Mapping Cerebellar Anatomical Heterogeneity in Mental and Neurological Illnesses

Authors: Milin Kim[1,2], Esten Leonardsen[1,2], Saige Rutherford [5,6,7], Geir Selbæk[10,11], Karin Persson[10,11], Nils Eiel Steen[1,12], Olav B. Smeland[1], Torill Ueland[1,2], Geneviève Richard[1], Christian F. Beckmann[6,14], Andre F. Marquand[6,13,], For the Alzheimer's Disease Neuroimaging Initiative (ADNI)[*], Ole A. Andreassen[1,3], Lars T. Westlye[1,2,3], Thomas Wolfers[1,8,9 **] & Torgeir Moberget[1,4 **].

Affiliations:

[1]Norwegian Centre for Mental Disorders Research (NORMENT), Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [2] Department of Psychology, Faculty of Social Sciences, University of Oslo, Norway; [3] KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway; [4]Department of Behavioral Science, School of Health Sciences, Oslo Metropolitan University - OsloMet, Oslo, Norway; [5] Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, Netherlands; [6]Donders Institute, Radboud University, Nijmegen, Netherlands; [7]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States; [8]Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Germany; [9]German Center for Mental Health (DZPG); [10]Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway; [11]The Norwegian National Centre for Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway; [12]Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway, [13]Department of Neuroimaging, Centre of Neuroimaging Sciences,

Institute of Psychiatry, King's College London, London, UK, [14]Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

**These authors contributed equally.

Keywords: Cerebellum, Normative modelling, Magnetic Resonance Imaging, Individual-level Inference, Heterogeneity Mapping, Schizophrenia, Alzheimer, Autism Spectrum Disorder, Mental Disorders, & Neurological Diseases

Corresponding authors
Milin Kim, Norwegian Centre for Mental Disorders Research (NORMENT), Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. E-mail: milink@student.sv.uio.no

Abstract

The cerebellum has been linked to motor coordination, cognitive and affective processing, in addition to a wide range of clinical illnesses. To enable robust quantification of individual cerebellar anatomy relative to population norms, we mapped the normative development and aging of the cerebellum across the lifespan using brain scans of > 54k participants. We estimated normative models at voxel-wise spatial precision, enabling integration with cerebellar atlases. Applying the normative models in independent samples revealed substantial heterogeneity within five clinical illnesses: autism spectrum disorder, mild cognitive impairment, Alzheimer's disease, bipolar disorder, and schizophrenia. Notably, individuals with autism spectrum disorder and mild cognitive impairment exhibited increased numbers of both positive and negative extreme deviations in cerebellar anatomy, while schizophrenia and Alzheimer's disease predominantly showed negative deviations. Finally, extreme deviations were associated with cognitive scores. Our results provide a voxel-wise mapping of cerebellar anatomy across the human lifespan and clinical illnesses, demonstrating cerebellum's nuanced role in shaping human neurodiversity across the lifespan and in different clinical illnesses.

Introduction

The cerebellum accounts for 10 to 15% of the brain's volume[1] and contains approximately 80% of all neurons[2]. The cerebellum has long been recognized for its involvement in motor functions[3], but has also been linked to human cognitive capacities such as language and social intelligence[4–7]. In line with this expanded view of cerebellar function, the cerebellum has also been implicated in a wide range of mental and neurological illnesses characterized by both motor and cognitive deficits[8]. However, empirical findings on cerebellar structure and function in different mental and neurological illnesses are varied[9,10]. A growing literature[11,12] suggests a key role for cerebellum in autism spectrum disorder (ASD). However, a meta-analysis[13] of human MRI studies revealed no significant group differences in cerebellar volume between individuals with ASD and individuals without diagnosis, suggesting the importance of replication and the inherent uncertainty that comes with small studies[14]. Such inconsistent findings may also be due to the high degree of heterogeneity within clinical groups, a feature which can be explicitly investigated through the construction of normative models.

Normative modelling draws its inspiration from paediatric growth charts, which chart key aspects of a child's development, such as weight or height, over the early years of life. As a machine learning framework, normative modelling is designed to estimate a prototypical and representative developmental trajectory, by mapping different types of variables onto each other, e.g. such as mapping age and sex onto different brain measures[15]. Once such models are established, individuals can be placed in reference to the resulting norms, from which individual-level normative probability maps can be derived. These represent the extent to which an individual deviates from the estimated norm in locations across the brain. In this respect the use

of the normative models enables comparison across studies, scanner sites and lifespan stages by binding to a common reference framework and thus makes studies more comparable[16]. In earlier applications of this approach, we charted the normative trajectory of cortical thickness and subcortical volume[17] across the lifespan. When applied to clinical samples, these normative models revealed a high degree of heterogeneity for individual-level profiles within same condition. This pattern has now been demonstrated and replicated[18] across many different clinical illnesses, such as ASD[19–21], attention deficit hyperactivity disorder (ADHD)[22], dementia including AD (Alzheimer's Disease)[23,24], first-episode psychosis[25], bipolar disorder (BD) and schizophrenia (SZ)[18,26]. Of note, a recent paper demonstrated that while phenotypic heterogeneity within same order may be due to heterogeneity in regional deviations, phenotypic similarities can be coupled to common functional circuits and networks[27]. Together, these findings highlight the need for mapping heterogeneity of complex clinical illnesses using voxel-wise, regional, and network-based approaches.

In the current study, we chart the normative development and aging of the human cerebellum at lobular and voxel-wise spatial precision in a sample of > 54k individuals. We map the heterogeneity of five distinct clinical phenotypes accounting for 132 scanning sites (Table 1 and Supplementary Table 1) and > 143k voxels, providing a generalisable and state-of-the-art reference model for cerebellar measures across the lifespan. Further, and based on cytoarchitectonic, functional and regional parcellations of the cerebellum, we link percentage of extreme deviations to cognitive measures and symptom scores across the groups and in this way shed light on the role of individual-level deviations from normative cerebellar anatomy across five mental and neurological illnesses.

[Table 1]

Results

Lifespan trajectories of cerebellar morphology

Normative models were trained on MRI-derived cerebellar features from the training portion of the overall dataset (Table 1), comprising more than 27k individuals without diagnosis across the lifespan (Fig. 1A). All normative models included age and sex as covariates, while controlling for scanning sites related variability. Cerebellar morphology was quantified as lobular volumes, or voxel-wise grey matter probability values (multiplied by the Jacobian of the transformation matrix to preserve volumetric information after normalization) (see Online Methods and Supplementary Fig. 2 and 3 for details). Figure 1B shows selected normative trajectories of cerebellar lobules. Most of the lobules exhibit a pattern of volume increase until around age 19, followed by a gradual decrease. Notably, we observed flatter trajectories in specific lobules, namely Left X, Left VIIIB, Corpus Medullare, Right VIIIB, Vermis VI, and Vermis VIII. To validate our model fit, we report the performance of these models, in terms of explained variance, plotted on the cerebellar flat maps (Figure 1B and 1C). The lobule with the best model fit was in Right V, while the worst fit was found in Vermis X. Additional evaluation metrics such as kurtosis, skew, and mean squared logarithmic loss (MSLL) are shown in Supplementary Figure 1. Additionally, Figure 1C presents voxel-wise trajectories, as chosen from the 143k models estimated for individual cerebellar grey matter voxels.

