Title: Using normative age modelling to isolate subsets of individuals with autism expressing highly age-atypical cortical thickness features

Authors: *Bethlehem, R.A.I.*[1,2,±], *Seidlitz, J.*[1,3], *Romero-Garcia, R.*[1] *& Lombardo, M.V.* [1,4]

Affiliations:

[1]Department of Psychiatry, University of Cambridge, Cambridge CB2 0SZ, United Kingdom

[2]Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge CB2 8AH, United Kingdom

[3] Developmental Neurogenomics Unit, National Institute of Mental Health, Bethesda, MD 20892, USA.

[4] Department of Psychology, Center for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus;

± Corresponding author:

Dr. Richard A.I. Bethlehem

Department of Psychiatry

Douglas House

18B Trumpington Road

CB5 8AH Cambridge

United Kingdom

rb643@medschl.cam.ac.uk

Abstract

Understanding heterogeneity in neural phenotypes is an important goal on the path to precision medicine for autism spectrum disorders (ASD). Age is a critically important variable in normal structural brain development and examining structural features with respect to age-related norms could help to explain ASD heterogeneity in neural phenotypes. Here we examined how cortical thickness (CT) in ASD can be parameterized as an individualized metric of deviance relative to typically-developing (TD) age-related norms. Across a large sample (n=942 per group) and wide age range (5-40 years), we applied a normative modelling approach that provides individualized whole-brain maps of age-related CT deviance in ASD. This approach isolates a highly age-deviant CT subtype with a median prevalence of 7.6% across all brain regions and prevalence within specific regions that can be greater than 10%. Individuals in this ASD subtype are statistical outliers in case-control models and this small subset of individuals drives a large majority of small effect results from case-control comparisons. Testing age-normed CT scores also highlights on-average differentiation and associations with behavioural symptomatology that is separate from insights gleaned from traditional case-control approaches. This work showcases a novel individualized approach for understanding ASD heterogeneity that could further prioritize work on a subset of individuals with significant cortical pathophysiology represented in age-related CT deviance.

Introduction

Autism spectrum disorder (ASD) is a clinical behavioural consensus label we give to a diverse collection of patients with social-communication difficulties and pronounced repetitive, restricted, and stereotyped behaviours [1]. Beyond the single label of ASD, patients are in fact widely heterogeneous in phenotype, but also with regards to the diversity of different aetiologies [2]. Even within mesoscopic levels of analysis such as examining brain endophenotypes, heterogeneity is the rule rather than the exception [3]. At the level of structural brain variation, neuroimaging studies have identified various neuroanatomical features that might help identify individuals with autism or reveal elements of a common underlying biology [3]. However, the vast neuroimaging literature is also considerably inconsistent, with reports of hypo- or hyper-connectivity, cortical thinning versus increased grey or white matter, brain overgrowth, arrested growth, etc, leaving stunted progress towards understanding mechanisms driving cortical pathophysiology in ASD.

Multiple explanations could be behind this inconsistency across the literature. Methodology widely differs across studies (e.g., low statistical power, different ways of estimating morphology or volume) and is likely a very important factor [4,5]. Initiatives such as the ABIDE [6] have made it possible to significantly boost sample size by pooling together data from several different studies. However, within-group heterogeneity in the autism population also immediately stands out as another factor obscuring consistency in the literature, particularly when the dominant approach of case-control models largely ignores heterogeneity within the ASD population. In particular, some autism-related heterogeneity reported in literature might be explained by factors such as age [7,8]. Indeed, with regards to structural brain features of interest for study in ASD (e.g., volume, cortical thickness, surface area), these features change markedly over development [9–12]. However, typical approaches towards dealing with age revolve around group statistical modelling of age as the variable of interest or removing age as a covariate and then parametrically modelling on-average differences between case versus controls. While these are common approaches in the literature, they do not immediately provide individualized estimates of age-related deviance. In contrast, normative models of age-related variation may likely be an important alternative to these approaches and may mesh better with some conceptual views of deviance in ASD as being an extreme of typical population norms [13]. In contrast to the canonical case-control model, normative age modelling allows for computation of individualized metrics that can hone in on the precision information we are interested in – that is, deviance expressed in specific ASD individuals relative to non-ASD norms. Such an approach may be a fruitful way forward in isolating individuals whom are ‘statistical outliers’. The reasons behind why these individuals are

outliers relative to non-ASD norms may be of potential clinical and/or mechanistic importance. Indeed, if we are to move forward towards stratified psychiatry and precision medicine for ASD [14], we must go beyond case-control approaches and employ dimensional approaches that can tell us information about which individuals are atypical and how or why they express such atypicality.

In the present study, we employ normative modelling on age-related variability as a means to individualize our approach to isolating specific subsets of patients with very different neural features. Here we focus specifically on a neural feature of cortical morphology known as cortical thickness (CT). CT is a well-studied neuroanatomical feature thought to be differentially affected in autism and has received increasing attention in recent years [12,15–19]. Recent work from our group also identified a genetic correlate for autism specific CT variation despite considerable heterogeneity in group specific CT in children with autism (Romero-Garcia, Warrier, Bullmore, Baron-Cohen, & Bethlehem, in press). A study examining ABIDE I cohort data discovered case-control differences in CT, albeit very small in effect size (Haar et a.., 2014). Similarly, the most recent and largest study to date, a mega-analysis combining data from ABIDE and the ENIGMA consortium, also indicated very small on-average case-control differences in cortical thickness restricted predominantly to areas of frontal and temporal cortices, and indicate very subtle age-related between-group differences and substantial within-group age-related variability [19]. Overall, these studies emphasize two general points of importance. First, age or developmental trajectory is extremely important [8,21–23]. Second, given the considerable within-group age-related variability and the presence of a large majority of null and/or very small between-group effects, rather than attempting to find on-average differences between all cases versus all controls, we should shift our focus to capitalize on this dimension of large age-related variability and isolate autism cases that are at the extremes of this dimension of normative variability.

Given our novel approach of age-related normative CT modelling, we first compare the utility of age-related normative modelling directly against case-control models. We then describe the prevalence of ASD cases that show clinically significant age-related deviance in CT (i.e. > 2 SD from age-related norms) and show how a metric of continuous variability in age-related deviance in CT is expressed across the cortex in autism. Finally, we identify important age-deviant CT-behaviour associations and assess whether such dimensional analyses associated with behaviour identify similar or different regions than typical case-control analyses.

Methods

Participants

In this study, we first sought to leverage large neuroimaging datasets to yield greater statistical power for identifying subtle effects. To achieve this, we utilized the ABIDE datasets (ABIDE I and II) (see Supplementary Table S1 and S2 for full list of sites used in the current analyses). Groups were subsequently matched on age using the non-parametric nearest neighbour matching procedure implemented in the Matchit package in R (https://cran.r-project.org/web/packages/MatchIt/index.html) [24]. After matching case and control groups we were left with a sample size N=942 per group (Table 1). Because of power limitations in past work with small samples, we conducted an a priori statistical power analysis indicating that a minimum case-control effect size of d = 0.168 could be detected at this sample size with 80% power at a conservative alpha set to 0.005 [25]. For analyses looking at brain-behaviour associations, we examined a subset of patients with the data from the SRS (N=418) and ADOS total scores (N=521). With same power and alpha levels the minimum effect for SRS is r = 0.177 and r = 0.158 for the ADOS.

