Schizophrenia Bulletin
doi:10.1093/schbul/sbx137

Multisite Machine Learning Analysis Provides a Robust Structural Imaging Signature of Schizophrenia Detectable Across Diverse Patient Populations and Within Individuals

Martin Rozycki[1,8], Theodore D. Satterthwaite[1,2,8], Nikolaos Koutsouleris[3], Guray Erus[1], Jimit Doshi[1], Daniel H. Wolf[1,2], Yong Fan[1], Raquel E. Gur[2], Ruben C. Gur[2], Eva M. Meisenzahl[4], Chuanjun Zhuo[5], Hong Ying[6], Hao Yan[7], Weihua Yue[7], Dai Zhang[7], and Christos Davatzikos*[,1]

[1]Center for Biomedical Image Computing and Analytics, Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; [2]Brain Behavior Laboratory, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; [3]Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University, Munich, Germany; [4]Department of Psychiatry and Psychotherapy, Heinrich Heine University Dusseldorf, Dusseldorf, Germany; [5]Tianjin Anning Hospital, Tianjin, China; [6]Department of Radiology, Xijing Hospital, Fourth Military Medical University, Xi'an, China; [7]Institute of Mental Health, Key Laboratory of Mental Health, Ministry of Health & National Clinical Research Center for Mental Disorders, Sixth Hospital, Peking University, Beijing, China

[8]These authors contributed equally to the article.

*To whom correspondence should be addressed; University of Pennsylvania, Richards Building, 7th Floor, 3700 Hamilton Walk, Philadelphia, PA 19104; tel: 215-746-4067, fax: 215-746-4060, e-mail: christos@rad.upenn.edu

Past work on relatively small, single-site studies using regional volumetry, and more recently machine learning methods, has shown that widespread structural brain abnormalities are prominent in schizophrenia. However, to be clinically useful, structural imaging biomarkers must integrate high-dimensional data and provide reproducible results across clinical populations and on an individual person basis. Using advanced multi-variate analysis tools and pooled data from case–control imaging studies conducted at 5 sites (941 adult participants, including 440 patients with schizophrenia), a neuroanatomical signature of patients with schizophrenia was found, and its robustness and reproducibility across sites, populations, and scanners, was established for single-patient classification. Analyses were conducted at multiple scales, including regional volumes, voxelwise measures, and complex distributed patterns. Single-subject classification was tested for single-site, pooled-site, and leave-site-out generalizability. Regional and voxelwise analyses revealed a pattern of widespread reduced regional gray matter volume, particularly in the medial prefrontal, temporolimbic and peri-Sylvian cortex, along with ventricular and pallidum enlargement. Multivariate classification using pooled data achieved a cross-validated prediction accuracy of 76% (AUC = 0.84). Critically, the leave-site-out validation of the detected schizophrenia signature showed accuracy/AUC range of 72–77%/0.73–0.91, suggesting a robust generalizability across sites and patient cohorts. Finally, individualized patient classifications displayed significant correlations with clinical measures of negative, but not positive, symptoms. Taken together, these results emphasize the potential for structural neuroimaging data to provide a robust and reproducible imaging signature of schizophrenia. A web-accessible portal is offered to allow the community to obtain individualized classifications of magnetic resonance imaging scans using the methods described herein.

Key words: psychosis/neuroimaging/MRI/brain/biomarker

Introduction

Schizophrenia is an often devastating illness that results in substantial morbidity and mortality worldwide.[1] Structural brain abnormalities were noted by early anatomists,[2,3] and results from in vivo magnetic resonance imaging (MRI) have supported the notion that schizophrenia is a brain disorder.[4,5] Despite such advances, neuroimaging is typically not used as part of standard clinical care of psychotic disorders, in part due to heterogeneous results. Advances in clinical practice require robust imaging biomarkers of disease that can be used for diagnosis, determination of prognosis, and integration within trials of novel therapeutics.

As imaging technology has advanced and costs have declined, both the sample sizes and number of studies examining structural brain abnormalities in schizophrenia have expanded dramatically. Results from small single-site studies have often been heterogeneous, leading investigators to integrate findings across studies using meta-analyses of published data from hundreds of studies and thousands of patients.[6–12] These studies provide increasingly consistent evidence of widespread patterns of structural deficits in schizophrenia, including gray matter loss that is prominent in the insular, cingulate, prefrontal, and temporal cortex.

However, such retrospective, literature-based meta-analyses are inherently limited by methodological variation. Scanner effects, processing choices, and analytic strategy can have a dramatic impact on results of individual studies, adding noise to results. The ENIGMA consortium was created to overcome such obstacles through the use of prospective meta-analyses, where all sites perform the same data processing, quality assurance, and group-level analyses.[13,14] Using this approach, the ENIGMA-schizophrenia working group recently examined subcortical volumetric abnormalities in a sample of over 2000 patients and a similar number of controls, and reported volume reductions of small-moderate effect size (Cohen's *d*: 0.2–0.4). These findings have subsequently been replicated by the COCORO consortium, which used a similar approach.[15] Despite the enormous sample size provided by ENIGMA, such regional analyses ultimately provide a coarse-grained account of structural brain abnormalities associated with schizophrenia, and at present are limited to the subcortex. In contrast, one large recent study directly pooled data in a voxel-based, whole-brain mega-analysis, and delineated widespread loss of gray matter density that was maximal in insular cortex.[16]

Importantly, despite accelerating efforts to pool structural imaging data to study psychosis, such meta- and mega-analyses have focused on mass-univariate approaches applied at the group level. Alternatively, brain-wide multivariate neuroanatomical patterns can be used to classify individuals as cases or controls using machine learning methods.[17–19] By concisely summarizing high-dimensional data, multivariate classification has clear advantages for clinical translation.[20,21] Single-site studies over the past decade have demonstrated that multivariate analyses can accurately discriminate patients and controls, as well as predict progression.[18,19,22–26] One recent meta-analysis of single-site multivariate classification studies reported pooled classification sensitivity of 76%, with 79% specificity.[27] However, no study to date has demonstrated single-subject classification in a large, multi-site, and multi-ethnic setting. In particular, it remains unknown whether classifiers trained on data acquired at one site will perform similarly on data acquired from a new site that was not included in the training set. Any practical clinical application would require an imaging biomarker that is robust to site differences, so that algorithms developed at academic centers could be applied in the community.

Here, we report results from a large-scale analysis of structural MRI data from 941 participants across 5 sites. Importantly, by pooling images instead of derived measures, and uniformly processing all data via identical pipelines, we were able to harmonize structural measurements and attenuate intersite differences. Building on prior efforts, our first hypothesis was that widespread structural brain abnormalities in adult patients with schizophrenia would be detectable across all cohorts. Our second hypothesis was that a reproducible neuroanatomical signature of schizophrenia would be present across sites and diverse populations, and be detectable at the individual patient level. Our approach included mass-univariate analyses of regional volumes defined using state-of-the-art multi-atlas segmentation,[28] optimally discriminative voxel based analyses (ODVBA)[29,30] for local-multivariate pattern analysis, and individual level classification using consensus-based machine learning.

Methods

Participants and Image Acquisition

Data from 5 MRI studies were included in the current multi-site mega-analysis. All subjects were part of previous studies overseen by local institutional policies.[31–39] Sample demographics of these data sets are given in table 1. Detailed MRI acquisition protocol information for each site is given in supplementary table 1. Anonymized data provided by the 5 sites were provided to and analyzed by the University of Pennsylvania's Center for Biomedical Image Computing and Analytics. Raw MR scans were initially examined for motion, image artifacts, or restricted field-of-view. Scans were also checked for lesions, but this population of patients being relatively young, there were few subjects with lesions, and the total lesion load of these subjects was low. Further details of quality control procedures are given in supplementary methods.

Image Processing

Following magnetic field inhomogeneity correction,[40] MR images were first segmented into a hierarchically organized set of 259 anatomical regions, ranging from total brain volume down to individual cortical gyri and deep structures. To parcellate the brain into anatomically defined regions, we used a highly accurate multi-atlas consensus labeling procedure,[28] which was recently the top-ranked method at an independent international competition.[41] Besides regional volumes, we also calculated a set of rich regional descriptors based on intensity, shape, size, and texture properties of each region, to be used as an additional set of input features in subsequent multivariate classification. See supplementary methods for more details.

Table 1. Sample Demographics

	Harmonized (n = 835)		SET1 (n = 325)		SET2 (n = 180)		SET3 (n = 330)		SET4 (n = 42)		SET5 (n = 64)	
	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients
Group, No. (%)	448 (53.7)	387 (46.3)	170 (52.3)	155 (47.7)	105 (58.3)	75 (41.7)	173 (52.4)	157 (47.6)	24 (57.1)	18 (42.9)	29 (45.3)	35 (54.7)
Female sex, no. (%)	215 (48.0)	148 (38.2)	100 (58.8)	70 (45.2)	61 (58.1)	39 (52.0)	54 (31.2)	39 (24.8)	11 (45.8)	12 (66.7)	13 (44.8)	15 (42.9)
Age, mean (SD)	33.9 (11.6)	35.3 (11.0)	35.7 (13.5)	39.5 (12.0)	35.0 (11.3)	35.5 (8.7)	31.5 (9.2)	31.0 (9.1)	28.2 (4.9)	24.0 (4.8)	23.1 (3.7)	23.5 (4.7)
ICV, mean (SD), cm^3	1415.8 (148.5)	1398.2 (145.4)	1410.8 (166.5)	1378.8 (149.3)	1382.3 (122.9)	1361.6 (153.1)	1441.0 (139.0)	1434.9 (128.7)	1405.1 (93.3)	1409.1 (161.3)	1470.8 (108.9)	1438.5 (148.8)
CPZ dosage, mean (min–max)[a]	N/A	436.0 (0.0–2766.7)	N/A	506.8 (25.0–2766.7)	N/A	478.0 (0.0–1775.0)	N/A	375.9 (10.0–1850.0)	Not available	Not available	Not available	Not available
Duration of illness, mean (min–max), years[b]	N/A	10.78 (0.00–43.90)	N/A	18.21 (0.60–43.90)	N/A	9.58 (0.00–27.00)	N/A	4.44 (0.04–39.41)	Not available	Not available	Not available	Not available
Clinical symptom scores, no. (%), type	N/A	357 (92.2), Various	N/A	143 (92.3), SAPS/SANS	N/A	59 (78.7), PANSS	N/A	155 (98.7), PANSS	Not available	Not available	Not available	Not available
Acquisition protocol	Various		MPRAGE		BRAVO		MPRAGE		MPRAGE		MPRAGE	
Field strength	Various		3 T		3 T		1.5 T		3 T		3 T	
Location	Various		University of Pennsylvania, Philadelphia, PA		Tianji Anning Hospita, Tianjin, China		Ludwig-Maximilian-University, Munich, Germany		Xijing Hospital, Xi'an, China		Peking University, Beijing, China	

Note: N/A = not applicable.
[a]CPZ dosage equivalents available for 71 of SET1 patients, 59 of SET2 patients, and 125 of SET3 patients.
[b]Duration of illness available for 130 of SET1 patients, 59 of SET2 patients, and 153 of SET3 patients.

To perform high-resolution voxel-based analysis, each image's gray matter segmentation was obtained using a previously reported and validated method, MICO.[42] Segmented maps were warped into a common template space using a highly accurate deformable registration (DRAMMS),[43–45] yielding regional volumetric maps (RAVENS) that quantify gray matter volume at each voxel.[46] We masked out subcortical gray matter structures on RAVENS maps, because available structural T1-weighted images do not provide sufficient contrast for reliable intensity-based segmentation;[47] analysis of subcortical structures was limited to the more robust atlas-based regional parcellation described above.

Intersite Image Harmonization

Scanner and demographic differences across MRI samples offer challenges for combining data. Accordingly, we harmonized data across sites by estimating intracranial volume (ICV), site, age, and sex effects on each imaging measure within a pooled sample of controls using a linear model. The coefficients estimated from this model were then applied to the whole sample including patient data. Effectively, this removed the influence of site and demographic effects on the difference between patients and controls. Importantly, this control-based harmonization model was always cross-validated, ie, it was only derived from the training set, and subsequently applied to the test set. Due to the small sample-sizes of the SET4 and SET5 sites, these sites were not included in the harmonization procedure and were set aside for use as validation cohorts in the classification experiments described later. The harmonization procedure was applied independently on each input image feature, including regional features and voxelwise RAVENS maps.

Group-Level Statistical Analyses

All group-level analyses were conducted within the harmonized cohort (n = 835). Regional case-control comparisons were conducted using 2 sample t-tests. Results are reported as effect sizes (Cohen's d-statistic). Effect of medication and disease duration were each evaluated separately using linear regression. Voxelwise analyses of RAVENS maps used ODVBA,[29,30] a local-multivariate pattern analysis method that has been shown in prior studies to have greater sensitivity than both standard voxel-based morphometry and searchlight-based local-multivariate analyses. Type I error was controlled using the false discovery rate (FDR) correction ($Q < 0.05$). Voxelwise effect sizes were reported in regions that were found to be significant by ODVBA.

Multivariate Classification

The above analyses provide a description of structural brain differences between schizophrenia and control groups. As a final step, we used machine learning techniques to create a multivariate classifier that provides individual-level prediction of group status. We used a linear support vector machine (SVM) classifier,[48] a widely used method with extensive validation in various similar high dimensional pattern classification problems. The input features included voxelwise values from RAVENS maps, regional volumes, and an extended set of shape, intensity and texture features computed for each region. A classifier was run separately for each feature type using 10-fold cross-validation. A consensus prediction was then obtained by combining the predictions of each classifier through a weighted averaging based on their cross-validated individual performances. This procedure was performed using an external 10-fold cross-validation (supplementary figure 4), except in experiments that use independent training and testing sets.

To maximize generalizability, we applied the SVM classification using default parameters (C = 1), without feature selection. (Optimization of such parameters would likely further improve accuracy, albeit at the risk of overfitting the data.) See supplementary methods for further details.

Multivariate models were trained under 3 different conditions to evaluate the impact of pooling data across sites and training/testing on different sites. First, models were trained and tested separately within each site using 10-fold cross-validation. Second, a single 10-fold cross-validated SVM classification analysis was conducted using the complete sample of harmonized pooled data. Third, we performed training/testing on different sites through *leave-one-site-out* cross-validation. This procedure included applying a model trained on the complete pooled data to the 2 independent validation cohorts. This procedure aimed to evaluate the generalizability of the classifier to a totally new data set sampled from an entirely different population and scanner. Importantly, in this procedure harmonization was done only within the training set, and the control-based regression model was applied to the test set. Before training the classifier, the training and testing sets were z-scored separately, with the assumption that the test set has a similar patient to control ratio as the training set. For all analyses, individual-subject classification performance was evaluated using receiver operating characteristic (ROC) curves. Classification performance was summarized using both the area under the curve (AUC) and the classification accuracy metrics.

Finally, to assess the clinical relevance of the MRI-based classification, we investigated correlations between the classifier output and clinical symptom scores of individual subjects. The symptom scores for different sites were obtained using common symptom rating scales in schizophrenia research: the Scale for the Assessment of Positive Symptoms,[49] the Scale for the Assessment of Negative Symptoms,[50] and the Positive And Negative Syndrome Scale.[51,52] Table 1 details the available type of scores for each site. The scores were normalized by z-scoring them within each site and pooled together. For this analysis, individualized pseudo-probabilities of having the neuroanatomical

signature of schizophrenia were estimated from the classifier's output using sigmoid fits. These pseudo-probabilities correspond to the degree to which a given individual's brain appears to match the schizophrenia pattern.

Results

Regional and Voxelwise Analyses Reveal Distributed Structural Brain Abnormalities in Schizophrenia

Mass-univariate regional analyses of pooled data delineated robust abnormalities of brain structure that survived FDR correction. The structures that show the most significant group differences, both in positive and negative directions, are shown in figure 1. Patients with schizophrenia had marked ventricular expansion, as well as larger pallidum volumes. Cortical gray matter volume loss was also quite evident, with maximal effects in the prefrontal cortex (superior frontal gyrus; $d = 0.51$) as well as temporal cortex, parietal cortex, insula, and amygdala. The effect sizes of volumetric changes were strongly correlated across sites (pairwise correlation coefficient range: 0.71–0.81), indicating strong consistency despite protocol, scanner, and demographic differences across sites. A complete list of all regions with significant group differences ($q < 0.05$) is provided in supplementary table 2. Site-specific analysis results are provided in supplementary table 3.

High-resolution voxelwise analyses of harmonized RAVENS maps in the pooled sample using ODVBA delineated a distributed pattern of volume loss (figure 2), and showed good agreement with the results of the ROI-based analysis (supplementary figure 1). The results of the ODVBA analyses performed using each data set separately are shown in supplementary figure 2. The effects found for individual data sets were consistent with each other and with the pooled analysis, but less significant than the pooled analysis results, emphasizing the enhanced power of the pooled analysis compared to the smaller single-site analyses.

We also evaluated the impact of disease duration and antipsychotic load (chlorpromazine equivalents[53]) on the regional data (supplementary table 4). In contrast to the very robust and distributed group differences, medication effects were only seen in the frontal operculum, where higher medication dose was associated with lower volume. Duration of illness was associated with larger lateral ventricles and pallidum, as well as reduced volume of regions including the middle frontal gyrus, parahippocampal gyrus, and hippocampus.

Multivariate Classification is Accurate Using Data Pooled Across Sites

Classification results are detailed in table 2 for the within-site, pooled, and between-site experiments. Compared to individual within-site classification, pooled classification resulted in higher classification accuracy for each of SET1, SET2, and SET3, with 0.02 to 0.045 increase in AUC. The ROC curve for the cross-validated classification within the pooled-data is shown in figure 3a. These results indicate that pooling data across sites is not only viable, but it actually improves the classification results. The leave-site-out cross-validation procedure on SET1, SET2 and SET3, ie, using models trained exclusively on data collected at different sites, achieved AUC values between 0.78 and 0.80, which were comparable to pooled or within site cross-validation results. Besides, combination of training data from 2 separate external sites consistently outperformed training on each external site individually, suggesting

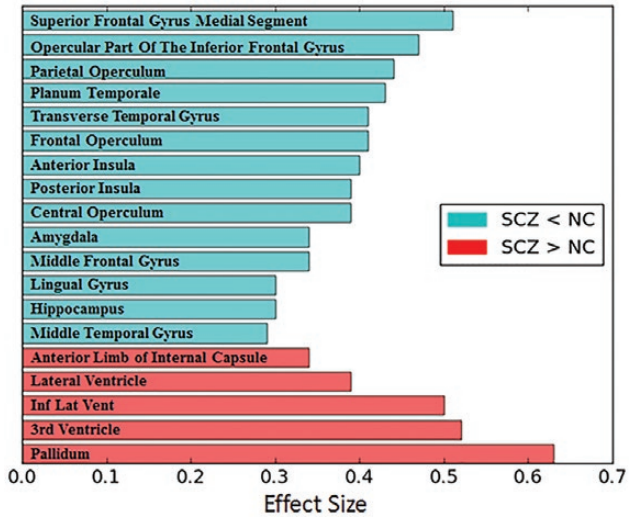


Fig. 1. Key regional volume differences between patients (SCZ) and controls (NC). ROIs with the highest effect sizes (absolute effect size ≥ 0.29) between controls ($N = 448$) and patients ($N = 387$), calculated using pooled, harmonized data. For a complete list of all regions with significant group differences ($q < 0.05$), see supplementary table 1.

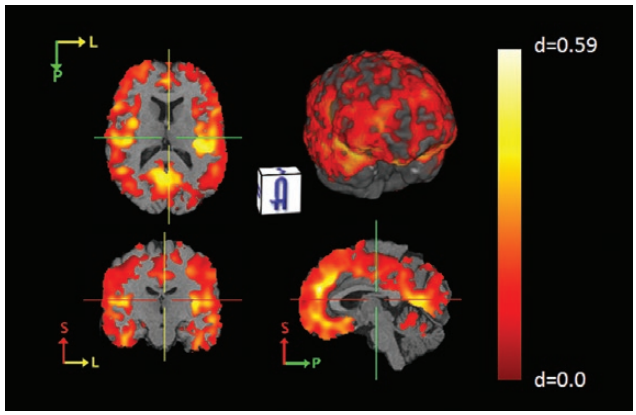


Fig. 2. Voxelwise gray matter volume differences between patients (SCZ) and controls (NC). Effect-size maps between controls ($N = 448$) and patients ($N = 387$) calculated using pooled, harmonized data. The highlighted regions that show significant group differences were calculated using the output of ODVBA thresholded at the FDR-corrected significance value of $q < 0.05$. *Note:* The location of the crosshairs is unchanged between each of the cross-sections. A color figure is available online at *Schizophrenia Bulletin*.

Table 2. Classification Results

	Test Site, AUC/Accuracy				
Train Site[a]	SET1	SET2	SET3	SET4	SET5
SET1	0.767/0.696	0.765/0.694	0.801/0.707	0.588/0.643	0.831/0.758
SET2	0.770/0.720	0.872/0.808	0.791/0.710	0.826/0.810	0.931/0.806
SET3	0.782/0.701	0.793/0.706	0.802/0.732	0.639/0.524	0.876/0.806
Other sites[b]	0.785/0.723	0.805/0.739	0.804/0.731	0.731/0.738	0.912/0.774
Pooled	0.813/0.736	0.893/0.803	0.829/0.762	N/A	N/A

Note: Color codes show the set of results for the 3 different settings in classification experiments (blue: within site classification with leave 10% out cross-validation; green: classification with pooled data with leave 10% out cross-validation; red: classification with training/testing on different sites). N/A = not applicable.
[a]Sites 1, 2, and 3 comprised the discovery cohort. Sets 4 and 5 comprised the validation cohort.
[b]For sites 1, 2, and 3, other sites indicates that 2 of the 3 sites were pooled for training, and the third site was used for testing. For sites 4 and 5, other sites indicates the pooled data from sites 1, 2, and 3 used for training.

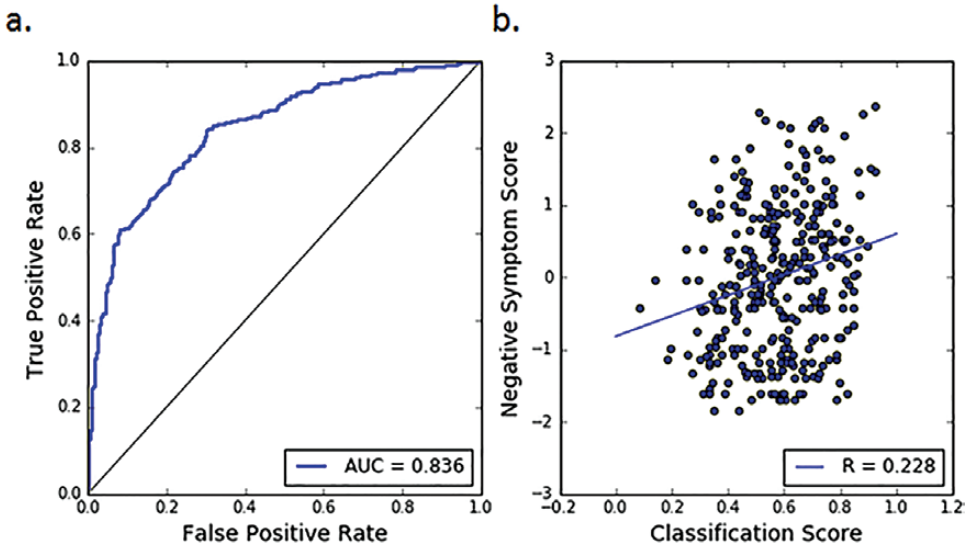


Fig. 3. Classification results. (a) ROC curve for patient vs control classification. The classification was performed using the pooled data and with leave-10%-out cross-validation through consensus-voting of all input feature sets, ie, RAVENS maps, regional volumes, and region-based descriptors. (b) Correlation between classifier output and negative clinical symptom scores (SANS for SET1, PANSS Negative for sets 2 and 3). Classification scores were obtained by converting the distances of the test samples from the discrimination hyperplane to pseudo-probabilities using sigmoid fits. Negative symptom scores were *z*-scored within data sets and then pooled together. The correlation for classifier score with negative symptom score was equal to $r = .228$ ($P = .00003$). No correlations were found between positive symptom scores and the classifier output ($r = -.019$, $P = .73$).

that pooling a larger number of data sets for training helps to increase accuracy and robustness of the classifier. Classification results for the independent validation cohorts, SET4 and SET5 were also high (accuracies 0.738 and 0.774), further emphasizing the feasibility of using models trained exclusively on data collected at a different site. A comparison of the relative performance of different types of input feature sets used in the classification experiments is shown in supplementary figure 3.

Classification Probability Is Related to Negative Symptoms

We found a significant positive correlation ($r = .228$, $P = .00003$) between the individual pseuoprobabilities of having neuroanatomical pattern and the negative symptom scores of the subjects (figure 3b). No significant correlations were found between the positive symptom scores and the classification pseudoprobabilities, nor was there a significant relationship to the duration of illness.

Discussion

In a large-scale analysis of data pooled across sites, we demonstrated that the widespread structural brain abnormalities associated with schizophrenia can be used for accurate single-subject classification using machine learning techniques. Application of state-of-the-art consensus-based classifiers on different types of imaging features yielded high performance, and indicated that multivariate

neuroanatomical signatures have the potential to become a quantitative imaging biomarker for schizophrenia. Importantly, these procedures remained accurate when trained and tested on data from completely different sites, populations, and scanners, highlighting the reproducibility of this imaging signature and, therefore, its translational potential.

Structural Brain Abnormalities in Schizophrenia Are Widespread and Visible at Multiple Scales

Capitalizing on a sample of nearly 1000 participants amassed via a multi-site, mega-analytic design, we provide robust evidence of widespread structural brain abnormalities in schizophrenia. These were present on multiple scales, and were evident in both analyses of regional volumes as well as high-resolution voxelwise analyses. Brain abnormalities in schizophrenia have been studied using brain imaging for 3 decades.[2,4] While abnormalities have been consistently documented by single-site studies, the relative degree to which different brain regions were impacted has varied considerably by sample. Large-scale meta-analyses have pooled data,[8,10–12,54] but they have been limited by important differences in data processing across studies. The ENIGMA and COROCO consortia, as well the study of Gupta et al.[16] have overcome this limitation with rigorous meta-analyses of data processed in consistent fashion (albeit not always with identical pipelines harmonized on raw data), and recently provided evidence of subcortical volume loss of moderate effect size.[14,15] Our analysis of subcortical structures provided convergent results to those provided by ENIGMA and other studies, including loss of hippocampus, thalamus, and amygdala volume; expansion of the pallidum; and ventricular enlargement. However, we build upon such findings from the subcortex and demonstrate cortical abnormalities of often large effect size,[11,16] with a maximal deficit in the prefrontal cortex. In line with previous studies, duration of illness was associated with increasing expansion of the pallidum and ventricles.[9,11,14,55] In contrast to the lack of medication effects found in ENIGMA, but in line with a previous large-scale retrospective meta-analysis, higher dose of antipsychotics was associated with larger pallidum and reduced gray matter volume in several frontal regions.[9,11,18]

Multivariate Classification of Individuals Is Accurate Even Using Data From Different Scanners and Sites

While mass-univariate analyses provide a valuable description of the structural brain abnormalities associated with schizophrenia at the group level, they cannot function as biomarkers in individual patients. Accordingly, our primary focus was to use machine learning techniques to create a classifier that leveraged the complex multivariate pattern of structural deficits.[21] The classifier we described here could differentiate patients and controls with a high degree of accuracy, with levels comparable to those previously reported in single-site studies.[18,24,25,27,56]

A highly significant positive correlation was found between the classification output and negative symptom scores. Thus, the probability to which an individual's multivariate pattern of structural anatomy was classified as consistent with schizophrenia was correlated with the burden of negative symptoms. In contrast, there was not a significant correlation with positive symptoms. This result is consistent with extensive prior research linking negative symptoms to structural brain abnormalities in schizophrenia.[57] To the degree to which both structural brain abnormalities and negative symptoms portend a poor response to standard pharmacotherapies,[58] these results may assist in stratification within clinical trials of new therapies which integrate structural neuroimaging and machine-learning tools.

Critically, classification results were obtained with established algorithms and methods, to ensure that the resultant biomarker would be accessible for widespread use. Nonetheless, it is likely that better classification could potentially be achieved via parameter tuning, feature selection, and use of nonlinear kernels or deep-leaning approaches. However, such gains might also carry lead to greater model complexity, higher risk of over-fitting to training data, and reduced generalizability to new data sets. Such generalizability is of paramount importance for clinical translation.

Indeed, the single most important finding from this approach was that multivariate classifiers could retain accurate prediction even when trained and tested on data from separate scanners. Machine learning can be quite robust to intersite variations, by virtue of depending on the relative weights, ie, contrasts, of a complex pattern rather than on absolute volume-based thresholds, which can be impacted by slight variations of image contrast across scanners. These results suggest that models trained on data collected at specialized academic centers have the potential to be applied to data acquired in the community in order to calculate an imaging biomarker of schizophrenia. Importantly, leave-one-site-out was tested without the need for harmonization of the testing data with the training data set, but only via *z*-scoring within the new test site data. The classification pipeline is available in our image processing portal as a web-accessible application (CBICA Image Processing Portal: https://ipp.cbica.upenn.edu/). The web application allows users to submit single or multiple T1-weighted MRI images and outputs individualized scores, computed using automatically extracted imaging features and the presaved classification model from the harmonized training data. Thus, for any new data set, this web portal provides the scientific and clinical community with a freely accessible quantitative index of the imaging-based biomarker described here.

Limitations and Future Directions

Although this study benefited from a large sample size and advanced analytics, certain limitations should be noted. While developing a distinctive imaging signature of schizophrenia is an important endeavor, its value in predicting disease progression, treatment response, and in assisting differential diagnosis needs to be investigated in future studies. We note that the proposed imaging index, which can be calculated at an individual level using the provided online tool, cannot be used as a direct diagnostic index, as it has not been validated in heterogeneous clinical populations. Applications to samples that are not enriched for schizophrenia may result in a high false-positive rate. Like most other psychiatric conditions, schizophrenia is associated with substantial co-morbidity of other psychiatric conditions (such as mood disorders). Unfortunately, comorbid conditions were not consistently recorded across the participating sites, and thus could not be evaluated in the present analysis. Future studies should evaluate comorbidity of other psychiatric conditions on classification performance. Expanded data sets that seek to parse heterogeneity within diagnostic groups will be critical.[39,59–61] Such work is particularly important for studies of youth with psychosis-spectrum or prodromal symptoms, where there is substantial diversity of clinical outcomes.[55,62–66] Recent work[67] using heterogeneity analyses offers hope that semi-supervised machine learning methods can provide a better understanding of the heterogeneity of neuroanatomical signatures of schizophrenia. Finally, future work that combines large-scale imaging data sets in youth with well-characterized longitudinal clinical follow up with advanced multivariate analytics will be critical areas of focus moving forward.

Conclusions

Using a pooled mega-analytic strategy, the present data provide among the most robust evidence to date of structural brain abnormalities in adults with schizophrenia. Furthermore, these results emphasize that such signals can be used to derive highly accurate multivariate models that allow for discrimination at the level of individual patients, thereby providing a robust neuroanatomical signature of schizophrenia. Critically, this signature remains accurate even when the classifier is trained using data from different sites and scanners. Taken together, our findings highlight the accelerating promise of imaging-based biomarkers in major neuropsychiatric illnesses such as schizophrenia. The classifier used in this paper for classification is publicly available online at the CBICA Image Processing Portal.

Supplementary Material

Supplementary data are available at *Schizophrenia Bulletin* online.

Funding

The research at the University of Pennsylvania site was supported in part by the National Institute of Health (P50MH096891 to R.E.G., R01MH107703 to T.D.S., R01MH101111 to D.H.W.) and the Dowshen Program for Neuroscience.

Acknowledgments

At the University of Pennsylvania site, thanks to Monica Calkins, PhD for assistance with clinical phenotyping.
Disclosures: R.E.G. participated in an advisory board for Otsuka Pharmaceuticals and RCG receives royalties from the Brain Resource Center and serves without compensation on an advisory board for Lumosity. All other authors have no disclosures.

References

1. Freedman R. Schizophrenia. *N Engl J Med*. 2003;349:1738–1749.
2. Stevens JR. An anatomy of schizophrenia? *Arch Gen Psychiatry*. 1973;29:177–189.
3. Gross G, Huber G, Schüttler R. Computerized tomography studies on schizophrenic diseases. *Arch Psychiatr Nervenkr (1970)*. 1982;231:519–526.
4. Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. *Schizophr Res*. 2001;49:1–52.
5. Insel TR. Rethinking schizophrenia. *Nature*. 2010;468:187–193.
6. Bora E, Fornito A, Radua J, et al. Neuroanatomical abnormalities in schizophrenia: a multimodal voxelwise meta-analysis and meta-regression analysis. *Schizophr Res*. 2011;127:46–57.
7. Cooper D, Barker V, Radua J, Fusar-Poli P, Lawrie SM. Multimodal voxel-based meta-analysis of structural and functional magnetic resonance imaging studies in those at elevated genetic risk of developing schizophrenia. *Psychiatry Res*. 2014;221:69–77.
8. Ellison-Wright I, Glahn DC, Laird AR, Thelen SM, Bullmore E. The anatomy of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. *Am J Psychiatry*. 2008;165:1015–1023.
9. Fusar-Poli P, Smieskova R, Kempton MJ, Ho BC, Andreasen NC, Borgwardt S. Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neurosci Biobehav Rev*. 2013;37:1680–1691.
10. Glahn DC, Laird AR, Ellison-Wright I, et al. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. *Biol Psychiatry*. 2008;64:774–781.
11. Haijma SV, Van Haren N, Cahn W, Koolschijn PC, Hulshoff Pol HE, Kahn RS. Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. *Schizophr Bull*. 2013;39:1129–1138.
12. Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*. 2005;162:2233–2245.

13. Thompson PM, Stein JL, Medland SE, et al.; Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav*. 2014;8:153–182.
14. van Erp TG, Hibar DP, Rasmussen JM, et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry*. 2016;21:585.
15. Okada N, Fukunaga M, Yamashita F, et al. Abnormal asymmetries in subcortical brain volume in schizophrenia. *Mol Psychiatry*. 2016;21:1460–1466.
16. Gupta CN, Calhoun VD, Rachakonda S, et al. Patterns of gray matter abnormalities in schizophrenia based on an international mega-analysis. *Schizophr Bull*. 2015;41:1133–1142.
17. Bansal R, Staib LH, Laine AF, et al. Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. *PLoS One*. 2012;7:e50698.
18. Davatzikos C, Shen D, Gur RC, et al. Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. *Arch Gen Psychiatry*. 2005;62:1218–1227.
19. Koutsouleris N, Meisenzahl EM, Davatzikos C, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry*. 2009;66:700–712.
20. Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Trans Med Imaging*. 2007;26:93–105.
21. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P. Machine learning approaches: from theory to application in schizophrenia. *Comput Math Methods Med*. 2013;2013:867924.
22. Nakamura K, Kawasaki Y, Suzuki M, et al. Multiple structural brain measures obtained by three-dimensional magnetic resonance imaging to distinguish between schizophrenia patients and normal subjects. *Schizophr Bull*. 2004;30:393–404.
23. Kawasaki Y, Suzuki M, Kherif F, et al. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *Neuroimage*. 2007;34:235–242.
24. Koutsouleris N, Gaser C, Bottlender R, et al. Use of neuroanatomical pattern regression to predict the structural brain dynamics of vulnerability and transition to psychosis. *Schizophr Res*. 2010;123:175–187.
25. Nieuwenhuis M, van Haren NE, Hulshoff Pol HE, Cahn W, Kahn RS, Schnack HG. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage*. 2012;61:606–612.
26. Schnack HG, Nieuwenhuis M, van Haren NE, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage*. 2014;84:299–306.
27. Kambeitz J, Kambeitz-Ilankovic L, Leucht S, et al. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology*. 2015;40:1742–1751.
28. Doshi J, Erus G, Ou Y, et al.; Alzheimer's Neuroimaging Initiative. MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *Neuroimage*. 2016;127:186–195.
29. Zhang T, Davatzikos C. ODVBA: optimally-discriminative voxel-based analysis. *IEEE Trans Med Imaging*. 2011;30:1441–1454.
30. Zhang T, Davatzikos C. Optimally-discriminative voxel-based morphometry significantly increases the ability to detect group differences in schizophrenia, mild cognitive impairment, and Alzheimer's disease. *Neuroimage*. 2013;79:94–110.
31. Satterthwaite TD, Wolf DH, Loughead J, et al. Association of enhanced limbic response to threat with decreased cortical facial recognition memory response in schizophrenia. *Am J Psychiatry*. 2010;167:418–426.
32. Wolf DH, Satterthwaite TD, Kantrowitz JJ, et al. Amotivation in schizophrenia: integrated assessment with behavioral, clinical, and imaging measures. *Schizophr Bull*. 2014;40:1328–1337.
33. Wang H, Zeng LL, Chen Y, Yin H, Tan Q, Hu D. Evidence of a dissociation pattern in default mode subnetwork functional connectivity in schizophrenia. *Sci Rep*. 2015;5:14655.
34. Zhu J, Zhuo C, Liu F, Xu L, Yu C. Neural substrates underlying delusions in schizophrenia. *Sci Rep*. 2016;6:33857.
35. Zhuo C, Ma X, Qu H, Wang L, Jia F, Wang C. Schizophrenia patients demonstrate both inter-voxel level and intra-voxel level white matter alterations. *PLoS One*. 2016;11:e0162656.
36. Yan H, Tian L, Wang Q, et al. Compromised small-world efficiency of structural brain networks in schizophrenic patients and their unaffected parents. *Neurosci Bull*. 2015;31:275–287.
37. Yan H, Tian L, Yan J, et al. Functional and anatomical connectivity abnormalities in cognitive division of anterior cingulate cortex in schizophrenia. *PLoS One*. 2012;7:e45659.
38. Huang P, Xi Y, Lu ZL, et al. Decreased bilateral thalamic gray matter volume in first-episode schizophrenia with prominent hallucinatory symptoms: A volumetric MRI study. *Sci Rep*. 2015;5:14505.
39. Zhang T, Koutsouleris N, Meisenzahl E, Davatzikos C. Heterogeneity of structural brain changes in subtypes of schizophrenia revealed using magnetic resonance imaging pattern analysis. *Schizophr Bull*. 2015;41:74–84.
40. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29:1310–1320.
41. Doshi JJ EG, Ou Y, Davatzikos C. Ensemble-based medical image labeling via sampling morphological appearance manifold. In: *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications*. Nagoya, Japan; 2013.
42. Li C, Gore JC, Davatzikos C. Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation. *Magn Reson Imaging*. 2014;32:913–923.
43. Yangming Ou AS, Aristeidis Sotiras. DRAMMS Distribution General Website; 2009. http://www.med.upenn.edu/sbia/dramms.html
44. Ou Y, Sotiras A, Paragios N, Davatzikos C. DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. *Med Image Anal*. 2011;15:622–639.
45. Ou Y, Akbari H, Bilello M, Da X, Davatzikos C. Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights. *IEEE Trans Med Imaging*. 2014;33:2039–2065.
46. Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *Neuroimage*. 2001;14:1361–1369.

47. Helms G, Draganski B, Frackowiak R, Ashburner J, Weiskopf N. Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. *Neuroimage*. 2009;47:194–198.
48. Vapnik VN. An overview of statistical learning theory. *IEEE Trans Neural Netw*. 1999;10:988–999.
49. Andreasen N. *The Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City: University of Iowa; 1984.
50. Andreasen N. *The scale for the assessment of negative symptoms (SANS)*. Iowa City: University of Iowa; 1983.
51. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–276.
52. Kay SR, Opler LA, Lindenmayer JP. The Positive and Negative Syndrome Scale (PANSS): rationale and standardisation. *Br J Psychiatry Suppl* 1989(7):59–67.
53. Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. *J Clin Psychiatry*. 2003;64:663–667.
54. Chan RC, Di X, McAlonan GM, Gong QY. Brain anatomical abnormalities in high-risk individuals, first-episode, and chronic schizophrenia: an activation likelihood estimation meta-analysis of illness progression. *Schizophr Bull*. 2011;37:177–188.
55. Fraguas D, Díaz-Caneja CM, Pina-Camacho L, Janssen J, Arango C. Progressive brain changes in children and adolescents with early-onset psychosis: a meta-analysis of longitudinal MRI studies. *Schizophr Res*. 2016;173:132–139.
56. Fan Y, Gur RE, Gur RC, et al. Unaffected family members and schizophrenia patients share brain structure patterns: a high-dimensional pattern classification study. *Biol Psychiatry*. 2008;63:118–124.
57. Ren W, Lui S, Deng W, et al. Anatomical and functional brain abnormalities in drug-naive first-episode schizophrenia. *Am J Psychiatry*. 2013;170:1308–1316.
58. Wolf DH. Anhedonia in schizophrenia. *Curr Psychiatry Rep*. 2006;8:322–328.
59. Bleich-Cohen M, Jamshy S, Sharon H, Weizman R, Intrator N, Poyurovsky M, Hendler T. Machine learning fMRI classifier delineates subgroups of schizophrenia patients. *Schizophr Res*. 2014;160:196–200.
60. Ivleva EI, Bidesi AS, Keshavan MS, et al. Gray matter volume as an intermediate phenotype for psychosis: Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry*. 2013;170:1285–1296.
61. Narayanan B, Ethridge LE, O'Neil K, et al. Genetic sources of subcomponents of event-related potential in the dimension of psychosis analyzed from the B-SNIP study. *Am J Psychiatry*. 2015;172:466–478.
62. Koutsouleris N, Gaser C, Patschurek-Kliche K, et al. Multivariate patterns of brain-cognition associations relating to vulnerability and clinical outcome in the at-risk mental states for psychosis. *Hum Brain Mapp*. 2012;33:2104–2124.
63. Karageorgiou E, Schulz SC, Gollub RL, et al. Neuropsychological testing and structural magnetic resonance imaging as diagnostic biomarkers early in the course of schizophrenia and related psychoses. *Neuroinformatics*. 2011;9:321–333.
64. Koutsouleris N, Davatzikos C, Bottlender R, et al. Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull*. 2012;38:1200–1215.
65. Pina-Camacho L, Garcia-Prieto J, Parellada M, et al. Predictors of schizophrenia spectrum disorders in early-onset first episodes of psychosis: a support vector machine model. *Eur Child Adolesc Psychiatry*. 2015;24:427–440.
66. Satterthwaite TD, Wolf DH, Calkins ME, et al. Structural brain abnormalities in youth with psychosis spectrum symptoms. *JAMA Psychiatry*. 2016;73:515–524.
67. Dong A, Toledo JB., Nicolas H., et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression, and biomarkers. *Brain*. 2016;140(3):735–747. doi:10.1093/brain/aww319

Supplementary Online Content

Supplementary Methods. Quality check method and results; Calculation of additional image features for multivariate classification; consensus-voting based classification procedure.

Supplementary Table 1. MRI scanner and image-acquisition protocol information.

Supplementary Table 2. Regional effect sizes and significance values for the group comparisons between controls and patients.

Supplementary Table 3. Regional effect sizes for the group comparisons between controls and patients, using pooled data versus using each dataset individually.

Supplementary Figure 1. Comparison of regions with significant group differences between controls and patients, for voxelwise versus regional analyses.

Supplementary Figure 2. Voxelwise group comparisons between controls and patients using GM RAVENS maps, for individual sites versus pooled data.

Supplementary Figure 3. Classification performance of individual feature sets and the ensemble classifier on the pooled data.

Supplementary Figure 4. Outline of image processing and cross-validated classification using ensemble of features

Supplementary Methods.

Quality Check Method and Results

A set of well-defined quality control (QC) procedures were applied to all datasets. These QC procedures included both automated flags and manual verification, which were systematically applied on both raw and processed images. Specifically, as part of the manual QC, all raw images were examined for motion, image artifacts, or restricted field-of-view. They were also checked for lesions, but this population of patients being relatively young, there were few subjects with lesions. Similarly, processed images were manually evaluated for pipeline failures, including failures of brain extraction, poor segmentation, or errors in image registration. In addition to manual QC, automated QC procedures flagged images based on outlying values of quantified metrics (for example, regional volume); flagged images were subsequently re-evaluated by an image analyst. A few subjects were also excluded for lacking basic demographic data. In total, scans of n=95 subjects were excluded during the initial or final QC process (45 from SET 1, 30 from SET 2, 11 from SET 3, 6 from SET 4, and 3 from SET 5), to give the final sample shown in Table 1 in the main manuscript.

Calculation of Additional Features for Multivariate Classification

We computed an extended set of region-based features derived from the anatomical regions of interest (ROIs) segmented using MUSE[1]. Specifically, the region-based feature sets consisted of the following sets:

- Intensity: Minimum, maximum, mean, median, standard deviation, skewness, and kurtosis values for each ROI.

- Shape: Elongation, flatness and roundness for each ROI.

- Size: Volume, surface area and spherical radius for each ROI. The size features were corrected for individual intra-cranial volume (ICV) to account for inter-subject differences in head size.

- Texture: Energy, entropy, homogeneity, inertia, cluster shape, cluster prominence for each ROI[2-5].

As additional input feature sets for the multivariate classification, we also used voxel values of regional volumetric maps (RAVENS)[6], which allow accurate quantification of local anatomical differences or changes, on a common atlas space. The categories of regional volumetric features used in the classification were the following:

- RAVENS: CSF, GM and WM RAVENS maps calculated via deformable registration (DRAMMS) [7, 8] to common atlas space and normalized for ICV.

-Affine RAVENS: CSF, GM and WM RAVENS maps calculated via affine-registration with 12 degrees of freedom to common atlas space and normalized for ICV.

Consensus-Voting Classification Procedure

A consensus-voting procedure is used for the classification, where classifiers are first trained on each feature set independently to generate an ensemble, and the results are fused into a single classification score through weighted voting of individual scores from each classifier. The weight of a feature set is determined based on the leave-10%-out cross-validated classification accuracy using only that single feature set. The ensemble classifier is applied with external leave-10%-out cross-validation. All experiments are repeated 10 times. The classification performance for individual feature sets and the ensemble are shown in Supplementary Figure 3.

Supplementary Table 1. MRI Scanner and Image-Acquisition Protocol Information

Site	Field Strength	Brand	Protocol Name	Repetition Time	Echo Time	Inversion Time	Flip Angle	Field of View	Slice Thickness
SET1	3T	Siemens	MPRAGE	1810 ms	3.51 ms	1100 ms	9	240 x 180	1 mm
SET2	3T	GE	BRAVO	8.2 ms	3.2 ms	450 ms	12	256 x 256	1 mm
SET3	1.5T	Siemens	MPRAGE	11.6 ms	4.9 ms	N/A	N/A	512 x 512	1.5 mm
SET4	3T	Siemens	MPRAGE	2530 ms	3.5 ms	N/A	7	256 x 256	1 mm
SET5	3T	Siemens	MPRAGE	2350 ms	3.44 ms	N/A	7	256 x 256	1 mm

Supplementary Table 2. Regional effect sizes[1] and significance values for the group comparisons between controls and patients. All values are corrected for multiple comparisons using false discovery rate (FDR) correction.

ROI	Effect Size (Left / Right)	Corrected significance level (q-Value)	Percent Difference
Pallidum	0.63 (0.59 / 0.59)	5.30E-17	5.74
3rd Ventricle	0.52 (n/a)	6.25E-12	37.37
Superior Frontal Gyrus Medial Segment	-0.51 (-0.42 / -0.42)	1.09E-11	-3.89
Inf Lat Vent	0.50 (0.47 / 0.47)	4.30E-11	29.85
Opercular Part Of The Inferior Frontal Gyrus	-0.47 (-0.41 / -0.41)	4.23E-10	-5.49
Parietal Operculum	-0.44 (-0.42 / -0.42)	6.64E-09	-5.68
Planum Temporale	-0.43 (-0.38 / -0.38)	7.39E-09	-4.91
Frontal Operculum	-0.41 (-0.36 / -0.36)	3.38E-08	-4.37
Transverse Temporal Gyrus	-0.41 (-0.41 / -0.41)	3.71E-08	-6.54
Anterior Insula	-0.4 (-0.38 / -0.38)	1.25E-07	-2.98
Lateral Ventricle	0.39 (0.39 / 0.39)	1.73E-07	30.49
Central Operculum	-0.39 (-0.35 / -0.35)	1.95E-07	-3.24
Posterior Insula	-0.39 (-0.31 / -0.31)	1.95E-07	-3.55
Middle Frontal Gyrus	-0.34 (-0.30 / -0.30)	4.31E-06	-2.36
Amygdala	-0.34 (-0.26 / -0.26)	7.56E-06	-3.64
Anterior Limb Of Internal Capsule	0.34 (0.36 / 0.36)	7.56E-06	3.33
Hippocampus	-0.30 (-0.32 / -0.32)	8.54E-05	-2.46
Lingual Gyrus	-0.30 (-0.23 / -0.23)	8.54E-05	-2.43
Middle Temporal Gyrus	-0.29 (-0.24 / -0.24)	1.12E-04	-2.06
Planum Polare	-0.29 (-0.28 / -0.28)	1.47E-04	-2.67
Fusiform Gyrus	-0.29 (-0.23 / -0.23)	1.56E-04	-2.27
Posterior Limb Of Internal Capsule	0.28 (0.26 / 0.26)	1.57E-04	2.31
Superior Temporal Gyrus	-0.28 (-0.25 / -0.25)	1.64E-04	-2.34
Precuneus	-0.28 (-0.27 / -0.27)	2.33E-04	-2.04
Posterior Orbital Gyrus	-0.27 (-0.27 / -0.27)	2.78E-04	-3.05
Medial Frontal Cortex	-0.27 (-0.19 / -0.19)	3.76E-04	-3.27
Posterior Cingulate Gyrus	-0.27 (-0.26 / -0.26)	3.76E-04	-2.47
Postcentral Gyrus	-0.26 (-0.27 / -0.27)	6.11E-04	-2.34
Calcarine Cortex	-0.25 (-0.22 / -0.22)	7.29E-04	-3.87
Cuneus	-0.23 (-0.18 / -0.18)	2.32E-03	-2.51
Occipital Fusiform Gyrus	-0.23 (-0.13 / -0.13)	2.82E-03	-2.31
Precentral Gyrus	-0.23 (-0.17 / -0.17)	2.90E-03	-1.71
Parahippocampal Gyrus	-0.22 (-0.17 / -0.17)	2.90E-03	-1.89
Thalamus Proper	-0.22 (-0.20 / -0.20)	3.56E-03	-1.32
Angular Gyrus	-0.22 (-0.18 / -0.18)	3.85E-03	-1.96
Anterior Orbital Gyrus	-0.21 (-0.12 / -0.12)	4.38E-03	-2.92
Inferior Occipital Gyrus	-0.21 (-0.20 / -0.20)	4.69E-03	-2.46
Superior Parietal Lobule	-0.20 (-0.19 / -0.19)	6.38E-03	-1.95

[1] Cohen's d. A negative value indicates reduced volume in schizophrenic patients relative to healthy controls.

Supplementary Table 2. Regional effect sizes[1] and significance values for the group comparisons between controls and patients (continued)

ROI	Effect Size (Left / Right)	Corrected significance level (q-Value)	Percent Difference
Basal Forebrain	-0.20 (-0.13 / -0.13)	6.86E-03	-3.12
Supplementary Motor Cortex	-0.20 (-0.20 / -0.20)	6.86E-03	-1.96
Superior Frontal Gyrus	-0.20 (-0.22 / -0.22)	8.88E-03	-1.63
Accumbens Area	-0.19 (-0.21 / -0.21)	9.96E-03	-2.35
Entorhinal Area	-0.17 (-0.09 / -0.09)	2.85E-02	-2.00
Supramarginal Gyrus	-0.17 (-0.11 / -0.11)	2.85E-02	-1.59
Middle Cingulate Gyrus	-0.16 (-0.14 / -0.14)	3.92E-02	-1.38
Putamen	0.15 (0.16 / 0.16)	4.17E-02	1.27
Subcallosal Area	-0.15 (-0.14 / -0.14)	4.17E-02	-2.25
4th Ventricle	0.15 (n/a)	4.29E-02	3.91
Middle Occipital Gyrus	-0.15 (-0.11 / -0.11)	4.29E-02	-1.67
Triangular Part Of The Inferior Frontal Gyrus	-0.15 (-0.18 / -0.18)	4.29E-02	-2.11

[1] Cohen's d. A negative value indicates reduced volume in schizophrenic patients relative to healthy controls.

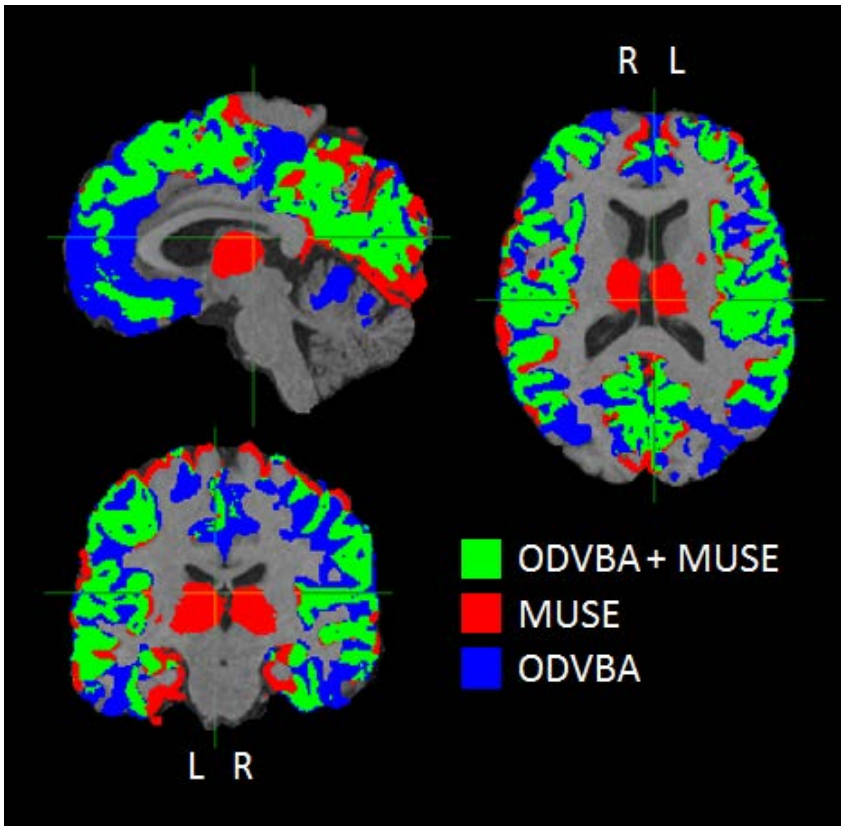
Supplementary Table 3. Regional effect sizes[1] for the group comparisons between controls and patients, using pooled data versus using each dataset individually. All values are corrected for multiple comparisons using false discovery rate (FDR) correction.

	Site			
ROI	**Pooled**	**SET1**	**SET2**	**SET3**
Pallidum	0.63	0.77	0.97	0.35
3rd Ventricle	0.52	0.47	0.59	0.58
Superior Frontal Gyrus Medial Segment	-0.51	-0.40	-0.87	-0.51
Inf Lat Vent	0.50	0.33	0.84	0.47
Opercular Part Of The Inferior Frontal Gyrus	-0.47	-0.60	-0.50	-0.35
Parietal Operculum	-0.44	-0.36	-0.55	-0.50
Planum Temporale	-0.43	-0.58	-0.45	-0.31
Frontal Operculum	-0.41	-0.42	-0.55	-0.34
Transverse Temporal Gyrus	-0.41	-0.36	-0.66	-0.39
Anterior Insula	-0.40	-0.32	-0.74	-0.31
Lateral Ventricle	0.39	0.37	0.71	0.26
Central Operculum	-0.39	-0.40	-0.55	-0.30
Posterior Insula	-0.39	-0.37	-0.83	-0.24
Middle Frontal Gyrus	-0.34	-0.27	-0.41	-0.42
Amygdala	-0.34	-0.39	-0.18	-0.33
Anterior Limb Of Internal Capsule	0.34	0.46	0.17	0.28
Hippocampus	-0.30	-0.43	-0.24	-0.15
Lingual Gyrus	-0.30	-0.29	-0.47	-0.23
Middle Temporal Gyrus	-0.29	-0.34	-0.45	-0.14

[1]Cohen's d. A negative value indicates reduced volume in schizophrenic patients relative to healthy controls.

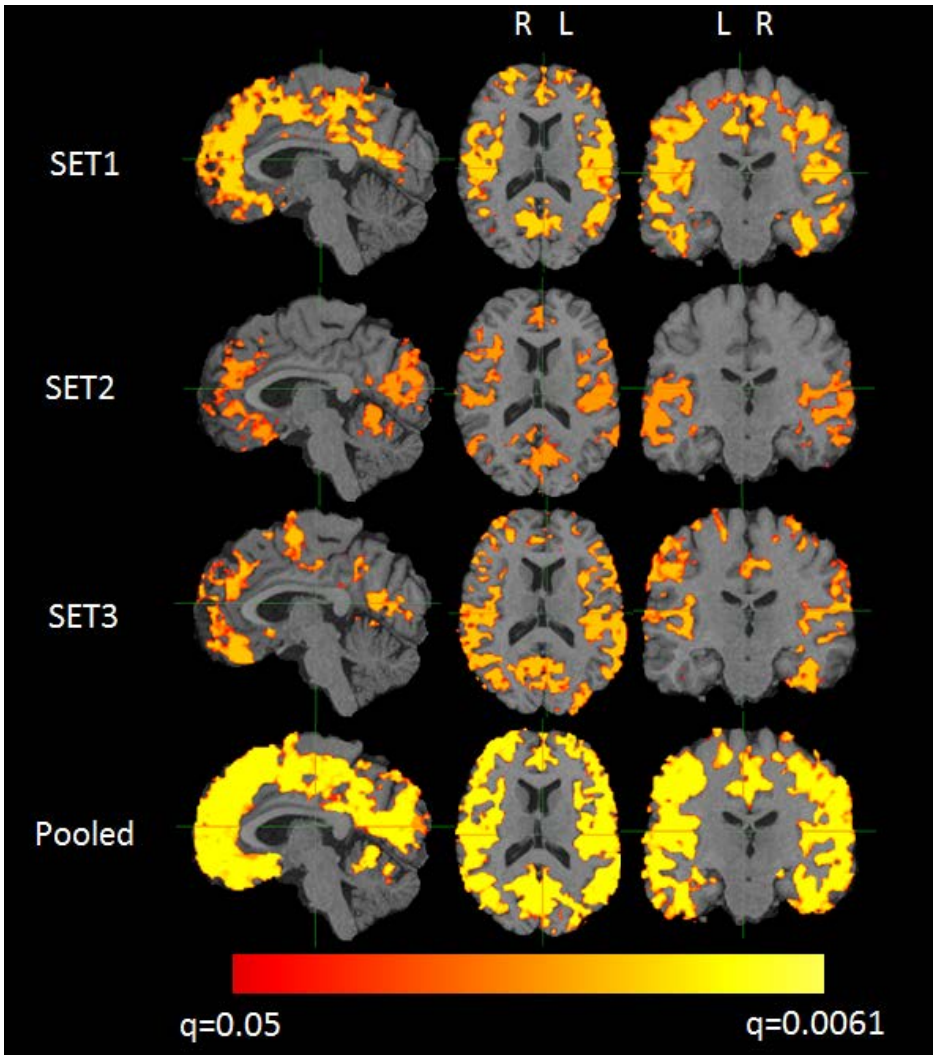
Supplementary Figure 1. Comparison of regions with significant group differences between controls and patients, for voxelwise versus regional analyses.

The maps show regions where significant gray matter differences were found between patients and controls, at the FDR-corrected threshold of q = 0.05. Voxelwise differences were found via ODVBA[9, 10] using GM RAVENS[6] maps. Regional differences were obtained by running univariate t-tests on regional segmentations provided by MUSE[1]. Regions shown in green were found to be significant in both analyses. In red areas only region-based analysis found significance, while blue areas were significant only in the ODVBA analysis. When interpreting the spatial specificity of these maps, please note that for the regional analysis results the significance level of an ROI was assigned uniformly to all voxels within that ROI.



Supplementary Figure 2. Voxelwise group comparisons between controls and patients using GM RAVENS[6] maps, for individual sites versus pooled data

Color maps show regions with significant group differences obtained via ODVBA[9, 10], corrected for multiple comparisons (FDR-corrected q<0.05). In the highlighted areas controls have higher RAVENS values, indicating higher GM volume.



Supplementary Table 4. Associations of disease duration and CPZ dosage equivalents with regional volumes

	Duration Illness		CPZ Dose Equivalents	
ROI	**correlation**	**q-Value [1]**	**correlation**	**q-Value [1]**
Frontal Operculum	-0.15	0.038*	-0.227	0.018*
Medial Orbital Gyrus	0.022	0.784	-0.209	0.018*
Middle Frontal Gyrus	-0.016	0.855	-0.203	0.018*
Pallidum	0.234	0.001*	0.203	0.018*
Subcallosal Area	-0.075	0.349	-0.203	0.018*
Temporal Lobe Wm	0.036	0.678	0.176	0.052
Central Operculum	-0.103	0.189	-0.174	0.052
Opercular Part Of The Inferior Frontal Gyrus	-0.168	0.018*	-0.164	0.074
Lateral Ventricle	0.205	0.005*	0.154	0.098
Anterior Limb Of Internal Capsule	0.024	0.773	-0.147	0.111
Angular Gyrus	-0.113	0.150	-0.145	0.111
Middle Cingulate Gyrus	0.088	0.282	-0.143	0.111
Postcentral Gyrus Medial Segment	-0.184	0.008*	-0.142	0.111
3rd Ventricle	0.126	0.103	0.141	0.111
Inf Lat Vent	0.006	0.947	0.139	0.114
Entorhinal Area	-0.015	0.855	0.133	0.133
Accumbens Area	-0.096	0.233	-0.128	0.151
Middle Temporal Gyrus	-0.191	0.007*	-0.124	0.168
Basal Forebrain	-0.037	0.673	-0.122	0.171
Posterior Orbital Gyrus	-0.109	0.173	-0.116	0.203
Superior Frontal Gyrus Medial Segment	0.024	0.773	-0.115	0.209
Postcentral Gyrus	-0.122	0.115	-0.111	0.219
Triangular Part Of The Inferior Frontal Gyrus	0.12	0.120	-0.111	0.219
Parietal Operculum	0.048	0.565	-0.108	0.229
Medial Frontal Cortex	0.007	0.942	-0.105	0.240
Lateral Orbital Gyrus	-0.073	0.349	-0.102	0.253
Lingual Gyrus	-0.104	0.189	-0.1	0.253
Anterior Orbital Gyrus	-0.059	0.479	-0.099	0.253
Posterior Insula	-0.118	0.125	-0.099	0.253
Planum Polare	0.003	0.987	-0.089	0.325
Supramarginal Gyrus	-0.056	0.512	-0.089	0.325
Superior Parietal Lobule	-0.079	0.347	-0.088	0.325
Inferior Occipital Gyrus	0.001	0.998	-0.083	0.346
Supplementary Motor Cortex	-0.071	0.362	-0.083	0.346
4th Ventricle	0.045	0.606	0.08	0.364
Anterior Insula	-0.074	0.349	-0.074	0.408
Superior Temporal Gyrus	0.037	0.673	-0.069	0.450
Precuneus	-0.165	0.019*	-0.064	0.489
Thalamus Proper	-0.022	0.784	-0.059	0.518
Hippocampus	-0.148	0.038*	-0.058	0.518
Posterior Cingulate Gyrus	-0.145	0.044*	-0.056	0.526
Amygdala	-0.096	0.233	0.054	0.526
Posterior Limb Of Internal Capsule Inc. Cerebral Peduncle	0.03	0.720	0.046	0.611

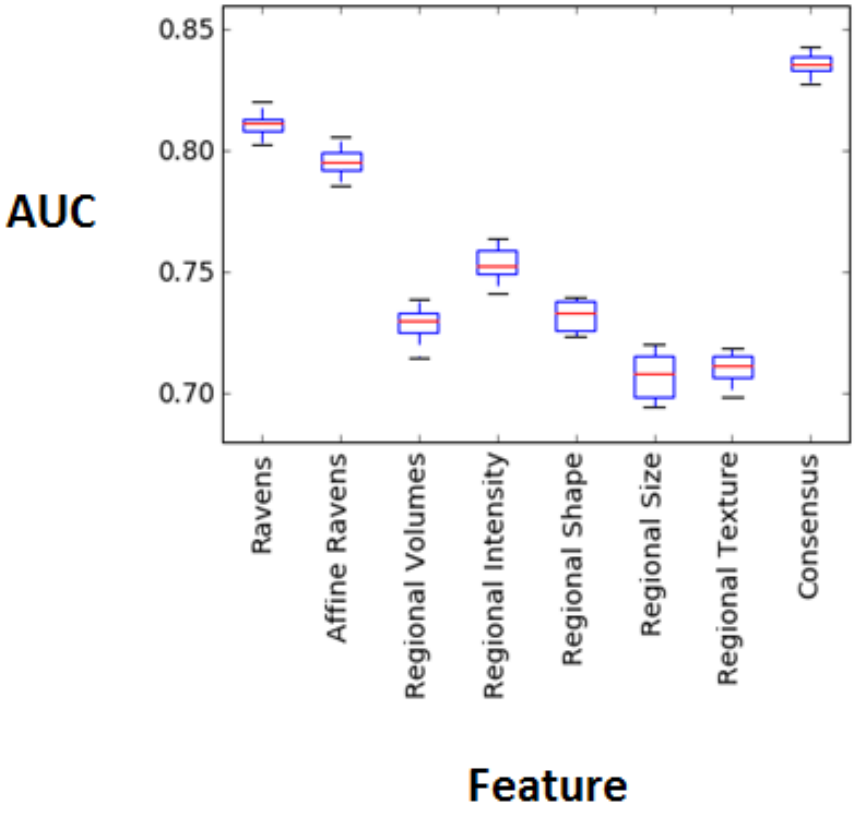
1 Correlations significant at FDR-corrected q<0.05 denoted with an asterisk.

Supplementary Table 4. Associations of Disease Duration and CPZ Dosage Equivalents with Regional Volumes (Continued)

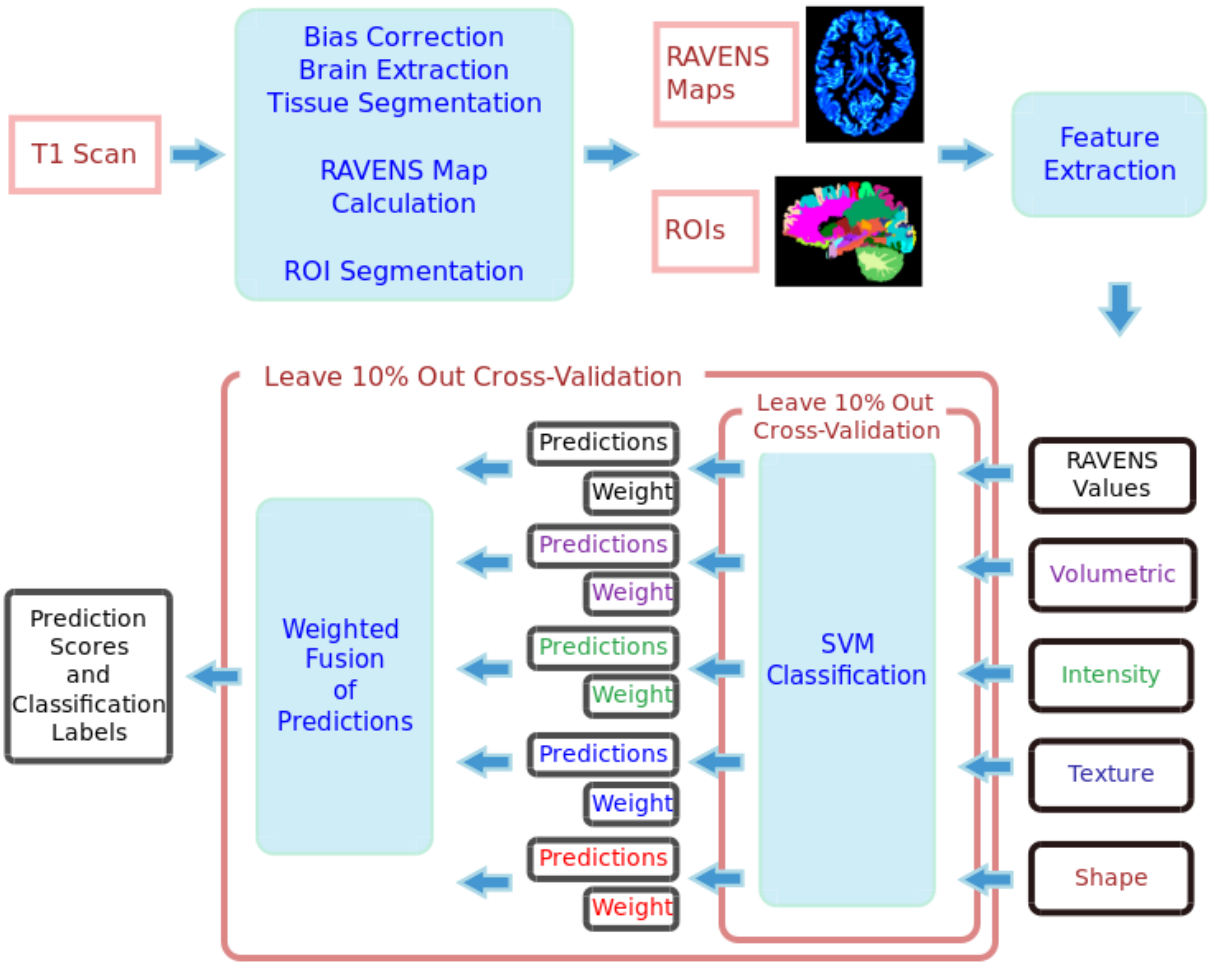
	Duration Illness		CPZ Dose Equivalents	
ROI	**correlation**	**q-Value [1]**	**correlation**	**q-Value [1]**
Calcarine Cortex	0.05	0.565	-0.043	0.641
Superior Frontal Gyrus	0.085	0.307	-0.042	0.641
Precentral Gyrus	-0.03	0.720	-0.041	0.651
Inferior Temporal Gyrus	-0.164	0.019*	-0.029	0.794
Planum Temporale	-0.083	0.317	-0.027	0.804
Cuneus	0.077	0.349	-0.026	0.804
Parahippocampal Gyrus	-0.193	0.007*	0.026	0.804
Putamen	0.089	0.282	0.022	0.838
Temporal Pole	-0.135	0.067	-0.022	0.838
Frontal Lobe Wm	0.04	0.673	0.018	0.882
CSF	0.186	0.008*	0.015	0.889
Precentral Gyrus Medial Segment	-0.054	0.518	0.015	0.889
Occipital Fusiform Gyrus	0.07	0.365	-0.014	0.889
Transverse Temporal Gyrus	0.026	0.759	0.012	0.889
Fusiform Gyrus	-0.076	0.349	0.009	0.897
Middle Occipital Gyrus	-0.069	0.378	-0.008	0.897

1 Correlations significant at FDR-corrected q<0.05 denoted with an asterisk.

Supplementary Figure 3. **Classification performance of individual feature sets and the ensemble classifier on the pooled data.** Each boxplot shows the result of running 10 separate leave-10%-out cross-validation experiments on a particular feature set. The consensus classifier was constructed by combining individual classifiers using each feature set through weighted voting. See Supplementary Methods for details.



Supplementary Figure 4. Outline of image processing and cross-validated classification using ensemble of features



Bibliography

1. Doshi J, Erus G, Ou Y, et al. MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *Neuroimage* Feb 15 2016;127:186-195.

2. Haralick RM, K. Shanmugam, I. Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics* 1973;3(6):610-620.

3. Haralick RM. Statistical and Structural Approaches to Texture. *Proceedings of the IEEE* 1979;67:786-804.

4. Harlow RWCaCA. A Theoretical Comaprison of Texture Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1980;2:204-222.

5. R.W. Conners MMT, C.A. Harlow. Segmentation of a High-Resolution Urban Scene using Texture Operators. *Computer Vision, Graphics and Image Processing* 184;25:273-310.

6. Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *Neuroimage* Dec 2001;14(6):1361-1369.

7. Ou Y, Sotiras A, Paragios N, Davatzikos C. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Med Image Anal* Aug 2011;15(4):622-639.

8. Yangming Ou AS, Aristeidis Sotiras. DRAMMS Distribution General Website; 2009.

9. Zhang T, Davatzikos C. ODVBA: optimally-discriminative voxel-based analysis. *IEEE Trans Med Imaging* Aug 2011;30(8):1441-1454.

10. Zhang T, Davatzikos C. Optimally-Discriminative Voxel-Based Morphometry significantly increases the ability to detect group differences in schizophrenia, mild cognitive impairment, and Alzheimer's disease. *Neuroimage* Oct 01 2013;79:94-110.