Molecular Psychiatry
https://doi.org/10.1038/s41380-019-0420-6

ARTICLE

Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder

Saashi A. Bedford[1,2] · Min Tae M. Park[1,3] · Gabriel A. Devenyi[1,4] · Stephanie Tullo[1,2] · Jurgen Germann[1] · Raihaan Patel[1,5] · Evdokia Anagnostou[6] · Simon Baron-Cohen[7] · Edward T. Bullmore[8] · Lindsay R. Chura[7] · Michael C. Craig[9,10] · Christine Ecker[9,11] · Dorothea L. Floris[7,12] · Rosemary J. Holt[7] · Rhoshel Lenroot[13] · Jason P. Lerch[14,15] · Michael V. Lombardo[7,16] · Declan G. M. Murphy[9] · Armin Raznahan[17] · Amber N. V. Ruigrok[7] · Elizabeth Smith[18] · Michael D. Spencer[7] · John Suckling[8] · Margot J. Taylor[19] · Audrey Thurm[18] · MRC AIMS Consortium · Meng-Chuan Lai[7,14,20,21,22] · M. Mallar Chakravarty[1,2,4,5]

Received: 9 March 2018 / Revised: 25 March 2019 / Accepted: 26 March 2019

Abstract

Significant heterogeneity across aetiologies, neurobiology and clinical phenotypes have been observed in individuals with autism spectrum disorder (ASD). Neuroimaging-based neuroanatomical studies of ASD have often reported inconsistent findings which may, in part, be attributable to an insufficient understanding of the relationship between factors influencing clinical heterogeneity and their relationship to brain anatomy. To this end, we performed a large-scale examination of cortical morphometry in ASD, with a specific focus on the impact of three potential sources of heterogeneity: sex, age and full-scale intelligence (FIQ). To examine these potentially subtle relationships, we amassed a large multi-site dataset that was carefully quality controlled (yielding a final sample of 1327 from the initial dataset of 3145 magnetic resonance images; 491 individuals with ASD). Using a meta-analytic technique to account for inter-site differences, we identified greater cortical thickness in individuals with ASD relative to controls, in regions previously implicated in ASD, including the superior temporal gyrus and inferior frontal sulcus. Greater cortical thickness was observed in sex specific regions; further, cortical thickness differences were observed to be greater in younger individuals and in those with lower FIQ, and to be related to overall clinical severity. This work serves as an important step towards parsing factors that influence neuroanatomical heterogeneity in ASD and is a potential step towards establishing individual-specific biomarkers.

Introduction

Early brain overgrowth was one of the earliest neural phenotypes reported in autism spectrum disorder (ASD) [1, 2]. However, subsequent studies examining advanced cortical phenotypes have reported diverse and conflicting neuroanatomical findings. For example, increases, as well as decreases have been reported in both cortical thickness (CT) [3–6] and surface area (SA) [7–10]. This may, in part, be attributable to factors that influence phenotypic heterogeneity in ASD, such as age, sex and intelligence [11–14]. However, these potential sources of heterogeneity are commonly regressed out as nuisance variables in statistical modelling or not considered in ASD studies. In the face of limited sample sizes, previous studies have omitted females altogether [5, 15, 16], or examined limited age ranges [8, 9, 17–19], while others typically do not examine

The Medical Research Council Autism Imaging Multicentre Study Consortium (MRC AIMS Consortium) is a UK collaboration between the Institute of Psychiatry, Psychology and Neuroscience (IoPPN) at King's College, London, the Autism Research Centre, University of Cambridge, and the Autism Research Group, University of Oxford. Members of MRC AIMS Consortium are listed at the end of the article.

Supplementary information The online version of this article (https://doi.org/10.1038/s41380-019-0420-6) contains supplementary material, which is available to authorised users.

✉ Saashi A. Bedford
saashi.bedford@mail.mcgill.ca

✉ M. Mallar Chakravarty
mallar@cobralab.ca

Extended author information available on the last page of the article.

Published online: 26 April 2019

associations with intelligence. The limited studies considering these factors have observed that ASD-related atypical neuroanatomy varies greatly by age [5, 20–25], sex (see Lai et al. [26] for a review) and estimated intelligence [27], suggesting a need to reconcile the association between factors that contribute to clinical heterogeneity and neuroanatomical differences. It is also possible that previous findings may be further confounded by biases in morphological estimates related to movement during image acquisition (particularly given the observation that neurotypical and ASD males are most likely to move during scanning) [28, 29], and variations in the quality control (QC) of image processing outputs [30]. Here we sought to reconcile the impact of sex, age and estimated intelligence on heterogeneity in ASD cortical morphology by performing a large-scale neuroimaging study using magnetic resonance imaging data acquired from multiple sources (initial dataset of 3145 subjects, 1327 subjects after rigorous QC).

Based on previous findings reported in the literature, we expected to observe overall greater CT in individuals with ASD relative to neurotypical controls [5, 31, 32]. Given known clinical, behavioural and neuroanatomical sex differences in ASD, we expected cortical alterations to differ in regional composition by sex [26, 33, 34]. We also expected differences to be more pronounced in younger [5, 35] and lower IQ individuals [17, 27].

Methods

Sample

Cross-sectional data included here were acquired from previous studies by the National Institute of Mental Health (USA), the Hospital for Sick Children (Canada), the Cambridge Family Study of Autism (UK) and the UK Medical Research Council Autism Imaging Multicentre Study (UK MRC AIMS). We also included publicly available data from the Autism Brain Imaging Data Exchange (ABIDE) I and II [36, 37].

The total initial sample size amounts to 3145 individuals: 1415 individuals with ASD (1165 male/250 female) and 1730 controls (1172 male/558 female), aged 2–65 years. See Supplementary Methods section S1 "Sample details" for imaging parameters and participant demographics.

QC and site elimination

Rigorous QC was performed by two independent raters (SAB, and either ST or MMC) at both the level of the raw input images (for motion and scan quality), and on processed outputs (see Supplementary Methods section S2 "Quality control and site elimination"; Supplementary Figs. S1 and S2). Sites with three or more individuals per sex and diagnostic group remaining after QC were included (final dataset of 1327 individuals; 491 individuals with ASD (362 male/129 female) and 836 neurotypical controls (481 male/355 female) (Supplementary Tables S1 and S2). All analyses, unless otherwise indicated, were performed using this dataset of 1327 individuals.

Image processing

All T1-weighted images were pre-processed using the minc-bpipe-library pre-processing pipeline (https://github.com/CobraLab/minc-bpipe-library), and then submitted to the CIVET processing pipeline [38] (version 1.1.12; Montreal Neurological Institute), to estimate CT, SA and volume. Image processing and QC was standardised across all data, and conducted within a single laboratory. For details, see Supplementary Methods sections S3 "Image pre-processing" and S4 "Image processing".

Statistical analysis

To account for differences in scanners, acquisitions and sample characteristics, statistical analysis was conducted using a prospective meta-analytic technique, where each site is initially treated as an independent study and results are pooled to define significance (see van Erp et al. [39]). First, multiple linear regressions were conducted to derive per site Cohen's *d* effect sizes for the main effect of each variable of interest. An aggregate statistic representing all sites was derived by pooling effect sizes in a random-effects meta-analysis [40, 41] (*metafor* 2.0-0 package in R 3.4.0). For examples of statistical models employed, see Supplementary Methods section S5 "Statistical models used".

Case–control comparisons: global measures

Differences in mean CT, total SA, cortical volume (CV), total grey matter (GM), total white matter (WM) and total brain volume (TBV) were compared between individuals with ASD and controls by examining the main effect of diagnosis, while including age (linear term) and sex in the model. Results were Bonferroni corrected with $p < 0.008$ (based on six tests) being considered significant. GM and WM analyses were reanalysed while controlling for TBV, to determine if these were differentially affected when accounting for global measures.

Case–control comparisons: vertex-wise analysis

Regional alterations in CT and SA were examined using the same meta-analytic technique and model described above for global measures, but extended to a vertex-wise level

(81,924 vertices across the brain), and corrected for multiple comparisons using the false discovery rate [FDR] [42]. To control for multiple comparisons both across vertices and across the various analyses done, *p*-values from all vertices of all main analyses were pooled (including each interval of the age- and full-scale intelligence [FIQ]-centred analyses described in the subsequent sections), and a 5% FDR threshold was used to control for multiple comparison across all statistical tests conducted. This stringent FDR correction was applied separately for CT and SA analyses.

Case–control comparisons were also examined using a mixed-effects model with site as a random factor to determine if our results diverge from previous large-scale studies that used this methodology (e.g., van Rooij et al. [6]).

Heterogeneity-focused analyses: importance of sex, age and FIQ

To assess the significance of sex, age and FIQ in our vertex-wise analysis of cortical alterations, we fitted two models for each variable: one including the variable of interest (i.e., sex, age or FIQ), plus an interaction term between that variable and diagnosis, and the other without the variable of interest, or the interaction, in the model. Please see Supplementary Methods section S5 "Statistical models used" for details. We then used Akaike information criterion [43] (AIC, representing the best model fit) to determine the importance of the variable at each vertex, within each site separately. At each vertex, we determined the number of sites for which each model was shown to be the best fit, and calculated a weighted average (based on site size) to determine the best model, on average, at that vertex, taking into account all sites.

