nature mental health

Article

https://doi.org/10.1038/s44220-023-00195-w

Symptom dimensions of resting-state electroencephalographic functional connectivity in autism

Received: 23 June 2023

Accepted: 12 December 2023

Published online: 10 January 2024

Check for updates

Xiaoyu Tong[1], Hua Xie[2], Gregory A. Fonzo[3], Kanhao Zhao[1], Theodore D. Satterthwaite[4,5], Nancy B. Carlisle[6] & Yu Zhang[1,7]

Autism spectrum disorder (ASD) is a common neurodevelopmental disorder characterized by social and communication deficits (SCDs), restricted and repetitive behaviors (RRBs) and fixated interests. Despite its prevalence, development of effective therapy for ASD is hindered by its symptomatic and neurophysiological heterogeneities. To comprehensively explore these heterogeneities, we developed a new analytical framework combining contrastive learning and sparse canonical correlation analysis that identifies symptom-linked resting-state electroencephalographic connectivity dimensions within 392 ASD samples. We present two dimensions with multivariate connectivity basis exhibiting significant correlations with SCD and RRB, confirm their robustness through cross-validation and demonstrate their conceptual generalizability using an independent dataset (n = 222). Specifically, the right inferior parietal lobe is the core region for RRB, while connectivity between the left angular gyrus and the right middle temporal gyrus show key contribution to SCD. These findings provide a promising avenue to parse ASD heterogeneity with high clinical translatability, paving the way for ASD treatment development and precision medicine.

Autism spectrum disorder (ASD) is a common neurodevelopmental disorder that manifests in differences in social interaction, communication and restricted and repetitive behaviors (RRBs) or fixated interests[1]. ASD affects quality of life in both children and adults, yet its therapeutics are limited. Heterogeneity in ASD, including variability in behavioral symptoms[2], genetic profiles[3] and neurobiology[4], is a major obstacle to developing effective treatment. To address this challenge, a better dissection of the psychopathological dimensions of ASD is needed to pave the way for the quantification of this heterogeneity and potential effective treatment[5].

The core diagnostic criteria of ASD comprises two symptom domains: social and communication deficits (SCDs) and fixated interests or restrictive and RRBs[1]. Using these criteria, many studies have refined the ASD-relevant clinical behavior measures[6,7] stemming from the gold standard tests for ASD[8,9], aiming to construct a better representation of the symptomatic heterogeneity of ASD. However, the psychopathological essence of these predefined behavioral domains is vague. For example, the RRB domains in the Autism Diagnostic Observation Schedule (ADOS)[8] and Social Responsiveness Scale (SRS)[10]

[1]Department of Bioengineering, Lehigh University, Bethlehem, PA, USA. [2]Center for Neuroscience Research, Children's National Hospital, Washington, DC, USA. [3]Center for Psychedelic Research and Therapy, Department of Psychiatry and Behavioral Sciences, Dell Medical School, The University of Texas at Austin, Austin, TX, USA. [4]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [5]Penn Lifespan Informatics and Neuroimaging Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [6]Department of Psychology, Lehigh University, Bethlehem, PA, USA. [7]Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, USA.
e-mail: yuzi20@lehigh.edu

share the same name (both named as RRB), while only showing modest correlation with each other[11]. Encouragingly, recent work suggested that machine learning techniques, such as sparse canonical correlation analysis (sCCA), can successfully establish data-driven symptom dimensions tailored to a neurophysiological basis[12], providing a promising approach to circumvent limitations on symptom quantification imposed by clinical measures typically constrained by arbitrary diagnostic criteria. Furthermore, another recent study established the data-driven dimensions of ASD symptoms and identified potential subtypes based on brain functional connectivity (FC) associated with genotypes[13], demonstrating the advantage of data-driven symptom dimensions over predefined behavioral traits.

In addition to the symptomatic heterogeneity, neuropathological heterogeneity—the variability in neurobiological phenotypes of individuals with ASD—is another important component of ASD heterogeneity hindering effective treatment development. Given the high variability observed across individual-level brain activity metrics, distilling disorder-specific information from overall variability would greatly facilitate the parsing of neuropathological heterogeneity. To this end, a machine learning technique termed contrastive learning, which extracts information that is unique in one dataset compared with the other, can be used to derive patient-specific neuroimaging features, thus eliminating the background information that is shared in both individuals with ASD and typically developing individuals (Methods). Notably, a recent study used a contrastive learning technique to identify principal neuroanatomical dimensions in ASD using structure-based neuroimaging data, demonstrating the potential for contrastive learning to reveal potential dimensions of neuropathology in individuals with ASD[14]. Together, these studies provided evidence that neuroimaging data can be successfully used to establish data-driven symptom scores and links between symptoms and their corresponding neurophysiological bases.

However, all the aforementioned studies evaluated structural and functional magnetic resonance imaging (MRI). Despite its popularity and demonstrated utility in studies defining neural circuitry biomarkers[12,14,15], the clinical utility of MRI is limited because of substantial requirements in terms of expertise, specialized equipment and prohibitive cost. Alternatively, electroencephalography (EEG) provides another neuroimaging data modality that is more cost-effective and easy to operate, facilitating the translation of research findings to improved clinical practice[16–18]. In the current study, using resting-state EEG (rsEEG) FC as neurophysiological features, we integrated contrastive learning with sCCA to parse heterogeneity within ASD by identifying symptom-linked neurophysiological dimensions (Fig. 1). Source-level reconstruction of rsEEG was used for its larger effect size of relationships with phenotypes[19] and the generalizability to selection of reference channel[20]. Thus, we identified two pairs of linked dimensions between neurophysiology and ASD symptoms, which corresponded to the SCD and RRB scales, respectively. We confirmed the robustness of these dimensions through cross-validation and further demonstrated their conceptual generalizability using an independent dataset with different behavior assessments. Moreover, we further investigated the rsEEG FC signature associated with SCD or RRB symptoms. Together, our findings suggested two distinct brain mechanisms underlying SCD and RRB symptoms, providing a dissection of both symptomatic and neuropathological heterogeneities in ASD. This may more effectively motivate the development of effective ASD treatment and the achievement of precision medicine in clinical practice.

Results

Anchoring ASD symptom dimensions using rsEEG connectivity

To extract disorder-specific features in individuals with ASD, we first performed contrastive principal component analysis (cPCA)[15] using the rsEEG FC features on the Autism Biomarker Consortium for Clinical Trials (ABC-CT) dataset[21] (n_{TD} = 239, n_{ASD} = 392). Specifically, the covariance of typically developing individuals quantified the subclinical (background) variability, which was then subtracted from the overall patient variability measured as the covariance of patient data. Rigorously, the subclinical variability we defined in this way included both the disorder-unrelated information and the symptomatology below the diagnosis threshold. The resulting difference in covariance between overall patient variability and subclinical variability quantified the disorder-specific (foreground) variability, yielding contrastive FC features. To confirm the advantage of contrastive features, we conducted representational similarity analysis to assess the association between each type of connectivity features (standard, subclinical, contrastive) and clinical measures or demographic features. Thus, we found that the standard variability of individuals mostly represented the subclinical variability, demonstrating the necessity of distilling disorder-specific information with contrastive learning techniques. Moreover, contrastive features were significantly more associated with symptom severity for most clinical measures compared with standard or subclinical features, suggesting the superiority of contrastive features in representing disorder-related variability in individuals with ASD (Supplementary Figs. 1 and 2).

Then, we conducted sCCA[22] to identify linked dimensions between contrastive FC features and symptom profiles in individuals with ASD. To establish data-driven symptom scores for ASD, we incorporated a total of 24 subscales from the Autism Diagnostic Observation Schedule (ADOS)[8], Autism Impact Measure (AIM)[7], SRS[10] and Vineland Adaptive Behavior Scale (VABS)[23] to quantify the symptom profile of each individual (Supplementary Fig. 3). Thus, two pairs of generalizable linked dimensions were identified and cross-validated (tenfold) between symptom profiles and contrastive FC features in theta band (correlation between symptom and FC dimensions, r = 0.70; the same correlation in cross-validation (CV), r_{CV} = 0.16; permutation test for cross-validation, $P_{permutation}$ = 0.005; Fig. 2a) and alpha band (r = 0.45, r_{CV} = 0.28, $P_{permutation}$ = 0.037; Fig. 2b), respectively. The stability of feature loadings was further confirmed by intra-class correlation (ICC) analysis on the feature weights of FC and symptom features across cross-validation runs (social communication deficits (SCDs): FC loading, ICC = 0.8873; symptom loading, ICC = 0.9907; RRB: FC loading, ICC = 0.6805; symptom loading, ICC = 0.8436; Supplementary Fig. 4). Notably, no generalizable dimensions were identified using standard principal components (PCs) of connectivity features.