[Figure 1]

Altered cerebellar morphology across clinical groups

Figure 2A display nonparametric comparison, Mann-Whitney U-test, of voxel-wise model-derived z-scores between patient groups and individuals without diagnosis in the test datasets (for lobule-wise results, see Supplementary Figure 4A and Supplementary Table 3). All reported effects in the result section are corrected for multiple comparisons using Bonferroni correction (across regions of interest or voxels, number of clinical groups and directions) and only significant results are shown in the figures (corrected $p < 0.05$). Of note, we replicated previous reports of reduced cerebellar volumes in SZ by observing significant negative effect sizes (rank biserial correlations) of deviations in the negative direction, indicating greater strength of SZ than individuals without diagnosis. Similar patterns were seen for SZ in lobular volumes, while AD and MCI exhibited a notably different pattern of negative deviations in lobular volumes compared to voxel-wise maps and furthermore in uncorrected maps (Supplementary Fig. 5). Significant negative effects were observed in Right V for MCI and AD in addition to Right Crus I for AD when corrected. ASD showed negative deviations in the anterior regions for lobular volumes while showed small negative effects in voxel-wise maps in the posterior regions. BD did not show noticeable effects in terms of z-scores with the individuals without diagnosis after multiple comparisons.

[Figure 2]

Individual-level extreme deviations are patient specific and transdiagnostic

The normative modelling approach allows us to quantify extreme positive and negative deviations, here defined as $|z| > 1.96$, at the voxel-wise and lobular levels for each individual. Figures 2B-C show examples of normative probability maps for

different individuals across clinical diagnostic groups. These maps reveal high cerebellar heterogeneity within and across groups, especially when we look at the overlaps of these individual-level probability maps. The overlap maps calculate the percentage of extreme deviations within the same group for each voxel. Even within the same diagnostic group only a marginal percentage of individuals show extreme deviations in the same cerebellar regions. Moreover, some cerebellar regions (e.g., in posterior lobules) show increased numbers of extreme deviations across several diagnostic groups (e.g., SZ, ASD and AD). Analyses of lobular volumes yielded similar results (see Supplementary Fig. 4B-C).

Norms summed across existing cerebellar atlases aid functional interpretation

Voxel-wise normative models can be projected onto any existing - or future - cerebellar atlas morphed into Montreal Neurological Institute (MNI) space (Fig. 3A). Here, the voxel-wise normative probability maps are summed across 28 lobules (anatomical atlas), 10 functionally defined cerebellar regions (task-based atlas)[28], and 17 regions defined by their resting-state connectivity with functional networks of the cerebral cortex (resting-state atlas)[29,30]. Percentage of deviations for each region is calculated across parcellations for each individual. The figures (Fig. 3B-C and Supplementary Table 4-6) show effect sizes (rank biserial correlation) of percentage of both positive and negative extreme deviations for various clinical cohorts compared to the individuals without diagnosis. As the percentage of negative deviations increases, the strength of the clinical groups also increases. SZ, mild cognitive impairment (MCI) and AD groups showed small to medium effects in the negative deviations compared to the individuals without diagnosis globally across all three atlases. For all three groups, the effects are visibly greater in the posterior regions of

the cerebellum and display consistent patterns across the atlases. However, ASD and MCI groups reveal significant small effects in positive deviations across numerous regions. For ASD, the medium effects in the negative deviations are seen in the region related to action observation and divided attention (region 4) in the task-based atlas and linked to the dorsal attention network region (network 6) in the resting state atlas.

[Figure 3]

Extreme deviations are associated with intelligence

Figure 4 shows correlations obtained from three distinct atlases when examining the relationship between intelligence scores. In ASD, regions associated with Right IV-VI in anatomical atlas and somatomotor B (network 4) in resting state atlas showed positive correlations between percentage of extreme positive deviations per participant and performance IQ (PIQ) (Supplementary Table 7-9) that survived multiple comparisons of number of tests, regions of interest and directions. Weak positive correlations between the percentage of extreme positive deviations per participant and intelligence scores were shown in patients with SZ across three atlases and vice versa for negative deviations (Figure 4B and Supplementary Table 13-15). However, symptom scores Autism Diagnostic Observation Schedule (ADOS) for ASD and Positive and Negative Syndrome Scale (PANSS) for SZ did not show significant associations with percentage of extreme deviations per participant after correcting for multiple comparisons (Supplementary Table 10-12 and 16-18).

[Figure 4]

Discussion

Leveraging brain MRI data from > 54k individuals, we chart cerebellar lifespan trajectories with voxel-wise precision using normative modelling. These models are estimated across 132 scanning sites and contribute to the field by: i) providing > 143k normative models of the cerebellar lobules and voxels across the lifespan, ii) confirming and extending previous reports of altered cerebellar structure in various mental and neurological illnesses, iii) demonstrating considerable structural cerebellar heterogeneity within all clinical groups, iv) coupling normative models with existing cerebellar atlases to enhance interpretability and v) demonstrating functional significance of extreme deviations in terms of associations with measures of cognitive function. Conditions marked by cognitive deficits, like schizophrenia, ASD, and dementia, exhibit cerebellar differences, although not universally across all individuals with such illnesses. The substantial individual differences even within the same diagnostic groups underscore the multifaceted role of the cerebellum in these clinical phenotypes and highlight the need for fine grained analytical procedures at scale.

As prolonged developmental and aging windows may render the cerebellum susceptible to cellular, morphological and circuit abnormalities[28], understanding normal and abnormal development of the cerebellum is a major research priority. The lifespan normative models developed here allow us to combine datasets and perform analyses in reference to a common population cohort, making further structural investigations of the cerebellum more comparable. Our results are generally in line with previous reports on cerebellar aging[29,30] showing a rapid growth of most cerebellar regions during childhood (with volumes typically peaking in late adolescence), followed by a more gradual decline In addition to boosting the sample size relative to these previous studies about 10-fold, we also increase the spatial

precision by analysing lifespan trajectories at voxel-wise spatial resolution. This allows for a more fine-grained understanding of cerebellar structure and more precise delineation of functional units. Importantly, these voxel-wise models can readily be integrated with any existing or future cerebellar atlas, and may thus be sensitive to structural deviations that do not necessarily align well with traditional anatomical borders (i.e., lobules)[31].

