Cerebellar atypicalities in autism?

Charles Laidi, Dorothea L Floris, Julian Tillmann, Yannis Elandaloussi, Mariam Zabihi, Tony Charman, Thomas Wolfers, Sarah Durston, Caroline Moessnang, Flavio Dell'acqua, et al.

Cerebellar atypicalities in autism?

Charles Laidi [1,2,3,4,5], Dorothea L Floris [6,7], Julian Tillmann [8], Yannis Elandaloussi [1,2,4], Mariam Zabihi [6], Tony Charman [8], Thomas Wolfers [9], Sarah Durston [10], Caroline Moessnang [11], Flavio Dell'Acqua [12], Christine Ecker [13], Eva Loth [12] , Declan Murphy [12], Simon Baron-Cohen [14], Jan K Buitelaar [15], Andre F Marquand [6], Christian F Beckmann [6], Vincent Frouin [4], Marion Leboyer [1,2,3], Edouard Duchesnay [4], Pierrick Coupé [15,16] , Josselin Houenou [1,2,3,4]; EU-AIMS LEAP Group

1. Univ Paris Est Créteil, INSERM U955, IMRB, Translational Neuro-Psychiatry, F-94010 Créteil, France
2. AP-HP, Hôpitaux Universitaires Henri Mondor, Département Médico-Universitaire de Psychiatrie et d'Addictologie (DMU IMPACT), Fédération Hospitalo-Universitaire de Médecine de Précision en Psychiatrie (FHU ADAPT) F-94010, France
3. Fondation FondaMental, F-94010 Créteil, France
4. Neurospin neuroimaging platform CEA, Saclay
6. Donders Institute for Brain, Cognition, and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands; Department for Cognitive Neuroscience, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands.
7. Methods of Plasticity Research, Department of Psychology, University of Zurich, Zurich, Switzerland
8. Department of Psychology, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom
9. Department of Psychology, University of Oslo, Norway; Norwegian Center for Mental Disorders Research (NORMENT), Division of Mental Health and Addiction,
10. Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
11. Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany.
12. Sackler Institute for Translational Neurodevelopment, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom; Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom.
13. Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University Hospital Frankfurt am Main, Goethe University, Frankfurt, Germany; Department of Psychology, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom.
14. Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom.
Donders Institute for Brain, Cognition, and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands; 15. Department for Cognitive Neuroscience, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands; Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands.
16. Pictura Research Group, Unité Mixte de Recherche Centre National de la Recherche Scientifique (UMR 5800), Laboratoire Bordelais de Recherche en Informatique, Centre National de la Recherche Scientifique, Talence, France.

Running title: No cerebellum atypicalities in autism

Corresponding author: Charles Laidi, MD, PhD ; charles.laidi@childmind.org; 101 E 56th St, New York, NY 10022; +1 (212) 308-3118

Keywords: Cerebellum, MRI ; Volumetry ; Parcellation ; Normative Modelling ; Autism

Abstract

Background. The cerebellum contains more than 50% of the brain neurons and is involved in a broad range of cognitive functions, including social cognition. Anatomical atypicalities in the cerebellum have repeatedly been reported in individuals with autism spectrum disorder (autism). However, most studies have yielded inconsistent findings - likely because of a lack of statistical power - and did not capture the clinical and neuroanatomical diversity of autism. The aim of our study was to better understand cerebellar anatomy but also its diversity in individuals with autism. **Methods**. We studied the cerebellar grey matter morphology in a sample of 274 individuals with autism and 219 controls recruited in a multicentre European cohort (EU-AIMS LEAP study). To ensure the robustness of our results, we conducted cerebellar automated parcellation with two different pipelines and also voxel-based morphometry. We performed statistical analyses with both linear and normative models to capture the diversity of cerebellar anatomy in individuals with autism and controls. **Results.** We did not find any significant differences in the cerebellum when comparing individuals with autism and controls using linear models. In addition, there were no significant deviations in our normative models in the cerebellum, suggesting that there might be no difference in the anatomy of individuals with autism and controls. Last, the cerebellar anatomy was not associated with the symptoms of ASD. **Conclusions.** Despite positive results published in the last decade from relatively small samples, our results suggest that there is no striking difference in cerebellar anatomy of individuals with autism.

Word count:
Abstract: 249
Main Text: 3961

Introduction

The cerebellum contains more than 50% of the neurons of the brain (1), has almost 80% of the surface area of the neocortex (2) and is known to be involved in a broad range of cognitive functions including social cognition (3,4)

A large body of literature, including preclinical (5,6), histopathology, genetics and neuroimaging studies (see Fatemi et al. 2012 (7), Wang et al. 2014 (8) for a full review) has established the implication of the cerebellar circuits in social cognition and the physiopathology of autism spectrum disorder (hereafter "autism").

More than 40 prior studies reported anatomical atypicalities in the cerebellum in autism in relatively small samples. In a meta-analysis, Traut et al. (9) reported a significantly larger global cerebellar volume in individuals with autism compared to controls, though with a small effect size. Further, they did not replicate this finding in a large sample of 681 subjects from the Autism Brain Imaging Data Exchange (ABIDE) database. Because the studies included in the meta-analysis were generally underpowered, the authors found that the number of significant findings was larger than expected. To date, despite many studies (see Traut et al. 2017 (9) for a complete meta-analysis) investigating the cerebellar structure in autism, no consistent atypicalities have been found.

Several reasons may explain such discrepant findings. First, it may be that there really are no group differences in the cerebellum between individuals with autism and controls and that previous positive findings are the consequence of a publication bias from a large number of small underpowered studies (9). Second, there might be differences, but cerebellar morphological alterations in autism are subtle and located only in specific parts of the cerebellum, such as the vermis or Crus I (5,10), which was not investigated in large multicentre studies (9). Third, because autism is a heterogeneous condition (11–15), there might be distinct subgroups of individuals with autism with different pathophysiological mechanisms and different cerebellar morphological patterns that might be correlated with clinical dimensions such as sensory motor atypicalities (16,17). In that case, absent or discrepant findings might be related to the heterogeneity of subjects included in the studies, diluting consistent neural features across all subjects. Fourth, different segmentation methods could account for the variability in the results across studies - as different parcellations algorithms have been developed for the cerebellum with various outcomes. To the best of our knowledge, no study has investigated lobular cerebellar atypicalities in autism comparing different parcellation techniques. Last, there is a need to employ novel methods that can quantify individual deviations from a normative pattern without relying on group means, such as normative modelling approaches (14,16).

Our goal, therefore, was to study cerebellar anatomy of individuals with autism taking into account these methodological and clinical considerations, in a large multicentre sample. First we compared the cerebellar anatomy of individuals with autism and controls using three different standard approaches, at both lobular and voxel levels. Next, to move beyond standard case-control paradigms, we used normative modelling to quantify deviations from a normative pattern to best characterize sample heterogeneity. Finally, we studied how, within autism, cerebellar anatomy was associated with variation in clinical features.

Methods and Materials

Participants

We analyzed data from the EU-AIMS Longitudinal European Autism Project (LEAP) (18) - the largest European multicentre initiative aimed at identifying biomarkers in autism. This cohort is a large multi-centre, observational study, which aims to identify and validate stratification biomarkers in autism. We included individuals from the LEAP cohort (individuals with autism and controls recruited in six centres), with an IQ > 70.

Parcellation of cerebellar lobules

All participants were scanned with a 3T MRI scanner at 6 different sites, in (i) Cambridge, UK (Siemens Verio), (ii) London, UK (General Electrics Discovery mr750), Mannheim, Germany (Siemens TimTrio), Nijmegen, Netherlands (Siemens Skyra), Rome, Italy (GE, Signa HDxt) and Utrecht (Phillips Medical System, Achieva / Ingenia CX). High-resolution structural T1-weighted volumetric images were acquired with full head coverage, at 1.2-mm thickness with 1.2x1.2-mm in-plane resolution. Acquisition protocols are reported in supplementary material 1. A flow chart describing the quality control steps supplementary material 2), a description of the motion-related artifacts (supplementary material 3), the most frequent parcellation errors (supplementary material 4) and an example of parcellation defects related to a cyst of the posterior fossa (supplementary material 5) are reported in the supplementary material section. A comparison of the subjects included and excluded from the analyses is reported in supplementary material 6. There was no difference in the proportion of autism/TD in the excluded/included subjects, however full-scale IQ and age were lower in the participants excluded from the analyses.

We used two different methods to perform parcellation at a lobular level (Figure 1a, 1b and supplementary material 7). First, we used the CERES pipeline (25) (Figure 1b and supplementary material 7). This pipeline relies on an atlas that has been compared to manual tracing (26). Second, we used the SUIT pipeline (26) (supplementary material 8). Although the SUIT pipeline has not been compared to manual tracing, this toolbox provides a segmentation of the cerebellar vermis, a region previously linked to autism (10,27–29). We performed a visual quality check (supplementary material 9) and compared the outcome of both parcellation measures. The description of the parcellation features is reported in supplementary material 9.

Voxel based morphometry analysis

We used VBM SUIT procedure (26) to look at finer-grained differences that parcellation-based approaches might not provide. This method relies on a normalization to a probabilistic atlas of the cerebellar lobules in the anatomical space and is further described in supplementary material 8.

Phenotyping and clinical features

IQ was assessed using the Wechsler Abbreviated Scales of Intelligence. The Autism Diagnostic Observation Schedule measured clinical core symptoms of autism (19). The full clinical assessment of the LEAP cohort is described elsewhere (20).

To assess if the clinical dimensions of autism were correlated with the cerebellar morphology while limiting the number of multiple comparisons, we selected 3 clinical variables based on the literature. First, we selected the social responsiveness score (SRS) in order to see how the cerebellar structure was correlated to the severity of autism (21). Second, we selected the sensory atypicalities, based on the Short Sensory Profile scale in the EU-AIMS sample (17). Last, we selected the ADHD DSM-IV rating scale to measure severity of ADHD symptoms (supplementary material 10), a frequent comorbidity of autism. Atypicalities of the cerebellum

(notably in the vermis of the cerebellum) have been reported in ADHD (22–24). These results suggested that cerebellar atypicalities reported in autism might be at least partly related to comorbid ADHD symptoms. We performed secondary analyses to assess the effect of ADOS Calibrated Severity Score (ADOS-CCS, Shumway et al. 2012) and the Restricted Behaviour Scale Revised (RBS-R) on cerebellar structure.

Normative modelling

This novel method has been described in detail elsewhere (16) and successfully applied to mental disorders and autism in particular (12,14). Structural T1-weighted images were pre-processed with the SUIT pipeline (see voxel based morphometry section) which is specifically adapted to the cerebellum (26). A Gaussian process regression model was trained at each voxel in the control cohort using age, sex, site of inclusion, full scale IQ and intracranial volume as covariates to predict cerebellar grey matter volume. We generated normative probability maps, which quantify the deviation of each participant from the normative model for cerebellar gray matter volume at each voxel. These maps were then compared in patients and controls (see a description of our normative model in supplementary material 11).

Meta-analytical approach and homogeneity measure

To study the variability across the different sites, we repeated our analyses in each site of inclusion. We compared the volume of each cerebellar sub-region with linear models, considering age, sex, intracranial volume and full-scale IQ as covariates. The results were not corrected for multiple tests since our goal was to understand the variability across the different sites. Next, we conducted a meta-analysis across the sites of inclusion and estimated the q and the I2 statistics to study the heterogeneity of our results.

Supervised learning

We conducted classification analyses to predict which individuals belonged to the autism / TD groups based on the cerebellar anatomy. Methods are reported in supplementary material 12.

Statistical analyses

For parcellation analyses, we conducted linear models, using site of inclusion as a random effect and sex, age, full-scale IQ and intracranial volume as covariates.

To decide how to consider linear covariates (age and iq) in our model, we tested the best model fit in each site of inclusion between a linear a cubic and a quadratic model (Bedford et al. 2020). Details on the statistical analyses are reported in supplementary material 13. Before performing pairwise comparisons (t-tests) between patients and controls, we ensured that the standardized residuals were normally distributed, as assessed by the Shapiro–Wilk test ($P > 0.05$) and a QQ-plot. We conducted a Pearson correlation test to assess the correlation between volumetric measures of CERES and SUIT pipeline.

Statistical analyses of VBM is reported in supplementary material 8.

To test if clinical features (SRS score, short sensory profile and diagnosis of ADHD) were associated with the cerebellar structure, we conducted linear models only in the group of individuals with autism and regressed out the effect of age, sex, intracranial volume, full-scale IQ and MRI scanning site.

We conducted heterogeneity focused analyses to understand the influence of age, sex and iq in our statistical models. We compared two models: the first one including the variable of interest (age, sex or iq) and its interaction with diagnosis, and the second not including the variable of interest and the interaction term, as described in Bedford et al. 2020 (supplementary material 14). We studied the effect of diagnosis on cerebellar structure in IQ and age-centred intervals (supplementary material 15).

We used the pymatch package (https://github.com/benmiroglio/pymatch) to create a sample of individuals with ASD matched to the neurotypical subjects by age, iq and sex. We subsequently repeated our case control analyses to ensure that the results from the main analyses were not driven by difference between individuals with ASD and controls (see supplementary material 16). In the normative modelling analyses, given the non-parametric distribution of positive and negative deviations in each region of interest, we compared autism and TD with non-parametric Mann-Whitney tests.

Dimensional analyses in a transdiagnostic paediatric cohort

We repeated our dimensional analyses in the Healthy Brain Network cohort (Alexander et al. 2017). In a more heterogenous population of individuals with autism-related symptoms without a formal diagnosis of autism, we studied the influence of the dimensions of autism using the SRS-2 scale and two of its subscales, measuring social / communication impairments or repetitive and restrictive behaviors. A complete description of the study population and its clinical features is reported in supplementary material 17.

Results

Population of the study
Demographics of the study population are reported in Table 1. There was no significant difference in age between patients and controls; however, there were more males than females in patients compared with controls and a higher full-scale IQ in controls compared to patients. Thus, full-scale IQ and sex were included as covariates in our statistical analyses.

Comparison of CERES and SUIT pipeline.
We performed cerebellar parcellation using SUIT and CERES pipeline and compared parcellation outcomes of both pipelines (supplementary material 8). We found a strong positive relationship in Crus I ($r : 0.72$; $pval < 0.001$), the anterior lobe of the cerebellum ($r : 0.73$; $pval < 0.001$), the postero-inferior lobe of the cerebellum ($r : 0.75$, $pval < 0.001$) between SUIT and CERES except for Crus II ($r : 0.55$; $pval < 0.001$) where we found a moderate positive relationship (supplementary material 18).

Case-control analyses.

Effect of age and IQ. We found that modelling age and IQ with a linear effect was more accurate, as opposed to a cubic or quadratic effect (supplementary material 13).

Parcellation and voxel-based morphometry. We did not find any significant effect of autism diagnosis in our regions of interest using the CERES or SUIT pipeline. Effect of the diagnosis of autism (CERES pipeline) on cerebellar sub-volumes are reported in Table 2 and supplementary material 9. There was no effect of autism diagnosis in the cerebellar vermis or lobule VI-VII part of the vermis (supplementary material 20), a region previously involved in autism. When conducting the analyses at the voxel level (voxel-based morphometry - SUIT pipeline, supplementary material 8), we did not find any effect of diagnosis or any significant sex by diagnosis interaction, which was consistent with our findings using a Region-of-Interest approach.