Table 1: Sample characteristics of age

	Autism				Control				Group difference	
	n	**mean**	**median**	**sd**	**n**	**mean**	**median**	**sd**	**W**	**p**
Male	816	16.15	13.55	8.89	714	16.52	13.57	8.87	285120	0.472
Female	126	15.05	12.58	8.28	228	13.32	11.02	5.48	15816	0.115

Imaging processing and quantification

For the ABIDE I release we used the available FreeSurfer reconstructions. For the ABIDE II release cortical surface reconstruction was performed using the MPRAGE (T1) image of each participant in FreeSurfer v5.3.0 (http://surfer.nmr.mgh.harvard.edu/). The reconstruction pipeline performed by FreeSurfer “recon-all” involved intensity normalization, registration to Talairach space, skull stripping WM segmentation, tessellation of the WM boundary, and automatic correction of topological defects. Briefly, non-uniformity intensity correction algorithms were applied before skull stripping [26], resulting in resampled isotropic images of 1mm. An initial segmentation of the white matter tissue was performed to generate a tessellated representation of the WM/GM boundary. The resulting surface is deformed outwards to the volume that maximize the intensity contrast between GM and cerebrospinal fluid, generating the pial surface [27]. Resulting surfaces were constrained to a

spherical topology and corrected for geometrical and topological abnormalities. Cortical thickness of each vertex was defined as the shortest distance between vertices of the GM/WM boundary and the pial surface [28]. Because surface reconstruction failed for 13 subjects (out of 1114), we chose to not conduct manual segmentations and excluded these subjects from any subsequent analysis.

Across both ABIDE I and ABIDE II cortical thickness was extracted for each subject using two different parcellations schemes: an approximately equally-sized parcellation of 308 regions (~500mm^2 each parcel) [29,30] and a parcellation of 360 regions derived from multi-modal features extracted from the Human Connectome Project (HCP) dataset [31]. The 308-region parcellation was constructed in the FreeSurfer fsaverage template by subdividing the 68 regions defined in the Desikan-Killiany atlas [32]. Thus, each of the 68 regions was sequentially sub-parcellated by a backtracking algorithm into regions of ~500mm^2, resulting in a high resolution parcellation that preserved the original anatomical boundaries defined in the original atlas [33]. Surface reconstructions of each individual were co-registered to the fsaverage subject. The inverse transformation was used to map both parcellation schemes into the native space of each participant.

Statistical analyses

There are likely many variables that contribute to variability in CT between individuals and across the brain. In order to visually assess the contribution of some prominent sources of variance we adopted a visualization framework derived from gene expression analysis (http://bioconductor.org/packages/variancePartition) [34] and included the most commonly available covariates in the ABIDE dataset: Age, Gender, Diagnosis, Scanner Site, Full-scale IQ, Verbal IQ, Handedness, SRS and Set (ABIDE I or ABIDE II). Given that ABIDE was not designed as an integrated dataset from the outset, it seems plausible that scanner site might be related to autism or autism-related variables (e.g., some sites might have different case-control ratios or only recruited specific subgroups). Figure 1 shows the ranked contribution of those covariates. Perhaps unsurprisingly, scanner site and age proved to be the most dominant sources of variance (each explaining on average around 15% of the total variance). Our initial conventional analysis was aimed to delineate potential broad case-control differences, as has been done in previous studies [5,19]. We used a linear mixed effects model with scanner site as random effect. Given the potentially strong contribution of age and gender we chose to include both as fixed effect covariates in the model. Multiple comparison correction was implemented with Benjamini-Hochberg FDR at $q<0.05$ [35].

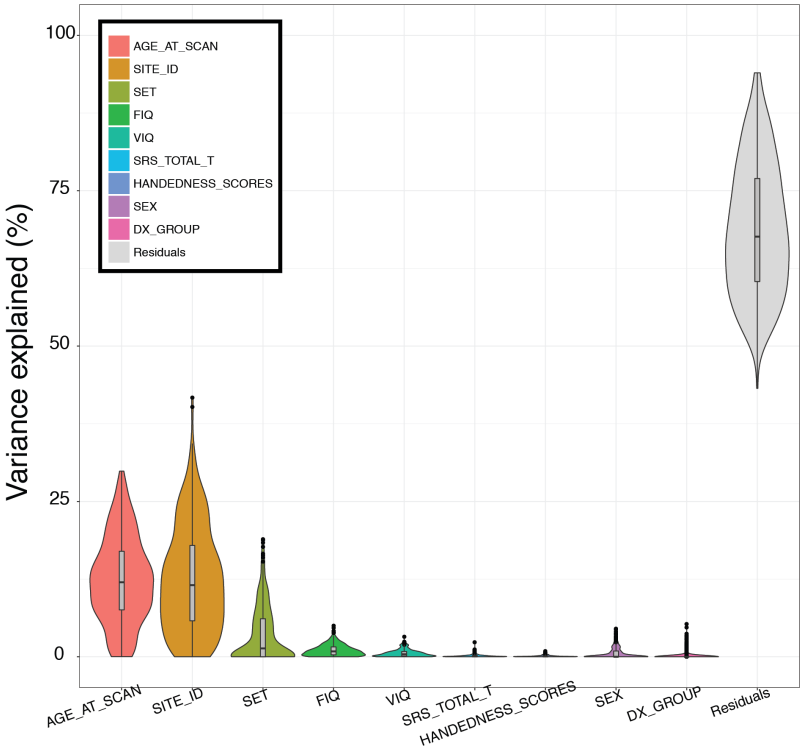

Figure 1: Explained variance in cortical thickness for each covariate

Age-related normative modelling

Normative modelling of age-related CT effects was done utilizing data from the typically-developing group (TD) (see Figure 2 for a schematic overview). We used a local polynomial regression fitting procedure (LOESS), where the local width or smoothing kernel of the regression was determined by the model that provided the smallest sum of squared errors. To align the TD and ASD groups, both were binned into one year age bins. For each age bin and every brain region we computed a normative (TD) mean and standard deviation. These statistical norms were then used to compute a w-score (analogous to a z-score) for every individual with autism. The w-score for an individual reflects how far away their CT is from TD norms in units of standard deviation. Because w-scores are computed for every brain region, we get a w-score map for each participant showing at a precise spatial level how each brain region for that individual is atypical relative to TD norms. Age bins that contained fewer than 2 data-points in the TD group were excluded from subsequent analysis as the standard deviations for these bins would essentially be zero (and thus the w-score could not be computed). Because w-score maps are computed for each individual, we ran hypothesis tests at each brain region to identify regions that show on-average non-zero w-scores – that is, a linear mixed effects model of w-score deviation from zero with scanner site as random effect and stratified by gender (FDR corrected at $q<0.05$). To assess the effect of age-

related individual outliers on the global case-control differences we re-ran the hypotheses tests on w-scores after removing region-wise individual outliers (based on a 2SD cut-off).

