Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Functional Imaging Derived ADHD Biotypes Based on Deep Clustering May Guide Personalized Medication Therapy

Jing Sui (kittysj@gmail.com)
Georgia Institute of Technology, Emory University and Georgia State University https://orcid.org/0000-0001-6837-5966

Aichen Feng
School of Artificial Intelligence, University of Chinese Academy of Sciences

Yuan Feng
Peking University Sixth Hospital/Institute of Mental Health, National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital)

Dongmei Zhi
State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University

Rongtao Jiang
Yale School of Medicine https://orcid.org/0000-0003-4657-0079

Zening Fu
Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University

Ming Xu
School of Artificial Intelligence, University of Chinese Academy of Science

Min Zhao
School of Artificial Intelligence, University of Chinese Academy of Sciences

Shan Yu
School of Artificial Intelligence, University of Chinese Academy of Sciences

Michael Stevens
Department of Psychiatry, Yale University School of Medicine

Li Sun
Peking University Sixth Hospital/Institute of Mental Health, National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital)

Vince Calhoun
Georgia Institute of Technology, Emory University and Georgia State University

---

Article

Keywords: Attention deficit hyperactivity disorder (ADHD), graph convolutional network (GCN), deep clustering, biological subtype detection, ABCD

Posted Date: September 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3272441/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Functional Imaging Derived ADHD Biotypes Based on Deep Clustering May Guide Personalized Medication Therapy

Aichen Feng[1,2#], Yuan Feng[3#], Dongmei Zhi[4#], Rongtao Jiang[5], Zening Fu[6], Ming Xu[1,2], Min Zhao[1,2], Shan Yu[1,2], Michael Stevens[7,8], Li Sun[3*], Vince Calhoun[6*], Jing Sui [4,6*]

[1] *Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190*
[2] *School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, 100049*
[3] *Peking University Sixth Hospital/Institute of Mental Health, National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Beijing, China, 100191*
[4] *State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China, 100875*
[5]*Department of Radiology and Biomedical imaging, Yale University, New Haven, Connecticut, USA,*
[6] *Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia Institute of Technology, Emory University and Georgia State University, Atlanta, Georgia, United States, 30303*
[7] *Department of Psychiatry, Olin Neuropsychiatry Research Center, Institute of Living, Hartford Healthcare Corporation, Hartford, CT, United States.*
[8] *Department of Psychiatry, Yale University School of Medicine, New Haven, CT, United States.*

Corresponding author:

Jing Sui Ph.D. Professor, Email: kittysj@gmail.com

Vince D Calhoun, Ph.D. Professor, Email: vcalhoun@gsu.edu

Li Sun, Ph.D. Professor, Email: sunlioh@bjmu.edu.cn

[#*]**These authors have contributed equally to this work.**

Keywords: Attention deficit hyperactivity disorder (ADHD), graph convolutional network (GCN), deep clustering, biological subtype detection, ABCD.

ABSTRACT

Attention deficit hyperactivity disorder (ADHD) is one prevalent neurodevelopmental disorder with childhood onset, however, there is no clear correspondence established between clinical ADHD subtypes and primary medications. Identifying objective and reliable neuroimaging markers for categorizing ADHD biotypes may lead to more individualized, biotype-guided treatment. Here we proposed graph convolutional network plus deep clustering for ADHD biotype detection using functional network connectivity (FNC), resulting in two biotypes based on 1069 ADHD patients selected from Adolescent Brain and Cognitive Development (ABCD) study, which were well replicated on independent ADHD adolescents undergoing longitudinal medication treatment (n=130). Interestingly, in addition to differences in cognitive performance and hyperactivity/impulsivity symptoms, biotype 1 treated with methylphenidate demonstrated significantly better recovery than biotype 2 treated with atomoxetine ($p<0.05$, FDR corrected). This imaging-driven, biotype-guided approach holds promise for facilitating personalized treatment of ADHD, exploring possible boundaries through innovative deep learning algorithms aimed at improving medication treatment effectiveness.

INTRODUCTION

Disease heterogeneity has been a critical challenge for precision diagnosis and treatment, especially in neuropsychiatric diseases. As a highly complex and heterogeneous disorder with childhood onset[1], attention deficit hyperactivity disorder (ADHD) is usually divided into three subtypes based on core specific symptoms (ICD-9/10;DSM-4/5)[2,3], *i.e.*, Hyperactivity-Impulsivity (ADHD-HI), Inattention (ADHD-I) and the Combined (ADHD-C) subtype[2]. Meanwhile, stimulant medications (e.g., methylphenidate (MPH), amphetamine) and non-stimulant medications (e.g., atomoxetine (ATX)) are two main categories of medications approved for treating ADHD. However, currently there is no clear medication treatment plan corresponding to each subtype, which is decided mainly relying on clinical experience and not always being effective. Reports have demonstrated that MPH can equally improve inattention symptoms in both ADHD-C and ADHD-I subtypes, while predominantly reducing hyperactivity in the former[4]. A rat study reported that both MPH and ATX improved overall attention, but the highest dose of MPH significantly increased impulsivity, while ATX induced decrease in impulsivity[5]. Therefore, for ADHD patient with pure inattention, the optimal medication selection is challenging in clinical practice. Thus, the identification of imaging-based ADHD biological subtypes (biotypes) that may exhibit greater sensitivity to specific drug therapies holds promise for guiding a more personalized treatment approach with enhanced effectiveness.

Meanwhile, multiple functional magnetic resonance imaging (fMRI) analyses have been performed to study ADHD subtype[6-9], though in a relatively highly piecemeal way, which most relied heavily on prior clinical subtype divisions based on pure symptoms[3] and highlighted group-level differences via univariate statistic methods[10-12]. For example, ADHD-C were found to have hyper-connectivity in the pre-default mode network and lower network modularity compared to health controls (HC), whereas no such differences were found in the ADHD-I subtype[13], whilst ADHD-C/HI to ADHD-I were often compared on the level of comorbidity, treatment response, and possible etiological factors[14].

In contrast, understanding ADHD heterogeneity how defined by multimodal neuroimaging and demographic features rather than symptoms can offer a novel perspective for unraveling the ADHD complexity and pathology, as it may provide critical information for developing a more targeted, effective treatment program based on objective measures. The existing studies on ADHD biotypes encompass investigations into brain function, structure and neuropathology using diverse methodologies[15-18]. For instance, Fair *et al.* adopted graph theory and community detection to clarify neuropsychological heterogeneity in typically developing children and children with ADHD, which yielded four unique subgroups each of which had unique patterns of factor scores[19]. Dias et al. performed community detection analysis and characterized heterogeneity in children with and without ADHD based on reward system connectivity[20]. These studies helped disentangle the disease heterogeneity, however, the replicability and utility of the identified ADHD biotypes require further research.

Motivated by this, we aim to develop innovative MRI-compatible, deep learning approaches to identify reliable ADHD biotypes in addressing the afore-mentioned medical issues. Deep learning (DL) has gained

wide attention in mental disorder analysis due to its capability in automatically learning useful feature expression from high dimensional neuroimaging data, though its application in ADHD is still rare. The time is ripe to capitalize on these sophisticated methodologic advances and foster the convergence between clinical psychiatry and imaging neuroscience. In biotype division, compared to traditional clustering methods, deep clustering is able to learn nonlinear properties, represent low-dimensional embeddings from high-dimensional neuroimaging input, and constrains these learned features to adapt to the clustering target, which maximizes the between-class distance and minimizes the within-class distance[21-24]. Specifically, both convolutional neural network[25] and generative adversarial network[26] have been employed to identify biotypes, *e.g.*, from mild cognitive impairment. However, one issue is that clustering methods often ignored the topological information of population networks[21-24], *i.e.*, complex associations between individuals were not fully explored, which may lead to sub-optimal performance in biotype identification. Considering graph convolutional network (GCN) is able to encode topological information of population networks to improve disease prediction[27], and has performed well for community detections such as social networks[28] and protein interaction networks[29], we are motivated to integrate deep learning advances such as GCN and deep K-means to identify reliable, multi-factorial signatures for imaging-derived ADHD biotypes; so we can take an initial step towards developing biotype-guided treatment recommendations. Furthermore, public imaging data of children offers unprecedented opportunities for ADHD research. With the aid of publicly shared data from Adolescent Brain Cognitive Development (ABCD)[30] study and local longitudinal medication treatment data, we have an excellent platform with sufficient statistical power and enriched multi-factors (fMRI, demographics, cognition, symptoms) for testing and validating the proposed method and the identified ADHD biotypes.

Consequently, we proposed a graph convolution network for biological subtype detection (GCN-BSD) using both functional imaging and non-imaging phenotypic data. As illustrated in **Fig. 1**, we first construct a population graph based on functional network connectivity (FNC, **Fig. 1a**) and phenotypic information (age, gender) to build individual mappings; where FNCs serve as the feature of the nodes and the similarities between two subjects, which were abstracted from gender and age serve as edges (see detailed FNC and network derivation in Method section). Then we applied GCN-BSD to learn embeddings that are both group-discriminative between ADHD and controls, as well as adapted to the clustering constraint through K-Means loss (**Fig. 1b**). We selected 1069 ADHD patients from 11875 children aged 9-11 from the ABCD study as the discovery dataset, identified K=2 for biotype division via the cluster sum of square (CSS) using the elbow method[31], and evaluated the clustering performance of 4 popular algorithms with proposed method based on Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI)[32] (**Fig. 1c**). As a result, two ADHD biotypes were identified, manifesting with different FNC patterns and distinguishing cognitive abilities.

More importantly, we used a valuable ADHD dataset with longitudinal medication treatment to validate the reliability and utility of the identified ADHD biotypes from ABCD. This data was collected from Peking University Sixth Hospital (PKU6, 130 ADHD, 105 controls, aged 9-15) with fMRI, cognitive metrics, and

symptom records. Interestingly, we found that ADHD biotypes identified in ABCD and PKU showed high similarity and replicability in FNC patterns. The most contributing FNCs and clinical records were compared dedicatedly between two biotypes (**Fig. 1d**). Specifically, biotype 1 presented milder symptoms while biotype 2 manifested more severe hyperactivity/impulsivity symptoms and worse cognitive levels. Finally, we compared the symptom relief and treatment outcome of two biotypes from 44 ADHD patients either treated by MPH or ATX at PKU6 according to our division (**Fig. 1e**). Interestingly, biotype 1 treated with methylphenidate (MPH) showed significantly better recovery than biotype 2 treated with atomoxetine (ATX) ($p < 0.05$). Collectively, this is a novel attempt to identify and validate ADHD biotypes that might facilitate ADHD personalized treatment, pushing the boundaries of what is possible through innovative deep learning algorithms and assisting in clinical drug selection to enhanced cure rates.

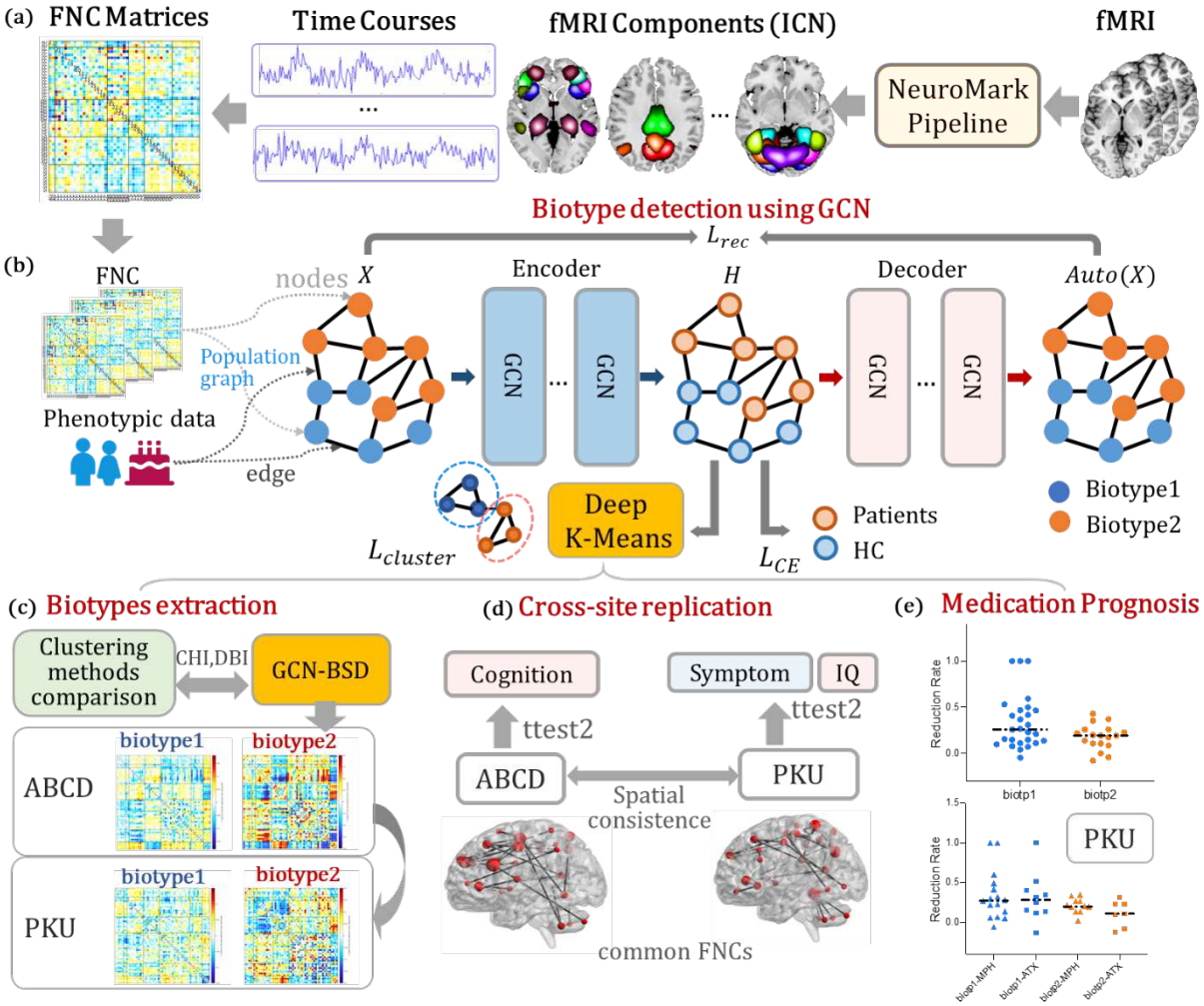


Fig. 1 Research outline. (a) Functional imaging feature extraction——functional network connectivity(FNC) deriviation. (b)Biotype detection based on GCN+deep k-means using population graph constructed via FNC and phenotypic data from the ABCD study. (c) Clustering method comparison using Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI), resulting in 2 biotypes, which were mapped to PKU. (d) biotype comparison and replication. (e) Comparison of the symptom relief and treatment outcome of two biotypes either treated by MPH or ATX using longitudinal data at PKU.

RESULTS

Two ADHD biotypes were identified, and the number of clusters was decided by the within-cluster sum of squared errors (SSE) using the elbow method[31], as shown in **Fig. 2a**. Through plotting the SSE as a function of the number of clusters k and picking the elbow of the curve, we chose $k = 2$ as the optimal number of clusters/subtypes. The performance of our GCN-BSD for significant characteristics along with conventional methods and deep methods with the input of FNC was reported in **Table 1** and **Fig. 2b**. The clustering performance is evaluated concerning two standard measures: Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI). As shown in **Fig. 2b**, GCN-BSD reached the lowest DBI and highest CHI. Meanwhile, GCN-BSD achieved the smallest P-values in several cognitive characteristics. In general, compared with the traditional clustering methods and naïve deep learning methods, GCN-BSD achieves better performance by its abstracted features and GCN module.

Table 1. Comparison of P-value in significant cognitive characteristics between 2 biotypes

Cognitive measures	P-values of 2 sample t-test between biotype 1 and biotype 2				
	Agg-Cluster	K-Means	DNN-KMeans	GCN-KMeans	GCN-BSD
PVT	0.07	0.02	2.83×10^{-3}	0.01	$\mathbf{1.22 \times 10^{-4}}$
LSWMT	0.03	0.01	8.83×10^{-3}	0.02	$\mathbf{1.35 \times 10^{-4}}$
DCCST	2.71×10^{-3}	1.67×10^{-4}	3.01×10^{-4}	1.21×10^{-4}	$\mathbf{2.87 \times 10^{-5}}$
PCPST	0.02	4.36×10^{-5}	2.27×10^{-5}	2.38×10^{-4}	$\mathbf{1.89 \times 10^{-6}}$
PSMT	0.10	0.01	0.03	0.04	$\mathbf{8.57 \times 10^{-4}}$
Fluid Cognition	4.66×10^{-4}	1.29×10^{-6}	7.25×10^{-6}	1.86×10^{-7}	$\mathbf{3.09 \times 10^{-8}}$
Total Cognition	8.92×10^{-4}	1.12×10^{-5}	1.25×10^{-5}	3.49×10^{-6}	$\mathbf{2.61 \times 10^{-7}}$

Note: PVT=picture vocabulary test, LSWMT=list sorting working memory test, DCCST=dimensional change card sort test, PCPST=pattern comparison processing speed test, PSMT=picture sequence memory test.