The composition of the symptom dimension score for the theta band was attributed to the communication and socialization subscales in the VABS (Fig. 2a). Notably, the VABS assesses the functionality of adaptive behaviors, thus being negatively correlated with the SCD. Therefore, greater dimension scores indicated more severe symptoms of SCD. Moreover, the symptom and FC dimension scores for the theta band were strongly correlated with SCD (Fig. 2c and Supplementary Fig. 5a). These results suggest that the FC dimension identified in the theta band is linked to SCD as quantified by the VABS; thus, it is hereafter referred to as the SCD-linked dimension. Owing to the intercorrelations between behavioral scales in individuals with ASD (Supplementary Fig. 3), the symptom and FC dimensions also showed modest correlations with certain other behavioral measures (Fig. 2c and Supplementary Fig. 5a).

By contrast, the composition of the symptom dimension score for the alpha band was primarily driven by the RRB score in ADOS, with a small contribution from atypical behavior measured by the AIM (Fig. 2b). Greater dimension scores were associated with more severe symptoms for RRB. Importantly, the symptom and FC dimension scores indeed showed strong correlations with ADOS-RRB (symptom: r = 0.74, false discovery rate $P(P_{FDR}) = 6.44 \times 10^{-69}$; Supplementary Fig. 5b; FC: r = 0.40, $P_{FDR} = 7.28 \times 10^{-5}$; Fig. 2d). Like SCD-linked dimensions, the RRB-linked symptom and FC dimension scores showed modest correlations with certain other behavioral measures (Fig. 2d and Supplementary Fig. 5b). Notably, these dimension scores did not show as strong

https://doi.org/10.1038/s44220-023-00195-w

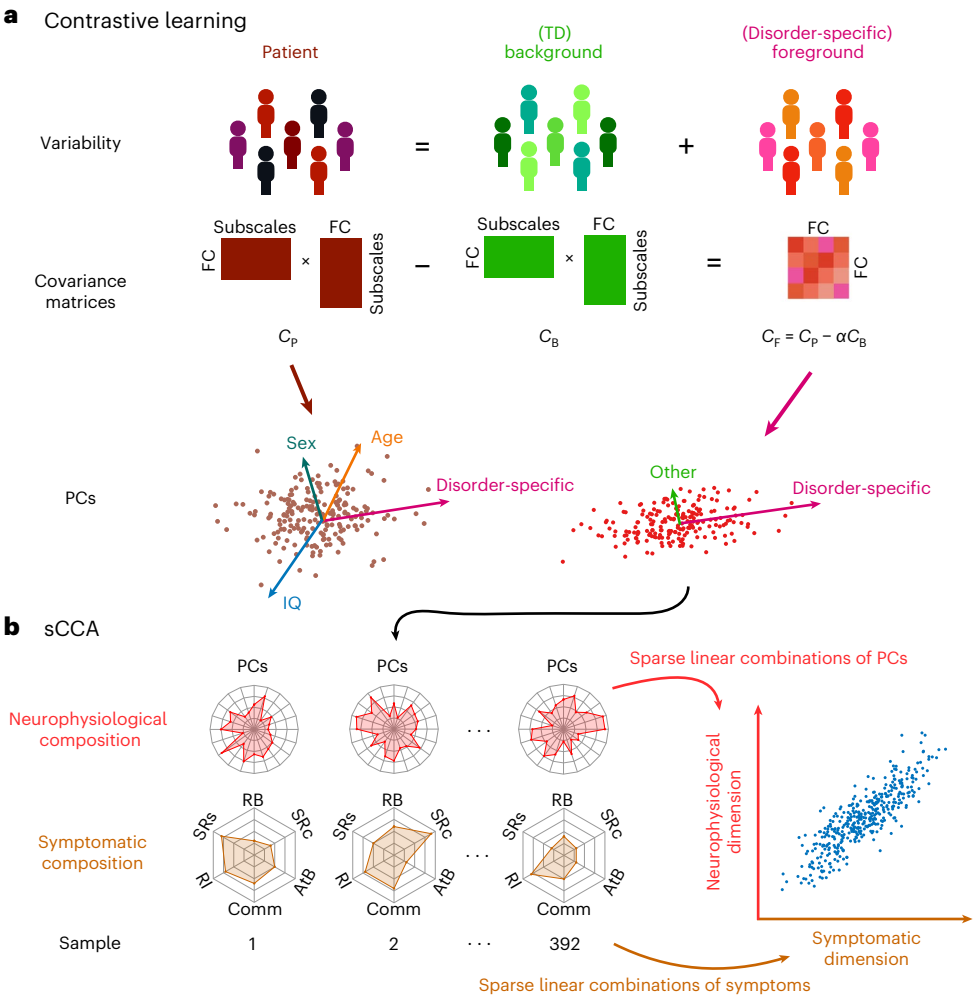


Fig. 1 | Flowchart of data-driven dissection of symptom-linked neurophysiological dimensions in individuals with ASD based on contrastive rsEEG FC features. a, Contrastive learning. Conceptually, variability in individuals with ASD consists of a background variability (shared by typically developing (TD) individuals) and a foreground variability (disorder-specific). Overall patient variability consisted of not only disorder-related variance, but also normative differences across individuals, such as age, sex and intellectual capacity (IQ). To improve the signal-to-noise ratio of FC features, we extracted contrastive features by subtracting background variability from overall patient variability. Specifically, we used covariance matrices to represent variability in individuals. Hence, the difference between the (weighted) covariance matrices of individuals with ASD (C_P) and typically developing individuals (C_B) represents the foreground variability (C_F) and is hypothetically related to mainly disorder-specific dimensions. Lastly, PCs were derived from the contrastive covariance matrix and used as the contrastive features for subsequent analyses. **b**, sCCA was then conducted with the obtained contrastive features to correlate the neurophysiological profiles of individuals with ASD with their symptom traits, including repetitive behavior (RB), social reciprocity (SRc), atypical behavior (AtB), communication (Comm), restricted interests (RI) and social responsiveness (SRs). The linked dimensions between neurophysiology and symptoms were then identified, which are essentially the sparse linear combination of contrastive PCs and symptom subscales, respectively.

correlations with the RRB score quantified using the SRS, partly because of the modest correlation between ADOS-RRB and SRS-RRB (Supplementary Fig. 3). Overall, these results demonstrated that the FC dimension identified in the alpha band is linked to the RRB score as measured by ADOS; thus, it is hereafter referred to as the RRB-linked dimension.

Subsequently, we conducted analyses to confirm that the identified dimensions were not driven by demographic differences. Specifically, we found that both SCD-linked and RRB-linked symptom and FC dimensions were not significantly correlated with age (SCD: symptom, $r = -0.084$, $P = 0.10$; FC, $r = -0.046$, $P = 0.36$; RRB: symptom, $r = -0.016$, $P = 0.76$; FC, $r = -0.023$, $P = 0.65$; Supplementary Fig. 6); there was no significant group difference in dimension scores between the sexes (SCD: symptom, $P_{ranksum} = 0.30$; FC, $P_{ranksum} = 0.38$; RRB: symptom, $P_{ranksum} = 0.16$; FC, $P_{ranksum} = 0.85$; Supplementary Fig. 7). Moreover, we examined the correlation between SCD-linked and RRB-linked dimension scores. While no significant correlation was observed between SCD-linked and RRB-linked FC dimensions ($r = 0.09$, $P = 0.07$), a significant positive correlation was observed between the two identified symptom dimensions ($r = 0.29$, $P = 4.13 \times 10^{-9}$). These results suggested that the SCD-linked and RRB-linked dimensions we identified (in the theta and alpha bands, respectively) may correspond to distinct underlying mechanisms of ASD symptoms in different EEG frequency bands. Meanwhile, these FC dimensions may be associated with co-occurring symptoms that vary in severity and contribute to the heterogeneity across individuals with ASD.