Based on the AIC comparison of the models, sex, age and FIQ were demonstrated to be important explanatory variables at a substantial proportion of vertices across the brain for both CT and SA, motivating our further examination of these factors and their impact on cortical alterations in ASD (see Supplementary Figs. S3 and S4).

Sex-focused analyses

Sex-specific patterns were examined using the case–control analysis described above separately in males and females (for global and vertex-wise measures), with diagnosis and age (linear term) included in the model.

Age-focused analyses

Given variable reporting of best-fit trajectories in ASD and typical neurodevelopment in general [30, 44], we tested the best model fit between linear, quadratic and cubic models of age (all models also included diagnosis, each age term, interaction between diagnosis and each age term, and sex; see Supplementary Methods section S5 for statistical models used). To do this, at each vertex, the minimum AIC was determined for each site, and a weighted average across sites was calculated per vertex, as described above.

The AIC for age revealed the linear model to be the best fit at most sites (range across vertices: 22–100% of sites) across most of the cortex for both CT (Supplementary Fig. S5A) and SA (Supplementary Fig. S6A).

Next, an age-centred analysis was used to examine age-dependent changes in patterns of vertex-wise CT and SA alterations by centreing age at intervals of 2 years, accounting for age as a linear term. This allows us to illustrate the differential effects on CT at different ages, and allows interpretation of group differences at the centred age interval. Essentially, this provides a "snapshot" of the groups' regression lines at that interval, without having to split the dataset into age ranges, thereby maximising power, and case–control differences were examined at each age interval [5, 45]. This was done by calculating the per site Cohen's *d* effect size for the main effect of diagnosis from each model (each age interval), and pooling these effect sizes in the random-effects meta-analysis in the same manner as the case–control comparisons.

FIQ-focused analyses

The best model fit for the FIQ analyses was tested in the same way as the age analyses described above: the best model fit was tested between linear, quadratic and cubic models of FIQ (all models also included diagnosis, each FIQ term, interaction between diagnosis and each FIQ term, age and sex).

The AIC for FIQ revealed the linear model to be the best fit at most sites (range across vertices: 24–100% of sites) across most of the cortex for CT (Supplementary Fig. S5B) and SA (Supplementary Fig. S6B).

An FIQ-centred analysis was performed in the same fashion as the age-centred analysis, with FIQ centred at intervals of 10 points, and using a linear term for FIQ. Results are examined at intervals of FIQ = 80 and above, as there are very few controls with an FIQ < 80. As FIQ data were not available for all individuals, this analysis was performed on a slightly smaller subset of 1214 individuals.

Associations between CT and ASD symptoms/characteristics

As consistent measures of autistic symptoms or characteristics were not available across all sites, analyses were performed on subsets of individuals who had the same measures, as in previous studies [6, 46]. We chose the measures which had the largest number of individuals

available, which included the ADOS-2 Calibrated Severity Scores (CSS) [47] to examine overall symptom severity ($N = 279$; also conducted separately in males [$N = 224$] and females [$N = 55$]), the ADOS-G reciprocal social interaction domain score, communication domain score, and restricted, repetitive behaviour [RRB] domain score (module 4; $N = 151$), and the ADOS-2 RRB domain score and social affect domain score (module 3; $N = 143$), all in individuals with ASD only. In both ASD and control individuals, we examined associations between CT and scores of the Social Responsiveness Scale (SRS; $N = 413$) and Autism Spectrum Quotient [48] (AQ; $N = 171$), as well as their interaction with diagnosis.

These analyses were conducted using a meta-regression technique. See Supplementary Methods section S6 "Associations between CT and ASD severity and symptoms" for meta-analysis details and subset sample characteristics.

Finally, we also performed a separate analysis to examine the potential effects of comorbid diagnoses on cortical alterations related to ASD. We repeated the case–control analysis excluding data from individuals with comorbid diagnoses (limiting our analyses to sites with this type of data recorded; resulting in a dataset of N = 519; 144 ASD/375 Controls).

Case–control, sex-stratified and age-centred analyses including FIQ in the model

Based on the results of the AIC analysis assessing the importance of FIQ as an explanatory variable, we examined the diagnosis, age-centred and sex-stratified analyses including FIQ in the model. This was done in the subset of individuals for whom FIQ data were available ($N = 1214$). For these analyses, FDR correction was conducted across all analyses (including all age intervals) together, but separately from the main set of analyses.

Impact of QC

We examined the impact of QC on both the neuroanatomy and demographics of our sample (see Supplementary Methods section S7 "Quality control analysis").

Power calculation

We used G*Power version 3.1.9.4 to determine the minimum detectable effect sizes given our sample size of 491 individuals with ASD and 836 controls. At a power level of 0.8 and a significance threshold of 0.05 (two tailed), we determined we would be have the statistical power to detect effect sizes of 0.1463 and greater. However, this is based on a simple multiple linear regression analysis that pools all data together, ignoring the differences between sites, and not accounting for this in the analysis. It is unclear how the meta-analytic technique employed here would affect these estimates.

Results

Results of case–control comparisons and sex-focused analysis are presented in Fig. 1, the age-focused analysis in Fig. 2, the FIQ-focused analysis in Fig. 3 and symptom/severity-focused analysis in Fig. 4.

Greater CV and mean and regional CT in ASD

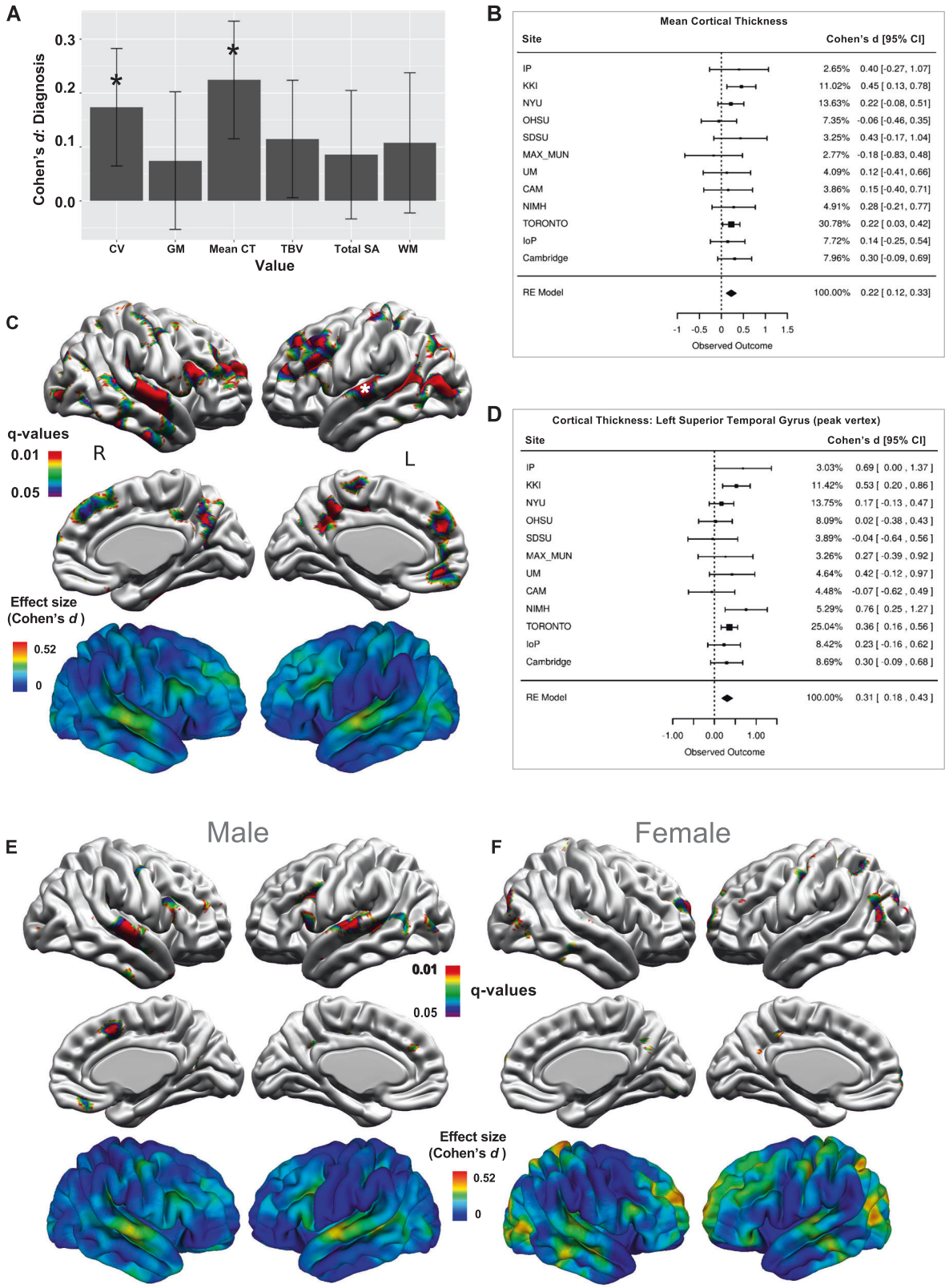
We observed significantly greater CV ($p < 0.008$; Cohen's $d = 0.17$) and mean CT ($p < 0.0001$; Cohen's $d = 0.22$) and a trend towards enlarged TBV ($p < 0.05$; Cohen's $d = 0.11$) in individuals with ASD (Fig. 1a). No differences were observed in total SA, WM or GM. When controlling for TBV, both GM and WM remain nonsignificant; however, WM seemed to be slightly more affected, changing from $p = 0.1$ to $p = 0.9$ when controlling for TBV, whereas GM was barely affected ($p = 0.25$ in original analysis and $p = 0.24$ when controlling for TBV).