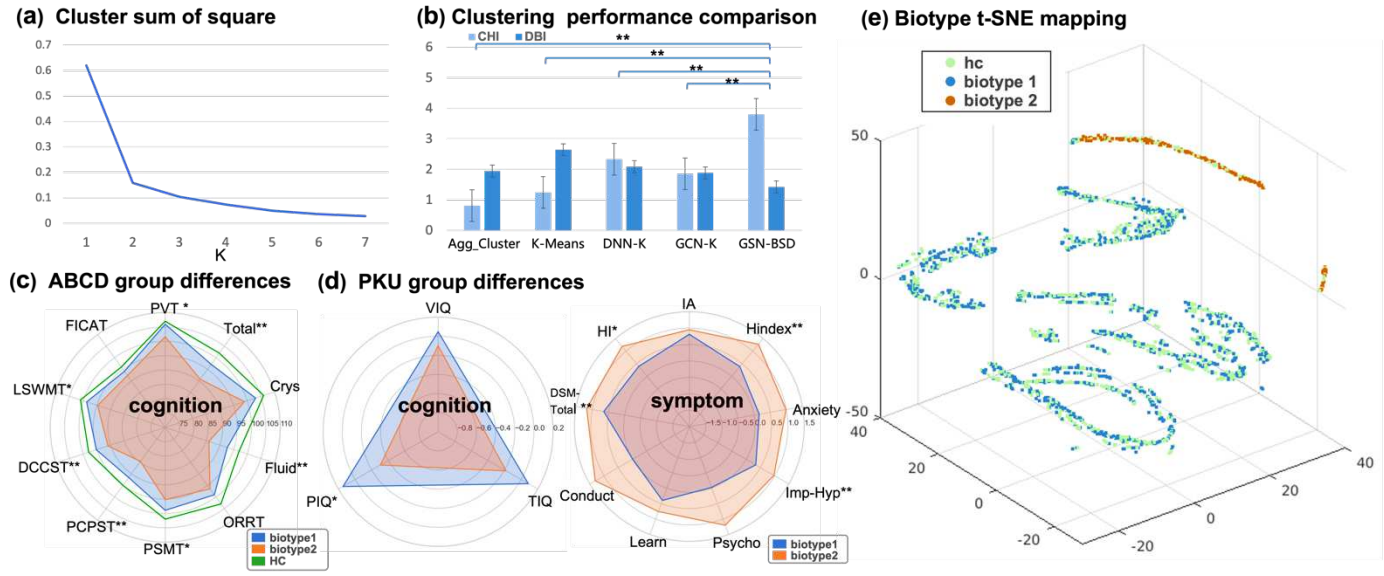


Fig. 2 ADHD biotypes detection results. (a) Cluster sum of square (CSS) across the number of clusters. **(b)** GCN-BSD shows the best clustering performance among the 5 methods, i.e., highest CHI and lowest DBI. **(c)** Mean cognitive scores for biotype 1 (blue), biotype 2 (yellow), and HC (green) in ABCD dataset, note that a higher score means better cognitive abilities. **(d)** Mean cognitive scores for biotype 1 (blue) and biotype 2 (yellow) in the PKU dataset, note that a higher score means better cognitive abilities in the left subgraph and a higher score means severer symptoms in the right subgraph. **(e)** Clustering results of the embedding features learned from GCN-BSD for ADHD biotypes and HC visualized by t-SNE.

ADHD biotypes extracted by GCN-BSD from ABCD.

Biotype 1 and biotype 2 have different connectivity patterns from HC. The mean differences of FNC between 2 biotypes and HC are shown in **Fig. 3a**. Fifty-three maximally independent component networks were arranged into 7 functional domains, including the subcortical (SC) (5 components), auditory (AU) (2 components), sensorimotor (SM) (9 components), visual (VI) (9 components), cognitive control (CC) (17 components), default mode (DM) (7 components), and cerebellar (CB)(4 components) networks. Biotype 1 shows positive connectivity within DM and between SM, CC, and CB, between VI and SM, SC, and negative connectivity between SM and DM, VI and CB. Biotype 2 shows more severe abnormality among all networks, compared to biotype 1. It shows positive connectivity between SM and DM, VI and CB, and negative connectivity between VI, CB, and SM, between DM and CB.

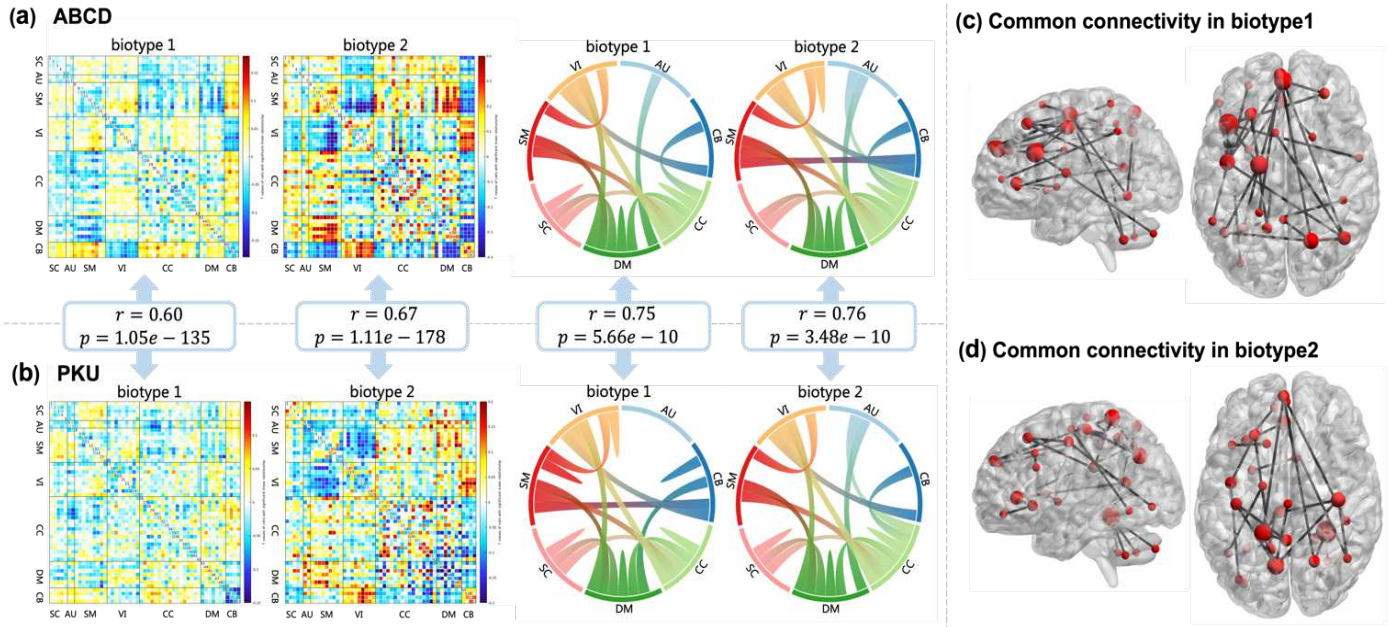


Fig. 3 ADHD biotypes detection results. (a) The mean differences of FNC between 2 biotypes and HC in ABCD and Top 100 connectivity patterns of biotype 1 and biotype 2 in the view of the contribution of functional networks. **(b)** The mean FNC differences between 2 biotypes-HC in PKU show high correlation with results from ABCD ($r>0.6$, $p<10^{-100}$ for both biotypes). Top 100 connectivity patterns of biotype 1 and biotype 2 in the view of the contribution of functional networks in PKU, showing high correlation with results from ABCD ($r=0.75$ for biotype 1 $r=0.76$ for biotype 2, $p<6\times10^{-10}$). **(c)** Overlapped top 100 discriminative FNCs for biotype 1 between ABCD and PKU. **(d)** Overlapped top 100 discriminative FNCs for biotype 2 between ABCD and PKU.

The two biotypes and demographically matched healthy controls were compared in cognitive measures (**Table 2**). Overall, across all ten cognitive scores, HC has the highest mean values, and biotype 1 has the lowest mean values among the three groups (**Fig. 2c**). The two biotypes are significantly different in seven out of 10 cognitive measures. Notably, fluid cognition ($p = 3.09 \times 10^{-8}$) and total cognition ($p = 2.61 \times 10^{-7}$) showed the most significant difference between the two biotypes.

ADHD biotypes replicated in PKU cohort

We further tested whether the identified connectivity pattern was consistent in the independent data cohort. By projecting the identified ADHD biotypes from ABCD into the PKU dataset, subjects were divided into 2 subsets. We detected the top 100 discriminative connectivity from 1378 FNCs in two biotypes by using

two-sample t-tests to estimate the group differences of FNC features. To explicitly characterize the contribution of each functional network, we grouped the 53 nodes into 7 networks according to the NeuroMark pipeline, as shown in **Fig. 3a and 3b**. Connectivity patterns of biotype 1 and biotype 2 were highly replicable from ABCD to PKU (correlation: $r = 0.75,\ p = 5.66 \times 10^{-10}$ for biotype 1, $r = 0.76$, $p = 3.48 \times 10^{-10}$ for biotype 2), and regions in the DM and CC exhibited the greatest involvement. For biotype 1, the results showed that the connectivity among CC, DM, SC, SM, and VI is overrepresented and repeatable. While for biotype 2, apart from the connectivity among multiple networks and VI, the overrepresented connections between AU and CC, DM, between CC and CB are also replicated between datasets. Then we expanded these to the whole FNC, the mean differences between biotypes and HCs in PKU still reveal a clear correlation to ABCD, where $r = 0.60, p = 1.05 \times 10^{-135}$ for biotype 1 and $r = 0.67, p = 1.11 \times 10^{-178}$ for biotype 2, as shown in **Fig. 3b**. Similar to the results from the ABCD dataset, biotype 1 shows positive connectivity between SC, AU, SM, CC, and CB, as well as SC and SM. And it also shows negative connectivity within CB and between DM and SM. As for biotype 2, consistent with the result from ABCD, it presents more impaired regions than biotype 1. Biotype 2 has positive connectivity between SM and DM, VI and CB, and negative connectivity between SM and VI, CB and DM, and within DM.

