Subgrouping autism and ADHD based on structural MRI population modelling centiles

Pecci-Terroba, C. ORCID: https://orcid.org/0000-0002-6379-582X, Lai, M.-C. ORCID: https://orcid.org/0000-0002-9593-5508, Lombardo, M. V. ORCID: https://orcid.org/0000-0001-6780-8619, Chakrabarti, B. ORCID: https://orcid.org/0000-0002-6649-7895, Ruigrok, A. N. V. ORCID: https://orcid.org/0000-0001-7711-8056, Suckling, J. ORCID: https://orcid.org/0000-0002-5098-1527, Anagnostou, E. ORCID: https://orcid.org/0000-0002-3455-9887, Lerch, J. P. ORCID: https://orcid.org/0000-0001-6164-2881, Taylor, M. J. ORCID: https://orcid.org/0000-0002-3534-9750, Nicolson, R. ORCID: https://orcid.org/0000-0001-7086-7038, Georgiades, S., Crosbie, J. ORCID: https://orcid.org/0000-0002-8710-3322, Schachar, R. ORCID: https://orcid.org/0000-0002-2015-4395, Kelley, E. ORCID: https://orcid.org/0000-0001-7742-6542, Jones, J. ORCID: https://orcid.org/0000-0002-1116-4321, Arnold, P. D. ORCID: https://orcid.org/0000-0003-2496-4624, Seidlitz, J. ORCID: https://orcid.org/0000-0002-8164-7476, Alexander-Bloch, A. F. ORCID: https://orcid.org/0000-0001-6554-1893, Bullmore, E. T. ORCID: https://orcid.org/0000-0002-8955-8283, Baron-Cohen, S. ORCID: https://orcid.org/0000-0001-9217-2544, Bedford, S. A. ORCID:

Molecular Autism

RESEARCH

Open Access

Subgrouping autism and ADHD based on structural MRI population modelling centiles

Clara Pecci-Terroba[1], Meng-Chuan Lai[2,3,4,5,6], Michael V. Lombardo[7], Bhismadev Chakrabarti[8], Amber N. V. Ruigrok[2,9], John Suckling[10], Evdokia Anagnostou[11,12], Jason P. Lerch[13,14,15], Margot J. Taylor[13,16], Rob Nicolson[17], Stelios Georgiades[18], Jennifer Crosbie[4,5,13,19], Russell Schachar[4,5,13,19], Elizabeth Kelley[20,21,22], Jessica Jones[21,22], Paul D. Arnold[23,24], Jakob Seidlitz[25,26,27], Aaron F. Alexander-Bloch[25,26,27], Edward T. Bullmore[10], Simon Baron-Cohen[2,28], Saashi A. Bedford[1,2,10*] and Richard A. I. Bethlehem[1,2*]

Abstract

Background Autism and attention deficit hyperactivity disorder (ADHD) are two highly heterogeneous neurodevelopmental conditions with variable underlying neurobiology. Imaging studies have yielded varied results, and it is now clear that there is unlikely to be one characteristic neuroanatomical profile of either condition. Parsing this heterogeneity could allow us to identify more homogeneous subgroups, either within or across conditions, which may be more clinically informative. This has been a pivotal goal for neurodevelopmental research using both clinical and neuroanatomical features, though results thus far have again been inconsistent with regards to the number and characteristics of subgroups.

Methods Here, we use population modelling to cluster a multi-site dataset based on global and regional centile scores of cortical thickness, surface area and grey matter volume. We use HYDRA, a novel semi-supervised machine learning algorithm which clusters based on differences to controls and compare its performance to a traditional clustering approach.

Results We identified distinct subgroups within autism and ADHD, as well as across diagnosis, often with opposite neuroanatomical alterations relatively to controls. These subgroups were characterised by different combinations of increased or decreased patterns of morphometrics. We did not find significant clinical differences across subgroups.

Limitations Crucially, however, the number of subgroups and their membership differed vastly depending on chosen features and the algorithm used, highlighting the impact and importance of careful method selection.

Conclusions We highlight the importance of examining heterogeneity in autism and ADHD and demonstrate that population modelling is a useful tool to study subgrouping in autism and ADHD. We identified subgroups with distinct patterns of alterations relative to controls but note that these results rely heavily on the algorithm used and encourage detailed reporting of methods and features used in future studies.

Keywords Autism, ADHD, Population modelling, Subgrouping, Neuroimaging, Structural MRI

*Correspondence:
Saashi A. Bedford
ajb349@cam.ac.uk
Richard A. I. Bethlehem
rb643@cam.ac.uk
Full list of author information is available at the end of the article

Background

Autism and attention deficit hyperactivity disorder (ADHD) are heterogeneous neurodevelopmental conditions whose biology and developmental patterning are not yet fully understood. The highly variable nature and presentation of these conditions make it more challenging to identify biomarkers that could enhance our understanding and aid in their diagnosis, which currently depends solely on behavioural assessment. So far, case–control approaches have been the main technique used to study these conditions; however, they overlook phenotypic and likely, neurobiological diversity in such heterogeneous disorders [1–3] by combining all individuals into a single group, thereby masking potential differences behind the group average. This can lead to inconclusive or contradictory findings. Identifying clear subgroups with distinct biological differences could significantly improve our understanding of the development of these conditions [4], as looking only at on-average case–control comparisons risks obscuring heterogeneity within the population. This would particularly affect subgroups that exhibit opposing effects relative to the comparison, such as increased or decreased total brain volume in autism. These opposing effects could cancel each other out when analysing the entire population, leading to inconclusive or null findings.

Many attempts to parse this heterogeneity have, for example, focused on delineating the neurobiology of autism by examining differences in total brain volume, in particular in early childhood. Early neuroimaging studies appeared to provide robust support for early brain overgrowth in very young autistic children [5–7] though with some contradictory findings [8], and varied results as to whether this overgrowth continued into late childhood or normalised with age [5, 9]. Crucially, the vast majority of these studies were conducted on cross sectional datasets of varying ages, and more recent studies have indicated that these findings and discrepancies are likely underpinned by the existence of distinctly defined subgroups with distinct neuroanatomical and clinical profiles [10]. Indeed, there is now strong evidence that this early brain overgrowth is observed only in a subset of young autistic (mostly male) children, with longitudinal studies suggesting that it does persist into later childhood [10–12], and that a smaller subset also presents with microcephaly [13]. Importantly, the brain overgrowth observed in a subset of autistic children does not appear to be uniform but may be driven by distinct profiles of enlarged cortical surface area [14] and gyrification [15], highlighting the importance of investigating multiple neuroanatomical features. There is a similar disparity across studies, albeit less well reproduced, in the study of ADHD, with some studies finding lower cortical thickness relative to controls [16], whilst others find no or very small differences [17], and some even report greater thickness, [18, 19]. Although these conditions are known to be associated with alterations in neurodevelopment, there is no canonical neuroanatomical profile that clearly or uniformly captures them nor have clear subgroups been defined based on diverging neuroanatomical trajectories.

Previous studies have also used clustering approaches based on behaviour and clinical features, [20–22]. For example, a recent study that clustered autistic individuals using Autism Diagnostic Observation Scores Generic (ADOS-G) identified two subgroups with distinct restrictive and repetitive behaviours which showed more homogeneous neuroanatomical differences relative to controls [23]. Other studies indicate the existence of brain-based subgroups [24–28], which may span diagnostic categories [29], have distinct clinical characteristics [25, 30], and be associated with different clinical outcomes later in life [10]. Across this literature, results are heterogenous, and subgroups rarely align between studies. Novel clustering algorithms, which make fewer assumptions about the identifiability of linear boundaries, hold promise for better capturing the heterogeneity in developmental conditions. One such example is HYDRA (HeterogeneitY through DiscRiminative Analysis) [31], a semi-supervised machine learning algorithm that clusters individuals based on differences relative to a control sample. Previous research using HYDRA on structural MRI measurements of cortical thickness and surface area in an autism and ADHD cohort yielded two subgroups with distinct neuroanatomical profiles relative to controls, but with no clinical difference across subgroups [29].

Population modelling, also known as normative modelling, is a recently popularised statistical technique to quantify individual variation in relation to population norms, implicitly embracing individual heterogeneity. This individual variation is quantified in the form of centile or "deviation" scores, analogous to those used in conventional paediatric growth charts [32]. Crucially, population modelling allows us to examine individual subjects (i.e., as opposed to group averages) from different sites and/or with different diagnoses, in a common space, and relative to a common reference group. Our group previously established population models to characterise trajectories of structural brain changes throughout the lifespan ("Brain Charts") and explored their use to study brain development and ageing [19, 33]. Specifically in the context of neurodevelopmental conditions, it has been used to examine how much a given individual diverges from a typical trajectory. In the context of autism and ADHD, it has allowed researchers to explore the heterogeneous anatomy of individuals with these conditions [34], as well as the modulatory effects of age and sex [19].

Population modelling has already been implemented to assist with the identification of subgroups; for example, one previous study used measures of cortical thickness in a dataset of 316 autistic participants to identify 5 separate clusters with distinct neuroanatomical profiles and associations with different clinical presentations [28]. Another study used population modelling to cluster autism based on grey matter volume measurements and identified two subgroups with increased volume compared to controls, and one with decreased volume [35]. Sample sizes in previous studies have been relatively small, and no studies thus far have used population modelling to cluster autism and ADHD transdiagnostically.

Here we sought to combine population modelling with a novel clustering algorithm to explore the existence of subgroups in a cross-diagnostic developmental context using a large, aggregated dataset of individuals with and without neurodevelopmental conditions [19]. We first applied HYDRA to the dataset to examine subgroups based explicitly on differences within the clinical cohort relative to controls, whilst controlling for the innate heterogeneity of a large multi-site, transdiagnostic dataset that spans 2–60 years of age. We expanded the work carried out by Itahashi et al. [29] by applying HYDRA to a larger cohort of autistic and ADHD participants, including both male and female, from multiple sites. In addition, we used centile scores instead of raw structural values with the aim of leveraging their ability to capture an individual's specific structural variation with respect to same-sex and same-age controls. Furthermore, we clustered individuals based on the different measurements collectively, instead of separately, to better understand the relationship between these features. We then compared this novel method to an alternative and more commonly used data-driven clustering approach, k-medoids [36], combined with the dimensionality reduction technique Uniform Manifold Approximation and Projection (UMAP) [37], to examine the consistency of results. Additionally, we explored subgroups both within diagnosis (autism and ADHD separately), and across diagnoses (both groups combined). Given the high heterogeneity of both conditions, as well as phenotypic and genetic overlap between them, this allowed us to examine the potential existence of transdiagnostic subgroups which may improve our understanding of the similarities and differences between these conditions.

Methods

Dataset

T1-weighted MRI scans were obtained from 48 different sites across 7 datasets as previously described in [19]. Quality control was conducted using FreeSurfer Quality Control (FSQC) [38]; based on previous work by our group, where any participants with a FSQC > 2.5 were removed. Our sample included autistic individuals, individuals with ADHD, and controls. Here, we use 'controls' to refer to individuals in our sample with no known neurodevelopmental, psychiatric, or physical diagnosis. While there were some participants who had a dual diagnosis of autism and ADHD, we did not have information of secondary diagnosis for all participants, thus individuals were included in the group of their primary diagnosis. The final dataset comprised 4115 participants (1823 controls [1151 male, 672 female], 987 individuals with ADHD [717 male, 270 female], and 1305 autistic individuals [1035 male, 270 female]). A summary of the participant demographic per site can be found in Table S1. Not all sites contained individuals across the three groups (autism, ADHD and control). Because the age range of the ADHD group was much narrower than the autism and control groups, in analyses on ADHD only, only controls within the same age range were used.