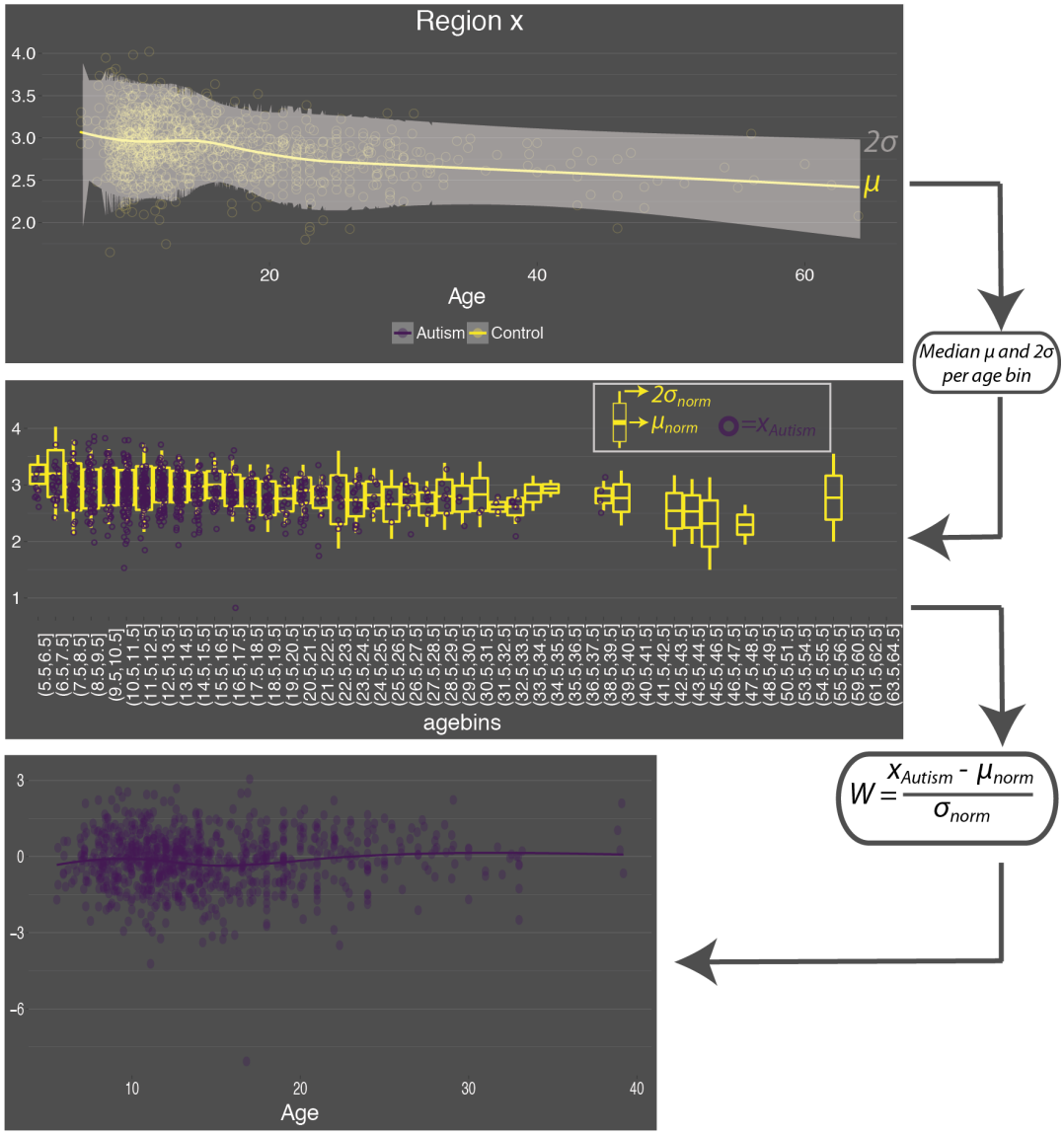

Figure 2: Schematic overview of normative modelling. In first instance LOESS regression is used to estimate the developmental trajectory on CT for every individual brain region to obtain an age specific mean and standard deviation. Then we computed median for each one-year age-bin for these mean and median neurotypical estimates to align them with the ASD group. Next, for each individual with autism and each brain region the normative mean and standard deviation are used to compute a w-score relative to their neurotypical age-bin.

To isolate subsets of individuals with significant age-related CT deviance, we used a cut-off score of 2 standard deviations (i.e. w >= 2 or w<=2). This cut-off allows us to isolate specific ASD patients with markedly abnormal CT relative to age-norms. We then calculated sample prevalence (percentage of all ASD patients with atypical w-scores), in order to describe how frequent such individuals are in the ASD population and across each brain region. A sample prevalence map can then be computed to show the frequency of these patients across each brain region. We also wanted to assess how many patients have markedly atypical w-scores

(beyond 2SD) across a majority of brain regions. This was achieved by computing an individual global w-score ratio as follows:

$$gW = \frac{\sum abs(w) > 2}{\sum abs(w) < 2}$$

We also computed global w-score ratios for positive and negative w regions separately.

Age-related CT deviance relationships with SRS and ADOS

To explore whether the w-scores reflect a potentially meaningful phenotypic feature we next computed Spearman correlations for each brain region between the most commonly shared phenotypic features in ABIDE: ADOS, SRS, SCQ, AQ, FIQ and Age. Resulting p-values matrices were corrected for multiple comparisons using Benjamini-Hochberg FDR correction and only regions surviving and FDR corrected p-value of $p < 0.05$ are reported. Given the large contribution to the explained variance by scanning site we also replicated all correlational analyses on w-scores where scanner site was regressed out (and the region means were added to the residuals to enable ease of interpretation). Given the reduced sample size in the female group and the known interaction between autism and biological sex [36], we conducted normative modelling on the male group only.

Multivariate clustering

Finally, we explored whether the raw CT values could be used in a multivariate fashion to separate groups by diagnosis or illuminate stratification within ASD into subtypes. Here we used k-medoid clustering on t-Distributed Stochastic Neighbour Embedding (tSNE) [37]. Barnes-Hut tSNE was used to construct a 2-dimensional embedding for all parcels in order to be able to run k-medoid clustering in a 2D representation and in order to visually assess the most likely scenario within the framework suggested by Marquand and colleagues [13]. Next, we performed partitioning around medoids (PAM), estimating the optimum number of clusters using the optimum average silhouette width [38].