Group comparisons using model-derived z-scores confirmed previous reports of altered cerebellar morphology in certain mental and neurological illnesses, while normative probability maps revealed high cerebellar heterogeneity within these same illnesses. First, patients with AD exhibited a significant reduction in lobular measures compared to the individuals without diagnosis while voxel-wise did not survive multiple comparison correction. In line with meta-analysis study on grey matter loss[9], group effects were particularly pronounced in the Right Crus I which have previously been associated with cognitive processing[32]. A recent study[33] reported that cerebellum volume is associated with cognitive decline in individuals with MCI but not in those with AD. Conversely, in MCI, the presence of extreme bidirectional deviations in the cerebellum might suggest a possible compensatory mechanism in some individuals and decline in others during the initial phase of the disease[34], potentially serving as a cognitive reserve[35]. Such “cerebellar reserve”[36], might thus mitigate some of the cognitive decline associated with neurodegeneration through compensatory reorganization. Our findings further suggest that only a proportion of all diagnosed individuals with AD exhibit extreme negative normative deviations in the cerebellum. These findings highlight the heterogeneity in phenotypes and pathophysiology in AD[37] and the importance of looking into the variability of the disease.

In line with a previous mega-analysis[10], we observed significantly lower normative model z-scores (indicating lower volumes) in patients with SZ relative to individuals without a diagnosis. Compared to individuals without diagnosis, some patients with SZ showed evidence of smaller regional volumes in reporting small but reliable reduction in cerebellar volume in SZ, particularly in areas associated with high-level cognitive function. Across atlases, our analyses revealed extreme negative deviations primarily in the posterior regions. Recent meta-analysis study[38] found brain-predicted age difference in SZ by average of +3.55 years compared to the individuals without diagnosis. In this study, findings of SZ stand out as brain age study[39] display more pronounced changes in full brain and cerebellar subcortical in comparison to AD. Moreover, disruptions in cerebello-thalamo-cortical circuit may lead to impairment in synchrony of mental processes, possibly generating symptoms of schizophrenia[40].

The present results are in line with previous findings of substantial heterogeneity in ASD[13,41,42]. We observed medium effects in percentage of negative deviations associated with action observation in task-based atlas and dorsal attention network in resting-state atlas. This is of interest as individuals with ASD frequently report difficulties with social interaction and restricted, repetitive behaviour[43]. D'Mello and colleagues[11] reported that reduced regional and lobular grey matter volumes in right VII (Crus I/II) correlated with the severity of social, communication and repetitive behaviours, based on ADOS scores. While we observe some nominally significant associations with ADOS scores om the current study, these did not survive correction for multiple comparisons.

Previous studies have established an association between intelligence and larger brain size or greater grey matter volume[44] as well as total brain volume[45]. Likewise, cerebellar morphology has been associated with cognitive ability and

psychiatric symptoms[46]. In patients with SZ, weak yet significant positive and negative associations were found with FIQ across the cerebellum, clearly showing that positive deviations are associated with higher scores and negative deviations with lower scores. In ASD, higher performance intelligence scores were associated with positive deviation in the regions related to Right IV-VI in anatomical atlas and somatomotor network (network 4) in task-based atlas.

While our work represents one of the largest investigations into the cerebellar structures and heterogeneity at the level of the individual, and provides a resource for researchers for future investigations, some limitations require consideration. First, we lack coherent and detailed behavioural, cognitive, genetic, phenotypic, and medical history information for both individuals without diagnosis and clinical samples. This is usually the case when combining information across multiple datasets from different projects as documentation and assessment varies which may preclude certain associations from being tested. Additionally, the coverages of lifespan normative models in the very young age and in age range of 30 to 40 were relatively low. Moreover, our data primarily represents western populations, potentially limiting its generalizability to other populations. Due to its location in the posterior fossa, its intricate arrangement, and motions artifacts, the cerebellum has posed challenges in imaging studies and therefore inferior regions of the cerebellum may result in poorer model fits. The cerebellar topography shows substantial individual differences[47], and therefore does not perfectly align with existing “average” atlases. Therefore, there are limitations to the interpretation of functional implications. The resource established here can be used in future studies, to further elucidate the complex aetiology of mental and neurological illnesses. By including comprehensive longitudinal datasets or

316 dividing the heterogeneous clinical population into subtypes, we can aim to enhance
317 biomarker development using cognitive and behavioural measures[21,48–50].
318
319 ***Conclusion***
320 We report the largest multi-site investigation of heterochronic development and aging
321 of the cerebellum at both voxel-wise and lobular spatial precision. Through normative
322 modelling, we observe individualized patterns of deviation across five different mental
323 and neurological illnesses. Several clinical phenotypes exhibited negative deviations
324 at the group level, but with notable individual differences even within the same clinical
325 groups. Overall, this study charts cerebellar morphology across the lifespan, provides
326 evidence for the differential involvement of the cerebellum across brain illness, and
327 links extreme deviations from population norms to cognitive function.
328
329
330
331
332
333
334
335
336
337
338
339
340

Online Methods

Population data

The cohort of individuals without diagnoses was obtained from 132 scanning sites, including various studies such as ABIDE, ADHD200, AOMIC ID1000, Beijing Enhanced, CAMCAN, CoRR, DLBS, DS000119, DS000202, DS000222, Fcon1000, HBN, HCP, MPI Lemon, NKI-Rockland, OASIS-3, PING, SALD, SLIM, and UK Biobank. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Further details about each study can be found in the associated publications. (Supplementary Table 1). The total number of participants in the individuals without diagnosis population was 54102 (53% females), and the clinical set was 1757, encompassing > 56k in total. The age range spanned from 3 to 85 years (Figure 1). Detailed descriptions of each site, including sample size, mean age, standard deviation, and sex ratio, can be found in Supplementary Table 2. If longitudinal scans were available in the studies, only the baseline scans were used. Participants who had withdrawn from the studies or had missing demographic information and T1-weighted MRI data were excluded from the analyses.

Clinical data

As for the clinical datasets, we combined data from ABIDE, ADNI, AIBL, DEMGEN, and TOP (Figure 1). Apart from the requirement of having available clinical diagnoses, clinical groups with more than 100 participants were included in the study. Among the clinical groups, we selected AD (Alzheimer's disease), ASD (autism spectrum disorder), BD (bipolar disorder), MCI (mild cognitive impairment), and SZ (schizophrenia) as the clinical cohorts.

Lobular-level processing

The T1-weighted images were skull stripped using the FreeSurfer 5.3 auto-recon pipeline[51] and reoriented to the standard FSL orientation using the *fslreorient2std*[52]. The linear registration was performed using *flirt*[53], which utilized linear interpolation and the default 1 mm FSL template (version 6.0). The borders were cropped in [6:173, 2:214, 0:160] coordinates to minimize size while retaining the complete volume. Lastly, the voxel intensity values were normalized to the range of [0,1], adjusting the intensity values of each voxel to a standardized scale.

In our study, we utilized the ACAPULCO algorithm[29], which is a state-of-the-art cerebellum parcellation algorithm based on convolutional neural networks. This algorithm is part of the ENIGMA Cerebellum Volumetric Pipeline and provides speedy and accurate quantitative in-vivo regional assessment of the cerebellum at the highest fidelity[54]. We used pre-processed T1-weighted image for better quality control and alignment. The inhomogeneity of the images was corrected using the N4[55] and registered to the 1mm isotropic ICBM 2009c template in MNI space using the ANTs registration suite[56]. The ACAPULCO algorithm was trained using 15 expert manual delineations of an adult cohort[54]. It performs per-voxel labelling of the cerebellum and applies post-processing to remove isolated pieces, ensuring accurate segmentation.

