bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Delineating Transdiagnostic Subtypes in Neurodevelopmental Disorders via Contrastive Graph Machine Learning of Brain Connectivity Patterns

Xuesong Wang[1], Kanhao Zhao[2], Lina Yao[1,3], Gregory A. Fonzo[4], Theodore D. Satterthwaite[5], Islem Rekik[6], and Yu Zhang[2,7,*]

[1]Data 61, Commonwealth Scientific and Industrial Research Organisation, New South Wales, Australia
[2]Department of Bioengineering, Lehigh University, Bethlehem, PA, USA
[3]School of Computer Science and Engineering, University of New South Wales, New South Wales, Australia
[4]Center for Psychedelic Research and Therapy, Department of Psychiatry and Behavioral Sciences, Dell Medical School, The University of Texas at Austin, Austin, TX, USA
[5]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, PA, USA
[6]BASIRA Lab, Imperial-X and Department of Computing, Imperial College London, London, UK
[7]Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, USA
[*]Corresponding author, email: yuzi20@lehigh.edu

ABSTRACT

Neurodevelopmental disorders, such as Attention Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD), are characterized by comorbidity and heterogeneity. Identifying distinct subtypes within these disorders can illuminate the underlying neurobiological and clinical characteristics, paving the way for more tailored treatments. We adopted a novel transdiagnostic approach across ADHD and ASD, using cutting-edge contrastive graph machine learning to determine subtypes based on brain network connectivity as revealed by resting-state functional magnetic resonance imaging. Our approach identified two generalizable subtypes characterized by robust and distinct functional connectivity patterns, prominently within the frontoparietal control network and the somatomotor network. These subtypes exhibited pronounced differences in major cognitive and behavioural measures. We further demonstrated the generalizability of these subtypes using data collected from independent study sites. Our data-driven approach provides a novel solution for parsing biological heterogeneity in neurodevelopmental disorders.

Introduction

Attention deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are neurodevelopmental disorders that typically manifest in childhood, impact social function, and impair academic achievement[1,2]. ADHD across clinical subtypes is characterized by inattention, hyperactivity and impulsivity, while ASD presents with deficits in social communication and interaction accompanied by restricted, repetitive behaviors, and sensory symptoms[3]. Despite these seemingly independent core diagnostic criteria, there is a significant clinical overlap between ADHD and ASD[2,4]. Evidence from research indicates shared genetic risk factors and neuropsychological vulnerabilities, particularly in executive function and emotional recognition domains[2,5,6]. Adding to the complexity, both ADHD and ASD diagnoses, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases, encompass a wide spectrum of symptom combinations. This symptom-based approach results in significant clinical heterogeneity, which aggregates patients with diverse symptoms, severity and mechanisms within a single medical diagnosis[7,8]. Beyond such clinical heterogeneity within each disorder, there is a significant co-morbidity of the two conditions: an estimated 30 to 50 % of individuals diagnosed with ASD also meet the criteria for ADHD[9], and roughly 14% of the children with ADHD also have ASD[10]. By definition, co-morbidity suggests the co-occurrence of many diseases in any one patient may share similar causal mechanisms, which can not only exhibit distinct neurofunctional deficits but also complicate the treatments of patients[11]. Together, this combination of heterogeneity and co-morbidity complicates the task of discerning the underlying mechanisms of these disorders and devising effective interventions[12]. As such, there is an urgent need to identify subtypes within these broad clinical categories.

Research using functional magnetic resonance imaging (fMRI) to study brain functional connectivity may provide a promising avenue for identifying subtypes that explain biological heterogeneity in ADHD and ASD. Previous investigations have highlighted that resting-state fMRI connectivity involving the ventral attention network can predict treatment response in

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ADHD[13] and assist in ASD diagnosis[14]. A meta-analysis further uncovered abnormalities in functional connections involving the frontal and posterior cingulate cortices from studies of group-level comparison between both ADHD and ASD with a healthy population[15]. Nonetheless, many studies have traditionally examined functional connectivity features independently (via vectorization), overlooking the potential benefits of considering the connectivity topology in understanding disorder vulnerability and resilience[16]. In response to this gap, some research efforts have shifted to analyze high-level brain graph features, such as participation coefficient and connectedness coefficient derived from the binarized functional connectivity graph[17,18]. Yet, focusing solely on these graph features may overlook essential global brain network properties. To utilize the complete brain graph information for downstream brain dysfunction analysis, deep learning-based end-to-end graph convolutional networks (GCNs) offer substantial advances in brain network representation, holistic mapping, diagnosis and generation[19]. Increasing numbers of research studies have shown the promising diagnostic ability of GCNs for ADHD and ASD[20,21]. Nevertheless, most of the existing studies have been conducted under a case-control framework by examining the differences between a single disease and healthy controls, overlooking the considerable heterogeneity and comorbidity inherent in ADHD and ASD[22].

Building on the hypothesis that ADHD and ASD present varying manifestations of an overarching neurodevelopmental disorder[23], our objective was to uncover novel transdiagnostic subtypes within these diagnoses. To this end, we employed graph neural networks to extract diagnostic-independent representations shared across both disorders. Recent advancements in self-supervised learning, especially contrastive learning, have paved the way to discern patterns from high dimensional functional connectivity data without being anchored to clinical diagnoses[24,25]. Unlike supervised learning models that optimize parameters based on categorical clinical labels, contrastive learning establishes a metric that enhances similarities between two functional connectivity representations derived from the same individual while fostering contrasts between representations from various individuals. Integrating contrastive learning with GCNs has the potential to delineate shared functional connectivity patterns between ADHD and ASD, while also parsing heterogeneity within each clinical diagnostic group. Accordingly, we developed a contrastive graph convolutional model to extract the ADHD and ASD representative graph features from brain functional connectivity networks. Our modelling framework learns disease-agnostic representations, which then serves as the foundation for building a clinical population graph spanning both ADHD and ASD. Using spectral clustering on the graph data, we were able to delineate distinct transdiagnostic subtypes. With three large-scale resting-state fMRI cohorts, including ADHD-200[26], Autism Brain Imaging Data Exchage (ABIDE)-I and ABIDE-II[27,28], our approach successfully identified two robust subtypes spanning both ADHD and ASD. Furthermore, we evaluated differences in independent functional and clinical measures that were not used in subtype identification. We also ensured the generalizability of our findings by conducting subtyping analyses on a separate, held-out dataset. Collectively, our approach seeks to fully harness the brain graph information, revealing novel and insightful transdiagnostic subtypes to disentangle the heterogeneity and comorbidity of ADHD and ASD. In doing so, we provided initial evidence for the neuropathological categorization of neurodevelopmental disorders, thus advancing the development of precision mental health.

Results

We first conducted the contrastive brain graph analysis and subtype identification using composite discovery data and then validated our results using held-out data. As we resolve to analyze the transdiagnostic relationships, we strategically combined two multi-sites datasets that cover ASD or ADHD patients for discovery analysis. The composite discovery data (including 268 ADHD, 531 ASD, and 881 neurotypical controls) combines a subset of ADHD-200 and the whole ABIDE-I. This subset of ADHD patients was selected from three sites with relatively balanced patient-to-healthy-control ratios. The ratio of training-validation-testing set was 7:1:2 and we tested the average model performance by randomly splitting dataset using six random seeds. The subtyping algorithm was trained on the training subset of the discovery set (N=1176), and the best model was saved by evaluation on the validation subset (N=168) and tested on the testing subset (N=336), while the subtype labels were assigned to all of the train-validation-test subsets. Such process was implemented on the discovery set (ABIDE-I and ADHD-discovery set). To further verify the transferability of the acquired subtypes, we introduced a held-out dataset (including 45 ADHD, 453 ASD, and 744 neurotypical controls) which contains the remaining site patients of ADHD-200 and the whole ABIDE-II dataset. This composite held-out dataset was employed for prediction rather than training the subtyping algorithm (see Participants section for more details).

Contrastive functional connectivity representation defines two subtypes across ADHD and ASD

In order to learn generalizable subtypes across ASD and ADHD, the proposed model should focus on transdiagnostic features rather than clinical labels. In light of this, our proposed contrastive graph learning approach (see Supplementary Figure 1 and section Methods) was applied to brain graphs constructed using functional connectivity together with personal characteristic data (PCD) on the clinical population. The functional connectivity was extracted using Schaefer parcellation of 100 regions of interest (ROIs). Site effects were mitigated by applying the well-established statistical harmonization tool ComBat [29]

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to functional connectivity inputs meanwhile preserving biological variability such as age, sex, handedness, and three IQ measures. The extraction of graph representations was based on self-resemblance of two non-overlapping fMRI series acquired for the same subject, meanwhile maximizing intra-subject contrast regardless of their clinical labels. Neurotypical controls were also utilized to guide the learning of disease-specific contrastive graph representation. We then employed the resulting representations to induce a population graph whose adjacency matrix was computed using similarities between subject features. Spectral clustering along the population graph was further conducted to identify transdiagnostic subtypes. As the graph representation was learnt in a diagnosis-agnostic manner, the resulting subtypes can be applicable to both ASD and ADHD subjects.

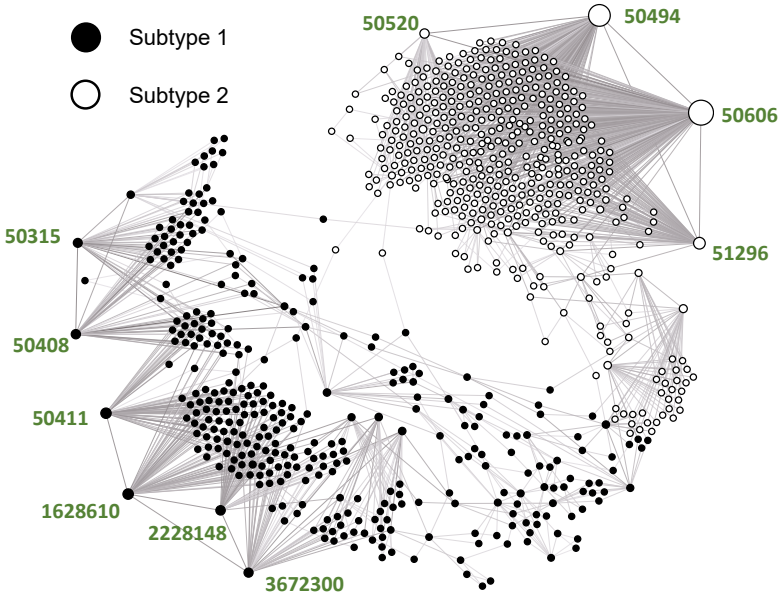


Figure 1. Contrastive functional connectivity representation identifies two distinct subtypes across ADHD and ASD. Each node represents an individual in the ADHD and ASD population, whereas the edge connecting two individuals indicates strong feature similarities between their contrastive features extracted by our proposed method. The node size depicts the degree of each node and the node colors differentiate the subtype labels. The colored number indexes represent the subject names of those "hub patients" who share the most feature similarities with other individuals.

Our approach discovered two distinct subtypes across ADHD and ASD populations(Figure 1). where the optimal number of the subtypes was suggested by the highest Calinski-Harabasz index for the clustering algorithm (Supplementary Figure 2). To understand which neural circuits are crucial for identifying the two subtypes, we examined connectivity differences in terms of gradient-based feature weights (see section Methods for more details of the feature weights calculation). Gradient-based methods were originally proposed to improve interpretability of deep learning methods[30]. For instance, recent work[31] computed the partial derivatives of the Austism labels with respect to the individual functional connectivity links from the brain connectom to identify discriminative functional connections. Many studies have emerged since to adopt gradient-based methods for brain imaging analysis[32–34]. We hereby compute the gradients of our subtyping model predictions to individual connectivity and group the resulting connectivity feature weights by subtyping labels. The resulting mean connectivity feature weights (Figure 2) revealed critical neural circuits that contributed to defining each of the two subtypes. As the number of the identified subtypes is 2, the gradient weight for one connectivity would be positive for one subtype and negative for the other subtype, indicating such connectivity is contributing positively to one subtype and has a negative contribution to identifying the other subtype. Hence, the resulting feature weight map on two subtypes are inversely correlated. The neural circuit pattern of subtype 1 mainly involved the somatomotor network (SMN), visual network (VN), ventral attention network (VAN), and dorsal attention network (DAN). Both the inter- and intra-network connections of these networks have acquired large feature weights (Figure 2 a, b, c). By contrast, the neural circuit pattern of subtype 2 was dominated by both inter- and intra-network connections involving the default mode network and frontoparietal control network (Figure 2 a, b, c). As the model is capable of predicting the individual connectivity feature weights, we further examined the subtype differences by computing two-sample t-tests on the where the feature weights of one subject was treated as a sample in a subtype group. Such comparison can provide a reliable measure as opposed to the simple difference of the averaged feature weights of two subtypes. Our results showed that connectivity gradients captured significant differences in the crucial networks of subtype 1 including SMN, VN, VAN, and DAN (Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a), which were not identifiable using raw connectivity features (Figure 4 b). The mean FC inputs show great resemblance (Supplementary Figure 4), which justifies the infeasibility of detecting subtypes on raw connectivity features.

To explore whether our subtyping method can be used to identify similar subtypes in individual diagnoses, we evaluated our analysis pipeline on ABIDE-I and the ADHD-discovery set separately. The experiment aims to test whether both subtypes can be identified separately in ASD and ADHD. Again, two subtypes were discovered from each dataset (Supplementary Figure 5 a,b). We quantified the similarity between these subtypes by computing Pearson correlation coefficients between the individual connectivity feature weights of each subtype and the weights of two transdiagnostic subtypes (see Supplementary Table 2). One of subtypes obtained from the individual ABIDE-I dataset suggested moderate resemblance towards transdiagnostic subtype 1 with a significant correlation coefficient of 0.34 (Supplementary Table 2 ABIDE-I subtype 2) and emphasis on connectivities within SMN, VN, and DAN (Supplementary Figure 5 a). For the ADHD-discovery set, both identified subtypes from this individual set (Supplementary Figure 5 b) showcased strong subtype similarities with the corresponding transdiagnostic subtypes with high correlation coefficient ($r = 0.58$, $p < 10^{-5}$ for subtype 1 and $r = 0.66$, $p < 10^{-5}$ for subtype 2 (see Supplementary Table 2 ADHD discovery set subtype 1 and 2). The crucial networks discovered by ADHD-discovery set subtype 1 are SMN, VN, DAN, while the subtype 2 showed distinct connectivity patterns in DMN and FPCN. Both echo the conclusions obtained on the transdiagnostic subtypes.