Heterogeneity of cerebellar anatomy in ASD
We found no case control difference in our analyses, which could be explained by a variability of cerebellar anatomy in ASD related either to sex, age, IQ or the clinical features of ASD. We tested in heterogeneity focused analyses the effect of sex, age and IQ on cerebellar anatomy. Results are reported in supplementary material 14. We found no evidence for a strong heterogeneity related to these variables in our linear models. In addition, we found no significant sex by diagnosis interaction for any of the cerebellar regions. We conducted our analyses only among males or females, which did not change our results (supplementary material 14). Regarding age and IQ , there was no significant (i) age by diagnosis interaction (supplementary material 15), (ii) effect of diagnosis on cerebellar structure in age-centred intervals, (iii) age by IQ interaction or (iv) effect of diagnosis on cerebellar structure in IQ-centred intervals (analyses are reported in supplementary material 15). Repeating our analyses in a sample of ASD individuals matched with controls based on sex and age and IQ did not change our results (supplementary material 16). Within the autism sample and contrary to our hypotheses, there was no significant correlation between the SRS2 t-score (supplementary material 21) or the diagnosis of ADHD or sensory alterations and cerebellar structure. Similarly, there was no association between restricted and repetitive behaviors (supplementary material 22) or ADOS CCS score (supplementary material 23).

Last, we selected individuals from an independent transdiagnostic mental health pediatric cohort, with symptoms related to autism, as defined by SRS-2 t-score (SRS-2) > 60. Individuals were recruited in 3 sites of inclusion (supplementary material 17). We defined both a relaxed (n = 352, SRS-2 > 60) and a restricted (n = 79, SRS > 76) sample. We found no effect of (i) total SRS-2, (ii) SRS-SCI (social communication questionnaire), (iii) SRS-RRB (restrictive and repetitive behaviors) on cerebellar structure, nor SRS-2 by FSIQ and SRS-2 by age interaction in both samples. Results are described in supplementary material 17.

Multivariate statistics

Normative Modelling. Despite a good fit of our normative model (Figure 2), we found no increased or decreased deviation in the cerebellar lobules (Figure 3 and supplementary material 24), when comparing individuals with autism to neurotypical control. There was no significant difference either at the whole cerebellum level or at a lobular level. Removing FSIQ from our model did not change our results.

Support vector machine to predict diagnosis. Our model did not predict the diagnostic category (individual with autism vs TD) above the level of chance, when considering as predictive features either regions of interest extracted from the cerebellar parcellation or VBM cerebellar maps. Balanced accuracy did not exceed 52% (supplementary material 25).

Meta-analytical approach and site by site analysis

We compared cerebellar anatomy in individuals with autism and TD in each site of inclusion (see supplementary material 27 for a description of clinical features across each site of inclusion and Figure 4 for the results of the meta-analysis). Using the CERES parcellation, two regions of interest were different in two sites: compared to TD there was an increase in the volume of the lobule VI in individuals with autism recruited in Rome ($p = 0.035$, uncorrected) and an increase in the volume of Crus II ($p = 0.048$, uncorrected) in individuals with autism recruited in Cambridge. Using the SUIT parcellation, we found only a reduced volume of the vermis in individuals with autism compared to TD in Nijmegen ($p = 0.041$).

It is important to note that the significant results obtained in two sites with the CERES pipeline were not significant using the SUIT pipeline (despite a good correlation between both measures reported in supplementary material 18). For all significant results, the confidence interval (Figure 4) was large compared to the full sample analysis and very close from 0. The meta-analysis metrics (q statistics and I2, see supplementary material 27) suggested that there was no strong heterogeneity in the results across sites.

Discussion

A broad range of atypicalities in different cerebellar regions have been inconsistently reported in autism in the last decades, mostly in small sample size studies. Our goal was to study the cerebellar anatomy in a large harmonized multicentric European cohort to reconcile previous results from the literature. We combined complementary statistical (both 'traditional' group case control paradigms and individual deviations using normative modelling, supervised learning) and neuroimaging (parcellation, voxel-based morphometry adapted to the cerebellum) methods to fully understand the cerebellar anatomy in autism.
We found that regardless of the analytical technique we employed, there was no difference in the cerebellar anatomy. In addition, within autism, there was no correlation between cerebellar anatomy and clinical features. We discuss these results in the context of neuroimaging of autism and replicability / reproducibility issues in neuroimaging.

Like many neurodevelopmental and psychiatric disorders, autism is clinically heterogeneous and conceptualized as a spectrum rather than a condition. In cortical regions, there have been recent attempts to identify subgroups of individuals with autism using neuroanatomical features (37,38). However, to date, reports from MRI studies on cerebellar anatomy in autism are based on "traditional" case-control analysis, and mostly from relatively small samples. These studies typically reported cerebellar alterations in the Crus I region (5,24), in the anterior lobe (10) or in the vermis (27,28,39), which were correlated to clinical dimensions of autism. In 2017, the group of Roberto Toro (9) published a meta-analysis on 30 studies on cerebellar anatomy in autism and reported a weak but significant association between autism diagnosis and increased global (overall) cerebellar volume ($p = 0.049$, uncorrected). In addition, Traut et al. (9) studied the cerebellar volume in a larger sample (ABIDE dataset) but did not conduct a parcellation analysis and studied the global volume of the cerebellum. However, the cerebellar cortex can be divided between an anterior part - connected to the sensory motor cortex - and a posterior / cognitive part - connected to the associative cortex. Because of this functional topography, it is critical to study the anatomy of the cerebellum at a lobular level.
We did not find a difference in terms of cerebellar sub-volume in individuals with autism compared to neurotypical controls. These results were consistent across two different parcellations methods and a voxel wise analysis. All parcellations were visually inspected by an expert rater blind of the diagnosis. These results are consistent with the meta-analysis of Traut et al. (9) that reported inconclusive results at a lobular level. Thus, we believe that there is no consistent difference in cerebellar morphology when using a classic case-control approach.

The discrepancy of previous results in the literature could be explained by different methods of segmentation and parcellation that they employed (10,40,41). For example, cerebellar segmentation can be performed manually, semi-automatically and fully automatically. Because of the heterogeneity of autism, it is critical to investigate its neuroanatomy in large multicentre samples to avoid false positive results (42). Also, manual and semi-automated segmentation methods are difficult to apply to large samples and there is a need to develop fully automated segmentation algorithms. However, fully automated parcellation methods rely on different atlases (26,43). To the best of our knowledge, our study is the first to compare different parcellation algorithms in a clinical population of individuals with autism.
We found a moderate to strong positive relationship between CERES (25) and SUIT (26). It is important to note that the definition of the lobules differs between both techniques, which rely on different atlases. The CERES pipeline relies on the atlas of Park et al. (43) where the vermis is merged into the cerebellar hemisphere. Thus, the Crus II region - where we only found a

moderate correlation between both methods - encompasses part of the vermis in the CERES pipeline as compared to the SUIT pipeline where the vermis is isolated from the hemisphere. This difference of definition in the cerebellar parcellation might partly explain the discrepant findings from previous studies. In our study, we analyzed the cerebellar volumetry with both techniques in order to ensure the robustness of our results and in both cases, we did not find differences between autism and neurotypical controls. In addition, we also used the SUIT pipeline (26) to perform analyses at a voxel level.

Our study has several strengths. Most of the prior studies investigating the anatomy of individuals with autism focused on the entire brain and did not investigate the cerebellum specifically. Because of the position of the cerebellum (distinct from the neo-cortex, in the posterior fossa) and its specific anatomical structure (high degree of folding), the analysis of the cerebellum requires specific tools and parcellations algorithms. In addition, it is very difficult to perform an effective quality check in the entire brain, including the cerebellum. In this paper, an expert rater, blind to diagnosis, visually inspected all cerebellar parcellations.
To ensure the robustness of our results, we used different parcellation methods and statistical analyses to fully understand how the cerebellar structure might differ in individuals with autism and controls. We believe that to date, this is the most exhaustive study investigating the structural anatomy of the cerebellum in autism.

Several reasons could explain our negative results. One possibility might be lack of statistical power. However, all previous results on cerebellar anatomy included smaller samples (Laidi et al. 2017 (10), see Traut et al. 2017 for a meta-analysis(9)) suggesting that, if present, atypicalities could have been detected. In addition, cerebellar atypicalities have been repeatedly reported in other brain disorders such as schizophrenia in samples of the same size as this study (30,33).
An increased heterogeneity in individuals with autism could also explain our negative results in the case control analyses. In that case, only a subgroup of individuals with ASD, such as individuals with a more severe form of autism, sensory atypicalities, a specific range of IQ or age or more severe restrictive / repetitive behavior, would present cerebellar anatomical atypicalities. In that case, those might not be detected when using classic group mean comparison. To fully explore this hypothesis, we conducted a wide range of analysis to investigate the effect of sex, age, IQ, severity of autism, sensory atypicalities, diagnosis of ADHD, repetitive and restrictive behaviors. In an independent cohort of individuals with autism-related symptoms, with higher heterogeneity compared to the EU-AIMS sample, the severity of autism had no influence on cerebellar structure. Last, we also conducted a normative model analysis to investigate differences at the individual level. However, we detected no significant positive or negative deviations from the norm despite a good fit of our model. Although this approach has been successfully applied and found positive results in the cerebral cortex (12), our results were negative in the cerebellum when using a similar sample.

Our meta-analytical approach revealed marginally significant results (Figure 4). These results were not replicated when using a different parcellation method. This suggests that interpreting results in small samples is not relevant and leads to inconsistent results that are sensitive to parcellation methods. This was the case of the studies published to date on cerebellar parcellation (including a study published by our group (10)). These results explain how false positive results might arise from the literature with real-life data.

Several limits should be considered before interpreting our results. Concerns have been raised regarding the validity of psychometric properties of the Short Sensory Profile scale (Williams

et al. 2018). While our paper is focused on cerebellar volumetry using 3T MRI, this approach has limitations. The cerebellum is a highly folded structure with almost 80% of the surface area of the neocortex. Partial volume issues are thus more prominent for the cerebellum. Thus, 7T MRI (2) might be more able to detect atypicalities in cerebellar anatomy of individuals with autism.

To the best of our knowledge, this is the largest study to investigate the anatomy of the cerebellum in autism. Our results strongly suggest that there is no significant difference in cerebellar anatomy between individuals with autism and controls. In the context of replicability and reproducibility issues in science, our paper underlines the interest of using different statistical / neuroimaging methods and a large sample to address the same research question and avoid inconsistent results. Beyond structural anatomy, functional MRI studies with large sample size could unravel altered cerebellar connectivity in autism.

Acknowledgement

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 115300 for the project EU-AIMS and no. 777394 for the project AIMS-2-TRIALS. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme, the European Federation of Pharmaceutical Industries and Associates, Autism Speaks, Autistica, and SFARI.
CL is supported by the Bettencourt Schueller Foundation (CCA-INSERM-BETTENCOURT-2018)
DLF is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101025785.

Disclosure

Dr. Tillmann has served as a consultant for Hoffmann–La Roche. Dr. Baron-Cohen has served as an author, consultant, or lecturer for Ability Partner, Clarion Healthcare, Expo Medica, Eli Lilly, GLGroup, Kompetento, Medice, Prima Psychiatry, Prophase, Roche, Shire, and System Analytic; he receives royalties for textbooks and diagnostic tools from Huber/Hogrefe, Kohlhammer, and UTB. Dr. Charman has received research grant support from the Medical Research Council (UK), the National Institute for Health Research, Horizon 2020 and the Innovative Medicines Initiative (European Commission), MQ, Autistica, FP7 (European Commission), the Charles Hawkins Fund, and the Waterloo Foundation; he has served as a consultant for F. Hoffmann–La Roche and Servier; and he has received royalties from Guilford Publications and Sage Publications. Dr. Beckmann is co-founder of SBGneuro. Dr. Buitelaar has served as a consultant, advisory board member, and/or speaker for Angelini, Janssen Cilag BV, Novartis, Medice, Roche, Servier, and Takeda/Shire. Dr. Murphy has received honoraria from Roche and Servier, and he has received grant support from the Medical Research Council (UK), the National Institute for Health Research, and Horizon 2020 and the Innovative Medicines Initiative (European Commission). The other authors report no financial relationships with commercial interests.

Tables

Table 1: Study population

	ASD n = 274	Controls n = 219	Statistics	P value
Site of inclusion (C , K, M, N, R, U)	*C : 45, K : 78, M : 21, N : 77, R : 20, U : 33*	*C : 32, K : 55, M : 18, N : 57, R : 18, U : 39*	*chi-2*	*NS.*
Mean Age (SD) [min - max]	*17 (5) [7.00 - 30]*	*17 (5) [6 - 30]*	*t-test*	*NS.*
Sex ratio (% of Males)	*0.73*	*0.64*	*chi-2*	*< 0.05*
Mean full scale iq (SD) [min - max]	*103 (19) [70 - 148]*	*104 (18) [70 - 142]*	*t-test*	*< 0.05*
Mean verbal iq (SD) [min - max]	*102 (20) [70 - 160]*	*104 (18) [70 - 158]*	*t-test*	*< 0.05*
Mean performance iq (SD) [min - max]	*104 (19) [70 - 150]*	*105 (20) [70 - 147]*	*t-test*	*< 0.05*
Mean ADOS 2 (SD) [min - max]	*Social-affect: 6 (2) [1 - 10] Communication: 4 (2) [1 - 10] CSS total: 4 (3) [1 - 10]*	*NA*	*NA*	*NA*
Mean ADOS CSS (SD) [min- max]	*5.2 (3) (1-10)*	*NA*	*NA*	*NA*
Mean ADI - R (SD) [min - max]	*Social-affect: 16 (6) [1 - 28] Communication: 12 (5) [0 - 25] RRB: 4 (2) [0 - 12]*	*NA*	*NA*	*NA*
Mean SRS score (SD) [min - max]	*70 (12) [43 - 90]*	*NA*	*NA*	*NA*
Diagnosis of ADHD (yes / no)	*98 / 144*	*NA*	*NA*	*NA*
Mean RBS-R score (SD) [min - max]	*15 (13) [0 - 73]*	*NA*	*NA*	*NA*
Mean SSP score (SD) [min – max]	*130 (36) [4 – 189]*	*NA*	*NA*	*NA*

ASD : individuals with autism spectrum disorders
NS : non significant ; SRS : social responsiveness scale ; RBS : restricted and repetitive behaviors – revised; SSP : Short Sensory Profile scale, ADOS-CSS : ADOS Calibrated Severity Score ; ADI-R: Autism Diagnostic Interview - Revised ; ADOS 2 : Autism Diagnostic Observation Schedule ; Sites of inclusion : C : Cambridge, K : King's College London, M : Mannheim, N : Nijmegen ; R : Rome, U : Utrecht

Table 2: Effect of autism on cerebellar anatomy - CERES analysis

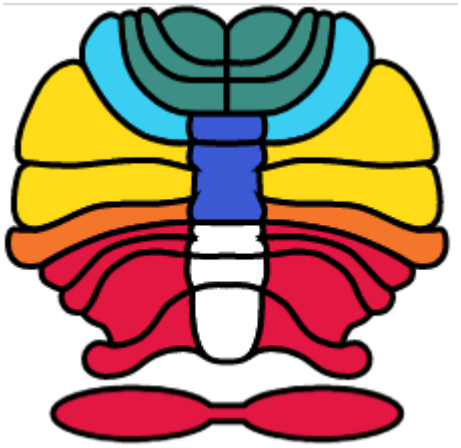
ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f^2					
					Group	Sex	Age	Site	ICV	FSIQ
Cerebellum	0.3824	-0.87	[-2.26 ; 0.87]	10	0.002	0.022	0.040	0.117	0.293	0.005
Ant. lobe	0.8042	0.25	[-0.24 ; 0.31]	10	0.000	0.024	0.015	0.179	0.184	0.005
Lobule VI	0.3276	-0.98	[-0.59 ; 0.20]	10	0.002	0.010	0.024	0.046	0.137	0.000
Crus I	0.3076	-1.02	[-0.91 ; 0.29]	10	0.002	0.007	0.005	0.046	0.118	0.004
Crus II	0.3616	-0.91	[-0.59 ; 0.21]	10	0.002	0.002	0.034	0.107	0.074	0.004
Lobule VIIb	0.3627	-0.91	[-0.32 ; 0.12]	10	0.002	0.006	0.017	0.062	0.109	0.011
Post. Inf. lobe	0.8685	0.17	[-0.51 ; 0.61]	10	0.000	0.016	0.015	0.151	0.215	0.000