The algorithm divides the cerebellum into 28 cerebellar lobules, including bilateral Lobules I–VI; Crus I and II; Lobules VIIB, VIIIA, VIIIB, and IX-X; Vermis VI, VII, VIII, IX, and X; and Corpus Medullare (CM). It also calculates the volume (mm^3) of each region. The automated quality control generates segmented images for each participant, and we removed the extreme outliers where the number of lobules exceeded a certain threshold (e.g., n > 2).

Voxel-level processing

In our study, we utilized the SUIT (Spatially Unbiased Infratentorial Toolbox) toolbox to perform segmentation of cerebellar grey and white matter voxel-based morphometry (VBM) maps. This segmentation process involved using the outputs from ACAPULCO, which include the N4 bias-corrected and MNI-aligned T1 image[57,58], as well as an averaged mask derived from randomly selected 300 individuals without diagnosis. The ACAPULCO mask plays a crucial role in correcting and refining overinclusion errors in the segmentation process due to variations in the segmentation algorithm. Following segmentation, the grey matter maps were normalized and resliced to align with a standardized space. This normalization step ensures that the data can be compared across different individuals and studies. Additionally, the grey matter maps were modulated by the Jacobian to preserve the value of each voxel in proportion to its original volume. This modulation accounts for individual differences in brain size and helps to retain the relative intensity values within the mapped brain regions. By using the SUIT toolbox and incorporating the ACAPULCO outputs, we were able to obtain accurate and spatially unbiased segmentation of cerebellar grey and white matter, enabling further analysis and comparison of VBM maps within and across individuals.

Normative modelling

We split our individuals without diagnosis sample into training and test sets based on scanning site, sex, and age. This split is important to account for the potential confounding effects of MRI scanners on the data[17,59,60]. The individuals without diagnosis were first stratified based on the scanning sites. This ensures that the training and testing datasets include a representative distribution of participants from each location. To achieve this, we evenly split the control participants from each scanning site between the training and testing sets. A minimum requirement of 5 participants from the same scanners sites was required. However, only the test set consisted of diagnostic groups (e.g., AD, ASD, BD) with minimum of 100 participants. This criterion ensures that there is an adequate number of participants in each diagnostic group to provide reliable statistical analyses. By employing this stratification approach, we aimed to create balanced and representative training and testing datasets that account for MRI scanners and sex, while also including enough participants in each diagnostic group.

We used the PCNtoolkit package (version 0.24)[15,61] in Python 3.8 to estimate a normative model for predicting regional cerebellar volumes and voxel-wise intensity based on sex and age, while correcting for scanning site. Results that deviated more than 5 standard of deviation were imputed by the mean. We employed Bayesian Linear Regression (BLR) with likelihood warping approach[62], specifically using the 'sinarcsinsh' transformation[60,63]. This approach is well-suited for handling non-linear basis functions and non-Gaussian predictive distributions for large datasets as well as correcting for outer centiles. A detailed mathematical description on BLR for normative modelling can be found in the following paper Fraza et al. (2021)[60]. To account for

scanner effects, we treated the scanning site as a fixed effect in our analysis[16,64]. This approach has been shown to yield relatively good performance, as demonstrated in previous work[16]. To assess how each participant's (i) deviate from the individuals without diagnosis pattern at each lobule or voxel (j) in the cerebellum, we calculated the z-score:

$$z_{ij} = \frac{y_{ij} - \widehat{y_{ij}}}{\sqrt{\sigma_{ij}^2 + \sigma_{nj}^2}}$$

The computation of the z-score includes predicted mean $\widehat{y_{ij}}$ (lobule or voxel), true response y_{ij}, predicted variance σ_{ij} and normative variance σ_{nj}. For model fit, the normative model provided point estimates and evaluation metrics, including explained variance, mean squared log-loss, skew, and kurtosis[63]. These evaluation metrics were computed in the test set that did not include any clinical groups. To determine participants with extreme deviations, we set a threshold at z > |1.96|, corresponding to the 95% confidence interval. For instance, deviations with z-scores greater than 1.96 were identified as extreme positive deviations, indicating significantly increased volume compared to the control pattern and vice versa for extreme negative deviations.

Group comparisons

We performed classical nonparametric test on the z-scores of clinical cohorts and individuals without diagnosis. To assess the statistical significance, we performed Mann-Whitney U-tests[65], a non-parametric test that is suitable for comparing two independent samples that are not normally distributed. To account for multiple comparisons, the resulting p-values were corrected, using Bonferroni correction[66] and

465 calculated rank biserial correlation to see its effect.

466

467 ***Atlas-based analyses***

468 Additionally, the normative model can be applied to various research questions and is
469 compatible with existing atlases by using registration methods such as FSL *flirt* and
470 *fnirt* [53,67]. This makes it an attractive versatile tool that can be utilized in different
471 studies and across different brain regions. By mapping the deviations onto specific
472 anatomical regions, such as the 28 cerebellar regions, King's 10 regions of interest
473 from the multi-domain task battery (MDTB)[31], and 17 regions of interest from resting-
474 state connectivity[68,69], we gain insights into the specific areas where deviations occur.
475 We separately calculated the percentage of extreme positive and negative deviations
476 for each participant in the regions of interest in reference to the existing atlases. We
477 divided this by the size of the region and multiplied the resulting proportion by 100 (i.e.,
478 deriving a percentage of extreme positive or negative deviations per region). To
479 compare the extreme deviations observed in different cohorts to the individuals without
480 diagnosis group, we used Mann-Whitney U-tests and calculated the rank biserial
481 correlation (r) for significant results.

482 To investigate potential associations between measured intelligence and
483 symptom scores and clinical cohorts, Spearman correlation analyses were performed
484 using voxel-wise extreme deviation scores that were mapped onto the existing atlases.
485 The Spearman correlation coefficient is used to quantify the strength and direction of
486 the association between the variables, allowing for the examination of potential
487 relationship (see Supplementary Methods for tests used). Only correlations with a
488 corrected p-value below 0.05 ($p < 0.05$) were considered statistically significant and
489 reported.

Data availability

In this study we used brain imaging from ABIDE, ADHD200, AOMIC ID1000, Beijing Enhanced, CAMCAN, CoRR, DLBS, DS000119, DS000202, DS000222, Fcon1000, HBN, HCP, MPI Lemon, NKI-Rockland, OASIS-3, PING, SALD, SLIM and UK Biobank, ADNI, AIBL, DEMGEN, PNC, and TOP. The ROI models from this work are available on via PCNportal[70]: https://pcnportal.dccn.nl/.

Code availability

All code used in this work is publicly available at FreeSurfer (https://surfer.nmr.mgh.harvard.edu), FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation), ACAPULCO (https://gitlab.com/shuohan/acapulco), and SUIT (https://github.com/jdiedrichsen/suit). Code for normative model is available as open-source python package, Predictive Clinical Neuroscience (PCN) toolkit (https://github.com/amarquand/PCNtoolkit). Further codes are available on https://github.com/milinkim/mapping_cerebellar_hetereogeneity.