To further validate whether our model can be used for identifying resemblant subtypes from unseen patients with ASD or ADHD, we further tested our pipelines on held-out data comprising ABIDE-II and ADHD-validation set. The resulting FC gradients of both subtypes in ABIDE-II (Figure 3) are highly correlated with the corresponding subtypes in the transdiagostic discovery set with positive Pearson coefficients ($r = 0.54$, $p < 10^{-5}$ for subtype 1 and $r = 0.48$, $p < 10^{-5}$ for subtype 2, see Supplementary Table 2 ABIDE-II results), whereas the gradient of the ADHD-validation set did not support such conclusion (Supplementary Table 2 ADHD-validation set results). Such divergence may be attributed to the small sample size since only 45 patients are available from the ADHD-validation set. Overall, transdiagnostic subtype 2 shared higher correlation coefficients on the held-out set, which showed distinct patterns on DMN and FPCN.

Subtype differences on demographic, cognitive and behavioral data

We further examined differences between identified subtypes in demographic, cognitive and behavioral measures. Three demographic features between the two diseases were firstly compared: age, gender, handedness (Supplementary Figure 6 a,b). Subtype 1 is younger than subtype 2 with significant difference(Supplementary Table 1). We then investigated differences between subtypes in the cognitive and behavioral assessments for the discovery set. Three transdiagnostic behaviour measures are included as training features: full-scale intelligence quotient (FIQ), verbal intelligence quotient (VIQ), and performance intelligence quotient (PIQ). Patients in subtype 2 achieved higher scores on these IQs (see Supplementary Table 1, 5, 4). Several disease-specific behavioral measures were also compared between subtypes. For the ADHD-discovery set, the two subtypes behave similarly with $p > 0.05$ on behavioral scores including ADHD index, inattentive and hyperactivity / impulsivity scores and their medical status (Supplementary Table 5). For the ABIDE-I dataset, subtype 1 constitutes the major ASD and Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) patients while in subtype 2 the dominant patients are Aspergers and Aspergers or PDD-NOS groups (Supplementary Table 4). Within ABIDE-I, subtype 2 showed significantly lower scores on three ASD clinical measures: Abnormality of Development Evident at or Before 36 Months (D) Total (ADI_R_ONSET_TOTAL_D), Classic Total Autism Diagnostic Observation Schedule Score (Communication subscore + Social Interaction subscore, ADOS_TOTAL) and Social Total Subscore of the Classic ADOS (ADOS_SOCIAL) (Supplementary Table 4). Subtype 2 also achieved significant higher scores on most of cognitive measures (starting with WISC_*, Wechsler Intelligence Scale for Children) except for WISC_IV coding scaled.

We further examined the demographic, cognitive and behavioral results on the held-out datasets (ABIDE-II and the ADHD-validation set). For ABIDE-II, subtype 1 has significant younger populations which echoed the conclusion of the transdiagnostic discovery set (Supplementary Table 6). The results from the DSM_IV measurements were also consistent with the discovery set where significant differences are observed between subtype 1 and 2 with more ASD participants in subtype 1, and higher Aspergers proportion in subtype 2. Similar to the discovery set, significant differences were recognized on FIQ and VIQ but not PIQ (Supplementary Table 6). Three clinical measurements on the held-out datasets also echoed the conclusion of the discovery set (ABIDE-I) where subtype 1 achieved higher scores: ADI_R_ONSET_TOTAL_D, ADOS_G_TOTAL, and ADOS_G_SOCIAL. Surprisingly, we discovered that subtype 1 has achieved significant higher scores on two ADHD-related clinical measurements: CBCL 6-18 Attention Problems T Score and CBCL 6-18 Attention Deficit/Hyperactivity Problems T Score.

In order to verify whether subtypes defined by the connectivity features can be identified using the demographic, cognitive and behaviour features alone, we ran a K-means clustering algorithm using the following six features that are transdiagnostic between ADHD and ASD: gender, handedness, age and three IQs: FIQ, VIQ, and PIQ (See Supplementary Figure 7 a, b). While these features shared an high accuracy on the ADHD-discovery set, the validation results on the ADHD-replication set

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and ABIDE-II are only 60% and 66% respectively as apposed to our results of 68.9% and 79.9%, suggesting that the subtypes acquired by PCD features were less transferable and replicable than our contrastive features.

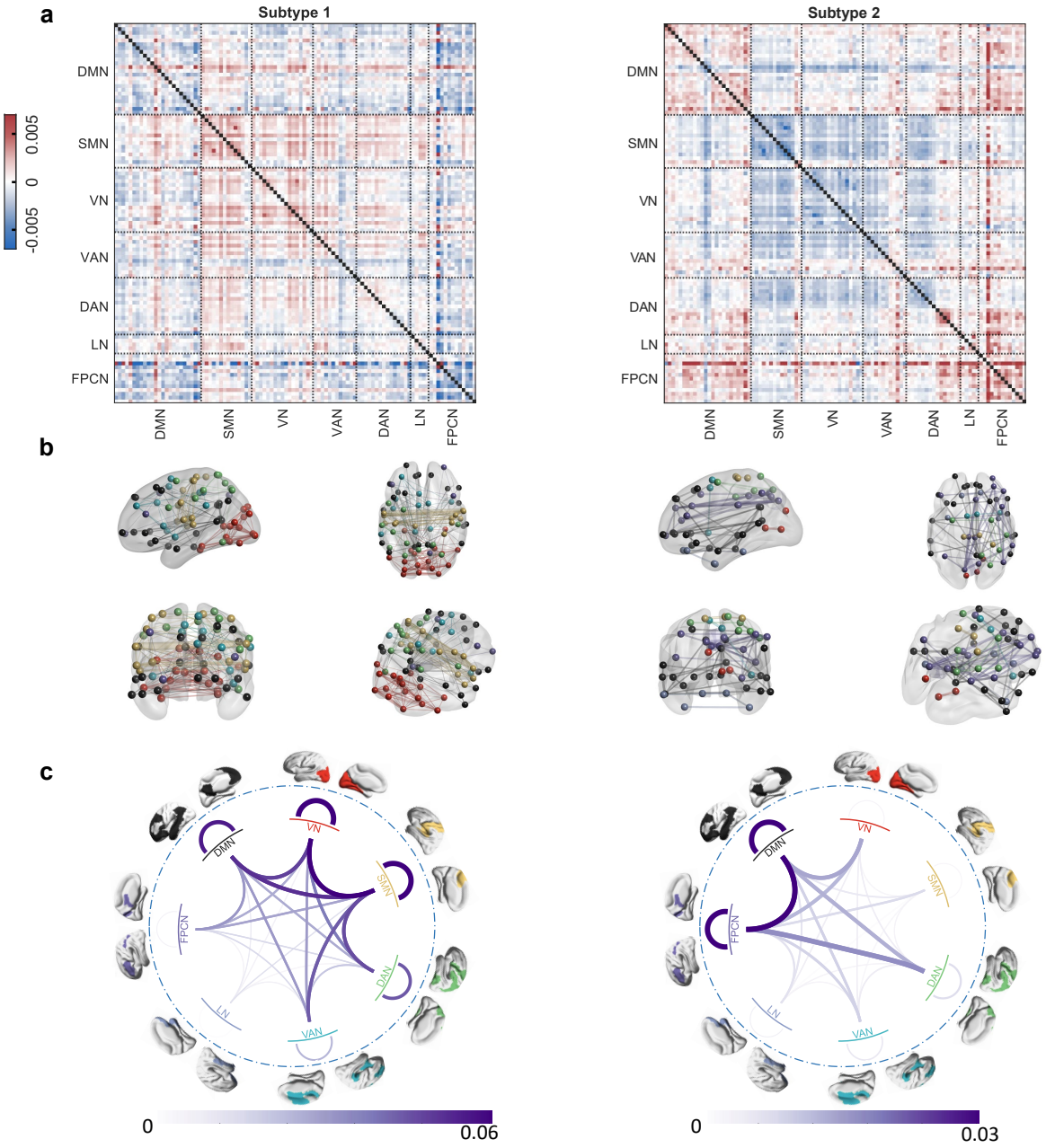


Figure 2. Mean gradient-based connectivity feature weights of the two subtypes on the transdiagnostic discovery set comprised of ABIDE-I and the ADHD-discovery set. The gradients are obtained from a subtyping classifier trained on the observed subtypes as ground truth, **a.** Mean connectivity gradient matrices where a positive gradient indicates the positive contribution of a connectivity feature to the subtype. The ROIs are reordered by seven brain networks: DMN: default mode network; SMN: somatomotor network; VN: visual network; VAN: ventral attention network; DAN: dorsal attention network; LN: limbic network; FPCN: frontoparietal control network. **b.** Visualization of connectivity patterns on the brain. Each ROI is represented by a node whose color was decided by the corresponding network. The edge width indicates absolute gradient value for a specific connectivity. The edges with strength below the 0.03 $\times$ maximum value are hidden for visual clarity. **c.** Inter- and Intra connectivity gradients on seven brain networks. The cord width describes the network-level connectivity gradients. The network color matched with the visualization on the brain.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

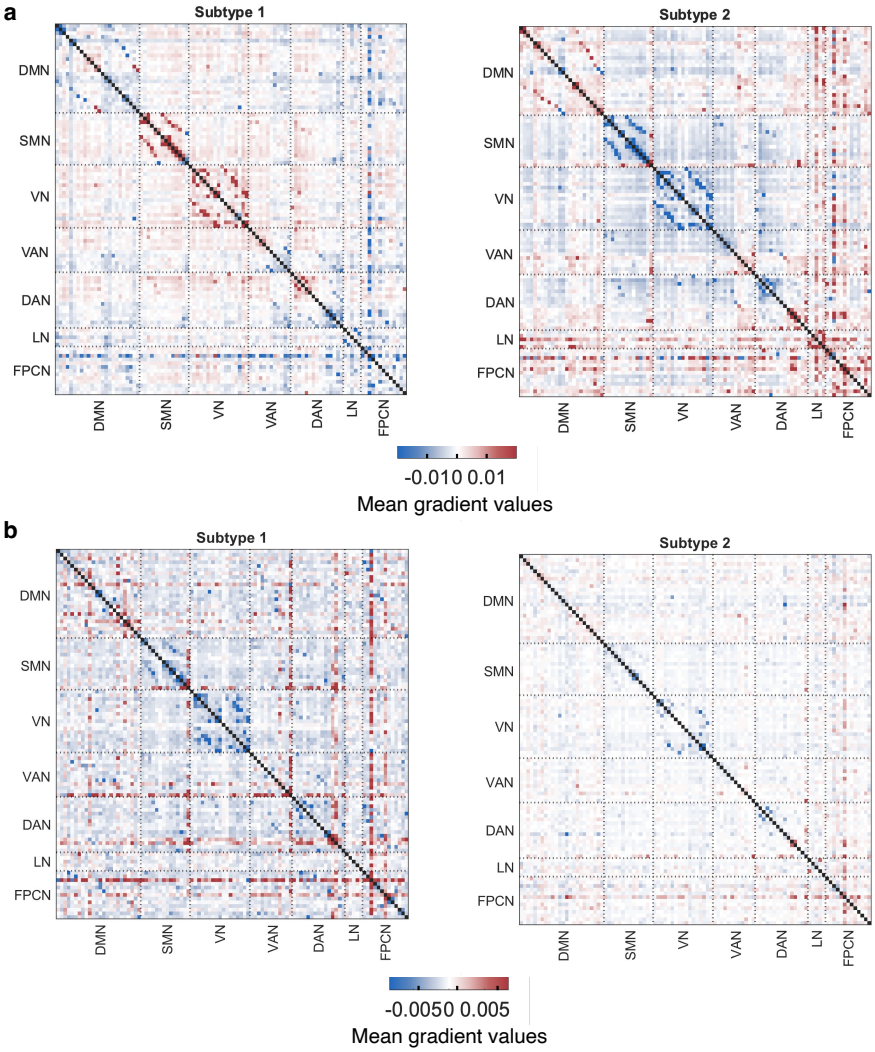


Figure 3. Connectivity gradients of two subtypes in individual replication sets. The gradients are obtained by retyping on two replication sets separately. Positive gradients indicate a positive contribution of connections to the subtype, while negative gradients signify a negative contribution **a**. Mean connectivity gradients of the two subtypes on ABIDE-II. **b.** Mean connectivity gradients of the two subtypes on the ADHD-validation set.

ASD and ADHD population graph investigation

As the node sizes depict the node degree in the population graph (Figure 1) and the edges represent strong subject similarities, a large node with many connecting edges suggested that this particular patient showed resemblance with many other patients on contrastive features and therefore their subtype labels would affect the results for most of the patients in the population graph. We therefore investigated whether the "hub patients" functional connectivity, demographic or biological features contribute to the subtyping results. We sorted the patients by their node degree and selected the top-10 hubs in both subtypes. Most Subtype 2 patients were linked to four hubs from ABIDE-I: 50606, 50494, 50520, 51296 (subject id), whereas Subtype 1 patients are connected by more scattered hubs from both ADHD-discovery set and ABIDE-I patients: 1628610, 3672300, 2228148 from ADHD-discovery set and 50411, 50408, 50315 from ABIDE-I. We first placed the demographic and behavioral features of these hubs into the corresponding distributions of the non-hub patients and evaluated their likelihood (Supplementary Figure 8). Two patients 50606 and 50494 in subtype 2 were identified as outliers, since either their ages were substantially older or they achieved much lower IQ scores. We also compared hub patients' raw FC inputs with the average connectivities of non-hub patients in the corresponding subtype (Supplementary Figure 9). Most of the hub patients were found to have marginally higher connectivities on the visual networks or somato-motor networks than non-hub patients.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

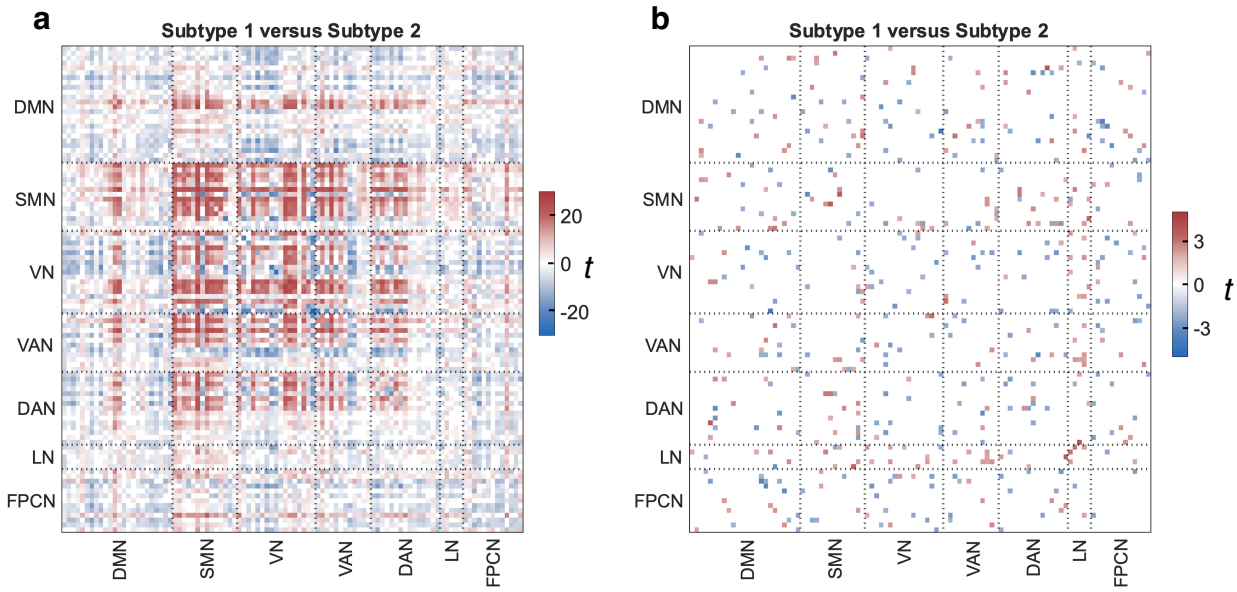


Figure 4. Connectivity differences between the two subtypes detected by independent t-test (FDR-corrected P < 0.05.a. Significant difference of gradient of raw functional connections (partial correlation). **b.** Significant difference of raw connectivity features.