rsEEG connectivity signatures associated with ASD symptoms

Subsequently, we investigated the rsEEG FC signatures that constitute the symptom dimension of SCD and RRB. First, we examined the importance of network-level FC metrics (Methods). For SCD, FC within the default mode network (DMN) showed the highest importance (Fig. 3a). For RRB, FC within the DMN and FC between the frontoparietal control network and visual network showed the highest importance (Fig. 3b). We then examined which regions of interest (ROIs)

https://doi.org/10.1038/s44220-023-00195-w

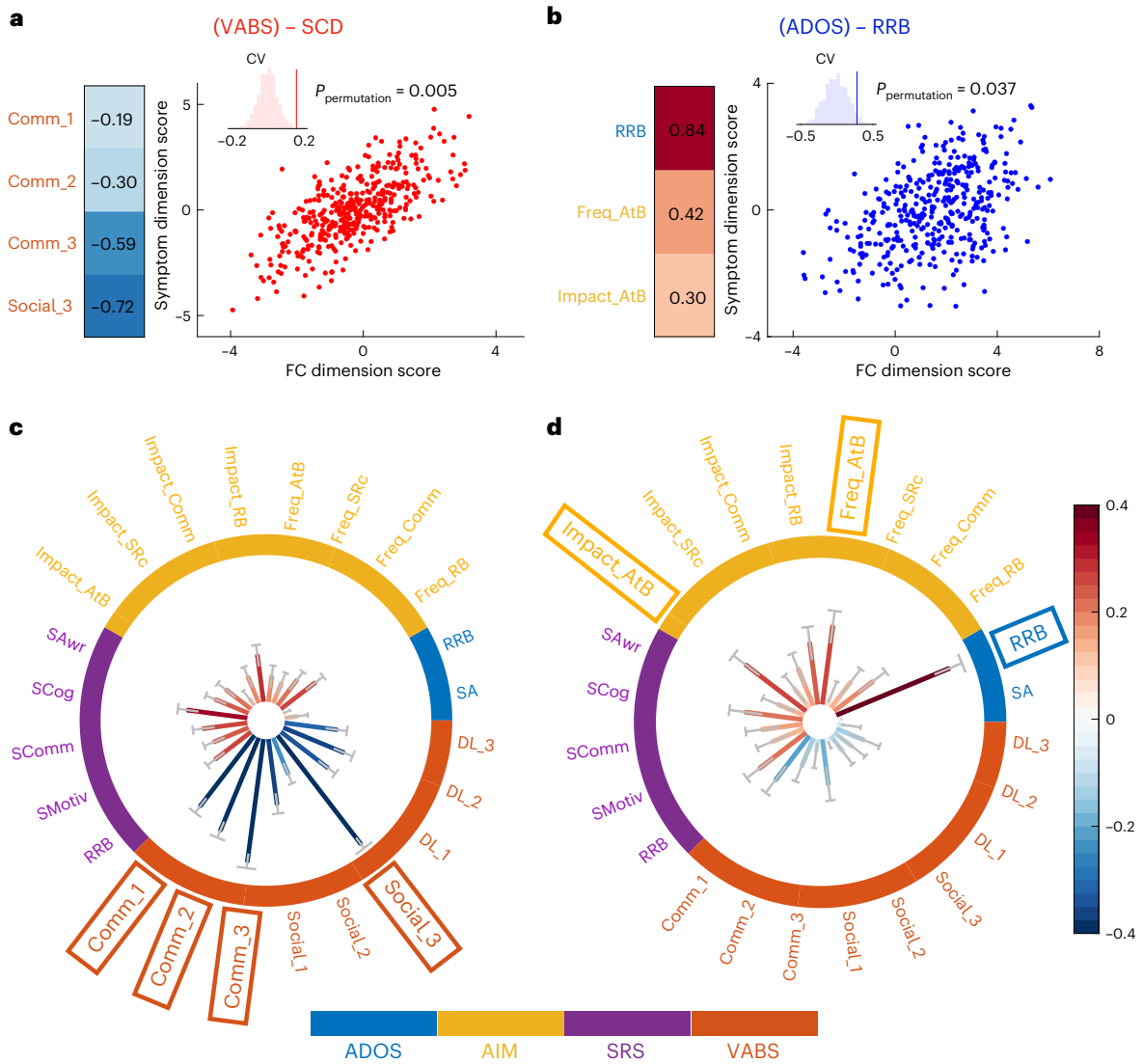


Fig. 2 | Linked dimensions between contrastive rsEEG features and symptom traits in individuals with ASD. a, The SCD-linked dimension. The symptom dimension score consists of deficits in communication (Comm_1, receptive; Comm_2, expressive; Comm_3, written) and socialization (Social_3, coping skills) functions measured using the VABS. The FC dimension score showed strong correlation with the symptom dimension score in the ASD samples ($r = 0.70$, $n = 392$). The robustness of this correlation was confirmed by cross-validation ($r_{CV} = 0.16$) and a permutation test ($P_{permutation} = 0.005$). **b**, The same figure as in **a** with the RRB-linked dimension. The FC dimension score showed strong correlation with the symptom dimension score ($r = 0.45$). The robustness of this correlation was confirmed by cross-validation ($r_{CV} = 0.28$) and a permutation test ($P_{permutation} = 0.037$). The histogram in the inset shows the null distribution of this correlation coefficient in cross-validation, where the vertical line indicates the actual value without permutation. Permutation tests were conducted 1,000 times. **c**, Correlations between SCD-linked FC dimension scores and clinical measures. The FC dimension score showed strong negative correlation with the level of communication and socialization skills while only showing modest correlations with other clinical measures, demonstrating its specific association with SCD symptoms. **d**, Correlations between RRB-linked FC dimension scores and clinical measures. The FC dimension score showed particularly strong positive correlation with RRB severity, while showing relatively modest correlations with other clinical measures, demonstrating its specific association with RRB symptoms. Each bar represents the correlation strength derived for all individuals with ASD ($n = 392$). The gray markers indicate the s.e.m. of the correlation coefficient, which is derived for Fisher's z test and then transformed to the space of the correlation coefficient. The color bar shows the value of correlation coefficients. For a complete list of clinical measures, see Supplementary Fig. 3.

contributed most substantially to each of the ASD symptom dimensions by calculating the node strength of each ROI. Node strength was quantified as the average absolute feature weights involving the particular ROI. Thus, we found that the left inferior parietal lobe, bilateral posterior middle frontal gyri and medial prefrontal cortex showed comparable contributions to the SCD-linked FC dimension (Fig. 3c). For the RRB-linked FC dimension, the right inferior parietal lobe (RIPL) showed the strongest contribution, with outstanding node strength compared to other ROIs (Fig. 3d). Afterward, we further examined the importance of ROI-level FC, as quantified by the feature weights for FC. Overall, the FC dimensions showed a multivariate basis of their compositions (Fig. 3e,f). Specifically, FC between the left angular gyrus (LANG) and right middle temporal gyrus (RMTG) showed great importance for SCD (Fig. 3g), while the composition of the SCD-linked FC dimension was mainly distributed across the frontoparietal control network (Fig. 3e and Supplementary Fig. 8a,c,e). Meanwhile, we identified FC between the left frontal eye field and the RIPL as particularly important for the RRB-linked FC dimension (Fig. 3h). Among the top five most important ROI-level FC metrics for RRB, four of them involved the RIPL (Fig. 3f,h and Supplementary Fig. 8b,d,f), indicating the strong

association of alterations in the neurophysiological activity of the RIPL with RRB symptoms.

Conceptual generalization analysis on an independent dataset

To confirm the generalizability of the identified symptom-linked FC dimensions, we conducted a validation analysis using an independent sample (n_{ASD} = 222) from the Healthy Brain Network (HBN) dataset[24]. We focused the generalization analysis on the FC dimensions because the HBN dataset used a different set of behavioral assessments. Specifically, we derived the transformation matrices (for SCD and RRB) from the ABC-CT dataset to convert standard rsEEG FC to the symptom-linked FC dimension scores. We then applied the same transformation matrices on the HBN samples without further training or tuning to calculate the symptom-linked FC dimension scores on the independent dataset.

Subsequently, we compiled available ASD-related behavioral scales in the HBN dataset with sufficient individuals (Methods), including SRS and Repetitive Behavior Scale (RBS)[25]. Afterward, we examined the correlations between the SRS subscales and symptom-linked FC dimension scores (Fig. 4a,c). As the SRS was the clinical measure shared by the ABC-CT and HBN datasets, we then compared these correlation coefficients with the results for the ABC-CT dataset, which indicated that there was no significant difference in correlation strengths when using the ABC-CT dataset or the HBN dataset (Fisher's *z* test: SCD: social awareness, P_{FDR} = 0.29; social cognition, P_{FDR} = 0.07; social communication, P_{FDR} = 0.29; social motivation (SMotiv), P_{FDR} = 0.45; RRB, P_{FDR} = 0.94. RRB: social awareness, P_{FDR} = 0.94; social cognition, P_{FDR} = 0.31; social communication, P_{FDR} = 0.83; social motivation, P_{FDR} = 0.88; RRB, P_{FDR} = 0.12). These results suggest that the symptom-linked FC dimensions did not have substantially different correlative relationships with common behavioral scales between cohorts.

Furthermore, to evaluate the generalizability of the association between identified FC dimensions and ASD symptoms, we examined the correlations between symptom-linked FC dimensions and the subscales of the RBS on the HBN dataset (Fig. 4b,d) because the subscales of the RBS have significant correlations with both the ADOS[26] and VABS[27]. Encouragingly, we found that the SCD-linked FC dimension showed significant positive correlations with social cognition (r = 0.11, P = 0.048) and RRB (r = 0.15, P = 0.012) in the RBS. Again, it is worth remembering that there is discrepancy between the RRB scores from the different scales (RBS versus ADOS), which might affect the correlation of RBS-RRB and the SCD-linked dimension score. The RRB-linked FC dimension also showed significant positive correlations with compulsive behavior (r = 0.15, P = 0.024) and restricted interests (r = 0.14, P = 0.031). Together, these results provide the generalizability of the association between symptom-linked FC dimensions and ASD symptom traits across unique samples and measures.

Discussion

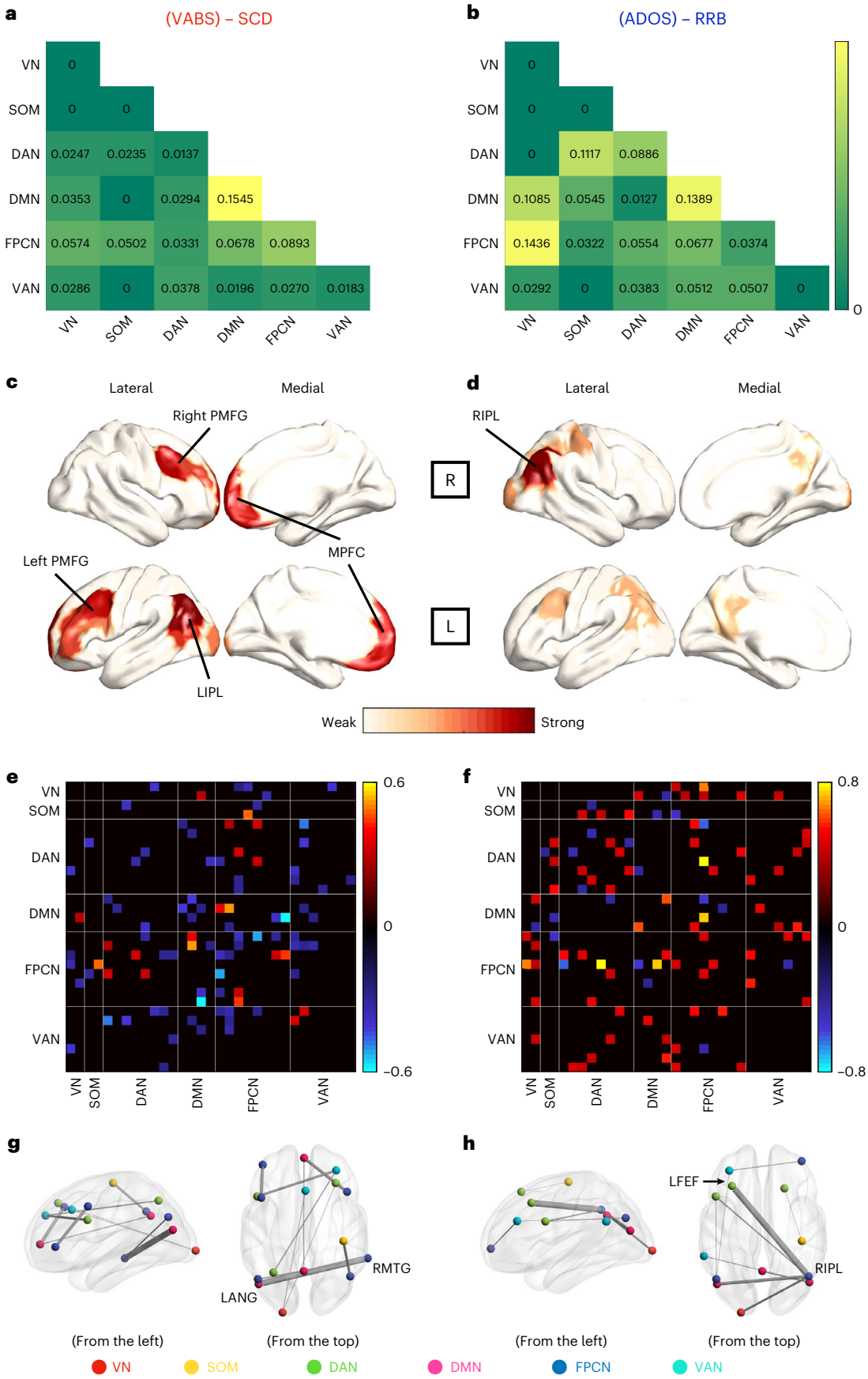
In this study, we identified two symptom-linked rsEEG FC dimensions in individuals with ASD by leveraging a combination of cutting-edge computational tools[12,15]. We showed that these two dimensions were mostly strongly associated with SCD and RRB symptoms, two of the core symptom domains based on the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition[1]. Importantly, the symptom-linked rsEEG FC dimensions showed generalizable correlations with the corresponding ASD symptom dimension scores, as confirmed by cross-validation and a conceptual generalization analysis on an independent dataset, demonstrating initial evidence of reliability and generalizability. These findings provided a promising avenue for understanding symptomatic and neuropathological heterogeneity in ASD, which may guide the discovery of neuroimaging-informed treatment targets.

Despite the popularity and success of structural and functional MRI in identifying the dimensions of brain structure and activity or psychopathology[12–15], its clinical utility is limited because of substantial requirements in terms of expertise, specialized equipment and high cost. We used rsEEG as an alternative data modality, which is more cost-effective and easy to operate. Additionally, thanks to its high temporal resolution, rsEEG FC can capture the typical or abnormal background oscillations (mainly alpha band oscillation) that are highly indicative of ASD diagnosis[28]. Moreover, our method also enables the simultaneous identification of latent symptom dimensions and corresponding neurophysiological dimensions. Taken together, these innovations distinguish our current study from existing work, demonstrating unique translational potential to improve clinical practice.

Notably, while we defined the neurophysiological dimensions as SCD-linked and RRB-linked, both SCD and RRB consisted of subscales[10,23,25] and showed considerable inconsistency with themselves across assessments[26]. For example, both the ADOS and SRS had RRB subscales, but they showed only modest correlations (r = 0.11, P = 0.023) in individuals with ASD in the ABC-CT dataset (n = 392). This discrepancy between subscales purporting to measure the same construct introduces additional heterogeneity to ASD symptom assessment. Therefore, we emphasize that the SCD and RRB symptom dimension scores we constructed are specifically associated with the RRB score measured by the ADOS and the SCD score composed by the VABS. Moreover, as the identified ASD symptom dimension scores (SCD and RRB) consisted of loadings from multiple subscales, they represent data-driven dimensions unique from any existing measures. As the generalizability analysis on dataset 2 was conducted on a different set of behavioral measures, future studies conducted on datasets with more comprehensive clinical assessments, and studies with more participants and potentially different rsEEG data preprocessing pipelines, are needed to confirm the composition of the SCD and RRB dimensions identified in this study. On verification of the reproducibility of these findings, we propose that these new symptom dimensions

Fig. 3 | Signature alterations in rsEEG FC for ASD symptoms. a,**b**, Importance of network-level FC for SCD (**a**) and RRB (**b**). The importance of network-level FC was calculated as the average absolute feature weights of ROI-level FC involving the particular network(s). The FC within the DMN showed high importance for the severity of SCD. The FC between the visual network (VN) and the frontoparietal control network (FPCN) and FCs within the DMN showed the highest importance for RRB severity, while the FCs between the DAN and SOM and FCs between the DMN and the visual network also showed substantial contributions. The color bar indicates the network-level FC importance. **c**,**d**, Node strength for the SCD (**c**) and RRB (**d**). Node strength was calculated as the average absolute feature weights of all ROI-level FC involving the particular ROI. The left inferior parietal lobe (LIPL), posterior middle frontal gyri (PMFG) and medial prefrontal cortex (MPFC) showed comparably strong contribution to the SCD-linked dimension. The right inferior parietal lobe (RIPL) showed the strongest contribution to the RRB-linked dimension, with outstanding node strength compared to other ROIs. The color bar indicates the ROI importance. The top ten ROIs are shown for clarity. The importance of network-level FCs and node strength was derived by averaging the absolute feature weights of the top 10% strongest ROI-level FCs for reliability and consistency. **e**,**f**, ROI-level FC pattern for the SCD-linked (**e**) and RRB-linked (**f**) dimensions. The hypo-connectivity between the LANG in the DMN and RMTG in the FPCN showed the strongest contribution to the SCD-linked FC dimension score. The hyper-connectivity between the left frontal eye field (LFEF) (in the dorsal attention network (DAN)) and the RIPL (in the FPCN) showed the strongest contribution to the RRB-linked FC dimension score. The top 10% ROI-level FCs are shown for visualization purposes. **g**,**h**, Important ROI-level FCs for SCD (**g**) and RRB (**h**). The FC between the LANG and RMTG showed the strongest association with SCD. The RIPL showed particularly strong association with the severity of RRB (four of the top five strongest ROI-level FC values involved the RIPL). The FC with the strongest association with RRB severity was the one between the RIPL and the LFEF. The top 10 ROI-level FCs are shown for clarity. SOM, somatomotor network; VAN, ventral attention network.

https://doi.org/10.1038/s44220-023-00195-w



should be used in future research to quantify ASD heterogeneity, as they exhibit optimal correlations with the underlying neurophysiological features (see Fig. 2a,b for their compositions). If all involved psychiatric assessments are not available, we suggest approximating these symptom dimension scores using the loadings of the corresponding measures from the VABS and ADOS for best performance.

https://doi.org/10.1038/s44220-023-00195-w

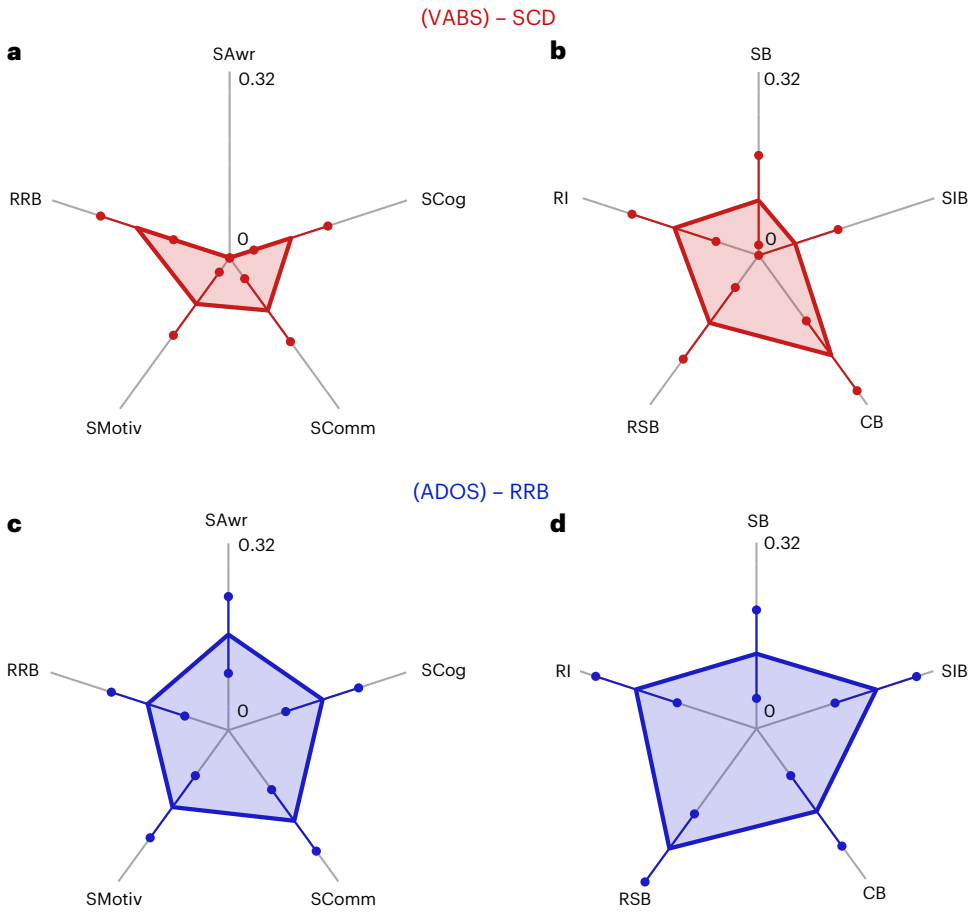


Fig. 4 | Association between symptom-linked FC dimensions and behavioral traits in the independent HBN dataset. a, Between the SCD-linked FC dimension and the SRS subscales. **b**, Between the SCD-linked FC dimension and RBS subscales. **c**, Between the RRB-linked FC dimension and the SRS subscales. **d**, Between the RRB-linked FC dimension and the RBS subscales. Symptom-linked dimensions showed comparable (no significant difference) correlation strengths with the SRS subscales between dataset 1 and the independent cohort; they also showed significant positive correlations with the RBS subscales. The circle markers indicate the s.e. of the correlation coefficients. The s.e. was derived for a Fisher's *z* test and then transformed to the space of the correlation coefficient. CB, compulsive behavior; RI, restricted interests; RSB, ritualistic/sameness behavior; SAWr, social awareness; SB, stereotypical behavior; SCog, social cognition; SComm, social communication; SIB, self-injurious behavior; SMotiv, social motivation.

Remarkably, the symptom-linked rsEEG FC dimensions identified in this study are well aligned with previous studies of rsEEG and MRI abnormalities in ASD[28,29]. Specifically, we found that the decrease in FC between LANG and RMTG contributed most substantially to SCD dimension symptom severity. While hypo-connectivity between the LANG[30,31] and RMTG[32–34] has been widely reported to be associated with ASD diagnosis, our results offer more precise insights into their association with specific ASD symptom dimensions. For RRB, we found that the increase in FC between the left frontal eye field and the RIPL contributed substantially to RRB domain symptom severity, echoing the findings of enhanced visual functioning[35] and altered anatomical cortical networks in ASD[36]. We also identified that alpha band RIPL activity was particularly important for RRB severity. Coinciding with our results, a previous pilot study[37] applied repetitive transcranial magnetic stimulation (rTMS) over the bilateral IPL to modulate alpha and low beta band activities; the authors found that rTMS modulation differed according to self-reported ASD symptom severity. These results support our findings that the RIPL may be a key region mediating RRB domain severity and offer a potential treatment target that can be tested to alleviate RRB symptoms.

Along with the findings from neuroimaging-based analyses, neurobiological studies showed that disrupted development and connectivity of GABA (gamma-aminobutyric acid)-producing (GABAergic) interneurons in the prefrontal and temporal cortices may be associated with ASD pathology[38,39]. Also, the alpha band is commonly associated with active inhibition during rest[40,41], a function achieved by GABAergic circuits[42]. Therefore, alpha band alterations in line with RRB severity could be attributed to the malfunction of GABAergic circuits in the RIPL. Notably, Parkinson's disease is another disorder that is attributed to dysfunction of GABAergic circuit-based active inhibition, although the dysfunction is proposed to be mostly neurodegenerative instead of neurodevelopmental as in ASD. Deep brain stimulation is used as the first-line therapy for Parkinson's disease[43], but rTMS has also shown promising treatment outcomes[44]. In ASD, rTMS is preferred over deep brain stimulation for its noninvasive characteristics. Considering the aforementioned pilot study applying rTMS over the bilateral IPL in individuals with ASD[37], rTMS could be a promising potential therapy for ASD. Although other rTMS trials for individuals with ASD have shown less compelling results[45–47], they applied rTMS to the prefrontal regions, which might not be as effective a target as the IPL, based on these and other findings. Also, previous studies did not consider heterogeneity in ASD, which may have reduced effect sizes. Taken together, we propose that rTMS over the IPL (specifically the RIPL) might be a promising and effective therapy for individuals with ASD with high RRB traits.

In summary, we successfully parsed symptomatic and neurophysiological heterogeneity in ASD and provided initial evidence of generalizability with cross-validation and an independent dataset. We identified SCD and RRB as two principal ASD neurophysiology-linked

symptom dimensions. The RIPL was identified as a key region associated with RRB symptom severity and might be targeted with rTMS as an intervention approach for individuals with ASD with high RRB traits. FC between the LANG and RMTG was the most highly associated with SCD symptom severity, as well as the comparably strong contribution of the left inferior parietal lobe, bilateral medial prefrontal cortex and medial prefrontal cortex, might also be targeted by potential therapies targeting challenges in social functioning. Our work provided a clinically translatable dissection of ASD heterogeneity, which offers a new constellation of symptom scores that measures clinical severity with neurophysiological perspectives and its corresponding brain biomarkers, which may lead to neuroimaging-guided treatment for ASD. Together, these findings hold great potential in motivating the discovery of new, effective treatments targeting specific ASD traits and guiding precision medicine for ASD tailored to individual-level symptom profiles.

Methods

Dataset 1 (ABC-CT)

Participants. The ABC-CT dataset[21] consisted of 280 children with ASD and 119 typically developing children whose ages ranged from 6 to 11. The study was approved by the Yale Institutional Review Board and registered with ClinicalTrials.gov (registration NCT02996669) with the study protocol available online[21]. Each individual had 1–3 EEG recordings and psychiatric assessments, resulting in 703 ASD samples and 364 typically developing samples. Because of low EEG data quality, 166 samples were excluded from our analysis (see 'EEG preprocessing' for details). Another 270 samples were also excluded because of incomplete psychiatric assessments, yielding 392 ASD samples and 239 typically developing samples that were included in our study. Diagnosis was made using the ADOS[8]. The full-scale IQ of individuals spanned from 60 to 150 as measured using the Differential Ability Scales[48]. Exclusion criteria included a known genetic syndrome, a neurological disorder putatively causally related to ASD, a known metabolic disorder and mitochondrial dysfunction. No treatment was administered in this clinical trial. However, medications were allowed to ensure the sample was representative with regard to participants with treatment for better generalizability of biomarker measurement, while a stable regimen (at least 8 weeks before enrollment) was required for inclusion. Written informed consent was provided by the legal guardian of each participant. Monetary compensation was provided for each participant.

Psychiatric assessments. The ABC-CT dataset provided comprehensive measurements for individuals with ASD and typically developing individuals. The diagnostic characterization relied on ADOS[8]. While the Autism Diagnostic Interview-Revised[9] scales were also proposed as diagnostic measures[21], they were excluded from our study because of incomplete data entries. The Differential Ability Scales and VABS were included as clinician-administered assessments for intellectual and cognitive function. Additionally, this dataset also provided the AIM, the Pervasive Developmental Disorder Behavior Inventory and SRS as caregiver questionnaires. However, the Pervasive Developmental Disorder Behavior Inventory was excluded from our study because of its limited data entries of only 50. The Clinical Global Impression Scale was also proposed as an outcome measure but was not included because of lack of available data entries. Taken together, based on data availability and relevance to ASD, we included the ADOS, VABS, AIM and SRS as ASD-related symptom measures in our study.

Remarkably, the ABC-CT study used a longitudinal design, aligning with the structure and timeline of actual clinical trials[21]. Each individual was assessed at three time points (baseline, 6 weeks after baseline and 24 weeks after baseline) for all the aforementioned behavioral measures, aiming for the evaluation of short-term test–retest reliability and developmental stability or change over a time period typical of clinical trials. Our results confirmed that the associations between symptom and FC dimension scores are general to all time points (Fisher's z test for the highest and lowest correlations; SCD, $P = 0.28$; RRB, $P = 0.15$; Supplementary Fig. 9). For information on the distributions of the demographics and psychiatric assessments for the ABC-CT dataset, see Supplementary Table 1.

EEG acquisition. The ABC-CT study used four EEG paradigms, including resting state (viewing abstract videos), human face event-related, biological motion event-related and visually evoked[21]. In this study, we focused on the rsEEG features to establish neuroimaging biomarkers for ASD symptoms in naturalistic conditions. Because the eyes closed session may not be reliably recorded for the participant age range (6–11), only the eyes-open sessions were collected. The rsEEG data were recorded using a 128-channel EEG geodesic HydroCel system by Electrical Geodesics. The recordings were at a sampling rate of 1,000 Hz. The recording reference was set to be Cz (head vertex) and the impedance of each electrode was checked before recording and was kept below 100 kΩ.

Dataset 2 (HBN)

Participants. The HBN study recruited children and adolescents aged between 5 and 21 at four study sites in the New York City area[24]. The study was approved by the Chesapeake Institutional Review Board. The recruitment criterion was having adequate verbal communication ability with the help of parents or guardians. Exclusion criteria included severe neurological disorders, severe impairment in cognition (IQ < 66), acute encephalopathy, known neurodegenerative disorder or other abnormalities that may prevent full participation in the protocol. Overall, the HBN incorporated several psychiatric disorders; we focused on ASD in this study. When we accessed the dataset, it included 185 typically developing individuals and 222 individuals with ASD (regardless of comorbidity) with available behavioral measures and neuroimaging data (one for each individual). Written informed consent was obtained from participants aged 18 or older and provided by the legal guardians of participants younger than 18. Monetary compensation was provided for each participant.

Psychiatric assessments. The HBN dataset provided comprehensive assessments for diagnostic psychiatry, intelligence, language and other cognitive functions in a transdiagnostic population[24]. In this study, we focused on ASD-related assessments. Although ADOS was proposed as the diagnostic assessment in the HBN dataset[24], it was unavailable when we acquired the data. We included the SRS and RBS[25] for the evaluation of ASD symptoms in the HBN dataset. Notably, the Gendered Autism Behavioural Scale and VABS were also related to the generalization analysis; however, there were not enough available data entries (no data were available for the Gendered Autism Behavioural Scale; only 24 individuals with ASD had both VABS assessments and EEG data available). Additionally, individuals aged 6–17 were also assessed with the Wechsler Intelligence Scale for Children to assess their intellectual capacity. Supplementary Table 1 provides information on the distributions of demographics and psychiatric assessments for the HBN dataset.

EEG acquisition. Unlike the psychiatric assessments, EEG data were collected at a single study site[24]. For the resting-state paradigm, which was designed to measure intrinsic brain activity during rest[49], participants were asked to view a fixation cross in the center of the computer screen and to open or close their eyes at various points as instructed. High-density EEG data were recorded in a sound-shielded room using a 128-channel EEG geodesic HydroCel system by Electrical Geodesics. The recordings were at a sampling rate of 500 Hz with a bandpass of 0.1–100 Hz; the recording reference was set to Cz (head vertex). The impedance of each electrode was checked before recording and was

kept below 40 kΩ. To match the EEG paradigm used in dataset 1, only eyes-open rsEEG data were included in this study.

EEG preprocessing

The recorded rsEEG data were cleaned offline with an EEGLAB (v.2019.1)-based fully automated artifact rejection pipeline[18]. The entire procedure included the following steps: (1) resampling of EEG to 250 Hz; (2) removal of 60 Hz AC line noise artifact[50]; (3) removal of nonphysiological low frequency in the EEG signals using a 0.01-Hz high-pass filter; (4) rejection of bad epochs by thresholding the magnitude of each epoch; (5) rejection of bad channels by thresholding the spatial correlations among channels; (6) exclusion of individuals with more than 20% bad channels; (7) estimate of the EEG signals from bad channels from the adjacent channels via spherical spline interpolation[51]; (8) independent component analysis to remove the remaining artifacts, including scalp muscle, ocular and ECG artifacts[52]; (9) re-referencing EEG signals to the common average; and (10) filtering of EEG signals to four canonical frequency ranges: theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz) and low gamma (31–50 Hz).

EEG source-space FC calculation

Using the Brainstorm toolbox (v.3)[53], we first implemented source localization using the minimum norm estimation approach[54] to convert the channel space EEG into the source-space signals of 3,003 vertices. Afterward, a three-layer (scalp, skull and cortical surface) boundary element head model was computed with the OpenMEEG (v.2.4) plugin[55] based on the FreeSurfer average brain template[56]. A total of 3,003 dipoles with unconstrained orientations were generated on the cortical surface. The lead field matrix relating the dipole activities to the EEG was obtained by projecting the standard electrode positions onto the scalp. For each individual, an imaging kernel that mapped from the channel space EEG to the source-space current density was then estimated using the minimum norm estimation approach with depth weighting and regularization. PCA was then used to reduce the three-dimensional estimated source signal at each vertex to the one-dimensional time series of the PC.

To capture the brain functional architecture, we extracted connectomic features by calculating the power envelope connectivity (PEC)[16] because it has demonstrated strength in mitigating spurious correlations resulting from volume conduction[57] and does not rely on precise lags between signals[58]. Hilbert transform was first applied to convert source estimates into analytical time series. The analytical time series of each pair of brain signals were then orthogonalized to remove the zero-phase-lag correlation[57]. Afterward, the power envelopes were measured by calculating the square of the orthogonalized analytical signals. A logarithmic transform was subsequently applied to enhance normality. The PEC was then calculated as the Pearson's correlation coefficient between the log-transformed power envelopes of each pair of brain regions. Finally, a Fisher's *r*-to-*z* transformation was performed to enhance normality[57]. The regional pairwise PEC features were further extracted for the 31 ROIs in the Montreal Neurological Institute template space[59]. For each pair of ROIs, connectivity was calculated by averaging the PEC values over all possible vertex pairs.

Contrastive learning framework for individuals with ASD

The core premise of the contrastive learning approach in studying individuals with ASD revolves around recognizing two distinct components within the neurophysiological variability measured by the rsEEG: the disorder-specific component and the background component. Unlike conventional covariate analysis, we did not explicitly specify the elements encompassed by the background component. Instead, we discerned the disorder-specific and background components of patient variability through a data-driven approach. Specifically, we defined the background component as the variability that is identifiable in typically developing individuals, and the disorder-specific component as the variability that is only identifiable in individuals with ASD (so that typically developing individuals cannot have it). In essence, this perspective posits that not all of the variability seen in individuals with ASD is attributable to the disorder, and the variability seen in typically developing individuals also exists in individuals with ASD.

Hence, the conventional rsEEG features for individuals with ASD may be noised by disorder-irrelevant information. To improve the specificity of the FC features to ASD symptoms, we extracted contrastive FC features using the cPCA technique[15]. The contrastive features essentially distilled the disorder-specific information from the overall variability. Specifically, 239 typically developing samples in the ABC-CT dataset were used as the background data to identify subclinical variability in FC features shared across typically developing individuals (quantified by their covariance matrix C_B). Meanwhile, 392 ASD samples in the ABC-CT dataset were used to identify variability in FC features shared across individuals with ASD (quantified by their covariance matrix C_B). The cPCA was jointly conducted on both groups to quantify the representative dimensions for disorder-specific variability (quantified by the contrastive covariance matrix $C_F = C_P - \alpha C_B$). The contrastive PCs were then calculated as the eigenvectors of the covariance matrix. As the variability in individuals with ASD can be conceptualized as the superposition of subclinical and ASD-specific variabilities, contrastive FC features were quantified as the difference between patient and subclinical PCs of the FC features. Lastly but importantly, some subclinical variability might share the same directions as ASD-indicative dimensions—they were subclinical only because typically developing individuals had smaller magnitude in those dimensions. Therefore, a contrastive parameter (α, between 0 and 1) was included to control the quantity of subtracted subclinical variability. The contrastive parameter was optimized using cross-validation for each EEG state with increments of 0.1. Supplementary Figs. 1 and 2 show the comparison across standard, subclinical and contrastive PCs for theta and alpha band EEG FC, respectively. Three hundred contrastive PCs were included in the subsequent analysis according to the number of positive eigenvalues of the contrastive covariance matrix (Supplementary Fig. 10) because the eigenvectors corresponding to negative eigenvalues are associated with noise components[15].

sCCA

sCCA is a multivariate procedure that seeks maximal correlations between linear combinations of variables in a pair of data matrices. Sparsity constraint was applied by L1 norm regularization to reduce the model complexity and alleviate the overfitting issue[22]. Collectively, these computational constraints defined a minimization problem:

$$\arg\min_{u,v}(-\mathrm{Cov}(Xu, Yv) + \lambda_1|u|_1 + \lambda_2|v|_1)$$

$$\text{subject to } \mathrm{Var}(Xu) < 1 \text{ and } \mathrm{Var}(Yv) < 1$$

where X and Y denote the data matrices, u and v denote the corresponding feature loadings and λ_1 and λ_2 are the regularization parameters. sCCA was used to identify generalizable linked dimensions between contrastive rsEEG FC and ASD-related symptom scores. Specifically, the contrastive PCs of rsEEG FC yielded by cPCA were first scaled to have a unit variance. This step was to ensure that the sparsity constraint was fair for all input features. No de-mean step was implemented after the cPCA step to make sure that the overall procedure (including the cPCA and sCCA steps) could be formularized as a linear transformation, to facilitate the generalization analysis on the independent dataset. Symptom measures were also normalized to have a zero mean and unit variance before the sCCA procedure.

Grid search for hyperparameters with cross-validation

Three hyperparameters were required for the entire analysis, including one contrastive parameter used in the cPCA and two sparsity

parameters used in the sCCA for contrastive FC features and symptom scores, respectively. These hyperparameters were optimized simultaneously within five rounds of tenfold cross-validation. For the cross-validation, the correlation between canonical dimensions of contrastive FC features and symptom scores was used as the metrics for the evaluation. The samples from the same individuals were either all in the training dataset or all in the testing dataset.

Conceptual generalizability analysis on the independent dataset

For the generalization analysis of symptom-linked dimension scores, we first multiplied the cPCA loading, scaling for unit variance, and the sCCA loading together to derive the equivalent linear transformation to convert standard rsEEG FC features to symptom-linked FC dimension scores for both RRB and SCD. Specifically, given the standard FC matrix X and the cPCA loading β, the contrastive PC score S is given by $S = X\beta$. Subsequently, the contrastive PCs were standardized to unit variance to ensure the fairness of the regularization of the sCCA. Denoting the ith PC as S_i and its s.d. across samples as σ_i, the standardized PC was $S_0 = \{S_{i,0}\}$, where $S_{i,0} = S_i/\sigma_i$. Lastly, given the sCCA loadings for contrastive PCs and behavioral assessments as W_S and W_Y, the FC (neurophysiological) dimension score was $U = S_0 W_S$ and the symptom dimension score was $V = YW_Y$, where Y is the data matrix for the original behavioral assessments. Because the cPCA loading, scaling for unit variance, and the sCCA loading are all linear transformations, they can be combined to an equivalent overall transformation matrix of $W_X = (\beta \odot (1/\sigma))W_S$, where $\odot$ represents the element-wise multiplication. Thus, the FC dimension score of each sample can be directly calculated as $U = XW_X$.

Afterward, we directly applied the same linear transformation on the rsEEG FC features in the HBN dataset, where the rsEEG FC features were also normalized for each individual as in the ABC-CT dataset. We then conducted correlation analysis on the derived FC dimensions with the symptom scores on the HBN dataset to test the aforementioned hypotheses. Both datasets were normalized for each individual before the procedure.

Importance of network-level FC

In principle, the importance of network-level FC is calculated as the average absolute feature weights of all ROI-level FC within and between the brain networks. To improve reliability, we only included the top 10% important ROI-level FC to derive the importance of network-level FC because the contribution of ROI-level FC with small loadings for the dimension score could be nonsignificant and unstable.

Representational similarity analysis

Representational similarity analysis is a data analytical framework that quantifies the mutual representation between brain activity measures and behavioral assessments[60], whose capability in probing the information encoded in latent features was demonstrated by a recent study[14]. We first calculated the standard PCs for individuals with ASD and typically developing individuals, respectively, in addition to the contrastive PCs. Subsequently, we derived the representational dissimilarity matrices for each PCA paradigm (ASD, typically developing, contrastive) by calculating the L2 norm of interindividual differences in PC scores. Specifically, each representational dissimilarity matrix was of a size of nIndividual by nIndividual, with each element as the L2 norm of pairwise differences in PC scores. (Because there are many PCs in each paradigm, the difference in PC scores is a vector.) Afterward, we also derived a representational dissimilarity matrix for the behavioral assessments by calculating the L2 norm of the interindividual difference in behavioral scores. In accordance with the previous study[14], the model fit of representational similarity analysis was defined as the pairwise Spearman's correlation coefficient between the representational dissimilarity matrices of PCs and behavioral assessments.

Statistical analysis

The significance of the correlation between linked dimensions of symptom and neurophysiology was assessed by conducting 1,000 permutations of the cross-validated sCCA step. Specifically, the clinical assessments were shuffled and assigned to random samples, resulting in mismatched neurophysiological and symptom profiles. The correlations between linked dimensions were evaluated using the Pearson's correlation coefficient. The representational similarity analysis used a Wilcoxon's signed-rank test to compare the information conveyed by subclinical, contrastive and PC features. The Wilcoxon's rank-sum test was used to assess the potential group difference between the sexes.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data supporting the results in this study are available within this paper and its supplementary materials. The ABC-CT dataset is publicly available via the National Institute of Mental Health Data Archive at https://nda.nih.gov/edit_collection.html?id=2288. The HBN dataset was obtained from the Child Mind Institute Biobank. The EEG data of the HBN dataset are publicly available through the International Neuroimaging Data-sharing Initiative at http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network. The phenotypical data of the HBN dataset can be accessed under a Data Usage Agreement with the Child Mind Institute Biobank, from the Longitudinal Online Research and Imaging System located at http://data.healthybrainnetwork.org/. The manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute.

Code availability

The cPCA step was implemented in Python (v.3.8.8) with a publicly available package (https://github.com/abidlabs/contrastive). The sCCA step was implemented in MATLAB (v.R2022a) using custom code. The statistical analyses were conducted in MATLAB with built-in functions. The code used in this study is available at https://github.com/Xiaoyu-Tong/Disorder-Specific-Symptom-Linked-rsEEG-FC-Dimension.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (2013).
2. Zheng, S., Hume, K. A., Able, H., Bishop, S. L. & Boyd, B. A. Exploring developmental and behavioral heterogeneity among preschoolers with ASD: a cluster analysis on principal components. *Autism Res.* **13**, 796–809 (2020).
3. An, J. Y. & Claudianos, C. Genetic heterogeneity in autism: from single gene to a pathway perspective. *Neurosci. Biobehav. Rev.* **68**, 442–453 (2016).
4. Hong, S.-J. et al. Toward neurosubtypes in autism. *Biol. Psychiatry* **88**, 111–128 (2020).
5. Wolfers, T. et al. From pattern classification to stratification: towards conceptualizing the heterogeneity of autism spectrum disorder. *Neurosci. Biobehav. Rev.* **104**, 240–254 (2019).
6. Hus, V., Gotham, K. & Lord, C. Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. *J. Autism Dev. Disord.* **44**, 2400–2412 (2014).
7. Kanne, S. M. et al. The Autism Impact Measure (AIM): initial development of a new tool for treatment outcome measurement. *J. Autism Dev. Disord.* **44**, 168–179 (2014).
8. Lord, C. et al. The autism diagnostic observation schedule–generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223 (2000).

9. Le Couteur, A, Lord, C & Rutter, M. *Autism Diagnostic Interview-Revised* (Western Psychological Services, 2003).
10. Constantino, J.N. & Gruber, C.P. *Social Responsiveness Scale: SRS-2* (Western Psychological Services, 2012).
11. Bölte, S., Poustka, F. & Constantino, J. N. Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). *Autism Res.* **1**, 354–363 (2008).
12. Xia, C. H. et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 3003 (2018).
13. Buch, A. M. et al. Molecular and network-level mechanisms explaining individual differences in autism spectrum disorder. *Nat. Neurosci.* **26**, 650–663 (2023).
14. Aglinskas, A., Hartshorne, J. K. & Anzellotti, S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science* **376**, 1070–1074 (2022).
15. Abid, A., Zhang, M. J., Bagaria, V. K. & Zou, J. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nat. Commun.* **9**, 2134 (2018).
16. Zhang, Y. et al. Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nat. Biomed. Eng.* **5**, 309–323 (2021).
17. Zhang, Y. et al. Machine learning-based identification of a psychotherapy-predictive electroencephalographic signature in PTSD. *Nat. Mental Health* **1**, 284–294 (2023).
18. Wu, W. et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nat. Biotechnol.* **38**, 439–447 (2020).
19. Nentwich, M. et al. Functional connectivity of EEG is subject-specific, associated with phenotype, and different from fMRI. *Neuroimage* **218**, 117001 (2020).
20. Mahjoory, K. et al. Consistency of EEG source localization and connectivity estimates. *Neuroimage* **152**, 590–601 (2017).
21. McPartland, J. C. et al. The autism biomarkers consortium for clinical trials (ABC-CT): scientific context, study design, and progress toward biomarker qualification. *Front. Integr. Neurosci.* **14**, 16 (2020).
22. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534 (2009).
23. Sparrow, S. S. & Cicchetti, D. V. *The Vineland Adaptive Behavior Scales* (Allyn & Bacon, 1989).
24. Alexander, L. M. et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* **4**, 170181 (2017).
25. Bodfish, J. W., Symons, F. J., Parker, D. E. & Lewis, M. H. Varieties of repetitive behavior in autism: comparisons to mental retardation. *J. Autism Dev. Disord.* **30**, 237–243 (2000).
26. Hooker, J. L., Dow, D., Morgan, L., Schatschneider, C. & Wetherby, A. M. Psychometric analysis of the repetitive behavior scale-revised using confirmatory factor analysis in children with autism. *Autism Res.* **12**, 1399–1410 (2019).
27. Gabriels, R. L., Cuccaro, M. L., Hill, D. E., Ivers, B. J. & Goldson, E. Repetitive behaviors in autism: relationships with associated clinical features. *Res. Dev. Disabil.* **26**, 169–181 (2005).
28. Wang, J. et al. Resting state EEG abnormalities in autism spectrum disorders. *J. Neurodev. Disord.* **5**, 24 (2013).
29. Rafiee, F., Rezvani Habibabadi, R., Motaghi, M., Yousem, D. M. & Yousem, I. J. Brain MRI in autism spectrum disorder: narrative review and recent advances. *J. Magn. Reson. Imaging* **55**, 1613–1624 (2022).
30. Kennedy, D. P. & Courchesne, E. The intrinsic functional organization of the brain is altered in autism. *Neuroimage* **39**, 1877–1885 (2008).
31. Wang, Z. et al. Resting-state brain network dysfunctions associated with visuomotor impairments in autism spectrum disorder. *Front. Integr. Neurosci.* **13**, 17 (2019).
32. Itahashi, T. et al. Alterations of local spontaneous brain activity and connectivity in adults with high-functioning autism spectrum disorder. *Mol. Autism* **6**, 30 (2015).
33. Yerys, B. E. et al. Neural correlates of set-shifting in children with autism. *Autism Res.* **8**, 386–397 (2015).
34. Xu, J. et al. Specific functional connectivity patterns of middle temporal gyrus subregions in children and adults with autism spectrum disorder. *Autism Res.* **13**, 410–422 (2020).
35. Samson, F., Mottron, L., Soulières, I. & Zeffiro, T. A. Enhanced visual functioning in autism: an ALE meta-analysis. *Hum. Brain Mapp.* **33**, 1553–1581 (2012).
36. Watanabe, T. & Rees, G. Anatomical imbalance between cortical networks in autism. *Sci. Rep.* **6**, 31114 (2016).
37. Puzzo, I., Cooper, N. R., Cantarella, S., Fitzgerald, P. B. & Russo, R. The effect of rTMS over the inferior parietal lobule on EEG sensorimotor reactivity differs according to self-reported traits of autism in typically developing individuals. *Brain Res.* **1541**, 33–41 (2013).
38. Casanova, M. F., Buxhoeveden, D. P., Switala, A. E. & Roy, E. Minicolumnar pathology in autism. *Neurology* **58**, 428–432 (2002).
39. Levitt, P. Disruption of interneuron development. *Epilepsia* **46**, 22–28 (2005).
40. Klimesch, W., Sauseng, P. & Hanslmayr, S. EEG alpha oscillations: the inhibition–timing hypothesis. *Brain Res. Rev.* **53**, 63–88 (2007).
41. Mathewson, K. E. et al. Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Front. Psychol.* **2**, 99 (2011).
42. Jensen, O. & Mazaheri, A. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* **4**, 186 (2010).
43. Limousin, P. & Foltynie, T. Long-term outcomes of deep brain stimulation in Parkinson disease. *Nat. Rev. Neurol.* **15**, 234–242 (2019).
44. Wagle Shukla, A. et al. Repetitive transcranial magnetic stimulation (rTMS) therapy in Parkinson disease: a meta-analysis. *PM R* **8**, 356–366 (2016).
45. Casanova, M. F. et al. Repetitive transcranial magnetic stimulation (rTMS) modulates event-related potential (ERP) indices of attention in autism. *Transl. Neurosci.* **3**, 170–180 (2012).
46. Sokhadze, E. M. et al. Prefrontal neuromodulation using rTMS improves error monitoring and correction function in autism. *Appl. Psychophysiol. Biofeedback* **37**, 91–102 (2012).
47. Enticott, P. G. et al. A double-blind, randomized trial of deep repetitive transcranial magnetic stimulation (rTMS) for autism spectrum disorder. *Brain Stimul.* **7**, 206–211 (2014).
48. Elliott, C. D. *Differential Ability Scales,* 2nd edn (The Psychological Corporation, 2007).
49. Fox, M. D. & Greicius, M. Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* **4**, 19 (2010).
50. Mullen, T. CleanLine: Tool/Resource Info. *NITRC* https://www.nitrc.org/projects/cleanline (2012).
51. Perrin, F., Pernier, J., Bertrand, O. & Echallier, J. F. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* **72**, 184–187 (1989).
52. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **7**, 1129–1159 (1995).
53. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **2011**, 879716 (2011).

54. Hauk, O. Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *Neuroimage* **21**, 1612–1621 (2004).
55. Gramfort, A., Papadopoulo, T., Olivi, E. & Clerc, M. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomed. Eng. Online* **9**, 45 (2010).
56. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
57. Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M. & Engel, A. K. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* **15**, 884–890 (2012).
58. Siems, M., Pape, A.-A., Hipp, J. F. & Siegel, M. Measuring the cortical correlation structure of spontaneous oscillatory activity with EEG and MEG. *Neuroimage* **129**, 345–355 (2016).
59. Rolle, C. E. et al. Cortical connectivity moderators of antidepressant vs placebo treatment response in major depressive disorder: secondary analysis of a randomized clinical trial. *JAMA Psychiatry* **77**, 397–408 (2020).
60. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis-connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

Acknowledgements

This work was supported by National Institutes of Health grant nos. R21MH130956, R01MH129694 and R21AG080425 to Y.Z., grant R01MH110512 to H.X., grants K23MH114023 and R01MH125886 to G.A.F., grants. R37MH125829, R01EB022573, R01MH112847 and R01MH120482 to T.D.S., as well as Lehigh University FIG (FIGAWD35), CORE and Accelerator grants to Y.Z. