the BOLD signal associated with rest, see Aron et al. (2006) for full details. We used the associated fMRI *z*-statistic images in the following analyses.

Segmented region-based measures

We performed two types of segmented region-based analysis. First, we used SIENAX tools implemented in FSL (Smith et al., 2001, 2002) to produce coarse whole-brain segments from which we derived measures of normalized volumes (in mm^3) of total brain, grey and white matter. Second, we used the Freesurfer (3.0.5) software package (http://surfer.nmr.mgh.harvard.edu) to create segments of the cortical surface and subcortical volumes (Fischl et al., 2002). From these segments, we derived measures of mean cortical thickness (in mm), and volumetric measures of various subcortical structures (in percent intracranial volume). There were 35 total segmented cortical areas (left and right) for which a measure of cortical thickness was obtained, and there were 11 subcortical structures for which volumes were obtained. These subcortical structures include basal ganglia (caudate, putamen, and pallidum), as well as accumbens area, thalamus, hippocampus, and amygdala.

Learning algorithms

SVM learning algorithms were trained to create models that could be used to a) classify individuals as belonging to a particular group (pre-HD or control) and b) make quantitative predictions about the level of disease progression in a pre-HD individual (based on age/CAG-estimated YTO). A detailed description of these algorithms is available elsewhere (Pereira et al., 2009; Smola and Schölkopf, 2004; Schölkopf and Smola, 2002). Here we provide a brief description.

SVM classification algorithms create a model that constructs an optimal separating hyperplane within the multidimensional dataset such that data from individuals belonging to one class are on one side of the hyperplane in multidimensional space, while those of the other class are on the other side of the hyperplane. Construction of this hyperplane is achieved by training the model using a labeled subset of the data (the training-set). Using these labeled training data, the optimal separating hyperplane is found by maximizing the margin of separation of the two classes while minimizing the amount of classification error. Once the model has been trained, the remainder of the data (the test-set) is used to test the ability of the model to accurately generalize to novel data examples, thus providing an estimate of model accuracy.

Regression algorithms using support vector regression (SVR) work much the same way as the SVM classification algorithms, with the goal of constructing a regression line that fits the data within some chosen level of error (ε). In this case, the optimization problem is solved by finding a solution that minimizes a cost function associated with making errors larger than (the ε-insensitive loss function). Once the regression model is trained, a test-set can be used to derive quantitative (continuous) estimates of a variable such as predicted YTO. The accuracy of the regression model can then be evaluated by comparing the model-predicted YTO (for pre-HD subjects in the test-set) with age/CAG-estimated YTO for these same subjects using both the Aylward and Langbehn methods (YTO-A and YTO-L).

For datasets where the number of features was substantially smaller than the number of observations, we also applied simpler methods (linear discriminant analysis for classification, multiple linear regression for regression) in order to confirm the validity of applying SVM algorithms to these datasets.

Feature selection

A fundamental problem in the analysis of high-dimensional datasets is the curse of dimensionality, meaning that as the number of dimensions in the dataset (e.g., voxels in the image) increases, the number of observations necessary to adequately sample the entire space increases exponentially. Accordingly, it is often preferable to try to reduce the dimensionality of neuroimaging datasets. One approach is to select features based on their relative usefulness for classification (e.g. Guyon et al., 2002), but this requires splitting the data such that the feature selection is based on a different set of observations from the test data. This was not a good option in our case because of the relatively small number of subjects in our 4-fold cross-validation training and testing sets. Instead, we took two approaches to dimensionality reduction that incorporate independent knowledge about brain anatomy. First, we used segmented region-based measures derived from the structural MRI images. Second, we parcellated the whole-brain images into regions of interest (ROIs) and created models using only voxels within a given ROI or set of ROIs.

This ROI parcellation served a dual purpose: in addition to creating smaller feature sets, it also allowed us to explore the potential informativeness of each of the various ROIs separately. ROIs for the GM and fMRI datasets were determined using the Harvard–Oxford probabilistic atlas (distributed with FSL). The Harvard–Oxford atlas includes 48 cortical and 7 subcortical structural areas. We created feature sets from our whole-brain data by extracting the data from voxels within each individual ROI in the atlas (55 in total, combining left and right hemisphere ROIs). In addition, we created three combined-ROI feature sets based on the fronto-striatal network known to be affected in HD. The combined-ROI feature sets included frontal cortex (a combination of the frontal pole, superior frontal gyrus, middle frontal gyrus, and inferior frontal gyrus ROIs), basal ganglia (a combination of the caudate, putamen, and pallidum ROIs), and frontal cortex and basal ganglia (a combination of the previous frontal cortex and basal ganglia combined-ROI feature sets). Therefore a total of 58 ROI feature sets were created for the GM and fMRI data: the 55 individual Harvard–Oxford probabilistic ROIs, and the three combined-ROI feature sets described above.

ROIs for the WM dataset were determined using the Johns Hopkins University white matter tractography atlas (distributed with FSL), which covers 10 probabilistically identified white matter structures. Again, feature sets were created for each individual ROI within the atlas (combining left and right hemisphere ROIs). Therefore a total of 10 ROI feature sets were created for the WM data.

The total number of models was large. For the grey matter voxel-based analysis, we tested one whole-brain model plus 58 ROI-based models (see above). For the fMRI voxel-based analysis, we also tested one whole-brain model plus 58 ROI-based models (see above). For the white matter voxel-based analysis, we tested one whole-brain model plus 10 ROI-based models (see above). For the segmented region-based analysis we tested one model based on whole-brain grey and white matter volume measures (SIENAX), one model based on cortical thickness measures (including measures of 70 cortical regions), and three models based on volumetric measures from subcortical regions (one including 6 measures from the basal ganglia, one including 5 measures from non-basal ganglia subcortical structures, and one including all 11 subcortical structures).

Model assessment

For each model we derived a training set and a test set using a 4-fold balanced cross-validation strategy. For each model, the data were randomly split into four approximately equal groups, ensuring that each group constituted an equivalent representation of the overall data distribution for the variable being predicted (e.g. proportion of pre-HD and controls was preserved across groups). Because results using this method can vary depending on how the data are split, we performed 100 random data splits and ran the 4-fold cross-validation on each split. The SVM learning algorithms were implemented in Matlab (2008b, The MathWorks, Natick, MA) using the libsvm distribution of algorithms available online (http://www.csie.ntu.edu.tw/ cjlin/libsvm). C-SVM was used for classification and ε-SVR was used for regression. Linear kernels and default model parameters ($C=1$, $\varepsilon=0.1$) were used in all of the following analyses. Neuroimaging data were always normalized to standard (MNI) space prior to being input to the algorithm.

Results of the classification models are described in terms of balanced accuracy values. The balanced accuracy measure takes into account both the sensitivity and specificity of the models, and is described by the equation:

$$\begin{aligned} \textit{Balanced accuracy} &= \frac{(TP/(TP+FN)) + (TN/(TN+FP))}{2} \\ &= \frac{\textit{Sensitivity} + \textit{Specificity}}{2} \end{aligned}$$

where TP, TN, FP, and FN are the number of true positives, true negatives, false positives and false negatives respectively. There are 100 balanced accuracy values for each model set, and we report the mean balanced accuracy value as well as the standard deviation of the model set. For the regression models, results are described in terms of the mean and standard deviation of the Pearson's correlation coefficient between the model-predicted YTO and the age/CAG-estimated YTO for each model set.

p-values for both classification and regression models were determined via permutation analysis. For each model set, 500 new models were created using a random permutation of the class labels (or YTO values), such that the imaging data was dissociated from its corresponding class label (or YTO value). Because there should be no systematic relation between labels and features in these datasets, this method therefore allows estimation of chance levels of accuracy (either balanced accuracy or correlation coefficient value). We verified that the mean balanced accuracies were near 50% and the mean correlation coefficients were near 0 in these analyses using permuted labels; thus, there was no systematic bias in the models. The *p*-values we report were calculated as the proportion of accuracy values in the null distribution that were equal to (or greater than) the mean accuracy of the true model set. Because of the large number of non-independent tests performed, we computed the false discovery rate (FDR) for each of the resulting significance tests to correct for multiple comparisons across the full set of classification and regression models. All results with FDR ≤ 0.1 are reported here.

Results

Classifying pre-HD vs. controls

The classification models successfully discriminated pre-HD individuals from controls using voxel-based GM, WM, and fMRI data as well as the segmented region-based morphometric data. Several models achieved classification accuracies of 62% or higher. Many of these models were significant with $p<0.05$ uncorrected, however a few remained significant following FDR-correction for multiple comparisons. We report all models that remain significant (FDR ≤ 0.1) and also display results of other select models for comparative purposes.

Voxel-based grey matter (GM)

There was significantly above-chance accuracy in classifying pre-HD vs. controls in 13 of the 59 model sets assessed using this GM data (feature sets from the 55 individual ROIs in the Harvard–Oxford atlas, 3 combined-ROI feature sets, and 1 whole-brain feature set). Of these 13 models, 6 remained significant following FDR-correction for multiple comparisons. We analyzed GM images with a range of smoothing (0, 2, 3, and 4 mm FWHM Gaussian kernel). The unsmoothed whole-brain GM image achieved 62% accurate classification, while the 2, 3, and 4 mm smoothed images achieved 67%, 66%, and 65% respectively. Thus, subsequent GM analyses used the 2 mm smoothed data. Table 2 describes all GM ROIs that provided significantly above-chance accuracy. Among all defined ROIs in the Harvard–Oxford anatomical atlas, the single ROI that provided the highest accuracy was the putamen at 70%. Combining all basal ganglia ROIs (putamen, caudate, and pallidum) provided 73% accuracy.

Voxel-based white matter (WM)

Of the 11 model sets created using the WM data (feature sets from the 10 individual tracts in the Johns Hopkins University atlas, and 1 whole-brain feature set), 5 produced significantly above-chance accuracy in classifying pre-HD vs. controls. Of these 5models, 1 remained significant following FDR-correction. The single voxel-based WM model that survived FDR-correction was the model based on data from the cingulum (cingulate gyrus portion) ROI with 71% accurate classification. The whole-brain WM dataset produced a model set averaging 65% classification accuracy, but failed to reach significance following FDR-correction (FDR = 0.16). Table 2 describes all WM models that produced near significant classification accuracy.

Voxel-based fMRI

There was significantly above-chance accuracy in classifying pre-HD vs. controls in 3 of the 59 model sets assessed using the fMRI data (feature sets from the 55 individual ROIs in the Harvard–Oxford atlas, 3 combined-ROI feature sets, and 1 whole-brain feature set). None of these models survived FDR-correction. However, one model that remained near significance following FDR-correction was the model based on data from the combined basal ganglia ROIs (putamen, caudate, and pallidum), achieving 64% accurate classification (FDR = 0.16).

Segmented region-based measures

The normalized volume measures of total grey and white matter achieved 62% accurate classification using only 2 features and the normalized subcortical volume measures of basal ganglia regions (left and right caudate, putamen, and pallidum) achieved 73% accurate classification using only 6 features. However, only the model based on basal ganglia subcortical volume measures remained significant following FDR-correction (FDR <0.05). Recognizing that, with such a small set of features, it may be possible to obtain similar results using a much simpler classification scheme, we used the same 4-fold cross-

Table 2
Model assessment measures for each of the successful classification model sets. The GM dataset is a VBM image of grey matter distribution derived from T1 structural scans. The WM dataset is an image of FA values within skeletonized white matter tracts derived from DTI scans. The fMRI dataset is an image derived from BOLD signal contrast of task minus baseline. Whole-brain indicates entire segmented grey matter image for GM, FA image of all skeletonized tracts for WM, and entire contrast image for fMRI. Various brain regions listed come from ROIs determined via atlases (see Methods). Segmented region-based data are derived measures of volume from whole-brain and subcortical structures (caudate, putamen, and pallidum), as well as measures of mean cortical thickness for 35 segmented cortical areas. *Refers to models created using simple linear discriminant analysis. Note: with 500 permutations of the labels, the minimum p-value we can report is $p<0.002$. FDR-corrected values appear in parenthesis next to reported p-values.

	Balanced accuracy		# Features	p-value (FDR)
Voxel-based data	Mean	SD		
Grey matter dataset				
Basal ganglia and frontal cortex	73.78	3.83	34,810	<0.002 (<0.05)
Basal ganglia	72.86	4.18	3382	<0.002 (<0.05)
Putamen	69.83	4.08	1721	0.006 (0.10)
Heschl's gyrus	68.7	5.15	604	0.004 (0.09)
Pallidum	68.36	3.82	578	0.006 (0.10)
Postcentral gyrus	67.1	4.66	6895	0.008 (0.11)
Whole-brain (smoothed with 2 mm Gaussian kernel)	67.09	4.8	196,147	0.008 (0.11)
Whole-brain (unsmoothed)	62.55	3.42	196,147	0.002 (0.05)
White matter dataset				
Cingulum (cingulate gyrus portion)	71.17	4.11	2877	<0.002 (<0.05)
Corticospinal tract	66.27	5.07	10,739	0.016 (0.16)
Whole-brain	65	3.13	99,100	0.018 (0.16)
fMRI dataset (z-statistic image)				
Basal ganglia	63.92	4.82	3382	0.028 (0.16)
Segmented region-based data				
Whole-brain volume (grey matter, white matter)	61.77	4.68	2	0.01 (0.12)
Whole-brain volume (grey matter, white matter)*	67.00	2.64	2	0.012 (0.13)
Subcortical volume (basal ganglia)	73.42	3.14	6	<0.002 (<0.05)
Subcortical volume (basal ganglia)*	76.16	3.37	6	<0.002 (<0.05)

validation scheme and applied simple linear discriminant analysis (LDA) (as implemented in the MATLAB Statistics Toolbox) to these segmented region-based measures in order to confirm results found using SVMs. LDA results were also significant for the model based on basal ganglia subcortical volume measures, slightly surpassing the comparable SVM model with 76% accuracy. Thus, the LDA model created using just 6 measures of normalized volume from the basal ganglia regions produced the highest accuracy across all models tested.

Further inspection of misclassified subjects

In order to determine if there was any pattern to the misclassifications, we further examined the subject data. Subjects were defined as being 'frequently misclassified' if they were misclassified >20% of the time by a given model set, and 'frequently correctly classified' otherwise. There were no significant differences seen between frequently correctly classified and frequently misclassified control subjects. However, upon examination of the frequently misclassified pre-HD subjects, we found that they have significantly greater YTO according to simple T-tests for both whole-brain grey matter and white matter model sets ($p<0.001$). Combining results from the whole-brain grey and white matter model sets, the average (mean ± SD) YTO-L of frequently misclassified pre-HD subjects was 21.2 ± 9.2 as compared to 12.0 ± 4.8 for frequently correctly classified subjects; and the average YTO-A of frequently misclassified pre-HD subjects was 13.1 ± 7.2 as compared to 3.1 ± 5.5 for frequently correctly classified subjects. Additionally, logistic regression models can predict if a subject will be frequently misclassified by whole-brain grey matter and white matter model sets based on their YTO measures ($p<0.01$). These results confirm that, as expected, it is more difficult for the classification models to distinguish between control subjects and pre-HD subjects with many estimated years to disease onset, as opposed to pre-HDs with comparatively fewer years to onset.

Regression models of pre-HD to determine years to onset

The regression models successfully predicted YTO measures using voxel-based GM, WM, and fMRI data as well as the segmented region-based morphometric data. Many of the successful regression models were able to effectively predict both YTO-L and YTO-A measures, but some worked well only on one or the other (see Table 3). Several models produced model-predicted YTO values that were significantly correlated with the age/CAG-estimated YTO values ($r=0.36$ or higher). Again, while many of these models were significant with $p<0.05$ uncorrected, a few remained significant following FDR-correction for multiple comparisons. We report all models that remain significant (FDR ≤ 0.1) and also display results of other select models for comparative purposes (Figs. 1 and 2).

Table 3
Model assessment measures for each of the successful regression model sets. The labels are the same as Table 2. *Refers to models created using simple linear regression analysis. †Refers to measures for which both YTO-A and YTO-L produce significantly successful regression models. Results are displayed only for the measure producing the best results. Note: with 500 permutations, the minimum p-value we can report is $p<0.002$. FDR-corrected values appear in parenthesis next to reported p-values.

	Correlation coefficient		# Features	p-value (FDR)	YTO measure
Voxel-based data	Mean	SD			
Grey matter dataset					
Caudate	0.6591	0.0531	1083	<0.002 (<0.02)	YTO-A†
Amygdala	0.586	0.0776	677	<0.002 (<0.02)	YTO-L
Whole-brain	0.5829	0.0383	196,147	<0.002 (<0.02)	YTO-A†
Insular cortex	0.5369	0.0951	2341	0.006 (0.06)	YTO-L
Supramarginal gyrus, anterior division	0.5271	0.0761	1742	0.004 (0.04)	YTO-L†
Paracingulate gyrus	0.477	0.0758	2944	0.004 (0.04)	YTO-A
Middle frontal gyrus	0.4558	0.0815	5316	0.006 (0.06)	YTO-A†
White matter dataset					
Corticospinal tract	0.6613	0.0627	10,739	<0.002 (<0.02)	YTO-L†
Forceps minor	0.622	0.055	19,407	<0.002 (<0.02)	YTO-A
Uncinate fasciculus	0.5407	0.0603	1985	0.002 (0.02)	YTO-L
Anterior thalamic radiation	0.4961	0.0672	16,720	0.004 (0.04)	YTO-A†
Whole-brain	0.4946	0.0481	99,100	0.002 (0.02)	YTO-A†
Inferior fronto-occipital fasciculus	0.4455	0.0794	12,751	0.008 (0.08)	YTO-L†
fMRI dataset (z-statistic image)					
Insular cortex	0.6847	0.0562	2341	<0.002 (<0.02)	YTO-L
Frontal pole	0.4967	0.0671	15,397	0.002 (0.02)	YTO-L
Caudate	0.4276	0.0771	1083	0.012 (0.15)	YTO-L
Whole-brain	0.3603	0.0684	356,102	0.024 (0.13)	YTO-L
Segmented region-based data					
Whole-brain volumes (grey and white matter)	0.5486	0.0451	2	0.002 (0.02)	YTO-A†
Whole-brain volumes (grey and white matter)*	0.5493	0.0455	2	<0.002 (<0.02)	YTO-A†
Cortical thickness	0.4512	0.0616	70	0.022 (0.13)	YTO-L
Subcortical volumes	0.574	0.0492	11	<0.002 (<0.02)	YTO-A
Subcortical volumes (basal ganglia)	0.5579	0.045	6	0.002 (0.02)	YTO-A
Subcortical volumes (basal ganglia)*	0.4456	0.1048	6	0.018 (0.18)	YTO-A

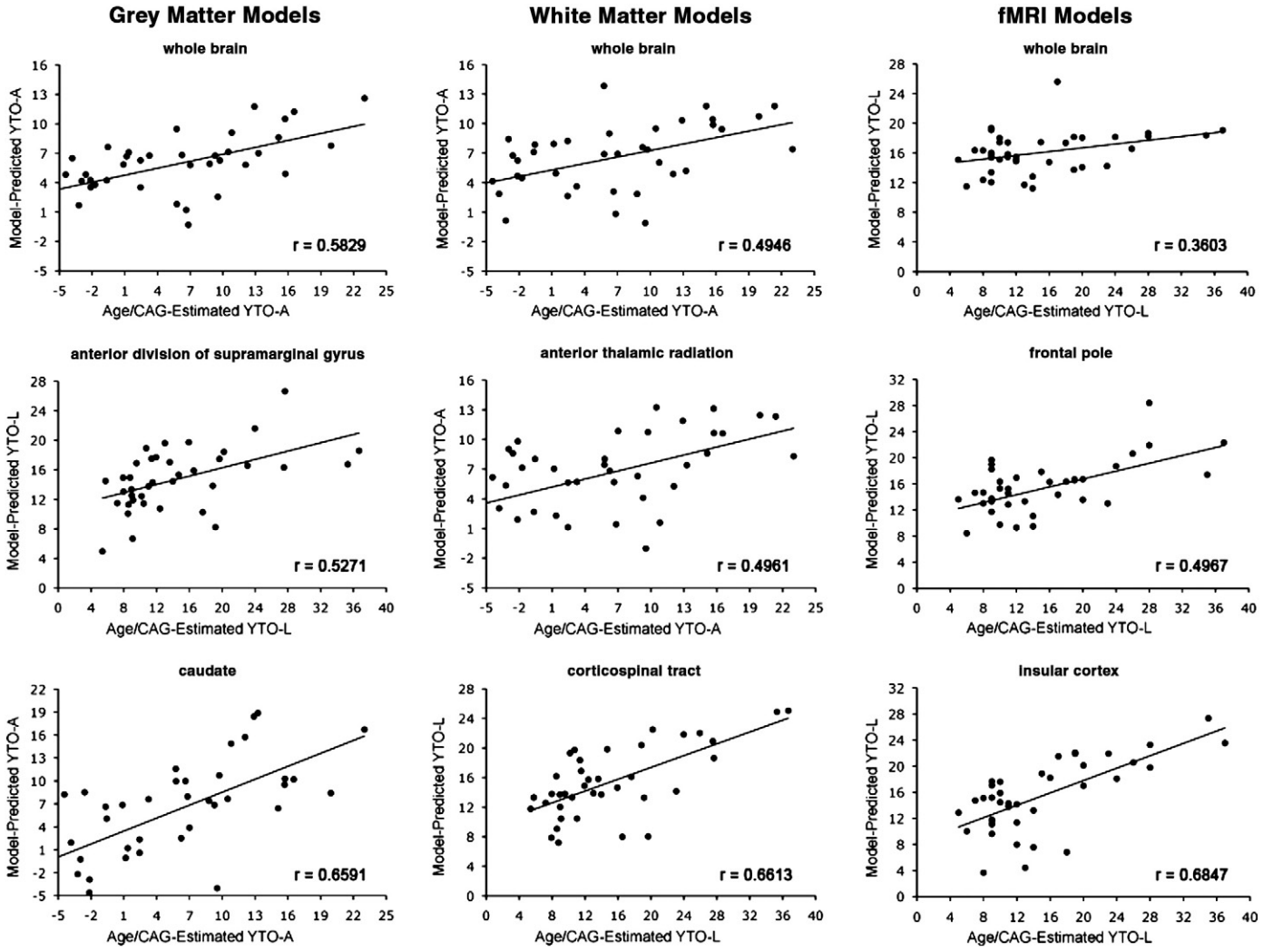

Fig. 1. Scatter plots of model-predicted vs. age/CAG-calculated YTO estimates from models produced using various subsets of voxel-based imaging data. Labels are the same as described in Table 2.

Voxel-based grey matter (GM)

Of the 59 GM model sets assessed, 21 produced model-predicted YTO values that were significantly correlated with the age/CAG-estimated YTO values (either YTO-A, YTO-L, or both). Of these 21 models, 7 remained significant following FDR-correction for multiple comparisons. The model set using the whole-brain GM dataset produced model-predicted YTO values that were significantly correlated with the age/CAG-estimated YTO-A values ($r = 0.5829$, FDR < 0.02); as did the model set using only voxels within the caudate ROI of the GM data ($r = 0.6591$, FDR < 0.02). Other GM ROI feature sets also produced regression models with significant correlations of predicted and age/CAG-estimated YTO measures (Table 3).

Voxel-based white matter (WM)

Of the 11 WM model sets assessed, 9 produced model-predicted YTO values that were significantly correlated with the age/CAG-estimated YTO values (either YTO-A, YTO-L, or both) and 6 remained significant following FDR-correction for multiple comparisons. The model set using the whole-brain WM dataset produced predictions that were significantly correlated with age/CAG-estimated YTO-A measures ($r = 0.4961$, FDR $= 0.02$), and several of the WM ROI feature sets were also capable of producing successful regression models. Some of the WM ROI regression models did particularly well, including the corticospinal tract ($r = 0.6613$, FDR < 0.02), uncinate fasciculus ($r = 0.5407$, FDR $= 0.02$), and forceps minor ($r = 0.622$, FDR < 0.02) (Table 3).

Voxel-based functional MRI (fMRI)

Of the 59 fMRI model sets assessed, 6 produced model-predicted YTO values that were significantly correlated with the age/CAG-estimated YTO-L values and 2 remained significant following FDR-correction for multiple comparisons. The model set created using the whole-brain fMRI *z*-statistic dataset produced predictions that were moderately correlated with age/CAG-estimated YTO-L measures ($r = 0.3603$), but the significance level did not survive FDR-correction (FDR $= 0.13$). A few of the fMRI ROI feature sets were able to do much better. The best performing fMRI regression model set came from the insular cortex ROI feature set ($r = 0.6847$, FDR < 0.02). The other fMRI ROI feature set that produced regression models with significant correlations of predicted and age/CAG-estimated YTO-L measures was the frontal pole ($r = 0.4967$, FDR $= 0.02$). The fMRI caudate ROI feature set produced moderately correlated predictions as well ($r = 0.4276$), but again the significance level did not survive FDR-correction (FDR $= 0.15$).

Segmented region-based measures

These models were also successful in predicting YTO measures. The predictions generated from model sets created using whole-brain volumes ($r = 0.5486$, FDR $= 0.02$) and subcortical volumes ($r = 0.574$, FDR < 0.02) both were significantly correlated with age/CAG-estimated YTO measures. The full set of subcortical volume measures included 6 basal ganglia volumes, as well as volumes of 5 other subcortical

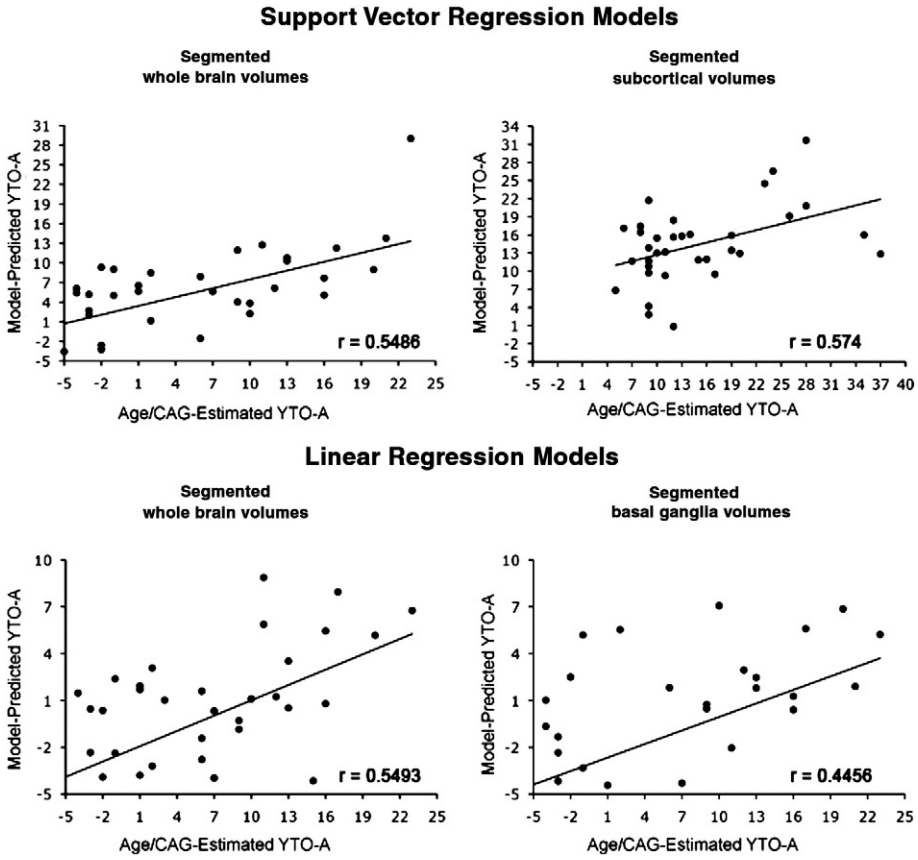

Fig. 2. Scatter plots of model-predicted vs. age/CAG-calculated YTO estimates from models produced using various subsets of segmented region-based imaging data. These data are derived measures of volume from whole-brain (grey and white matter) and subcortical structures (caudate, putamen, pallidum, accumbens area, thalamus, hippocampus, and amygdala), as well as measures of mean cortical thickness for 35 segmented cortical areas.

structures. It was determined that models including only the basal ganglia volumes produced successful SVM regression models ($r = 0.5579$, FDR = 0.02) while models excluding the basal ganglia produced moderate correlations ($r = 0.3595$) whose significance level did not survive FDR-correction (FDR = 0.16). In addition, the cortical thickness measures also produced moderate correlations ($r = 0.4512$) whose significance level did not survive FDR-correction (FDR = 0.13).

Again recognizing that, with such a small set of features, it may be possible to obtain similar results using a much simpler regression procedure, we used the same 4-fold cross-validation scheme and applied simple multiple linear regression analysis procedures to these segmented region-based measures in order to confirm results found using SVMs. The linear regression-based models created using the two normalized volume measures (grey matter and white matter volume) produced predictions that were significantly correlated with age/CAG-estimated YTO-A ($r = 0.5493$, FDR < 0.02). However, while the full set of subcortical volume measures produced successful SVM regression-based models (as described above), they did not produce successful linear regression-based models; and when the full set of subcortical measures was separated, the linear regression-based models created using only the 6 basal ganglia volume measures produced moderately correlated predictions ($r = 0.4456$) whose significance level did not survive FDR-correction (FDR = 0.18).

Discussion

There are many potential neuroprotective strategies now available for neurodegenerative disorders. Evaluating these requires measuring the degree of disease progression in patients well before any clinical symptoms emerge, since by then, a substantial amount of brain tissue may already be irreparably lost. Thus, it is critical to identify biomarkers that can identify and quantify the disease process in-vivo, especially at the pre-symptomatic stage. HD has become a test-bed for such biomarker development because individuals can be studied who are gene-positive yet still asymptomatic. Much recent work has shown that MRI is capable of identifying differences between pre-HD and control individuals (reviewed by Bohanna et al., 2008). Other studies have also shown that some MRI measures, such as caudate and putamen GM volume, are sensitive to longitudinal change in pre-HD individuals (Aylward et al., 2004). Many investigators are now examining other kinds of MRI measures in longitudinal studies to determine which may be most sensitive to the disease progress, and how to power potential neuroprotective studies (Hersch and Rosas, 2008). Most of these studies involve conventional cross-sectional and longitudinal analyses of MRI data (but see Klöppel et al. (2008, 2009) for exception). Here we highlight the usefulness of machine learning approaches. Using classification models, we show that multiple types of MRI data contain information that discriminates pre-HD from controls. Using regression models, we show that multiple types of MRI data contain information about the level of disease progression in pre-HD individuals. These findings strongly support the utility of machine learning approaches for biomarker development and evaluation.

Classification models can accurately discriminate pre-HD from controls

We set up classification models for a wide range of MRI data. This included: voxel-based whole-brain GM and WM; voxel-based GM and WM ROIs; segmented region-based whole-brain and subcortical volumes; segmented region-based cortical thickness; fMRI whole-brain *z*-statistic images; and fMRI *z*-statistic image ROIs. For each model we derived a training-set and a test-set, using a 4-fold balanced cross-validation approach. We then evaluated the models accuracy by

comparing the classification labels of individuals in the test-set against the true labels.

For the T1 segmented grey matter images, we replicated the findings of Klöppel et al. (2009) by showing that we could correctly discriminate pre-HD individuals from controls with up to 76% accuracy (versus 56% for the entire sample and 69% for the near-conversion subjects in the 2009 Klöppel et al. study). Notably, we were able to replicate this result with a much smaller dataset (64 subjects vs.191 subjects). However it is important to note that our data were taken from a single imaging center and therefore did not suffer the disadvantage of additional variability inherent in data collected from multiple centers. For this reason we were able to preserve more of the fine-grained spatial information by using less smoothing (2 mm Gaussian kernel vs. 8 mm kernel).

For the FA white matter images, we also replicated the results of Klöppel et al. (2008), showing that we could correctly discriminate pre-HD individuals from controls with up to 71% accuracy (versus 82% in Klöppel et al. (2008)). We find no difference in the distribution of subject demographic information between the current study and Klöppel et al. (2008), although there are a greater number of pre-HD subjects in the current study (39 vs. 25 in Klöppel et al. (2008)). The discrepancy in the level of classification accuracy may be due to specific image processing parameters, or could perhaps be due to the fact that the FA datasets in Klöppel et al. (2008) included much more brain volume than the skeletonized FA dataset in the present study.

Additionally, we found that other forms of MRI data also carried information that could discriminate groups (see Table 2). We created models able to successfully differentiate pre-HD from control participants using BOLD data from fMRI, and also using a handful of volumetric measures from segmented T1-weighted images.

Some of the best performing models were created using grey matter feature sets from the fronto-striatal ROIs. This provides further evidence indicating the presence of structural grey matter changes in the basal ganglia of pre-HD individuals (Aylward et al., 1996, 2000, 2004). However, it is notable that a number of grey matter ROIs outside of the fronto-striatal network also produced models able to discriminate pre-HD from control, including regions within the temporal and parietal lobes (Gómez-Ansón et al., 2009; Kipps et al., 2005; Paulsen et al., 2006b; Rosas et al., 2005; Thieben et al., 2002).

Other models that performed very well were produced using white matter ROI feature sets. Fractional anisotropy data from the cingulum (cingulate gyrus portion) seemed to be the most informative white matter tract, as it produced the most accurate classification model among all white matter feature sets.

Further examination of misclassified subjects suggests that the classification models are sensitive to increasing disease-state risk, because while no significant differences are seen between frequently correctly classified and frequently misclassified control subjects, the frequently misclassified pre-HD subjects have significantly greater estimated years to disease onset as compared to frequently correctly classified pre-HD subjects. These results indicate that it is more difficult for the classification models to distinguish between control subjects and pre-HD subjects with many estimated years to disease onset, an issue supported by the results of Klöppel et al. (2009) in which satisfactory classification accuracy could only be achieved using a subset of subjects that were determined to be near disease conversion. While it was still possible in the present study to achieve significantly successful classification using the entire sample of pre-HD subjects, these results may suggest that there are limitations on the ability of classification models to identify pre-HD subjects who have >22 YTO (according to Langbehn method) or >14 YTO (according to Aylward method).

Regression models can accurately predict years to clinical onset

An important and novel aspect of our results relates to the use of regression models to predict quantitative disease progression. We did this by training the regression model on a set of the pre-HD MRI data along with the associated age/CAG-estimated YTO measures (which are calculated based on CAG length and age), and then deriving the model-predicted YTO in the test-set. We did this for all the same MRI data as for the classification models (see above) using the same 4-fold cross-validation procedure. Comparisons of model-predicted and age–CAG-estimated YTO, using Pearson's correlation (Figs. 1 and 2), showed that several forms of MRI data were highly accurate (with some r values exceeding 0.65).

We note that age/CAG-estimated YTO is itself an estimate with its own variability. It is based on age and CAG repeat using either the method of Langbehn et al. (2004) or Aylward et al. (1996). These different methods give somewhat different YTO estimates. The true gold standard for each pre-HD individual in our study would be the age at which he/she meets diagnostic criteria for movement disorder, something that is very far in the future for some individuals. Thus the comparison of model-estimated YTO and age/CAG-estimated YTO is simply a proxy for establishing whether multivariate information in the images relates to level of disease progression. Our findings are that the regression models performed very well at detecting disease progression. The best performing models included those based on the whole-brain GM dataset, the GM caudate, the fMRI insular cortex, and the WM corticospinal tract.

Model complexity and performance

Among all of the sophisticated models and enormous datasets employed in these various classification schemes, the best performing model was one of the simplest. Using 6 measures of normalized basal ganglia volumes (left and right caudate, putamen, and pallidum) and LDA, we were able to produce, on average, 76% accurate classification of pre-HD vs. control subjects. Because of the central role of the basal ganglia in HD, this is hardly surprising, but it is interesting that adding additional data only reduced the degree of classification accuracy. For regression, there seemed to be a greater benefit to using more complex models, though again the most predictive models were those using relatively smaller numbers of features (voxels). The strong performance of ROI-based analyses in comparison to whole-brain analyses suggests that this may be an optimal approach for the development of biomarkers.

Because our analyses focused almost exclusively on the use of Support Vector algorithms, it is an open question as to whether other high-dimensional classifier/regression methods might perform better with the present data. For example, there is great interest in methods using L1-based regularization (Johnstone and Titterington, 2009; Ravikumar et al., 2009) that can provide sparse solutions to classification and regression problems. It is possible that such methods could more efficiently detect the relatively few features that turn out to provide the best prediction. Further analysis is needed to examine this question.

Future directions

Our current contribution is to show that the pattern analysis approach could be useful in pre-HD with many types of imaging data. It will be very important to validate the pattern analysis approach with longitudinal follow-up — something we are currently doing. For example, with a longitudinal dataset it will be possible to train regression models using MRI-based data from multiple timepoints and then, in a test-set, attempt to make a quantitative prediction regarding an individual's change in disease progression based on their MRI data from timepoint 1 alone. The accuracy of this approach could be compared against the actual change in brain data. If machine learning techniques could be developed that are highly accurate in this longitudinal projection aspect, then this could be useful for neurotherapeutics. Based on timepoint 1 with no treatment, and change in data from later timepoints following treatment, it would be possible to

evaluate the degree of neuroprotection rendered by the treatment. Classifiers could be useful in using timepoint 1 data alone to predict change across time, and the likely effects of neuroprotection. Multivariate models such as those used here could prove to be more sensitive to complex distributed changes than conventional approaches when evaluating the effect of a given treatment. These models would also be capable of detecting potentially informative changes in data from multiple imaging modalities within a single model.

Importantly, in order for these tools to be used in large-scale multicenter clinical trials, they must show reliability, ease of use, and manageable operator time. With this in mind, all of the models in the present study were fit to MRI data which we analyzed automatically using robust and reliable tools (FSL and Freesurfer). Future refinement of the regression models, in terms of dimension reduction and parameter optimization, will afford users a relatively fast and easy interface via development of customized software tools.

Summary

These results clearly demonstrate the utility of machine learning for evaluating MRI-based biomarkers for neurodegeneration in pre-HD. We have verified that several different measures derived from both structural and functional MRI data can be used to successfully classify between pre-HD and controls. We have also shown that such MRI measures can be used to create regression models that are able to accurately predict an established quantitative measure of disease progression.

Acknowledgments

We gratefully acknowledge CHDI (A. Aron, PI) and the Office of Naval Research (Grant # N00014-07-1-0116, R. Poldrack, PI) for the financial support. Additional thanks to Yaroslav Halchenko for the helpful comments on this manuscript.

References

Aron, A.R., Shohamy, D., Clark, J., Myers, C., Gluck, M.A., Poldrack, R.A., 2004. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. J. Neurophysiol. 92 (2), 1144–1152.

Aron, A.R., Gluck, M.A., Poldrack, R.A., 2006. Long-term test–retest reliability of functional MRI in a classification learning task. Neuroimage 29 (3), 1000–1006.

Ashburner, J., Friston, K., 2000. Voxel-based morphometry—the methods. Neuroimage 11 (6), 805–821.

Aylward, E., Codori, A., Barta, P., Pearlson, G., Harris, G.J., Brandt, J., 1996. Basal ganglia volume and proximity to onset in presymptomatic Huntington disease. Arch. Neurol. 53 (12), 1293–1296.

Aylward, E.H., Codori, A.M., Rosenblatt, A., Sherr, M., Brandt, J., Stine, O.C., Barta, P.E., Pearlson, G.D., Ross, C.A., 2000. Rate of caudate atrophy in presymptomatic and symptomatic stages of Huntington's disease. Mov. Disord. 15 (3), 552–560.

Aylward, E.H., Sparks, B.F., Field, K.M., Yallapragada, V., Shpritz, B.D., Rosenblatt, A., Brandt, J., Gourley, L.M., Liang, K., Zhou, H., Margolis, R.L., Ross, C.A., 2004. Onset and rate of striatal atrophy in preclinical Huntington disease. Neurology 63 (1), 66–72.

Beglinger, L.J., Nopoulos, P.C., Jorge, R.E., Langbehn, D.R., Mikos, A.E., Moser, D.J., Duff, K., Robinson, R.G., Paulsen, J.S., 2005. White matter volume and cognitive dysfunction in early Huntington's disease. Cogn. Behav. Neurol. 18 (2), 102–107.

Beglinger, L., Paulsen, J., Watson, D., Wang, C., 2008. Obsessive and compulsive symptoms in prediagnosed Huntington's disease. J. Clin. Psychiatry 69 (11), 1758–1765.

Bohanna, I., Georgiou-Karistianis, N., Hannan, A.J., Egan, G.F., 2008. Magnetic resonance imaging as an approach towards identifying neuropathological biomarkers for Huntington's disease. Brain Res. Rev. 58 (1), 209–225.

Duff, K., Paulsen, J.S., Beglinger, L.J., Langbehn, D.R., Stout, J.C., of the Huntington Study Group, P.-H. I, 2007. Psychiatric symptoms in Huntington's disease before diagnosis: the predict-HD study. Biol. Psychiatry 62 (12), 1341–1346.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Georgiou-Karistianis, N., 2009. A peek inside the Huntington's brain: will functional imaging take us one step closer in solving the puzzle? Exp. Neurol. 220 (1), 5–8.

Gómez-Ansón, B., Alegret, M., Muñoz, E., Monté, G.C., Alayrach, E., Sánchez, A., Boada, M., Tolosa, E., 2009. Prefrontal cortex volume reduction on MRI in preclinical Huntington's disease relates to visuomotor performance and CAG number. Parkinsonism Relat. Disord. 15 (3), 213–219.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (3), 389–422.

Hersch, S.M., Rosas, H.D., 2008. Neuroprotection for Huntington's disease: ready, set, slow. Neurotherapeutics 5 (2), 226–236.

Johnstone, I.M., Titterington, D.M., 2009. Statistical challenges of high-dimensional data. Philos. Trans. A Math. Phys. Eng. Sci. 367 (1906), 4237–4253.

Kipps, C.M., Duggins, A.J., Mahant, N., Gomes, L., Ashburner, J., McCusker, E.A., 2005. Progression of structural neuropathology in preclinical Huntington's disease: a tensor based morphometry study. J. Neurol. Neurosurg. Psychiatry 76 (5), 650–655.

Klöppel, S., Draganski, B., Golding, C.V., Chu, C., Nagy, Z., Cook, P.A., Hicks, S.L., Kennard, C., Alexander, D.C., Parker, G.J.M., Tabrizi, S.J., Frackowiak, R.S.J., 2008. White matter connections reflect changes in voluntary-guided saccades in pre-symptomatic Huntington's disease. Brain 131 (Pt 1), 196–204.

Klöppel, S., Chu, C., Tan, G.C., Draganski, B., Johnson, H., Paulsen, J.S., Kienzle, W., Tabrizi, S.J., Ashburner, J., Frackowiak, R.S.J., of the Huntington Study Group, P.-H. I, 2009. Automatic detection of preclinical neurodegeneration: presymptomatic Huntington disease. Neurology 72 (5), 426–431.

Knowlton, B.J., Mangels, J.A., Squire, L.R., 1996. A neostriatal habit learning system in humans. Science 273 (5280), 1399–1402.

Langbehn, D., Brinkman, R., Falush, D., 2004. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin. Genet. 65 (4), 267–277.

Marshall, J., White, K., Weaver, M., Wetherill, L.F., Hui, S., Stout, J.C., Johnson, S.A., Beristain, X., Gray, J., Wojcieszek, J., Foroud, T., 2007. Specific psychiatric manifestations among preclinical Huntington disease mutation carriers. Arch. Neurol. 64 (1), 116–121.

Paulsen, J.S., 2009. Functional imaging in Huntington's disease. Exp. Neurol. 216 (2), 272–277.

Paulsen, J., Zimbelman, J., Hinton, S., 2004. fMRI biomarker of early neuronal dysfunction in presymptomatic Huntington's disease. Am. J. Neuroradiol. 25 (10), 1715–1721.

Paulsen, J.S., Hayden, M., Stout, J.C., Langbehn, D.R., Aylward, E., Ross, C.A., Guttman, M., Nance, M., Kieburtz, K., Oakes, D., Shoulson, I., Kayson, E., Johnson, S., Penziner, E., of the Huntington Study Group, P.-H. I, 2006a. Preparing for preventive clinical trials: the predict-HD study. Arch. Neurol. 63 (6), 883–890.

Paulsen, J.S., Magnotta, V.A., Mikos, A.E., Paulson, H.L., Penziner, E., Andreasen, N.C., Nopoulos, P.C., 2006b. Brain structure in preclinical Huntington's disease. Biol. Psychiatry 59 (1), 57–63.

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45 (1 Suppl), S199–S209.

Ravikumar, P., Lafferty, J., Liu, H., Wasserman, L., 2009. Sparse additive models. J. R. Stat. Soc. 71 (5), 1009.

Reading, S.A.J., Dziorny, A.C., Peroutka, L.A., Schreiber, M., Gourley, L.M., Yallapragada, V., Rosenblatt, A., Margolis, R.L., Pekar, J.J., Pearlson, G.D., Aylward, E., Brandt, J., Bassett, S.S., Ross, C.A., 2004. Functional brain changes in presymptomatic Huntington's disease. Ann. Neurol. 55 (6), 879–883.

Reading, S.A.J., Yassa, M.A., Bakker, A., Dziorny, A.C., Gourley, L.M., Yallapragada, V., Rosenblatt, A., Margolis, R.L., Aylward, E.H., Brandt, J., Mori, S., van Zijl, P., Bassett, S.S., Ross, C.A., 2005. Regional white matter change in pre-symptomatic Huntington's disease: a diffusion tensor imaging study. Psychiatry Res. 140 (1), 55–62.

Rosas, H.D., Hevelone, N.D., Zaleta, A.K., Greve, D.N., Salat, D.H., Fischl, B., 2005. Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. Neurology 65 (5), 745–747.

Rosas, H.D., Tuch, D.S., Hevelone, N.D., Zaleta, A.K., Vangel, M., Hersch, S.M., Salat, D.H., 2006. Diffusion tensor imaging in presymptomatic and early Huntington's disease: Selective white matter pathology and its relationship to clinical measures. Mov. Disord. 21 (9), 1317–1325.

Schölkopf, B., Smola, A., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Shohamy, D., Myers, C.E., Grossman, S., Sage, J., Gluck, M.A., Poldrack, R.A., 2004. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 127 (4), 851–859.

Smith, S.M., Stefano, N.D., Jenkinson, M., Matthews, P.M., 2001. Normalized accurate measurement of longitudinal brain change. J. Comput. Assist. Tomogr. 25 (3), 466–475.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., Stefano, N.D., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17 (1), 479–489.

Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31 (4), 1487–1505.

Smola, A., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.

Stoffers, D., Sheldon, S., Kuperman, J., Goldstein, J., Corey-Bloom, J., Aron, A.R., 2010. Contrasting grey- and white matter changes in preclinical Huntington's disease: an MRI study. Neurology 74 (Apr 13) (15), 1208–1216.

Thieben, M.J., Duggins, A.J., Good, C.D., Gomes, L., Mahant, N., Richards, F., McCusker, E., Frackowiak, R.S.J., 2002. The distribution of structural neuropathology in pre-clinical Huntington's disease. Brain 125 (Pt 8), 1815–1828.

Vonsattel, J.P., DiFiglia, M., 1998. Huntington disease. J. Neuropathol. Exp. Neurol. 57 (5), 369–384.

Wolf, R.C., Vasic, N., Schönfeldt-Lecuona, C., Landwehrmeyer, G.B., Ecker, D., 2007. Dorsolateral prefrontal cortex dysfunction in presymptomatic Huntington's disease: evidence from event-related fMRI. Brain 130 (Pt 11), 2845–2857.

Zimbelman, J.L., Paulsen, J.S., Mikos, A., Reynolds, N.C., Hoffmann, R.G., Rao, S.M., 2007. fMRI detection of early neural dysfunction in preclinical Huntington's disease. J. Int. Neuropsychol. Soc. 13 (5), 758–769.