References

1. Hofman, M. A. Size and shape of the cerebral cortex in mammals. II. The cortical volume. *Brain. Behav. Evol.* **32**, 17–26 (1988).
2. Lent, R., Azevedo, F. A. C., Andrade-Moraes, C. H. & Pinto, A. V. O. How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *Eur. J. Neurosci.* **35**, 1–9 (2012).
3. Kelly, R. M. & Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci. Off. J. Soc. Neurosci.* **23**, 8432–8444 (2003).
4. Barton, R. A. & Venditti, C. Rapid evolution of the cerebellum in humans and other great apes. *Curr. Biol. CB* **24**, 2440–2444 (2014).
5. Leiner, H. C., Leiner, A. L. & Dow, R. S. Does the cerebellum contribute to mental skills? *Behav. Neurosci.* **100**, 443–454 (1986).
6. Schmahmann, J. D. The Role of the Cerebellum in Cognition and Emotion: Personal Reflections Since 1982 on the Dysmetria of Thought Hypothesis, and Its Historical Evolution from Theory to Therapy. *Neuropsychol. Rev.* **20**, 236–260 (2010).
7. Strick, P. L., Dum, R. P. & Fiez, J. A. Cerebellum and nonmotor function. *Annu. Rev. Neurosci.* **32**, 413–434 (2009).
8. Schmahmann, J. D. The cerebellum and cognition. *Neurosci. Lett.* **688**, 62–75 (2019).
9. Gellersen, H. M., Guell, X. & Sami, S. Differential vulnerability of the cerebellum in healthy ageing and Alzheimer's disease. *NeuroImage Clin.* **30**, 102605 (2021).
10. Moberget, T. *et al.* Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Mol. Psychiatry* **23**, 1512–1520 (2018).

538 11. D'Mello, A. M., Crocetti, D., Mostofsky, S. H. & Stoodley, C. J. Cerebellar gray matter
539 and lobular volumes correlate with core autism symptoms. *NeuroImage Clin.* **7**, 631–639
540 (2015).
541 12. Wang, S. S.-H., Kloth, A. D. & Badura, A. The Cerebellum, Sensitive Periods, and
542 Autism. *Neuron* **83**, 518–532 (2014).
543 13. Traut, N. *et al.* Cerebellar Volume in Autism: Literature Meta-analysis and Analysis of
544 the Autism Brain Imaging Data Exchange Cohort. *Biol. Psychiatry* **83**, 579–588 (2018).
545 14. Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B. & Marquand, A. F. From
546 estimating activation locality to predicting disorder: A review of pattern recognition for
547 neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* **57**, 328–349
548 (2015).
549 15. Marquand, A. F. *et al.* Conceptualizing mental disorders as deviations from normative
550 functioning. *Mol. Psychiatry* **24**, 1415–1424 (2019).
551 16. Kia, S. M. *et al.* Federated Multi-Site Normative Modeling using Hierarchical Bayesian
552 Regression. 2021.05.28.446120 Preprint at https://doi.org/10.1101/2021.05.28.446120
553 (2021).
554 17. Rutherford, S. *et al.* Charting brain growth and aging at high spatial precision. *eLife* **11**,
555 e72904 (2022).
556 18. Wolfers, T. *et al.* Replicating extensive brain structural heterogeneity in individuals with
557 schizophrenia and bipolar disorder. *Hum. Brain Mapp.* **42**, 2546–2555 (2021).
558 19. Floris, D. L. *et al. Atypical brain asymmetry in autism – a candidate for clinically*
559 *meaningful stratification*. http://biorxiv.org/lookup/doi/10.1101/2020.03.24.000349
560 (2020) doi:10.1101/2020.03.24.000349.

20. Zabihi, M. *et al.* Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 567–578 (2018).

21. Zabihi, M. *et al.* Fractionating autism based on neuroanatomical normative modeling. *Transl. Psychiatry* **10**, 384 (2020).

22. Wolfers, T. *et al.* Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol. Med.* **50**, 314–323 (2020).

23. Verdi, S., Marquand, A. F., Schott, J. M. & Cole, J. H. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. *Brain* **144**, 2946–2953 (2021).

24. Verdi, S. *et al.* Revealing Individual Neuroanatomical Heterogeneity in Alzheimer Disease Using Neuroanatomical Normative Modeling. *Neurology* (2023) doi:10.1212/WNL.0000000000207298.

25. Remiszewski, N. *et al.* Contrasting Case-Control and Normative Reference Approaches to Capture Clinically Relevant Structural Brain Abnormalities in Patients With First-Episode Psychosis Who Are Antipsychotic Naive. *JAMA Psychiatry* **79**, 1133–1138 (2022).

26. Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* **75**, 1146–1155 (2018).

27. Segal, A. *et al.* Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. *Nat. Neurosci.* 1–17 (2023) doi:10.1038/s41593-023-01404-6.

28. Sathyanesan, A. *et al.* Emerging connections between cerebellar development, behaviour and complex brain disorders. *Nat. Rev. Neurosci.* **20**, 298–313 (2019).

29. Han, S., An, Y., Carass, A., Prince, J. L. & Resnick, S. M. Longitudinal analysis of regional cerebellum volumes during normal aging. *NeuroImage* **220**, 117062 (2020).

30. Romero, J. E. *et al.* Toward a unified analysis of cerebellum maturation and aging across the entire lifespan: A MRI analysis. *Hum. Brain Mapp.* **42**, 1287–1303 (2021).

31. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat. Neurosci.* **22**, 1371–1378 (2019).

32. Guo, C. C. *et al.* Network-selective vulnerability of the human cerebellum to Alzheimer's disease and frontotemporal dementia. *Brain* **139**, 1527–1538 (2016).

33. Lin, C.-Y., Chen, C.-H., Tom, S. E., Kuo, S.-H., & for the Alzheimer's Disease Neuroimaging Initiative. Cerebellar Volume Is Associated with Cognitive Decline in Mild Cognitive Impairment: Results from ADNI. *The Cerebellum* **19**, 217–225 (2020).

34. Jacobs, H. I. L. *et al.* The cerebellum in Alzheimer's disease: evaluating its role in cognitive decline. *Brain J. Neurol.* **141**, 37–47 (2018).

35. Stern, Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol.* **11**, 1006–1012 (2012).

36. Mitoma, H. *et al.* Consensus Paper. Cerebellar Reserve: From Cerebellar Physiology to Cerebellar Disorders. *The Cerebellum* **19**, 131–153 (2020).

37. Graff-Radford, J. *et al.* New insights into atypical Alzheimer's disease in the era of biomarkers. *Lancet Neurol.* **20**, 222–234 (2021).

38. Constantinides, C. *et al.* Brain ageing in schizophrenia: evidence from 26 international cohorts via the ENIGMA Schizophrenia consortium. *Mol. Psychiatry* **28**, 1201–1209 (2023).