Data processing

All scans were processed using FreeSurfer version 6.0.1 [39] and parcellations were extracted using the Desikan-Killiany atlas [40], as the Brain Charts [33] model used to derive the centiles is only available for this parcellation. Estimates for each cortical region were averaged across hemispheres due to the availability of Brain Chart models. Analyses were conducted on centile scores derived using Generalised Additive Models of Location Scale and Shape (GAMLSS), based on the Brain Chart models previously generated and published by our group [33]. These models were generated and extensively validated based on over 75,000 typically developing individuals [33], and out-of-sample centile scores for each participant were then generated for the current sample based on these reference models. In the original paper, a leave-one-study-out analysis demonstrated high reliability and stable and unbiased estimates for out-of-sample centile scores (supplementary methods section S1.1). Additional details can be found in supplementary methods Sect. 1.8 of [33] and supplementary methods Sect. "Methods" of [19].

Given the additional variability in the clinical data, and the small size of some sites, the data was first harmonised using ComBat, a common site harmonisation tool [41], as described in [19, 42], before generating centile scores with GAMLSS. For each individual, we obtained centile scores for six global measurements: total grey matter volume (GMV), subcortical grey matter volume (sGMV), white matter volume (WMV), surface area (SA), mean cortical thickness (CT), and ventricular volume, as well as regional cortical measures of thickness, volume and surface area (34 regions per measure, based on the Desikan-Killiany parcellations).

Measures of autistic and ADHD traits were used for comparison between subgroups. This clinical data was only available for a subset of participants in our dataset, and included the Autism Diagnostic Observation Schedule Calibrated Severity Score (ADOS CSS) (764 participants: 13 ADHD, 732 autism, 19 controls), the Repetitive Behaviour Scale—Revised (RBS—R) (1492 participants: 456 ADHD, 594 autism, 442 control), the Social Responsiveness Scale (SRS) (1676 participants: 372 ADHD, 602 autism, 702 control), and the Strengths and Weaknesses of ADHD Symptoms and Normal Behaviour Inattentiveness subscale (SWAN I) (801 participants: 326 ADHD, 286 autism, 189 control), and hyperactivity subscale (SWAN HI) (799 participants: 325 ADHD, 285 autism, 189 controls).

Clustering

We performed clustering analyses on both global and regional cortical measures to identify potential subgroups of autistic and ADHD participants with similar neuroanatomical profiles, based on normalised centile scores. Given previous findings [19] showing that variations in cortical thickness related to autism are localised to the superior temporal cortex, whereas ADHD individuals show more global level variations in cortical thickness, surface area and volume, we decided to apply clustering at both the regional and global level, using a data-driven approach to identify regions and features that may be particularly relevant for identifying subgroups. We used two different algorithms, as outlined below, to explore different methods of potential subgroup identification, and the extent to which this impacts the resulting subgroups. Clustering was performed both within each diagnosis (autism or ADHD only), and across diagnosis (in a combined dataset including autistic and ADHD individuals together), to examine the extent to which subgroups fall along diagnostic boundaries. For each analysis, clustering was first carried out using only the global features and then using regional measures. To determine the regional effect, we corrected regional centiles for total tissue centile scores by means of subtraction, to aid interpretation and following from the normal distribution of the original centiles. This was done by subtracting the corresponding global centile score from each regional value: i.e. mean cortical thickness (CT), total surface area (SA), and total grey matter volume (GMV) were subtracted from regional cortical thickness, surface area, and grey matter volume values, respectively.

HYDRA

To examine subgroups of a given condition relative to a reference control group, we used the HYDRA clustering algorithm, a novel non-linear, semi-supervised machine learning algorithm [31]. Based on the assumption that there is no single pattern that distinguishes a given condition, HYDRA uses multiple hyperplanes to first separate the case and control groups using supervised learning, based on the input imaging features, and then cluster the diagnosed individuals based on their differences relative to controls. By using the control group as a reference, HYDRA avoids identifying clusters that reflect variability due to noise or demographic factors such as age or sex, rather than actual heterogeneity related to the condition in question [31]. The similarity between the different cluster solutions is assessed using the Adjusted Rand Index (ARI), which is computed through cross-validation, and the solution with the highest ARI represents the optimal number of clusters. The ARI score is a measure of how similar two clustering solutions are. It can take any value from -1 to 1 where 0 means that agreement between solutions is what we would expect by chance, 1 means they perfectly overlap, and negative values indicate that the agreement between solutions is less than the expected by chance. Although the centiles are already derived from comparisons against the controls, clustering using them remains useful. The controls still span the entire centile range, allowing algorithms like HYDRA to identify differences between them and the disorder cohort beyond inherent population variations. HYDRA has been validated using both imaging and genetic data, in comparison to other supervised and unsupervised machine learning techniques, and on both simulated and clinical data. For more details on HYDRA methods, see [31]. We implemented HYDRA on MATLAB, following instructions and parameters found here https://github.com/evarol/HYDRA. Specifically, we implemented HYDRA with 20 clustering consensus steps, and 50 iterations between estimating hyperplanes and cluster estimation.

Here, we ran three-fold cross-validation and evaluated HYDRA from $k=2$ to $k=5$ clusters in line with previously reported ranges of clustering solutions in literature [25, 26, 28, 29]. Given that HYDRA is a computationally expensive algorithm, especially when applied to large datasets, we initially tested clustering solutions up to $k=10$ on the global autism dataset. Since the performance of clusters 6–10 did not surpass that of clusters 2–5, we decided to run all remaining datasets with k ranging from 2 to 5 (Figure S1).

K—medoids and UMAP

Next, in order to compare and contrast HYDRA-generated subgroups with a relatively simpler and more widely used technique, we used k-medoids clustering combined with the dimensionality reduction technique Uniform Manifold Approximation (UMAP) [37] to cluster the

data, again both within diagnosis and transdiagnostically. Dimensionality reduction transforms high dimensional data into low dimensional space while retaining meaningful properties of the original data but allowing for easier interpretation and downstream analysis or clustering. UMAP is a nonlinear dimensionality reduction technique, similar to the commonly used t-distributed Stochastic Neighbour Embedding (tSNE) [43], but faster and with better preservation of the global data structure [37].

K-medoids clustering was then performed using the *pamk* (partitioning around medoids) package in R, using the *pam* algorithm with no scaling and Euclidean distance [44]. K-medoids clustering attempts to divide the data into groups, by designating a centre point (medoid) in each cluster, and minimising the distance between each point and the medoid. We selected the solution with the highest silhouette score as the optimal number of clusters. The silhouette score measures how well-separated clusters are by comparing the distances between points within the same cluster and those in neighbouring clusters. The scores range from -1 to 1, with higher average scores indicating better-defined clusters. To choose the optimal number of clusters, we ran each analysis from k = 2 to k = 10 and selected the solution with the highest silhouette score.

K-medoids is a fully unsupervised clustering method and is not designed to distinguish between controls and other groups, unlike HYDRA. In order to compare both techniques, we removed the control group from each dataset when using this implementation, and clustered only on autistic and/or ADHD individuals. We then compared resulting subgroups to controls on neuroanatomical and clinical measures.

Post clustering analysis

After determining the optimal number of clusters, we examined neuroanatomical, clinical and demographic differences, both between subgroups and against controls. Permutation tests were used to calculate significant differences between the neuroanatomical profiles of controls and each subgroup with a significance threshold of $p < 0.05$. All results were corrected for the false discovery rate (FDR) at 0.05. In addition, for the neuroanatomical analyses, we calculated effect sizes using Cohen's d to examine the magnitude of the differences at each regional location. Chi-squared tests were implemented to examine differences in diagnosis and site between subgroups and Kruskall-Wallis tests with a significant threshold of $p < 0.05$ were used to examine differences in IQ, age, and FSQC.

Benjamini–Hochberg FDR correction [45] was also applied to these measures at a threshold 0.05. We also calculated the Normalised Mutual Information (NMI) between different clustering solutions, both across techniques and across corrected and uncorrected datasets, to measure the overlap between them.

For both clustering methods, autism and ADHD were initially investigated separately, to attempt to maximise clinical interpretability and utility, and then they were combined in a single dataset to examine transdiagnostic clustering solutions.

Results

Subgroup number and stability

Global features

When using global features, HYDRA identified two clusters as the optimum number across all datasets: autism ARI = 0.799, ADHD ARI = 0.717, autism and ADHD ARI = 0.905. Similarly, k-medoids identified two clusters as the optimal solution across all global datasets: autism silhouette score = 0.483, ADHD silhouette score = 0.473, autism and ADHD silhouette score = 0.459. Visualising the clustered individuals projected onto the UMAP embeddings showed a clear divide along the second UMAP embedding (Fig. 1). The ARI score of the optimal solution for HYDRA was considerably higher than the next best solution (Figure S2a). This was not the case for k-medoids, where the silhouette score was similar across all possible clustering solutions (Figure S2b).

Regional features

In order to gain further insight into the differences between the subgroups, we performed clustering on the regional centile scores, based on Desikan-Killiany parcellations. Using these measures, HYDRA still identified two clusters as the optimal number for autism (ARI = 0.503). However, the optimal number of clusters for the ADHD and the autism and ADHD combined regional datasets increased to three (ARI = 0.509, ARI = 0.494 respectively). The ARI scores were overall lower for the regional dataset than for the global dataset, and the difference between the optimal solution and the others slightly decreased, suggesting lower stability of clusters (Figure S2a).

