Cerebral Cortex, April 2016;26: 1440–1452

doi: 10.1093/cercor/bhu242
Advance Access Publication Date: 14 October 2014
Original Article

ORIGINAL ARTICLE

Anatomical Abnormalities in Autism?

Shlomi Haar[1], Sigal Berman[3], Marlene Behrmann[4], and Ilan Dinstein[1,2]

[1]Department of Brain and Cognitive Sciences, [2]Department of Psychology, [3]Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer Sheva 84105, Israel, and [4]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Address correspondence to Dr Ilan Dinstein. Email: dinshi@bgu.ac.il

Abstract

Substantial controversy exists regarding the presence and significance of anatomical abnormalities in autism spectrum disorders (ASD). The release of the Autism Brain Imaging Data Exchange (~1000 participants, age 6–65 years) offers an unprecedented opportunity to conduct large-scale comparisons of anatomical MRI scans across groups and to resolve many of the outstanding questions. Comprehensive univariate analyses using volumetric, thickness, and surface area measures of over 180 anatomically defined brain areas, revealed significantly larger ventricular volumes, smaller corpus callosum volume (central segment only), and several cortical areas with increased thickness in the ASD group. Previously reported anatomical abnormalities in ASD including larger intracranial volumes, smaller cerebellar volumes, and larger amygdala volumes were not substantiated by the current study. In addition, multivariate classification analyses yielded modest decoding accuracies of individuals' group identity (<60%), suggesting that the examined anatomical measures are of limited diagnostic utility for ASD. While anatomical abnormalities may be present in distinct subgroups of ASD individuals, the current findings show that many previously reported anatomical measures are likely to be of low clinical and scientific significance for understanding ASD neuropathology as a whole in individuals 6–35 years old.

Key words: anatomy, autism, MRI, thickness, volume

Introduction

Considerable effort has been devoted to the identification of anatomical abnormalities in individuals with autism spectrum disorder (ASD) (Courchesne et al. 2007; Amaral et al. 2008). In the current study, we analyzed anatomical data acquired from individuals older than 6 years of age for whom MRI scans are available in the Autism Brain Imaging Data Exchange (ABIDE) database. Previous studies of individuals in this age range have reported that, in comparison to controls, ASD individuals exhibit numerous abnormalities including larger gray matter (Lotspeich et al. 2004; Hazlett et al. 2006; Ecker et al. 2013), white matter (Hazlett et al. 2006), amygdala (Bellani et al. 2013a), and hippocampus (Groen et al. 2010) volumes, smaller cerebellum (Scott et al. 2009; Fatemi et al. 2012) and corpus callosum (CC; Bellani et al. 2013b) volumes, and abnormal cortical thickness (Raznahan et al. 2010; Wallace et al. 2010). These findings have been interpreted as supporting evidence for different theories of ASD including, for example, the "amygdala theory of autism" (Baron-Cohen et al. 2000) and the "underconnectivity" theory of ASD (Just et al. 2007). These findings, however, have not been replicated consistently in the literature and heterogeneous results across studies with small samples of participants have demonstrated a critical need for analyzing larger cohorts (Amaral et al. 2008).

More recent anatomical studies have also utilized multivariate classification techniques to identify patterns of anatomical measures that differ across ASD and control individuals instead of focusing on just one measure at a time. These studies have reported remarkable accuracies in decoding the group identity of single subjects (above 85%) when utilizing measures of cortical thickness, geometry, curvature, and/or surface area (Ecker, Marquand et al. 2010; Jiao et al. 2010; Uddin et al. 2011), thereby implicitly suggesting that anatomical measures may have clinical diagnostic value for ASD. While these initial results seem promising, it is important to note that previous studies sampled data from only 20–30 subjects in each group and

reported successful classification with entirely different sets of anatomical measures. The lack of consistency across studies raises critical questions regarding the generalizability and reproducibility of findings to larger cohorts.

The release of the ABIDE data exchange (Di Martino et al. 2014) offers a unique opportunity to perform more definitive anatomical comparisons by analyzing aggregated MRI data, collected in 20 different international studies, from ~1000 ASD and control participants, ages 6–65 years. This dataset, which is an order of magnitude larger than any anatomical dataset published to date, offers substantial statistical power and the ability to dissociate anatomical variability related to age, IQ, and scanning site from true anatomical differences across control and autism groups. As such, this dataset offers an unprecedented opportunity to determine the existence (or lack thereof) of anatomical abnormalities in children, adolescents, and adults with ASD.

Materials and Methods

The data analyzed in the current study are part of the ABIDE database (http://fcon_1000.projects.nitrc.org/indi/abide). All data are fully anonymized as required by HIPAA regulations and all participating sites received local Institutional Review Board approval for acquisition of the contributed data.

Subjects

The ABIDE database contains aggregated MRI scans of 539 ASD individuals and 573 typical controls aged 6–65 years who were scanned as part of 20 international studies. We performed analyses on 2 samples from the database.

Strict Sample

Here we excluded ASD subjects older than 35 (21 subjects) and all females (65 subjects) given their low preponderance. Individuals without IQ scores (34 subjects), those with low-quality MRI scans (26 subjects, see below), and those with gross anatomical volumes that exceeded 3 standard deviations from the mean of the site (8 ASD and 8 control subjects in total) were also excluded. The remaining 401 ASD subjects were matched to controls within the same site based on age (±5 years) and IQ (±10 points). ASD subjects for whom there was no matched control were excluded. Final analyses were performed with 295 ASD and 295 control subjects (Table 1).

Relaxed Sample

Here we excluded subjects with low-quality MRI scans and equated the number of ASD and control subjects within each site by selecting control subjects that were closest in age to the ASD subjects. Left over subjects from sites with an unequal number of ASD and control subjects were excluded. Final analyses were performed with 453 ASD and 453 control subjects (Supplementary Table 1). In this sample subjects were NOT matched with respect to gender, IQ, or age.

MRI Scans

All MRI data were acquired using 3-Tesla scanners with T_1 weighted scans (1 × 1 × 1-mm resolution). The specific scanning parameters of each sample are available at http://fcon_1000.projects.nitrc.org/indi/abide.

Brain Segmentation

Cortical reconstruction and volumetric segmentation was performed with Freesurfer (http://surfer.nmr.mgh.harvard.edu/). The technical details of these procedures are described elsewhere (Fischl 2012). Briefly, this processing includes removal of nonbrain tissue and segmentation of subcortical and cortical gray and white matters based on image intensity. Low-quality MRI scans where the segmentation procedures failed or showed geometric inaccuracies were excluded (26 ASD subjects).

ROI Parcellation

Parcellation was also performed using Freesurfer. Individual brains were registered to a spherical atlas which utilized individual cortical folding patterns to match brain geometry across subjects. Each brain was then parcellated into 148 gray matter and 32 subcortical ROIs using the Destrieux anatomical atlas (Destrieux et al. 2010). Finally, ROI labels were transformed back into each subject's native space to compute the volume of each ROI.

Cortical Thickness and Surface Area

Both measures were computed using Freesurfer by estimating the gray/white matter boundary and the pial surface (Fischl and Dale 2000). Cortical thickness was computed for each triangular vertex as the distance from one surface to the other. The number of vertices varied across subjects according to individual cortical surface area. ROI thickness was computed by averaging the vertices belonging to each ROI in each subject's native anatomical space. Cortical surface area was computed per ROI using 2D flattened representation of the cortical surface.

Regressing Out Site Age and IQ

When performing the mixed-model analysis using group and site as main factors and age and IQ as covariates, we found that site, age, and IQ had significant effects on most anatomical measures. To eliminate differences due to these confounding factors we regressed out the effects of site, age, and IQ from each of the ROIs for each of the anatomical measures separately. This was done by performing a multiple regression analysis of the anatomical data (including both ASD and control subjects) with the 3 variables (site, age, and IQ) and extracting the residuals. Since anatomical measures may change in a linear (e.g., gray matter volume), exponential (e.g., white matter volume), or quadratic (e.g., atrophy at older ages) manner with age, we included a linear, an exponential, and a quadratic predictor for age in the multiple regression analysis. This analysis was performed for each ROI separately.