In the vertex-wise analysis, regional group differences of CT (greater CT in ASD compared with controls) were observed in the inferior frontal and prefrontal cortex, superior temporal, postcentral, and posterior cingulate gyri and precuneus, bilaterally, surviving 5% FDR (peak Cohen's $d = 0.32$). Effect sizes showed some variability by site, however, were largely positive (Fig. 1b, d; Supplementary Fig. S7). The mixed-effects model yielded similar results to the meta-analytic approach, however, the results were less significant and over a smaller proportion of the cortex (Supplementary Fig. S8). No significant differences were observed in SA.

Sex-specific cortical alterations

ASD males had significantly greater CV ($p < 0.008$; Cohen's $d = 0.19$) and mean CT ($p < 0.008$; Cohen's $d = 0.21$) compared with male controls (Supplementary Fig. S9). WM volume trended towards being greater in ASD males relative to controls ($p < 0.05$; Cohen's $d = 0.18$). No differences in total SA or GM volume were observed. In females with ASD, mean CT trended towards being greater compared with controls ($p < 0.05$; Cohen's $d = 0.21$). No differences were observed in TBV, total SA, CV, GM or WM in the females (Supplementary Fig. S10).

Both males and females with ASD presented with regions of significantly greater CT relative to controls, surviving 5% FDR, however, the observed patterns of CT differences were distinct between males and females



(Fig. 1e, f). In ASD males, regions of greater CT were observed in bilateral superior temporal, inferior frontal, and right precentral gyri (peak Cohen's $d = 0.39$). In ASD females, these differences were observed in bilateral prefrontal and occipital cortices, and left posterior parietal cortex and pre- and postcentral gyri (peak Cohen's

◀ **Fig. 1** Case–control comparisons. Individuals with autism spectrum disorder (ASD) presented with overall greater cortical volume and mean cortical thickness (CT), and a trend towards greater total brain volume, as well as regionally specific differences in CT. These group differences were observed in sex-specific patterns of regional involvement, and were of a larger magnitude in the females. **a** Cohen's *d* effect sizes for case–control comparisons of cortical volume (CV), total grey matter (GM), mean CT, total brain volume (TBV), total surface area (SA) and total white matter (WM) (* denotes $p < 0.008$; error bars represent 95% confidence intervals). Positive effect sizes denote greater values in individuals with ASD compared with controls. Significantly greater CV ($p < 0.008$) and mean CT ($p < 0.0001$), were observed in individuals with ASD. **b** Forest plot of Cohen's *d* effect sizes of mean CT per site. **c** Significant vertex-wise group differences in CT across all subjects, shown at a false discovery rate (FDR) threshold of 5% (top), and effect size maps (bottom). Individuals with ASD show greater CT relative to controls. **d** Forest plot showing effect sizes per site at a peak vertex in the left superior temporal gyrus. **e** Significant vertex-wise group differences in CT in males, shown at an FDR threshold of 5% (top) and effect size maps (bottom). Males with ASD show greater CT relative to controls, primarily in bilateral inferior frontal and superior temporal regions. **f** Significant vertex-wise group differences in CT in females, shown at an FDR threshold of 5% (top) and effect size maps (bottom). Females with ASD show greater CT relative to controls, primarily in left prefrontal, parietal and occipital regions. Effect sizes in females are greater than those seen in males

$d = 0.45$). Sex-specific effect sizes were overall stronger than in the combined sample, and larger effect sizes were observed in the females compared with males (Supplementary Figs. S11 and S12).

In the both males and females, for CT, the mixed-effects model yielded similar but less diffuse results, and only survived 5% FDR in the left hemisphere (for both, see Supplementary Fig. S13).

No significant differences in SA were observed in the males or females (meta-analytic model used for both).

Subtle age-specific cortical alterations

In the age-centred analyses, subtle but significant group differences in CT were maximal in childhood (8–10 years), with individuals with ASD presenting demonstrating greater CT relative to controls in small regions of the cortex. Figure 2 shows differences between individuals with ASD and controls at age intervals of 4 years, accounting for age using a linear model. Foci of significance were most apparent in the age range of 8–12 years, but the linear fits suggested steadily larger effect sizes for diagnosis on CT as one moves towards younger ages. Between the ages of 6 and 14 years, regions of significantly greater CT were observed primarily in lateral temporal and frontal regions, and the posterior cingulate cortex. After 12–14 years, less difference was observed between groups, and these differences were observed only in medial prefrontal regions.

In the age-centred SA analysis, no significant differences in SA were observed at any age interval.

FIQ-specific cortical alterations

Individuals with ASD with lower FIQ were observed to have much greater and more widespread differences in CT relative to controls than those with higher FIQ (Fig. 3), spanning large regions of the frontal, temporal and occipital cortices. Foci of significance were most apparent in the FIQ range of 100–110, but the linear fits suggested steadily larger effect sizes for diagnosis on CT as one moves towards lower FIQ. At FIQ of 120, only minimal significant group differences in CT were observed. Higher than this, no significant differences were seen.

In the FIQ-centred SA analysis, no significant differences in SA were observed at any FIQ interval.

Associations between CT and ASD symptoms/characteristics

A significant, positive correlation between CT and ADOS-2 CSS was observed in ASD individuals, primarily in the right hemisphere. This relationship was observed in regions in which individuals with ASD presented with significantly greater CT relative to controls, including the right superior temporal gyrus (STG) and inferior frontal sulcus, right orbitofrontal cortex and bilateral posterior cingulate cortices. Furthermore, motivated by our findings of sex-specific regions of CT alterations in subjects with ASD, we explored the relationship between CT and CSS in males and females separately. In the female sample, we observed a significant positive relationship between CT and severity, primarily in prefrontal and temporal regions. Conversely, in the males, only very minimal regions showed this significant relationship, despite the much larger sample size compared with the females (Fig. 4). Males and females in this sample did not differ significantly in severity or FIQ.

No significant associations were observed between the SRS or AQ and CT. Only very minimal significant associations were observed for ADOS domain scores with CT, in very small cortical regions. Please see Supplementary Results section S7 and Supplementary Figs. S14 and S15 "Associations between neuroanatomy and ASD symptoms/characteristics" for details.

Based on our analysis of the potential impact of comorbidities, including only individuals with ASD with no comorbid features does not seem to change the spatial extents of our results, but does impact the number of vertices surviving 5% FDR, and increases the overall effect size. Please see Supplementary Results section S8 and Supplementary Fig. S16 for details.