Validation of the subtype transferability across datasets

To investigate whether the subtypes acquired from the transdiagnostic population are transferable to unseen patients, we assigned labels for new individuals with two solutions: either by choosing the subtype of the most similar subject in the transdiagnostic population as the prediction or prediction using a classifier trained by regarding transdiagnostic subtypes as ground truth. First, we applied the same subtyping pipeline to individual diagnoses and map their subtyping results to the transdiagnostic discovery set (see Supplementary Figure 3). For subtypes acquired from ADHD alone, patients in ASD in the transdiagostic popultion were unseen by the model and therefore can be used for validation purpose. Both subtypes from the transdiagnostic discovery set were recognized in the individual diseases. To assess the robustness of the obtained subtypes, we conducted several transferability experiments on individual discovery set (ABIDE-I and ADHD-discovery set) as well as replication set (ABIDE-II + ADHD-validation set) (see Figure 5). We initially evaluated the transferability from the transdiagnostic discovery set to individual discovery sets. The subtypes obtained from the trans-discovery set (Figure 1) were utilized as predictions and the ground truth subtypes were acquired from training and testing on individual discovery sets, i.e., ABIDE-I, or ADHD-discovery set. The result on the ADHD-discovery set achieved a higher accuracy of 85.8% compared with 79.7 % on the ABIDE-I, which indicates that each subtype can be identified in individual diagnoses with high transferable accuracy (see the confusion matrix in Figure 5 a).

Additionally, we validated the subtype findings using ABIDE-II and ADHD-validation set to confirm the replicability of the identified subtypes. The analysis was implemented from two perspectives, KNN-based method and subtyping classifier. KNN-based method employed the trans-discovery subjects as neighbours to label subtypes of the unseen patients, whereas subtyping classifier explicitly train a classifier on the trans-discovery labels to predict subtyping labels for new patients. Transferability is more robust in ABIDE-II than in ADHD-validation set, where the average accuracy for ABIDE-II was 79.9% for KNN method (Figure 5 b) and 80.6% for subtyping classifiers (Figure 5 c), and the average accuracy on ADHD-validation set was 68.9% for KNN-based method (Figure 5 b) and 60% for classifiers (Figure 5 c). Such difference may be attributable to the small dataset size of ADHD that consequently may generate less reliable subtyping labels.

Discussion

Heterogeneity and comorbidity have been frequently reported in ADHD[35–37] and ASD[38], which precludes delivery of effective and customized treatment to distinct subgroups within the same clinical diagnosis. In light of this, we have developed a novel subtyping framework utilizing contrastive graph machine learning, which successfully identified two transdiagnostic subtypes characterized by differing functional connectivity gradients, and cognitive and behavioral profiles. In particular, subtype 1 exhibited significantly increased FC gradients in somatomotor network, visual network, ventral attention network and dorsal

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

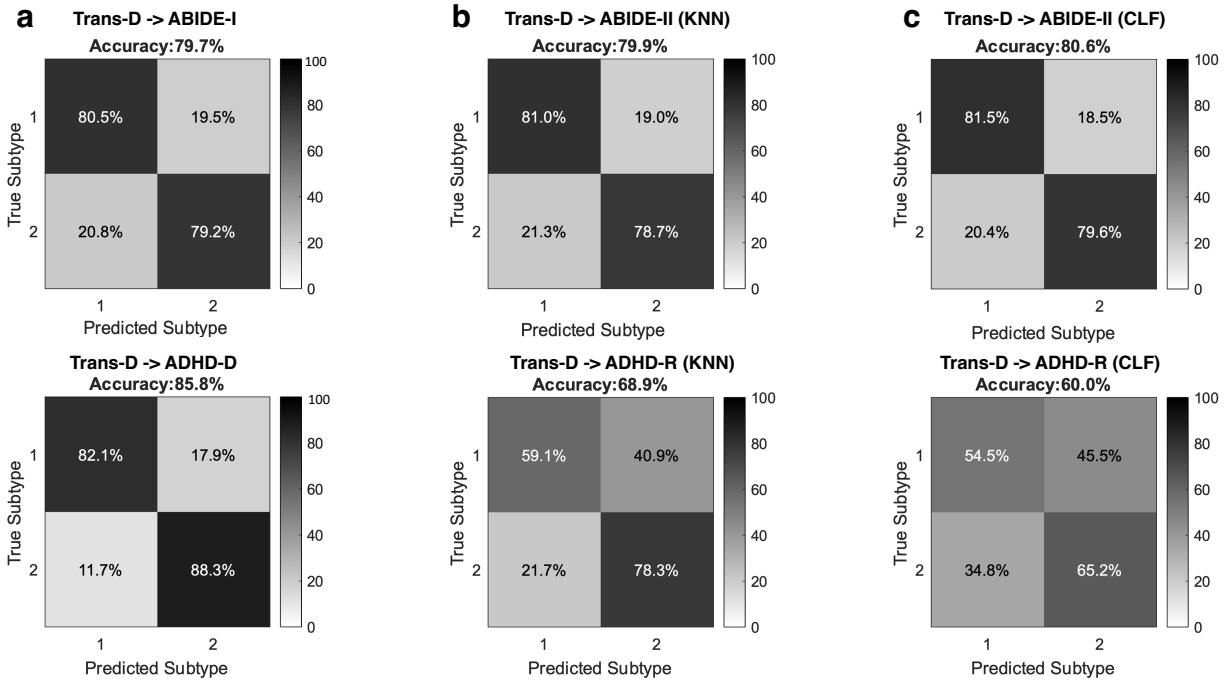


Figure 5. Confusion matrix for model transferability. Subtypes obtained from the transdiagnostic discovery set (Trans-D) are utilized to predict the subtypes on the individual discovery set and replication set. **a.** Confusion matrix from Trans-D to ABIDE-I and ADHD-discovery set (ADHD-D). We run the same pipeline on the individual disease and obtained their subtypes as ground truths and the subtypes from the transdiagnostic population are treated as predictions. **b.** Confusion matrix on the replication dataset of ABIDE-II and ADHD-validation set (ADHD-R) using KNN-based method. We run the same pipeline on the replication set to obtain subject embeddings, the subtyping predictions are then assigned based on their neighbours in the trans-D set. The ground truth is obtained by running the pipeline on the replication set separately. **c.** Confusion matrix on ABIDE-II and ADHD-validation set using a subtyping CLassiFier (CLF) which was trained using the subtyping labels from the Trans-D set.

attention network. Subtype 1 also showed significantly lower cognitive scores (including FIQ, VIQ and PIQ) than subtype 2 and healthy controls. For ASD patients in subtype 1, they displayed higher total ADOS scores, where an ADOS score ranging from 30-36 suggests mild to moderate autism, and a score greater than 37 is classified as severe autism[39]. Subtype 1 patients also achieved lower cognitive behaviour scores with the WISC assessments including Verbal Comprehension Index (VCI), Perceptual Reasoning Index(PRI), Working Memory Index (WMI) and the Processing Speed Index (PSI), where a lower score below 90 suggests below average intellectual ability. In contrast, subtype 2 exhibited significantly increased FC gradients in the default mode network and frontoparietal control network. Subtype 2 displayed clinical profiles that resembled those of healthy controls. Furthermore, the validation of these subtypes on the independent dataset ABIDE-II and ADHD200 indicated high transferability across datasets. In particular, both subtypes showed consistent functional connectivity gradients and intellectual capacity measurements, such as FIQ and VIQ, between the discovery data and held-out data. These findings provide further support for underlying neurobiological mechanisms that may contribute to substantial comorbidity between ASD and ADHD. These insights have the potential to facilitate future precision treatment efforts through stratifying individuals by subgroups across diseases that have more homogeneous characteristics, which may, in turn, improve treatment outcomes if specific treatments are better suited to a particular subtype. This moves the field going beyond reliance solely on clinical diagnoses and trial-and-error treatment approaches.

Our findings provide valuable insight into the heterogeneity of functional connectivity observed in individuals with ASD and ADHD, confirming previous research and highlighting distinct neural circuit patterns associated with specific subtypes. For ASD, our identified subtype 2 confirms previous evidence of brain dysfunction in DMN[40–42]. A study[40] associated DMN with the main social deficit in children with ASD and identified fundamental aspects of posteromedial cortical heterogeneity. Interestingly, two recent studies of functional connectivity heterogeneity in ASD have shown very similar subtypes to ours [43,44]. Other work [43] detected three subtypes in a combined dataset, including ABIDE-II. Similar to this study, their important

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

connectivities in subtype 1 and subtype 2 are almost complementary with subtle differences. Regarding ADHD, we found supporting evidence from various studies that pinpoint highly related regions[45–52]. These studies have consistently identified functional connectivities within and between DMN and FPCN, which is consistent with the patterns observed in our subtype 2. One study [45] also suggests hypoconnectivity in VAN, DAN, and SMN, which we observed in our subtype 1. There is also an increasing trend to parse heterogeneity within the ADHD diagnosis. For instance, three discrete subtypes are identified by Damien A 2015[53] on a group of children with and without ADHD. Atypicality in DMN is discovered in all three subtypes. Their subtype 1 is specifically connected to regions of DAN and VN which parallels our subtype 1. Altogether, these findings justify the importance of pursuing subtyping efforts on the ADHD and autism diagnoses and converge with the findings in this investigation regarding the distinct networks or regions associated with specific subtypes.

Emerging evidence points to a shared behavioral dysfunction between ADHD and ASD, as highlighted by various studies[54–56]. Our observations regarding FPCN dysfunction in subtype 2 find support from several significant research efforts. For instance, literature [6] posited that the impairment of executive function could be a shared root for neurodevelopmental disorders like ADHD and ASD. Additionally, a subsequent study[57] highlighted that executive function characteristics not only explain the co-occurrence of ADHD and ASD, but also underscore the intrinsic heterogeneity across these disorders, thereby hypothesizing subgroups with distinct mechanisms across disorders. This research delineated differences between subtypes and neurotypical controls, particularly in VAN and SMN, which corroborates our findings in subtype 1. While distinctions based on hyperactive/inattention symptoms were discernible, no functional connectivity differences were identified between their subtypes, implying a more complex mechanism underlying these disorders. We surmise that this might stem from the emphasis on direct functional connectivity measurements, rather than on connectivity gradients - an approach that our results affirmed as essential for discerning key connectivity features. In a separate endeavor, Lake et. al.[58] confirmed the significance of FPCN and also pinpointed an accentuated hubness in VN. Furthermore, they identified overlapping domains in representation/recognition, and language or semantics across these disorders. Notably, literature[54] detected shared classification markers between ASD and ADHD across LN, DMN, VAN, and VN. In sum, these insights pave the way for a more comprehensive understanding of both the inherent comorbidity shared aspects between ADHD and ASD.

To assess the association between transdiagnostic subgroups, demographic data, and behavior scores, we examined shared features spanning age, gender, handedness, FIQ, VIQ, and PIQ across diseases. Additionally, disease-specific scores were evaluated to elucidate subtype distinctions in the brain. Intriguingly, neither subtype predominated in ADHD or ASD, suggesting our identification of two genuinely transdiagnostic subgroups. Our age-related findings converge with literature [59] reporting that younger ASD individuals exhibited greater impairment in lateral temporal (a component of VN) and frontal (a part of FPCN) regions. This aligns with the FC gradients we observed in subtype 1. These effects might stem from enhanced neurogenesis, reduced synaptic pruning, diminished neuronal cell death, or abnormal myelination, as noted in literature[60]. Furthermore, another study[59] established a link between FIQ values and CTs, where individuals with lower FIQ (akin to subtype 1 in our study) exhibited abnormalities in frontal (VAN, DAN), temporal (VN and part of SMN), and occipital (VN, SMN) cortices.

Regarding the disease-specific symptoms, both subtypes presented similarly in terms of the ADHD index, inattention, and hyper-impulsivity scores. We did not observe the distinct patterns suggested in literature[57] for inattention and hyperactivity/impulsivity scores. In contrast, these subtypes showed varied results for ASD clinical diagnoses and behaviour scores. Our analysis suggests the model might capture a broader variance of ASD symptoms compared to ADHD. Specifically, subtype 1 comprises a higher percentage of ASD patients, while subtype 2 predominantly includes Asperger's patients. This observation aligns with Lotspeich (2004)[61] which posited that Asperger's syndrome manifests milder neuroanatomical deviations compared to lower IQ individuals (as seen with ASD patients in our subtype 1). Additionally, the Asperger's subgroup showed fewer abnormalities in Autism-related scores. This was echoed by Anderson (2011)[42], which identified significant correlations between individuals with ASD and behavioral measures such as ADOS_SOCIAL, ADOS_COMM, the Social Responsiveness Scale (SRS), and VIQ. These scores depict verbal language deficits and deficits in social interaction. Another study[62] found an association between atypical FC patterns and clinical severity in ASD, pinpointing regions similar to our findings. They revealed that heightened clinical severity was linked to enhanced connectivity regions within DMN such as the angular gyrus. In contrast, connectivity diminished in areas vital for multimodal integration like the superior frontal and parietal zones of the DAN. Additionally, subtype 2, including primarily Asperger's patients achieved higher scores in IQ and verbal assessments, which showcased the distinction of language skills between Autism and Asperger's syndrome. We also observed a subtype distinction on the Child Behaviour Checklist attention problems subscale (CBCL_6-18) where subtype 2 presented fewer anomalies than subtype 1. A similar study[63] conducted subtyping on a comorbid group of ADHD and ASD patients and found that healthy controls exhibited fewer attention problems relative to both diseases. No discernible difference was detected between the two diagnostic groups, lending further weight to the notion of shared patterns across these disorders.