K-medoids regional results were consistent with those identified using global features: two clusters was the optimal solution across all datasets (autism silhouette score = 0.526, ADHD silhouette score = 0.442, autism and ADHD silhouette score = 0.465), however the silhouette scores were similar across all potential numbers of clusters (Figure S2b). The distribution of the clusters across the UMAP embeddings looked very similar to the global result (Fig. 1b). The first embedding appears to be capturing cortical thickness, and the second embedding is

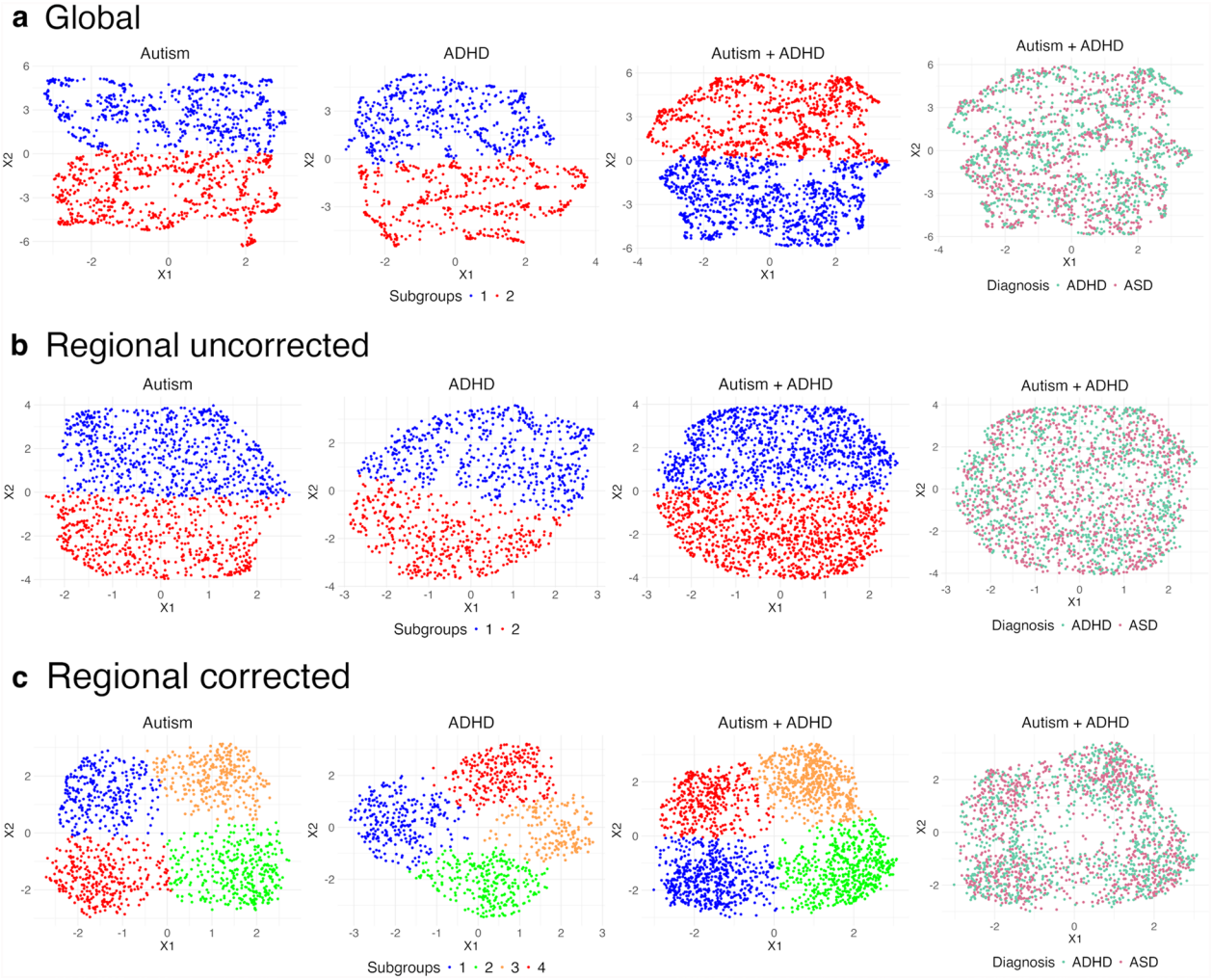

Fig. 1 Scatter plot showing distribution of individuals across UMAP embeddings (X1 and X2), coloured by subgroup. Each row corresponds to a different set of features: **a** global, **b** regional uncorrected, **c** regional corrected. Scatter plots for the combined dataset are also coloured in by diagnostic

capturing total grey matter volume and total surface area (Figure S3a and b).

Given the similarity between these regional results and those based solely on the global features, we decided to rerun the clustering after correcting for global effects. This allowed us to examine whether the initial local results were driven by global effects and examine any regionally-specific effects. With this corrected dataset, HYDRA identified three optimal clusters across all datasets: autism ARI = 0.602, ADHD ARI = 0.497, autism and ADHD ARI = 0.458. K-medoids, on the other hand, identified four clusters for all datasets: autism silhouette score = 0.497, ADHD silhouette score = 0.476, autism and ADHD silhouette score = 0.505. The UMAP embeddings plot showed clear divides across both axes in the autism and combined datasets (Fig. 1c). Similar to the global results, each UMAP embedding seems to be capturing different features: cortical thickness appears to be represented in the first embedding and surface area and volume in the second one (Figure S3c).

There is moderate overlap between the solutions from the different techniques for all global datasets: autism NMI = 0.4869, ADHD NMI = 0.5359, combined NMI = 0.577 (Figure S4).

Subgroup differences relative to controls

We next examined differences in neuroanatomical and clinical features between each subgroup and the control group. When using HYDRA, all datasets were clustered directly against the controls; while this was not the case for k-medoids, we compared the subgroups to the controls post hoc.

First, we examined the differences between subgroups and controls for the global dataset. Overall, all datasets

showed a clear pattern where one subgroup presented with significantly increased SA and GMV, and the other subgroup with decreased SA and GMV, relative to controls (Fig. 2). In addition, across all HYDRA results, there was also a significant difference in cortical thickness between the controls and each subgroup, where the direction of cortical thickness alterations was opposite to SA and GMV. WMV and sGMV followed similar patterns to SA and GMV, but ventricular volume showed a wider centile range in each subgroup (Figs. 2 and S5). Using k-medoids, we found consistent significant differences between subgroups and controls in the combined dataset, while the individual datasets exhibited varying significant differences across features and subgroups.

We then explored further differences between the controls and subgroups by looking at the results based on regional clustering.

Autism

Using HYDRA, we identified subgroups with similar patterns to the global results in the autism dataset. The first subgroup had increased SA and GMV, whilst the second subgroup showed decreased SA and decreased GMV. These effects appeared to be global across the cortex, with stronger effect sizes than CT. Additionally, differences in CT were significant in fewer regions and exhibited a less consistent direction within a subgroup (Fig. 3a). Given the similarities with the global results, this solution could have been driven by global effects in the dataset. When the regional features were corrected for global effects, however, the number of subgroups increased to three and the effect sizes decreased considerably, with no clear differences remaining between subgroups. The NMI across uncorrected and corrected datasets was small (NMI = 0.0017) indicating that the solutions were highly dissimilar (Figure S6a).

Similar to HYDRA, the k-medoids solution for the regional uncorrected dataset comprised two distinct subgroups, one with greater surface area and volume, and one with the opposite effect. However, in the corrected case, four subgroups were identified instead of two. The four subgroups displayed a similar pattern to the uncorrected case: two subgroups had higher SA and GMV relative to controls, and two subgroups had decreased SA and GMV. For each of these pairs of subgroups, one subgroup presented with increased CT, and the other one had decreased CT. This can be observed in the alluvial diagram (Figure S6b) where both subgroups in the uncorrected case contained individuals with both high and low CT, but after the correction, subgroups contained either high or low CT individuals. The NMI between regional solutions was higher than the HYDRA solution, but still low (NMI = 0.2686), indicating small overlap between solutions.

The overlap across techniques was high for the regional uncorrected dataset where the different solutions agreed in the optimal number of clusters, but very low in the corrected case where they differed (NMI Regional uncorrected = 0.4289, NMI Regional corrected = 0.0052, Fig. 3b).

Individual tables summarising all clinical and demographic results for the autism subgroups across techniques and datasets can be found in the supplementary, Tables S2–S7. There were no significant differences in clinical or demographic features in the global dataset for HYDRA (Figure S7). For this dataset, the k-medoids subgroups had a significant difference in age ($q-value = 0.026$, Cohen's d = 0.106), RBS-R score ($q-value = 0.026$, Cohen's d = -0.206) and SWAN ADHD Inattention ($q-value = 0.026$, Cohen's d = -0.313). In the regional uncorrected HYDRA subgroups, there was a significant difference in age ($q-value = 0.044$, Cohen's d = 0.090) and IQ ($q-value = 0.032$, Cohen's d = 0.144). Finally, for the regional corrected solution obtained with k-medoids, there was a significant difference across subgroups in FSQC ($q-value < 0.0001$) and site ($q-value < 0.0001$).

ADHD

For the uncorrected regional dataset, HYDRA identified three subgroups with reduced SA and GMV relative to controls (Fig. 4a). The effect sizes and significant regions varied across subgroups with subgroup 1 having the highest number of significant regions, and none observed in subgroup 3. After correcting for global effects, the pattern across subgroups shifted, with all subgroups showing slight increases in SA and GMV relative to controls and smaller effect sizes. There were only a few significant regions across the subgroups, which were primarily localised to the superior frontal cortex and the insula. In both cases, there was more local variability in the direction of the differences across regional measures of CT than GMV or SA. Similarly to the autism solution, there were little similarities between the uncorrected and corrected solutions (NMI = 0.0026, Figure S6a).

The k-medoids solution demonstrated the same pattern as the k-medoids autism subgroups: the two uncorrected subgroups presented with decreased and increased SA and GMV respectively, which then split into four different subgroups in the corrected case, capturing differences in cortical thickness (Fig. 4a). In both the corrected and uncorrected cases, the effect sizes were higher than using HYDRA. The similarity between uncorrected and corrected solutions was slightly higher than for HYDRA (NMI = 0.2545, Figure S6b).

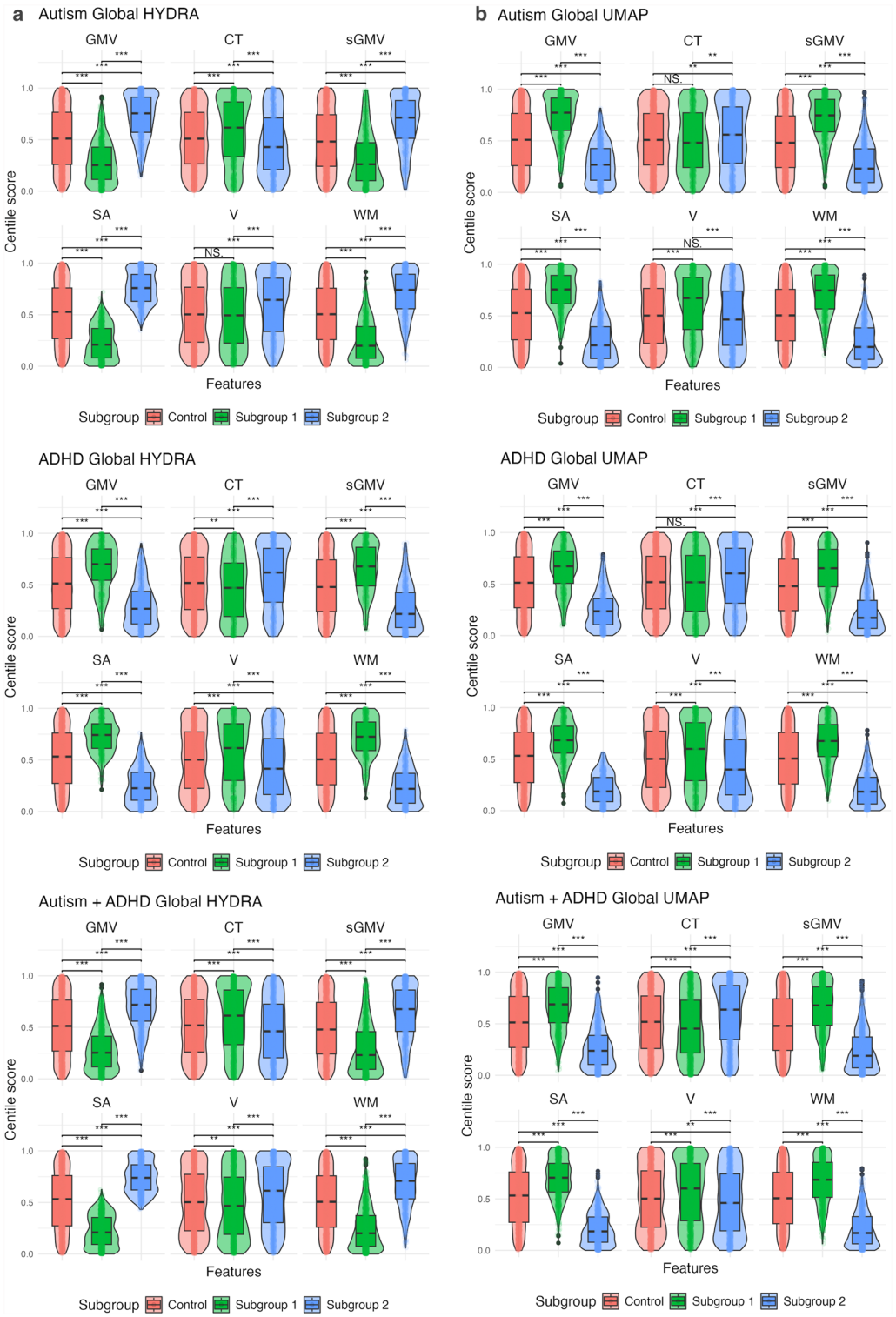

Fig. 2 Violin plots showing the difference between controls and subgroups for the main global features: total grey matter volume (GMV), mean cortical thickness (CT), total surface area (SA), white matter volume (WMV), subcortical grey matter volume (sGMV), ventricular volume (V). Each row corresponds to a different dataset. **a** HYDRA, **b** UMAP and K-medoids