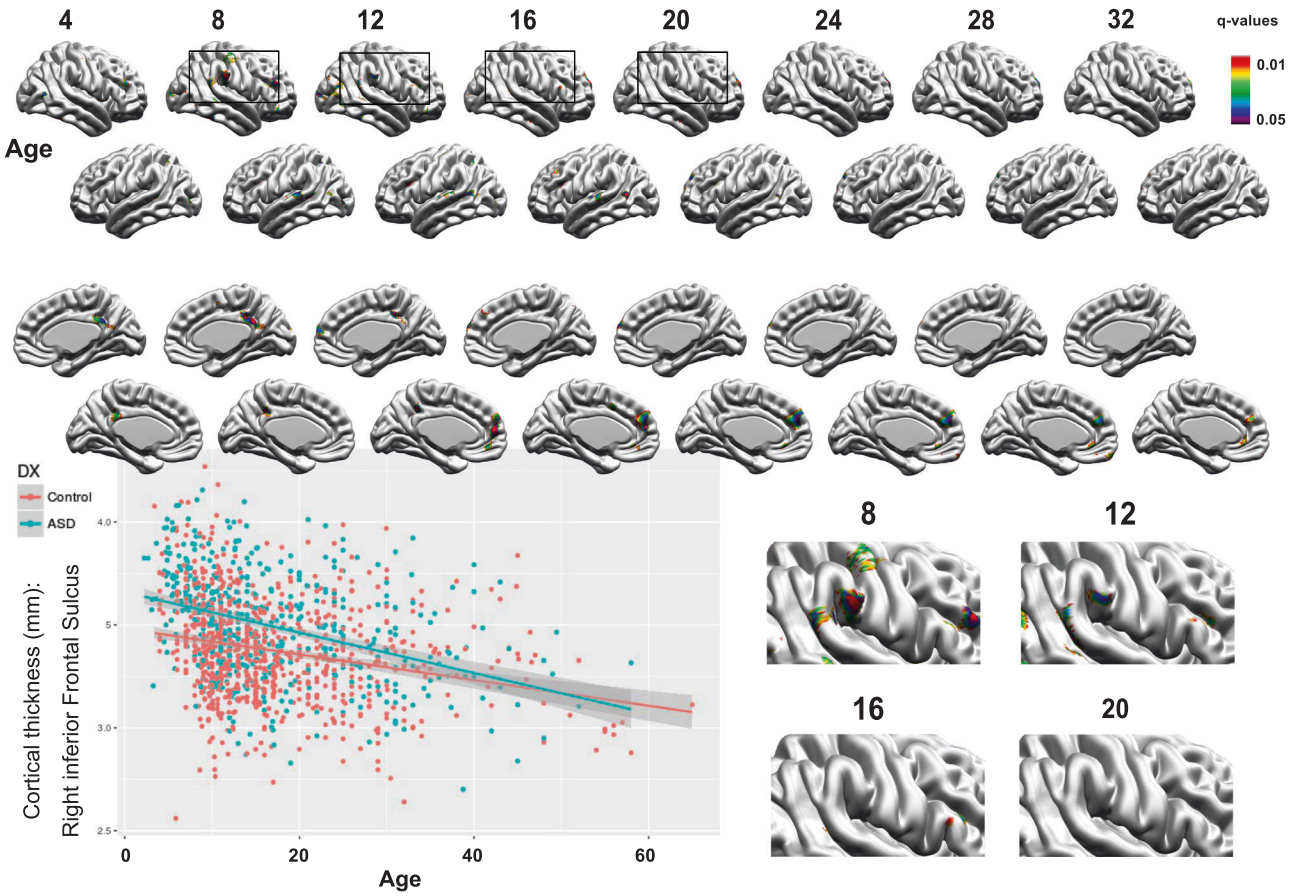


Fig. 2 Age-centred analysis. Main effect of diagnosis shown at 4-year intervals, using a linear model for age, shown at 5% false discovery rate (FDR) up until the age of 32, after which no significant differences are seen. Only minimal group differences were seen, primarily in right superior temporal and inferior frontal regions. Cortical thickness (CT) at a peak vertex in the left inferior frontal sulcus is plotted against age (bottom)

Case–control, sex-stratified and age-centred analyses including FIQ in the model

Including FIQ in the model did not substantially alter the results for the diagnosis main effect, sex-stratified analyses or age-centred analyses. Please see Supplementary Results section S9, and Supplementary Figs. S17–S20.

Discussion

In this study, we use a large dataset that has been strictly QC'd and analysed using harmonised image processing and statistical methods in order to study variation in cortical anatomy in ASD. Our results show greater CT in widespread cortical regions in individuals with ASD, primarily in the frontal and superior temporal cortex, as well as the precuneus and posterior cingulate cortices. Cortical alterations were observed to be differentially impacted by sex, age and FIQ. Greater CT was observed in largely different regions between males and females, with females demonstrating potentially greater magnitude of CT alterations than males, relative to same-sex controls. Group differences were greatest in childhood, and differences lessen after early adulthood. Alterations were observed in largest regions and were more significant in individuals with FIQ of 80–110, with almost no significant group differences observed in individuals with FIQ of 120 and higher. In ASD individuals, greater CT was positively correlated with symptom severity measured by ADOS-2 CSS, in regions which also showed greater CT relative to controls, and these correlations were stronger, and seen in distinct regions, in females compared with males.

Greater TBV in very young children with ASD is one of the most consistently reported findings in the ASD neuroimaging literature [49–51], and some studies show that this larger brain volume persists into adolescence [52]. Mechanisms potentially underlying increased TBV include increased neurogenesis, decreased synaptic pruning and neuronal cell death, and abnormal myelination [53]. Our results suggest that the larger TBV phenotype observed in ASD can also be recapitulated at levels of local and global

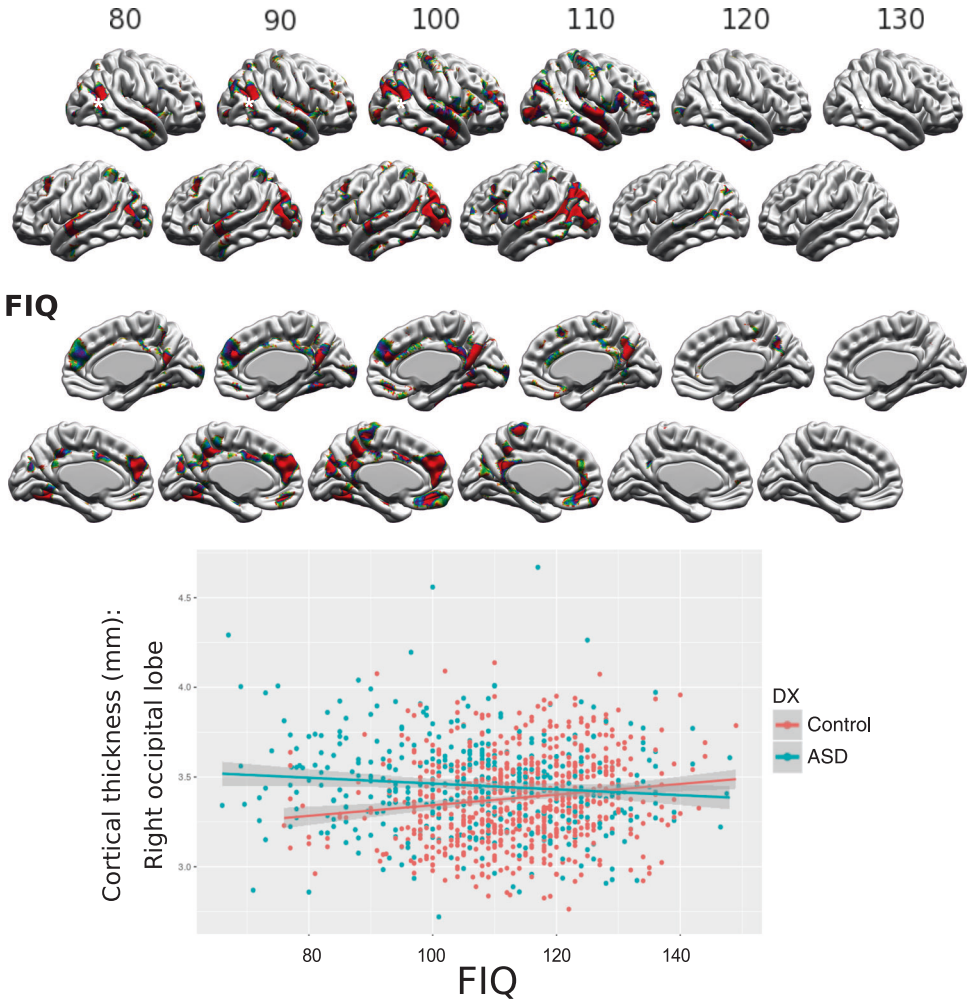


Fig. 3 Full-scale intelligence (FIQ)-centred analysis. Main effect of diagnosis at intervals of 10 FIQ points, using a linear model for FIQ (shown at 5% false discovery rate (FDR)), from an FIQ of 80, up until an FIQ of 130, after which no significant differences were seen. Maximal differences were observed around an FIQ of 100. Cortical thickness (CT) at a peak vertex in the right occipital lobe is plotted against FIQ (bottom)

CT (though here we only observed greater TBV in ASD at a trend level). Increased cell proliferation in the ventricular zone during development has been suggested as underlying abnormalities in the number and width of cortical columns (resulting in increased cortical SA), as well as increased neuronal density [54] (resulting in increased CT). Both cortical column abnormalities and increased neuronal density have both been reported in ASD [55], thus it is unclear why we do not observe the alterations in SA in individuals with ASD reported by other studies [7–9]. It is also unclear how QC may impact results (see below for further discussion on this). Thus, this relationship warrants further investigation. Deficiencies in synaptic pruning [56], which begins in early life and continues into adolescence, have also been proposed as underlying the greater CT observed in ASD [57]. This is supported by studies reporting reduced synaptic pruning during development in children with ASD [58], and could explain the differences that persist into adulthood, as observed here.

It should be noted that other factors can affect CT measurements; for example, altered cortical myelination or reduced integrity of the GM–WM boundary, potentially resulting from deficits in neuronal migration during early development. Specifically, this blurring of the cortical interface has been demonstrated in individuals with ASD in both histological post-mortem [59] and in vivo neuroimaging studies [60, 61], and could potentially lead to inaccuracies in CT estimates due to misplacement of the cortical boundary, with apparent increases in CT.