Several limitations and potential avenues for future exploration emerge from the current study. By selecting multiple scans or segmenting the ROI time series into non-overlapping windows, we created multiple views of the subjects. Such an approach may limit model flexibility, especially when additional scans are unavailable, or the data length is insufficient for segmentation.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Future studies might consider graph topology augmentation as a strategy to generate positive pairs for contrastive learning[64]. Furthermore, integrating behavioral and cognitive measures as training features could lead to biased feature importance on them and consequently weakened the effects of functional connectivity. Solely focusing on functional connectivity in subtyping analyses might overcome such limitations. Finally, given the limitations of data acquired in the same samples utilized here to represent neurodevelopmental disorders, we were unable to assess the treatment effects on the various subgroups or detail their responses, which worth future studies with data from clinical trials.

In summary, our study developed a contrastive graph representation learning framework to identify transdiagnostic subtypes across ADHD and ASD based on rs-fMRI functional connectivity. Our comprehensive analyses revealed two replicable subtypes that showed distinct functional connectivity gradients and associations with intrinsic large-scale brain networks. The neurobiological heterogeneity and comorbidity across the two diseases were underscored by subtype validation on held-out data. These subtypes further demonstrated differences in demographic characteristics and clinical severity. Additionally, both subgroups exhibited diverging patterns of intellectual performance for the ASD patients. Our subtype discovery efforts illuminate the diversity and overlapping features between ADHD and ASD and offer a generalizable methodology for delineating transdiagnostic neurobiological variation in psychiatric populations.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Participants and Data Acquisition

Datasets

The ADHD-200, ABIDE-I, and ABIDE-II cohorts were utilized for our study. The ADHD-200 cohort[26] includes 830 participants recruited from seven different scan sites. To facilitate independent validation of experimental findings, we split the whole dataset into a discovery and a replication sets. The ADHD-discovery set was formed using data from the three sites, including Kennedy Krieger Institute (KKI), Peking University (PKU), and New York University Child Study Center (NYU), since they provided a relatively balanced patient-control ratio in sample size (268 ADHD patients and 310 neurotypical controls). The ADHD-validation set was formed by 45 ADHD patients and 207 neurotypical controls from the remaining four sites: NeuroIMAGE Sample, Oregon Health and Science University (OSHU), University of Pittsburgh, and Washington University in St. Louis. The ABIDE cohort[27,28] contains two independent datasets (ABIDE-I and ABIDE-II). The ABIDE-I dataset including 531 ASD patients and 571 neurotypical controls from 20 different sites was combined with the ADHD-discovery set for our discovery analysis of transdiagnostic subtypes. The ABIDE-II dataset including 453 ASD patients and 537 neurotypical controls from 16 different sites was combined with the ADHD-validation set to validate the generalizability of our subtype findings.

Phenotypical assessments

In the ADHD-200 cohort, phenotypical measures including FIQ, PIQ, and VIQ were assessed on one of the following scales: Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV), Wechsler Abbreviated Scale of Intelligence (WASI), and Wechsler Intelligence Scale for Chinese Children-Revised (WISCC-R). In the ABIDE cohort, FIQ scores are assessed based on one of the following scales: Differential Ability Scales II - School age (DAS-II), WASI, WISC, Wechsler Adult Intelligence Scales (WAIS), Hamburg-Wechsler Intelligence Test for Children (HAWIK-IV), and Wortschatztest(WST). VIQ scores are measured using DAS-II, WASI, WISC, WAIS, Peabody Picture Vocabulary Test (PPVT) and Stanford Binet (STANFORD). PIQ scores are evaluated by referring to DAS-II, WASI, WISC, WAIS, STANFORD and Raven's Standard Progress Matrices (RAVENS).

Functional MRI acquisition and preprocessing

For both the ADHD-200 and ABIDE cohorts, rsfMRI data were collected with varying protocols and scanner parameters specific to each of the study sites. Details of imaging parameters can be found in literature[26–28]. All available rsfMRI data were preprocessed using the well-established fMRIPrep pipeline[65]. The T1-weighted image was corrected for intensity non-uniformity and then stripped skull. Spatial normalization was done through nonlinear registration, with the T1w reference[66]. Using FSL, brain features such as cerebrospinal fluid, white matter, and grey matter were segmented from the reference, brain-extracted T1 weighted image[67]. The fieldmap information was used to correct distortion in low-frequency and high-frequency components of fieldmap. Then, a corrected echo-planar imaging reference was obtained from a more accurate co-registration with the anatomical reference. The blood-oxygenation-level-dependent (BOLD) reference was then transformed to the T1-weighted image with a boundary-based registration method, configured with nine degrees of freedom to account for distortion remaining in the BOLD reference[68]. Head-motion parameters (rotation and translation parameters of volume-to-reference transform matrices) were estimated with MCFLIRT (FSL). BOLD signals were slice-time corrected and resampled onto the participant's original space with head-motion correction, susceptibility distortion's correction, and then resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA)[69] was performed on the preprocessed BOLD time-series on MNI space after removal of non-steady-state volumes and spatial smoothing with an isotropic Gaussian kernel of 6 mm FWHM (full-width half-maximum).

Functional connectivity calculation and brain graph construction

With the preprocessed BOLD signals, we extracted regional time series based on the Schaefer parcellation[70] of 100 regions of interest (ROIs). For each subject, functional connectivity was then calculated between these regional time series using Pearson's correlation and partial correlation, respectively. We regressed out mean framewise displacement from functional connectivity features to alleviate head motion effects and then applied the well-established statistical harmonization tool ComBat [29] to reduce site effects. We included six personal characteristic data (PCD) comprising gender, age, handedness, and three IQ measures (FIQ, VIQ, PIQ) along with functional connectivity for the subsequent graph construction. Following literature[21,71], each subject's brain graph input is defined as $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ where node features for each ROI is $\mathscr{V}_i \in \mathbb{R}^{ROIs+PCD} = \mathbb{R}^{106}$which concatenates the i-th row of the Pearson's correlation matrix with the corresponding PCD. The edge feature $\mathscr{E}_{ij} \in \mathbb{R}$ indicates the partial correlation connectivity between ROI i and j. A range of sparsity values are tested on the partial correlation and the optimal sparsity is set as 0.6 to trade off between efficiency and precision. As contrastive learning requires multiview inputs

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from the same subject to form a positive pair, we generate three graphs for each subject using either multiple scans of the same subject or three non-overlapping ROI slices if the subject only had a single scan.

Contrastive graph representation learning

Our proposed contrastive representation framework consists of two components (Supplementary Figure 1): contrastive graph learning (CGL) and dynamic graph classification (DGC). CGL aims to train a self-supervised neural network that can represent high dimensional subject features. DGC utilizes diagnostic label information to further improve the representation by quantifying patient-specific variations.

The motivation behind CGL is that two functional connectivity views coming from the same subject should form a "homogeneous" pair whereas different patient views should form a "heterogeneous" pair. More specifically, we can treat multiple fsfMRI slices obtained from one individual as different views of such subject, and therefore the functional connectivity extracted by those slices should share higher similarities than two random functional connectivity matrices chosen from two different subjects. Such inductive bias avoids the adoption of labels to train deep networks, which effectively alleviates overfitting and has been testified in many scenarios[72]. For implementation, the brain graph $\mathscr{G}$ was firstly passed to several spectral graph convolutional networks[73] defined as: $\mathscr{V}' = \sum_{k=1}^{K} Z^k(\mathscr{V}) \cdot \theta^k$ where θ^k is a neural network and $Z^k(\mathscr{V})$ is associated with the Laplacian matrix $\mathscr{L}$ of the graph $\mathscr{G}$. k denotes the filter size. Following[74], K = 3 for computational efficiency, with $Z^3(\mathscr{V}) = 2\tilde{\mathscr{L}}Z^2(\mathscr{V}) - Z^1(\mathscr{V})$, $Z^1(\mathscr{V}) = \mathscr{V}, Z^2(\mathscr{V}) = \tilde{\mathscr{L}}\mathscr{V}$. $\tilde{\mathscr{L}}$ is the Laplacian $\tilde{\mathscr{L}} = \frac{2\mathscr{L}}{\lambda_{max}} - I$ with λ_{max} denoting the largest eigenvalue of $\mathscr{L}$. $\mathscr{L} = I - D^{-1/2}AD^{-1/2}$ where I is the identity matrix, $A \in \mathscr{R}^{ROIs \times ROIs}$ is the adjacency matrix and $D \in \mathscr{R}^{ROIs \times ROIs}$ refers to the diagonal degree matrix. top-K graph pooling and global pooling operations are employed to vectorize the subject features after the graph convolution. The similarity between the outputs of two views can be modeled by the Euclidean distance $\mathscr{V}_1', \mathscr{V}_2'$ after the representation: $\mathscr{M}(\mathscr{V}_1', \mathscr{V}_2') = \frac{\mathscr{V}_1'^{\top}\mathscr{V}_2'}{\|\mathscr{V}_1'\|\|\mathscr{V}_2'\|}$ We sample two views from a batch of N patients and thereby construct a similarity matrix of $\mathscr{M} \in \mathbb{R}^{2N \times 2N}$ for any pair of two views. To differentiate different subject, the contrastive loss defined on the i-th subject is:

$$\mathscr{L}(\mathscr{V}_1'^{(i)}, \mathscr{V}_2'^{(i)}) = -\log \frac{\exp(\mathscr{M}(\mathscr{V}_1'^{(i)}, \mathscr{V}_2'^{(i)})/\tau)}{\sum_{n=1}^{N}\sum_{j=1}^{2} 1_{n \neq i} \exp(\mathscr{M}(\mathscr{V}_1'^{(i)}, \mathscr{V}_j'^{(n)})/\tau)} \tag{1}$$

where τ is a temperature factor to control the desired attractiveness strength and $1(\cdot) = \{0,1\}$. The numerator aims to enhance homogeneous attraction whereas the denominator minimize the overall heterogeneous similarities across the population. The overall loss function is aggregated over all the pairs: $L_{CGL} = \frac{1}{2N}\sum_{n=1}^{N}(\mathscr{L}(\mathscr{V}_1'^{(n)}, \mathscr{V}_2'^{(n)}) + \mathscr{L}(\mathscr{V}_2'^{(n)}, \mathscr{V}_1'^{(n)}))$ The resulting representations of patients $\mathscr{V}'$ are then fed to the DGC to utilize label information meanwhile alleviating overfitting.

DGC transforms the disease diagnosis to a node classification on the population graph $\mathscr{G}^N$ whose structure is dynamically adapted to achieve higher precision. When predicting the disease label for an unknown subject in the graph, one can make inference with from its supervised neighbours. The initial population graph nodes are the representations learnt from contrastive learning: $\mathscr{G}^N = (\mathscr{V}'^{(n)})$ with $n = \{0,1,\cdots N\}$. The graph edges are constructed by channelling top-20 neighbours for each subject: $\mathscr{E}_{1,n}^{N} = 1$ if $\mathscr{V}'^{(n)} \in \mathscr{K}NN(\mathscr{V}'^{(1)})$. The node features are then updated by passing the information through the neighbours: $\mathscr{V}'^{(1)} = \sum_{n \in \mathscr{E}_{1,n}^{N}} \phi(\mathscr{V}'^{(1)}; \mathscr{V}'^{(n)} - \mathscr{V}'^{(1)})$, where $\phi(\cdot)$ is a fully connected network. It transform the previous node features $\mathscr{V}'^{(1)}$ as well as its neighbors$(\mathscr{V}'^{(n)} - \mathscr{V}'^{(1)})$ to new node features. which will in turn build new edges dynamically for the next update. The final classifier is a fully connected network to predict the probability of a class:$P_{0/1}(\mathscr{V}'^{(n)})$. Focal loss is employed for the classification to balance the metrics and focus more on the sensitivity: $L_{DGC} = -\sum_{n=1}^{N}\sum_{c=0}^{1}(1 - P_c(\mathscr{V}'^{(n)}))^{\gamma}\log(P_c(\mathscr{V}'^{(n)}))$. We extract the second last layer of $\phi(\mathscr{V}'^{(1)}; \mathscr{V}'^{(n)} - \mathscr{V}'^{(1)})$(Figure 1) as the final representation of the 1-st subject. All the patient representations excluding the healthy controls are eventually exploited for subtype identification.

Subtype identification

After acquiring patient representation with the label-aimed contrastive learning, we apply spectral clustering on the population graph $\mathscr{G}^P = (\mathscr{V}'^{(p)})$ with $p = \{0,1,\cdots P\}$ denoting all the patients. Spectral clustering can explicitly model patient similarities and offer the customized cluster count by selecting the top-K eigenvectors. Given the Laplacian matrix of $\mathscr{G}^P : \mathscr{L}^P = I - D^{-1/2}AD^{-1/2}$, spectral clustering firstly computes its eigenvalues and eigenvectors: $\mathscr{L}^P\boldsymbol{u} = \lambda\boldsymbol{u}$ where $u \in \mathbb{R}^{P \times T}$ with $T \leq P$ indicates the eigenvectors and λ is the eighvalue. We rank the eigenvalues and choose top-K corresponding eigenvector $u' \in \mathbb{R}^{P \times K}$as a low-dimensional orthogonal mapping of the patients where a K-means is applied. The clustering labels $S \in \mathbb{R}^{P \times N}$ are mapped back to the patients as subtype results. The optimal number of subtypes was selected based on the highest Calinski-Harabasz index.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discriminative functional connectivity identification through partial derivatives

The subtypes are observed on the representation space $f(x)$: $\mathbb{E}_m[P(f(x_m) \mid y_m = 1)] \neq \mathbb{E}_n[P(f(x_n) \mid y_n = 1)]$, where two communities M and N from the disease group do not perform consistently. If we aim to analyse the differences on the FC inputs, which can showcase medical significance and interpretation, the gradient-based method offers a promising solution. It has been shown to demonstrate feature importance, also known as the saliency map[75,76]. The motivation is that suppose a prediction $y = f(x)$ can be approximated with a linear function $y = Wx + b$, then the partial derivative $\partial y/\partial x = W$ showcases the crucial inputs with the greatest influence on the outputs. Based on this, we train a subtype classifier using the subtyping of the training set as the ground-truth. The classifier replaces the last layer of the original contrastive framework with a new subtyping classifier. The reason for such layer modification rather than utilizing the gradients of original framework is the shift of focus from the disease to the subtypes. The gradients of the original disease classifier can only indicate the connectivities contributing to the disease, which would present spurious similarities between both subtypes and fail to showcase the subtyping differences. Finally, the gradients of the subtyping outputs over the connectivity inputs including both the partial and 's connectivities are derived to highlight important connectivities.