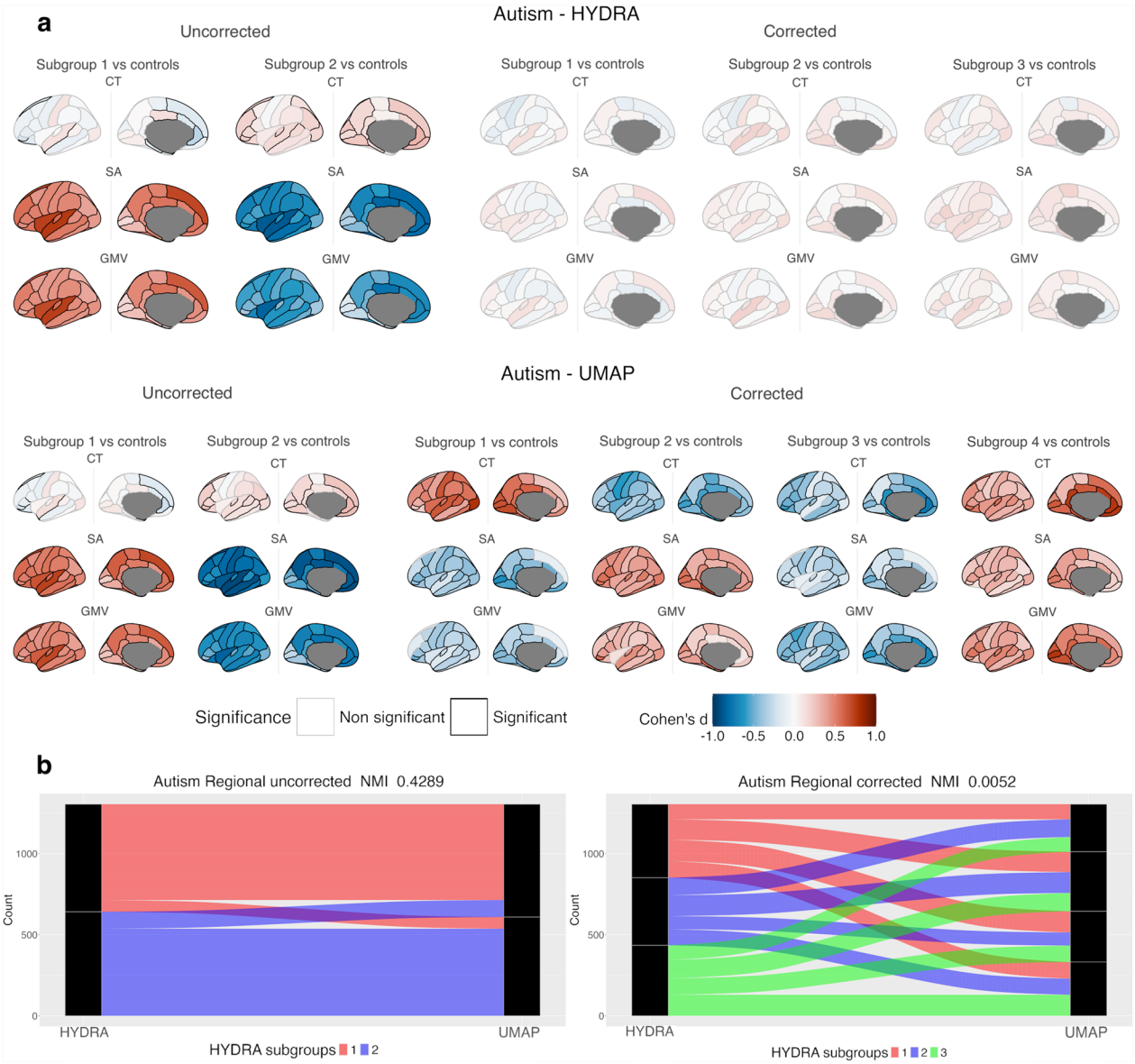

Fig. 3 Autism regional features results. **a** Differences in regional features between controls and autism subgroups, both uncorrected and corrected for global effects. The brain plots show Cohen's d effect size, where red represents a positive effect size (subgroup > control), and blue a negative effect size (subgroup < control). Significant regions are outlined in black. **b** Alluvial showing flow of participants across techniques for the uncorrected and corrected cases. The colour indicates the subgroups identified by HYDRA

The overlap of participants across HYDRA and k-medoids subgroups was minimal across both regional datasets (NMI = 0.0003 and NMI = 0.0016 for uncorrected and corrected respectively, Fig. 4b). Individual tables summarising all clinical and demographic results for the ADHD subgroups across techniques and datasets can be found in the supplementary, Tables S8–S13. For both HYDRA and UMAP, there was a significant difference in IQ between subgroups in the global dataset (HYDRA $q-value = 0.003$, Cohen's d = 0.281, and k-medoids $q-value = 0.038$, Cohen's d = 0.224, Figure S7). There were no further differences in clinical features across datasets. As in the autism k-medoids subgroups, there was a significant difference in site and FSQC across regional corrected k-medoids subgroups (FSQC $q-value < 0.001$ and site $q-value < 0.001$).

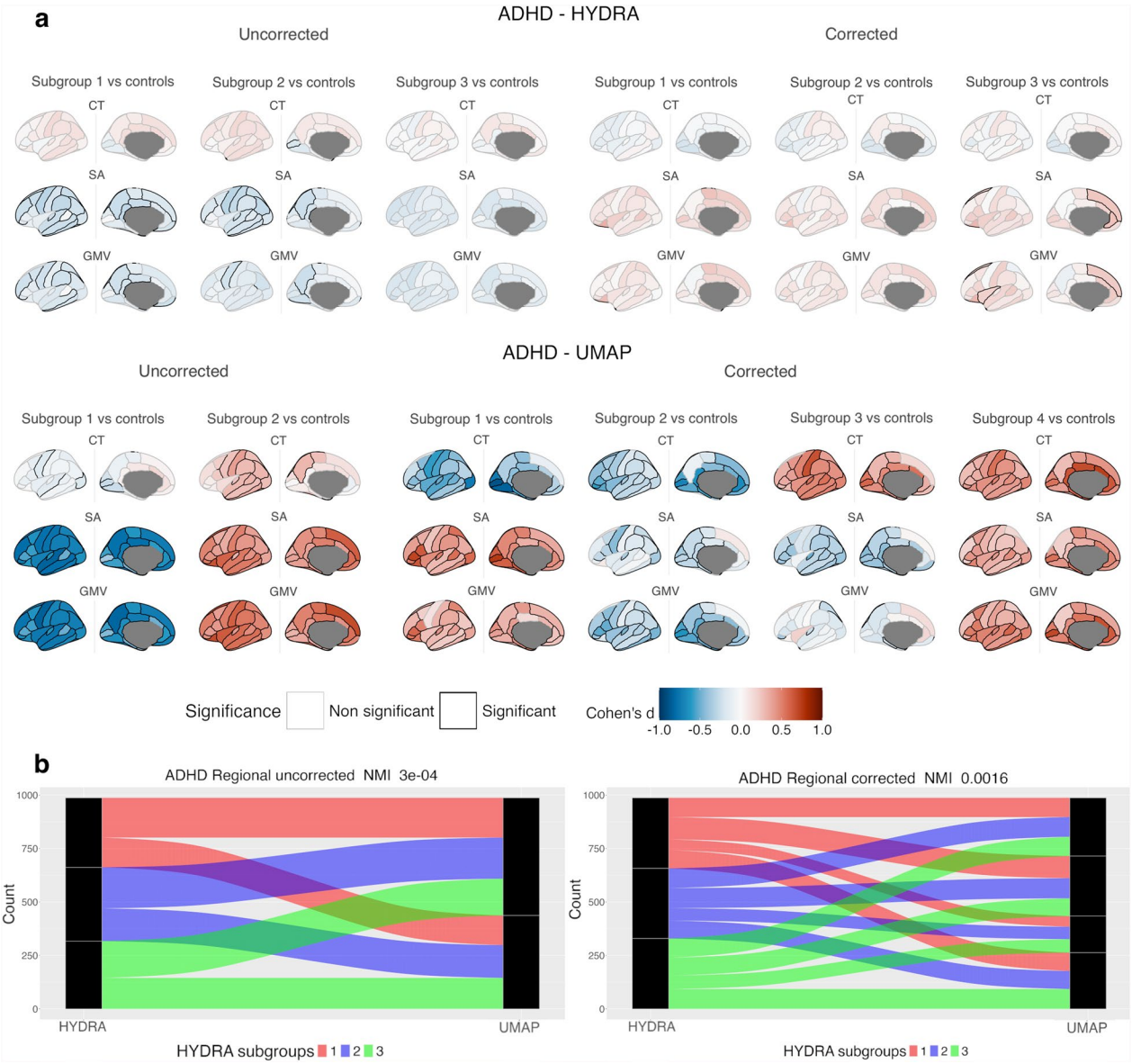

Fig. 4 ADHD regional features results. **a** Differences in regional features between controls and ADHD subgroups, both uncorrected and corrected for global effects. The brain plots show Cohen's d effect size, where red represents a positive effect size (subgroup > control), and blue a negative effect size (subgroup < control). Significant regions are outlined in black. **b** Alluvial showing flow of participants across techniques for the uncorrected and corrected cases. The colour indicates the subgroups identified by HYDRA

Combined dataset

The HYDRA subgroups for the combined dataset displayed similar patterns to the ones identified for ADHD (Fig. 5a). The initial dataset with uncorrected regional values was split into three subgroups, all with decreased surface area and volume relative to controls, though these differences were not significant in subgroup 3. After correcting for global effects, the significant effects disappeared, and the direction of the effect size was reversed, with subgroups having higher SA and GMV than the controls, though not significantly so. The uncorrected and corrected solutions were highly dissimilar (NMI = 0.0004, Figure S6a).