Previous studies [8, 10, 51] have reported very early expansion of the cortical surface and increased SA in young children (2–5 years) and infants (6–24 months) with ASD, and suggest this may drive the early brain overgrowth that has been observed in ASD. In keeping with our results, other studies have found no group differences in SA in preschoolers [62], or children and adolescents [24]. However, lower SA has been observed in children with ASD aged 9–20 years, normalising in adulthood [9], as well as in a sample of male adults with ASD [7]. There is evidence that CT peaks around 1 or 2 years of age, and gradually declines thereafter into adolescence [63], whereas SA develops rapidly in the first year of life, and continues to

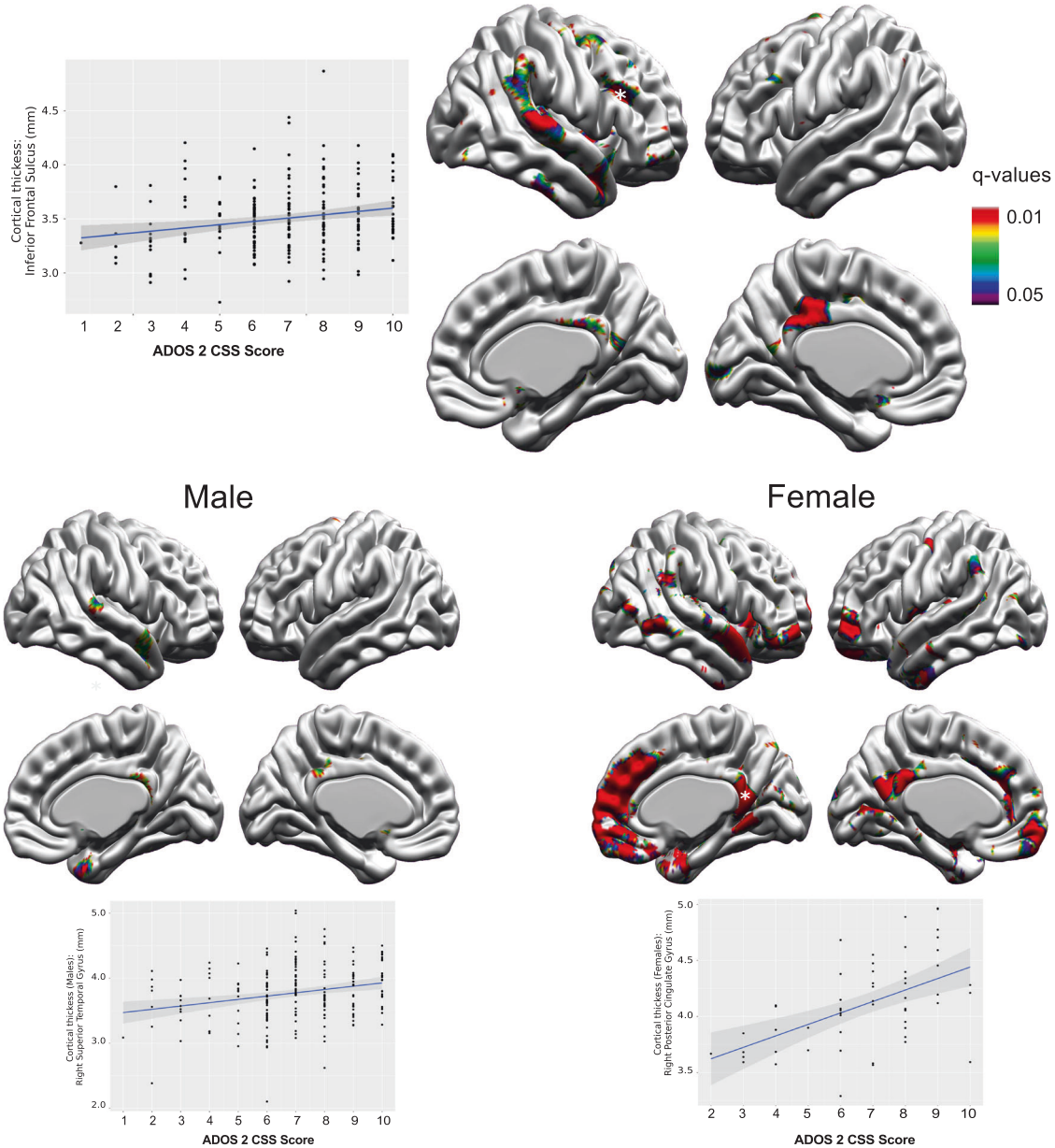


Fig. 4 Relation between cortical thickness (CT) and ADOS-calibrated severity scores (CSS). Relationship between ADOS-2 CSS and CT in individuals with autism spectrum disorder (ASD), shown at a peak vertex in the inferior frontal sulcus (IFS). ADOS-2 severity was positively correlated with CT, primarily in the right hemisphere, in regions which show significantly greater CT in individuals with ASD relative to controls. Correlations between CT and CSS were observed in distinct regions between males and females. In the female sample, there was a significant positive relationship between CT and severity, primarily in prefrontal and temporal regions. In the males, only very minimal regions showed this significant relationship, observed in the superior temporal gyrus and temporal pole. Shown at 5% false discovery rate (FDR)

gradually expand into late childhood or adolescence, before declining [63, 64]. Therefore, it is possible that the early increases in SA in ASD observed in previous studies normalise after this period of rapid development, and thus were not captured in our sample, which has only very few individuals between the ages of 2 and 5 years.

With this large dataset, we hoped to reconcile some of the inconsistencies reported in the literature with regard to cortical phenotypes of ASD. While many other neuroimaging studies of ASD have reported greater CT values [5, 15, 31], others have reported lower thickness [65], or no differences [9]. Our findings of greater CT in ASD are largely in agreement with other large-scale neuroimaging studies, including studies using the ABIDE dataset [5, 15, 16, 66] and recent findings by the ENIGMA consortium [6]. However, the recent ENIGMA study, in addition to greater CT in ASD in the frontal and posterior cingulate cortices, also reports significantly less CT in ASD

in the temporal and parahippocampal cortices. We found no regions of significantly lower CT values; conversely, we observed greater CT in multiple temporal regions. Methodological differences may account for the disparity between our results and those of the ENIGMA study, as well as others reporting decreased CT in ASD. These differences include our rigorous QC (more discussion on this below), the analysis of region-specific differences using the vertex-wise extension of the prospective meta-analysis technique (instead of the regions of interest approach in the ENIGMA study), differences in image processing pipelines and differences in sample characteristics (despite some overlap between our and the ENIGMA sample [ABIDE sample and ~150 controls from the Toronto sample]). Interestingly, the ENIGMA study found that the mixed-effect models strategy yielded more significant results than the meta-analytic technique, whereas we found the opposite. It is unclear how the choice of statistical method interacts with these other factors, however, we believe that the meta-analytic model better deals with the possible confounding variables and variability between sites, and that the mixed-effects model may be less sensitive to capturing small effect sizes through the noise introduced by this variability.

Other studies using the ABIDE dataset have likewise found abnormalities in CT in ASD, in regions overlapping with our results, but of varying magnitude and direction [5, 15, 16]. Most consistent between these studies is the observation of greater CT in individuals with ASD in the STG, as well as frontal regions. However, it should be noted that most of these studies examine males only, and thus are more appropriately interpreted in comparison with our male-specific results.

QC likely greatly contributes to the inconsistencies in the literature; many studies do not describe their QC procedures in detail, rendering it difficult to assess the impact that motion or inaccurate segmentation may have on reported results. In our study, particular attention was given to motion artefact at the level of the raw input images, as in-scanner motion is known to cause apparent cortical thinning due to blurring of the GM–WM boundary [28, 67]. Thus, inadequate QC could lead to results of greater CT in individual with ASD, a population likely to move while being scanned, being attenuated or obscured by this effect. Importantly, in our sample, when no or minimal QC was implemented, CT differences (greater in ASD) that were observed in the QC sample were greatly attenuated. In addition, regions of decreased CT in the bilateral temporal poles and left orbitofrontal cortex were also observed in individuals with ASD (Supplementary Fig. S21). Decreased CT in these regions has previously been reported to be associated with motion [29, 67], and these results highlight the potential for motion to confound results.

In addition to the issue of QC, it is often unclear to what extent case–control differences reported in the literature are influenced by factors contributing to the heterogeneity observed, such as age, sex, FIQ and severity [68]. Thus, another primary objective of this work was to begin to parse this heterogeneity observed in ASD, and determine to what extent these factors influence the reported diagnostic differences in neuroanatomy observed in previous studies, and the variability in these results. While these factors have been demonstrated to impact the neuroanatomical alterations in ASD [17, 22, 26], many studies do not take them into account when examining case–control differences.