Data availability

The data supporting the results in this study are available within the paper and its Supplementary Information. The neuroimaging datasets are publicly available from ADHD-200 (http://fcon_1000.projects.nitrc.org/indi/adhd200/), ABIDE-I and ABIDE-II (https://fcon_1000.projects.nitrc.org/indi/abide/).

Code availability

The custom code used in this study is available for research purpose from the corresponding author upon reasonable request.

Acknowledgements

Y.Z. was supported by NIH grant nos. R21MH130956, R01MH129694, and R21AG080425, and Lehigh University FIG (FIGAWD35), CORE, and Accelerator grants. G.A.F. was supported by NIH grant nos. R01MH132784 and R01MH125886 and grants from the Brain and Behavior Research Foundation and One Mind – Baszucki Brain Research Fund. T.D.S. was supported by NIH grant nos. R37MH125829, R01EB022573, R01MH112847, and R01MH120482.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Thapar, A., Cooper, M. & Rutter, M. Neurodevelopmental disorders. *The Lancet Psychiatry* **4**, 339–346 (2017).
2. Kernbach, J. M. *et al.* Shared endo-phenotypes of default mode dysfunction in attention deficit/hyperactivity disorder and autism spectrum disorder. *Transl. psychiatry* **8**, 133 (2018).
3. American Psychiatric Association, D., Association, A. P. *et al. Diagnostic and statistical manual of mental disorders: DSM-5*, vol. 5 (American psychiatric association Washington, DC, 2013).
4. Brookman-Frazee, L., Stadnick, N., Chlebowski, C., Baker-Ericzén, M. & Ganger, W. Characterizing psychiatric comorbidity in children with autism spectrum disorder receiving publicly funded mental health services. *Autism* **22**, 938–952 (2018).
5. Antshel, K. M. & Russo, N. Autism spectrum disorders and adhd: Overlapping phenomenology, diagnostic issues, and treatment considerations. *Curr. psychiatry reports* **21**, 1–11 (2019).
6. Vaidya, C. J. *et al.* Data-driven identification of subtypes of executive function across typical development, attention deficit hyperactivity disorder, and autism spectrum disorders. *J. Child Psychol. Psychiatry* **61**, 51–61 (2020).
7. Kofler, M. J. *et al.* Heterogeneity in adhd: Neurocognitive predictors of peer, family, and academic functioning. *Child Neuropsychol.* **23**, 733–759 (2017).
8. Wåhlstedt, C., Thorell, L. B. & Bohlin, G. Heterogeneity in adhd: Neuropsychological pathways, comorbidity and symptom domains. *J. abnormal child psychology* **37**, 551–564 (2009).
9. Davis, N. O. & Kollins, S. H. Treatment for co-occurring attention deficit/hyperactivity disorder and autism spectrum disorder. *Neurotherapeutics* **9**, 518–530 (2012).
10. Danielson, M. L. *et al.* Prevalence of parent-reported adhd diagnosis and associated treatment among us children and adolescents, 2016. *J. Clin. Child & Adolesc. Psychol.* **47**, 199–212 (2018).
11. Hours, C., Recasens, C. & Baleyte, J.-M. Asd and adhd comorbidity: what are we talking about? *Front. Psychiatry* **13**, 154 (2022).
12. Harkins, C. M., Handen, B. L. & Mazurek, M. O. The impact of the comorbidity of asd and adhd on social impairment. *J. Autism Dev. Disord.* 1–11 (2021).
13. Norman, L. J., Sudre, G., Bouyssi-Kobar, M., Sharp, W. & Shaw, P. A longitudinal study of resting-state connectivity and response to psychostimulant treatment in adhd. *Am. J. Psychiatry* **178**, 744–751 (2021).
14. Yahata, N. *et al.* A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. communications* **7**, 11254 (2016).
15. Kern, J. K. *et al.* Shared brain connectivity issues, symptoms, and comorbidities in autism spectrum disorder, attention deficit/hyperactivity disorder, and tourette syndrome. *Brain connectivity* **5**, 321–335 (2015).
16. Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).
17. Duffy, K. A. *et al.* Increased integration between default mode and task-relevant networks in children with adhd is associated with impaired response control. *Dev. Cogn. Neurosci.* **50**, 100980 (2021).
18. Ray, S. *et al.* Structural and functional connectivity of the human brain in autism spectrum disorders and attention-deficit/hyperactivity disorder: A rich club-organization study. *Hum. brain mapping* **35**, 6032–6048 (2014).
19. Bessadok, A., Mahjoub, M. A. & Rekik, I. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis Mach. Intell.* **45**, 5833–5848 (2022).
20. Cao, M. *et al.* Using deepgcn to identify the autism spectrum disorder from multi-site resting-state data. *Biomed. Signal Process. Control.* **70**, 103015 (2021).
21. Zhao, K. *et al.* A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in adhd. *NeuroImage* **246**, 118774 (2022).
22. Schirmer, M. D. *et al.* Neuropsychiatric disease classification using functional connectomics-results of the connectomics in neuroimaging transfer learning challenge. *Med. image analysis* **70**, 101972 (2021).
23. Rommelse, N., Buitelaar, J. K. & Hartman, C. A. Structural brain imaging correlates of asd and adhd across the lifespan: a hypothesis-generating review on developmental asd–adhd subtypes. *J. Neural Transm.* **124**, 259–271 (2017).
24. Wang, Z., Liu, Q. & Dou, Q. Contrastive cross-site learning with redesigned net for covid-19 ct classification. *IEEE J. Biomed. Heal. Informatics* **24**, 2806–2813 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Zheng, L., Xiong, J., Zhu, Y. & He, J. Contrastive learning with complex heterogeneity. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2594–2604 (2022).

26. consortium, A.-. The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Front. Syst. Neurosci.* **6**, 62 (2012).

27. Di Martino, A. *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**, 659–667 (2014).

28. Di Martino, A. *et al.* Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Sci. Data* **4**, 1–15 (2017).

29. Fortin, J.-P. *et al.* Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* **167**, 104–120 (2018).

30. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626 (2017).

31. Li, H., Parikh, N. A. & He, L. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. *Front. neuroscience* **12**, 491 (2018).

32. Hu, W. *et al.* Interpretable multimodal fusion networks reveal mechanisms of brain cognition. *IEEE transactions on medical imaging* **40**, 1474–1483 (2021).

33. Rieke, J., Eitel, F., Weygandt, M., Haynes, J.-D. & Ritter, K. Visualizing convolutional networks for mri-based diagnosis of alzheimer's disease. In *Machine Learning in Clinical Neuroimaging (MLCN)* (2018).

34. Dai, W. *et al.* Brainformer: A hybrid cnn-transformer model for brain fmri data classification. *arXiv preprint arXiv:2208.03028* (2022).

35. Fair, D. A. *et al.* Distinct neural signatures detected for adhd subtypes after controlling for micro-movements in resting state functional connectivity mri data. *Front. systems neuroscience* **6**, 80 (2013).

36. Nicholson, A. A. *et al.* The dissociative subtype of posttraumatic stress disorder: unique resting-state functional connectivity of basolateral and centromedial amygdala complexes. *Neuropsychopharmacology* **40**, 2317–2326 (2015).

37. Zhang, Y. *et al.* Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nat. biomedical engineering* **5**, 309–323 (2021).

38. Zerbi, V. *et al.* Brain mapping across 16 autism mouse models reveals a spectrum of functional connectivity subtypes. *Mol. psychiatry* **26**, 7610–7620 (2021).

39. Park, H. S., Yi, S. Y., Yoon, S. A. & Hong, S.-B. Comparison of the autism diagnostic observation schedule and childhood autism rating scale in the diagnosis of autism spectrum disorder: a preliminary study. *J. Korean Acad. Child Adolesc. Psychiatry* **29**, 172 (2018).

40. Lynch, C. J. *et al.* Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. *Biol. psychiatry* **74**, 212–219 (2013).

41. Hong, S.-J. *et al.* Toward neurosubtypes in autism. *Biol. Psychiatry* **88**, 111–128 (2020).

42. Anderson, J. S. *et al.* Functional connectivity magnetic resonance imaging classification of autism. *Brain* **134**, 3742–3754 (2011).

43. Tang, S. *et al.* Reconciling dimensional and categorical models of autism heterogeneity: a brain connectomics and behavioral study. *Biol. psychiatry* **87**, 1071–1082 (2020).

44. Guo, X. *et al.* Inter-individual heterogeneity of functional brain networks in children with autism spectrum disorder. *Mol. Autism* **13**, 1–13 (2022).

45. González-Madruga, K., Staginnus, M. & Fairchild, G. Alterations in structural and functional connectivity in adhd: Implications for theories of adhd. *New Discov. Behav. Neurosci. Attention-Deficit Hyperact. Disord.* 445–481 (2022).

46. Lin, H.-Y. *et al.* Brain–behavior patterns define a dimensional biotype in medication-naïve adults with attention-deficit hyperactivity disorder. *Psychol. medicine* **48**, 2399–2408 (2018).

47. Oldehinkel, M. *et al.* Functional connectivity in cortico-subcortical brain networks underlying reward processing in attention-deficit/hyperactivity disorder. *NeuroImage: Clin.* **12**, 796–805 (2016).

48. Yerys, B. E. *et al.* Functional connectivity of frontoparietal and salience/ventral attention networks have independent associations with co-occurring attention-deficit/hyperactivity disorder symptoms in children with autism. *Biol. Psychiatry: Cogn. Neurosci. Neuroimaging* **4**, 343–351 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

49. Mostert, J. C. *et al.* Characterising resting-state functional connectivity in a large sample of adults with adhd. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **67**, 82–91 (2016).

50. Qian, X. *et al.* Large-scale brain functional network topology disruptions underlie symptom heterogeneity in children with attention-deficit/hyperactivity disorder. *NeuroImage: Clin.* **21**, 101600 (2019).

51. Sörös, P. *et al.* Hyperactivity/restlessness is associated with increased functional connectivity in adults with adhd: a dimensional analysis of resting state fmri. *BMC psychiatry* **19**, 1–11 (2019).

52. Luo, Y., Weibman, D., Halperin, J. M. & Li, X. A review of heterogeneity in attention deficit/hyperactivity disorder (adhd). *Front. human neuroscience* **13**, 42 (2019).

53. Dias, T. G. C. *et al.* Characterizing heterogeneity in children with and without adhd based on reward system connectivity. *Dev. cognitive neuroscience* **11**, 155–174 (2015).

54. Jung, M. *et al.* Surface-based shared and distinct resting functional connectivity in attention-deficit hyperactivity disorder and autism spectrum disorder. *The Br. J. Psychiatry* **214**, 339–344 (2019).

55. Bethlehem, R. A., Romero-Garcia, R., Mak, E., Bullmore, E. & Baron-Cohen, S. Structural covariance networks in children with autism or adhd. *Cereb. Cortex* **27**, 4267–4276 (2017).

56. Linke, A. C., Olson, L., Gao, Y., Fishman, I. & Müller, R.-A. Psychotropic medication use in autism spectrum disorders may affect functional brain connectivity. *Biol. Psychiatry: Cogn. Neurosci. Neuroimaging* **2**, 518–527 (2017).

57. Cordova, M. *et al.* Heterogeneity of executive function revealed by a functional random forest approach across adhd and asd. *NeuroImage: Clin.* **26**, 102245 (2020).

58. Lake, E. M. *et al.* The functional brain organization of an individual allows prediction of measures of social abilities transdiagnostically in autism and attention-deficit/hyperactivity disorder. *Biol. psychiatry* **86**, 315–326 (2019).

59. Bedford, S. A. *et al.* Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. *Mol. Psychiatry* **25**, 614–628 (2020).

60. Bauman, M. L. & Kemper, T. L. Neuroanatomic observations of the brain in autism: a review and future directions. *Int. journal developmental neuroscience* **23**, 183–187 (2005).

61. Lotspeich, L. J. *et al.* Investigation of neuroanatomical differences between autism and aspergersyndrome. *Arch. general psychiatry* **61**, 291–298 (2004).

62. Martínez, K. *et al.* Sensory-to-cognitive systems integration is associated with clinical severity in autism spectrum disorder. *J. Am. Acad. Child & Adolesc. Psychiatry* **59**, 422–433 (2020).

63. Safar, K. *et al.* Shared and distinct patterns of functional connectivity to emotional faces in autism spectrum disorder and attention-deficit/hyperactivity disorder children. *Front. Psychol.* 563 (2022).

64. You, Y. *et al.* Graph contrastive learning with augmentations. *Adv. neural information processing systems* **33**, 5812–5823 (2020).

65. Esteban, O. *et al.* fmriprep: a robust preprocessing pipeline for functional mri. *Nat. methods* **16**, 111–116 (2019).

66. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. image analysis* **12**, 26–41 (2008).

67. Zhang, Y., Brady, J. M. & Smith, S. Hidden markov random field model for segmentation of brain mr image. In *Medical Imaging 2000: Image Processing*, vol. 3979, 1126–1137 (SPIE, 2000).

68. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).

69. Pruim, R. H. *et al.* Ica-aroma: A robust ica-based strategy for removing motion artifacts from fmri data. *Neuroimage* **112**, 267–277 (2015).

70. Schaefer, A. *et al.* Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cereb. Cortex* **28**, 3095–3114 (2018).

71. Li, X. *et al.* Braingnn: Interpretable brain graph neural network for fmri analysis. *Med. Image Analysis* **74**, 102233 (2021).

72. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607 (PMLR, 2020).

73. Defferrard, M., Bresson, X. & Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. neural information processing systems* **29** (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

74. Parisot, S. *et al.* Spectral graph convolutions for population-based disease prediction. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, 177–185 (Springer, 2017).