The k-medoids subgroups displayed the same pattern as the previous k-medoids results: two subgroups were identified for the uncorrected data, showing opposite patterns of SA and GMV, which then split into four when correcting for global effects. The effect sizes were stronger than for HYDRA with widespread significant differences across the cortex (Fig. 5a). The NMI across

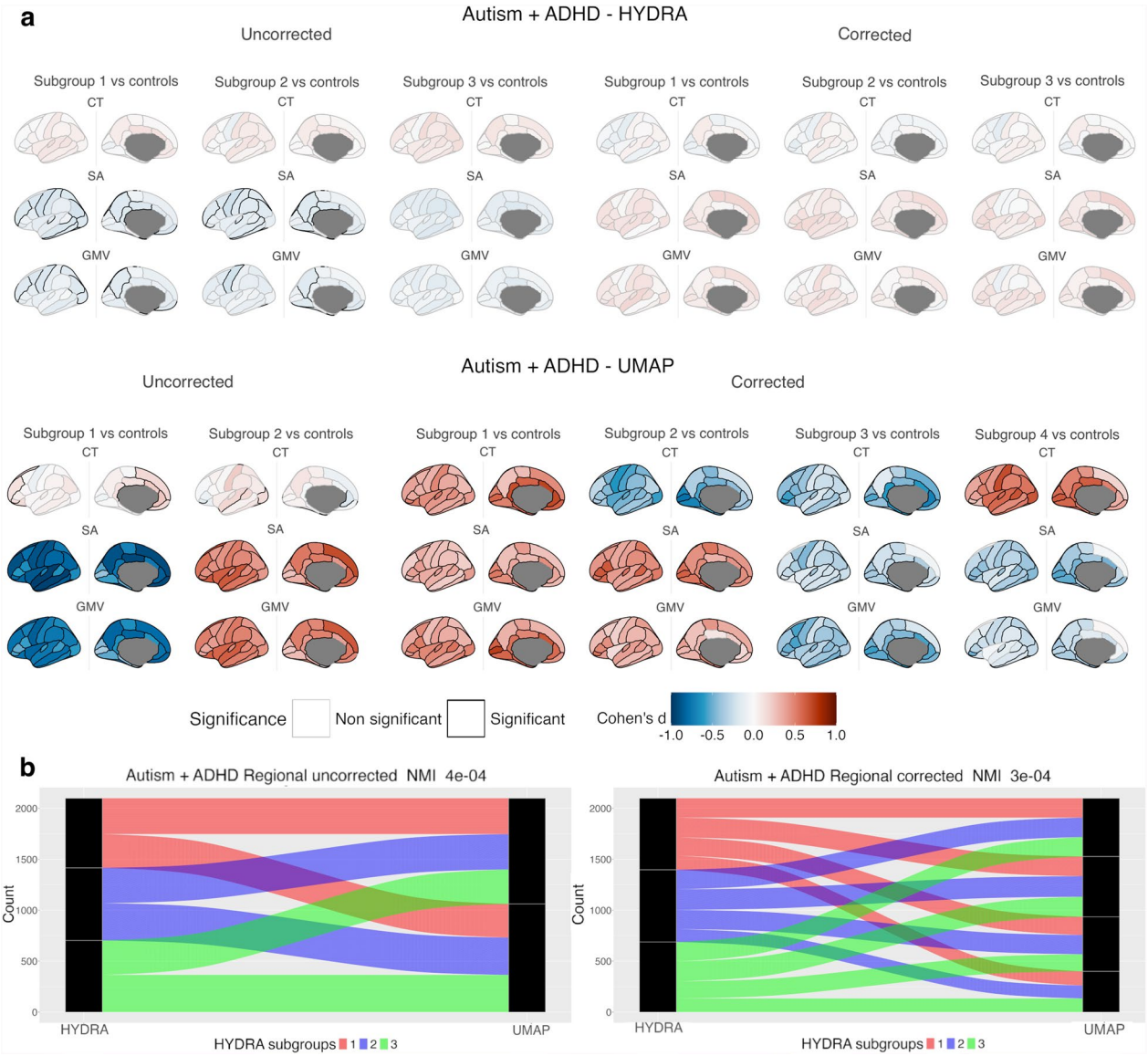

Fig. 5 Combined dataset regional features results. **a** Differences in regional features between controls and Autism and ADHD Subgroups, both uncorrected and corrected for global effects. The brain plots show Cohen's d effect size, where red represents a positive effect size (subgroup > control), and blue a negative effect size (subgroup < control). Significant regions are outlined in black. **b** Alluvial showing flow of participants across techniques for the uncorrected and corrected cases. The colour indicates the subgroups identified by HYDRA

regional uncorrected and corrected solutions was low, but higher than for the HYDRA solution (NMI = 0.2815, Figure S6b).

Similar to the ADHD case, the overlap in solutions across techniques was very small (NMI = 0.0004, NMI = 0.0003 for regional uncorrected and corrected respectively, Fig. 5b). Individual tables summarising all clinical and demographic results for the combined subgroups across techniques and datasets can be found in the supplementary, Tables S14–S19. For the global datasets, both HYDRA and k-medoids subgroups had significant differences in their IQ (HYDRA $q-value < 0.001$, Cohen's d = -0.205, and k-medoids $q-value < 0.0001$, Cohen's d = 0.217, Figure S7). In the regional uncorrected dataset, the k-medoids subgroups had a significant difference in age ($q-value = 0.026$, Cohen's d = -0.112) and IQ ($q-value = 0.025$, Cohen's d = -0.142). There was also a significant difference in the diagnosis across k-medoids subgroups in the regional corrected dataset ($q-value = 0.025$): autistic individuals were overrepresented in subgroup 4 compared against all other subgroups in a post hoc chi-squared test ($p-values = 0.01$,

0.01, 0.02 for subgroups 1, 2 and 3 respectively). IQ was also significantly different across these subgroups ($q-value=0.004$). Finally, similar to the ADHD case, there was a significant difference in FSQC and site for the combined regional subgroups identified with k-medoids (FSQC $q-value<0.0001$ and site $q-value<0.0001$). However, when visualising the distribution of sites across the two UMAP embeddings, no obvious clustering according to site is observed (Figure S8).

Comparison across methods

Overall, the patterns observed within a single clustering technique were more consistent than those observed across different techniques applied on the same dataset. Although there was agreement across k-medoids and HYDRA for the global datasets, there was little agreement across the regional ones (Figs. 3, 4, 5). When the regional measures were corrected for global effects, significant differences disappeared for the HYDRA solutions, and the subgroups did not display distinguishable neuroanatomical characteristics. On the other hand, for k-medoids, the effect sizes of the regional corrected dataset were still comparable to the regional uncorrected results. Generally, the ARI and silhouette scores observed were low, with no obvious peaks (Figures S2), indicating that the clusters were not well-separated or easily distinguishable into a clear number of subgroups. The largest difference between scores was observed for HYDRA solutions obtained using the global dataset, where the optimal cluster's ARI was considerably higher than that of the next best solution.

Discussion

In this paper, we used two different clustering algorithms to identify subgroups, both within and across autism and ADHD diagnoses, based on centile scores derived from population modelling. We compared our results across two different clustering implementations, as well as the effect of using global or regional features, and examining conditions individually or transdiagnostically. Most notably, we found that the optimal number of subgroups differs greatly depending on the algorithm chosen and features used, raising significant implications for the interpretation of studies focused on data-driven subgrouping methods, as well as the comparison of results between studies using different methods and algorithms.

Clustering based on global features

When clustering using global neuroanatomical features, results were similar across all datasets and methods. In all cases, two subgroups with opposite effects were identified: one with increased GMV and SA, and one with decreased GMV and SA relative to controls. Most datasets also had a significant difference in cortical thickness between the control group and the subgroups, but with a less consistent pattern. For the UMAP and k-medoids results, the UMAP embedding plot showed a clear separation across the second embedding for all datasets, which captured surface area and grey matter volume. For HYDRA, using the global features resulted in the highest ARI scores, as well as the greatest difference between the optimal solution and the others. This is likely caused by the lower dimensionality of the data when using global features compared to regional values which results in lower noise.

In some of our clustering solutions, we found certain individuals that, despite having a very large centile opposite to the subgroup, still get classified in that specific subgroup. This would be the case, for example, of an individual with a GMV centile of +0.75 getting classified into the 'small' brain subgroup. It is important to take into consideration that even in the global datasets, the data is clustered based on 6 global measurements: grey matter volume, surface area, cortical thickness, subcortical grey matter volume, ventricular volume, white matter volume. Therefore, the allocation to a specific subgroup is driven by the interaction of all of these variables, not just one in particular. It is also possible that an individual with a very high centile for GMV also has a low centile for CT (Figure S5). Although GMV, SA, sGMV and WMV are largely consistent in the directions across subgroups, CT and ventricular volume are more variable, and thus can have an impact on the subgrouping. Overall, most individuals do follow the general patterns of the subgroups.

The objective behind using all of the global measurements together during clustering was to capture patterns across them that could not be seen if the clustering had been carried out manually. In order to ensure that these purely data-driven methods can yield interpretable results, it is key to carry out appropriate post-clustering analysis, such as looking at the silhouette or ARI scores and comparing results across algorithms.

Clustering based on regional features

Regional uncorrected features

In order to further disentangle the differences between subgroups, we next clustered the data based on regional centile scores, with the aim of identifying more localised differences relative to controls. HYDRA still identified two subgroups as the optimum number for the autism dataset, with similar neuroanatomical profiles to the global results. For ADHD and for the combined dataset, however, the optimal number of clusters was three, suggesting that the ADHD individuals were potentially driving the results observed in the combined dataset. In

addition, the patterns across subgroups in the combined dataset looked more similar to the ADHD subgroups than the autism ones. For k-medoids, the optimal number of clusters identified with regional features was the same as for the global dataset for all groups, and the projections onto the two UMAP embeddings also showed the same partitioning as for the global result.

Regional features with global correction

Finally, to clarify whether these results were driven by global or regional effects, we clustered the data on regional features that were corrected for global effects by subtracting the corresponding global centile score from each regional value. Using the corrected dataset, HYDRA identified three optimal clusters across all datasets. Similar to the uncorrected case, the ARI scores were lower and closer to each other than for the global dataset. K-medoids, on the other hand, consistently identified four subgroups across all datasets.

Autism results

For both the global and regional uncorrected datasets, across both techniques, the optimal number of clusters identified for autism was two. These clusters showed distinct and opposite effects relative to the control group: one with increased surface area and grey matter volume, and the other with decreases in these features. In the regional uncorrected results, although the effect sizes and the number of significant regions were stronger for both SA and GMV, the subgroups also showed differences in cortical thickness: subgroup 1 presented with overall decreased CT, whereas subgroup 2 had an overall increase. This difference was further amplified when correcting for global effects in the k-medoids solution: the two clusters split into four, where two clusters had increased CT and two had decreased CT. In contrast, when correcting for global effects and implementing HYDRA, the effect sizes decreased greatly, and no significant regions remained. This indicates that HYDRA was not able to capture neuroanatomically different subgroups when the data was corrected for global effects. Our findings suggest that, rather than being restricted to specific brain regions, differences between autistic and non-autistic individuals manifest at a more global level (Fig. 2a). While the majority of autistic individuals have global brain measures within the typical range, our analyses indicate that the observed differences are not confined to discrete regions but instead reflect broader patterns of variation.