In particular, the issue of sex differences in ASD has been receiving more attention recently, yet still studies examining neuroanatomical sex differences are rare, and have largely been underpowered due to small samples sizes of females with ASD [26]. Of existing studies examining sex differences in CT specifically, results are varied: one such study found a sex-by-diagnosis interaction, with lower CT in ASD females, but greater CT in ASD males [34], while others report no difference [6, 33]. Even with our large sample and proportion of females with ASD (362 males and 129 females with ASD), we do not detect a significant sex-by-diagnosis interaction. However, when stratifying by sex, we demonstrate both qualitatively and quantitatively distinct diagnostic effects in males and females, as well as a sex-specific relationship between ASD symptom severity and CT. Our overall case–control results much more closely reflect those of the male-only findings, suggesting the female differences (observed in different regions, and with larger effect sizes) are obscured due to the small sample. Interestingly, the relationship between CT and ASD symptom severity seemed to be driven primarily by the females. This is in spite of the fact that in this sample, males and females do not differ significantly in ASD symptom severity or FIQ; suggesting that females perhaps need more substantial neuroanatomical alterations to result in the same level of clinical presentation as in males (in keeping with the female protective hypothesis [69, 70]). These results highlight the importance of taking biological sex into account when studying ASD, as well as the urgent need for studies examining neuroanatomical sex differences in ASD in larger samples.

Age has been a significant contributor to the heterogeneity observed in ASD. Results of studies examining different age ranges of ASD, in particular in those with small sample sizes, are often conflicting or inconsistent. Recent large-scale studies examining wide age ranges that have attempted to reconcile these inconsistencies have reported CT differences in childhood and early adolescence, followed by normalisation of group differences later in life [5, 6, 66]. While we cannot strictly make inferences about cortical development from our cross-sectional dataset, here,

we seem to recapitulate these results to an extent, though the results observed in our age-centred analysis are subtle. This possible attenuation and eventual disappearance of diagnostic group differences in adolescence and adulthood could be the result of accelerated cortical thinning in ASD after an initial period of overgrowth, as has been observed in previous longitudinal samples [5, 22], as well as post-mortem studies [71]. We also demonstrated a linear model to be the best fit for the majority of our dataset, across most of the cortex, as opposed to the curvilinear trajectories that have been reported by other studies [5, 6]. This may be due, in part, to the meta-analytic technique we chose to employ, which necessitated conducting the model fit on a per site basis, as some smaller sites may lack the power to model higher order trajectories. Improved QC in our study may also play a role, as a recent study demonstrated that after strict QC, previously observed higher order trajectories were mostly replaced by linear effects [30]. While some early general population studies reported a peak in CT in late childhood followed by a decline [72, 73], more recent studies, including those using generalised additive mixed models [74, 75], have reported a monotonic decrease in CT from around 2 years of age [22, 30, 44, 76, 77]. Our findings, though cross-sectional, seem to support this reported linear decline in CT, rather than a peak later in childhood. Taken together, our findings may help further clarify the recent changes in our understanding of neurotypical and atypical cortical developmental trajectories [72, 73] as these models continue to evolve in relation to the greater awareness of potential age-related biases related to motion and image processing QC. However, given that our data are not longitudinal, and the inclusion of limited number of adults, these results should be interpreted with these caveats in mind. Larger, longitudinal studies will be necessary to confirm these findings.

Few studies have examined the potential moderating effects of IQ on the neuroanatomy of ASD, though there is some evidence suggesting that individuals with a diagnosis of Asperger's syndrome (with average or above average IQ) present with milder neuroanatomical atypicalities compared with lower IQ individuals [27, 78]. Despite our sample being skewed to the cognitively higher functioning end of the spectrum, our results seem to align with these findings as we observed greater alterations in the lower FIQ part of our sample. Further, our observation of an inverse relationship between CT and FIQ in individuals with ASD, with the opposite or no relationship in controls, is aligned with previous studies of ASD [17], as well as in typically developing individuals [79]. Shaw et al. [80] also demonstrated that IQ is differentially associated with CT in children compared with adults; future larger-scale work should examine three way relationships between IQ, CT and age in the context of ASD, as well as the extent to which group differences observed may be attributable to lower intellectual functioning rather than simply ASD diagnosis.

As ADOS versions and modules were not consistent across sites, we could not directly test the relations between region-specific cortical alterations and specific ADOS symptom domains in the whole sample. However, the positive relationship between ADOS-2 CSS and CT observed in a subset of individuals with ASD, in regions where case–control differences were observed, suggests a functional relevance of these cortical alterations. Some of the strongest group differences in both the overall sample and in the symptom-based analyses were observed in the STG and inferior frontal gyrus (IFG) and might reflect the social communication deficits that are characteristic of ASD [81–84]. Interestingly, the IFG and STG were also the regions where the strongest case–control differences in CT were observed in males, but not in females.

The results presented here should be interpreted with respect to several limitations. First, in order to amass the significant amount of data presented here, we were required to pool already collected data from multiple sites. The lack of standardisation across sites of magnetic resonance imaging acquisition, inclusion criteria and clinical assessments should be considered. While the meta-analytic statistics used pool common effect sizes across sites, the impact of this lack of standardisation will certainly have an impact on our results. The lack of standardised measures across sites made examination of heterogeneity associated with specific ASD symptoms challenging. As a result, the impact of important factors such as socioeconomic status and parental education (which were not available for any of our sample) could not be ascertained. Similarly, we could not directly assess the impact of specific comorbid diagnoses (which were collected and coded inconsistently between sites); however, based on the results of our analysis including only individuals with no comorbidities, the inclusion of individuals with ASD with comorbid features did not seem to substantially impact our results, though this may have added further variability and attenuated the effect of group differences observed. More targeted investigations into the relationship between common ASD-specific comorbidities and the clinical and neurobiological heterogeneity commonly observed in ASD is necessary. Please see Supplementary Tables S3 and S4 for details on clinical and demographic data available per site. The statistical analysis method itself may also, in turn, be limited in its ability to detect small effects within each site, as well as curvilinear relationships with age or FIQ in the smaller samples.

In addition, there are two considerations which would have improved our ability to better understand factors impacting heterogeneity. The first is the absence of genetic data. While ASD is highly heritable, it has been associated with a diverse number of risk genes [85–87] and rare copy

number variants [88, 89]. These genotypes have been observed to impact the heterogeneity of ASD and require further consideration. The second is the use of longitudinal data to truly model intra-individual change over time to better define alterations in neuroanatomical trajectories [45, 90]. It is possible, given the large sample size used that we have partially overcome this limitation given that our results are consistent with at least one large, longitudinal study examining cortical development in ASD [22]. Nonetheless, further investigation with large longitudinal samples that include males and females are clearly needed.

Finally, further consideration of the demographics of our sample is needed when interpreting our findings. This includes being cautious regarding interpretation of findings in the part of the sample >30 years old, as this represents a smaller subset of the study cohort. Second, the unbalanced male/female distribution requires further consideration. It is likely that we are only detecting the largest effect size differences between ASD and control females and there are likely smaller effects that we are underpowered to detect. Finally, individuals excluded due to QC were younger, had lower IQ and higher severity scores, and included a higher proportion of male and ASD individuals; thus biasing and further skewing our sample towards higher IQ individuals (see Supplementary Table S5). We acknowledge that smaller studies might not have the option of excluding such a large proportion of their data. However, in light of the potential contribution of motion and data quality to inconsistencies in the literature, there are certain steps that should be taken to ensure proper quality, and thus reliability, of data. These include the use of prospective motion correction techniques such as vNavs volumetric navigators [91], the recruitment of larger samples with the knowledge that there may be a large proportion of data that could not be used in statistical analyses, to book sufficient scanner time so as to allow re-scanning where necessary, and, in the case of small samples, to augment the sample using publicly available or collaborator data for replication purposes. The potential exploration and subsequent use of automated motion/quality scores as confounding variables in analyses could also be considered [92, 93]. Our thorough and rigorous manual QC was initiated and performed prior to the availability of these kinds of methods, thus we have not included these methods in our analysis. Nonetheless, we believe that the final QC used in this sample is extremely thorough.

Our findings address limitations in the literature regarding cortical neuroanatomy in ASD by combining multiple datasets. Our sample of 1327 individuals allowed us to detect significant group differences in the whole sample, as well as to examine potential sources of heterogeneity in relation to sex, age and FIQ, and their impact on cortical alterations in ASD. These findings highlight the importance of taking into account factors contributing to the phenotypic heterogeneity in ASD when examining the neuroanatomy in a supervised manner [68], which could further our research of the neurobiology of ASD.

Code availability

R code used to conduct the prospective meta-analyses described here is available from the corresponding authors upon request.

MRC AIMS Consortium: Anthony J. Bailey (Oxford), Simon Baron-Cohen (Cambridge), Patrick F. Bolton (IoPPN), Edward T. Bullmore (Cambridge), Sarah Carrington (Oxford), Marco Catani (IoPPN), Bhismadev Chakrabarti (Cambridge), Michael C. Craig (IoPPN), Eileen M. Daly (IoPPN), Sean C. L. Deoni (IoPPN), Christine Ecker (IoPPN), Francesca Happé (IoPPN), Julian Henty (Cambridge), Peter Jezzard (Oxford), Patrick Johnston (IoPPN), Derek K. Jones (IoPPN), Meng-Chuan Lai (Cambridge), Michael V. Lombardo (Cambridge), Anya Madden (IoPPN), Diane Mullins (IoPPN), Clodagh M. Murphy (IoPPN), Declan G. M. Murphy (IoPPN), Greg Pasco (Cambridge), Amber N. V. Ruigrok (Cambridge), Susan A. Sadek (Cambridge), Debbie Spain (IoPPN), Rose Stewart (Oxford), John Suckling (Cambridge), Sally J. Wheelwright (Cambridge) and Steven C. Williams (IoPPN).