75. Adebayo, J. *et al.* Sanity checks for saliency maps. *Adv. neural information processing systems* **31** (2018).

76. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034* (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary

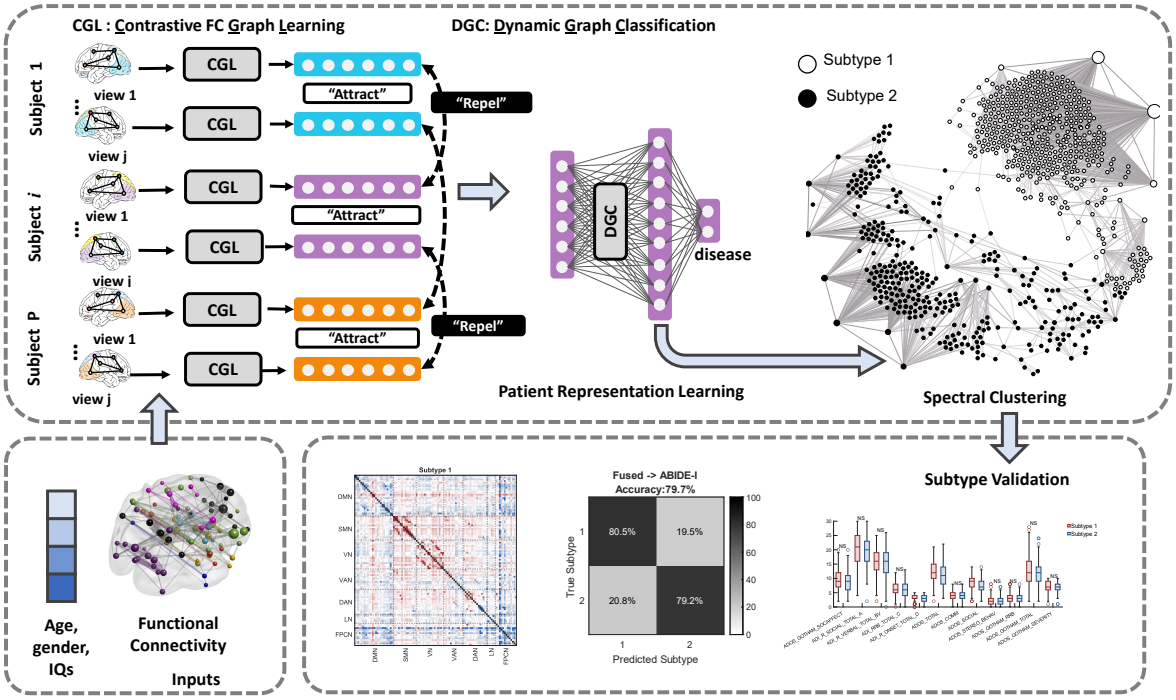


Figure 1. Illustration of contrastive learning framework for transdiagnostic subtype identification.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

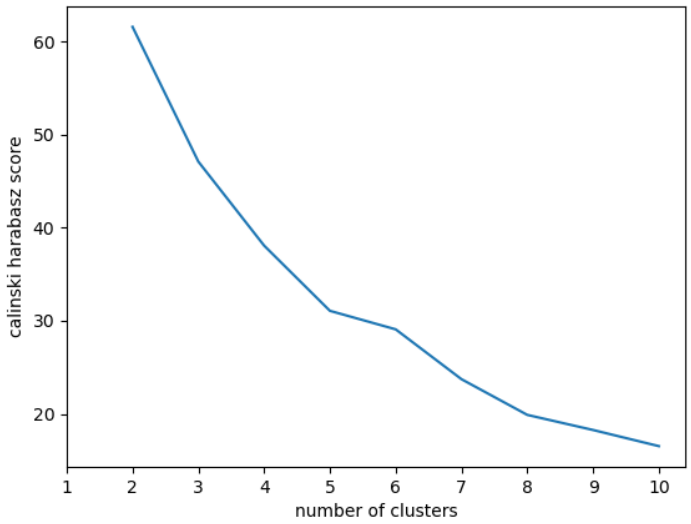


Figure 2. Selection of the optimal number of clusters. We run the spectral clustering method on the transdiagnostic ASD and ADHD population. We set varying number of clusters ranging from 2 to 10 for the clustering algorithm whose input is a distance matrix between subjects. For each choice of the number we compute the corresponding Calinski-Harabasz score that measures the ratio of inter-cluster distance and intra-cluster distance. Therefore the highest Calinski-Harabasz score suggests clear-defined clusters.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

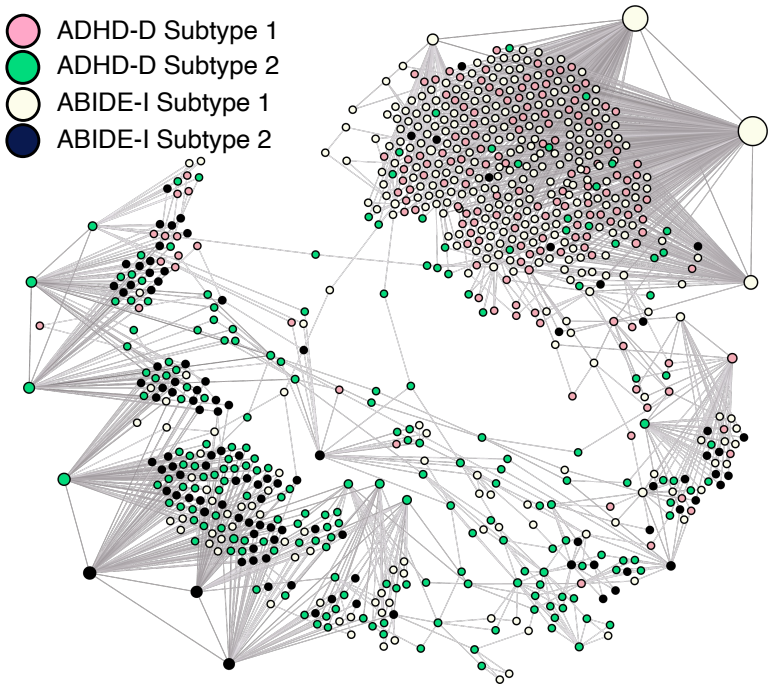


Figure 3. Subtype relationships between the individual disease. Two subtypes are obtained separately in ABIDE-I and ADHD-discovery set (ADHD-D) using the same pipeline to showcase transdiagnostic heterogeneity and comorbidity. The population graph was generated by the transdiagnostic discovery set but the node labels are obtained through clustering on individual diseases. Their results are mapped to the transdiagnostic population graph with different colors indicating the subtype labels acquired from individual diseases.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

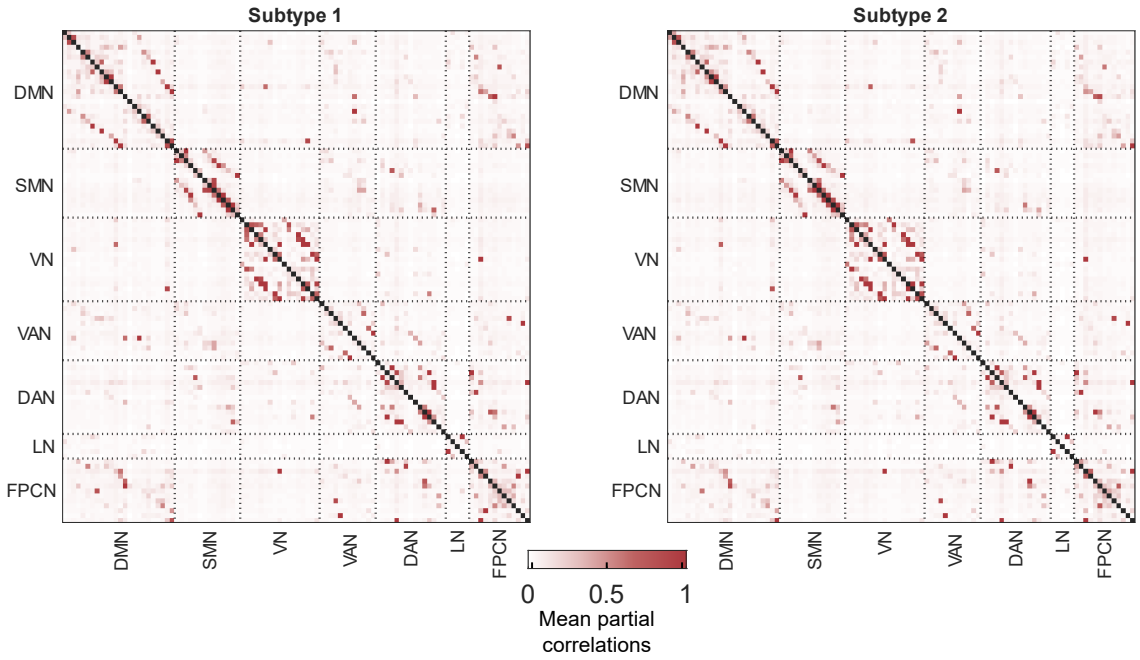


Figure 4. Mean connectivity inputs of the two subtypes extracted from the transdiagnostic set of ABIDE-I and ADHD discovery set. The color bar indicates the mean partial connectivity values.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

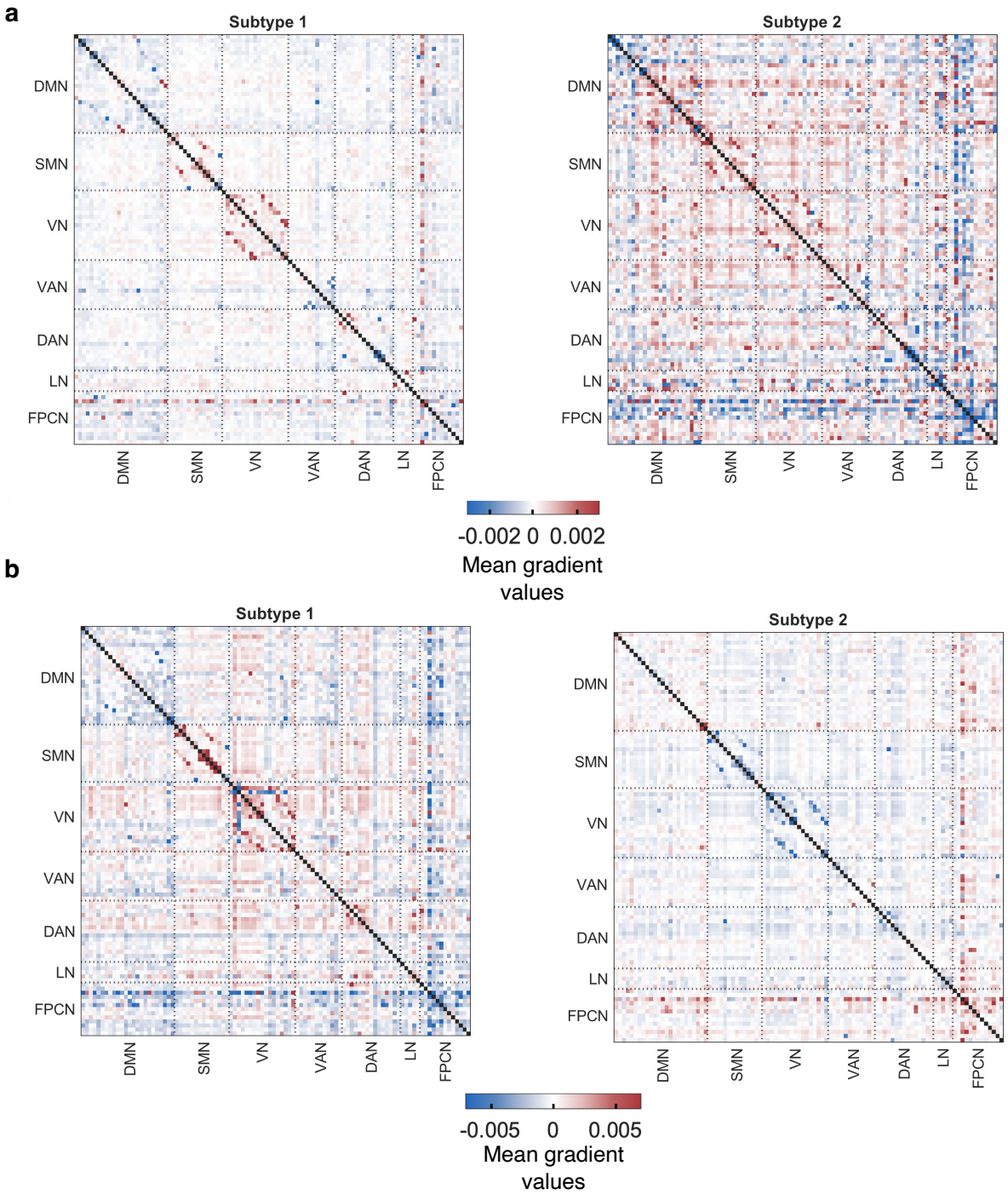


Figure 5. Connectivity gradients of two subtypes obtained from the individual discovery sets. The gradients are obtained from running the subtyping on individual sets separately, where positive gradients mean the positive contribution of a connectivity to the subtype. **a**. Mean connectivity gradients of the two subtypes on ABIDE-I. **b.** Mean connectivity gradients of the two subtypes on the ADHD-discovery set.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