Data and code availability

Data and code are available on GitHub: https://github.com/rb643/Normative_modeling

Results

Case-control differences versus age-related normative modelling

Our first analysis examined conventional case-control differences. As expected from prior papers utilizing large-scale datasets for case-control analysis (e.g., [5,19]), a small subset of regions (15%, 49/308 regions) pass FDR correction. Of these regions, most are of very small effect size, with 13 of the detected 49 regions showing an effect less than 0.1 standard deviations of difference (Figure 3A and 3B). We suspected that such small effects could be largely driven by a few ASD patients [39] with highly age-deviant CT. Because we also had computed w-scores from our normative age-modelling approach, we identified specific 'statistical outlier' patients with w-scores > 2 standard deviations from typical norms and excluded them from the case-control analysis. This analysis guards against the influence of these extreme outlier patients, and if there are true on-average differences in ASD, the removal of these outlier patients should have little effect on our ability to detect case-control differences. However, rather than continuing to identify 49 regions with small case-control differences, removal of outlier patients now only revealed 12 regions passing FDR correction - a 4-fold decrease in the number of regions detected. Indeed, the majority of case-control differences identifying small on-average effects were primarily driven by this small subset of highly-deviant patients (Figure 4A and 4B, table 2). These remaining 12 regions with small on-average effects were restricted to areas near primary auditory cortex in the superior temporal gyrus, posterior cingulate cortex, and areas of visual cortex.

In contrast to a canonical case-control model, we computed normative models of age which resulted in individualized w-scores that indicate how deviant CT is for an individual compared to typical norms for that age. This modelling approach allows for computation of w-scores for every region and in every patient, thus resulting in a w-score map that can then itself be tested for differences from a null hypothesis of 0, indicating no significant on-average ASD deviance in age-normed CT. These hypothesis tests on normative w-score maps initially revealed no regions surviving FDR correction. However, like the original case-control analysis where we observed that statistical outlier patients significantly influenced the results, we removed these outliers and then re-ran the test for identifying such on-average differences in CT-normed scores. Removing such outliers now boosted the number of detected regions from 0 to 12 (Figure 4C and 4D, table 3). Interestingly, the resulting 12 detected regions with an on-average CT-normed difference are completely non-overlapping with the 12 regions detected in the canonical case-control analysis. These regions are restricted to areas of visual cortex, motor cortex, and para-hippocampal gyrus. Also of

particular note is the reversal of directionality of effect in visual cortex. Whereas, case-control analysis with outliers included signalled the possibility of thicker visual cortices in ASD, this analysis on w-scores and with outliers removed suggests that the majority of the ASD population actually shows a slight on-average decrease (not increase) in thickness. This indicates that a normative age-modelling indeed identifies different aspects of small on-average CT differences in ASD compared to canonical case-control modelling approaches (Figure 4B, 4D).

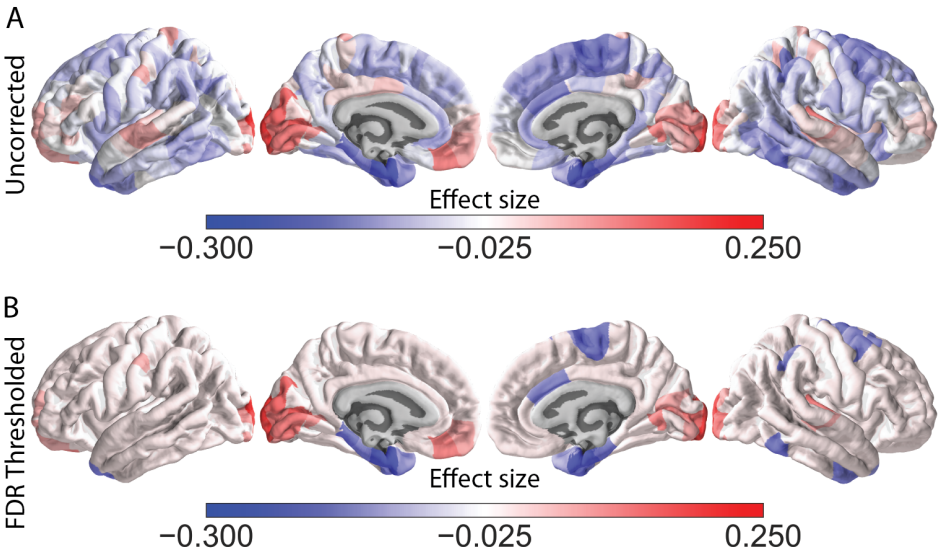

Figure 3: Effect sizes (A) and regions passing FDR correction (B) for linear mixed effect modelling of conventional case control difference analysis. Cohen's d values represent ASD – Control, thus blue denotes ASD<Control and red denotes ASD>Control. Cohen's d is computed using: https://github.com/mvlombardo/utils/blob/master/cohens_d.R

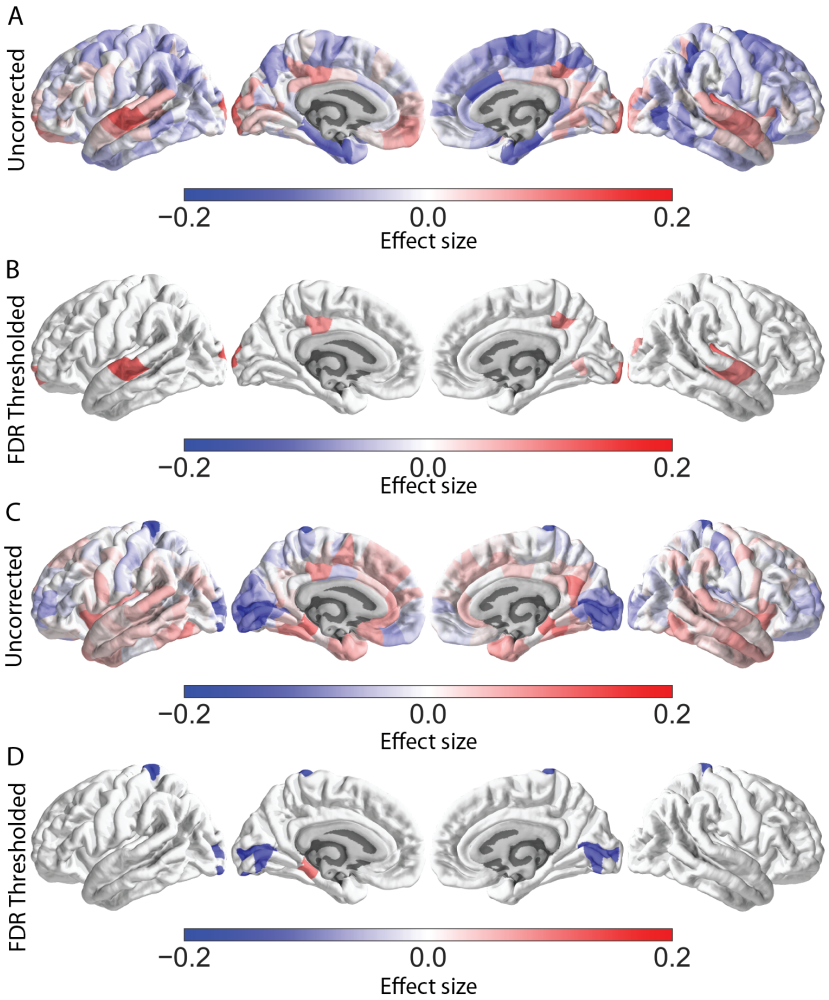

Figure 4: Effect sizes (A) and regions passing FDR correction after outlier removal (B) for linear mixed effect modelling of conventional case control difference analysis. Cohen's d values represent ASD – Control, thus blue denotes ASD<Control and red denotes ASD>Control. Effect sizes based on mean/standard deviation for w-scores (C) and regions passing FDR correction after outlier removal (D). In lower two panels blue denotes on average thinner regions and red denotes on average thicker cortex.