Table 2. Demographic and Cognitive Characteristics of the biotypes from ABCD.

Characteristics	biotype 1	biotype 2	P-value	Cohen's d
Num	**821**	**248**		
Age (months)	118.88±7.61	118.37±7.27	0.33	
Gender (M/F)	523/289	174/74	0.11	
PVT	105.97±15.81	101.62±15.37	1.22×10^{-4}	0.28
LSWMT	98.74±14.30	94.70±14.52	1.35×10^{-4}	0.28
DCCST	95.21±14.77	90.95±13.68	2.87×10^{-5}	0.30
PCPST	91.62±22.42	84.66±22.14	1.89×10^{-6}	0.31
PSMT	98.94±15.43	95.20±15.36	8.57×10^{-4}	0.24
Fluid Cognition	92.72±17.00	85.98±16.32	3.09×10^{-8}	0.40
Total Cognition	97.18±17.20	90.72±16.95	2.61×10^{-7}	0.38

Note: PVT=picture vocabulary test, LSWMT=list sorting working memory test, DCCST=dimensional change card sort test, PCPST=pattern comparison processing speed test, PSMT=picture sequence memory test.

Table 3. Demographic and Clinical Characteristics of the biotypes from PKU

Characteristics	biotype 1	biotype 2	P-value	Cohen's d
Num	**89**	**41**		
Age(years)	9.61±2.88	9.41±2.56	0.49	
Gender(M/F)	82/21	32/9	0.84	
RS_HI	18.64±4.27	22.22±5.16	4.97×10^{-4}	0.68
RS_Total	45.14±6.79	48.82±6.42	3.82×10^{-3}	0.56
PIQ	107.41±15.28	100.89±11.59	0.01	0.48
CPRS_IH	4.94±2.27	6.33±2.75	8.67×10^{-3}	0.55
CPRS_H-Index	12.96±4.18	15.23±4.66	0.01	0.51

Note: RS_HI= RS-IV_hyperactive/impulsive symptoms, RS_Total = RS-IV_ total ADHD symptoms, PIQ = Performance IQ, CPRS_IH= CPRS_impulsivity–hyperactivity factor, CPRS_H-Index =CPRS_ hyperactivity index.

In PKU, two-sample t-tests were used to estimate the group differences of the IQ scale and symptom scale, see **Table 3**. Performance IQ measured by the Wechsler Child/Adult Intelligence Scale distinguished itself ($p = 0.01$). Further, the severity of ADHD symptoms was evaluated by the ADHD Rating Scale-IV (ADHD RS-IV) and Conners' Parent Rating Scale (CPRS)[33]. And we noticed that hyperactive/impulsive

symptoms and total ADHD symptoms from RS-IV, and impulsivity–hyperactivity factor and hyperactivity index from CPRS showed significant differences between biotype 1 and biotype 2 ($p = 4.97 \times 10^{-4}$ for RS-hyperactive/ impulsive, $p = 3.82 \times 10^{-3}$ for RS-total, $p = 8.67 \times 10^{-3}$ for CPRS-impulsivity–hyperactivity, $p = 0.01$ for CPRS-hyperactivity index).

FNC pattern consistency across datasets.

Moreover, as shown in **Fig. 3c and d,** we extracted the overlapping FNCs between ABCD and PKU, resulting in 30 FNCs for biotype 1 and 27 for biotype 2. Both in the biotype 1 and biotype 2, the connections between the superior parietal lobule and posterior cingulate cortex, precuneus in DM, between the precuneus and anterior cingulate cortex, superior frontal gyrus, between the precentral gyrus and superior frontal gyrus are overrepresented. Separately, the key nodes in biotype 1 are the paracentral lobule, superior parietal lobule, cerebellum, frontal gyrus, and precentral gyrus, while in biotype 2 those are fusiform gyrus, posterior cingulate cortex, precuneus, anterior cingulate cortex and cerebellum. For biotype 1, the connectivity within the frontal gyrus and their connections with other regions including the cerebellum, middle temporal gyrus, paracentral lobule, precentral gyrus, and inferior parietal lobule are overrepresented. While for biotype 2, FNCs exhibited considerably denser patterns and greater degrees in DM regions including the precuneus, anterior cingulate cortex, and posterior cingulate cortex. And connectivity between the cerebellum and fusiform gyrus, and between the insula and anterior cingulate cortex are overrepresented for biotype 2.

The consistent discriminative FNCs between biotype 1 and biotype 2 were highly correlated with several cognition scales. For example, the connectivity between the superior parietal lobule and precuneus was correlated with total intelligence ($r = -0.16, p = 2.47 \times 10^{-7}$), fluid intelligence ($r = -0.14, p = 3.44 \times 10^{-6}$), crystallized intelligence ($r = -0.13, p = 8.46 \times 10^{-6}$) from ABCD baseline cognitive battery. For PKU, this connectivity was correlated with RS-hyperactive/impulsive ($r = 0.20, p = 0.03$), and CPRS-impulsivity–hyperactivity ($r = 0.22, p = 0.02$), CPRS-hyperactivity index ($r = 0.20, p = 0.03$). The correlation results of the connectivity between the middle temporal gyrus and paracentral lobule, between the superior parietal lobule and posterior cingulate cortex, and the specialized connectivity cerebellum-fusiform gyrus in biotype 2 are listed in Table S4.

Comparison of medication effect of two biotypes

One of the important applications of biotype detection is to observe their clinical progression and differences under treatment; therefore, we further evaluated symptom improvement within the ADHD patients under the treatment of ATX and MPH in week 1, 2, 3, 4, 8. For ADHD children under medication from PKU, 27 of 44 patients were classified as biotype 1 and the rest of them as biotype 2. To test whether there are differences in the benefits of the drug between 2 biotypes, we compared the symptom scores RS-IV and CPRS from baseline to week 8. As shown in **Fig. 4**, we found symptom reduction in both biotype 1 and biotype 2 from baseline to endpoint. Note that no significant group difference in symptom scales was identified at baseline, however, in week 8 the reduction rate of inattention ($p = 0.01$) and total ($p = 0.02$) scores in RS-IV and learning problem ($p = 0.04$) score in CPRS showed significant group difference. After

medication, biotype1 and biotype2 showed similar declines in symptoms from week 1 to week 4. After week 4, biotype 1 still showed a steady decline, while biotype 2 showed a minor reduction in multiple symptoms.

Furthermore, we analyzed the prognosis of using different medication in two biotypes. 16 of 26 subjects in biotype 1 and 11 of 18 subjects in biotype 2 took MPH, and the rest of them took ATX. The demographic information and differences in symptom reduction between the ATX group and MPH group were shown in the Supplementary file. Our results showed that there was a significant group difference between biotype1 with MPH and biotype 2 with ATX in inattention ($p = 0.05$), hyperactive/impulsive ($p = 0.05$), and the total score ($p = 0.03$) from RS-IV in week 8.