Given that when using HYDRA we are clustering autistic individuals against controls, we initially might expect 3 clusters to be found: one capturing microcephaly, one capturing macrocephaly, and one representing 'typically'-sized brains. However, the optimal number of clusters was two when clustering the global dataset using HYDRA. It is possible that HYDRA is sensitive to small deviations and clusters together individuals with the same trend, independent of the magnitude, as it is trying to find heterogeneity within the condition cohort.

In the case of k-medoids, where no controls were included in the clustering, we also found a 2-class solution for the global and regionally uncorrected autism datasets. As there were no controls in this dataset, and thus no label for 'typical', we might expect two clusters to arise capturing microcephaly and macrocephaly. The distribution of individuals across the UMAP embeddings does not show any clear clusters (Fig. 1), and the silhouette scores were close to each other. Therefore, although the solution was statistically optimal, the data was not highly separable. The overlap between HYDRA and k-medoids solutions was high for the global dataset, showing that similar patterns had been identified across methods for this particular dataset.

The two clusters identified here are also consistent with some previous work [23, 29, 36]. Finally, it is also possible that age related differences are impacting our findings, given the wide age range of our dataset, and that there is as of yet no consensus on findings of micro/macrocephaly later in life.

In the global dataset solution generated with k-medoids, the subgroups showed significant differences in age, RBS-R Total and SWAN ADHD Inattention. However, the effect sizes were small. IQ and age also differed significantly for the regional uncorrected HYDRA subgroups, again with small effect sizes. Overall, the significant differences in clinical features between subgroups were not consistent across datasets using different features or different clustering techniques.

Previous studies using neuroimaging phenotypes to cluster within autism have similarly found subgroups with increased and decreased cortical thickness and surface area, although the number of clusters varies between studies. In a recent study, Zabihi et al. observed 5 distinct subgroups of autism with different clinical profiles using population modelling [28]. On the other hand, Itahashi et al. reported two clusters implementing HYDRA within a male-only autistic and ADHD cohort using measures of CT and SA separately [29]. There was no relationship between the diagnosis and the subgroups identified in the study. These findings further reinforce that the number of subgroups identified is highly dependent on the data used and the algorithms chosen. A previous study that clustered autistic individuals based on their ADOS-G scores found two distinct subgroups. Using normative modelling, they were able to compare neuroanatomical differences between the subgroups and found that the

direction of the CT deviations varied within each subgroup, similar to our findings [23]. Additionally, the study highlights significant heterogeneity within subgroups as the accuracy of a support vector machine did not improve when clustering the controls against the subgroups instead of against all of the autistic individuals.

Our results are in line with research suggesting that previously inconsistent findings with regards to brain overgrowth in autism may in fact be capturing distinct subgroups within the autistic population, including one with macrocephaly [10, 11], highlighting the importance of examining heterogeneity and differences within clinical cohorts instead of focusing only on case-control analysis. Previous research suggests young autistic boys are much more likely than girls to fall into the macrocephaly subgroup [10, 11, 46]. Here, we did not find a significant sex difference, however, unfortunately, our female sample may have been underpowered, and this is an important future topic of investigation. Critically, these subgroups have been demonstrated to also have distinct behavioural and clinical profiles, as well as functional outcomes [10, 46]. Unfortunately, we were also limited in our ability to explore clinical differences between subgroups due to the consistent availability of clinical data across sites in our dataset; however, future research on better characterised samples will be critical to achieving clinical translation or utility from subgrouping research.

ADHD results

In the ADHD global dataset, both algorithms identified two distinct subgroups as the optimal solution, each one presenting with increased or decreased SA and GMV. However, the solutions for the regional datasets differed greatly across methods. For the regional uncorrected dataset, HYDRA identified 3 clusters, all with slight decreases in SA and GMV relative to the controls; however, these differences were only significant in some regions for subgroups 1 and 2, with very small effect sizes, and with no significant differences in subgroup 3. The direction of CT variations were less consistent within subgroups, and only one region was significant across all three groups. After correcting for global effects, the patterns across subgroups looked very different: all subgroups showed an overall relative increase in surface area and grey matter volume with respect to the controls, but with very few regions reaching significance. The regions which did show significant increases relative to controls were localised to the prefrontal cortex and insula, rather than the widespread effect across the cortex observed in other analyses. The prefrontal cortex is a highly relevant area for ADHD and has been consistently implicated in neuroimaging studies, and linked to difficulties in regulatory and reward behaviours, social and emotional processing, planning, and cognitive control in diagnosed individuals [17, 47–50]. The solution for k-medoids, however, captured two distinct subgroups for the uncorrected dataset that separated into four clusters in the corrected case, as was observed in the autism dataset. Although there were no clinical differences across subgroups, there was a significant difference in IQ in the global dataset for HYDRA and k-medoids, where in both cases subgroup 1 had a slightly higher IQ than subgroup 2. The effect sizes in both cases were small.

Combined dataset results

In the autism and ADHD combined dataset, the HYDRA solutions strongly resembled those obtained in the ADHD dataset. In addition, the k-medoids solution reflected the same pattern as for the single-diagnosis datasets, with two subgroups that split into four after correcting for global effects. Both methods identified a significant difference in IQ between subgroups in the global dataset, again with small effect sizes. In addition, the k-medoids derived subgroups showed significant differences in age and IQ in the regional uncorrected dataset, and differences in IQ and diagnostic in the regional corrected dataset. In this last case, although all subgroups contained participants from both diagnoses, subgroup 4 had a reduced ratio of ADHD to autistic participants and also a significantly increased IQ with respect to two other subgroups. The differences in diagnostic group, IQ and site across the combined dataset all appear to be linked as the ADHD participants had overall lower IQ and came from different sites than the autism participants. Thus, it is difficult to disentangle these results and understand which one was driving the clustering.

Itahashi et al. [29] used HYDRA to cluster a cohort of male ADHD and autistic individuals using regional measures of cortical thickness and surface area. They found a 2-class solution across both cortical thickness and surface area and two different atlases. Each class represents an inverse pattern with respect to controls, one shows an increase and the other a decrease. This is the case for both solutions clustered using CT and SA separately. For the equivalent dataset in our study (the combined regional datasets clustered using HYDRA) we instead find 3 clusters as the optimal solution. In our case, the three clusters show overall the same pattern, with decreased SA and GMV, but no significant differences in CT.

Differences in these results could be due to multiple factors. Firstly, there are differences in dataset characteristics: we are working with a larger number of individuals, across multiple sites, including both men and women, whereas they focused on a smaller cohort of only males. Secondly, they clustered using raw measurements of surface area and cortical thickness, obtained for both

Schaefer and Destrieux' atlases. We instead used centile scores obtained for the Desikan-Killiany atlas. Their findings do however support the hypothesis that HYDRA is sensitive to differences between subgroups and a 'typical' cluster might be unlikely.

Clinical and demographic differences between subgroups

IQ was the clinical feature with data available for most participants, which likely contributed to its higher statistical power relative to the other features and made the detection of significant differences more feasible. These results suggest that IQ may be a useful clinical feature but show that having sufficient data is key in obtaining adequate statistical power. In other studies with more extensive clinical data — either focusing on autism subgroups [23, 34] or using transdiagnostic datasets [51] — significant differences in clinical features were identified between subgroups, suggesting that these features could serve as potential markers for subgrouping.

Apart from the clinical features, all adjusted regional k-medoids solutions had significant differences across subgroups for both FSQC and site. Strict quality control was carried out to select participants for this study, therefore all the participants chosen had high quality scans. In addition, when plotting the individuals along the UMAP embeddings and visualising the distribution of sites, there were no clear differences or clusters observed (Figure S8). The site difference could again be linked to the fact that participants with different diagnoses were overrepresented in certain sites.

Comparison between algorithms

The stark differences across methods may stem from their distinct approaches to clustering. In addition to differences in the underlying algorithms, HYDRA clusters individuals directly against the controls, while the k-medoids method only clusters within the diagnostic group. As a result, k-medoids is able to capture differences in the subgroups that are inherent to the entire population. This could explain why effect sizes were overall higher in the k-medoids solutions than for HYDRA, specifically in the regional cases, as it could be capturing wider variations in the population. As HYDRA is actively looking for differences against the controls, it may also handle site or FSQC more effectively, which had significant effects for some of the k-medoids derived subgroups, but not for HYDRA.

It is also a challenge to compare across clustering methods that use different evaluation metrics, such as the ARI and the silhouette score. The ARI measures how reproducible a clustering solution is across trials. On the other hand, the silhouette score is a measure of how well separated clusters are. Thus, a direct comparison between both is not easily interpretable. We observed low silhouette scores, which may indicate some underlying structure in the data, but also suggest that the clusters are not well-defined or distinct. In this case, the ARI score could be high if cluster assignments were consistent across trials even though the clusters are not well separated.

Validation of clusters and algorithm selection

As noted above, while there was some agreement, in some cases the two algorithms resulted in drastically different clustering solutions. This raises important implications about the generalisability and validity of findings of data-driven subtypes in clinical conditions in the current literature, in particular in the absence of a ground truth. Critically, we also lacked consistent clinical measures across datasets to properly examine whether these brain-based subgroups map onto distinct clinical profiles or have tangible clinical utility. We observed only very minimal clinical differences between subgroups, though it is difficult to know to what extent this is attributable to insufficient data and statistical power. Given that this exploration of clusters was purely data-driven and there is no ground truth regarding the existence of specific subgroups, it is difficult to conclude which clustering approach is best. HYDRA has the advantage of clustering the data directly against the controls and therefore should account for inter-individual variability. HYDRA may be especially well-suited to study a condition with a distinct boundary with neurotypical individuals. The ARI scores we obtained, however, were not as high as those obtained in previous studies with fewer participants, all of which were from a single site [29]. In that study, individuals were clustered twice, one based on CT, and one based on SA measurements. As we clustered based on all CT, SA, GMV features simultaneously, the dataset may have contained more noise. Therefore, handling great amounts of data coming from different sites may still be a limitation of these clustering algorithms. Behaviour-based subgroups may be more clinically relevant and more reproducible [52–54]. Therefore, exploring neuroanatomical differences within a diagnosis should be guided by clinical differences to enhance clinical translation. This study highlights the fact that this kind of algorithms are highly sensitive to the input data and the parameters used, and thus researchers should ensure they are implementing validation methods to ensure robustness when identifying subgroups. An example of this could be implementing multiple clustering algorithms like in this study and then measuring the agreement in the solution across methods.

Future directions

Our study exemplifies the many considerations that need to be taken into account when carrying out clustering. Firstly, the datasets and features used to carry out the clustering, and how they are preprocessed, matter. We found different numbers of clusters depending on whether we used the global or the regional features. Although using regional features might capture more information about a given participant, it can also introduce more noise given the higher number of dimensions. This could explain why the ARI scores for HYDRA were much higher in the global dataset. When using k-medoids, we projected the data onto two dimensions using UMAP for both global and regional datasets, and we did not observe the same drop in silhouette score. Therefore, it is clear that the dimensionality of the data can impact the result. However, with the increasing availability of large-scale data both in terms of sample sizes as well in the number of phenotypes measured, it will become more important to find ways to reduce the data dimensionality while retaining interpretability and signal. While we have outlined several permutations of this here, these are by no means exhaustive, and we encourage future work to explore multiple approaches for identifying meaningful subgroups given the variable impact the chosen method can have on the outcome.