Table 2: *Main effect of diagnosis on LME after outlier removal*

Region	p	p_{FDR}	Cohens d
lh_lateraloccipital_part3	*2.38E-04*	*1.12E-02*	*0.14*
lh_posteriorcingulate_part1	*1.45E-04*	*1.11E-02*	*0.09*
lh_rostralmiddlefrontal_part4	*9.89E-04*	*3.38E-02*	*0.1*
lh_superiortemporal_part5	*1.23E-04*	*1.11E-02*	*0.13*
lh_superiortemporal_part6	*2.31E-04*	*1.12E-02*	*0.14*
rh_lateraloccipital_part1	*1.18E-03*	*3.65E-02*	*0.12*
rh_lateraloccipital_part4	*1.45E-03*	*4.07E-02*	*0.08*
rh_lingual_part5	*1.78E-03*	*4.56E-02*	*0.06*
rh_precuneus_part7	*1.16E-04*	*1.11E-02*	*0.1*
rh_superiortemporal_part4	*6.83E-04*	*2.63E-02*	*0.1*
rh_superiortemporal_part6	*2.54E-04*	*1.12E-02*	*0.1*
rh_transversetemporal_part1	*1.07E-04*	*1.11E-02*	*0.13*

Table 3: *Main effect of diagnosis on w-score after outlier removal*

Region	p	p_{FDR}	Cohens d
lh_lateraloccipital_part7	*4.78E-04*	*1.34E-02*	*-0.13*
lh_lateraloccipital_part8	*1.20E-04*	*7.42E-03*	*-0.14*
lh_lingual_part4	*1.55E-04*	*7.98E-03*	*-0.14*
lh_parahippocampal_part2	*1.44E-03*	*3.70E-02*	*0.12*
lh_pericalcarine_part1	*8.97E-06*	*1.38E-03*	*-0.17*
lh_pericalcarine_part2	*2.93E-04*	*1.00E-02*	*-0.13*
lh_postcentral_part1	*2.87E-06*	*8.83E-04*	*-0.18*
rh_lingual_part4	*3.55E-04*	*1.09E-02*	*-0.13*
rh_pericalcarine_part1	*8.58E-05*	*6.61E-03*	*-0.15*
rh_pericalcarine_part2	*2.44E-04*	*1.00E-02*	*-0.14*
rh_pericalcarine_part3	*2.80E-04*	*1.00E-02*	*-0.13*
rh_postcentral_part1	*1.61E-05*	*1.65E-03*	*-0.16*

Isolating ASD individuals with significant age-related CT deviance

While the normative modelling approach can be sensitive to different pathology than traditional case-control models, another strength of the approach is the ability to isolate individuals expressing highly significant CT-deviance. We operationalized 'clinically significant' deviance in statistical terms as w-scores greater than 2SD away from TD norms. By applying this cut-off, we can then describe what proportion of the ASD population falls into this CT subtype category. Over all brain regions the median prevalence for these patients is around 7.6%. This prevalence estimate is much higher than the expected 4.55% prevalence one would expect for greater than 2 standard deviations of difference. The distribution of prevalence across brain regions also has a positive tail indicating that for a small number of brain regions the prevalence can jump up to more than 10%. Expressed back into sample size numbers, if 10% of the ASD population had clinically significant CT abnormalities, with a sample size of n=942, this means that n=94 patients possess such significant issues. Underscoring the prevalence of these significant cases is important since as shown earlier, it is likely that primarily these 'statistical outlier' patients drive most of the tiny case-control differences observed.

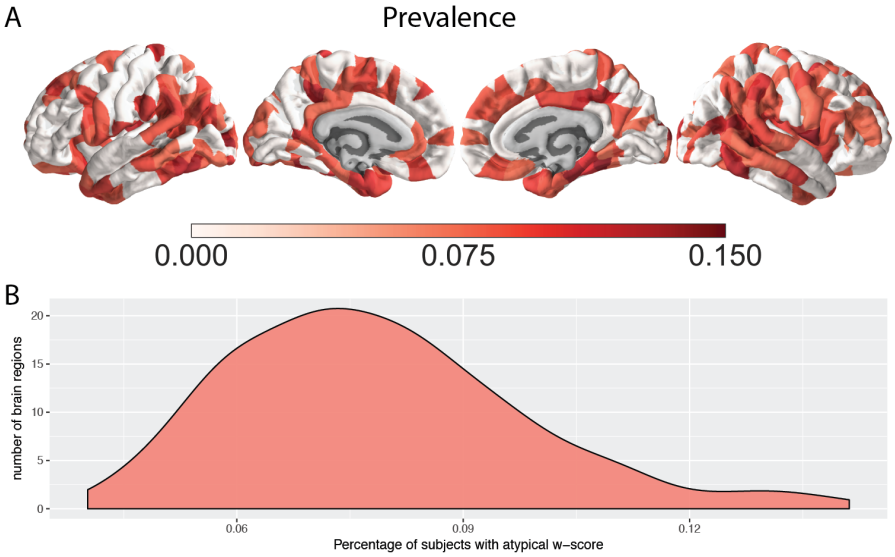

Figure 5: Region specific prevalence of atypical w-scores. Panel A shows the by region prevalence of individuals with a w-score of greater than +/-2. For visualization purposed these images are thresholded at the median prevalence of 0.076. Panel B shows the overall distribution of prevalence across all brain regions.

There are other interesting attributes about this subset patients. With regard to age, these patients were almost always in the age range of 6-20, and were much less prevalent beyond age 20. The median age of outliers across brain regions ranged from [10.6 – 20.2] years old, with an overall skewed distribution towards the younger end of the spectrum (supplementary Figure S1 and S3), showing that CT deviance potentially normalizes with increasing age in ASD. Patients with clinically significant CT deviance were also largely those that expressed such deviance within specific brain regions and were not primarily subjects with globally deviant CT. To show this we computed a w-score ratio across brain regions that helps us isolate patients that show globally atypical CT deviance across most brain regions. The small number of patients with a ratio indicating a global difference (n=18) were those that had globally thinner cortices. Upon visual inspection of the raw data for these participants, it is clear that the global thinning effect is not likely a true biological difference but rather one likely driven by artefact in the raw images (Supplementary Figure S2).