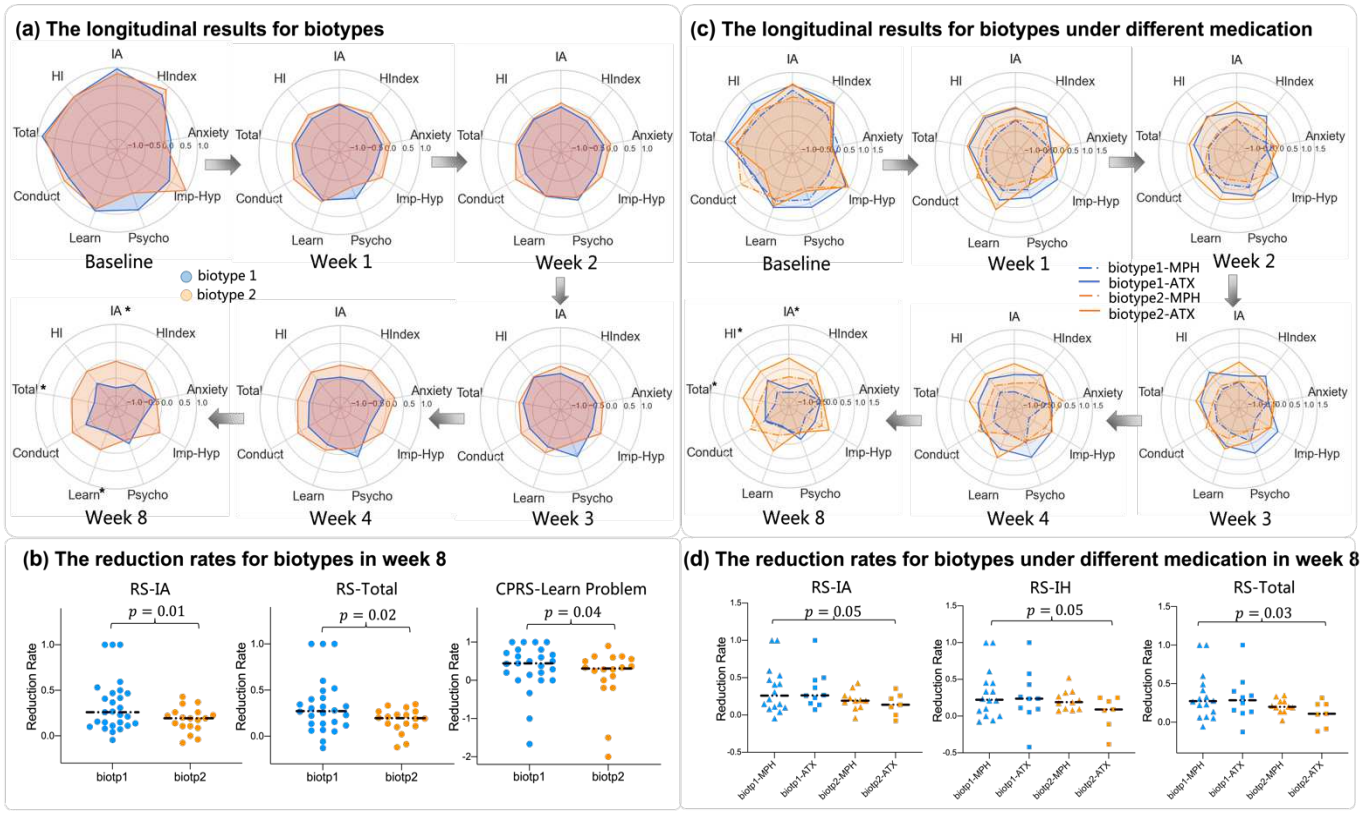


Fig. 4 The longitudinal results for patients under medication. (a) The longitudinal results for biotypes under medication from PKU in terms of the symptom scores RS-IV and CPRS from baseline to week 8. Note that a lower score means milder symptoms. From the baseline to week 8, biotype 1 showed a steady decline in all ADHD symptoms. While for biotype 2, the reduction after week 4 stagnated. **(b)** At week 8, the reduction rate showed significant group differences at RS-inattention ($p = 0.01$), RS-total ($p = 0.02$) and CPRS-learning problem ($p = 0.04$). **(c)** The longitudinal results for biotypes with ATX or MPH in terms of the symptom scores RS-IV and CPRS from baseline to week 8. **(d)** At week 8, we found significant group difference between biotype 1 with MPH and biotype 2 with ATX in inattention ($p = 0.05$), hyperactive/impulsive ($p = 0.05$) and total score ($p = 0.03$) from RS-IV, all FDR corrected.

DISCUSSION

In this study, we proposed a novel GCN-based deep clustering framework, compared it with several clustering methods, validated its utility in ADHD biotype detection, and identified two clinically discrepant biotype patterns. Biotype 1 contained more patients, presented milder symptoms, and overrepresented several wildly recognized brain aberrations including frontal gyrus and cerebellum. In contrast, biotype 2 included fewer patients, presented more severe symptoms especially hyperactive/impulsive, and showed greater degrees in regions from DM to SM, as well as the connectivity between the cerebellum and fusiform gyrus.

Based on this, we considered biotype 1 as a typical but mild ADHD pattern and biotype 2 as an atypical but more severe one. To the best of our knowledge, this is a novel attempt to introduce population association into the identification of disorder biotypes and identified two ADHD biotypes showing significant group differences.

Disease heterogeneity has attracted much interest, for it elucidates the implicit pathological mechanisms and makes it possible to offer individual treatment options to patients from different subgroups. For example, Varol et al. present an algorithm HYDRA based on SVM to simultaneously identify a class of pathological samples and separate them into coherent subgroups[34]. Building on HYDRA, Wen et al. proposed MAGIC to capture multi-scale representations of disorder subgroups and exploited a double-cycle optimization method[35]. In contrast to existing clustering approaches presented in[34-36], we considered the population association by introducing population graph GCN into deep clustering, for heterogeneity is a property that can only be expressed at the population level instead of the individual level. Population graph GCN has been proven effective in disease classification by a number of previous studies[37-39]. Our GCN-BSD not only outperformed conventional clustering methods but also significantly outperformed the deep learning-based clustering model. When using the publicly accessible dataset ABCD, lowest DBI and highest CHI were achieved in ADHD biotype detection. Both as traditional methods, K-Means showed decreased P-values in characteristics compared to agglomerative clustering. As expected, we observed that the DNN with K-Means and GCN with K-Means performed better than the two conventional approaches. In addition, our validation in the dependent dataset further indicated the effectiveness of our proposed method. Our ADHD biotype results are consistent across two independent cohorts, as shown in **Fig.3 a** and **b.** Consistent with the cognitive ability results, PKU biotypes were differentiable in performance IQ and further several symptoms including hyperactive/impulsive. ADHD biotypes identified by GCN-BSD from ABCD not only make sense in the objective index but also be reproducible and show significance in clinical scales.

The pattern of biotype 1 is in accord with the classical ADHD model as a disorder of deficient frontoparietal and multiple regions activation, while biotype 2 expresses an unconventional pattern. Regarding the specific abnormalities identified in biotype 1, this typical but mild biotype is overrepresented in multiple connectivity from DM to SM and CC, besides, the anterior cingulate cortex, cerebellum, paracentral lobule, and frontal gyrus also present as the key nodes. The frontal gyrus is hypo-activated in various tasks ranging from working memory to time discrimination[40,41]. ADHD has long been thought to reflect dysfunction of prefrontal–striatal–cerebellar circuitry, but accumulating evidence suggests that the prefrontal–striatal model of ADHD should be extended to include other circuits and their interrelationships from the perspective of systems neuroscience[42]. Consistent with the previous studies[43,44], the biotype 1 supported a model that ADHD-related dysfunctions are not only involved in higher-level cognitive-behavioral functions, such as the frontoparietal and default networks, but also in sensorimotor processes, including SM[45]. Note that all the remaining regions implicated in the prefrontal–striatal–cerebellar model of ADHD[46,47] are components of the frontoparietal circuit known as the executive control circuit[48,49]. As for the unconventional biotype 2, we identified connectivity which not overrepresented in biotype 1. We noticed that

besides the frontoparietal circuit, the special connectivity cerebellum-fusiform gyrus is highly correlated with multiple measures, including total intelligence, fluid intelligence, crystallized intelligence, inattentive and hyperactive/impulsive symptoms. The fusiform gyrus territory has received considerable attention for its role in reading[50,51], moreover, its damage might lead to the letter-by-letter reading strategies shown by neuroanatomical investigations[52,53]. Our identified biotype 2 is consistent with this investigation that though the cerebellum and fusiform gyrus do not connect physically, their functional connection contributes to reading abilities, phonological processing, and semantic memory[54]. This abnormal connectivity might further influence cognitive abilities, and lead to inattention and hyperactive symptoms.

As the most prominent network in the clinical neuroscience literature regarding spontaneous intrinsic brain activity, DM underlies some of the executive function deficits, working memory deficits, and attention lapses in ADHD patients[55,56]. Our results showed that the extracted common top 100 discriminative FNCs from two separate cohorts showed great overlap and correlated with multiple cognition and symptom scales. Among top 100 FNCs, the fundamental connection between the superior parietal lobule and posterior cingulate cortex, and precuneus are highly correlated with multiple cognitive abilities and symptoms, indicating that these connections from DM regions to the superior parietal lobule are associated with several attention problems[57,58]. The superior parietal lobule has been shown to play a major role in the voluntary shift of spatial visual attention and attention switching, and thus to be particularly important for regulated attentional processes and attention disengagement[59,60]. Besides, the common connection in DM from the precuneus to the anterior cingulate cortex is also considered a candidate locus of dysfunction in ADHD[61,62]. A functional connectivity analysis suggested that structural and functional circuits linking the anterior cingulate cortex to the precuneus may represent “small-world network” long-range connections[63]. Our results affirmed the importance of DM in ADHD analysis and put forward a presumption that the connection between DM and the superior parietal lobule might also be a key circuit.