Funding This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative, in the form of a graduate student fellowship to SAB. The Autism Imaging Multicentre Study Consortium was funded by the Medical Research Council United Kingdom grant G0400061. The Cambridge Family Study of Autism was funded by a Clinical Scientist Fellowship from the UK Medical Research Council (MRC) (G0701919). AR was supported by funding from the Intramural Research Program of the NIMH (Clinical trial reg. no. NCT00001246, clinicaltrials.gov; NIH Annual Report Number, ZIA MH002794, Protocol ID 89-M-0006). The Toronto sample was gathered from studies supported by grants MOP-119541, MOP-106582 and MOP-14237 from the Canadian Institutes of Health Research (to MT), and from the POND Network, funded by the Ontario Brain Institute (grant IDS-I l-02 to EA and JL), an independent non-profit corporation, funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred. JL received funding from the Canadian Institute for Health Research. MMC received funding from the Canadian Institute for Health Research, the Natural Sciences and Engineering Research Council, the Fonds de recherche du Québec – Santé and McGill University's Healthy Brains for Healthy Lives initative. SBC was supported by the Autism Research Trust. DGM was supported in this work by funding from the MRC UK, the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, and King's College London (Medical Research Council grant no. G0400061 to DGMM). DGM, SBC, AR, JS, CE, and RH are also supported by EU-AIMS and AIMS-2 TRIALS. EU-AIMS receives support from the Innovative Medicines Initiative (IMI) Joint Undertaking (JU) under grant agreement no. 115300, the resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (grant FP7/2007- 2013). AIMS-2 TRIALS received support from EFPIA and AUTISM SPEAKS, Autistica, and SFARI, and funding from the IMI 2 JU under grant agreement no. 777394, with support from the European Union's Horizon 2020 research and innovation programme. MVL was supported by an ERC Starting Grant (ERC-2017-STG; 755816). M-CL was supported by the O'Brien

Scholars Program within the Child and Youth Mental Health Collaborative at the Centre for Addiction and Mental Health and the Hospital for Sick Children, Toronto, and the Slifka-Ritvo Award for Innovation in Autism Research by the Alan B. Slifka Foundation and the International Society for Autism Research. The authors would like to thank the investigators and participants in the ABIDE dataset. Funding sources for each individual sites are provided on the official ABIDE website (http://fcon_1000.projects.nitrc.org/indi/abide/).

Compliance with ethical standards

Conflict of interest DGM reported receiving honoraria from Roche for being on a scientific advisory board. No other conflicts of interest were reported.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Courchesne E, Carper R, Akshoomoff N. Evidence of brain overgrowth in the first year of life in autism. JAMA. 2003;290:337–44.
2. Hazlett HC, et al. Magnetic resonance imaging and head circumference study of brain size in autism: birth through age 2 years. Arch Gen Psychiatry. 2005;62:1366–76.
3. Wallace GL, Dankner N, Kenworthy L, Giedd JN, Martin A. Age-related temporal and parietal cortical thinning in autism spectrum disorders. Brain. 2010;133:3745–54.
4. Raznahan A, et al. Mapping cortical anatomy in preschool aged children with autism using surface-based morphometry. NeuroImage Clin. 2013;2:111–9.
5. Khundrakpam BS, Lewis JD, Kostopoulos P, Carbonell F, Evans AC. Cortical thickness abnormalities in autism spectrum disorders through late childhood, adolescence, and adulthood: a large-scale MRI study. Cereb Cortex. 2017;27:1721–31.
6. van Rooij D, et al. Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: results from the ENIGMA ASD Working Group. Am J Psychiatry. 2018;175:359–69.
7. Ecker C, et al. Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. JAMA Psychiatry. 2013;70:59–70.
8. Ohta H, et al. Increased surface area, but not cortical thickness, in a subset of young boys with autism spectrum disorder. Autism Res. 2016;9:232–48.
9. Mensen VT, et al. Development of cortical thickness and surface area in autism spectrum disorder. NeuroImage Clin. 2017;13:215–22.
10. Hazlett HC, et al. Early brain development in infants at high risk for autism spectrum disorder. Nature. 2017;542:348–51.
11. Mandy W, et al. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. J Autism Dev Disord. 2012;42:1304–13.
12. Mandic-Maravic V, et al. Sex differences in autism spectrum disorders: does sex moderate the pathway from clinical symptoms to adaptive behavior? Sci Rep. 2015;5:10418.
13. Klin A, et al. Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: the Vineland and the ADOS. J Autism Dev Disord. 2007;37:748–59.
14. Vivanti G, Barbaro J, Hudry K, Dissanayake C, Prior M. Intellectual development in autism spectrum disorders: new insights from longitudinal studies. Front Hum Neurosci. 2013;7:354.
15. Haar S, Berman S, Behrmann M, Dinstein I. Anatomical abnormalities in autism? Cereb Cortex. 2016;26:1440–52.
16. Valk SL, Di Martino A, Milham MP, Bernhardt BC. Multicenter mapping of structural network alterations in autism. Hum Brain Mapp. 2015;36:2364–73.
17. Misaki M, Wallace GL, Dankner N, Martin A, Bandettini PA. Characteristic cortical thickness patterns in adolescents with autism spectrum disorders: interactions with age and intellectual ability revealed by canonical correlation analysis. Neuroimage. 2012;60:1890–1901.
18. Richter J, et al. Reduced cortical thickness and its association with social reactivity in children with autism spectrum disorder. Psychiatry Res. 2015;234:15–24.
19. Wallace GL, et al. Longitudinal cortical development during adolescence and young adulthood in autism spectrum disorder: increased cortical thinning but comparable surface area changes. J Am Acad Child Adolesc Psychiatry. 2015;54:464–9.
20. Raznahan A, et al. Cortical anatomy in autism spectrum disorder: an in vivo MRI study on the effect of age. Cereb Cortex. 2010;20:1332–40.
21. Greimel E, et al. Changes in grey matter development in autism spectrum disorder. Brain Struct Funct. 2013;218:929–42.
22. Zielinski BA, et al. Longitudinal changes in cortical thickness in autism and typical development. Brain. 2014;137:1799–812.
23. Lin H-Y, Ni H-C, Lai M-C, Tseng W-YI, Gau SS-F. Regional brain volume differences between males with and without autism spectrum disorder are highly age-dependent. Mol Autism. 2015;6:29.
24. Sussman D, et al. The autism puzzle: diffuse but not pervasive neuroanatomical abnormalities in children with ASD. NeuroImage Clin. 2015;8:170–9.
25. Zhang W. et al. Revisiting subcortical brain volume correlates of autism in the ABIDE dataset: effects of age and sex. Psychol Med. 2017. https://doi.org/10.1017/S003329171700201X.
26. Lai MC, et al. Imaging sex/gender and autism in the brain: etiological implications. J Neurosci Res. 2017;95:380–97.
27. Lotspeich LJ, et al. Investigation of neuroanatomical differences between autism and Asperger syndrome. Arch Gen Psychiatry. 2004;61:291–8.
28. Alexander-Bloch A, et al. Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. Hum Brain Mapp. 2016;2397:2385–97.
29. Pardoe HR, Kucharsky Hiess R, Kuzniecky R. Motion and morphometry in clinical and nonclinical populations. Neuroimage. 2016;135:177–85.
30. Ducharme S, et al. Trajectories of cortical thickness maturation in normal brain development--the importance of quality control procedures. Neuroimage. 2016;125:267–79.
31. Hardan AY, Muddasani S, Vemulapalli M, Keshavan MS, Minshew NJ. An MRI study of increased cortical thickness in autism. Am J Psychiatry. 2006;163:1290–2.
32. Hyde KL, Samson F, Evans AC, Mottron L. Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry. Hum Brain Mapp. 2010;31:556–66.
33. Schaer M, Kochalka J, Padmanabhan A, Supekar K, Menon V. Sex differences in cortical volume and gyrification in autism. Mol Autism. 2015;6:42.
34. Ecker C, et al. Association between the probability of autism spectrum disorder and normative sex-related phenotypic diversity in brain structure. JAMA Psychiatry. 2017;74:329.
35. Lange N, et al. Longitudinal volumetric brain changes in autism spectrum disorder ages 6-35 years. Autism Res. 2015;8:82–93.
36. Di Martino A, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry. 2014;19:659–67.

37. Di Martino A. et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Scientific Data. 2017;4:170010.
38. Zijdenbos AP, Forghani R, Evans AC. Automatic 'pipeline' analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans Med Imaging. 2002;21:1280–91.
39. van Erp TGM, et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Mol Psychiatry. 2016;21:547–53.
40. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007;82:591–605.
41. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1:97–111.
42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300.
43. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19:716–23.
44. Walhovd KB, Fjell AM, Giedd J, Dale AM, Brown TT. Through thick and thin: a need to reconcile contradictory results on trajectories in human cortical development. Cereb Cortex. 2017;27:1472–81.
45. Chakravarty MM, et al. Striatal shape abnormalities as novel neurodevelopmental endophenotypes in schizophrenia: a longitudinal study. Hum Brain Mapp. 2015;36:1458–69.
46. Schuetze M, et al. Morphological alterations in the thalamus, striatum, and pallidum in autism spectrum disorder. Neuropsychopharmacology. 2016;41:2627–37.
47. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. J Autism Dev Disord. 2009;39:693–705.
48. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord. 2001;31:5–17.
49. Courchesne E, Moses P, Pierce K, Pizzo S. Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. Neurology. 2001;57:245–54.
50. Schumann CM, et al. Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism. J Neurosci. 2010;30:4419–27.
51. Hazlett HC, et al. Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. Arch Gen Psychiatry. 2011;68:467–76.
52. Redcay E, Courchesne E. When is the brain enlarged in autism? A meta-analysis of all brain size reports. Biol Psychiatry. 2005;58:1–9.
53. Bauman ML, Kemper TL. Neuroanatomic observations of the brain in autism: a review and future directions. Int J Dev Neurosci. 2005;23:183–7.
54. Schumann C, Noctor SC, Amaral DG. Neuropathology of autism spectrum disorders: postmortem studies. Autism Spectrum Disorders 2012;1**:**62–74.
55. Casanova MF, et al. Minicolumnar abnormalities in autism. Acta Neuropathol. 2006;112:287–303.
56. Huttenlocher PR. Morphometric study of human cerebral cortex development. Neuropsychologia. 1990;28:517–27.
57. Ecker C. The neuroanatomy of autism spectrum disorder: an overview of structural neuroimaging findings and their translatability to the clinical setting. Autism. 2017;21:18–28.
58. Tang G, et al. Loss of mTOR-dependent macroautophagy causes autistic-like synaptic pruning deficits. Neuron. 2014;83:1131–43.
59. Avino TA, Hutsler JJ. Abnormal cell patterning at the cortical gray–white matter boundary in autism spectrum disorders. Brain Res. 2010;1360:138–46.
60. Andrews DS, et al. In vivo evidence of reduced integrity of the gray-white matter boundary in autism spectrum disorder. Cereb Cortex. 2017;27:877–87.
61. Bezgin G, Lewis JD, Evans AC. Developmental changes of cortical white–gray contrast as predictors of autism diagnosis and severity. Transl Psychiatry. 2018;8:249.
62. Smith E, et al. Cortical thickness change in autism during early childhood. Hum Brain Mapp. 2016;2629:2616–29.
63. Gilmore JH, Knickmeyer RC, Gao W. Imaging structural and functional brain development in early childhood. Nat Rev Neurosci. 2018;19:123–37.
64. Lyall AE, et al. Dynamic development of regional cortical thickness and surface area in early childhood. Cereb Cortex. 2015;25:2204–12.
65. Ecker C, et al. The effect of age, diagnosis, and their interaction on vertex-based measures of cortical thickness and surface area in autism spectrum disorder. J Neural Transm. 2014;121:1157–70.
66. Bethlehem RAI, Seidlitz J, Romero-Garcia R, Lombardo MV. Using normative age modelling to isolate subsets of individuals with autism expressing highly age-atypical cortical thickness features. bioRxiv. 2018. https://doi.org/10.1101/252593.
67. Reuter M, et al. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. Neuroimage. 2015;107:107–15.
68. Lombardo MV, Lai M-C, Baron-Cohen S. Big data approaches to decomposing heterogeneity across the autism spectrum. Mol Psychiatry. 2019. https://doi.org/10.1038/s41380-018-0321-0.
69. Werling DM, Geschwind DH. Understanding sex bias in autism spectrum disorder. Proc Natl Acad Sci USA. 2013;110:4868–9.
70. Cauvet É, et al. Sex differences along the autism continuum: a twin study of brain structure. Cereb Cortex. 2019;29:1342–50.
71. Hutsler JJ, Love T, Zhang H. Histological and magnetic resonance imaging assessment of cortical layering and thickness in autism spectrum disorders. Biol Psychiatry. 2007;61:449–57.
72. Raznahan A, et al. How does your cortex grow? J Neurosci. 2011;31:7174–7.
73. Shaw P, et al. Neurodevelopmental trajectories of the human cerebral cortex. J Neurosci. 2008;28:3586–94.
74. Tamnes CK, et al. Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. J Neurosci. 2017;37:3402–12.
75. Gennatas ED, et al. Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. J Neurosci. 2017;37:5065–73.
76. Brown TT, et al. Neuroanatomical assessment of biological maturity. Curr Biol. 2012;22:1693–8.
77. Amlien IK, et al. Organizing principles of human cortical development—thickness and area from 4 to 30 years: insights from comparative primate neuroanatomy. Cereb Cortex. 2016;26:257–67.
78. Schumann CM, et al. The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. J Neurosci. 2004;24:6392–401.
79. Narr KL, et al. Relationships between IQ and regional cortical gray matter thickness in healthy adults. Cereb Cortex. 2007;17:2163–71.
80. Shaw P, et al. Intellectual ability and cortical development in children and adolescents. Nature. 2006;440:676–9.
81. Redcay E. The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. Neurosci Biobehav Rev. 2008;32:123–42.
82. Verhoeven JS, De Cock P, Lagae L, Sunaert S. Neuroimaging of autism. Neuroradiology. 2010;52:3–14.

83. Herringshaw AJ, Ammons CJ, DeRamus TP, Kana RK. Hemispheric differences in language processing in autism spectrum disorders: a meta-analysis of neuroimaging studies. Autism Res. 2016;9:1046–57.
84. Lombardo MV, et al. Different functional neural substrates for good and poor language outcome in autism. Neuron. 2015;86:567–77.
85. Ellegood J, et al. Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. Mol Psychiatry. 2015;20:118–25.
86. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016;22:345–61.
87. Yuen RKC, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. 2017;20:602–11.
88. Marshall CR, Scherer SW. Detection and characterization of copy number variation in autism spectrum disorder. Methods Mol Biol. 2012;838:115–35.
89. Turner TN, et al. Genomic patterns of de novo mutation in simplex autism. Cell. 2017;171:710.
90. Shaw P, Gogtay N, Rapoport J. Childhood psychiatric disorders as anomalies in neurodevelopmental trajectories. Hum Brain Mapp. 2010;31:917–25.
91. Tisdall MD, et al. Prospective motion correction with volumetric navigators (vNavs) reduces the bias and variance in brain morphometry induced by subject motion. Neuroimage. 2016;127:11–22.
92. Rosen AFG, et al. Quantitative assessment of structural image quality. Neuroimage. 2018;169:407–18.
93. White T, et al. Automated quality assessment of structural magnetic resonance images in children: comparison with visual inspection and surface-based reconstruction. Hum Brain Mapp. 2018;39:1218–31.

Affiliations

Saashi A. Bedford[1,2] · Min Tae M. Park[1,3] · Gabriel A. Devenyi[1,4] · Stephanie Tullo[1,2] · Jurgen Germann[1] · Raihaan Patel[1,5] · Evdokia Anagnostou[6] · Simon Baron-Cohen[7] · Edward T. Bullmore[8] · Lindsay R. Chura[7] · Michael C. Craig[9,10] · Christine Ecker[9,11] · Dorothea L. Floris[7,12] · Rosemary J. Holt[7] · Rhoshel Lenroot[13] · Jason P. Lerch[14,15] · Michael V. Lombardo[7,16] · Declan G. M. Murphy[9] · Armin Raznahan[17] · Amber N. V. Ruigrok[7] · Elizabeth Smith[18] · Michael D. Spencer[7] · John Suckling[8] · Margot J. Taylor[19] · Audrey Thurm[18] MRC AIMS Consortium · Meng-Chuan Lai[7,14,20,21,22] · M. Mallar Chakravarty[1,2,4,5]

[1] Cerebral Imaging Centre, Douglas Mental Health University Institute, Montreal, QC, Canada

[2] Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada

[3] Department of Psychiatry, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

[4] Department of Psychiatry, McGill University, Montreal, QC, Canada

[5] Department of Biological and Biomedical Engineering, McGill University, Montreal, QC, Canada

[6] Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada

[7] Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK

[8] Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, UK

[9] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[10] National Autism Unit, Bethlem Royal Hospital, London, UK

[11] Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Goethe University, Frankfurt am Main, Germany

[12] Hassenfeld Children's Hospital at NYU Langone Department of Child and Adolescent Psychiatry, Child Study Center, New York City, NY, USA

[13] Department of Psychiatry, University of New South Wales, Sydney, NSW, Australia

[14] Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, ON, Canada

[15] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

[16] Department of Psychology, University of Cyprus, Nicosia, Cyprus

[17] Developmental Neurogenomics Unit, Human Genetics Branch, National Institute of Mental Health, Bethesda, MD, USA

[18] Section on Behavioral Pediatrics, National Institute of Mental Health, Bethesda, MD, USA

[19] Diagnostic Imaging, The Hospital for Sick Children, Toronto, ON, Canada

[20] The Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health and Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada

[21] Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[22] Department of Psychiatry, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan