SUPPLEMENTARY METHODS

Neuropsychological data

We selected tests from the battery that probe each of these domains: memory (California Verbal Learning Test (CVLT) – Short Form 10 minute recall, d-prime (a measure of delayed recognition accuracy), and 10 minute recall of the previously copied Benson figure), visuospatial ability (Benson figure copy), language (15-item Boston Naming Test, phrase repetition, verbal agility, semantic (animal) fluency, and the Peabody Picture Vocabulary Test (PPVT)) and executive function (modified Trails B time, Stroop inhibition, "D" word fluency, design fluency, and a composite rule-violation score derived from the sum of errors made on design fluency, Stroop inhibition, and modified Trails B). Current bvFTD diagnostic criteria (Rascovsky *et al.*, 2011) require a determination about the degree of executive function relative to episodic memory and visuospatial abilities. As not all patients received all tests, we analyzed the subset of patients who had available data for the CVLT, Benson Figure test, and 3/4 executive function measures (Stroop, D word fluency, design fluency, and modified Trails).

Image selection and pre-processing

We selected the first UCSF research MRI of sufficient quality for analysis. Pre-processing included segmentation into gray and white matter, alignment and normalization to the standard adult tissue probability maps in Montreal Neurological Institute (MNI) space distributed with SPM12 (http://www.fil.ion.ucl.ac.uk/spm/), modulation and smoothing with an 8 mm full width at half maximum (FWHM) Gaussian kernel.

Neuroimaging – Frequency maps

We first performed voxel-wise regression in SPM12 on a set of nuisance variables (age, sex, total intracranial volume (TIV), handedness, and scanner type) to estimate their effect on smoothed-

modulated-warped gray matter images in a group of 304 healthy controls selected to represent the range of values for the nuisance variables observed in the patient groups (mean age of 67.7 (8.4), 40.5 % male, mean TIV 1.43 (0.14), 87.2% right-handed, with 185 scanned on the 3T MRI, 64 at 1.5T, and 55 at 4T). This approach resulted in beta maps for the nuisance variables as well as a residual map for each control. We then used these values to compute a voxel-wise map of W scores for each patient using the formula: W score = (actual – expected)/SD, where actual is the smoothed-modulated-warped gray matter probability for a given patient at each voxel, expected is the predicated gray matter probability for that voxel using the nuisance factors and beta weights from the regression in healthy controls, and SD is the standard deviation of the residuals for that voxel among the controls.

Neuroimaging – Principal Component Analysis

All principal component analysis and linear discriminant analysis of gray and white matter voxel-wise W maps were performed in Python using nibabel (http://nipy.org/nibabel/), scikit-learn (http://scikit-learn.org/), and nilearn (http://nilearn.github.io/). Analyses were masked to either the standard adult gray or white tissue probability maps from SPM12, thresholded at 10%.

***A priori* classification algorithm**

For purposes of the algorithm, motor neuron disease was defined as present if the clinician evaluating the patient at the time had given a diagnosis of ALS or motor neuron disease or indicated that the individual met El Escorial ALS criteria (Brooks *et al.*, 2000). PSP syndrome was defined as present either if the clinician indicated they met PSP diagnostic criteria (Litvan *et al.*, 1996) or if the examination documented the presence of falls, axial rigidity, and characteristic eye movement abnormalities (either supranuclear gaze palsy or slowing or decreased amplitude of saccadic eye movements). To better capture the situation faced by clinicians seeing new patients we only separated out motor neuron disease or PSP syndromes in the algorithm if these features were present at the first visit, instead of

developing over the course of longitudinal follow-up. Blinded visual reads of MRI scans were performed by two authors (DCP and WWS). Imaging features assessed included knife edge frontotemporal atrophy, anterior temporal lobe-predominant atrophy, disproportionately severe caudate atrophy, whether atrophy is most severe in dorsal or ventral structures, and disproportionate atrophy of posterior cingulate/precuneus. Structures included in rating of a dorsal pattern of atrophy included dorsolateral prefrontal cortex, dorsal anterior insula, frontal operculum, anterior mid-cingulate cortex, pre-supplementary motor area, and supramarginal gyrus. Structures included in rating of a ventral pattern of atrophy included the medial orbitofrontal cortex/subgenual anterior cingulate cortex, ventral anterior insula (i.e. frontoinsula), pregenual anterior cingulate cortex, ventral striatum, amygdala, and temporal pole. In cases of disagreement the raters discussed each scan until reaching consensus.

SUPPLEMENTARY MATERIAL REFERENCES

Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders 2000; 1: 293-9.

Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology 1996; 47: 1-9.

Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 2011.

SUPPLEMENTARY FIGURE LEGEND

Supplementary figure 1 Linear Discriminant Analysis. The top 25% of gray matter voxels that contribute to the linear boundary between subjects with one diagnosis and subjects with 9 other diagnoses. For voxels that are part of the top 25% map for more than one diagnosis, the diagnosis displayed is the one that most strongly differentiates that diagnosis compared to the rest. N = 82. (A) Regions where each diagnosis has less gray matter than the remaining diagnoses by LDA. (B) Regions where each diagnosis has more gray matter than the other diagnoses.

Supplementary figure 2 Principal Component Analysis of white matter W-score maps. The top 25% of voxels contributing to each of the first 10 components from the principal component analysis. With each is a color bar representing the mean score for each component of all included subjects with each of the top 10 diagnoses. N=82.

Supplementary figure 3 Classification procedure involving *a priori* algorithm, simulated amyloid/tau imaging, and discriminant function analysis. At each branch point the simulated results of amyloid/tau imaging lead either to prediction of a single diagnosis, or to a narrow list of potential diagnoses, before selecting the top remaining diagnosis from the discriminant function analysis as the predicted diagnosis. The right side shows the correct and incorrect predictions at each step when applying this classification procedure to the bvFTD cohort.