Most interestingly, we illustrate the treatment record varying from 1, 2, 3, 4, 8 weeks after the treatment between the two biotypes. Though only part of the subjects received medication treatment, we still found that biotype 1 showed a steady decline in all ADHD symptoms, and biotype 2 approached healthy children at a smaller degree after 8 weeks, which reflected that biotype 2 suffered from severe symptoms from the side. Multiple studies revealed that MPH and ATX improve attention functions and upregulate abnormal fronto-cortical activation during executive function tasks in ADHD patients[64-66]. Specifically, both of the drugs have a moderating effect on the DM and cerebellum. DM serves as a core neuropathological mechanism of attention deficit behavior and cerebellar degree centrality was significantly associated with learning function performance[67,68]. Therefore, differential DM and cerebellar connectivity patterns between the two biotypes may influence drug responsiveness, consequently leading to varying symptom improvement outcomes, including inattention, total, and learning problem scores. Furthermore, we found reduction rate divisible in different biotypes with different medications, particularly, MPH treatment in biotype 1 significantly outperforms ATX treatment in biotype 2, implying that for patients with poorer treatment outcomes, MPH may be a preferable treatment option. Previous studies showed that stimulant medications like MPH are the

most effective treatments for ADHD, while the non-stimulant ATX shows slightly lower but good efficacy in reducing ADHD symptoms[69]. Both biotype 2 showing server abnormality and the effectiveness of MPH might illustrate our results. Due to the limitation of our dataset, we need further study with more subjects to support our hypothesis.

Some limitations should be considered. First, ABCD is an ongoing prospective longitudinal study starting at the ages of 9-10 and includes a diverse sample enrolled at 21 research sites across the U.S. The PKU dataset is collected from a local hospital, which contains potential ADHD patients. These two cohorts might be different in data distribution, even after regressing the factor of site and scanner. This limitation likely reduces the similarity of the ADHD biotype patterns from two datasets. Second, only some patients from PKU under medication have the longitudinal scale information. Therefore, our analysis of medication influence is an explorative attempt. We might include much more patients under medication to make a detailed analysis in the future.

Collectively, in this study, we proposed a novel framework, GCN-BSD, that can jointly characterize brain imaging data and phenotypic association and further use this knowledge to guide disease biotype detection. The population graph provides association information that the DNN-based model ignores. Importantly, the identified two ADHD biotypes exhibit significant group differences in functional networks and multiple cognitive abilities and symptoms, especially in fluid intelligence and hyperactive/impulsive. All the above findings indicate the validation of the frontoparietal circuits to serve as a key signature to ADHD and provide the first evidence for the connection from the cerebellum to the fusiform gyrus to be used as a biomarker in the uncommon subgroup. This study helps move forward from a conventional biotype detection approach to the use of a more flexible deep learning-based analysis.

METHODS and MATERIALS

Study design

A population graph was first constructed based on functional network connectivity (**Fig. 1a**) and phenotypic information (age, gender) to build individual mappings; where FNCs serve as the feature of the nodes and the similarities between two subjects, which were abstracted from gender and age serve as edges. Then we applied GCN-BSD to learn embeddings that are both group-discriminative between ADHD and controls, as well as adapted to the clustering constraint through K-Means loss (**Fig. 1b**). We selected 1069 ADHD patients from the ABCD study as the discovery dataset, identified K=2 for biotype division via the cluster sum of square (CSS) using the elbow method[31], and evaluated the clustering performance of 4 popular algorithms, including (1) agglomerative clustering, (2) conventional K-Means, (3) DNN with deep K-Means, and (4) autoencoder GCN with K-Means with GCN-BSD based on Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI)[32] (**Fig. 1c**). As a result, two ADHD biotypes were identified, manifesting with different FNC patterns and distinguishing cognitive abilities. Then we used 130 ADHD and 105 controls collected from Peking University Sixth Hospital as validation dataset to test the generalizability and potential clinical use of the identified biotypes. Interestingly, we found that ADHD biotypes identified in ABCD and PKU

showed high similarity and replicability in FNC patterns(**Fig. 1d**). The most contributing FNCs and clinical records were compared dedicatedly between two biotypes. Specifically, biotype 1 presented milder symptoms while biotype 2 manifested more severe hyperactivity/impulsivity symptoms and worse cognitive levels. Finally, we compared the symptom relief and treatment outcome of two biotypes from 44 out of 130 ADHD patients either treated by MPH or ATX at PKU6 according to our division (**Fig. 1e**).

The discovery ABCD cohort

The ABCD study (https://nda.nih.gov/abcd/) is an observation assessment of brain development in children in the US at 9-11 years of age to characterize psychological and neurobiological development from early adolescence to early adulthood[30]. Our data were analyzed from the ABCD Study curated annual release 3.0, which contains baseline data from 11,875 children. We determined ADHD same as previous studies[70,71] by past and present mental disorder diagnoses using parent-reported responses to the self-administered computerized Schedule for Affective Disorders and Schizophrenia for School-Age Children for DSM-5 (K-SADS-5).

Independent PKU cohort

Another in-house dataset was included in this study as a validation. This children's cohort was approved by the Ethics Committee of Peking University Sixth Hospital. A total of 130 children with ADHD and 105 HCs (6-15 years old) were recruited. 44 patients received medical treatment, 17 of them were treated with atomoxetine (ATX), and 27 of them were treated with methylphenidate (MPH), and we recorded their symptom scales at baseline, week 1, week 2, week 3, week 4, and week 8. The diagnosis was made by a senior psychiatrist based on the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL)[72], which is a clinical and semi-structured interview based on the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV). The severity of inattentive symptoms, hyperactive/impulsive symptoms, and total ADHD symptoms of all subjects were evaluated by the ADHD Rating Scale-IV (ADHD RS-IV). Conners' Parent Rating Scale (CPRS)[33] was used to assess the hyperactive/impulsive symptoms(see more details in the Supplementary file). The demographic information was listed in **Table S2**.

Image acquisition and processing

The resting-state fMRI data from two datasets were preprocessed with a combination of FMRIB Software Library v6.0 toolbox and Statistical Parametric Mapping 12 toolbox, including rigid body motion and distortion correction, spatial normalization into Montreal Neurological Institute space, and smoothing with a 6-mm full width at half maximum Gaussian kernel (please see more details in Supplementary file). We decomposed the preprocessed MRI by a fully automated spatially constrained ICA run on each individual subject separately based on the neuromark_fMRI_1.0 template (available in GIFT; http://trendscenter.org/software/gift and as a separate download at http://trendscenter.org/data) and pipeline used in[73,74]. For this template, fifty-three independent component networks (ICNs) were identified and

arranged into 7 functional networks, including the subcortical (SC) (5 components), auditory (AU) (2 components), sensorimotor (SM) (9 components), visual (VI) (9 components), cognitive control (CC) (17 components), default mode (DM) (7 components), and cerebellar (CB) (4 components) networks. FNC is calculated using Pearson correlation, resulting in a 53×53 matrix for each subject. We extracted the upper triangle elements of the matrix as FNC features, namely, each subject has an FNC vector in the dimension of (53×52)/2 = 1378. Participants from the ABCD dataset were selected based on the availability of non-imaging measures and high-quality neuroimaging data. The quality assessment of the neuroimaging data was based on the "Quality Control and Recommended Inclusion Criteria for Structural Data" provided as a document in the "Release Notes Imaging Instruments" folder. Participants were excluded from the present analyses because of incomplete rsMRI data at baseline and/or because brain scans did not pass the ABCD Study's quality control. After all quality control, we have 1069 ADHD participants included (see more details in the Supplementary file). And 1164 demographically matched HCs were included in our study. Their demographic and cognitive information were listed in **Table S1**.

Cognitive abilities and symptom scales

The ABCD baseline cognitive battery assessed by a well-validated National Institute of Health Toolbox was used as evaluation in this study[75,76], which covers Picture Vocabulary Test Score (PVT); Flanker Inhibitory Control and Attention Test Score (FICA); List Sorting Working Memory Score (LSWM); Dimensional Change Card Sort Test Score (DCCS); Pattern Comparison Processing Speed Test Score (PCPS); Picture Sequence Memory Test Score (PSM); Oral Reading Recognition Test Score (ORR); Cognition Fluid Composite Score; Crystallized Composite Score; Cognition Total Composite Score. Age-corrected T-scores were used in this study where higher values correspond to better performance.

For patients from PKU, the severity of inattentive symptoms, hyperactive/impulsive symptoms, and total ADHD symptoms of all subjects were evaluated by the ADHD Rating Scale-IV (ADHD RS-IV). Besides, the Conners' Parent Rating Scale (CPRS) was used to assess the symptoms in children participants with ADHD. The CPRS is a widely used instrument for screening and evaluating ADHD-related symptoms as well as other behavioral problems frequently associated with ADHD in children. It contains 48 items and can be divided into six factors: conduct problems, learning problems, psychosomatic problems, impulsivity–hyperactivity, anxiety, and hyperactivity index. Verbal IQ, performance IQ, and full-scale IQ were measured by the Wechsler Child/Adult Intelligence Scale, Third Edition.

GCN for biotype detection

GCN-BSD is an autoencoder-based GCN model. The proposed architecture consists of three main elements: an encoder, an embedding layer, and a decoder, as shown in **Fig. 1b**. The encoder and decoder are connected via the embeddings. By putting the embeddings into the fully connected layer, we adopted the binary cross entropy function to learn group-discriminative features, which identified the subject labels as patients or healthy controls for all subjects. Besides, we put the patient subgraph with the connection unchanged into the clustering layer for biotype detection. The deep K-Means loss function is adopted as the

clustering loss. To better extract high-resolution embeddings, we also adopted the reconstruction loss. The GCN-BSD can then be formulated as minimizing the following function:

$$L = \sum \gamma_1 L_{rec}(X) + \gamma_2 L_{CE}(X) + \gamma_3 L_{cluster}(X) \quad (1)$$

where $\gamma_1, \gamma_2, and\ \gamma_3$ regulate the trade-off among seeking good representation for X, which is set for biotype clustering, meanwhile is faithful to the original feature and maintains label information. More details can be found in the supplementary file.

Firstly, GCN-BSD was adopted in the ABCD dataset to identify ADHD biotypes. The number of clusters K was decided by the cluster sum of square (CSS) using the elbow method[31]. GCN-BSD was implemented with Pytorch and optimized using Adam optimizer[77]. Hyperparameters were set as: γ_1 =0.1, $\gamma_2 = 0.5,\ \gamma_3 = 1,$ and the learning rate was 0.001. To evaluate the clustering performance, we compared four alternative clustering methods with GCN-BSD, including (1) agglomerative clustering, (2) conventional K-Means, (3) DNN with deep K-Means, and (4) autoencoder GCN with K-Means. The clustering performance was measured by two internal cluster evaluation measures: Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI)[32], and one performance indicator: ABCD cognitive battery. Statistical analyses were performed among ADHD biotypes and HC. A two-sample t-test was performed for all measures between biotypes and HC. The P-values were corrected for multiple comparisons using Bonferroni correction.

Biotypes validation

To validate the constancy of the identified ADHD biotype patterns, we used the mean FNC feature of two biotypes as templates and projected them into an independent dataset from PKU to extract similar biotype patterns. Specifically, for each subject in PKU, we calculated the Euclidean distance between individual FNC and two FNC templates, and then we divided the subject into a subgroup where the template showed more similarity than the other. Two-sample t-tests were used to calculate the group differences between two biotypes of the symptom scales, including RS-IV, CPRS, and Wechsler Child/Adult Intelligence Scale. Note that sites and scanners were regressed out before statistical analysis.

The top 100 discriminative FNC in two biotypes were identified by two-sample t-tests in ABCD and PKU, respectively. We grouped the 100 FNCs into 7 brain networks according to the NeuroMark pipeline to analyze the functional connectivity on a large scale. Moreover, to discover the exact consistent brain patterns, we picked the overlapped top 100 FNCs in two separate datasets and obtained brain functional patterns for two biotypes. We further calculated the correlations between the identified overlapped FNC and ABCD cognitive battery, RS-IV, CPRS, Wechsler Child/Adult Intelligence Scale in two datasets, separately, in order to study the association between FNC and multiple symptoms.

Clinical progression of two biotypes

In order to explore whether the identified biotypes respond differently to medical treatment, we compared the treatment outcome under medication treatment of ATX and MPH after 1, 2, 3, 4 and 8 weeks. 44 out of 130 ADHD patients from PKU were recorded with longitudinal symptom and cognitive scores, who

were divided into two biotypes according to the previous results. Particularly, 27 of 44 patients were classified as biotype 1 and the rest of them as biotype 2. To test whether there are differences in the benefits of the drug between 2 biotypes, we compared the symptom scores RS-IV and CPRS from baseline to week 8 by repetive two sample t-test.

Acknowledgment

This work was supported by the scientific and technological innovation 2030 - the major project of the Brain Science and Brain-Inspired Intelligence Technology (2021ZD0200500), National key research and development program (2021YFE0202500), the National Natural Science Foundation of China (82022035, 61773380), China Postdoctoral Science Foundation (2022M710434), the National Institute of Health grants (R01MH117107, R01MH118695) and the National Science Foundation (2112455).

References

1. Association, A.P. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition - Coding Update. (2014).
2. Sobanski, E.*, et al.* Subtype differences in adults with attention-deficit/hyperactivity disorder (ADHD) with regard to ADHD-symptoms, psychiatric comorbidity and psychosocial adjustment. *European Psychiatry* **23**, 142-149 (2008).
3. Mentales, M. Kaplan & Sadock's synopsis of psychiatry. (2007).
4. Gorman, E.B., Klorman, R., Thatcher, J.E. & Borgstedt, A.D. Effects of methylphenidate on subtypes of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* **45**, 808-816 (2006).
5. Blondeau, C. & Dellu-Hagedorn, F. Dimensional analysis of ADHD subtypes in rats. *Biological psychiatry* **61**, 1340-1350 (2007).
6. Luo, N.*, et al.* Age-related structural and functional variations in 5,967 individuals across the adult lifespan. *Human brain mapping* **41**, 1725-1737 (2020).
7. Jiang, R.*, et al.* Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores. *Cerebral cortex* **30**, 888-900 (2020).
8. Zhi, D.*, et al.* Aberrant dynamic functional network connectivity and graph properties in major depressive disorder. *Frontiers in psychiatry* **9**, 339 (2018).
9. Guo, X.*, et al.* Shared and distinct resting functional connectivity in children and adults with attention-deficit/hyperactivity disorder. *Translational psychiatry* **10**, 1-12 (2020).
10. Molavi, P.*, et al.* ADHD subtype-specific cognitive correlates and association with self-esteem: a quantitative difference. *BMC psychiatry* **20**, 1-10 (2020).
11. Dobson-Patterson, R., O'Gorman, J.G., Chan, R.C. & Shum, D.H. ADHD subtypes and neuropsychological performance in an adult sample. *Research in Developmental Disabilities* **55**, 55-63 (2016).
12. Salvi, V.*, et al.* ADHD in adults: clinical subtypes and associated characteristics. *Rivista di Psichiatria* **54**, 84-89 (2019).
13. 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).
14. Grizenko, N., Paci, M. & Joober, R. Is the inattentive subtype of ADHD different from the combined/hyperactive subtype? *Journal of attention disorders* **13**, 649-657 (2010).
15. Li, T.*, et al.* Characterizing neuroanatomic heterogeneity in people with and without ADHD based on subcortical brain volumes. *Journal of child psychology and psychiatry* **62**, 1140-1149 (2021).
16. Karalunas, S.L.*, et al.* Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: toward biologically based nosologic criteria. *JAMA psychiatry* **71**, 1015-1024 (2014).
17. Kofler, M.J.*, et al.* Executive functioning heterogeneity in pediatric ADHD. *Journal of abnormal child psychology* **47**, 273-286 (2019).
18. Roberts, B.A., Martel, M.M. & Nigg, J.T. Are there executive dysfunction subtypes within ADHD? *Journal of attention disorders* **21**, 284-293 (2017).
19. Fair, D.A., Bathula, D., Nikolas, M.A. & Nigg, J.T. Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of the National Academy of Sciences* **109**, 6769-6774 (2012).
20. Dias, T.G.C.*, et al.* Characterizing heterogeneity in children with and without ADHD based on reward system connectivity. *Developmental cognitive neuroscience* **11**, 155-174 (2015).
21. Dong, A., Honnorat, N., Gaonkar, B. & Davatzikos, C. CHIMERA: Clustering of heterogeneous disease effects via distribution matching of imaging patterns. *IEEE transactions on medical imaging* **35**, 612-621 (2015).
22. Noh, Y.*, et al.* Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. *Neurology* **83**, 1936-1944 (2014).
23. Zhang, X.*, et al.* Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proceedings of the National Academy of Sciences* **113**, E6535-E6544 (2016).

24. Dong, A.*, et al.* Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain* **140**, 735-747 (2017).
25. Kwak, K.*, et al.* Subtyping of mild cognitive impairment using a deep learning model based on brain atrophy patterns. *Cell Reports Medicine* **2**(2021).
26. Yang, Z., Wen, J., Ahmed Abdulkadir, Y.C., Guray Erus, Elizabeth Mamourian, Randa Melhem, Dhivya Srinivasan, Sindhuja T. Govindarajan, Jiong Chen, Mohamad Habes, Colin L. Masters, Paul Maruff, Jurgen Fripp, Luigi Ferrucci, Marilyn S. Albert, Sterling C. Johnson, John C. Morris, Pamela LaMontagne, Daniel S. Marcus, Tammie L. S. Benzinger, David A. Wolk, Li Shen, Jingxuan Bao, Susan M. Resnick, Haochang Shou, Ilya M. Nasrallah, & Davatzikos, C. Gene-SGAN: a method for discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. *Axiv* **arXiv:2301.10772v1**(2023).
27. Parisot, S.*, et al.* Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease. *Medical Image Analysis* **48**, 117-130 (2018).
28. Devi, J.C. & Poovammal, E. An Analysis of Overlapping Community Detection Algorithms in Social Networks. in *12th International Conference on Communication Networks (ICCN) / 12th International Conference on Data Mining and Warehousing (ICDMW) / 12th International Conference on Image and Signal Processing (ICISP)*, Vol. 89 349-358 (Bangalore, INDIA, 2016).
29. Chen, J.C. & Yuan, B. Detecting functional modules in the yeast protein-protein interaction network. *Bioinformatics* **22**, 2283-2290 (2006).
30. Caseya, B.J., Cannoniera, T., Conleya, M.I., Cohenb, A.O. & Barchc, D.M. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroence* **32**(2018).
31. Marutho, D., Handaka, S.H., Wijaya, E. & Muljono. The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. in *2018 International Seminar on Application for Technology of Information and Communication (iSemantic)* (2018).
32. Maulik, U. & Bandyopadhyay, S. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on pattern analysis and machine intelligence* **24**, 1650-1654 (2002).
33. Conners, C.K., Sitarenios, G., Parker, J.D. & Epstein, J.N. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of abnormal child psychology* **26**, 257-268 (1998).
34. Varol, E., Sotiras, A., Davatzikos, C. & Initiative, A.s.D.N. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *Neuroimage* **145**, 346-364 (2017).
35. Wen, J.*, et al.* Multi-scale semi-supervised clustering of brain images: Deriving disease subtypes. *Medical Image Analysis* **75**, 102304 (2022).
36. Yang, Z.J.*, et al.* A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. *Nature Communications* **12**(2021).
37. Liu, J.*, et al.* MMGK: Multimodality Multiview Graph Representations and Knowledge Embedding for Mild Cognitive Impairment Diagnosis. *IEEE Transactions on Computational Social Systems* (2022).
38. Liu, L., Wang, Y.-P., Wang, Y., Zhang, P. & Xiong, S. An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Medical Image Analysis* **81**, 102550 (2022).
39. Song, X.*, et al.* Multi-center and multi-channel pooling GCN for early AD diagnosis based on dual-modality fused brain network. *IEEE Transactions on Medical Imaging* (2022).
40. Durston, S.*, et al.* Neural and behavioral correlates of expectancy violations in attention-deficit hyperactivity disorder. *Journal of child Psychology and Psychiatry* **48**, 881-889 (2007).
41. Konrad, K., Neufang, S., Hanisch, C., Fink, G.R. & Herpertz-Dahlmann, B. Dysfunctional attentional networks in children with attention deficit/hyperactivity disorder: evidence from an event-related functional magnetic resonance imaging study. *Biological psychiatry* **59**, 643-651 (2006).
42. Castellanos, F.X. & Proal, E. Large-scale brain systems in ADHD: beyond the prefrontal–striatal model. *Trends in cognitive sciences* **16**, 17-26 (2012).
43. Makris, N., Biederman, J., Monuteaux, M.C. & Seidman, L.J. Towards conceptualizing a neural systems-based anatomy of attention-deficit/hyperactivity disorder. *Developmental neuroscience* **31**, 36-49 (2009).
44. Peterson, B.S.*, et al.* An FMRI study of the effects of psychostimulants on default-mode processing during Stroop task performance in youths with ADHD. *American Journal of Psychiatry* **166**, 1286-1294 (2009).
45. Cortese, S.*, et al.* Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. *American Journal of Psychiatry* **169**, 1038-1055 (2012).
46. Nakao, T., Radua, J., Rubia, K. & Mataix-Cols, D. Gray matter volume abnormalities in ADHD: voxel-based meta-analysis exploring the effects of age and stimulant medication. *American Journal of Psychiatry* **168**, 1154-1163 (2011).
47. Valera, E.M., Faraone, S.V., Murray, K.E. & Seidman, L.J. Meta-analysis of structural imaging findings in attention-deficit/hyperactivity disorder. *Biological psychiatry* **61**, 1361-1369 (2007).
48. Vincent, J.L., Kahn, I., Snyder, A.Z., Raichle, M.E. & Buckner, R.L. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of neurophysiology* **100**, 3328-3342 (2008).
49. Menon, V. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in cognitive sciences* **15**, 483-506 (2011).
50. Cohen, L.*, et al.* Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* **125**, 1054-1069 (2002).
51. McCandliss, B.D., Cohen, L. & Dehaene, S. The visual word form area: expertise for reading in the fusiform gyrus. *Trends in cognitive sciences* **7**, 293-299 (2003).
52. Gaillard, R.*, et al.* Direct intracranial, FMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* **50**, 191-204 (2006).

53. Hirshorn, E.A., *et al.* Decoding and disrupting left midfusiform gyrus activity during word reading. *Proceedings of the National Academy of Sciences* **113**, 8162-8167 (2016).
54. Alvarez, T.A. & Fiez, J.A. Current perspectives on the cerebellum and reading development. *Neuroscience & Biobehavioral Reviews* **92**, 55-66 (2018).
55. Greicius, M.D., Krasnow, B., Reiss, A.L. & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences* **100**, 253-258 (2003).
56. Qiu, M.-g., *et al.* Changes of brain structure and function in ADHD children. *Brain topography* **24**, 243-252 (2011).
57. Fang, A., *et al.* Maladaptive self-focused attention and default mode network connectivity: a transdiagnostic investigation across social anxiety and body dysmorphic disorders. *Social cognitive and affective neuroscience* **17**, 645-654 (2022).
58. Tan, X., *et al.* Altered functional connectivity of the posterior cingulate cortex in type 2 diabetes with cognitive impairment. *Brain imaging and behavior* **13**, 1699-1707 (2019).
59. Donner, T., *et al.* Involvement of the human frontal eye field and multiple parietal areas in covert visual selection during conjunction search. *European Journal of Neuroscience* **12**, 3407-3414 (2000).
60. Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F.E. & Linden, D.E. Attentional systems in target and distractor processing: a combined ERP and fMRI study. *Neuroimage* **22**, 530-540 (2004).
61. Castellanos, F.X., *et al.* Cingulate-precuneus interactions: a new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biological psychiatry* **63**, 332-337 (2008).
62. Uddin, L.Q., *et al.* Network homogeneity reveals decreased integrity of default-mode network in ADHD. *Journal of neuroscience methods* **169**, 249-254 (2008).
63. Bassett, D.S. & Bullmore, E. Small-world brain networks. *The neuroscientist* **12**, 512-523 (2006).
64. Rubia, K., *et al.* Methylphenidate normalises activation and functional connectivity deficits in attention and motivation networks in medication-naive children with ADHD during a rewarded continuous performance task. *Neuropharmacology* **57**, 640-652 (2009).
65. Chou, T.-L., Chia, S., Shang, C.-Y. & Gau, S.S.-F. Differential therapeutic effects of 12-week treatment of atomoxetine and methylphenidate on drug-naive children with attention deficit/hyperactivity disorder: A counting Stroop functional MRI study. *European Neuropsychopharmacology* **25**, 2300-2310 (2015).
66. Bush, G., *et al.* Atomoxetine increases fronto-parietal functional MRI activation in attention-deficit/hyperactivity disorder: a pilot study. *Psychiatry Research: Neuroimaging* **211**, 88-91 (2013).
67. Cubillo, A., *et al.* Drug-specific laterality effects on frontal lobe activation of atomoxetine and methylphenidate in attention deficit hyperactivity disorder boys during working memory. *Psychological medicine* **44**, 633-646 (2014).
68. Fu, Z., *et al.* Shared and unique effects of long-term administration of methylphenidate and atomoxetine on degree centrality in medication-naïve children with attention-deficit/hyperactive disorder. *International Journal of Neuropsychopharmacology* **25**, 709-719 (2022).
69. Cortese, S., *et al.* Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *The Lancet Psychiatry* **5**, 727-738 (2018).
70. Lees, B., *et al.* Altered neurocognitive functional connectivity and activation patterns underlie psychopathology in preadolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **6**, 387-398 (2021).
71. Lees, B., *et al.* Association of prenatal alcohol exposure with psychological, behavioral, and neurodevelopmental outcomes in children from the adolescent brain cognitive development study. *American Journal of Psychiatry* **177**, 1060-1072 (2020).
72. Kaufman, J., *et al.* Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version. *Journal of the American Academy of Child & Adolescent Psychiatry* (1996).
73. Du, Y., *et al.* NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *Neuroimage Clin* **28**, 102375 (2020).
74. Fu, Z., *et al.* Dynamic functional network reconfiguration underlying the pathophysiology of schizophrenia and autism spectrum disorder. *Human brain mapping* **42**, 80-94 (2021).
75. Gershon, R.C., *et al.* NIH toolbox for assessment of neurological and behavioral function. *Neurology* **80**, S2-S6 (2013).
76. Gershon, R.C., *et al.* IV. NIH Toolbox Cognition Battery (CB): measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development* **78**, 49-69 (2013).
77. Kingma, D. & Ba, J. Adam: A Method for Stochastic Optimization. *Computer Science* (2014).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2022fengADHDsubtypesupplemetary.docx