39. Kaufmann, T. *et al.* Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617–1623 (2019).

40. Andreasen, N. C. & Pierson, R. The role of the cerebellum in schizophrenia. *Biol. Psychiatry* **64**, 81–88 (2008).

611 41. Laidi, C. *et al.* Cerebellar Atypicalities in Autism? *Biol. Psychiatry* **92**, 674–682 (2022).
612 42. Lombardo, M. V., Lai, M.-C. & Baron-Cohen, S. Big data approaches to decomposing
613 heterogeneity across the autism spectrum. *Mol. Psychiatry* **24**, 1435–1450 (2019).
614 43. Lai, M.-C., Lombardo, M. V. & Baron-Cohen, S. Autism. *Lancet Lond. Engl.* **383**, 896–
615 910 (2014).
616 44. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence
617 differences. *Nat. Rev. Neurosci.* **11**, 201–211 (2010).
618 45. Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M. & Deary, I. J. Structural
619 brain imaging correlates of general intelligence in UK Biobank. *Intelligence* **76**, 101376
620 (2019).
621 46. Moberget, T. *et al.* Cerebellar Gray Matter Volume Is Associated With Cognitive
622 Function and Psychopathology in Adolescence. *Biol. Psychiatry* **86**, 65–75 (2019).
623 47. Xue, A. *et al.* The detailed organization of the human cerebellum estimated by intrinsic
624 functional connectivity within the individual. *J. Neurophysiol.* **125**, 358–384 (2021).
625 48. Bethlehem, R. A. I. *et al.* A normative modelling approach reveals age-atypical cortical
626 thickness in a subgroup of males with autism spectrum disorder. *Commun. Biol.* **3**, 1–10
627 (2020).
628 49. Hong, S.-J. *et al.* Toward Neurosubtypes in Autism. *Biol. Psychiatry* **88**, 111–128 (2020).
629 50. Shan, X. *et al.* Mapping the Heterogeneous Brain Structural Phenotype of Autism
630 Spectrum Disorder Using the Normative Model. *Biol. Psychiatry* **91**, 967–976 (2022).
631 51. Ségonne, F. *et al.* A hybrid approach to the skull stripping problem in MRI. *NeuroImage*
632 **22**, 1060–1075 (2004).
633 52. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL.
634 *NeuroImage* **62**, 782–790 (2012).

635 53. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of
636 brain images. *Med. Image Anal.* **5**, 143–156 (2001).
637 54. Carass, A. *et al.* Comparing fully automated state-of-the-art cerebellum parcellation from
638 magnetic resonance images. *NeuroImage* **183**, 150–172 (2018).
639 55. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging*
640 **29**, 1310–1320 (2010).
641 56. Fonov, V. *et al.* Unbiased average age-appropriate atlases for pediatric studies.
642 *NeuroImage* **54**, 313–327 (2011).
643 57. Diedrichsen, J. A spatially unbiased atlas template of the human cerebellum. *NeuroImage*
644 **33**, 127–138 (2006).
645 58. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR
646 atlas of the human cerebellum. *NeuroImage* **46**, 39–46 (2009).
647 59. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533
648 (2022).
649 60. Fraza, C., Dinga, R., Beckmann, C. F., Marquand, A. F., & Andre Marquand. Warped
650 Bayesian Linear Regression for Normative Modelling of Big Data. *bioRxiv* (2021)
651 doi:10.1101/2021.04.05.438429.
652 61. Rutherford, S. *et al.* The normative modeling framework for computational psychiatry.
653 *Nat. Protoc.* **17**, 1711–1734 (2022).
654 62. Rios, G. & Tobar, F. Compositionally-warped Gaussian processes. *Neural Netw.* **118**,
655 235–246 (2019).
656 63. Dinga, R. *et al.* Normative modeling of neuroimaging data using generalized additive
657 models of location scale and shape. 2021.06.14.448106 Preprint at
658 https://doi.org/10.1101/2021.06.14.448106 (2021).

64. Bayer, J. M. M. *et al.* Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models. 2021.02.09.430363 Preprint at https://doi.org/10.1101/2021.02.09.430363 (2021).

65. Mann, H. B. & Whitney, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.* **18**, 50–60 (1947).

66. Bonferroni, C. E. *Teoria statistica delle classi e calcolo delle probabilità*. (Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936).

67. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* **17**, 825–841 (2002).

68. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).

69. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

70. Barkema, P. *et al.* Predictive Clinical Neuroscience Portal (PCNportal): instant online access to research-grade normative models for clinical neuroscientists. *Wellcome Open Res.* **8**, 326 (2023).

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Also, data used in preparation of this article were obtained from the Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL) databases (adni.loni.usc.edu), and the Pediatric Imaging, Neurocognition and Genetics (PING) study database (chd.ucsd.edu/research/ping-

709 study.html, now shared through the NIMH Data Archive (NDA). This publication is
710 solely the responsibility of the authors and does not necessarily represent the views
711 of the National Institutes of Health or PING investigators. The authors of this
712 manuscript gratefully acknowledge the following funding bodies. TW acknowledges
713 funding by the German Research Foundation (DFG; Projectnumber: 513851350) as
714 well as starting funding from the faculty of medicine at the University of Tübingen.
715
716 The European Research Council under the European Union's Horizon 2020 research
717 and Innovation program (ERC StG, Grant 802998), the Research Council of Norway
718 (300767, 324499), the South-Eastern Norway Regional Health Authority (2019101).
719 We performed this work on the *Services for sensitive data* (TSD), University of Oslo,
720 Norway, with resources provided by UNINETT Sigma2 - the National Infrastructure for
721 High-Performance Computing and Data Storage in Norway. We want to acknowledge
722 the Norwegian registry of persons assessed for cognitive symptoms (NorCog), for
723 providing access to patient data. We conducted this research using the UK Biobank
724 Resource under Application Number 27412.
725

726 ***Authorship Contributions***

727 T.M., T.W., and M.K. originally conceived of the project. M.K., T.W., T.M., and E.L.
728 performed the analyses. M.K. wrote the initial draft of the manuscript. O.A., G.R.,
729 K.P., G.S., N.E.S., O.B.S., A.F.M., C.F.B., T.U., and L.W. contributed to data
730 curation. All authors discussed the results and contributed to the final manuscript.
731

732 ***Ethics declarations***

733 The authors declare no competing interests.

734 **Table 1.** Sample description and demographics

		N (Participants	N (Scanners)	Age (Mean, S.D.)	Sex (%F:%M)
Full	All	54102	132		
	Training set	27117	132	54.36 (20.31)	0.53:0:47
	Testing set	26985	132	54.52 (20.19)	0.53:0:47
Clinical	Testing set	1757	53	29.40 (20.84)	0.30 0.70
	Alzheimer's Disease	146	13	72.42 (7.65)	0.53:0:47
	ASD	900	37	16.20 (9.00)	0.14:0:86
	Bipolar Disorder	277	3	32.73 (11.67)	0.60:0:40
	Mild Cognitive Impairment	122	3	67.25 (9.27)	0.42:0:58
	Schizophrenia	312	3	29.58 (9.52)	0.33:0:67

735

736

737

738

739

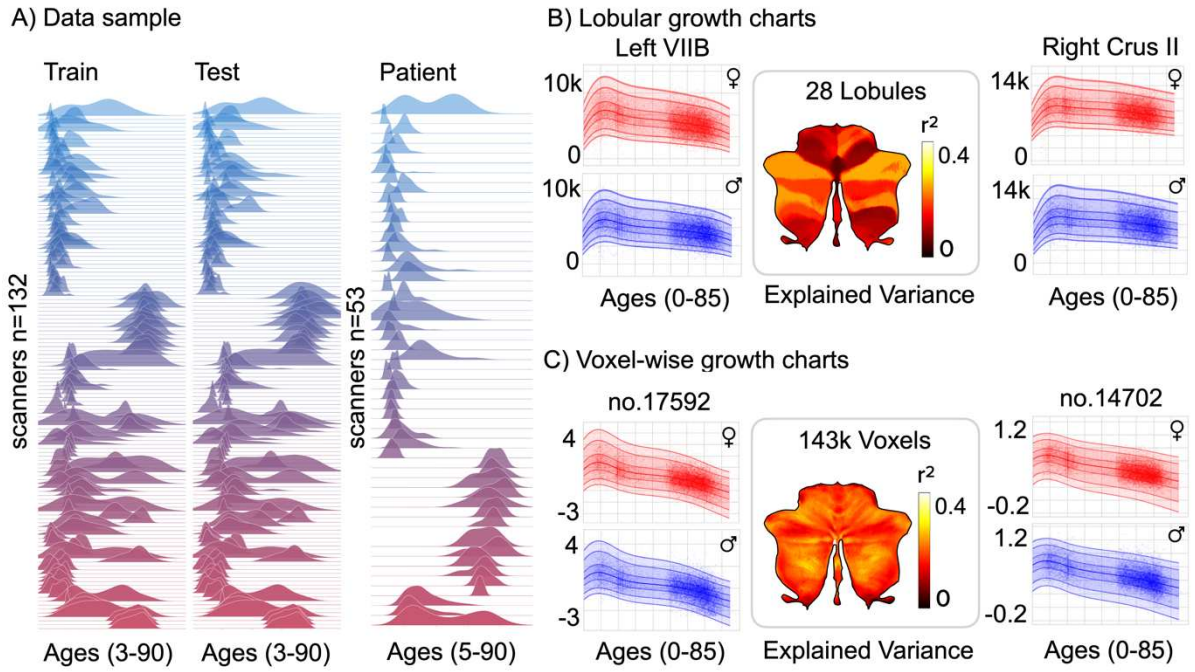


Figure 1. Normative models based on MRI data from > 54k participants describe the lifespan trajectories of cerebellar lobules and individual voxels. In panel (A), the age density distribution is displayed for each scanning sites in the training, test, and clinical sets. Panel (B-C) showcase two of the 28 regions representing the lobular growth charts and two of the 143k voxel-wise growth charts for each sex. The x-axis represents age, ranging from 3 to 85, while the y-axis represents the predicted cerebellar volume and grey matter probability values. Additionally, the figure includes the explained variance, indicating the goodness of fit.

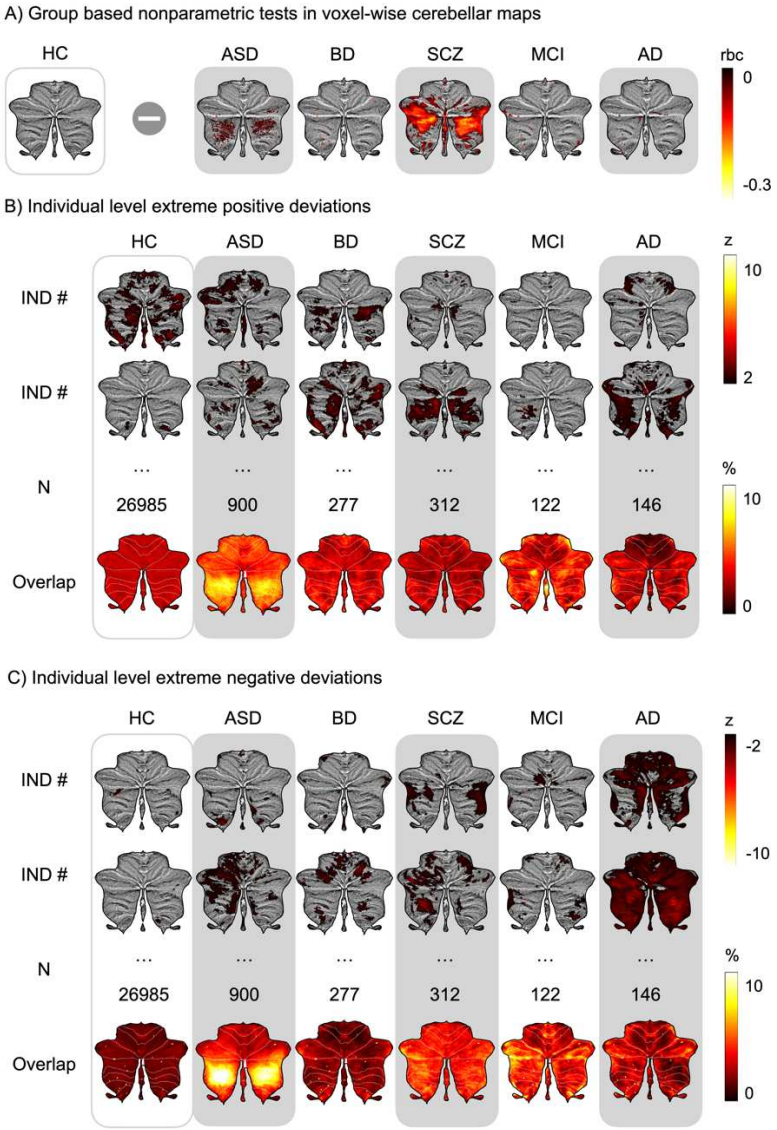


Figure 2. The voxel-wise deviations from estimated norms show high levels of heterogeneity within diagnostic groups (A) depict Mann-Whitney U-test rank biserial correlation (rbc) of z-scores between the clinical and the individuals without diagnosis in voxel-wise. The effects are corrected for multiple comparisons using Bonferroni correction (corrected $p < 0.05$). Specifically, patients diagnosed with schizophrenia (SZ) and autism spectrum disorder (ASD) exhibited significant effects compared to individuals without diagnosis (HC) in the voxel-wise maps. (B-C) The z-scores of extreme positive and negative deviations ($|z| > 1.96$) are shown for two individuals per cohort. Overlap maps calculate percentage of extreme deviations occurred in the same group. The clinical groups displayed a significantly higher occurrence of percentage of extreme deviations, even in cases where the group based nonparametric tests did not differ significantly. These results indicate that within the clinical groups, there were individuals who exhibited significant deviations from the normative patterns, regardless of the overall group effects.