Pval, P-value; tval, t-value; IC., confidence interval; dof, degrees of freedom.
(diagnosis, sex, site, age, site, ICV and IQ as covariates)

Figures

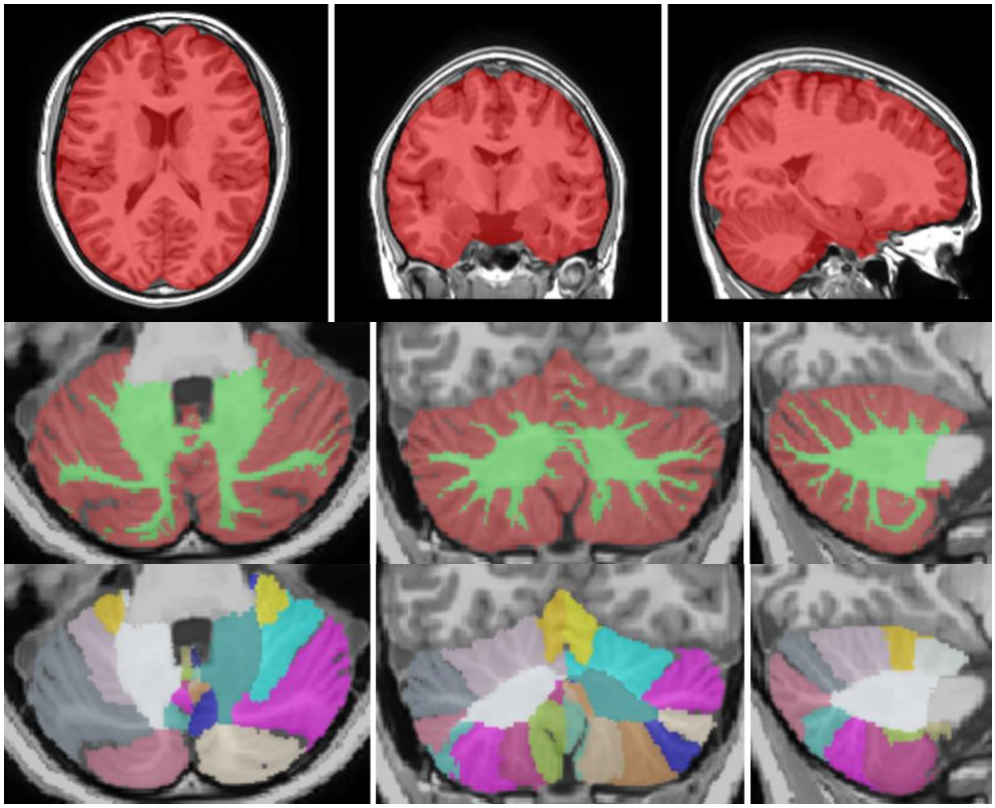
Figure 1: Cerebellar parcellation

Figure 1.a : Schematic of the cerebellar parcellation depicted in a cerebellar flat map



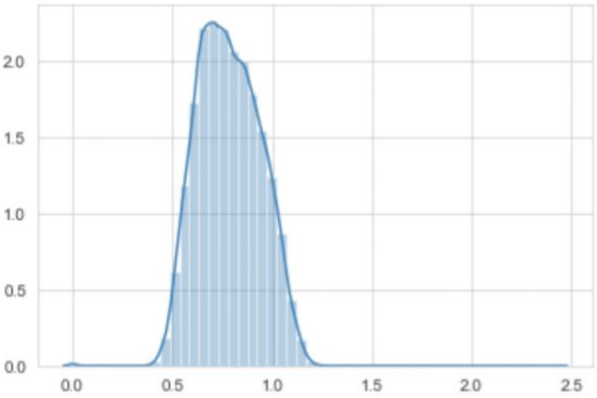
Legend : visualization of the cerebellum with cerebellum-value-map package https://gitlab.com/shan-utils/cerebellum-value-map (52)

Figure 1.b: Parcellation of the cerebellum with the CERES pipeline

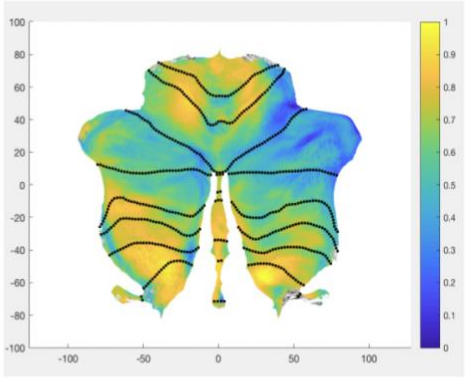


Legend: Fig 1.a : Anterior cerebellum (Lobules I - V) = green ; Lobule VI = light blue ; Crus I and Crus II = yellow ; Lobule VIIb = orange ; Postero-inferior lobe (red) ; Vermal portion of lobules VI and VII (dark blue) ; Vermal portion of postero-inferior lobe (white). Fig 2.a : Cerebellar parcellation and intracranial volume measured with the CERES pipeline. Top panel = Intracranial volume ; Middle panel = tissue classification ; Lower panel = cerebellar parcellation

Figure 2: Goodness of fit of the normative model in controls



Histogram of SMSE neurotypical subjects



Rho map in neurotypical subjects

SMSE: symmetric mean square error, evaluating the goodness of fit of the model.
Rho map showing the correlation between the predicted and the actual value for each voxel of the cerebellum. The relationship between the covariates (age, sex, ICV, full scale IQ, and site of inclusion) and cerebellar anatomy is very strong with nearly always positive correlation and SMSE < 1.

Figure 3: No difference of positive / negative deviations in the regions of interest of the cerebellum in autism vs neurotypical populations

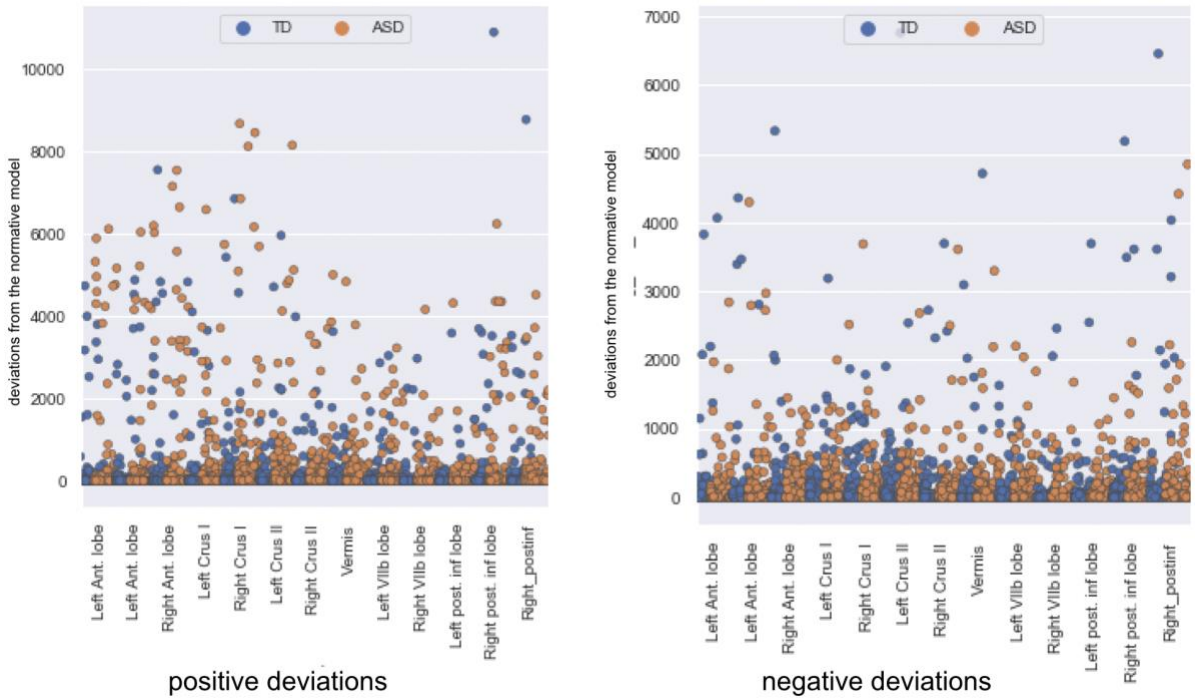
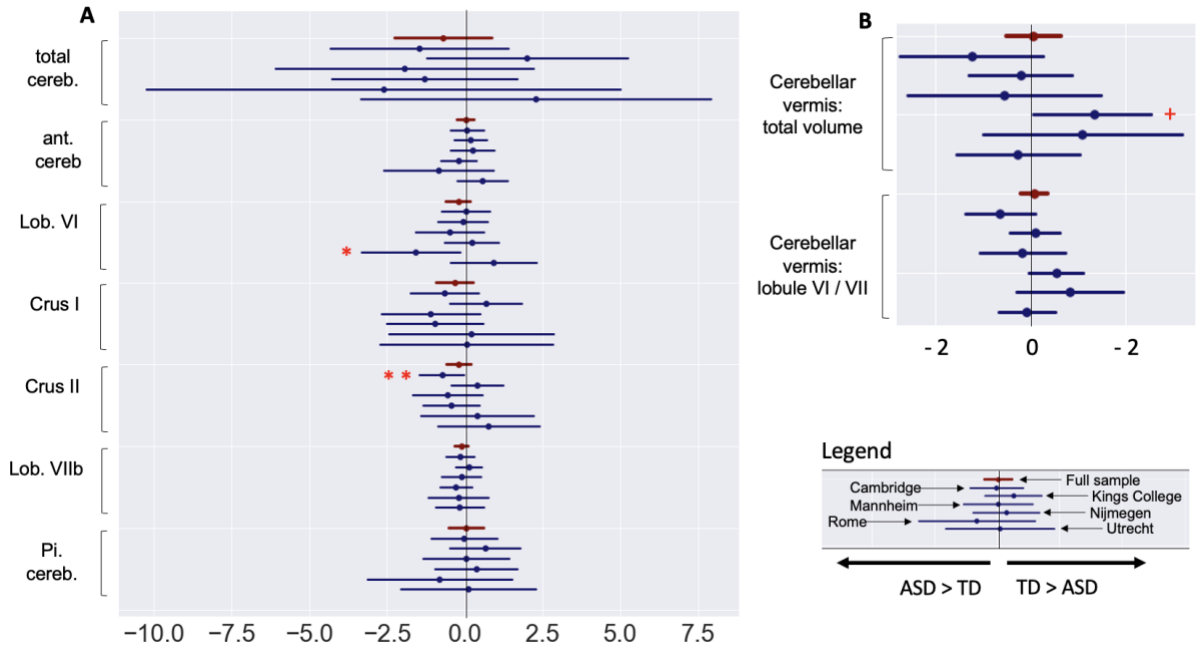


Figure 4: Cerebellar parcellation in each site of inclusion in individuals with autism and typically developing controls



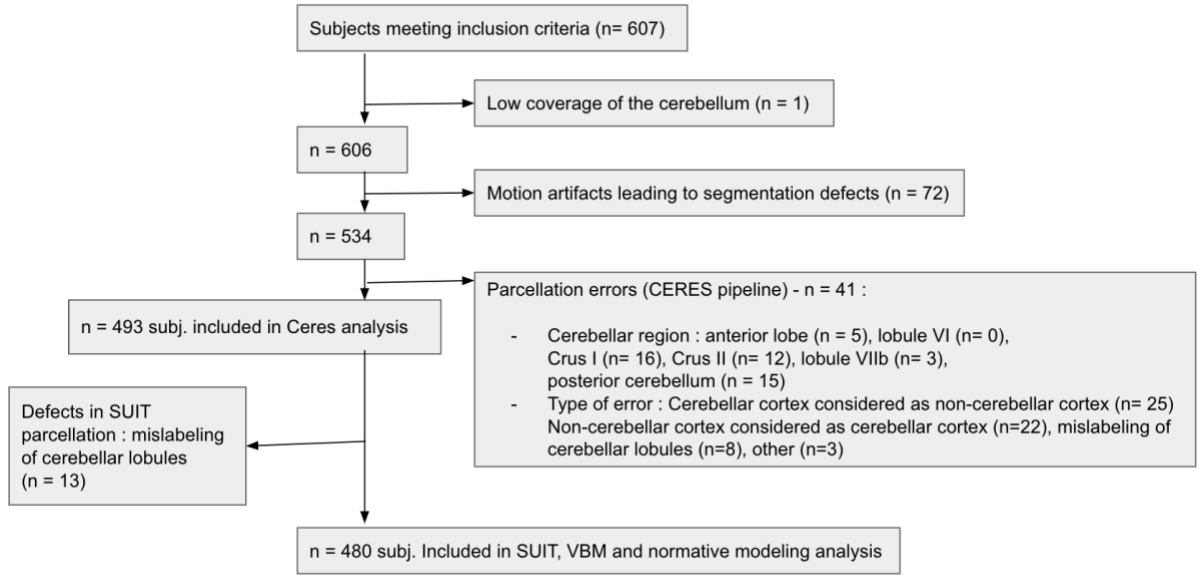
A. Effect of diagnosis (95% confidence intervals) for the total grey matter volume of the cerebellum and each subregion of the cerebellum (CERES pipeline), in the whole sample (dark red) and each site of inclusion (dark blue)
B. Effect of diagnosis (95% confidence intervals) for the cerebellar vermis and the vermal portion of lobule VI / VII (SUIT pipeline), in the whole sample (dark red) and each site of inclusion (dark blue)
ASD: individuals with Autism Spectrum Disorder ; TD: Typically developing individuals
* : uncorrected p-value = 0.035 ; ** : uncorrected p-value = 0.048 ; + : uncorrected p-value = 0.041
Y axis: regions of the cerebellum
X axis: residuals of linear models, including age, sex, intracranial volume and full-scale IQ for each site of inclusion (dark blue) and age, sex, intracranial volume, full-scale IQ and site of inclusion (dark red) for the whole sample
Pi : postero-inferior cerebellum ; Lob. : lobule ; ant. cereb : anterior lobe of the cerebellum

Supplementary Materials

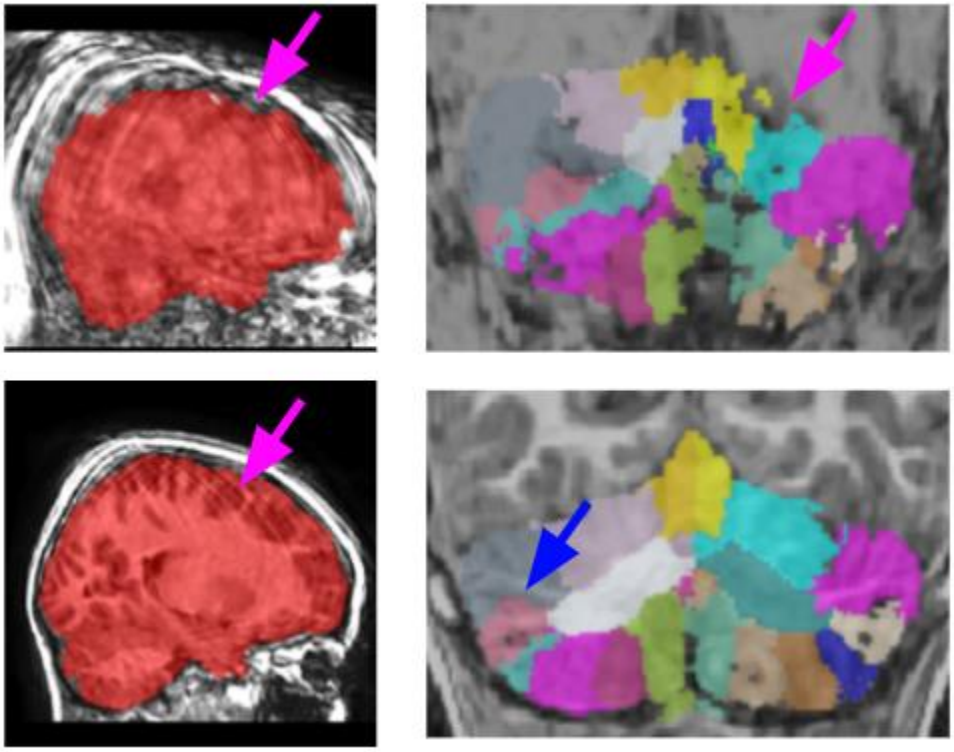
Supplementary Material 1: MR acquisition protocols

Site	Manufacturer	Model	Software Version	Acquisition sequence	Coverage	Slices	Thickness [mm]	Resolution [mm³]	TR [s]	TE [ms]	FA [°]	FOV
Cambridge	Siemens	Verio	Syngo MR B17	Tfl3d1_ns	256*256	176	1.2	1.1*1.1*1.2	2.3	2.95	9	270
London	GE Medical systems	Discovery mr750	LX MR DV23.1_V02_1317.c	SAG ADNI GO ACC SPGR	256*256	196	1.2	1.1*1.1*1.2	7.31	3.02	11	270
Mannheim	Siemens	TimTrio	Syngo MR B17	MPRAGE ADNI	256*256	176	1.2	1.1*1.1*1.2	2.3	2.93	9	270
Nijmegen	Siemens	Skyra	Syngo MRD13	Tfl3d1_16ns	256*256	176	1.2	1.1*1.1*1.2	2.3	2.93	9	270
Rome	GE Medical systems	Signa HDxt	24/LX/MR HD16.0_V02_1131.a	SAG ADNI GO ACC SPGR	256*256	172	1.2	1.1*1.1*1.2	5.96	1.76	11	270
Utrecht	Philips Medical Systems	Achieva/ Ingenia CX	3.2.3, 3.2.3.1	ADNI GO 2	256*256	170	1.2	1.1*1.1*1.2	6.76	3.1	9	270

Supplementary material 2: Flow chart

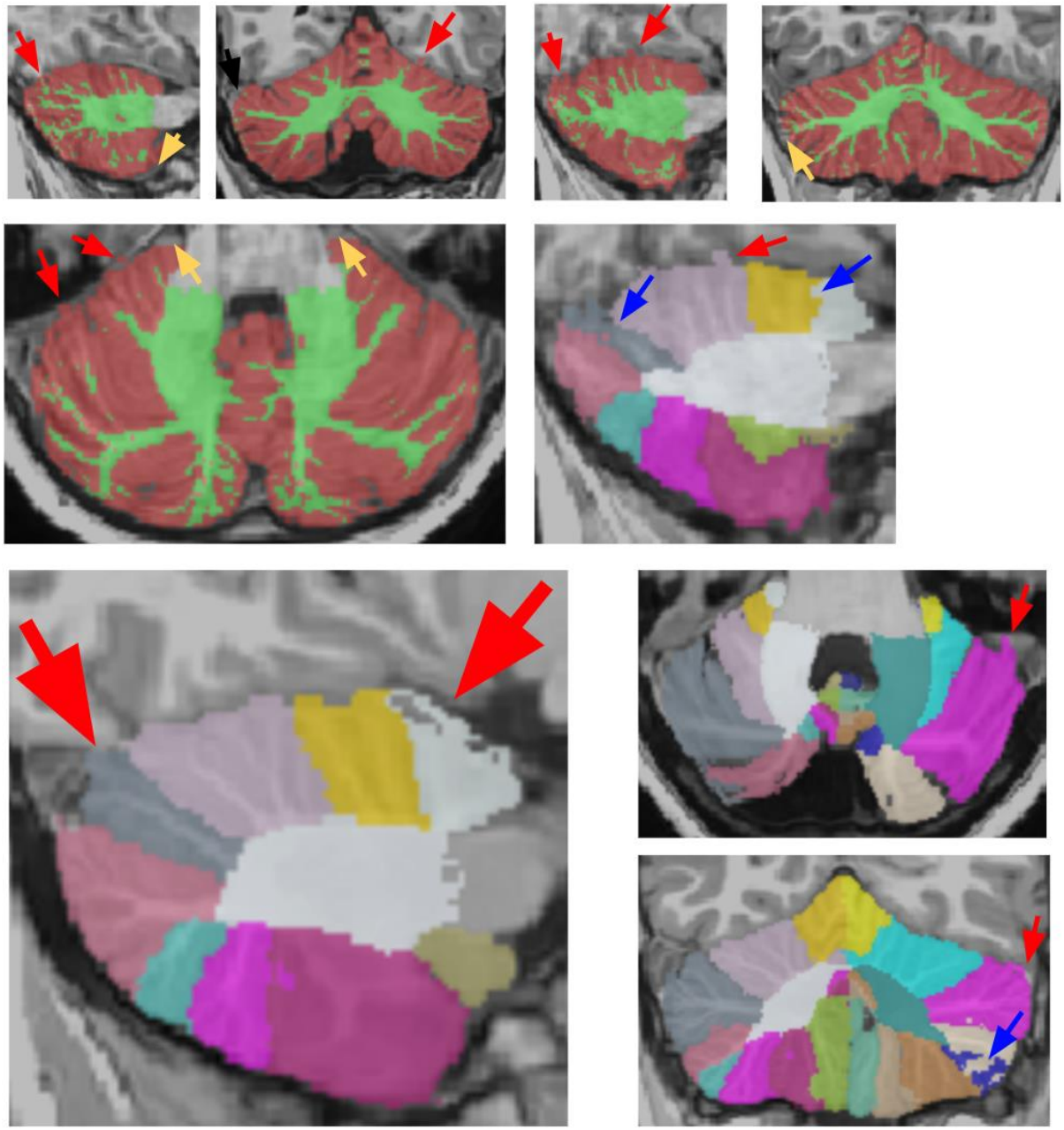


Supplementary Material 3: Motion-related artifacts

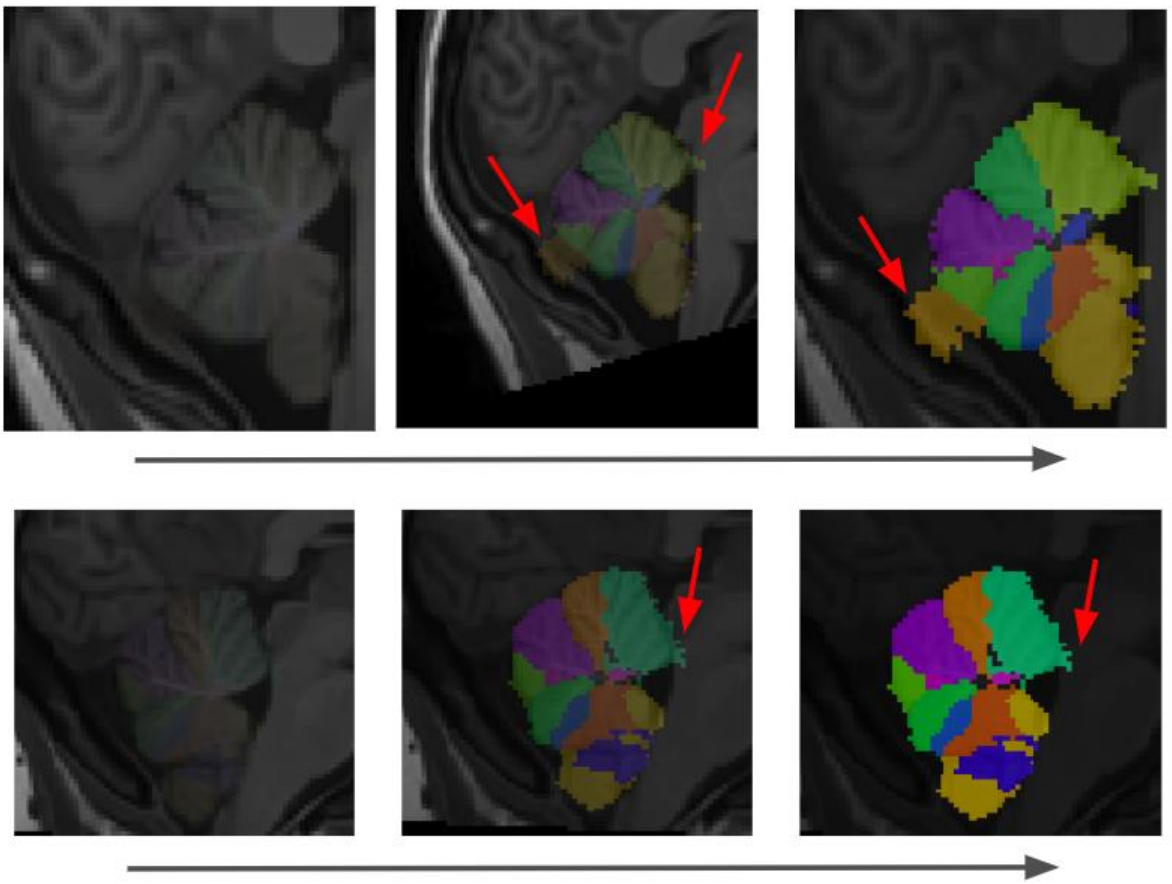


Legend :
Parcellation artifacts related to severe (top panels) or mild (bottom panels) motion.
Arrows refer to different artifacts : purple = motion artifacts ; blue = parcellation erros

Supplementary Material 4: Description of parcellation errors with the CERES pipeline

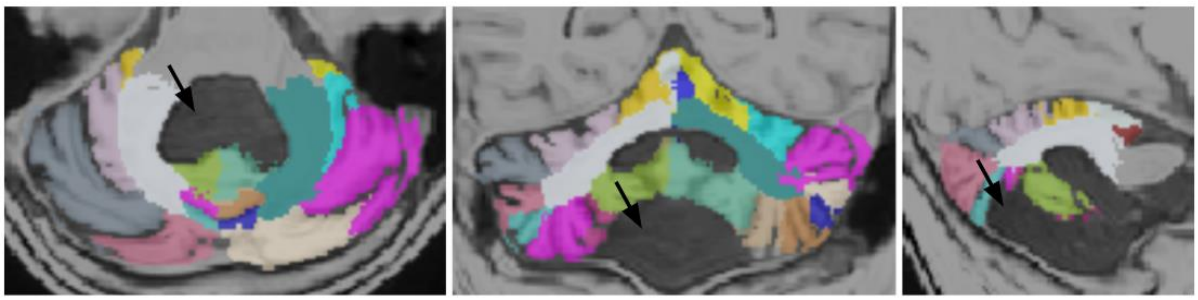


Legend:
Arrows refer to different types of artifacts :
blue = mislabeling of cerebellar lobules
red = parcellation of non cerebellar tissue
yellow = segmentation defect



Legend:
Parcellation defects (non cerebellar tissue labeled as cerebellar tissue = red arrow)
Black arrow indicates different levels of opacity of the cerebellar parcellation superimposed to the native image, from the left (low opacity) to the right (high opacity)

Supplementary Material 5: Cyst of the posterior fossa (CERES pipeline)

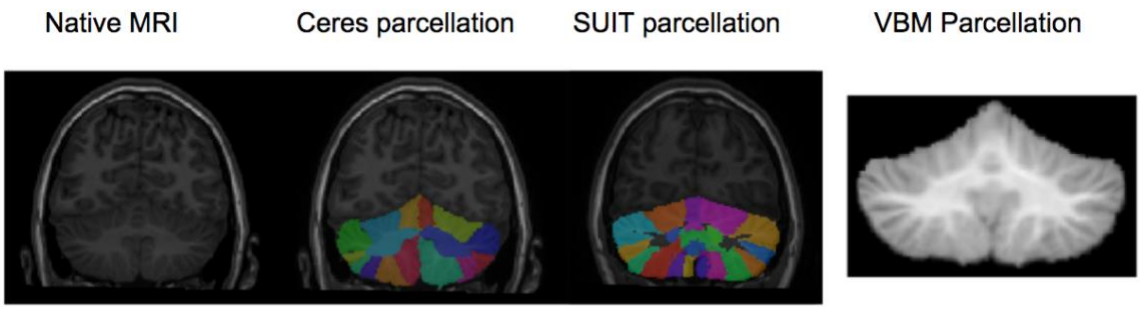


Legend :
Black arrow = cyst of the posterior fossa

Supplementary material 6: Comparison of subjects included and excluded from the analyses

	Subjects included in the analysis N = 480	Subjects excluded from the analysis N = 127	Statistics
ASD / TD (% of ASD)	269/211 (56%)	71/56 (60%)	Chi2 - pval = 0.97
M / F (% of M)	330/150 (69%)	92/35 (72%)	Chi2 - pval = 0.42
Mean age (SD)	17 (5)	14(5)	T-test - pval < 0.005
Mean IQ (SD)	105 (15)	101(15)	T-test - pval = 0.005

Supplementary Material 7: Neuroimaging methods used to study the cerebellar anatomy: CERES and SUIT cerebellar parcellation and SUIT voxel-based morphometry pipeline



CERES parcellation pipeline (25)**.** This pipeline is freely available online: https://volbrain.upv.es/index.php

Suit parcellation and VBM pipeline (27)**.** The pipeline can be downloaded online: http://www.diedrichsenlab.org/imaging/suit.htm

Supplementary Material 8: Description of the SUIT pipeline

In the SUIT pipeline, cerebellar structures are automatically isolated from the cerebral cortex based on an anatomical image. This toolbox provides an automated cortical segmentation of the cerebellum, using a probabilistic atlas to assign locations to different cerebellar lobules. We noticed defects in the SUIT pipeline at the isolation step, with non-cerebellar tissues labelled as the cerebellum. This defect has been reported in other studies (30). Thus, we created a mask based on the CERES segmentation that isolated the cerebellum more accurately and we used it in the SUIT pipeline. Using the SUIT pipeline, we performed a parcellation analysis and a voxel based morphometry analysis (VBM).

The VBM procedure used was part of the SUIT pipeline (which allows both ROI and VBM analyses). It relies on a normalization to a probabilistic atlas of the cerebellar lobules in the anatomical space. This template preserves the anatomical details of the cerebellum. The description of the template (Diedrichsen et al. 2006, 2009 *Neuroimage)* and the procedure is described on the SUIT website (https://www.diedrichsenlab.org/imaging/suit_function.htm#norm_dartel). This modulated VBM procedure relies on the Dartel engine which is based on the Jacobian determinants to compensated for volume modification (Shen et al. 2012) (https://onlinelibrary.wiley.com/doi/full/10.1002/jmri.23927).

We performed these analyses in the same sample that we previously analysed in the volumetric analyses (see above). We used a stringent nonparametric 10 000 permutations test (FSL toolbox : https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise) at a voxel level, in a search space including the total cerebellar grey matter, to assess the effect of diagnosis on cerebellar structures and limit possible false positive results. We conducted linear models, using MRI study site, sex, age, full-scale IQ and intracranial volume as covariates. Ten thousand permutations were performed for each contrast. Only voxels surviving a voxel-wize family-wise error-corrected statistical threshold of $p < 0.05$ (two tailed, permutation based) were considered significant (30).

Supplementary Material 9: Quality control procedure and parcellation outcomes

We performed cerebellar with two different methods : CERES and SUIT.
In both cases, all segmentation outcomes were visually inspected by an examiner (CL) with previous experience in cerebellar segmentation and parcellation (10,31,32) blind of the diagnosis. To ensure the consistency of parcellation measures with both methods, we studied the correlation between both pipelines (see statistical analysis section) for the anterior lobe, the postero-inferior lobe, Crus I and Crus II region.
We studied the grey matter volume of the following regions of interest using both the CERES and the SUIT pipeline: anterior lobe, lobule VI, Crus I, Crus II, lobules VIIb, Lobule VIII and lobules IX-X. We chose these regions of interest based on our previous work (10,33) ; we did not include in our analyses separately lobules I, II, III, IV and V (anterior lobe) since the volumes are too small to effectively assess visually the quality of the parcellation. In addition, we studied the vermal volume (lobule VI-VII of the vermis, total vermal volume) using the SUIT pipeline, which was not available in the CERES pipeline.

Supplementary Material 10: Assessment of ADHD diagnosis in our cohort

ADHD symptoms were assessed with the DSM-4 ADHD rating scale, covering both inattention and hyper-activity/impulsivity symptoms based on either self-or parent-report. Self-report scores were only used when parent-report scores were unavailable. A categorical variable was computed based on the DSM-4 criteria.

Supplementary Material 11: Normative modeling analysis

Normative modelling has been described in detail elsewhere (16) and successfully applied to mental disorders and autism in particular (12,14). Structural T1-weighted images were pre-processed with the SUIT pipeline (see voxel based morphometry section) which is specifically adapted to the cerebellum (26). A Gaussian process regression model was trained at each voxel in the control cohort using age, sex, site of inclusion, full scale IQ and intracranial volume as covariates to predict cerebellar grey matter volume. This resulted in a developmental model in the controls. To avoid overfitting, assess generalizability, and determine whether neurotypical individuals fall within the normative range, we used 10-fold cross-validation in neurotypical individuals before retraining the model in the entire sample to make predictions in individuals with autism. We generated normative probability maps, which quantify the deviation of each participant from the normative model for cerebellar gray matter volume at each voxel. These subject-specific *Z* score images provide a statistical estimate of how much each individual's true volume value differs from the predicted value with reference to the neurotypical pattern at each voxel given the participant's age, sex, and site. Normative probability maps were thresholded at an absolute value of $|Z| > 2.6$ (41,42). We extracted positive and negative deviations for each region of interest (defined with the SUIT atlas) and compared them using non-parametric tests (Mann-Whitney tests) given the non-normal distribution of the deviations. We first conducted the analysis in the whole cerebellum and then in every region of interest of the cerebellum, defined with the SUIT atlas. To evaluate the goodness of fit of our normative model in the neurotypical subjects, we computed a Rho map of the cerebellum showing the correlation between the predicted and the actual values in the cerebellum. In addition, we calculated the symmetric mean square error (SMSE) of our model in the neurotypical subjects.

Supplementary Material 12: Support Vector Machine Methods

Machine learning analyses were performed with support vector machine (SVM), implemented in the python library scikit-learn (https://scikit-learn.org).
We tested two different types of features in our machine learning model. First, we selected ROI-based features, based on the CERES parcellation conducted in our previous analyses (see *parcellation of cerebellar volume* section), in order to introduce anatomical priors in our model. Next, we selected grey matter VBM maps from the SUIT analysis (see *voxel based morphometry* section).
We regressed out the effect of age, sex, site and full-scale IQ that were considered as confounds in our previous analyses. Thus, the prediction was based on residuals after regression of the confounds.
We evaluated the performance with a double cross-validation (CV) scheme (see de Pierrefeu et al. 2018 (53) for more details). This design was chosen to avoid optimization of the hyperparameter that could overfit the dataset and provide an over-optimistic evaluation of the model. In the outer loop, a set of subjects is considered as the training data, while the remaining data are used as the train data, using a ten-fold cross validation. In the inner loop, the training sets are partitioned into sub-training and validation sets, using a nested five-fold cross validation to set the hyperparameter.
The classifier performances were assessed by computing the balanced accuracy, sensitivity and specificity in the test sample.

Supplementary Material 13: Comparison of linear, quadratic and cubic effect of age and IQ on cerebellar structure

We tested the best model fit between linear, quadratic and cubic effect of age across cerebellar regions. We calculated the percentage of site (as described above) at which the linear model was to best fit, compared to quadratic and cubic models across cerebellar regions. We included in our model diagnosis, each age term and its interaction with diagnosis and compared the following models across cerebellar regions:

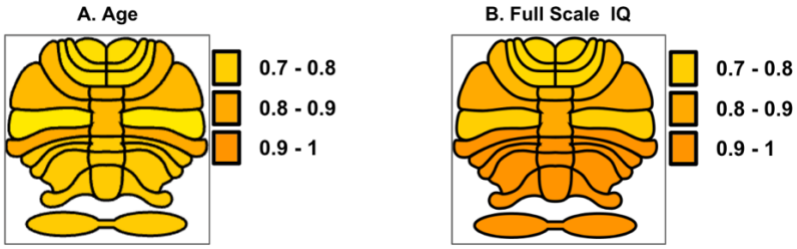
$CT_i = \beta_0 + \beta_1 Diagnosis + \beta_2 Age + \beta_3 (Diagnosis*Age) + \beta_4 Sex + \beta_5 ICV + \beta_6 \text{ full scale IQ} + \varepsilon$
$CT_i = \beta_0 + \beta_1 Diagnosis + \beta_2 Age + \beta_3 (Diagnosis*Age) + \beta_4 Age^2 + \beta_5 (Diagnosis*Age^2) + \beta_6 Sex + \beta_7 ICV + \beta_8 \text{ full scale IQ} + \varepsilon_i$
$CT_i = \beta_0 + \beta_1 Diagnosis + \beta_2 Age + \beta_3 (Diagnosis*Age) + \beta_4 Age^2 + \beta_5 (Diagnosis*Age^2) + \beta_6 Age^3 + \beta_7 (Diagnosis*Age^3) + \beta_8 Sex + \beta_9 ICV + \beta_{10} \text{ full scale IQ} + \varepsilon_i$

Results are reported in Figure A. The model including linear term was a better fit (as defined with a lower AIC) than cubic or quadratic models on average in 79% of the site of inclusion across cerebellar regions (from 70% to 92%, see Figure B1), suggesting that including age as a linear term was a good choice.

We repeated the analysis performed on age on full scale IQ. First, we tested the best model fit between linear, quadratic and cubic effect of IQ measures across cerebellar regions. We calculated the percentage of site (as described above) at which the linear model was to best fit, compared to quadratic and cubic models across cerebellar regions. We included in our model diagnosis, each iq term and its interaction with diagnosis and compared the following models across cerebellar regions :

$Cerebellar\ volumes_i = \beta_0 + \beta_1 Diagnosis + \beta_2 IQ + \beta_3 (Diagnosis*IQ) + \beta_4 Sex + \beta_5 ICV + \beta_6 Age + \varepsilon$
$Cerebellar\ volumes_i = \beta_0 + \beta_1 Diagnosis + \beta_2 IQ + \beta_3 (Diagnosis*IQ) + \beta_4 IQ^2 + \beta_5 (Diagnosis*IQ^2) + \beta_6 Sex + \beta_7 ICV + \beta_8 Age + \varepsilon_i$
$Cerebellar\ volumes_i = \beta_0 + \beta_1 Diagnosis + \beta_2 IQ + \beta_3 (Diagnosis* IQ) + \beta_4 IQ^2 + \beta_5 (Diagnosis*IQ^2) + \beta_6 IQ^3 + \beta_7 (Diagnosis* IQ^3) + \beta_8 Sex + \beta_9 ICV + \beta_{10} Age + \varepsilon_i$

For full scale IQ, the model including linear term was a better fit (as defined with a lower AIC) than cubic or quadratic models on average in 88 % of the sites of inclusion across cerebellar regions (see Figure B), suggesting that including age as a linear term was a good choice.



Supplementary Material 14: Heterogeneity focused analyses

We studied the influence of IQ, sex and age in heterogeneity focused analyses. Because IQ scores are well known to be heterogeneous in autism (with a frequent dissociation between verbal IQ and performance IQ), we studied verbal IQ (VIQ) and performance IQ (PIQ) in addition to full scale IQ (FSIQ).

We reproduced the methods of Bedford et al. 2020 to specifically study in "heterogeneous focused analyses" the importance of sex, age, FSIQ, PIQ and VIQ.
We fitted two linear models for each of those 5 variables:

- a model including the variable of interest (either sex, age, FSIQ, PIQ or VIQ) and an interaction term between this variable of interest and the diagnosis of autism
- a model not including the variable of interest or interactions in the model.

In order to compare those two models, we used the AIC (Akaike information criterion) to determine the influence of the variable in the cerebellum, separately in each site of inclusion. Following the same methods as in Bedford et al. 2019, we determine for each region of the cerebellum the number of site 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, in a region of the cerebellum taking into account all sites of inclusion.
We conducted these analyses using the results of the CERES pipeline described in our manuscript. The CERES pipeline has shown to perform well compared to other cerebellar parcellation techniques in both adult and pediatric populations. CERES was the overall winner of a cerebellar parcellation challenge (MICCAI 2017), published by Carass et al. 2018 (Carass et al. 2018 *Neuroimage*). More recently, an independent group (Soros et al). reported that CERES was an accurate and reproducible tool for fully automated segmentation and parcellation of the cerebellum Soros et al. 2021 (Soros et al. 2021 *Cerebellum*). In order to avoid increasing the risk of false positive findings, we conducted the following analyses using the outcomes of the CERES pipeline.

Heterogeneity focused analysis of sex
As described above and in Bedford et al. 2020, we compared the AIC of the two following models:

$$\text{Cereb. volume}_i = \beta_0 + \beta_1 \text{Diagnosis} + \beta_2 \text{Sex} + \beta_3 (\text{Diagnosis*Sex}) + \beta_4 \text{Age} + \beta_5 \text{Intracranial Volume} + \varepsilon_i$$
$$\text{Cereb. volume}_i = \beta_0 + \beta_1 \text{Diagnosis} + \beta_2 \text{Age} + \beta_3 \text{Intracranial Volume} + \varepsilon_i$$

Results are reported in Figure A
The model including sex term and its interaction with diagnosis was a better fit (as defined with a lower AIC) on average on 29% of sites of inclusion across cerebellar regions (from 15% to 42%, see Figure A1). Contrary to the work of Bedford et al. this result suggest a relatively low influence of sex in this heterogeneity-focused analysis.

Heterogeneity focused analysis of age
As described above and in Bedford et al., we compared the AIC of the two following models:

$$\text{Cereb. volume}_i = \beta_0 + \beta_1 \text{Diagnosis} + \beta_2 \text{Sex} + \beta_3 \text{Age} + \beta_4 (\text{Diagnosis*Age}) + \varepsilon_i$$
$$\text{Cereb. volume}_i = \beta_0 + \beta_1 \text{Diagnosis} + \beta_2 \text{Sex} + \varepsilon_i$$

Results are reported in Figure B
The model including age term and its interaction was a better fit (as defined with a lower AIC) on average in 30% of the site of inclusion across cerebellar regions (from 0 to 42%), suggesting a relatively modest influence of age in the heterogeneity focused analysis.

Bedford et al. found for instance that the model with age term and its interaction was a better fit in up to 92% of the site of inclusion across the cortex.

Heterogeneity focused analysis of full-scale IQ
As described above and in Bedford et al., we compared the AIC of the two following models:

$$\text{Cereb. volume}_i = \beta_0 + \beta_1\text{Diagnosis} + \beta_2\text{Sex} + \beta_3\text{Age} + \beta_4(\text{Diagnosis*Full scale IQ}) + \varepsilon_i$$
$$\text{Cereb. volume}_i = \beta_0 + \beta_1\text{Diagnosis} + \beta_2\text{Sex} + \varepsilon_i$$

Results are reported in Figure C
The model including full scale IQ term and its interaction was a better fit (as defined with a lower AIC) on average in 20% of the site of inclusion across cerebellar regions (from 0 to 50%, see Figure A3) suggesting a relatively modest influence of age in the heterogeneity focused analysis.

Heterogeneity focused analysis of performance IQ
As described above and in Bedford et al., we compared the AIC of the two following models:

$$\text{Cereb. volume}_i = \beta_0 + \beta_1\text{Diagnosis} + \beta_2\text{Sex} + \beta_3\text{Age} + \beta_4(\text{Diagnosis*performance IQ}) + \varepsilon_i$$
$$\text{Cereb. volume}_i = \beta_0 + \beta_1\text{Diagnosis} + \beta_2\text{Sex} + \varepsilon_i$$

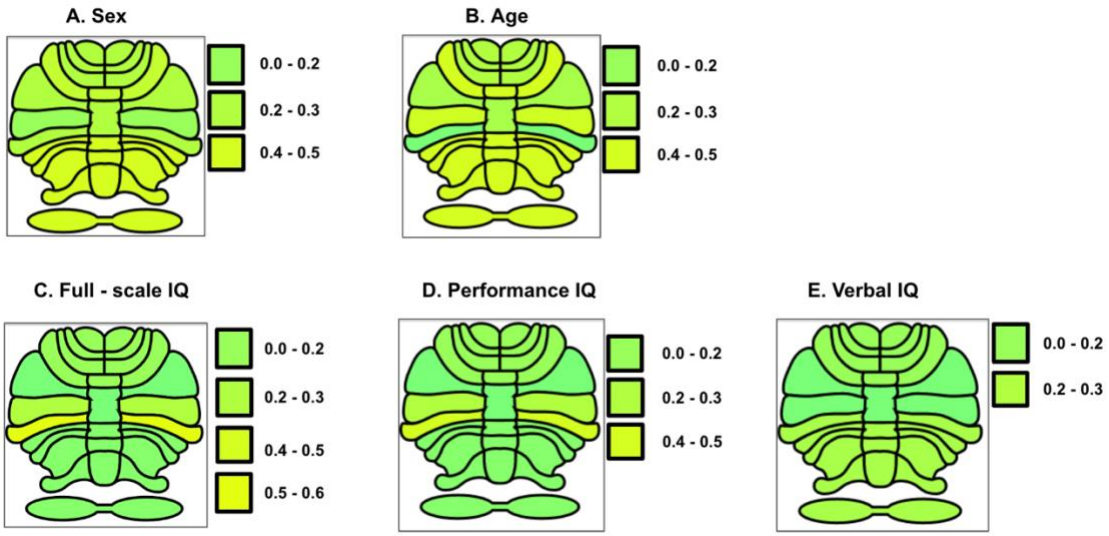
Results are reported in Figure D
The model including performance IQ term and its interaction was a better fit (as defined with a lower AIC) on average in 18% of the site of inclusion across cerebellar regions (from 0 to 42%, see Figure A4), suggesting a relatively modest influence of performance IQ in the heterogeneity focused analysis.

Heterogeneity focused analysis of verbal IQ
As described above and in Bedford et al., we compared the AIC of the two following models:

$$\text{Cereb. volume}_i = \beta_0 + \beta_1\text{Diagnosis} + \beta_2\text{Sex} + \beta_3\text{Age} + \beta_4(\text{Diagnosis * verbal IQ}) + \varepsilon_i$$
$$\text{Cereb. volume}_i = \beta_0 + \beta_1\text{Diagnosis} + \beta_2\text{Sex} + \varepsilon_i$$

Results are reported in Figure E
The model including verbal IQ term and its interaction was a better fit (as defined with a lower AIC) on average in 13% of the site of inclusion across cerebellar regions (from 0 to 23%, see Figure A5), suggesting a relatively modest influence of age in the heterogeneity focused analysis.

Figure Supp. Mat. 14: Percentage of subjects for which adding Sex (A), Age (B), Full scale IQ (C), Performance (D), Verbal IQ (E) + interactions improved AIC criteria.



To ensure that there was no negative effect of sex, we repeated our main analyses in both males and females separately.

Table A. Effect of diagnosis on cerebellar volume in males

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2				
					Group	Age	Site	ICV	FSIQ
Cerebellum	0.9455	-0.07	[-2.01 ; 1.88]	9	0.000	0.036	0.144	0.332	0.004
Ant. lobe	0.8957	0.13	[-0.32 ; 0.37]	9	0.000	0.007	0.210	0.208	0.005
Lobule VI	0.5286	-0.63	[-0.65 ; 0.33]	9	0.001	0.020	0.055	0.147	0.000
Crus I	0.8952	0.13	[-0.69 ; 0.79]	9	0.000	0.005	0.044	0.137	0.006
Crus II	0.5175	-0.65	[-0.67 ; 0.34]	9	0.001	0.030	0.113	0.086	0.002
Lobule VIIb	0.6747	0.42	[-0.22 ; 0.34]	9	0.001	0.026	0.063	0.118	0.013
Post. Inf lobe	0.7408	0.33	[-0.58 ; 0.81]	9	0.000	0.014	0.190	0.265	0.000

Pval, P-value; tval, t-value; IC., confidence interval; dof, degrees of freedom, FSIQ : full-scale IQ
(diagnosis, age, age, site, ICV and IQ as covariates)

Table B. Effect of diagnosis on cerebellar volume in females

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2				
					Group	Age	Site	ICV	FSIQ
Cerebellum	0.1645	-1.40	[-4.41 ; 0.76]	9	0.013	0.058	0.076	0.204	0.013
Ant. lobe	0.7211	0.36	[-0.39 ; 0.56]	9	0.001	0.043	0.127	0.115	0.002
Lobule VI	0.3751	-0.89	[-1.01 ; 0.38]	9	0.006	0.027	0.031	0.108	0.005
Crus I	0.4256	-0.80	[-0.15 ; 0.07]	9	0.004	0.052	0.330	0.025	0.002
Crus II	0.4562	-0.75	[-0.94 ; 0.42]	9	0.004	0.051	0.118	0.047	0.022
Lobule VIIb	0.4067	-0.83	[-0.16 ; 0.06]	9	0.005	0.008	0.426	0.003	0.001
Post. Inf lobe	0.9824	0.02	[-0.91 ; 0.93]	9	0.000	0.022	0.079	0.104	0.004

Pval, P-value; tval, t-value; IC., confidence interval; dof, degrees of freedom, FSIQ : full-scale IQ
(diagnosis, age, age, site, ICV and IQ as covariates)

Supplementary Material 15: Effect of diagnosis on cerebellar structure in IQ and age-centred intervals

To investigate the effect of age on the cerebellar anatomy of individuals with ASD, we splitted our sample in age centered analysis at the interval of 2 years at 8, 12, 16, 20, 24 and 28 years old, while considering age as a linear term (see Bedford et al.). We considered the following age range : 6 to 10 yo, 10 to 14 yo, 14 to 18 yo, 18 to 22 yo, 22 to 26 yo, 26 to 30 yo. Next we re-ran linear models at each age interval (including site, sex, full scale IQ and ICV as covariates) and found no significant difference in cerebellar volumes when comparing individuals with ASD and controls. All Cohen's f2 effect sizes for diagnosis across cerebellar regions of interest were inferior to 0.01. Regression lines comparing the effect of age on cerebellar volumes in individuals with ASD and controls are reported in Figure A.

To investigate the effect of full-scale IQ on the cerebellar anatomy of individuals with ASD, we divided our sample in IQ centered analyses at the interval of 10 IQ points at 80, 100, 120, while considering IQ measures as a linear term (see Bedford et al.). We considered the following IQ range: [70 - 90], [90 - 110], [110-130]. Next, we re-ran linear models separately at each IQ interval (including site, sex, ICV and age as covariates) and found no significant difference in cerebellar volume when comparing ASD and controls. All Cohen's f2 effect sizes for diagnosis across cerebellar regions of interest were inferior to 0.01. Regression lines comparing the effect of full-scale IQ on cerebellar volumes in individuals with ASD and controls are Reported in Figure B.

Figure A : Correlation between age and cerebellar volumes in individuals with ASD or TD

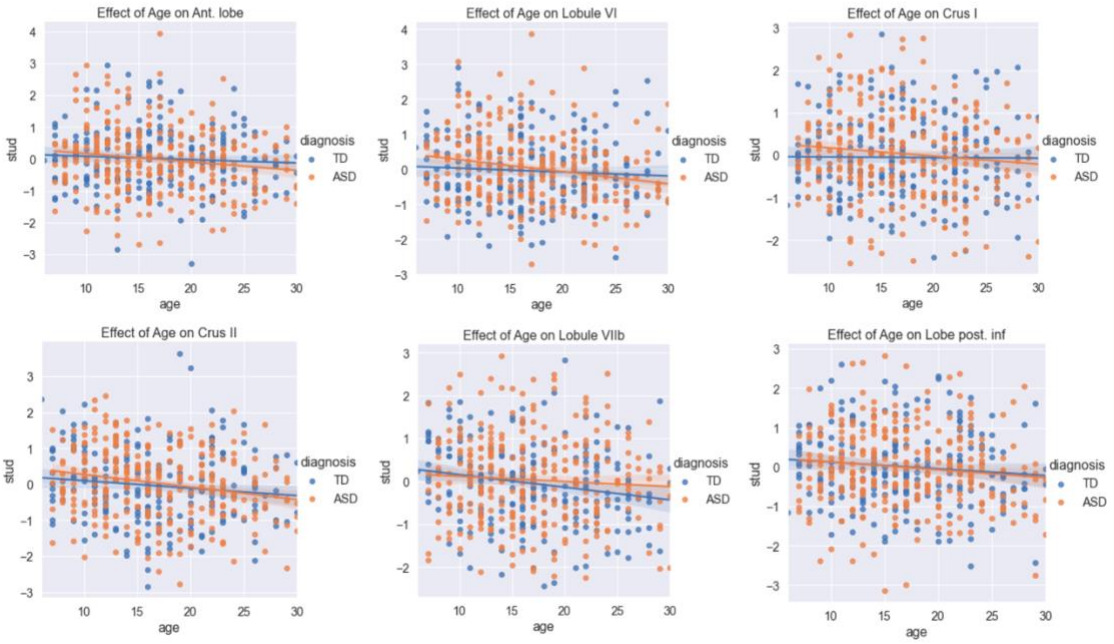
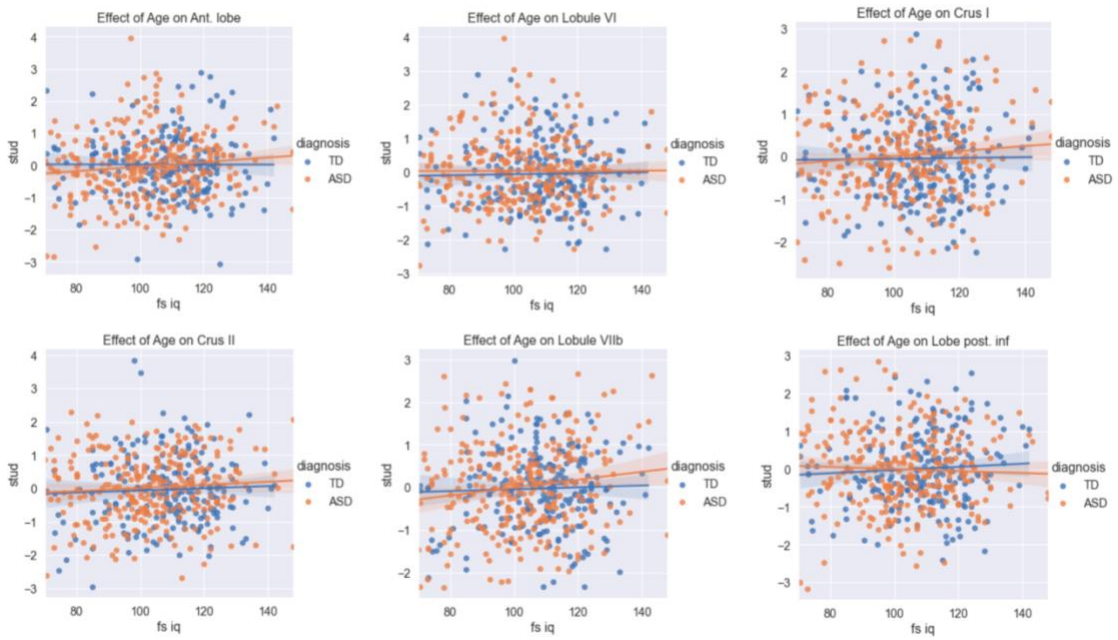


Figure B : Correlation between age and full scale IQ in individuals with ASD or TD



Supplementary Material 16: Matched sample analysis

In our sample, there was difference between patients and controls in sex ratio and full-scale IQ.

To ensure that this difference did not prevent us to found significant findings, we matched ASD and TD sample based on sex, iq and site. We use the pymatch package (https://pypi.org/project/pymatch/) to match patients and controls, based, on sex, age, and full-scale IQ. The result of this matching procedure is reported in Table A.
In this sample, we re-ran our analysis in this matched sample, including ICV and site as covariates. We did not include sex, age, and full-scale IQ as covariate since there was no significant difference in both group. However, adding those covariates did not change our results. Results are reported in Table B (restricted model) and C (with all covariates)

Table A. Study population - Matched sample

	ASD n = 211	Controls n = 211	Statistics	P value
Mean Age (SD) [min - max]	17 (5) [7.00 - 30]	17 (5) [6 - 30]	t-test	NS.
Sex ratio (% of Males)	0.65	0.65	chi-2	NS.
Mean full scale iq (SD) [min - max]	106 (19) [70 - 148]	107 (18) [70 - 140]	t-test	NS.

Table B – Effect of diagnosis in the matched sample (restricted model)

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2		
					Group	Site	FSIQ
Cerebellum	0.4477	-0.76	[-2.41 ; 1.06]	7	0.080	0.036	0.220
Ant. lobe	0.9796	0.03	[-0.30 ; 0.31]	7	0.051	0.152	0.141
Lobule VI	0.4038	-0.84	[-0.62 ; 0.25]	7	0.036	0.010	0.104
Crus I	0.2133	-1.25	[-1.06 ; 0.24]	7	0.022	0.011	0.078
Crus II	0.5858	-0.55	[-0.55 ; 0.31]	7	0.046	0.068	0.042
Lobule VIIb	0.2806	-1.08	[-0.36 ; 0.10]	7	0.029	0.038	0.092
Post. Inf lobe	0.6061	0.52	[-0.44 ; 0.75]	7	0.039	0.097	0.161

Table C– Effect of diagnosis in the matched sample including all covariates

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					Group	Sex	Age	Site	ICV	FSIQ
Cerebellum	0.3544	-0.93	[-2.47 ; 0.89]	10	0.002	0.027	0.037	0.110	0.279	0.009
Ant. lobe	0.9369	-0.08	[-0.31 ; 0.29]	10	0.000	0.027	0.017	0.193	0.183	0.002
Lobule VI	0.3285	-0.98	[-0.64 ; 0.21]	10	0.002	0.010	0.026	0.047	0.137	0.000
Crus I	0.1960	-1.30	[-1.07 ; 0.22]	10	0.004	0.013	0.004	0.036	0.102	0.007
Crus II	0.4975	-0.68	[-0.57 ; 0.28]	10	0.001	0.000	0.037	0.110	0.085	0.006
Lobule VIIb	0.2350	-1.19	[-0.37 ; 0.09]	10	0.003	0.006	0.019	0.068	0.120	0.004
Post. Inf lobe	0.6561	0.45	[-0.46 ; 0.72]	10	0.000	0.022	0.011	0.131	0.194	0.004

Supplementary Material 17: Dimensional analyses in the Healthy Brain Network cohort

The Healthy Brain Network cohort is a large transdiagnostic dataset of both brain imaging and clinical/behavioral assessments from children and adolescents (5-21 years) with psychiatric disorders or at risk for such disorders (Alexander et al., 2017).

Healthy Brain Network : inclusion and exclusion criteria

The inclusion criteria of HBN are broad and only require participants to be aged from 5 to 21 yo, to speak english and to be able to undergo a clinical evaluation. The Healthy Brain Network is a transdiagnostic pediatric cohort. Thus, there is only a limited number of individuals without a diagnosis (Alexander et al. 2017), which is not well suited for case-control comparison. Instead, the advantage of this cohort relies on the possibility to investigate associations between brain and behavior from a dimensional perspective.

The exclusion criteria include having a severe neurological disorder or suffering from an acute psychotic episode. Written informed consent was obtained from participants aged 18 years or older, and from legal guardians, in addition to themselves, for those under 18 years old. This protocol was approved by the Chesapeake Institutional Review Board, is conducted following the Declaration of Helsinki for human research and is described elsewhere (Alexander et al. 2017).

Subject selection

As we did in our initial study, we excluded all participants with an IQ below 70 measured with the Wechsler Adult Intelligence Scale (WASI-II) or the Wechsler Intelligence Scale for Children (WISC-V) depending on the age of the participants. The full clinical assessment of the HBN cohort is described elsewhere (Alexander et al., 2017).

Following our negative findings in the EU-AIMS cohort, our goal was to study how social impairments related to autism were associated with cerebellar anatomy in a transdiagnostic cohort. Symptoms related to autism were measured with the second version of the Social Responsiveness Scale (SRS-2, t-score) (Constantino et al., 2004).

The SRS T-score can be interpreted as follows : 60 to 65 indicate a mild to moderate deficit in social interactions, 66 to 75 can be interpreted with moderate deficiencies in reciprocal social interactions that are clinically significant and lead to interferences in everyday social interactions. A score above 76 is considered as severe and strongly associated with a diagnosis of ASD. We defined within the HBN cohort two different samples, based on the severity of the SRS. The first sample (referred to as “Relaxed-sample” - R-sample) included individuals with a SRS score above 60. The second sample (referred to as “Strict-sample”) included individuals with a SRS score above 76.

Data acquisition, data processing and quality assessment

MRI scans were acquired on three distinct sites in New York City: Staten Island, Rutgers University and Cornell Brain Imaging Center. Staten Island images were acquired on a 1.5T Siemens Avanto (*TR* = 2730 ms, *TE* = 1.64 ms, flip angle = 7°, slice number= 176, voxel dimensions = 1.0 × 1.0 × 1.0 mm^3). Rutgers University images were acquired on a 3T Siemens Tim Trio (*TR* = 2500 ms, *TE* = 3.15 ms, flip angle = 8°, slice number= 224, voxel dimensions = 0.8 × 0.8 × 0.8 mm^3). Cornell Brain Imaging Center images were acquired on a Siemens Prisma 3T MRI (*TR* = 2500 ms, *TE* = 3.15 ms, flip angle = 8°, slice number= 224, voxel dimensions = 0.8 × 0.8 × 0.8 mm^3).

As we did in our initial study, we conducted a visual quality check on the cerebellar parcellations. First, we inspected the T1 MRI and excluded those with evident motion. Subjects with evident motion artifacts were removed from further analyses. Second, we conducted a fully automated well validated (Carrasco et al. 2019) cerebellar parcellation with

the CERES pipeline (Romero et al., 2017). All data were processed on a high computing performance cluster in Bordeaux (PC) by the team that developed the CERES pipeline as in the EU-AIMS sample. Next, an expert rater (YE) - blind of the clinical features of each participant - visually assessed the quality of MRI scans in every slice for each spatial plan of the cerebellum. We identified subjects with non-cerebellar voxels labeled as voxels belonging to the cerebellum and vice versa and subjects with parcellation errors within the cerebellar lobules. All images with parcellation defects were excluded from further analyses.

Statistical analyses

In the relaxed and strict sample (see subject selection section), we investigated if there was a significant effect of the SRS scale on cerebellar anatomy. As we did in the EU-AIMS sample, we included in a linear model SRS2 T-score, age, sex, site of inclusion, ICV and full-scale IQ (FSIQ) as covariates. Following the remarks of Reviewer #1, we tested if (i) there was a SRS by FSIQ interaction and if our results changed without including FSIQ as a covariate.

In addition, following the remark of Reviewer #2 (comment n°14) we repeated our analyses with 2 subscales of the SRS. The SRS-2 can be divided into two scores corresponding to the DSM-V autism dimensions : a social communication/interaction score (SCI; sum of Awareness, Cognition, Communication, and Motivation subscales) and a restricted/repetitive behavior score (RRB; Mannerisms subscale) (Frazier et al., 2014; Prigge et al., 2018). In order to differentiate the social component related to autism from the restrictive / repetitive behaviors component, we studied separately the two components of the SRS scale.

Study population

Out of 1453 subjects with structural MRI data and complete clinical assessment (SRS total and subscores, IQ measurement and basic demographic characteristics - age, sex, site of inclusion), we excluded 280 subjects with excessive head motion and low quality of the anatomical image and 279 subjects with parcellation errors. Next we selected only subjects with > 60 SRS2 t-score as described above. The study population is reported in Table A.

Table A. Study population - Healthy Brain Network

Demographic and clinical characteristics	Loose sample n = 352	Strict sample n = 79
Mean age [min - max] (SD)	11 [6 - 18] (3)	11 [6 - 17] (3)
Mean SRS T-score [min - max] (SD)	70 [60 - 90] (8)	82 [76 - 90] (5)
Mean SRS SCI T-score [min - max] (SD)	70 [58 - 90] (8)	81 [73 - 90] (5)
Mean full scale IQ [min - max] (SD)	96 [70 - 147] (16)	93 [70 - 129] (15)
Males / Females	223 / 129	53 / 26
Site of inclusion : CBIC / RU / SI	120 / 150 / 82	30 / 23 / 26

Results

We find no significant effect on the SRS score on cerebellar anatomy in the relaxed and the strict sample. Results are reported in Tables B-1 to B-6. We found no FSIQ by SRS interaction and excluding FSIQ from our model did not change our results.

Likewise, there was no significant influence of the SRS social communication/interaction score and the SRS restricted/repetitive behavioral score.

Table B1. Effect of the total SRS score on cerebellar anatomy in the HBN Relaxed sample

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					SRS	Sex	Age	Site	ICV	IQ
Cerebellum	0.1232	1.55	[-0.02 ; 0.18]	7	0.007	0.008	0.008	0.036	0.338	0.000
Ant. lobe	0.7821	-0.28	[-0.00 ; 0.00]	7	0.000	0.000	0.016	0.067	0.098	0.015
Lob. VI	0.2947	1.05	[-0.01 ; 0.04]	7	0.003	0.000	0.002	0.002	0.145	0.001
Crus I	0.3042	1.03	[-0.02 ; 0.06]	7	0.003	0.002	0.001	0.010	0.167	0.000
Crus II	0.6312	0.48	[-0.02 ; 0.04]	7	0.001	0.009	0.007	0.003	0.063	0.001
Lob. VIIb	0.2579	1.13	[-0.01 ; 0.02]	7	0.004	0.000	0.011	0.054	0.093	0.000
Post. inf lobe	0.9137	0.11	[-0.03 ; 0.04]	7	0.000	0.012	0.010	0.069	0.219	0.002

Table B2. Effect of the total SRS score on cerebellar anatomy in the HBN Strict sample

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					SRS	Sex	Age	Site	ICV	IQ
Cerebellum	0.9458	0.07	[-0.35 ; 0.38]	7	0.000	0.005	0.002	0.144	0.192	0.001
Ant. lobe	0.8410	-0.20	[-0.01 ; 0.01]	7	0.001	0.013	0.052	0.030	0.065	0.010
Lob. VI	0.2443	1.17	[-0.04 ; 0.16]	7	0.019	0.001	0.000	0.062	0.065	0.004
Crus I	0.5078	-0.67	[-0.18 ; 0.09]	7	0.006	0.002	0.017	0.063	0.081	0.001
Crus II	0.7973	0.26	[-0.10 ; 0.13]	7	0.001	0.026	0.011	0.044	0.015	0.000
Lob. VIIb	0.4938	0.69	[-0.04 ; 0.08]	7	0.007	0.001	0.000	0.163	0.039	0.000
Post. inf lobe	0.2917	-1.06	[-0.20 ; 0.06]	7	0.016	0.007	0.020	0.156	0.129	0.007

Table B3. Effect of the total SRS-SCI score (communication and interactions) on cerebellar anatomy in the HBN Relaxed sample

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					SRS-SCI	Sex	Age	Site	ICV	IQ
Cerebellum	0.1341	1.50	[-0.02 ; 0.18]	7	0.007	0.008	0.008	0.037	0.338	0.000
Ant. lobe	0.9504	-0.06	[-0.00 ; 0.00]	7	0.000	0.000	0.016	0.066	0.099	0.015
Lob. VI	0.2190	1.23	[-0.01 ; 0.05]	7	0.004	0.001	0.002	0.002	0.146	0.001
Crus I	0.2888	1.06	[-0.02 ; 0.06]	7	0.003	0.002	0.002	0.010	0.167	0.000
Crus II	0.9919	-0.01	[-0.03 ; 0.03]	7	0.000	0.009	0.006	0.003	0.062	0.001
VIIb	0.3948	0.85	[-0.01 ; 0.02]	7	0.002	0.001	0.011	0.054	0.092	0.000
Post. inf lobe	0.9532	0.06	[-0.04 ; 0.04]	7	0.000	0.012	0.010	0.069	0.219	0.002

Table B4. Effect of the total SRS-SCI score (communication and interactions) on cerebellar anatomy in the HBN Strict sample

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					SRS-SCI	Sex	Age	Site	ICV	IQ
Cerebellum	0.5795	0.56	[-0.24 ; 0.43]	7	0.004	0.006	0.001	0.140	0.192	0.001
Ant. lobe	0.8890	0.14	[-0.01 ; 0.01]	7	0.000	0.014	0.052	0.030	0.064	0.009
Lob. VI	0.1148	1.60	[-0.02 ; 0.17]	7	0.035	0.001	0.000	0.055	0.069	0.005
Crus I	0.5599	-0.59	[-0.17 ; 0.09]	7	0.005	0.002	0.018	0.063	0.079	0.002
Crus II	0.9263	-0.09	[-0.11 ; 0.10]	7	0.000	0.025	0.011	0.046	0.016	0.001
VIIb	0.3758	0.89	[-0.03 ; 0.08]	7	0.011	0.000	0.000	0.157	0.040	0.000
Post. inf lobe	0.5262	-0.64	[-0.16 ; 0.08]	7	0.006	0.007	0.019	0.156	0.123	0.005

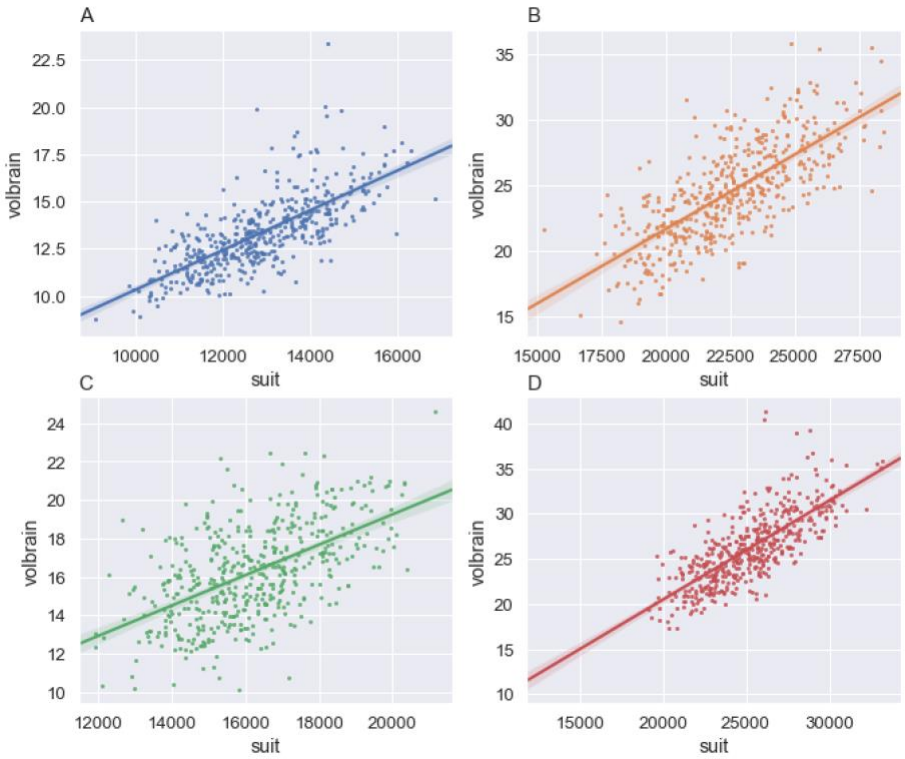
Table B5. Effect of the total SRS-RRB score (restrictive and repetitive behaviors) on cerebellar anatomy in the HBN Relaxed sample

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					SRS-RRB	Sex	Age	Site	ICV	IQ
Cerebellum	0.2522	1.15	[-0.03 ; 0.12]	7	0.004	0.008	0.006	0.037	0.336	0.001
Ant. lobe	0.4929	-0.69	[-0.00 ; 0.00]	7	0.001	0.000	0.016	0.068	0.097	0.015
Lob. VI	0.6949	0.39	[-0.02 ; 0.02]	7	0.000	0.000	0.001	0.002	0.143	0.002
Crus I	0.5501	0.60	[-0.02 ; 0.04]	7	0.001	0.002	0.001	0.010	0.166	0.000
Crus II	0.1434	1.47	[-0.01 ; 0.04]	7	0.006	0.008	0.006	0.003	0.066	0.001
VIIb	0.1401	1.48	[-0.00 ; 0.02]	7	0.006	0.000	0.009	0.054	0.094	0.000
Post. inf lobe	0.9162	0.11	[-0.02 ; 0.03]	7	0.000	0.012	0.010	0.069	0.219	0.002

Table B6. Effect of the total SRS-RRB score (restrictive and repetitive behaviors) on cerebellar anatomy in the HBN Strict sample

ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					SRS-RRB	Sex	Age	Site	ICV	IQ
Cerebellum	0.2407	-1.18	[-0.38 ; 0.10]	7	0.019	0.005	0.000	0.140	0.210	0.000
Ant. lobe	0.3919	-0.86	[-0.01 ; 0.00]	7	0.010	0.013	0.044	0.026	0.072	0.011
Lob. VI	0.5909	-0.54	[-0.09 ; 0.05]	7	0.004	0.002	0.000	0.072	0.076	0.001
Crus I	0.9350	-0.08	[-0.09 ; 0.09]	7	0.000	0.001	0.017	0.059	0.075	0.003
Crus II	0.4933	0.69	[-0.05 ; 0.10]	7	0.007	0.026	0.008	0.047	0.012	0.000
VIIb	0.6846	-0.41	[-0.04 ; 0.03]	7	0.002	0.001	0.000	0.169	0.045	0.001
Post. inf lobe	0.1066	-1.63	[-0.15 ; 0.02]	7	0.036	0.008	0.029	0.144	0.143	0.005

Supplementary Material 18: Comparison of parcellation using CERES (volbrain) and SUIT pipeline in 4 regions of interest: A. anterior lobe (blue), B. Crus I (orange), C. Crus II (green) and D. postero-inferior lobe (red)

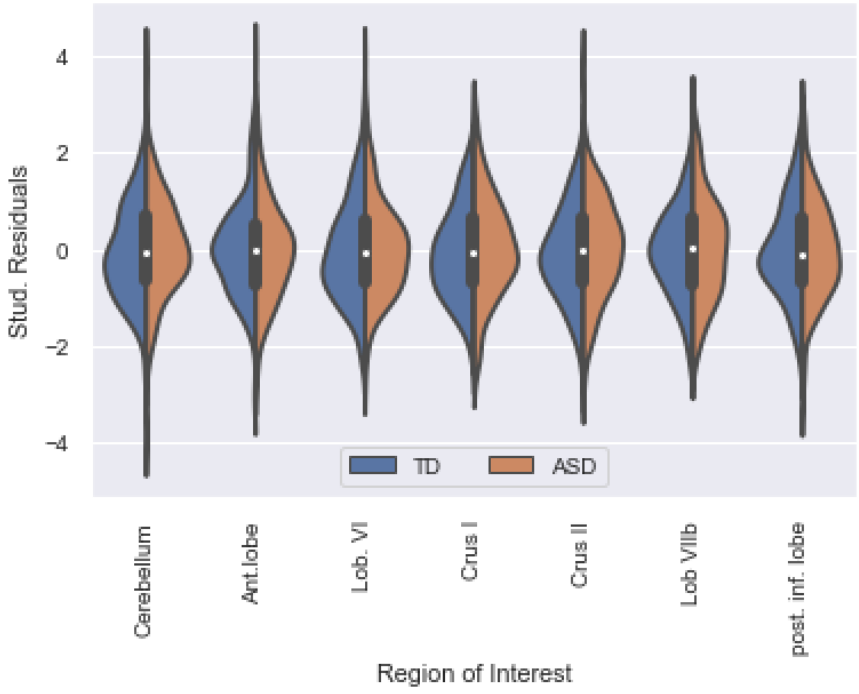


Legend
X-axis : volume of the cerebellum calculated with the SUIT pipeline
Y-axis : volume of the cerebellum calculated with the CERES (volbrain) pipeline

Supplementary Material 19: Effect of diagnosis on the cerebellar vermis

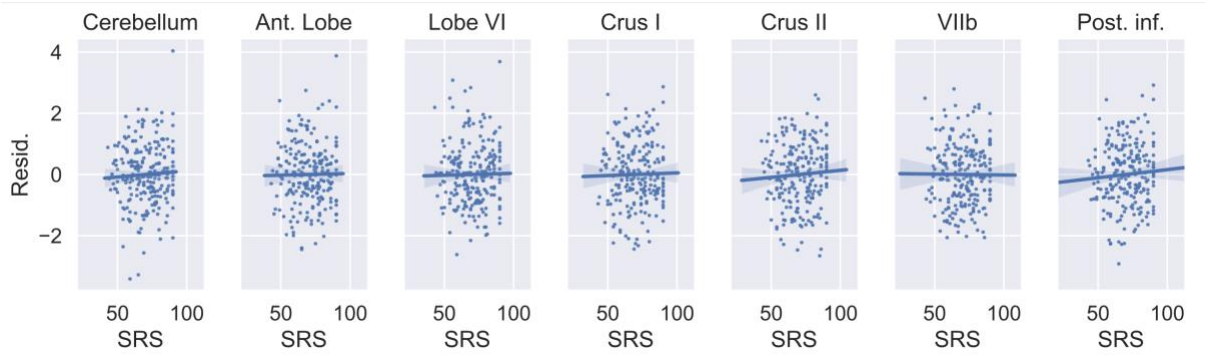
ROI	pval	tval	[ICinf ; ICsup]	dof	Cohen's f2					
					Diagnosis	Sex	Age	Site	ICV	FSIQ
Vermis	0.3824	-0.87	[-2.26 ; 0.87]	10	0.002	0.009	0.030	0.013	0.383	0.001
Vermis lob. VI-VII	0.8042	0.25	[-0.24 ; 0.31]	10	< 0.001	0.012	0.027	0.023	0.353	0.001

Supplementary Material 20: Comparison of cerebellar volumes in patients and controls (studentized residuals) with the CERES pipeline



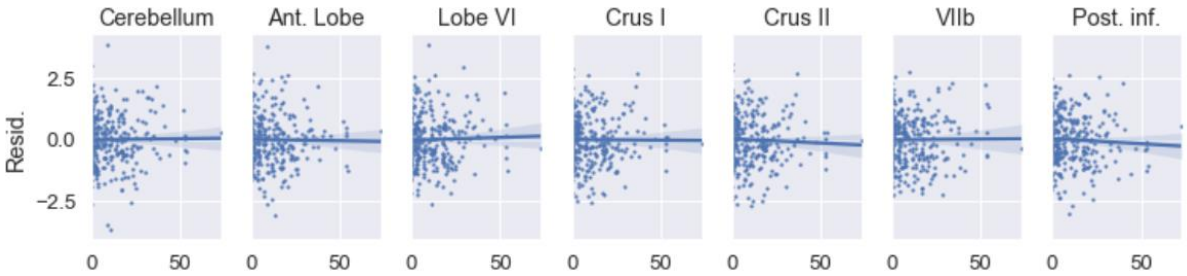
Legend
Stud. Residuals = studentized residuals
TD : typically developing subjects
ASD : individuals with Autism Spectrum Disorders
Ant. lobe : anterior lobule of the cerebellum
Post. inf. Lobe : postero-inferior lobe of the cerebellum

Supplementary Material 21: No correlation between SRS-2 score and cerebellar volumes (residuals) - CERES pipeline

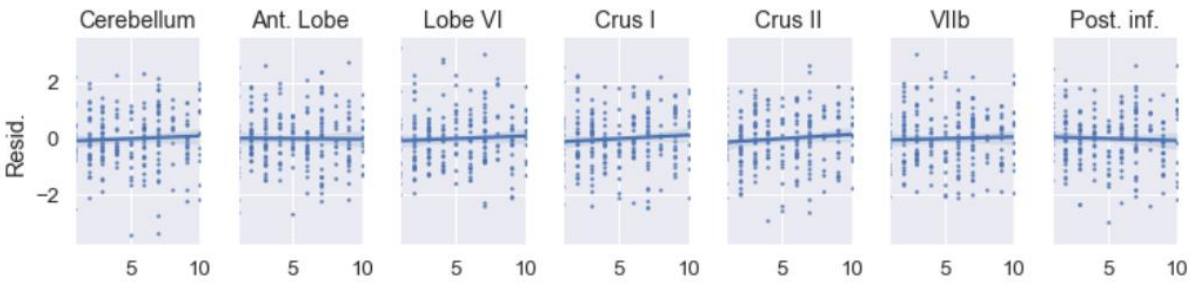


Legend
SRS-2: Social Responsiveness Scale

Supplementary Material 22: No correlation between RBS-R score and cerebellar volumes (residuals) - CERES pipeline



Supplementary Material 23: No correlation between ADOS-CCS score and cerebellar volumes (residuals) - CERES pipeline



Supplementary Material 24: Positive and negative deviations from the normative model in regions of interest of the cerebellum

Region of interest (Cerebellum)	p-value	ASD (mean / med)	TD (mean / med)
Left ant neg	0.05	75 / 3	110.0806 / 6
Left ant pos	0.119	245 /1	157.1137 / 1
Right ant neg	0.248	111 / 3	111.2227 / 4
Right ant pos	0.067	239 / 0	163.1374 / 1
Left VI neg	0.405	92 / 3	111.6493 / 5
Left VI pos	0.394	286 / 1	186.2133 / 1
Right VI neg	0.151	104 / 5	101.7962 / 7
Right VI pos	0.349	197 / 3	148.1422 / 3
Left CrusI neg	0.100	94 / 4	133.3223 / 10
Left CrusI pos	0.361	327/ 6	181.0995 / 7
Right CrusI neg	0.071	109 / 19	173.7488 / 35
Right CrusI pos	0.220	246/23	171.8389 / 21
Left CrusII neg	0.146	90/1	101.6777 / 4
Left CrusII pos	0.284	1662/1	103.1137 / 1
Right CrusII neg	0.384	82/3	106.5071 / 2
Right CrusII pos	0.314	1516/6.5	104.9763 / 5
Vermis neg	0.475	92/4	72.9005 / 4
Vermis pos	0.182	122/0	88.0284 / 1
Left VIIb neg	0.200	44/0	49.3507 / 0
Left VIIb pos	0.437	77/0	66.0995 / 0
Right VIIb neg	0.332	53/1	63.9668 / 0
Right VIIb pos	0.431	72/0	54.4313 / 1
Left post. inf neg	0.331	91/6	130.2133 / 8
Left post. inf pos	0.107	204/2	190.9526 / 4
Right post. inf neg	0.312	137/5	162.1137 / 9
Right post. inf pos	0.234	170/5	177.2654 / 3

Legend:
ASD: individuals with autism spectrum disorders
TD: typically developing individuals
Neg: negative deviation from the normative model
Pos: positive deviation from the normative model

Supplementary Material 25: Prediction performance of diagnostic category based on the cerebellar anatomy

Features	Sensitivity	Specificity	Balanced Accuracy
Regions of interest (CERES)	0.52	0.51	0.52
Regions of interest (SUIT)	0.53	0.49	0.51
Grey matter cerebellar Voxel based morphometry (SUIT)	0.51	0.52	0.51

Supplementary Material 26: Clinical characteristics in individuals with ASD across sites of inclusion

	Cam	KCL	Man	Nij	Rom	Utr	Statistics	P value
Sample size	45	78	21	77	20	33	*NA*	*NA*
Mean Age [min - max]	16 [7 - 30]	17 [7 - 27]	15 [10 - 24]	16.87 [6 - 30]	24 [19 - 30]	17 [8 - 28]	*ANOVA F = 11.7*	*< 0.0001*
Sex ratio (% of Males)	*0.66*	0.79	0.76	0.64	0.70	0.70	*Chi-2* Q = 2.96	0.56
Mean full scale iq [min - max]	106 [73 - 148]	105 [70 - 148]	101 [78 - 139]	104 [70 - 142]	103 [71 - 136]	106 [72 - 128]	ANOVA F = 1.83	*0.10*
Mean SRS t-score [min - max]	72 [43 - 90]	73 [47 - 90]	74 [62 - 90]	68 [44 - 90]	66 [50 - 83]	63 [45 - 90]	ANOVA F = 2.45	0.005

NS : non significant ; SRS : Social responsiveness score,
Sites of inclusion : Cam : Cambridge, KCL : King's College London, Man : Mannheim, Nij : Nijmegen ; Rom : Rome, Utr : Utrecht

Supplementary Material 27: Meta-analysis metrics across the sites of the study

Region of interest	Heterogeneity	I2 score	Parcellation method
Total cerebellum	Q = 5.20 - p = 0.39	3.84 %	CERES
Ant. lobe	Q = 4.20 - p = 0.52	0 %	CERES
Lobule VI	Q = 7.61 - p = 0.18	34.3 %	CERES
Crus I	Q = 1.81 - p = 0.87	0 %	CERES
Crus II	Q = 6.92 - p = 0.23	27.73 %	CERES
Lobule VIIb	Q = 1.97 - p = 0.85	0 %	CERES
Post. inf lobe	Q = 1.81 - p = 0.87	0 %	CERES
Vermis	Q = 8.65 - p = 0.13	42 %	SUIT
Lobule VI-VII of the vermis	Q = 8.88 - p = 0.43	43 %	SUIT

References

1. Li W-K, Hausknecht MJ, Stone P, Mauk MD (2013): Using a million cell simulation of the cerebellum: network scaling and task generality. *Neural Netw Off J Int Neural Netw Soc* 47: 95–102.
2. Sereno MI, Diedrichsen J, Tachrount M, Testa-Silva G, d'Arceuil H, De Zeeuw C (2020): The human cerebellum has almost 80% of the surface area of the neocortex. *Proc Natl Acad Sci U S A* 117: 19538–19543.
3. Van Overwalle F, D'aes T, Mariën P (2015): Social cognition and the cerebellum: A meta-analytic connectivity analysis. *Hum Brain Mapp* 36: 5137–5154.
4. Guell X, Gabrieli JDE, Schmahmann JD (2018): Triple representation of language, working memory, social and emotion processing in the cerebellum: convergent evidence from task and seed-based resting-state fMRI analyses in a single large cohort. *NeuroImage* 172: 437–449.
5. Stoodley CJ, D'Mello AM, Ellegood J, Jakkamsetti V, Liu P, Nebel MB, *et al.* (2017): Altered cerebellar connectivity in autism and cerebellar-mediated rescue of autism-related behaviors in mice. *Nat Neurosci* 20: 1744–1751.
6. Kelly E, Meng F, Fujita H, Morgado F, Kazemi Y, Rice LC, *et al.* (2020): Regulation of autism-relevant behaviors by cerebellar-prefrontal cortical circuits. *Nat Neurosci* 23: 1102–1110.
7. Fatemi SH (2013): Cerebellum and autism. *Cerebellum Lond Engl* 12: 778–779.
8. Wang SS-H, Kloth AD, Badura A (2014): The cerebellum, sensitive periods, and autism. *Neuron* 83: 518–532.
9. Traut N, Beggiato A, Bourgeron T, Delorme R, Rondi-Reig L, Paradis A-L, Toro R (2018): Cerebellar Volume in Autism: Literature Meta-analysis and Analysis of the Autism Brain Imaging Data Exchange Cohort. *Biol Psychiatry* 83: 579–588.

10. Laidi C, Boisgontier J, Chakravarty MM, Hotier S, d'Albis M-A, Mangin J-F, *et al.* (2017): Cerebellar anatomical alterations and attention to eyes in autism. *Sci Rep* 7: 12008.

11. Waterhouse L, London E, Gillberg C (2017): The ASD diagnosis has blocked the discovery of valid biological variation in neurodevelopmental social impairment. *Autism Res Off J Int Soc Autism Res* 10: 1182.

12. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, *et al.* (2019): Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biol Psychiatry Cogn Neurosci Neuroimaging* 4: 567–578.

13. Floris DL, Wolfers T, Zabihi M, Holz NE, Zwiers MP, Charman T, *et al.* (2021): Atypical Brain Asymmetry in Autism-A Candidate for Clinically Meaningful Stratification. *Biol Psychiatry Cogn Neurosci Neuroimaging* 6: 802–812.

14. Wolfers T, Floris DL, Dinga R, van Rooij D, Isakoglou C, Kia SM, *et al.* (2019): From pattern classification to stratification: towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neurosci Biobehav Rev* 104: 240–254.

15. Lombardo MV, Lai M-C, Baron-Cohen S (2019): Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol Psychiatry* 24: 1435–1450.

16. Marquand AF, Rezek I, Buitelaar J, Beckmann CF (2016): Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry* 80: 552–561.

17. Tillmann J, Uljarevic M, Crawley D, Dumas G, Loth E, Murphy D, *et al.* (2020): Dissecting the phenotypic heterogeneity in sensory features in autism spectrum disorder: a factor mixture modelling approach. *Mol Autism* 11: 67.

18. Loth E, Charman T, Mason L, Tillmann J, Jones EJH, Wooldridge C, *et al.* (2017): The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. *Mol Autism* 8: 24.

19. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, *et al.* (2000): The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30: 205–223.

20. Charman T, Loth E, Tillmann J, Crawley D, Wooldridge C, Goyard D, *et al.* (2017): The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. *Mol Autism* 8: 27.

21. Mihailov A, Philippe C, Gloaguen A, Grigis A, Laidi C, Piguet C, *et al.* (2020): Cortical signatures in behaviorally clustered autistic traits subgroups: a population-based study. *Transl Psychiatry* 10: 207.

22. Shaw P, Ishii-Takahashi A, Park MT, Devenyi GA, Zibman C, Kasparek S, *et al.* (2018): A multicohort, longitudinal study of cerebellar development in attention deficit hyperactivity disorder. *J Child Psychol Psychiatry* 59: 1114–1123.

23. Webb SJ, Sparks B-F, Friedman SD, Shaw DWW, Giedd J, Dawson G, Dager SR (2009): Cerebellar vermal volumes and behavioral correlates in children with autism spectrum disorder. *Psychiatry Res* 172: 61–67.

24. Stoodley CJ (2016): The cerebellum and neurodevelopmental disorders. *Cerebellum Lond Engl* 15: 34–37.

25. Romero JE, Coupé P, Giraud R, Ta V-T, Fonov V, Park MTM, *et al.* (2017): CERES: A new cerebellum lobule segmentation method. *NeuroImage* 147: 916–924.

26. Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N (2009): A probabilistic MR atlas of the human cerebellum. *NeuroImage* 46: 39–46.

27. Allen G, Courchesne E (2003): Differential effects of developmental cerebellar abnormality on cognitive and motor functions in the cerebellum: an fMRI study of autism. *Am J Psychiatry* 160: 262–273.

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

29. Pierce K, Courchesne E (2001): Evidence for a cerebellar role in reduced exploration and stereotyped behavior in autism. *Biol Psychiatry* 49: 655–664.

30. Moberget T, Doan NT, Alnæs D, Kaufmann T, Córdova-Palomera A, Lagerberg TV, *et al.* (2018): Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Mol Psychiatry* 23: 1512–1520.

31. Laidi C, d'Albis M-A, Wessa M, Linke J, Phillips ML, Delavest M, *et al.* (2015): Cerebellar volume in schizophrenia and bipolar I disorder with and without psychotic features. *Acta Psychiatr Scand* 131: 223–233.

32. Laidi C, Hajek T, Spaniel F, Kolenic M, d'Albis M-A, Sarrazin S, *et al.* (2019): Cerebellar parcellation in schizophrenia and bipolar disorder. *Acta Psychiatr Scand* 140: 468–476.

33. Laidi C, Hajek T, Spaniel F, Kolenic M, d'Albis M-A, Sarrazin S, *et al.* (2019): Cerebellar parcellation in schizophrenia and bipolar disorder. *Acta Psychiatr Scand* 140: 468–476.

34. Thissen D, Steinberg L, Kuang D (2002): Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *J Educ Behav Stat* 27: 77–83.

35. Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BTT (2011): The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol* 106: 2322–2345.

36. Buckner RL (2013): The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron* 80: 807–815.

37. Bethlehem RAI, Seidlitz J, Romero-Garcia R, Trakoshis S, Dumas G, Lombardo MV (2020): A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. *Commun Biol* 3: 486.

38. Hong S-J, Valk SL, Di Martino A, Milham MP, Bernhardt BC (2018): Multidimensional Neuroanatomical Subtyping of Autism Spectrum Disorder. *Cereb Cortex N Y N 1991* 28: 3578–3588.

39. Levitt JG, Blanton R, Capetillo-Cunliffe L, Guthrie D, Toga A, McCracken JT (1999): Cerebellar vermis lobules VIII — X in autism. *Prog Neuropsychopharmacol Biol Psychiatry* 23: 625–633.

40. D'Mello AM, Crocetti D, Mostofsky SH, Stoodley CJ (2015): Cerebellar gray matter and lobular volumes correlate with core autism symptoms. *NeuroImage Clin* 7: 631–639.

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

42. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013): Power failure: why small sample size undermines the reliability of neuroscience. ***Nat Rev Neurosci*** **14: 365–376.**

43. Park MTM, Pipitone J, Baer LH, Winterburn JL, Shah Y, Chavez S, ***et al.*** **(2014): Derivation of high-resolution MRI atlases of the human cerebellum at 3T and segmentation using multiple automatically generated templates.** ***NeuroImage*** **95: 217–231.**