Associations between individualized metrics of age-related CT deviance and behavioural measures of symptomatology

An additional advantage of the use of normative modelling over the traditional case-control modelling is that we can use the individualized deviation as a novel metric for finding associations with phenotypic features. Here we use w-scores to compute Spearman correlations for the most commonly shared phenotypic features in the ABIDE dataset: ADOS, SRS, SCQ, AQ, FIQ and Age. After correcting for multiple comparisons across phenotype and region (6 phenotypic measures * 308 regions = 1848 tests) we identified a

number of brain regions that survive multiple comparison corrections for the SRS and ADOS scores (Figure 5 and supplementary figure S4). SRS is associated with w-scores primarily in areas of lateral frontal and parietal cortex, while ADOS is associated with w-scores primarily in lateral and inferior temporal cortex. Notably, these regions are largely different from regions that appear to show on-average differentiation in case-control and w-score analyses. Given that scanning site explains a large proportion of overall variance we also re-ran our correlation analysis after regressing out scanner site from the w-scores (and adding the overall mean for ease of interpretation). Parcellation did not affect these results (see supplementary materials). As outlined before, in order to show that results were independent of the parcellation scheme used we performed the same analysis using the HCP parcellation [31]. Using a different parcellation scheme did not affect the regional correlations for the SRS score as the overall pattern largely overlapped (supplemental figure S5). For the ADOS correlations the patterns seemed less robust across parcellation schemes.

Interestingly, we also performed the same analysis after regressing out scanning site from the raw CT values (and computing w-scores after this regression). Perhaps somewhat unsurprisingly this approach removed all effects. This is most likely due to the fact that scanning site is known to interact with numerous other variables such as age (most scanning sites contributed data from a specific age range) and diagnosis (the case-control ratio varies substantially across scanning sites and in some cases regressing out site might thus also remove diagnostic signal).

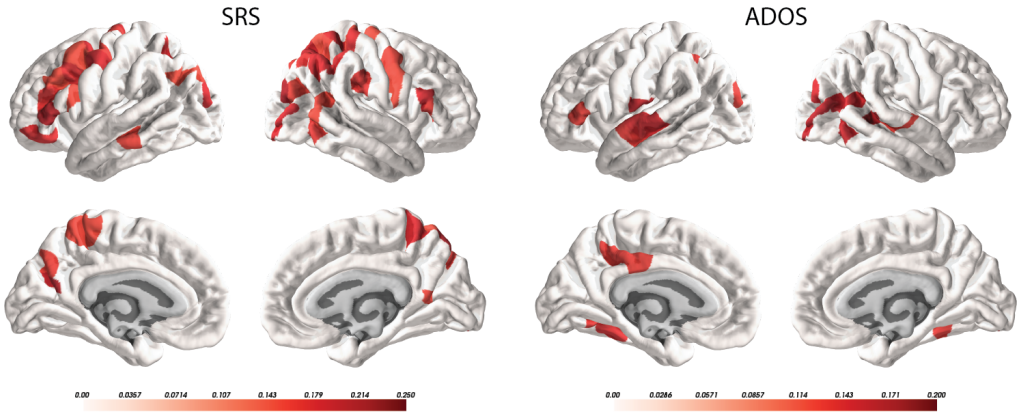

Figure 5: FDR Corrected Symptom correlations between w-scores and SRS and ADOS respectively.

Clustering

Despite the limited main diagnosis effect on CT over the majority of brain regions and the fact that only a small subset of individuals appears to contribute to this difference, it may still be possible that the multivariate patterning in CT may capture some diagnostic effect. Thus,

we performed exploratory clustering analysis to determine if raw CT values across the whole brain could be used to delineate the ASD group from the TD group. In addition, we reasoned a data-driven clustering approach might also reveal subgroups within each group (e.g. Lombardo et al., 2016). Results from clustering the neighbour embedded raw CT scores are shown in Figure 6. As can be observed in panel B, the within-group heterogeneity is entirely captured by normative heterogeneity and the overall density plots for both groups are close to identical. The pattern we find when projecting the whole brain raw cortical thickness into a 2-dimensional embedding most closely resembles the 3rd scenario outlined by Marquand and colleagues [13], whereby disease related variation is nested with the normal variation. Our results show that, when it comes to whole-brain cortical thickness, the condition related variation is entirely nested within the neurotypical variation. Obviously, the present clustering and embedding approaches only provide one way of clustering or segregating case-control variation in cortical thickness. Other multivariate approaches that do not take the same data reduction step might be able to more finely parse out sub-groups or potential case-control differences. Additionally, other measures than CT might provide a different picture. In is interesting however to note that both dimensions were correlated with age. Dimension one showed a significant negative correlation: R^2 = -0.26, p = 2.2e-16. Whereas dimension 2 showed a stronger positive correlation: R^2 = 0.43, p = 2.2e-16. No correlations were observed with any of the other common phenotypic measures. Thus, this 2-dimensional embedding likely captures the variability in cortical thickness expansion and thinning over the course of development, but is not sensitive enough to pick up potential alterations in the overall trajectory of this process between groups.

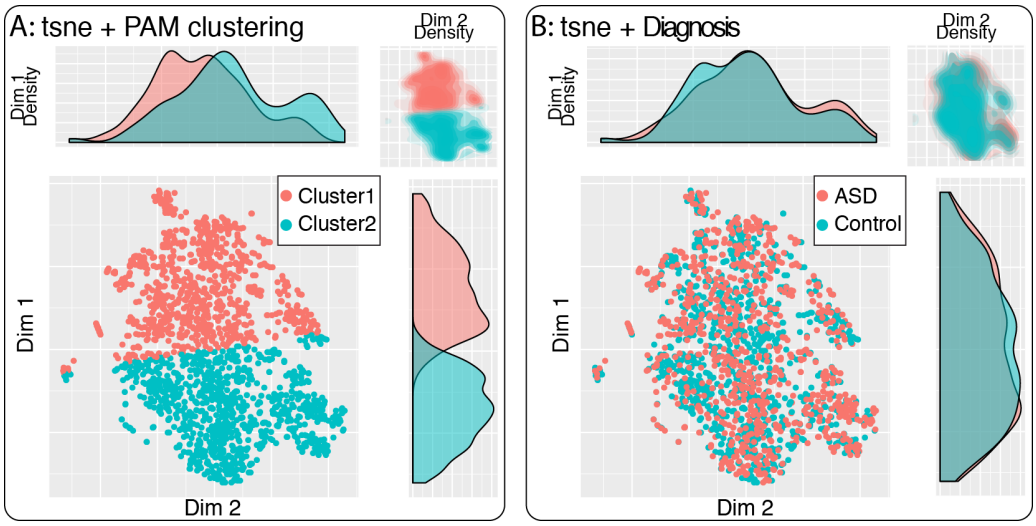

Figure 7: tSNE Clustering
Panel A shows the results from k-medoid clustering of the 2D embedding of the raw CT values as achieved by tSNE. K-medoid clustering clearly identifies 2 clusters. However, as Panel B shows, these clusters did not provide any meaningful distinction on diagnosis.