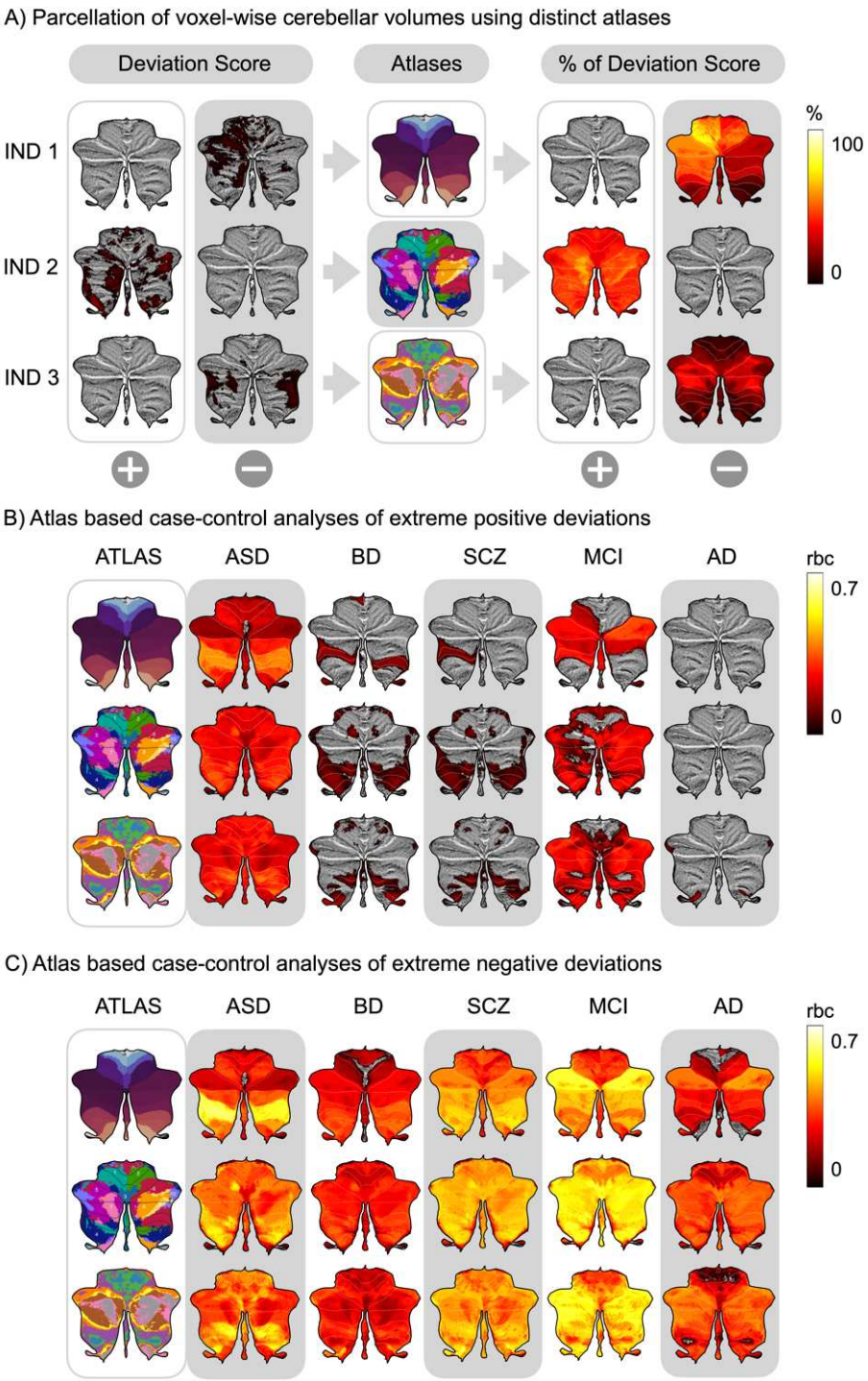
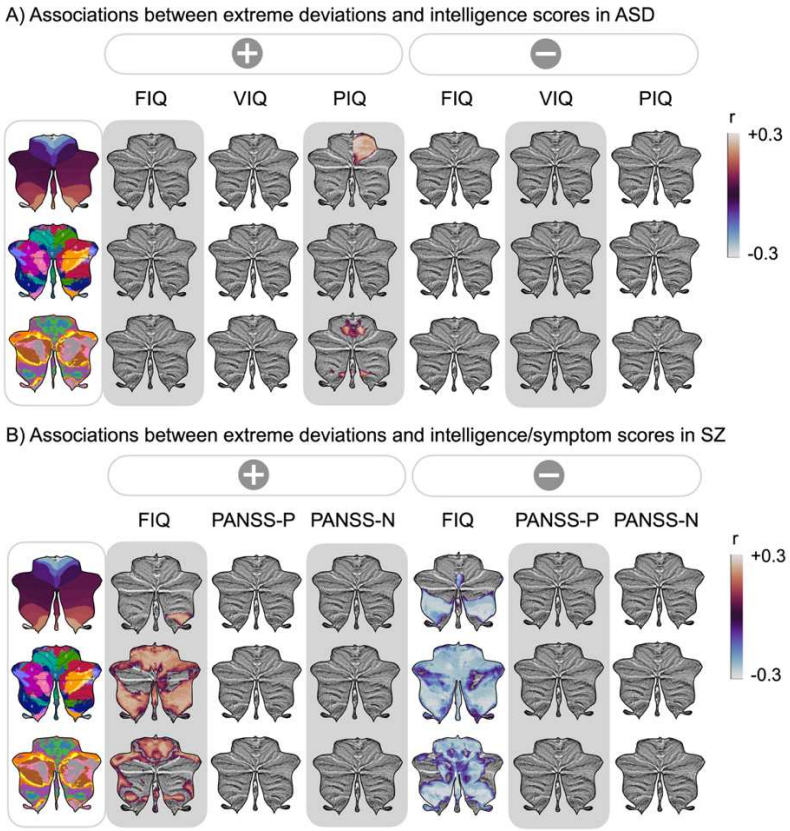


Figure 3. Voxel-wise normative models can be applied to existing or future cerebellar atlases. (A) The outputs from normative model, 143k features normative probability maps of an individual, are applied onto existing atlases of traditional anatomical regions, task-based regions, and resting state connectivity atlases. Panel (B-C) depict effect size of comparison of percentage of extreme positive and negative deviations of clinical cohorts to the individuals without diagnosis in voxel-wise per participant based on the three atlases. Scale indicates Mann-Whitney U-test rank biserial correlation (rbc) and only shows significant regions after multiple comparison corrections.



775

776 **Figure 4. When applied to different atlases, significant correlations were observed between the percentage**
777 **of extreme deviations per participant and IQ scores.** The panels (A-B) show significant correlations between
778 extreme positive (left) or negative (right) deviations per participant and intelligence or symptom scores mapped
779 onto three atlases. Panel (A) displays significant correlations between performance intelligence scores (PIQ) and
780 the percentage of extreme positive deviations per participant in autism spectrum disorder (ASD). (B) In
781 schizophrenia (SZ), positive associations are shown in percentage of extreme positive deviation while negative
782 associations in percentage of extreme negative deviations with full-scale IQ (FIQ).

783

784