In addition, another consideration would be the features and modalities used during clustering. We incorporated measures of regional and global grey matter volume, cortical thickness, surface area, as well as global white matter volume, subcortical grey matter volume and ventricular volume, together in the clustering, although previous studies have used these measures separately [29]. Although this implementation was part of our purely data-driven approach, future studies could identify which features are the most informative and potentially use only those for clustering to decrease noise in the dataset and improve the robustness of the findings. This would be more challenging in transdiagnostic studies but could be particularly useful in studies on a single condition, in which feature selection could be guided by previously reported case–control differences to identify participants which do and do not fit these neuroanatomical profiles of difference. Previous studies have also successfully identified subgroups of autism based on clinical or behavioural data [20, 21, 51] or employed methods to combine these and other multimodal features for clustering [55, 56]. In our study, given the reduced availability of clinical data, we clustered solely based on the neuroanatomical centiles and then investigated differences in clinical data between subgroups post hoc. For future studies, where clinical data is available, it would be interesting to explore combining it with the neuroimaging data to develop multimodal approaches to clustering. Such methods may better capture individual differences and help unravel heterogeneity across domains.

The choice of algorithm can also lead to different results. Using k-medoids requires using an additional dimensionality reduction technique, which introduces further preprocessing steps that can affect final results. On the other hand, HYDRA is deployed end-to-end, eliminating the need for separate dimensionality reduction before clustering the data. The second main difference between these implementations is the fact that HYDRA clusters the data against the controls. This can also be a trade-off to consider in future studies as it also requires collecting comparison/control participants data and involves more complex methods and longer processing time.

One ultimate goal of subgrouping is, of course, to inform clinical practice and support of autistic individuals and those with ADHD. For this to be possible, better characterised samples with more consistent clinical data will be necessary, either to validate and identify the clinical significance of subgroups, or as inputs to cluster on, as mentioned above. For example, being able to identify support needs and/or functional outcomes could be hugely valuable given the high phenotypic heterogeneity and variability in clinical support needed observed in neurodevelopmental conditions. Future work should explore the most informative and clinically significant features to this end.

Limitations

Further limitations to this study include other difficulties that arise from combining data from multiple sources. Here, we chose to capitalise on the use of available data and the undeniable benefits that come with large sample sizes that allow for robust data driven approaches, and which would be impossible to recruit and collect from a single site. While this of course comes with downsides of multi-site data acquisition, we have taken extensive methods to minimise the impact of site, harmonising data using a two-step procedure involving ComBat [42] and population modelling [19, 33]. Using a multi-site dataset allows us to leverage all available data and avoid issues associated with small sample sizes. To ensure the robustness of our results, we implemented cross-fold validation in HYDRA and calculated ARI scores to measure agreement between solutions. We also calculated silhouette scores for the k-medoids results to assess how well-separated the clusters were. Additionally, we evaluated the agreement across methods by calculating the NMI score between HYDRA and k-medoids, allowing us to assess the consistency of results across different clustering implementations, as a further robustness test.

However, some site-related differences will no doubt remain, in particular in demographic characteristics. For example, the age ranges of the autism and ADHD samples differed, and the proportion of diagnostic group was uneven between sites.

A restricted clinical characterization also likely resulted in a lack of statistical power, potentially explaining why very few significant differences were observed in clinical measures between subgroups, and further complicating comparisons to other studies with more complete clinical data. We were also unable to examine the effect of clinical factors such as co-occurring conditions, substance use, or medication status, due to lack of consistent availability of this data. In addition, given the relatively low number of female individuals with ADHD or autism in the dataset, we were not able to systematically examine sex differences across subgroups or conduct clustering separately by sex; of note, we did not find any differences in sex distribution across clusters. Gender identity data were also not available in these datasets, hence could not be accounted for. This highlights the importance of recruiting diverse participants across sites and measuring consistent features.

Conclusion

Our results highlight the high heterogeneity in autism and ADHD, and the importance of understanding differences within clinical categorical cohorts rather than relying solely on case–control comparison, as well as demonstrating that population modelling is a useful tool in facilitating this. However, our results also demonstrate that individuals have diverse neuroanatomical profiles that cannot be consistently clustered. Finally, and perhaps most importantly, we demonstrate that algorithm and feature selection significantly impact the subgroups identified, often giving starkly different solutions between methods and datasets. This has important implications for comparing results between studies in the current literature and highlights the care that must be taken when designing studies investigating subgroups. Based on these results, we encourage future studies to use multiple clustering approaches and report on the parameters and feature selection as well as using different forms of evaluation.

Abbreviations

ABIDE	Autism Brain Imaging Data Exchange
ADHD	Attention Deficit-Hyperactivity Disorder
ADOS	Autism Diagnostic Observation Schedule
ARI	Adjusted Rand Index
CMI	Child Mind Institute
CT	Cortical Thickness
FDR	False Discovery Rate
FSQC	FreeSurfer Quality Control
GAMLSS	Generalised Additive Models of Location, Scale and Shape
GMV	Grey Matter Volume
HYDRA	HeterogeneitY through DiscRiminative Analysis
HBN	Healthy Brain Network
MRC-AIMS	Medical Research Council Autism Imaging Multi-centre Study
MRI	Magnetic Resonance Imaging
POND	Province of Ontario Neurodevelopmental Network
RBS-R	Restricted Behaviours Scale (Revised)
SA	Surface Area
sGMV	subcortical Grey Matter Volume
SRS	Social Responsiveness Scale
SWAN	Strengths and Weaknesses of ADHD Symptoms and Normal Behaviour
tSNE	t-distributed Stochastic Neighbour Embedding
UMAP	Uniform Manifold Approximation and Projection
WMV	White Matter Volume

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13229-025-00667-z.

Supplementary file1.

Acknowledgements

We would like to thank all the participants that took part in these studies.

Author contributions

Conceptualization: CPT, SAB, RAIB. Implementation: CPT. Data collection: MC-L, MVL,BC, AR, JS, EA, JPL, MT, RB, GS, JC, RS, EK, JJ, PDA, SB-C. Data curation: SAB,PT. Contributed resources: ETB, JS, AFA-B. Writing—original draft: CPT, SAB, RAIB. Supervision: SAB, RAIB. Funding: RAIB. All authors reviewed the manuscript.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council (Grant No BBSRC BB/X010899/1 [to CPT]); the Trinity College Coutts-Trotter Studentship and the Canadian Centennial Scholarship Foundation (to SAB); the Wellcome Trust (Grant No. 214322\Z\18\Z [to SB-C]); the Autism Centre of Excellence, the Simons Foundation Autism Research Initiative (SFARI), the Templeton World Charitable Foundation, and the Medical Research Council (to SB-C); the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre and an NIHR Senior Investigator Award (to ETB); the Alberta Innovates Translational Health Chair in Child and Youth Mental Health, the Canadian Institutes of Health Research, Canadian Foundation for Innovation, and the Alberta Children's Hospital Foundation, and Biohaven Pharmaceuticals (to PDA); the Academic Scholars Award from the Department of Psychiatry at the University of Toronto, the Discovery Fund from the Centre for Addiction and Mental Health, and the Canadian Institutes of Health Research Sex and Gender Science Chair (Grant No. GSB 171373 [to M-CL]); the Ontario Brain Institute to (EA and JPL); National Institute on Deafness and Other Communication Disorders (Grant No. R01DC016385 [to EC and KP]); and National Institute of Mental Health (NIMH) (Grant Nos. R01MH104446 [to KP] and R01MH133843 [to AFA-B]).

Availability of data and materials

The code for HYDRA [31] can be found here: https://github.com/evarol/HYDRA. The code for clustering using k-medoids and UMAP can be found here: https://github.com/clarapecci/subgrouping_autism_adhd_population_modelling. All datasets included in this study, with the exception of MRC-AIMS, are publicly available upon request. For more information, see. ABIDE [57]: https://fcon_1000.projects.nitrc.org/indi/abide/. ADHD200 Consortium [58]: https://fcon_1000.projects.nitrc.org/indi/adhd200/. POND: https://pond-network.ca/. HNB-CMI [59]: https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/. Female ASD: https://nda.nih.gov/edit_collection.html?id=2021. MRC—AIMS [60–62]: https://www.autismresearchcentre.com/past-projects/cognition-and-neuroimaging-in-females-with-autism-spectrum-conditions

Declarations

Ethics approval and consent to participate
Ethical approval and informed consent were obtained for each primary study. The Cambridge Psychology Research Ethics Committee (PRE.2020.104) deemed that secondary analysis of deidentified data did not require ethical oversight.

Consent for publication
Not applicable.

Competing interests
RAIB, MVL, and M-CL are Associate Editors, and EA and BC are Editorial Board members of Molecular Autism. SBC is a former Editor-in-Chief of the journal. ETB reports consultancy work for Boehringer Ingelheim, Sosei Heptares, SR One, and GlaxoSmithKline. ETB, RAIB, JS, and AFA-B are cofounders of Centile Bioscience. PDA receives research support from Biohaven Pharmaceuticals. M-CL has received editorial honorarium from SAGE Publications. RN reported receiving grants from Brain Canada, Hoffman La Roche, Otsuka Pharmaceuticals, and Maplight Therapeutics outside the submitted work. EA reported receiving grants from Roche and Anavex; receiving nonfinancial support from AMO Pharma and CRA-Simons Foundation; and receiving personal fees from Roche, Impel, Ono, and Quadrant outside the submitted work.

Author details
[1]Department of Psychology, University of Cambridge, Cambridge, UK. [2]Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK. [3]Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health and Azrieli Adult Neurodevelopmental Centre, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada. [4]Department of Psychiatry, The Hospital for Sick Children, Toronto, ON, Canada. [5]Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada. [6]Department of Psychiatry, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan. [7]Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems, Istituto Italiano Di Tecnologia, Rovereto, Italy. [8]Centre for Autism, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK. [9]Division of Psychology and Mental Health, Faculty of Biology, Medicine and Health, School of Health Sciences, University of Manchester, Manchester, UK. [10]Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, UK. [11]Holland Bloorview Kids Rehabilitation Hospital, Bloorview Research Institute Toronto, University of Toronto, Toronto, ON, Canada. [12]Department of Pediatrics, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada. [13]Program in Neurosciences and Mental Health, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada. [14]Mouse Imaging Centre, Hospital for Sick Children, Toronto, ON, Canada. [15]Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [16]Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, ON, Canada. [17]Department of Psychiatry, University of Western Ontario, London, ON, Canada. [18]McMaster University, Hamilton, ON, Canada. [19]Genetics & Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada. [20]Department of Psychology, Queen's University, Kingston, ON, Canada. [21]Centre for Neuroscience Studies, Queen's University, Kingston, ON, Canada. [22]Department of Psychiatry, Queen's University, Kingston, ON, Canada. [23]Mathison Centre for Mental Health Research & Education, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. [24]Departments of Psychiatry and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. [25]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. [26]Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. [27]Lifespan Brain Institute, The Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA, USA. [28]Cambridge Lifetime Autism Spectrum Service, Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK.

Received: 29 November 2024 Accepted: 14 May 2025
Published online: 04 June 2025

References

1. Lenroot RK, Yeung PK. Heterogeneity within autism spectrum disorders: what have we learned from neuroimaging studies? Front Hum Neurosci. 2013;7:733.
2. Karalunas SL, Nigg JT. Heterogeneity and subtyping in attention-deficit/hyperactivity disorder—considerations for emerging research using person-centered computational approaches. Biol Psychiatry. 2020;88:103–10.
3. Lombardo MV, Lai MC, Baron-Cohen S. Big data approaches to decomposing heterogeneity across the autism spectrum. Mol Psychiatry. 2019;24:1435–50.
4. Lai MC, Lombardo MV, Chakrabarti B, Baron-Cohen S. Subgrouping the Autism "Spectrum": Reflections on DSM-5. PLoS Biol. 2013;11: e1001544.
5. Courchesne E. Brain development in autism: Early overgrowth followed by premature arrest of growth. Ment Retard Dev Disabil Res Rev. 2004;10:106–11.
6. Sacco R, Gabriele S, Persico AM. Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. Psychiatry Res Neuroimaging. 2015;234:239–51.
7. Courchesne E, Karns CM, Davis HR, Ziccardi R, Carper RA, Tigue ZD, et al. Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. Neurology. 2001;57:245–54.
8. Raznahan A, Wallace GL, Antezana L, Greenstein D, Lenroot R, Thurm A, et al. Compared to What? Early Brain Overgrowth in Autism and the Perils of Population Norms. Biol Psychiatry. 2013;74:563–75.
9. Yankowitz LD, Herrington JD, Yerys BE, Pereira JA, Pandey J, Schultz RT. Evidence against the "normalization" prediction of the early brain overgrowth hypothesis of autism. Mol Autism. 2020;11:51.
10. Amaral DG, Li D, Libero L, Solomon M, Van De Water J, Mastergeorge A, et al. In pursuit of neurophenotypes: The consequences of having autism and a big brain. Autism Res. 2017;10:711–22.
11. Lee JK, Andrews DS, Ozonoff S, Solomon M, Rogers S, Amaral DG, et al. Longitudinal evaluation of cerebral growth across childhood in boys and girls with autism spectrum disorder. Biol Psychiatry. 2021;90:286–94.
12. Libero LE, Nordahl CW, Li DD, Ferrer E, Rogers SJ, Amaral DG. Persistence of megalencephaly in a subgroup of young boys with autism spectrum disorder. Autism Res. 2016;9:1169–82.
13. Amaral DG, Andrews DS, Nordahl CW. Structural brain imaging biomarkers of autism spectrum disorder. Neurophysiol Biomark Neuropsychiatr Disord. 2024;40: 491509.
14. Ohta H, Nordahl CW, Iosif A, Lee A, Rogers S, Amaral DG. Increased surface area, but not cortical thickness, in a subset of young boys with autism spectrum disorder. Autism Res. 2016;9:232–48.
15. Libero LE, Schaer M, Li DD, Amaral DG, Nordahl CW. A longitudinal study of local gyrification index in young boys with autism spectrum disorder. Cereb Cortex. 2019;29:2575–87.
16 Narr KL, Woods RP, Lin J, Kim J, Phillips OR, Del'Homme M, et al. Widespread cortical thinning is a robust anatomical marker for attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2009;48:1014–22.
17. Hoogman M, Muetzel R, Guimaraes JP, Shumskaya E, Mennes M, Zwiers MP, et al. Brain Imaging of the cortex in ADHD: a coordinated analysis of large-scale clinical and population-based samples. Am J Psychiatry. 2019;176:531–42.
18. Levman J, Forgeron C, Shiohama T, MacDonald P, Stewart N, Lim A, et al. Cortical thickness abnormalities in attention deficit hyperactivity disorder revealed by structural magnetic resonance imaging: Newborns to young adults. Int J Dev Neurosci. 2022;82:584–95.
19. Bedford SA, Lai MC, Lombardo MV, Chakrabarti B, Ruigrok A, Suckling J, et al. Brain-charting autism and attention-deficit/hyperactivity disorder reveals distinct and overlapping neurobiology. Biol Psychiatry. 2025;97:517–30.
20. Bitsika V, Sharpley CF, Orapeleng S. An exploratory analysis of the use of cognitive, adaptive and behavioural indices for cluster analysis of ASD subgroups. J Intellect Disabil Res JIDR. 2008;52:973–85.
21. Hu VW, Steinberg ME. Novel clustering of items from the autism diagnostic interview-revised to define phenotypes within autism spectrum disorders. Autism Res Off J Int Soc Autism Res. 2009;2:67–77.
22. Nordahl CW, Iosif AM, Young GS, Hechtman A, Heath B, Lee JK, et al. High psychopathology subgroup in young children with autism: associations with biological sex and amygdala volume. J Am Acad Child Adolesc Psychiatry. 2020;59:1353-63.e2.

23. Meijer J, Hebling Vieira B, Elleaume C, Baranczuk-Turska Z, Langer N, Floris DL. Toward understanding autism heterogeneity: Identifying clinical subgroups and neuroanatomical deviations. J Psychopathol Clin Sci. 2024;133:667–77.
24. Bethlehem RAI, Seidlitz J, Romero-Garcia R, Trakoshis S, Dumas G, Lombardo MV. A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. Commun Biol. 2020;3:486.
25. Hong SJ, Valk SL, Di Martino A, Milham MP, Bernhardt BC. Multidimensional Neuroanatomical Subtyping of Autism Spectrum Disorder. Cereb Cortex. 2018;28:3578–88.
26. Li T, van Rooij D, Roth Mota N, Buitelaar JK, Hoogman M, et al. Characterizing neuroanatomic heterogeneity in people with and without ADHD based on subcortical brain volumes. J Child Psychol Psychiatry. 2021;62:1140–9.
27 Wolfers T, Beckmann CF, Hoogman M, Buitelaar JK, Franke B, Marquand AF. Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. Psychol Med. 2020;50:314–23.
28. Zabihi M, Floris DL, Kia SM, Wolfers T, Tillmann J, Arenas AL, et al. Fractionating autism based on neuroanatomical normative modeling. Transl Psychiatry. 2020;10:384.
29. Itahashi T, Fujino J, Hashimoto RI, Tachibana Y, Sato T, Ohta H, et al. Transdiagnostic subtyping of males with developmental disorders using cortical characteristics. NeuroImage Clin. 2020;27: 102288.
30. Bertelsen N, Landi I, Bethlehem RAI, Seidlitz J, Busuoli EM, Mandelli V, et al. Imbalanced social-communicative and restricted repetitive behavior subtypes of autism spectrum disorder exhibit different neural circuitry. Commun Biol. 2021;4:574.
31. Varol E, Sotiras A, Davatzikos C. HYDRA: revealing Heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage. 2017;145:346–64.
32. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biol Psychiatry. 2016;80:552–61.
33. Bethlehem RAI, Seidlitz J, White SR, Vogel JW, Anderson KM, Adamson C, et al. Brain charts for the human lifespan. Nature. 2022;604:525–33.
34. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. Dissecting the heterogeneous cortical anatomy of autism spectrum disorder using normative models. Biol Psychiatry Cogn Neurosci Neuroimaging. 2019;4:567–78.
35. Shan X, Uddin LQ, Xiao J, He C, Ling Z, Li L, et al. Mapping the heterogeneous brain structural phenotype of autism spectrum disorder using the normative model. Biol Psychiatry. 2022;91:967–76.
36. Kaufmann L, Rousseeuw P. Clustering by Means of Medoids. Data Anal Based L1-Norm Relat Methods. 1987:405–16.
37. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2020 http://arxiv.org/abs/1802.03426 Accessed 14 Mar 2024
38. Bedford SA, Ortiz-Rosa A, Schabdach JM, Costantino M, Tullo S, Piercy T, et al. The impact of quality control on cortical morphometry comparisons in autism. Imaging Neurosci. 2023;1:1–21.
39. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A. 2000;97:11050–5.
40. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31:968–80.
41. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage. 2018;167:104–20.
42. Schabdach JM, Schmitt JE, Sotardi S, Vossough A, Andronikou S, Roberts TP, et al. Brain Growth Charts for Quantitative Analysis of Pediatric Clinical Brain MRI Scans with Limited Imaging Pathology. Radiology. 2023;309: e230096.
43. van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9:2579–605.
44. pamk: Partitioning around medoids with estimation of number of clusters in fpc: Flexible Procedures for Clustering https://rdrr.io/cran/fpc/man/pamk.html Accessed 18 Mar 2024
45. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Stat Methodol. 1995;57:289–300.
46. Nordahl CW, Lange N, Li DD, Barnett LA, Lee A, Buonocore MH, et al. Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. Proc Natl Acad Sci. 2011;108:20195–200.
47. Rubia K, Alegria A, Brinson H. Imaging the ADHD brain: disorder-specificity, medication effects and clinical translation. Expert Rev Neurother. 2014;14:519–38.
48. Kasparek T, Theiner P, Filova A. Neurobiology of ADHD From Childhood to Adulthood: Findings of Imaging Methods. J Atten Disord. 2015;19:931–43.
49. AlbajaraSáenz A, Villemonteix T, Massat I. Structural and functional neuroimaging in attention-deficit/hyperactivity disorder. Dev Med Child Neurol. 2019;61:399–405.
50. Shaw P, De Rossi P, Watson B, Wharton A, Greenstein D, Raznahan A, et al. Mapping the Development of the Basal Ganglia in Children With Attention-Deficit/Hyperactivity Disorder. J Am Acad Child Adolesc Psychiatry. 2014;53:780-9.e11.
51. Vandewouw MM, Brian J, Crosbie J, Schachar RJ, Iaboni A, Georgiades S, et al. Identifying Replicable Subgroups in Neurodevelopmental Conditions Using Resting-State Functional Magnetic Resonance Imaging Data. JAMA Netw Open. 2023;6: e232066.
52. Nordahl CW, Andrews DS, Dwyer P, Waizbard-Bartov E, Restrepo B, Lee JK, et al. The Autism Phenome Project: Toward Identifying Clinically Meaningful Subgroups of Autism. Front Neurosci. 2022;15: 786220.
53. Lombardo MV, Pierce K, Eyler LT, Carter Barnes C, Ahrens-Barbeau C, Solso S, et al. Different Functional Neural Substrates for Good and Poor Language Outcome in Autism. Neuron. 2015;86:567–77.
54. Lombardo MV, Pramparo T, Gazestani V, Warrier V, Bethlehem RAI, Carter Barnes C, et al. Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. Nat Neurosci. 2018;21:1680–8.
55. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014;11:333–7.
56. Gao L, Qiao S, Zhang Y, Zhang T, Lu H, Guo X. Parsing the heterogeneity of brain structure and function in male children with autism spectrum disorder: a multimodal MRI study. Brain Imaging Behav. 2025;19:407–20.
57. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry. 2014;19:659–67.
58. HD-200 Consortium. The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. Front Syst Neurosci. 2012;6:62.
59. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci Data. 2017;4: 170181.
60. Lai MC, Lombardo MV, Suckling J, Ruigrok ANV, Chakrabarti B, Ecker C, et al. Biological sex affects the neurobiology of autism. Brain J Neurol. 2013;136:2799–815.
61. Ecker C. The neuroanatomy of autism spectrum disorder: An overview of structural neuroimaging findings and their translatability to the clinical setting. Autism. 2017;21:18–28.
62. Zeestraten EA, Gudbrandsen MC, Daly E, De Schotten MT, Catani M, Dell'Acqua F, et al. Sex differences in frontal lobe connectivity in adults with autism spectrum conditions. Transl Psychiatry. 2017;7: e1090.

Publisher's Note