Discussion

In the present study, we find that with a highly-powered dataset, conventional analyses reveal small case-control differences in cortical thickness in autism restricted to a small subset of regions. In general, this idea about subtle effect sizes for case-control comparisons is compatible with other recent papers utilizing partially overlapping data - Haar and colleagues utilized only ABIDE I data [5], while van Rooji and colleagues [19] utilized both ABIDE I and II dataset combined with further data from the ENIGMA consortium. While these statements about small effect sizes are not novel, we contribute a novel idea here in our findings that suggest that even these small effect sizes may be misleading and over-optimistic. Utilizing normative modelling as a way of identifying and removing CT-deviant outlier patients, we find here that most small case-control differences are driven by a small subtype of patients with highly CT-deviance for their age. In contrast, we further showed that analysis of CT-normed scores (i.e. w-scores) themselves reveals a completely different set of regions that are on-average atypical in ASD. The directionality of such differences also reverses in some cases. For instance, Haar and colleagues discovered that areas of visual cortex are thicker in ASD compared to TD in ABIDE I [5]. Our case-control analyses here largely mirror that finding. However, analysis of w-scores and having removed outliers, indicates a significant difference below 0, where 0 indicates no difference from age-norms. This difference below 0 is indicative of the ASD group showing on-average thinner visual cortices. This effect is apparent after the removal of highly CT-deviant ASD outliers and thus, is representative of a slight on-average difference in the majority of ASD patients. Thus, here is a clear case whereby our novel normative age modelling approach identifies effects that are completely obscured and reversed compared to what a traditional case-control analysis might reveal.

The revelation of new insights via normative age modelling, alongside cleaning up interpretations behind case-control models, both highlight the significant utility of such a novel approach. Because the presence of a small subtype in ASD shows highly age-deviant CT and that this subtype misleadingly drives on-average inferences from case-control models, it is important for the field to better understand how prevalent is this subtype and which brain regions this subtype might affect. With our normative modelling approach, we were able to quantify the prevalence of this CT outlier subtype at a median prevalence of 7.6% when taking into account all brain regions – an estimate much larger than the expected 4.55% for standard deviations greater than 2. However, there is some heterogeneity across brain regions, as a small proportion of brain regions are even more enriched in this subtype and can contain greater than 10% prevalence.

We also noted that this small subtype showing highly age-atypical CT was predominantly restricted to the childhood to early adult age range. In later adult ages, the prevalence of this subtype drops off. This could be a potential indicator that highly atypical CT is more prevalent and detectable at earlier ages. It will be important to assess even earlier age ranges such as the first years of life [21], as well as later adult years when significant aging processes begin to take effect [41]. Importantly, the interpretations behind why this subtype of patients is so atypical also needs to be addressed. Mirroring work in autism genetics, whereby discoveries are continually being made regarding very small proportions of the ASD population being explained by highly penetrant genetic mechanisms [42], it also may be the case that such individuals with highly age-deviant CT are individuals with specific highly penetrant biological mechanisms underlying them, and possibly related to neurogenesis and other factors that are implicated in CT changes (Romero-Garcia et al., In Press). With animal models of highly penetrant genetic mechanisms linked to autism, it is notable that such mechanisms have heterogeneous effects on brain volume [43]. Thus, it will be important for future work to parse apart explanations behind why such a small subset of individuals appear to have such highly age-deviant CT features.

Normative modelling was also applied towards identifying continuous associations between the degree of age-related CT deviance and symptom severity as measured by the SRS and ADOS. Here we find collections of areas that are largely different from regions detected as showing on-average case-control or on-average non-zero w-score differences. These results likely suggest that w-score can pick up on some aspect of dimensional variability related to behavioural symptoms. However, as was the case for the case-control analyses, the effect sizes for these associations with behavioural symptoms were very small. In addition to common univariate analysis approaches, we also explored novel multivariate approaches for assessing differentiation in CT patterning. However, these analyses largely reinforced the notion that at least within this large dataset, autism related variation of cortical thickness is in fact largely nested within the normal variation, and that a good explanatory predictor of such variation is age.

Finally, there are a number of caveats to consider in the present study. First and foremost, the present data are cross-sectional and the normative age modelling approach cannot make claims about trajectories at an individual level. With longitudinal data, this normative modelling approach could be extended. However, at the moment the classifications of highly age-deviant CT individuals are limited to static normative statistics within discrete age-bins rather than based on statistics from robust normative trajectories. The dataset also

represents ASD within an age range that misses very early developmental and also very late adulthood periods. The dataset also presents a post-hoc collection of sites accumulated through the ABIDE initiative, whereby scanners, imaging acquisition sequences and parameters, sample ascertainment, etc, are highly heterogeneous. As a result, we observed that site had a large effect on explaining variance in CT and this is compatible with observations made by other studies [5]. Furthermore, it is likely that there may systematic interaction between scanner site and some variables of interest such as age (e.g. different scanning sites will likely have recruited specific age cohorts). Future work with more homogenous imaging sequences, scanner hardware, etc, that bolster multi-site combination of data is warranted in order to reduce this pronounced site related issue.

In conclusion, the present study showcases a novel normative age modelling approach in ASD and one that can significantly impact the interpretation of conventional case-control modelling, but which can also shed significant new insight into heterogeneity in ASD. We show that results from case-control analyses, even within large datasets, can be highly susceptible to the influence of 'outlier' subjects. Removing these outlier subjects from analyses can considerably clean up the inferences being made about on-average differences that apply to a majority of the ASD population. Rather than only being nuisances for standard group-level analyses, these outlier patients are significant in their own light, and can be identified with our normative age modelling approach. Normative models may provide an alternative to case-control models that test hypotheses at a group-level, by allowing additional insight to be made at more individualized levels, and thus help further progress towards precision medicine for ASD.

Acknowledgments

This work was supported by a European Research Council (ERC) Starting Grant (755816; AUTISMS) awarded to MVL. RRG was funded by the Guarantors of Brain. JS was funded by the National Institutes of Health Oxford-Cambridge Scholars Program. RAIB was funded by the Medical Research Council and Autism Research Trust.

Author Contributions

RAIB, RRG, JS and MVL designed the experiment. RAIB, RRG, JS and MVL conceived and implemented all analyses. RAIB, RRG, JS, AR and MVL wrote the manuscript.

Financial Disclosures

None of the authors have any financial interests to declare.