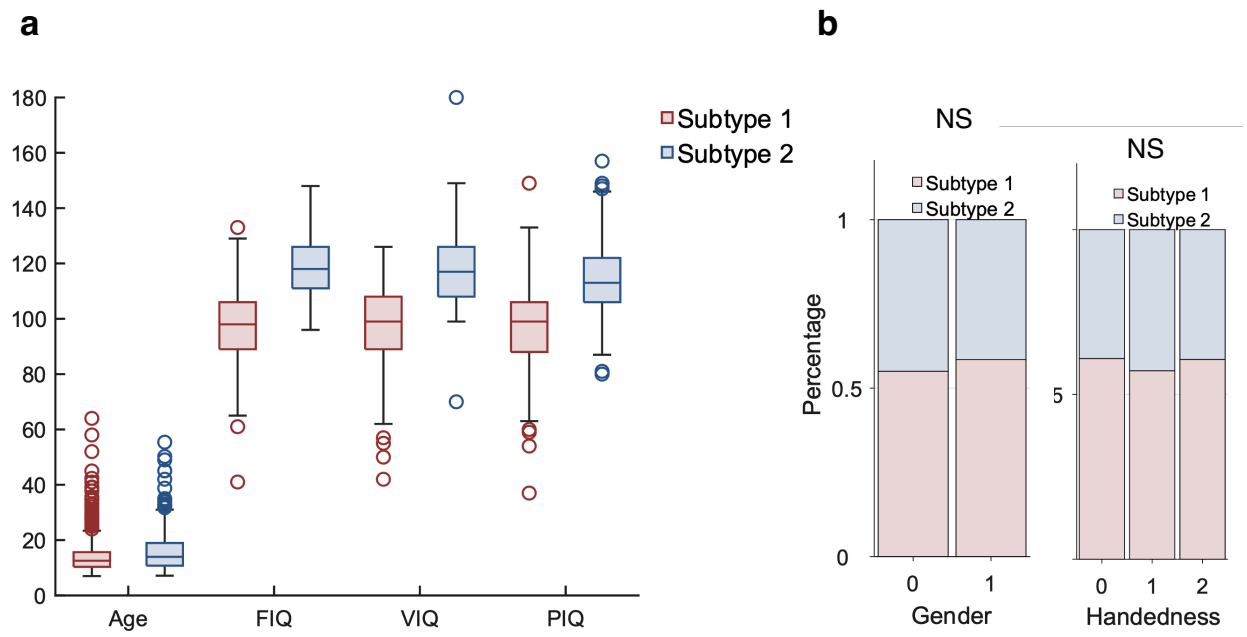


Figure 6. Demographic and clinical differences on the transdiagnostic discovery set (ABIDE-I + ADHD discovery set). **a.** Boxplot for numerical features including age and three IQ cognitive measurements,where 'NS' indicates non- significant t-test results. **b.** Stacked percentage barplot for categorical features that showcases the percentage of the two subtypes in each feature category, such as gender (0: Female, 1: Male) and handedness (0: Left, 1: Right, 2: Ambidextrous. Continuous values between 0 to 1 in ADHD are categorized as Ambidextrous). 'NS' shows non-significant chi-square test results.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

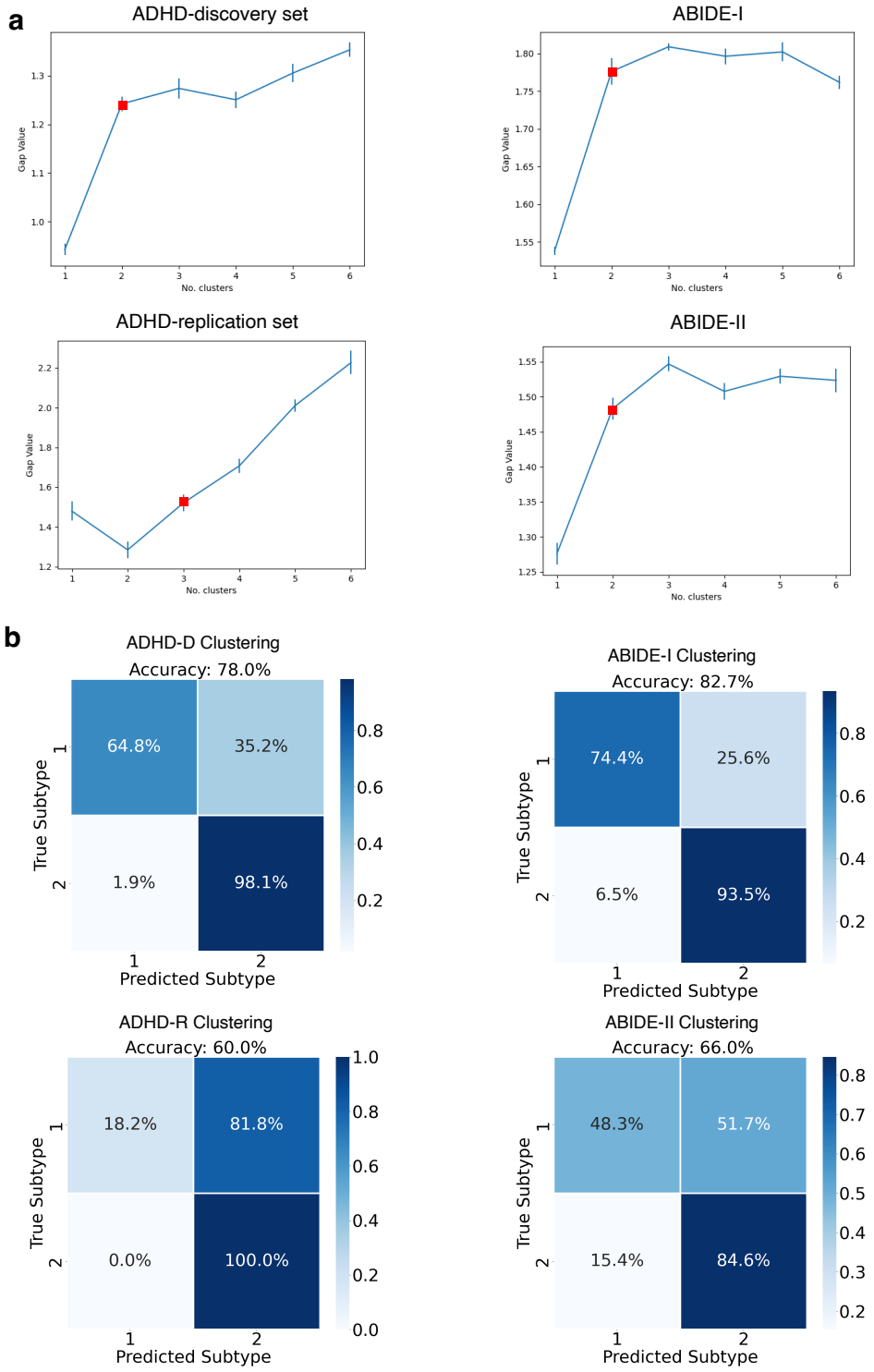


Figure 7. Clustering based on demographic information and cognitive measures for the discovery and replication datasets. Six features including Gender, Handedness, Age, FIQ, VIQ, PIQ are utilized to train the subtyping cluster. **a.** The optimal number of clustering (noted with red squares) found in 2 discovery and 2 replication sets using gap statistic criterion. The error bars indicate standard deviation. **b.** Confusion matrices of comparing the subtyping labels obtained by our pipeline (ground truths) and subtyping (predictions) using biological features as such age, gender are shown with confusion matrices.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

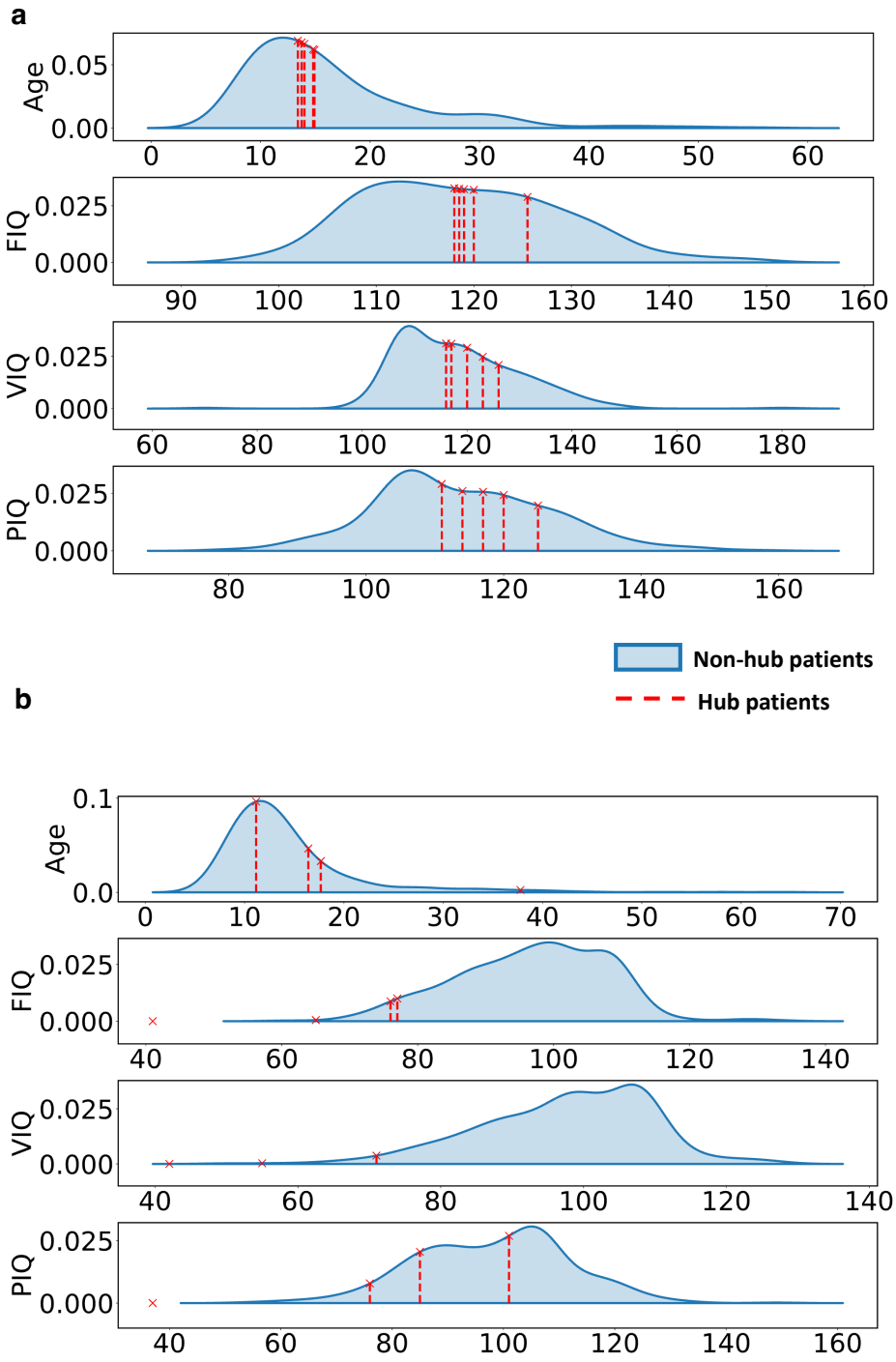


Figure 8. PCD differences between the hub and non-hub patients in the transdiagnostic discovery set. Hub patients are defined by the patients with the highest node degree in the population graph. We estimate the distributions of the age and three cognitive IQs for the non-hub patients in each subtype. Their estimates on the hub data are shown in crosses and the dashed line, where a long dashed line indicate normality and small value indicates abnormality. **a.** Six hub patients verus the rest non-hub patients in subtype 1. The non-hub distribution is fitted by the Gaussian kernel density estimation. **b.** Four hub patients versus the rest non-hub patients in subtype 2.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

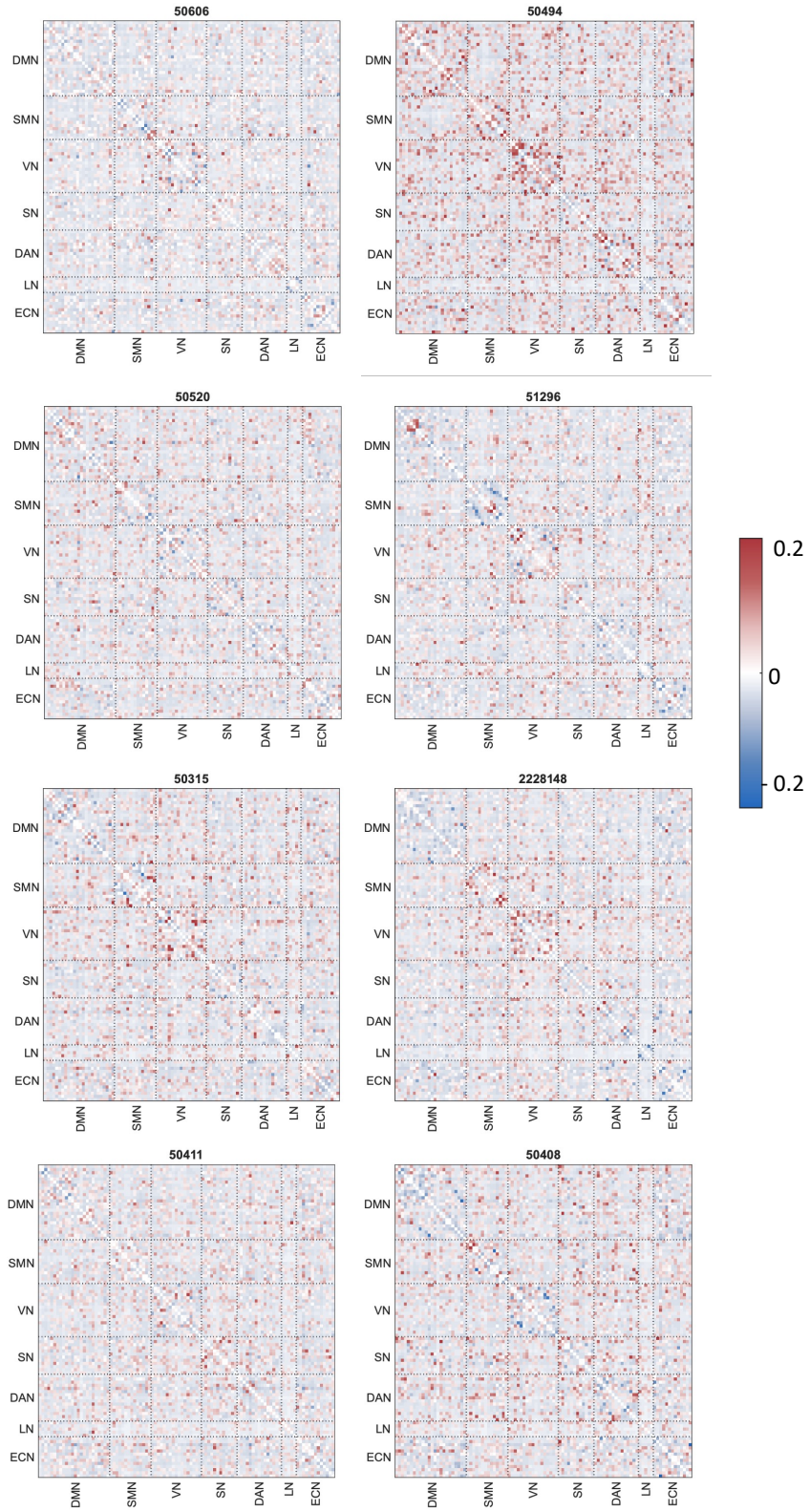


Figure 9. FC differences between the hub and non-hub patients in the transdiagnostic discovery set. The label indicate the names of the hub subjects. The matrix values show the different between the FC inputs of a hub patient and the mean FC values from the non-hub patients within its corresponding subtype category.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Demographic and clinical characteristic differences between the two subtypes on the transdiagnostic dataset of ADHD-discovery set + ABIDE-I. * indicates significant differences with p value < 0.05