Classification

We used linear and quadratic (nonlinear) discriminant analyses implemented in MATLAB to classify ASD and control subjects according to anatomical measures. Discriminant analysis is a supervised multivariate classification method where each subject is represented by a vector in a high dimensional space defined by the number of selected features (anatomical ROIs). Training data were classified into 2 classes (ASD and controls) by identifying a multidimensional hyperplane in the linear case and a multidimensional quadratic surface in the nonlinear case that optimally separates the 2 classes. The accuracy of the hyperplane/surface was tested by assessing its ability to separate independent "testing data" that was not included in the "training data" (see cross-validation below). Classification was applied

Table 1 Strict sample phenotypic information

Site	Scanner	*n*		Age				IQ				ADOS	
		ASD	TD	ASD		TD		ASD		TD		ASD	
				Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CALTECH	SIEMENS Trio	5	5	22.6	4.2	22.1	3.7	113	11.1	116	6.1	10.8	2.7
CMU	SIEMENS Verio	8	8	24.9	4.1	25.8	4.7	110	11.6	109	6.1	13.0	3.2
KKI	Philips Achieva	10	10	10.4	1.4	10.4	1.4	107	13.3	111	11.4	10.6	1.8
Leuven 1	Philips INTERA	12	12	20.4	1.9	22.7	2.6	110	12.5	113	10.5		
Max Munich	SIEMENS Verio	13	13	19.9	9.2	20.1	8.9	109	9.7	109	8.1	9.8	3.9
NYU	SIEMENS Allegra	56	56	13.6	5.9	14.3	4.8	110	16.5	112	13.2	11.3	4.1
OHSU	SIEMENS Trio	8	8	11.1	2.1	10.0	1.1	117	12.6	115	10.1	8.6	3.7
OLIN	SIEMENS Allegra	13	13	16.2	3.0	17.4	3.4	114	16.2	114	17.1	14.2	3.7
PITT	SIEMENS Allegra	18	18	18.3	6.0	17.4	5.3	109	13.7	110	9.4	12.8	3.0
SDSU	GE MR750	8	8	14.6	1.7	14.3	1.3	113	9.5	113	7.5	10.6	4.1
Stanford	GE Signa	9	9	9.5	1.6	9.4	1.3	113	13.1	114	11.7	11.9	3.8
Trinity	Philips Achieva	22	22	16.7	3.0	16.4	3.1	111	13.3	111	12.4	10.6	2.9
UCLA 1	SIEMENS Trio	27	27	13.2	2.6	13.3	2.2	104	13.0	105	10.4	10.5	3.3
UCLA 2	SIEMENS Trio	5	5	12.3	1.9	12.2	1.1	105	9.4	108	8.8	13.5	3.1
UM 1	GE Signa	26	26	12.8	2.4	13.1	3.2	105	15.9	108	9.8		
UM 2	GE Signa	11	11	14.6	1.6	15.6	1.7	114	13.0	112	9.1		
USM	SIEMENS Trio	30	30	21.5	5.7	21.1	5.9	107	13.9	110	10.3	12.9	2.7
Yale	SIEMENS Trio	14	14	12.6	3.0	12.2	2.8	99	18.4	100	16.5		

Number of subjects, mean and standard deviation of age, IQ scores, and ADOS scores (total communication and social scores from ADOS modules 3 and 4 only) are presented for each site along with the type of MRI scanner used at the site. Note that ADOS testing was not performed at 4 of the sites (63 subjects with ASD).

separately to five different anatomical feature sets: all ROI volumes (180 features), subcortical ROI volumes (32 features), cortical ROI volumes (148 features), thicknesses (148 features), and surface areas (148 features).

10-Fold Cross-validation

The data were randomly split into 10 even-sized subsets, which included an even number of subjects from ASD and control groups. Nine subsets (90% of the data) were used to train the classifier and the remaining subset (10%) was used to test classifier accuracy by assessing the proportion of subjects that were accurately decoded. This process was iterated 10 times such that each of the subsets was used once for "testing" and the decoding accuracies were averaged across iterations.

Leave-Two-Out Cross-validation

Here we used the same procedure as the "10-fold" validation but left out 2 exemplars (one with ASD and one control) instead of 10% of the data from each group for "testing" and the number of iterations equaled $n/2$.

Randomization Test

We used a randomization test to determine statistical significance in the different analyses. In the univariate analyses, we randomly shuffled the ASD and control labels across subjects and computed the difference in the particular measure (e.g., intracranial volume) across groups. This process was repeated 10 000 times to generate 10 000 difference values that represented a distribution of differences expected by chance (null distribution). For the true (un-shuffled) value to be considered significant, it had to surpass the 2.5th or 97.5th percentile of the null distribution (i.e., the equivalent of a $P < 0.05$ value in a two-tailed t-test). A separate randomization analysis was performed for each anatomical measure. In the multivariate analyses, we similarly shuffled the ASD and control labels and then performed identical classification and cross-validation procedures to those described above. This process was repeated 100 times to generate 100 decoding accuracy values that represented a distribution of accuracies expected by chance (null distribution). A separate randomization analysis was performed for each classification analysis with each set of anatomical features. For the real decoding accuracy to be considered statistically different from chance levels, it had to exceed the 95th percentile of the relevant null distribution.

Multiple Comparisons

We used false discovery rate (FDR) correction (Benjamini and Hechtlinger 2014) to control for the multiple comparisons problem when assessing anatomical differences across multiple ROIs (e.g., Fig. 2) . This correction is more relaxed than the Bonferroni method, and we used it in order to increase the sensitivity of the analyses so as not to miss any potential differences across groups.

Results

While the ABIDE database contains data from ~1000 individuals, behavioral scores (i.e., IQ, ADOS, ADI) were not available for all subjects and age and gender were not sufficiently matched within all sites. We, therefore, decided to perform the majority of our analyses with a subset of 590 participants (half controls) who were selected using the following stringent criteria. Since 98% of the subjects were under the age of 35 and 85% were male, we selected only males 6–35 years old. An equal number of ASD and control subjects were sampled from each site and each ASD subject was matched with a control subject on IQ and age within the site. This selection ensured that there were no major group differences on multiple demographic, biographic, and

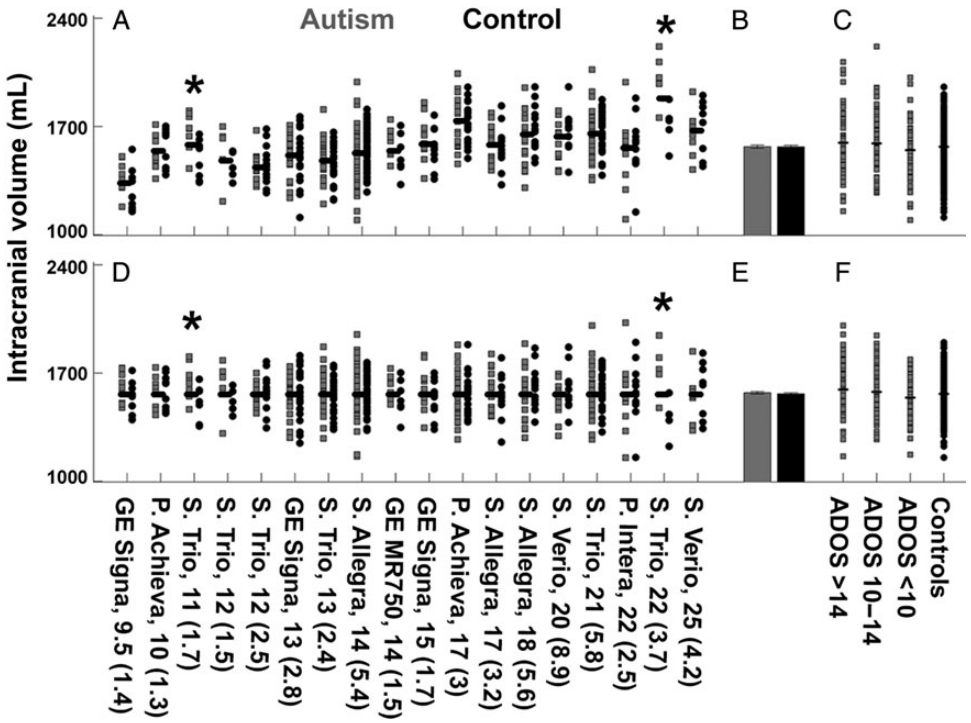

Figure 1. Scatter plot of intracranial volumes of individuals with ASD (gray) and controls (black) in each site (*A*). Sites are ordered according to age. Mean (SD) age and scanner type are noted for each site (GE, General Electric; P, Phillips; S, Siemens). Bars represent the mean intracranial volume of each group across sites (*B*). An additional scatter plot (*C*) represents intracranial volumes after separating ASD subjects into subgroups with ADOS <10, 10–14, and >14. Panels on bottom row show equivalent analyses after regressing out variability associated with site, age, and IQ (*D*–*F*). Asterisks: significant differences ($P < 0.05$, randomization test, FDR corrected for multiple comparisons). Horizontal black lines: mean across site (*A*, *D*) or group (*C*, *F*). Error bars: standard error of the mean across subjects.

environmental variables that may have varied across international sites. In addition, we repeated the analyses with larger sets of subjects while using more relaxed sampling criteria and found equivalent results to those described below. To demonstrate this, we present results from a larger sample containing 906 participants who were not matched with respect to age, gender, or IQ (Supplementary Figs 6 and 7).