References

1. Lai M-C, Lombardo M V., Baron-Cohen S. Autism. *Lancet*. 2014;383(9920):896-910. doi:10.1016/S0140-6736(13)61539-1.
2. Lai M-C, Lombardo M V., Chakrabarti B, Baron-Cohen S. Subgrouping the Autism "Spectrum": Reflections on DSM-5. Nestler E, ed. *PLoS Biol*. 2013;11(4):e1001544. doi:10.1371/journal.pbio.1001544.
3. Ecker C. The neuroanatomy of autism spectrum disorder: An overview of structural neuroimaging findings and their translatability to the clinical setting. *Autism*. 2016:1362361315627136-. doi:10.1177/1362361315627136.
4. Vissers ME, Cohen MX, Geurts HM. Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neurosci Biobehav Rev*. 2012;36(1):604-625. doi:10.1016/j.neubiorev.2011.09.003.
5. Haar S, Berman S, Behrmann M, Dinstein I. Anatomical Abnormalities in Autism? *Cereb Cortex*. 2016;26(4):1440-1452. doi:10.1093/cercor/bhu242.
6. Di Martino A, Yan C, Li Q, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry*. 2014;19(6):659-667. doi:10.1038/mp.2013.78.
7. Lord C, Bishop S, Anderson D. Developmental trajectories as autism phenotypes. *Am J Med Genet Part C Semin Med Genet*. 2015;169(2):198-208. doi:10.1002/ajmg.c.31440.
8. Georgiades S, Bishop SL, Frazier T. Editorial Perspective: Longitudinal research in autism - introducing the concept of "chronogeneity." *J Child Psychol Psychiatry*. 2017;58(5):634-636. doi:10.1111/jcpp.12690.
9. Raznahan A, Lerch JP, Lee N, et al. Patterns of coordinated anatomical change in human cortical development: a longitudinal neuroimaging study of maturational coupling. *Neuron*. 2011;72(5):873-884. doi:10.1016/j.neuron.2011.09.028.
10. Raznahan A, Shaw P, Lalonde F, et al. How Does Your Cortex Grow? *J Neurosci*. 2011;31(19):7174-7177. doi:10.1523/JNEUROSCI.0054-11.2011.
11. Mills KL, Goddings A-L, Herting MM, et al. Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *Neuroimage*. 2016;141:273-281. doi:10.1016/j.neuroimage.2016.07.044.
12. Smith E, Thurm A, Greenstein D, et al. Cortical thickness change in autism during early childhood. *Hum Brain Mapp*. 2016;37(7):2616-2629. doi:10.1002/hbm.23195.
13. Marquand AF, Rezek I, Buitelaar JK, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry*. 2016;80(7):552-561. doi:10.1016/j.biopsych.2015.12.023.
14. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17(12):1174-1179.

doi:10.1038/mp.2012.105.

15. 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. 2017;(February):1721-1731. doi:10.1093/cercor/bhx038.
16. Moradi E, Khundrakpam B, Lewis JD, Evans AC. NeuroImage Predicting symptom severity in autism spectrum disorder based on cortical thickness measures in agglomerative data. *Neuroimage*. 2017;144(February 2016):128-141. doi:10.1016/j.neuroimage.2016.09.049.
17. Zielinski B a., Prigge MBDD, Nielsen JA, et al. Longitudinal changes in cortical thickness in autism and typical development. *Brain*. 2014;137(6):1799-1812. doi:10.1093/brain/awu083.
18. Jiao Y, Chen R, Ke X, Chu K, Lu Z, Herskovits E. Predictive models of autism spectrum disorder based on brain regional cortical thickness. *Neuroimage*. 2011;50(2):589-599. doi:10.1016/j.neuroimage.2009.12.047.Predictive.
19. van Rooij D, Anagnostou E, Arango C, 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*. 2017;(7):appi.ajp.2017.1. doi:10.1176/appi.ajp.2017.17010100.
20. Romero-Garcia R, Warrier V, Bullmore ET, Baron-Cohen S, Bethlehem RAI. Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Mol Psychiatry*. doi:10.1101/208223.
21. Courchesne E, Campbell K, Solso S. Brain growth across the life span in autism: age-specific changes in anatomical pathology. *Brain Res*. 2011;1380:138-145. doi:10.1016/j.brainres.2010.09.101.
22. Courchesne E, Pierce K, Schumann CM, et al. Mapping early brain development in autism. *Neuron*. 2007;56(2):399-413. doi:10.1016/j.neuron.2007.10.016.
23. Schumann CM, Bloss CS, Barnes CC, et al. Longitudinal MRI Study of Cortical Development through Early Childhood in Autism. *J Neurosci*. 2010;30(12):4419-4427. doi:10.1523/JNEUROSCI.5714-09.2010.Longitudinal.
24. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal*. 2007;15(3):199-236. doi:10.1093/pan/mpl013.
25. Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. *Nat Hum Behav*. 2017. doi:10.1038/s41562-017-0189-z.
26. Ségonne F, Dale AM, Busa E, et al. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*. 2004;22(3):1060-1075. doi:10.1016/j.neuroimage.2004.03.032.
27. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. 1999;9(2):179-194. doi:10.1006/nimg.1998.0395.
28. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci*. 2000;97(20):11050-11055. doi:10.1073/pnas.200033797.

29. Whitaker KJ, Vértes PE, Romero-Garcia R, et al. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc Natl Acad Sci*. 2016;113(32):9105-9110. doi:10.1073/pnas.1601745113.
30. Bethlehem RAI, Romero-Garcia R, Mak E, Bullmore ET, Baron-Cohen S. Structural Covariance Networks in Children with Autism or ADHD. *Cereb Cortex*. 2017;27(8):4267-4276. doi:10.1093/cercor/bhx135.
31. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016;536(7615):171-178. doi:10.1038/nature18933.
32. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31:968-980. doi:10.1016/j.neuroimage.2006.01.021.
33. Romero-garcia R, Atienza M, Clemmensen LH, Cantero JL. Effects of network resolution on topological properties of human neocortex. *Neuroimage*. 2012;59(4):3522-3532. doi:10.1016/j.neuroimage.2011.10.086.
34. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*. 2016;17(1):483. doi:10.1186/s12859-016-1323-z.
35. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289-300.
36. Lai M-C, Lombardo M V., Auyeung B, Chakrabarti B, Baron-cohen S. Sex/Gender Differences and Autism: Setting the Scene for Future Research. *J Am Acad Child Adolesc Psychiatry*. 2015;54(1):11-24. doi:10.1016/j.jaac.2014.10.003.
37. Maaten L van der. Accelerating t-SNE using Tree-Based Algorithms. *J Mach Learn Res*. 2014;15:3221-3245.
38. Hennig C, Liao TF. How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *J R Stat Soc Ser C (Applied Stat*. 2013;62(3):309-369. doi:10.1111/j.1467-9876.2012.01066.x.
39. Byrge L, Dubois J, Tyszka JM, Adolphs R, Kennedy DP. Idiosyncratic Brain Activation Patterns Are Associated with Poor Social Comprehension in Autism. *J Neurosci*. 2015;35(14):5837-5850. doi:10.1523/JNEUROSCI.5182-14.2015.
40. Lombardo M V, Lai M-C, Auyeung B, et al. Unsupervised data-driven stratification of mentalizing heterogeneity in autism. *Sci Rep*. 2016;6:35333. doi:10.1038/srep35333.
41. Happé F, Charlton RA. Aging in Autism Spectrum Disorders: A Mini-Review. *Gerontology*. 2012;58(1):70-78. doi:10.1159/000329720.
42. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol*. 2015;14(11):1109-1120. doi:10.1016/S1474-4422(15)00044-7.
43. Ellegood J, Markx S, Lerch JP, et al. A highly specific pattern of volumetric brain changes due to 22q11.2 deletions in both mice and humans. *Mol Psychiatry*. 2014;19(1):6-6. doi:10.1038/mp.2013.179.