Categorical variables	Subtype 1 (N=452)	Subtype 2 (N=347)	χ^2 value	p value
Female, No (%)	71 (59.2)	49 (40.8)	0.27	0.60
Male, No (%)	381 (56.1)	298 (43.9)		
Handedness = Left, No (%)	26 (56.5)	20(43.5)	1.50	0.47
Handedness = Right, No (%)	345 (55.6)	276 (44.4)		
Handedness = Ambidextrous, No (%)	81 (61.4)	51 (38.6)		
Continuous variables			t value	p value
Age, Mean (Std)	14.63 (7.23)	15.56 (7.67)	-1.73	0.08
FIQ, Mean (Std)	92.89 (10.42)	115.87(10.21)	-31.21	<1e-4 *
VIQ, Mean (Std)	85.49 (14.19)	114.17 (11.85)	-20.23	<1e-4 *
PIQ, Mean (Std)	93.94 (13.71)	112.98 (10.99)	-21.75	<1e-4 *

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Similarity measure between individual dataset and the transdiagnostic discovery set using correlation coefficient. * indicates significant differences with p value < 0.05

	Transdiagnostic discovery set subtype1 correlation coefficient (p value)	Transdiagnostic discovery set subtype2 correlation coefficient (p value)
ADHD-discovery set subtype 1	0.5834 ($< 10^{-5}$)	-0.4985 ($< 10^{-5}$)
ADHD-discovery set subtype 2	-0.6561 ($< 10^{-5}$)	0.6552 ($< 10^{-5}$)
ABIDE-I subtype 1	0.0243 (0.0154)	-0.0460 ($< 10^{-5}$)
ABIDE-I subtype 2	0.3384 ($< 10^{-5}$)	-0.2991 ($< 10^{-5}$)
ADHD-validation set subtype 1	-0.2316 ($< 10^{-5}$)	0.2212 ($< 10^{-5}$)
ADHD-validation set subtype 2	-0.1482 ($< 10^{-5}$)	0.1217 ($< 10^{-5}$)
ABIDE-II subtype 1	0.5432 ($< 10^{-5}$)	-0.6004 ($< 10^{-5}$)
ABIDE-II subtype 2	-0.5856 ($< 10^{-5}$)	0.4755 ($< 10^{-5}$)

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Demographic characteristics and cognitive assessments of **healthy controls** from the transdiagnostic discovery set. * indicates significant differences with p value < 0.05

Continuous variable	Mean (std)	Categorical variable	No, (%)
Age	15.1 (6.9)	Gender= Female	237 (26.9)
FIQ	112.0 (12.3)	Gender= Male	644 (73.1)
VIQ	112.3 (12.5)	Handedness = Left	42 (4.8)
PIQ	108.8 (12.96)	Handedness = Right	736 (83.5)
		Handedness = Ambidextrous	103 (11.7)

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. Demographic characteristics and clinical variables between the two subtypes on the ABIDE-I dataset.

Categorical variables	Subtype 1 (N=301)	Subtype 2 (N=230)	χ^2 value	p value
Female, No (%)	33 (51.6)	31 (48.4)	0.56	0.45
Male, No (%)	268 (57.4)	199 (42.6)		
Handedness = Left, No (%)	25 (62.5)	15 (37.5)	0.62	0.73
Handedness = Right, No (%)	269 (56.2)	210 (43.8)		
Handedness = Ambidextrous, No (%)	7 (58.3)	5 (41.7)		
Medication Naïve, No (%)	137 (52.3)	125 (47.7)	0.48	0.49
Medication Not Naïve, No (%)	73 (56.6)	56 (43.4)		
DSM_IV_TR=control, No (%)	6 (30.0)	14 (70.0)	31.98	<1e-4 *
DSM_IV_TR=autism, No (%)	221 (65.2)	118 (34.8)		
DSM_IV_TR=Aspergers, No (%)	36 (38.7)	57 (61.3)		
DSM_IV_TR=PDD-NOS, No (%)	20 (55.6)	16 (44.4)		
DSM_IV_TR=Aspergers / PDD-NOS, No (%)	1 (16.7)	5 (83.3)		
Continuous variables			**t value**	**p value**
Age, Mean (Std)	16.07 (8.22)	18.37 (8.47)	-3.15	2e-3 *
FIQ, Mean (Std)	95.59 (12.09)	119.02 (10.71)	-23.2	<1e-4 *
VIQ, Mean (Std)	95.34 (12.88)	117.45 (11.74)	-20.32	<1e-4 *
PIQ, Mean (Std)	98.98 (14.77)	114.04 (12.45)	-12.43	<1e-4 *
WISC_IV_VCI, Mean (Std)	93.56 (13.66)	124.17 (11.08)	-4.69	0.0001 *
WISC_IV_PRI, Mean (Std)	98.50 (9.61)	119.83 (9.17)	-4.48	0.0002 *
WISC_IV_WMI, Mean (Std)	85.94 (14.91)	107.5 (11.53)	-3.05	0.006 *
WISC_IV_PSI, Mean (Std)	82.63 (9.85)	100.0 (17.15)	-2.82	0.01 *
WISC_IV_SIM_SCALED, Mean (Std)	9.69 (3.48)	14.0 (1.91)	-2.74	0.01 *
WISC_IV_VOCAB_SCALED, Mean (Std)	8.0 (3.41)	14.17 (2.48)	-3.86	0.0009 *
WISC_IV_INFO_SCALED, Mean (Std)	8.88 (2.64)	14.33 (1.97)	-4.39	0.0003 *
WISC_IV_BLK_DSN_SCALED, Mean (Std)	11.44 (1.06)	14.17 (3.85)	-2.47	0.02 *
WISC_IV_PIC_CON_SCALED, Mean (Std)	7.50 (2.29)	11.17 (1.86)	-3.35	0.003 *
WISC_IV_MATRIX_SCALED, Mean (Std)	10.25 (2.82)	13.50 (2.29)	-2.41	0.03 *
WISC_IV_DIGIT_SPAN_SCALED, Mean (Std)	8.06 (2.84)	11.00 (2.38)	-2.15	0.04 *
WISC_IV_LET_NUM_SCALED, Mean (Std)	7.19 (3.34)	11.83 (1.95)	-3.06 f	0.006 *
WISC_IV_CODING_SCALED, Mean (Std)	6.44 (1.90)	8.67 (3.25)	-1.89	0.07
WISC_IV_SYM_SCALED, Mean (Std)	7.19 (2.58)	10.67 (2.69)	-2.66	0.02 *
ADOS_GOTHAM_SOCAFFECT, Mean (Std)	9.84 (3.81)	9.05 (3.65)	1.56	0.12
ADI_R_SOCIAL_TOTAL_A, Mean (Std)	20.13 (5.51)	19.30 (5.47)	1.45	0.15
ADI_R_VERBAL_TOTAL_BV, Mean (Std)	15.79 (4.39)	15.75 (4.92)	0.09	0.92
ADI_RRB_TOTAL_C, Mean (Std)	6.19 (2.42)	5.95 (2.74)	0.89	0.38
ADI_R_ONSET_TOTAL_D, Mean (Std)	3.40 (1.23)	3.01 (1.28)	2.64	0.009 *
ADOS_TOTAL, Mean (Std)	12.35 (3.72)	11.25 (3.76)	2.84	0.005 *
ADOS_COMM, Mean (Std)	3.92 (1.54)	3.71 (1.57)	1.27	0.21
ADOS_SOCIAL, Mean (Std)	8.53 (2.60)	7.59 (2.78)	3.26	0.001 *
ADOS_STEREO_BEHAV, Mean (Std)	2.24 (1.65)	1.94 (1.48)	1.58	0.12
ADOS_GOTHAM_RRB, Mean (Std)	3.02 (1.90)	2.99 (1.57)	0.15	0.88
ADOS_GOTHAM_TOTAL, Mean (Std)	12.91 (4.87)	11.93 (4.40)	1.57	0.12
ADOS_GOTHAM_SEVERITY, Mean (Std)	7.25 (2.07)	6.84 (2.10)	1.47	0.14

Note: empty entries as well as NaNs are excluded during counting.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 5. Demographic characteristics and clinical variables between the two subtypes on the ADHD-discovery set. The subtyping labels are obtained from the transdiagnostic discovery set. * indicates significant differences with p value < 0.05

Categorical variables	Subtype 1 (N=162)	Subtype 2 (N=106)	χ^2 value	p value
Female, No (%)	33 (58.9)	23 (41.1)	0.01	0.91
Male, No (%)	129 (60.8)	83 (39.2)		
Handedness = Left, No (%)	3 (50)	3 (50)	0.28	0.88
Handedness = Right, No (%)	86 (60.6)	56 (39.4)		
Handedness = Ambidextrous, No (%)	73 (60.8)	47 (39.2)		
Medication Naïve, No (%)	113 (58.5)	80 (41.5)	0.78	0.38
Medication Not Naïve, No (%)	49 (65.3)	26 (34.7)		
Continuous variables			t value	p value
Age, Mean (Std)	11.28 (2.35)	11.50 (2.51)	-0.72	0.47
FIQ, Mean (Std)	98.28 (9.42)	118.25 (7.86)	-18.02	<1e-4 *
VIQ, Mean (Std)	100.47 (10.44)	120.29 (10.43)	-15.14	<1e-4 *
PIQ, Mean (Std)	96.02 (11.80)	114.00 (12.26)	-11.96	<1e-4 *
ADHD Index, Mean (Std)	63.52 (13.66)	63.58 (13.52)	-0.04	0.97
Inattentive, Mean (Std)	55.15 (22.12)	54.81 (22.39)	0.12	0.90
Hyper/Impulsive, Mean (Std)	51.54 (24.66)	51.68 (23.83)	-0.05	0.96

Note: empty entries as well as NaNs are excluded during counting.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 6. Demographic characteristics and clinical variables between the two subtypes on the ABIDE-II dataset. * indicates significant differences with p value < 0.05

Categorical variables	Subtype 1 (N=232)	Subtype 2 (N=221)	χ^2 value	p value
Female, No (%)	35 (50.7)	34 (49.3)	0.00	1.00
Male, No (%)	197 (51.3)	187 (48.7)		
Handedness = Left, No (%)	27 (56.3)	21 (43.7)	0.83	0.66
Handedness = Right, No (%)	190 (50.3)	188 (49.7)		
Handedness = Ambidextrous, No (%)	15 (55.6)	12(44.4)		
Medication Naïve, No (%)	146 (50.9)	141 (49.1)	0.95	0.33
Medication Not Naïve, No (%)	73 (56.6)	56 (43.4)		
DSM_IV_TR=autism, No (%)	65 (60.7)	42 (39.3)	8.27	0.02 *
DSM_IV_TR=Aspergers, No (%)	39 (41.9)	54 (58.1)		
DSM_IV_TR=PDD-NOS, No (%)	24 (61.5)	15 (38.5)		
Continuous variables			t value	p value
Age, Mean (Std)	13.86 (6.28)	17.46 (11.37)	-4.19	<0.0001 *
FIQ, Mean (Std)	97.68 (15.19)	116.34 (11.22)	-14.78	<0.0001 *
VIQ, Mean (Std)	98.97 (12.29)	116.87 (11.42)	-16.01	<0.0001 *
PIQ, Mean (Std)	104.09 (17.36)	105.90 (12.52)	-1.26	0.21
CBCL_6-18_ATTENTION_T	69.09 (10.65)	63.67 (8.72)	3.39	0.0008 *
CBCL_6-18_AGGRESSIVE_T	58.63 (8.36)	58.74 (8.15)	-0.08	0.94
CBCL_6-18_ATTENTION_DEFICIT_T	66.15 (8.22)	62.76 (8.03)	2	0.04 *
ADI_R_SOCIAL_TOTAL_A	19.56 (5.69)	18.47 (6.10)	1.52	0.13
ADI_R_VERBAL_TOTAL_BV	15.58 (4.73)	14.49 (4.68)	1.88	0.06
ADI_R_NONVERBAL_TOTAL_BV	8.66 (3.55)	8.38 (3.34)	0.61	0.54
ADI_R_RRB_TOTAL_C	5.67 (2.63)	5.64 (2.40)	0.12	0.91
ADI_R_ONSET_TOTAL_D	3.11 (1.47)	2.68 (1.25)	2.58	0.01 *
ADI_R_C4_TOTAL	1.93 (2.89)	1.00 (0.64)	1.73	0.08
ADOS_G_TOTAL	11.23 (4.14)	10.18 (3.71)	2.25	0.03 *
ADOS_G_COMM	3.45 (1.68)	3.17 (1.58)	1.38	0.17
ADOS_G_SOCIAL	7.54 (2.66)	6.82 (2.60)	2.28	0.02 *
ADOS_G_STEREO_BEHAV	1.68 (1.44)	1.55 (1.45)	0.71	0.48
ADOS_G_CREATIVITY	0.56 (0.70)	0.59 (0.70)	-0.14	0.89
ADOS_2_SOCAFFECT	9.67 (4.11)	9.00 (3.09)	1.22	0.23
ADOS_2_RRB	3.19 (1.71)	3.19 (1.80)	0.005	0.99
ADOS_2_TOTAL	12.87 (4.65)	12.19 (3.60)	1.07	0.29
ADOS_2_SEVERITY_TOTAL	7.14 (2.05)	6.93 (1.68)	0.72	0.47

Note: empty entries as well as NaNs are excluded during counting.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.29.582790; this version posted March 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 7. Demographic characteristics and clinical variables between the two subtypes on the ADHD-validation set. * indicates significant differences with p value < 0.05

Categorical variables	Subtype 1 (N=22)	Subtype 2 (N=23)	χ^2 value	p value
Female, No (%)	7 (87.5)	1 (21.5)	4.07	0.04 *
Male, No (%)	15 (40.5)	22 (59.5)		
Handedness = Left, No (%)	1 (16.7)	5 (83.3)	1.58	0.21
Handedness = Right, No (%)	21 (53.8)	18 (46.2)		
Handedness = Ambidextrous, No (%)	0	0		
Medication Naïve, No (%)	19 (45.2)	23 (54.8)	1.53	0.22
Medication Not Naïve, No (%)	3 (100)	0		
Continuous variables			t value	p value
Age, Mean (Std)	15.17 (4.03)	16.64 (2.26)	-1.28	0.21
FIQ, Mean (Std)	111.04 (8.45)	109.74 (1.22)	0.71	0.48
VIQ, Mean (Std)	108.05 (5.02)	107.61 (1.84)	0.38	0.70
PIQ, Mean (Std)	105.86 (4.00)	107.00 (0.00)	-1.33	0.19

Note: empty entries as well as NaNs are excluded during counting.