Intracranial Volume

To determine whether there were significant differences in intracranial volumes across the ASD and control groups, we performed a mixed-model analysis (MANCOVA) with 2 main factors (group and site) and 2 covariates (age and IQ). This analysis revealed significant differences across groups ($P = 0.04$) and sites ($P < 0.001$), which co-varied significantly with age and IQ ($P < 0.002$), and a significant interaction between site and group ($P = 0.02$). Since the interaction between group and site was significant, we performed an additional analysis per site and assessed whether differences between groups were consistent across sites (Fig. 1*A*). The results showed that significant differences across groups were found in only 2 of the 18 sites, which contained only 26 of the 590 (4.4%) subjects. Removing these sites from the analysis eliminated the significant difference across groups and the significant interaction between group and site ($P > 0.1$ for both) thereby demonstrating that initial significance was driven by the small subset of subjects from these 2 sites.

The results of the mixed-model analysis demonstrated that intracranial volumes varied significantly across sites, ages, and IQs. Since these are confounding factors in comparisons of anatomy across control and ASD groups, we reanalyzed the data after regressing out the variability associated with each factor. This normalization procedure eliminated mean volume differences across sites, ages, and IQ levels, but did not affect the differences between autism and control groups within each site. Hence the 2 sites with significant intracranial volume differences across groups before normalization also exhibit significant differences after normalization (Fig. 1*D* and Supplementary Fig. 1). Performing an independent *t*-test or randomization test on the normalized data showed that there were no significant differences across ASD and control groups ($P > 0.5$, independent *t*-test and randomization test, Fig. 1*B*,*E*) and demonstrated that the negative finding was not due to variability associated with these confounding variables.

In a final analysis, we also stratified the ASD individuals into subgroups with weak (ADOS <10), moderate (ADOS = 10–14), and severe (ADOS >14) symptoms as measured by the total social and communication ADOS scores. Both *t*-tests and randomization tests did not reveal any significant differences between each of the three ASD severity groups and their matched control groups ($P > 0.25$ for all comparisons, Fig. 1*C*,*F*).

Gross Anatomy and ROIs of Special Interest

Equivalent mixed-model analyses were performed separately for each of the following measures: white matter volume, gray matter volume, cortical thickness, and cortical surface area as well as cerebellar gray matter, cerebellar white matter, amygdala, hippocampus, CC, and ventricular volumes. All volumetric measures were first translated to percentages of intracranial volume in order to compensate for differences in intracranial volumes

across subjects. Here too, several measures exhibited significant differences across groups, but also significant interactions between group and site. We, therefore, normalized differences across sites, ages, and IQ levels and re-examined group differences using independent *t*-tests and randomization tests (Fig. 2).

These analyses revealed that only ventricular volume was significantly larger in the autism group when compared with controls ($P < 0.001$, randomization test, FDR corrected). All other anatomical measures did not differ significantly across groups ($P > 0.1$). A "per-site" analysis showed that the vast majority of sites did not exhibit significant differences across groups for any of the measures. In the case of cortical gray matter, 2 sites exhibited contradictory findings. Effect sizes across groups were small with the largest one evident in the comparison of ventricular volumes ($d = 0.34$).

Stratifying the ASD individuals into subgroups with weak (ADOS <10), moderate (ADOS = 10–14), and severe (ADOS >14) symptoms did not reveal any additional differences across groups. None of the autism severity groups showed significant differences from the control group for any of the anatomical measures including the ventricular volumes ($P > 0.07$, randomization test, FDR corrected, Supplementary Fig. 2). This shows that the difference in ventricular volumes across groups is relatively weak and does not appear reliably when sampling smaller subgroups of autism and control subjects.

Finally, we also examined whether there were volumetric group differences in specific segments of the CC. The CC of each subject was segmented into 5 equal parts along the anterior–posterior axis and the volume of each part was calculated (Fig. 3). The ASD group exhibited significantly smaller volumes in the central CC segment when compared with the control group ($P < 0.05$, randomization test, FDR corrected). Note that the effect size of this finding was small ($d = 0.2$) and that significant differences appeared in only 2 of the 18 sites.

Additional Cortical and Subcortical Areas

Univariate comparisons across groups were performed for each of 148 cortical and 32 subcortical ROIs that were defined in an automated manner for each subject using the Freesurfer atlas (see Materials and Methods). Volumetric measures were normalized by total intracranial volume and variability associated with site, age, and IQ was regressed out for each measure separately. Analysis of the cortical ROIs revealed that individuals with ASD exhibited significantly thicker cortex than controls in several ROIs including the right and left occipital poles, left middle occipital sulcus, left occipital-temporal sulcus, left cuneus gyrus, right and left subparietal sulci, and left superior temporal gyrus and sulcus ($P < 0.05$, independent *t*-test, FDR corrected, Fig. 4). There were no significant differences in volumetric or surface area measures between groups. There were also no significant volumetric differences between groups in any of the subcortical ROIs.

Correlations with IQ and ADOS

Intracranial, gray, and white matter volumes as well as cortical surface were positively and significantly correlated with IQ in both control and ASD groups ($P < 0.05$, randomization test, uncorrected to increase sensitivity, Fig. 5, top and middle rows). Individuals with ASD also exhibited significant positive correlations between gray and white matter volumes and cortical surface area measures and ADOS scores (Fig. 5, bottom row). Note that all correlations were computed after regressing out site and age as performed in the analyses described above.

Classification

We used a linear discriminant analysis (LDA) classifier to decode the group identity of individual subjects (Fig. 6). A 10-fold validation scheme was used to assess the decoding accuracy of the classifier such that 90% of the data were used to train the classifier and the left-out 10% was used to assess decoding accuracy. This procedure was performed 10 times, each time leaving out a different subset of subjects for testing. A randomization analysis was used to determine whether decoding accuracies exceeded chance levels (see Materials and Methods). Classification was performed using 5 anatomical feature sets: all ROI volumes (180 ROIs), only gray matter ROI volumes (148 ROIs), only subcortical ROI volumes (32 ROIs), cortical thickness (148 ROIs), and cortical surface area (148 ROIs).

Above-chance decoding accuracies were found when classifying all ASD and control subjects using subcortical volumes and cortical thickness measures (accuracies = 56% and 60%, respectively, $P < 0.05$, uncorrected to increase sensitivity). No other anatomical measures yielded significant, above-chance decoding accuracies. When performing the same analysis separately on ASD subgroups with different symptom severities, above-chance decoding accuracies were found only for the ASD subgroup with severe symptoms using subcortical volumes (accuracy = 60%, $P < 0.05$, uncorrected).

Additional analyses using a "leave-two-out" validation scheme with an LDA classifier or a 10-fold validation scheme with a quadratic discriminant analysis classifier (i.e., nonlinear classifier) revealed similar decoding accuracies that did not exceeded 60% and exceeded chance level only with subcortical volumes and cortical thickness measures (Supplementary Fig. 3).

Finally, we performed a per-site classification analysis within the 5 sites that contained at least 20 subjects in each group (Supplementary Fig. 4). This analysis revealed variable and weak decoding accuracies across sites, which did not exceed 60% for any of the sites when using any of the anatomical measures. This suggests that the poor decoding accuracies found across sites were not due to between-site variability, but rather to the lack of consistent separation across ASD and control individuals.

Classification of Different Group Sizes

We evaluated our classification procedures on smaller, randomly selected groups of subjects ($n = 20$, $n = 50$, $n = 100$, $n = 150$, and $n = 200$) to characterize the relationship between group size and the distribution of decoding accuracies. Randomly selected subgroups of ASD individuals were age and IQ matched to control individuals and the LDA classification analysis was performed using the 10-fold and leave-two-out validation schemes (Fig. 7). This procedure was repeated 100 times for each group size while selecting different subjects each time, thereby yielding a distribution of decoding accuracies for each group size. Smaller group sizes yielded wider decoding accuracy distributions regardless of the anatomical measures used to perform the classification. For example, performing this analysis with 148 cortical surface area measures revealed that ~30% of the randomly selected groups containing 20 subjects in each group (marked with a star) exhibited decoding accuracies >60%. Note that this effect was more prominent in the leave-two-out validation scheme and that the real accuracy computed on the entire ABIDE data in this case was indistinguishable from chance level (51%). Since

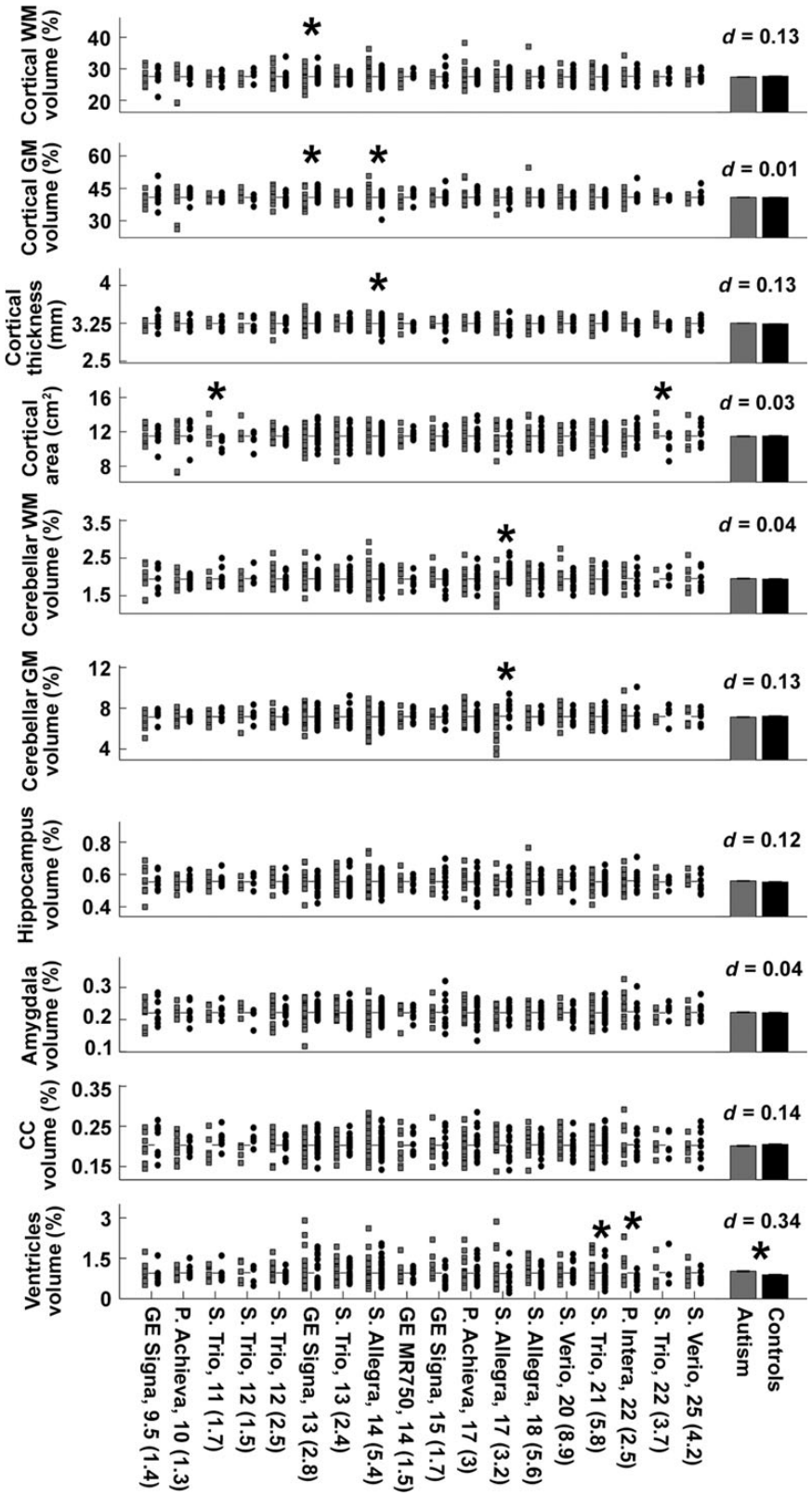

Figure 2. Scatter plots of cortical white matter volume, cortical gray matter volume, cortical thickness, cortical surface area, and cerebellar white matter, cerebellar gray matter, hippocampus, amygdala, corpus callosum (CC), and ventricular volumes in individuals with ASD (gray) and matched controls (black) separately for each site. Bars: means across sites for each subject group. Asterisks: significant differences ($P < 0.05$, randomization test, FDR corrected). Error bars: standard error of the mean across subjects. d = Cohen's d (effect size across groups).

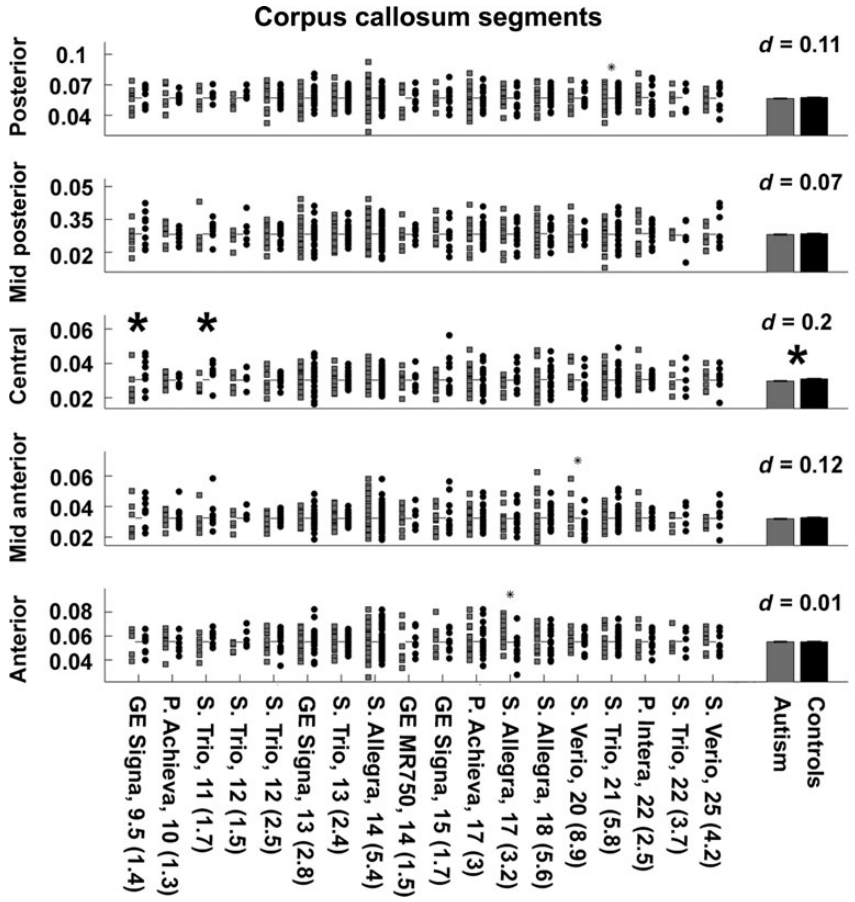

Figure 3. Scatter plots of corpus callosum segment volumes in individuals with ASD (gray) and matched controls (black) separately for each site. Bars: means across sites for each subject group. Asterisks: significant differences ($P < 0.05$, randomization test, FDR corrected). Error bars: standard error of the mean across subjects. d = Cohen's d (effect size across groups).

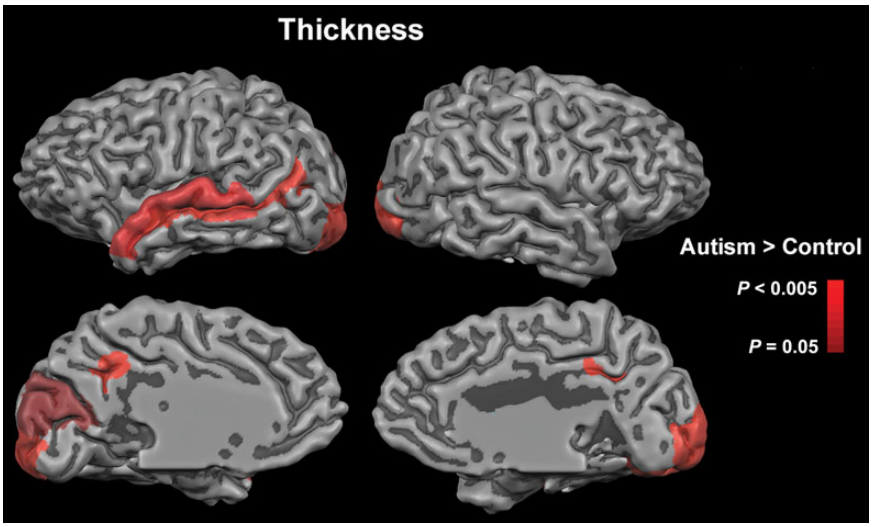

Figure 4. Cortical ROIs that were significantly thicker in the ASD group compared with the control group. There were no significant differences in the opposite comparison. Significance was assessed using a t-test and FDR was used to correct for multiple comparisons.

there are multiple feature sets that can be chosen, the use of small groups and the choice of an arbitrary feature set holds the potential of greatly inflating decoding accuracies.

Age and Data Sampling Effects

To assess the effects of age on the examined anatomical measures, we also computed growth curves for each of the gross

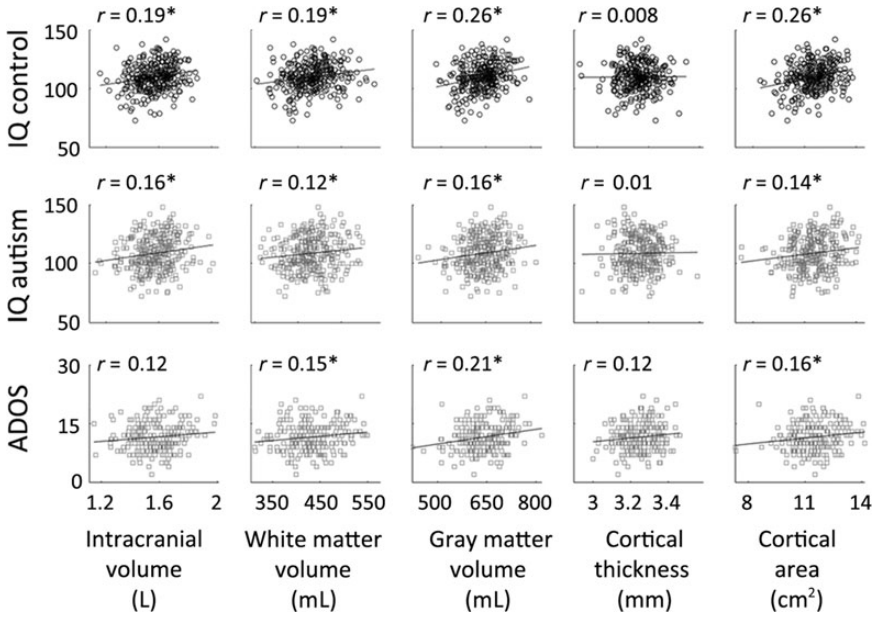

Figure 5. Correlation of intracranial volume, white matter volume, gray matter volume, cortical thickness, and cortical surface area, with IQ in the control (top row) and autism (middle row) groups or with ADOS (bottom row). Asterisks: significant correlation ($P < 0.05$, randomization analysis, uncorrected to increase sensitivity).

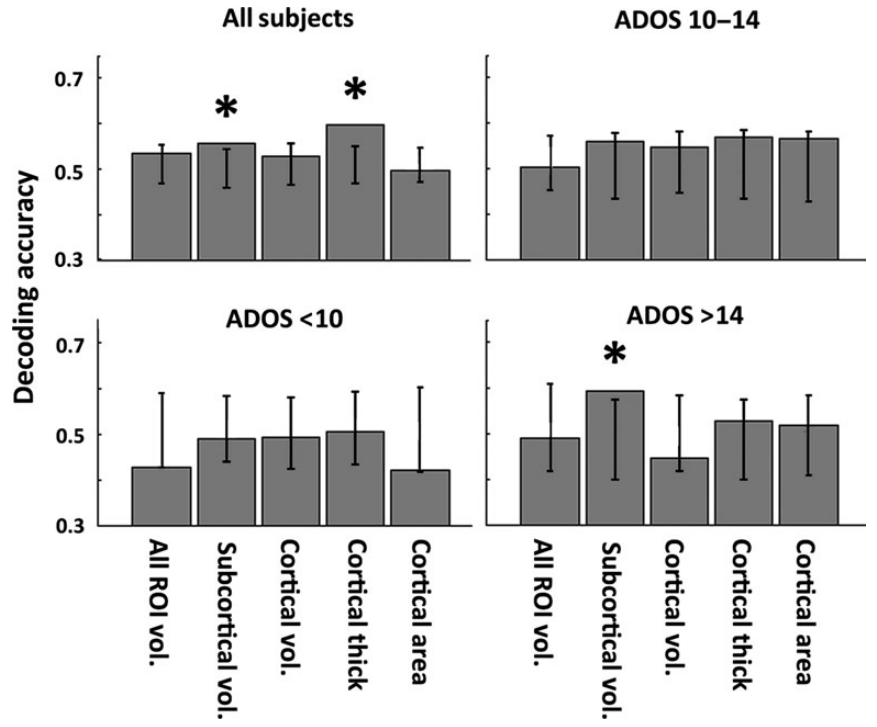

Figure 6. Decoding accuracies. Classification of group identity was performed separately for all subjects and three ASD subgroups who had ADOS scores <10, ADOS = 10–14, or ADOS >14. Bars show decoding accuracies for each of the following feature sets: all 180 ROI volumes, 32 subcortical ROI volumes, 148 cortical ROI volumes, 148 cortical ROI thicknesses, and 148 cortical ROI surface areas. Error bars: 5th and 95th percentiles of chance accuracy distribution as estimated with a randomization analysis. Asterisks: significant decoding accuracies ($P < 0.05$, uncorrected to increase sensitivity).

anatomy measures in each of the groups. We used a linear, an exponential, and a quadratic term to fit the data of the ASD and control subjects separately. The use of the 3 predictors enabled us to avoid making any assumptions regarding the shape of the fitted curve. The estimated parameters did not differ significantly across groups ($P > 0.3$, randomization test). We, therefore, performed a multiple regression analysis for the entire sample (ASD and controls) and regressed out the variability associated with age, effectively flattening the growth curves of both groups (Supplementary Fig. 5), and performed the analyses described above using the residuals of this regression analysis.

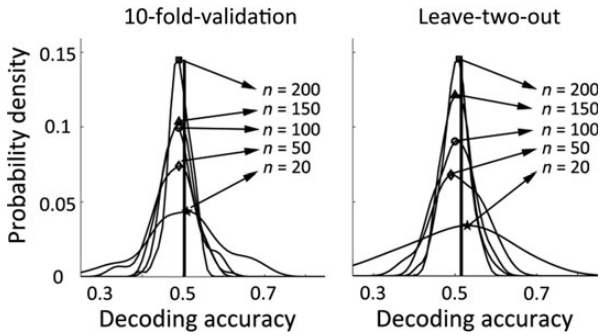

Figure 7. Decoding accuracy distributions when performing classification analyses with random samples of 20 (star), 50 (diamond), 100 (circle), 150 (triangle), and 200 (square) subjects in each group using cortical surface area measures. Black line: decoding rate for entire ABIDE data with 10-fold (left) and leave-two-out (right) validation schemes.

In a final set of analyses, we demonstrate that similar findings to those reported above are apparent when relaxing inclusion criteria so as to include the vast majority of viable MRI scans available in the ABIDE database. Here, we selected 906 subjects (half with ASD) without matching ASD and control subjects with respect to gender, IQ or age. Our only sampling criterion here was to ensure that the number of ASD and control subjects was equal at every site. Equivalent comparisons of the gross anatomical measures and ROIs of special interest (after regressing out the variance associated with site and age) revealed that the ASD group exhibited significantly smaller white matter volume, smaller cerebellar gray matter volumes, larger cortical thickness, and larger ventricles volume than the control group (Supplementary Fig. 6). Here too, effect sizes were small (cohen's $d \leq 0.32$ for all measures) with large between-subject variability apparent in both groups. An equivalent classification analysis using the LDA classifier and the 10-fold validation scheme showed that decoding accuracies were significantly above-chance when using all ROI volumes, subcortical volumes, and cortical thickness measures ($P < 0.05$, randomization analysis, uncorrected), yet all decoding accuracies were below 60% (Supplementary Fig. 7). Note that the slightly larger between-group differences apparent in this larger sample may have been driven by age, IQ, and gender differences across ASD and control groups that were not equated in this sample.

Choice of Segmentation Tool

The analyses above were performed using anatomical measures that were computed with the Freesurfer segmentation and parcellation algorithms. We also performed an additional analysis using the FAST tissue-type segmentation algorithm, which is part of the FSL toolbox (Jenkinson et al. 2012). This algorithm is limited in comparison to Freesurfer and enables extraction of only 3 anatomical measures: total gray matter (cortical and cerebellar), white matter (cortical and cerebellar), and CSF volumes. Comparisons of these measures across the 2 groups showed, again, no significant differences across groups and large within-group variability (Fig. 8).

Discussion

Analyses of anatomical MRI scans from the ABIDE database revealed several weak yet significant differences across ASD and control groups. Of the 180+ examined ROIs, significant differences were found only in ventricular volumes, the central segment of the CC, and in cortical thickness measures of several occipital and temporal ROIs (Figs 2 and 4). There was no evidence for between-group differences in any measures of gross anatomy or in specific brain regions including the amygdala, hippocampus, most segments of the CC, and the cerebellum, which have been implicated in previous anatomical studies of ASD (Figs 1–3). These results suggest that many of the previously reported anatomical abnormalities are likely to be of low scientific and clinical significance for explaining ASD neuropathology as a whole in individuals 6–35 years of age. Instead, it may be more useful to divide the heterogeneous ASD population into genetically, clinically, and/or behaviorally homogeneous subgroups and attempt to identify unique anatomical abnormalities in each.

Attempts to decode the group identity of single subjects with multivariate classification techniques using different sets of anatomical measures revealed remarkably weak decoding accuracies. While decoding accuracies using volumetric measures of subcortical areas and cortical thickness measures were significantly above-chance level, all accuracies were below 60% (Fig. 6 and Supplementary Figs 3 and 7). Poor decoding accuracies demonstrate that weak anatomical differences are apparent only at the group level and offer very limited diagnostic value at the single-subject level, likely due to the considerable anatomical variability across subjects of each group.

Taken together, our results suggest that individuals with ASD ages 6–35 years old who are capable of participating in MRI studies (i.e., high-functioning individuals with IQ >70), present anatomical profiles that are mostly indistinguishable from those of control individuals. Small differences across groups were greatly overshadowed by considerable within-group variability apparent in both ASD and control groups (Figs 1–3). While consistent anatomical abnormalities may be evident in toddlers with ASD during early development (Courchesne et al. 2004, 2007; Amaral et al. 2008; Stanfield et al. 2008), our results suggest that these abnormalities are not apparent in children, adolescents, and adults with ASD despite their persisting behavioral problems.

Gross Anatomical Findings in ASD

Numerous studies have reported that toddlers aged 1–4 years old who are later diagnosed with ASD exhibit gross anatomical differences including increased head circumferences and intracranial volumes in comparison to control toddlers (Courchesne et al. 2001, 2004, 2011; Amaral et al. 2008; Shen et al. 2013). In contrast to the studies in toddlers, there is considerable controversy regarding the presence and significance of gross anatomical abnormalities in individuals with ASD above the age of 6 years. While some studies have reported that individuals with ASD exhibit increased gray (Lotspeich et al. 2004; Palmen et al. 2004; Hazlett et al. 2006; Ecker et al. 2013) and white (Hazlett et al. 2006) matter volumes, others have reported no group difference (Ecker, Rocha-Rego et al. 2010; Ecker et al. 2012). Additional studies have reported that individuals with ASD exhibit significantly thicker cortices (Hardan et al. 2006) while others have reported the opposite (Scheel et al. 2011). The definitive findings reported here show that intracranial, gray matter, and white matter volumes, overall cortical thickness, and surface area measures are not significantly different across groups (Figs 1 and 2). These results suggest that persistent ASD behavioral symptoms are not ubiquitously associated with gross anatomical abnormalities in 6–35 year old individuals with ASD, regardless of their symptom severity (Supplementary Fig. 2). The transition from positive to negative findings when examining toddlers versus

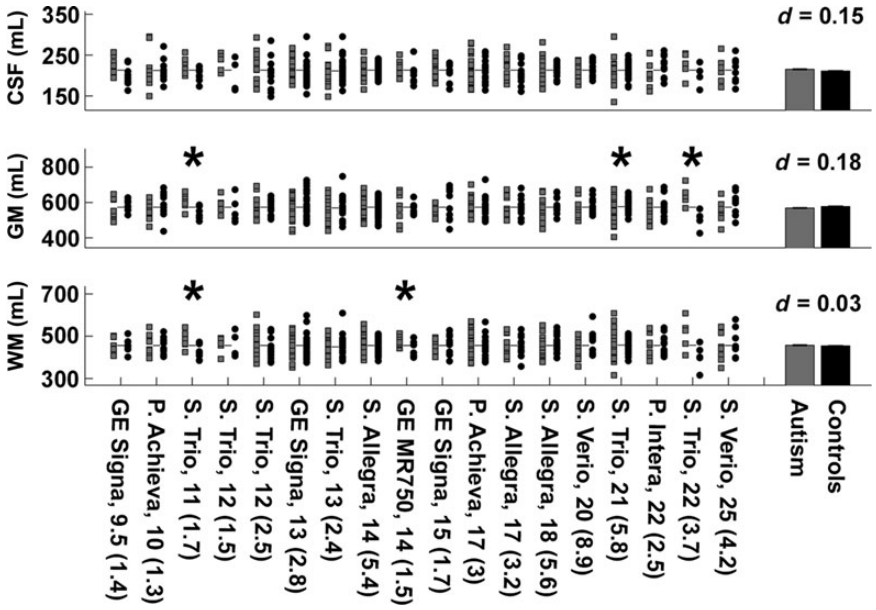

Figure 8. Results from the FSL FAST algorithm revealed no significant differences across groups. Scatter plots of cerebrospinal fluid (CSF), gray matter, and white matter volumes in individuals with ASD (gray) and matched controls (black) separately for each site. Bars: means across sites for each subject group. Asterisks: significant differences ($P < 0.05$, randomization test, FDR corrected). Error bars: standard error of the mean across subjects. d = Cohen's d (effect size across groups).

older individuals supports recent hypotheses, which have suggested that ASD is characterized by early brain overgrowth followed by arrested growth or even degeneration at later ages (Courchesne et al. 2011).

Brain Areas of Special Interest

Similar controversy exists regarding the presence and significance of anatomical abnormalities in several ROIs of specific interest to ASD research including the amygdala, hippocampus, cerebellum, and CC. The "Amygdala theory of autism" has suggested that abnormalities in the amygdala may explain the social and emotional difficulties found in ASD (Baron-cohen et al. 2000). While some studies have reported significantly larger amygdala volumes in ASD individuals than controls, others have reported no difference across groups, and one study has reported the opposite (Bellani et al. 2013a). There were no significant differences across groups in our study in amygdala volumes (Fig. 2) nor in hippocampus volumes, another limbic system region of particular interest for ASD research (Sweeten et al. 2002).

The CC has also been a focus of attention following several theories that have proposed poor long-range neural connectivity/synchronization in ASD (Gepner and Féron 2009). While some studies have reported significantly smaller CC volumes in ASD individuals than controls, others have reported no difference across groups (Bellani et al. 2013b). We found no significant between-group difference when assessing the entire CC volume (Fig. 2), yet a more detailed analysis revealed that the ASD group exhibited a significantly smaller volume in the central segment of the CC when compared with the control group (Fig. 3). The effect size of this difference, however, was very small (Cohen's $d = 0.2$) and stands in contrast to previous estimates of much larger effect sizes in all CC segments (Cohen's $d > 0.4$, Frazier and Hardan 2009).

Studies of the cerebellum in ASD individuals have reported mixed findings, as well, with some documenting abnormally large cerebellar volumes (Hardan, Minshew, Harenski et al. 2001; Scott et al. 2009) and abnormally small vermis volumes and midsagittal vermis areas, often found only in specific lobules (Courchesne et al. 1988, 1994). Here, we only assessed cerebellar gray and white matter volumes and found no significant differences across groups (Fig. 2).

Ventricles

While early ventriculomegaly (enlarged ventricles) is associated with multiple developmental delays (McKechnie et al. 2012), only a few studies have assessed ventricle volumes in ASD with some (Palmen et al. 2004) but not others (Hardan, Minshew, Mallikarjuhn et al. 2001) reporting significantly larger ventricle volumes in ASD. Our results revealed significantly larger ventricle volumes in ASD when compared with controls (Fig. 2), yet differences across groups exhibited a relatively small effect size (Cohen's $d = 0.34$).

Relationship Between Volume, IQ, and ADOS

Our findings are consistent with previous studies that have reported significant correlations between IQ and gross anatomical measures (Deary et al. 2010). Both control and ASD subjects exhibited significant positive correlations between IQ and intracranial, gray, and white matter volumes as well as cortical surface area (Fig. 5). ASD individuals also exhibited significant positive correlations between ADOS scores and gray and white matter volumes as well as cortical surface area. Considering that gross anatomical measures are a function of neuronal loss, synaptic pruning, and gray matter shrinkage, in conjunction with white matter maturation and growth (as evident in the growth curves of Supplementary Fig. 5), it is intriguing that similar positive correlations exist between gross anatomical developmental measures and IQ/ADOS scores. Our results, however, offer very limited insight regarding the meaning of these general associations.

When considering the IQ range of the current ASD sample, it is important to note that our results are limited to relatively high-functioning ASD individuals (IQ mean = 106, standard deviation, SD = 17). ASD individuals with lower function may potentially exhibit significant structural abnormalities and stronger correlations with behavioral measures. Studies with ASD individuals of lower function are generally missing from the ASD neuroimaging literature and are highly warranted.

Multivariate Classification of Autism

Multivariate classification techniques offer a method for decoding the group identity of ASD and control subjects based on patterns of anatomical features, rather than single anatomical measures. Several recent studies have reported exceptional decoding accuracies (over 85%) of group identity when using measures of cortical thickness, geometry, curvature, and/or surface area in small samples of 20–30 subjects from each group (Ecker, Marquand et al. 2010; Jiao et al. 2010; Uddin et al. 2011). These studies have suggested that distributed anatomical abnormalities can consistently be used to identify adults with ASD.

Performing similar multivariate analyses with the ABIDE data revealed much lower decoding accuracies (<60%), which exceeded chance levels only when using measures of subcortical volumes or cortical thickness (Fig. 6). Similarly weak decoding accuracies were found when using different validation schemes (i.e., leave-two-out instead of 10-fold validation) and when using linear or nonlinear classification algorithms (Supplementary Fig. 3). In addition, performing the same classification analyses within each of the 5 largest sites (>20 subjects in each group) revealed equivalent results with accuracy rates below 60% in all sites. These results suggest that weak classification accuracies were not generated by differences across sites or the choice of a particular classification algorithm or validation scheme. Instead, we believe that these weak decoding accuracies were the consequence of the considerable within-group variability, which was clearly evident in the univariate analyses (Figs 1–3). Variable and inconsistent anatomy across individuals of each group is expected to reduce classification accuracy across groups.

While classification techniques hold great promise, they also suffer from a major potential pitfall: overfitting. The accuracy and validity of a classifier depend on the number of features being considered and the number of subjects available for training and testing. In situations where there are more anatomical features than subjects, the classification algorithm may arbitrarily separate individual subjects into meaningless groups by overfitting noise/variability (Cawley and Talbot 2010). When only a small number of subjects are left out for testing, there is a risk of arbitrarily inflating decoding accuracies. This was clearly apparent when we performed classification analyses with small, randomly selected subgroups. For example, when decoding subject identity based on cortical thickness measures, the actual decoding accuracy was 51%, yet roughly 30% of randomly selected small groups (20 ASD/20 control subjects) exhibited decoding accuracies above 60%. When considering the multitude of anatomical features that can be used for classification, the potential of finding a set of features that yields accurate decoding seems extremely high. It is, therefore, critical to evaluate classification schemes on large samples and test their generalizability.

Automated Analyses Using Large Multicenter Samples

We believe that the controversy regarding the presence and significance of anatomical abnormalities in ASD is a result of the fact that the majority of previous studies recruited small ASD and control samples with distinct age, IQ, and ASD severity characteristics from general populations that have considerable anatomical variability. We suggest that studies with such small samples are unlikely to capture the true anatomical distributions of either population and are, therefore, unlikely to yield consistent results. The only way to identify robust differences across ASD and control groups and reconcile previous contradictions in the literature is to assess anatomical differences across large cohorts where age, IQ, gender, and autism severity factors can be controlled for (or their relative contribution assessed) and the true anatomical distribution of each measure can be estimated for each group as performed here using the ABIDE data.

The down side of evaluating the ABIDE data is that aggregated MRI scans from multiple international sites may contain large between-site variability, resulting from possible confounding factors, for example, autism diagnosis, and hence, subject selection may differ across sites such that sites with unique clinical criteria may "contaminate" the dataset. All of the participants in the current study, however, were characterized with the standardized ADOS assessment, which is the current "golden standard" for identifying individuals with ASD. Another possible confound is the heterogeneity of MRI scanners and scanning protocols across sites. Fortunately, however, automated estimations of cortical and subcortical volumes with Freesurfer have been shown to be remarkably robust across MRI scanners of different manufacturers (varying by 0–3%) as long as the field strength is kept constant at 3T (Jovicich et al. 2009). The third and likely largest source of variability across sites is generated by genetic and environmental heterogeneity, which are known to affect brain anatomy (Thompson et al. 2001).

To address this caveat, we performed several analyses to ensure that our results were not due to between-site variability. First, we used a standard mixed-model approach to examine the contribution of site differences to potential differences between ASD and control groups. In an additional step, we regressed out the differences across sites and performed several univariate and multivariate analyses with data where the mean of each site had been normalized. Finally, we performed per-site analyses to determine reliability of results across multiple sites. These analyses reassured us that the conclusions of this study were not based on potentially misleading between-site differences or on spurious findings that were apparent in only a small minority of sites. Perhaps the strongest impression one should gain from examining the presented data is that the vast majority of sites do not show any consistent anatomical differences across groups. Importantly, in order for group differences to be clinically relevant, they must be robust to differences in MRI scanning parameters as well as genetic and environmental heterogeneity.

Implications for Future ASD Research

The current study demonstrates that anatomical differences between high-functioning ASD and control groups (aged 6–35 years old) are very small in comparison to large within-group variability. This suggests that anatomical measures alone are likely to be of low scientific and clinical significance for identifying children, adolescents, and adults with ASD or for elucidating their neuropathology. Rather than expecting to find consistent anatomical abnormalities across the entire ASD population, it may be more reasonable to search for subgroups of ASD individuals with more homogeneous etiologies who may (or may not) exhibit common structural findings. Determining how to segregate

ASD individuals into meaningful subgroups based on genetic profiling, clinical comorbidities, sensory sensitivities, and other relevant measures seems to be the most urgent next step for future ASD research. Further efforts to aggregate broad clinical, genetic, and neuroimaging data from large ASD samples of different ages and IQ levels will be extremely helpful in achieving this goal.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work was supported by Simons Foundation SFARI grant 177638 (M.B. and I.D.).

Notes

We thank our colleague, Dr Nancy Minshew, for her consistent support of our research efforts. We also thank Dr Joel Greenhouse, Department of Statistics, Carnegie Mellon University, for his useful guidance on some of the analyses. *Conflict of Interest*: None declared .

References

Amaral DG, Schumann CM, Nordahl CW. 2008. Neuroanatomy of autism. Trends Neurosci. 31:137–145.

Baron-cohen S, Ring HA, Bullmore ET, Wheelwright S, Ashwin C, Williams SC. 2000. The amygdala theory of autism. Neurosci Biobehav Rev. 24:355–364.

Bellani M, Calderoni S, Muratori F, Brambilla P. 2013a. Brain anatomy of autism spectrum disorders II. Focus on amygdala. Epidemiol Psychiatr Sci. 22:309–312.

Bellani M, Calderoni S, Muratori F, Brambilla P. 2013b. Brain anatomy of autism spectrum disorders I. Focus on corpus callosum. Epidemiol Psychiatr Sci. 22:217–221.

Benjamini Y, Hechtlinger Y. 2014. Discussion: an estimate of the science-wise false discovery rate and applications to top medical journals by Jager and Leek. Biostatistics. 15:13–16; discussion 39–45.

Cawley G, Talbot N. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 11:2079–2107.

Courchesne E, Campbell K, Solso S. 2011. Brain growth across the life span in autism: age-specific changes in anatomical pathology. Brain Res. 1380:138–145.

Courchesne E, Karns CM, Davis HR, Ziccardi R, Carper RA, Tigue ZD, Chisum HJ, Moses P, Pierce K, Lord C, et al. 2001. Unusual brain growth patterns in early life in patients with autistic disorder. An MRI study. Neurology. 57:245–254.

Courchesne E, Pierce K, Schumann CM, Redcay E, Buckwalter Ja, Kennedy DP, Morgan J. 2007. Mapping early brain development in autism. Neuron. 56:399–413.

Courchesne E, Redcay E, Kennedy DP. 2004. The autistic brain: birth through adulthood. Curr Opin Neurol. 17:489–496.

Courchesne E, Townsend J, Saitoh O. 1994. The brain in infantile autism: posterior fossa structures are abnormal. Neurology. 44:214–223.

Courchesne E, Yeung-Courchesne R, Press GA, Hesselink JR, Jernigan TL. 1988. Hypoplasia of cerebellar vermal lobules VI and VII in autism. N Engl J Med. 318:1349–1354.

Deary IJ, Penke L, Johnson W. 2010. The neuroscience of human intelligence differences. Nat Rev Neurosci. 11:201–211.

Destrieux C, Fischl B, Dale A, Halgren E. 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage. 53:1–15.

Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K, Anderson JS, Assaf M, Bookheimer SY, Dapretto M, et al. 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry. 19:659–667.

Ecker C, Ginestet C, Feng Y, Johnston P, Lombardo MV, Lai M-C, Suckling J, Palaniyappan L, Daly E, Murphy CM, et al. 2013. Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. JAMA Psychiatry. 70:59–70.

Ecker C, Marquand A, Mourão-Miranda J, Johnston P, Daly EM, Brammer MJ, Maltezos S, Murphy CM, Robertson D, Williams SC, et al. 2010. Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. J Neurosci. 30:10612–10623.

Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG. 2010. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. Neuroimage. 49:44–56.

Ecker C, Suckling J, Deoni SC, Lombardo MV, Bullmore ET, Baron-Cohen S, Catani M, Jezzard P, Barnes A, Bailey AJ, et al. 2012. Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: a multicenter magnetic resonance imaging study. Arch Gen Psychiatry. 69:195–209.

Fatemi SH, Aldinger KA, Ashwood P, Bauman ML, Blaha CD, Blatt GJ, Chauhan A, Chauhan V, Dager SR, Dickson PE, et al. 2012. Consensus paper: pathological role of the cerebellum in autism. Cerebellum. 11:777–807.

Fischl B. 2012. FreeSurfer. Neuroimage. 62:774–781.

Fischl B, Dale AM. 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA. 97:11050–11055.

Frazier TW, Hardan AY. 2009. A meta-analysis of the corpus callosum in autism. Biol Psychiatry. 66:935–941.

Gepner B, Féron F. 2009. Autism: a world changing too fast for a mis-wired brain? Neurosci Biobehav Rev. 33:1227–1242.

Groen W, Teluij M, Buitelaar J, Tendolkar I. 2010. Amygdala and hippocampus enlargement during adolescence in autism. J Am Acad Child Adolesc Psychiatry. 49:552–560.

Hardan AY, Minshew NJ, Harenski K, Keshavan MS. 2001. Posterior fossa magnetic resonance imaging in autism. J Am Acad Child Adolesc Psychiatry. 40:666–672.

Hardan AY, Minshew NJ, Mallikarjuhn M, Keshavan MS. 2001. Brain volume in autism. J Child Neurol. 16:421–424.

Hardan AY, Muddasani S, Vemulapalli M, Keshavan MS, Minshew NJ. 2006. An MRI study of increased cortical thickness in autism. Am J Psychiatry. 163:1290–1292.

Hazlett HC, Poe MD, Gerig G, Smith RG, Piven J. 2006. Cortical gray and white brain tissue volume in adolescents and adults with autism. Biol Psychiatry. 59:1–6.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. 2012. FSL. Neuroimage. 62:782–790.

Jiao Y, Chen R, Ke X, Chu K, Lu Z, Herskovits EH. 2010. Predictive models of autism spectrum disorder based on brain regional cortical thickness. Neuroimage. 50:589–599.

Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, Pacheco J, Albert M, Killiany R, Blacker D, et al. 2009. MRI-

derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage. 46:177–192.

Just MA, Cherkassky VL, Keller TA, Kana RK, Minshew NJ. 2007. Functional and anatomical cortical underconnectivity in autism: evidence from an fMRI study of an executive function task and corpus callosum morphometry. Cereb Cortex. 17:951–961.

Lotspeich LJ, Kwon H, Schumann CM, Fryer SL, Goodlin-Jones BL, Buonocore MH, Lammers CR, Amaral DG, Reiss AL. 2004. Investigation of neuroanatomical differences between autism and Asperger syndrome. Arch Gen Psychiatry. 61:291–298.

McKechnie L, Vasudevan C, Levene M. 2012. Neonatal outcome of congenital ventriculomegaly. Semin Fetal Neonatal Med. 17:301–307.

Palmen SJMC, Hulshoff Pol HE, Kemner C, Schnack HG, Janssen J, Kahn RS, van Engeland H. 2004. Larger brains in medication naive high-functioning subjects with pervasive developmental disorder. J Autism Dev Disord. 34:603–613.

Raznahan A, Toro R, Daly E, Robertson D, Murphy C, Deeley Q, Bolton PF, Paus T, Murphy DGM. 2010. Cortical anatomy in autism spectrum disorder: an in vivo MRI study on the effect of age. Cereb Cortex. 20:1332–1340.

Scheel C, Rotarska-Jagiela A, Schilbach L, Lehnhardt FG, Krug B, Vogeley K, Tepest R. 2011. Imaging derived cortical thickness reduction in high-functioning autism: key regions and temporal slope. Neuroimage. 58:391–400.

Scott JA, Schumann CM, Goodlin-Jones BL, Amaral DG. 2009. A comprehensive volumetric analysis of the cerebellum in children and adolescents with autism spectrum disorder. Autism Res. 2:246–257.

Shen MD, Nordahl CW, Young GS, Wootton-Gorges SL, Lee A, Liston SE, Harrington KR, Ozonoff S, Amaral DG. 2013. Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. Brain. 136: 2825–2835.

Stanfield AC, McIntosh AM, Spencer MD, Philip R, Gaur S, Lawrie SM. 2008. Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. Eur Psychiatry. 23:289–299.

Sweeten TL, Posey DJ, Shekhar A, McDougle CJ. 2002. The amygdala and related structures in the pathophysiology of autism. Pharmacol Biochem Behav. 71:449–455.

Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen VP, Huttunen M, Lönnqvist J, Standertskjöld-Nordenstam CG, Kaprio J, Khaledy M, et al. 2001. Genetic influences on brain structure. Nat Neurosci. 4:1253–1258.

Uddin LQ, Menon V, Young CB, Ryali S, Chen T, Khouzam A, Minshew NJ, Hardan AY. 2011. Multivariate searchlight classification of structural magnetic resonance imaging in children and adolescents with autism. Biol Psychiatry. 70:833–841.

Wallace GL, Dankner N, Kenworthy L, Giedd JN, Martin A. 2010. Age-related temporal and parietal cortical thinning in autism spectrum disorders. Brain. 133:3745–3754.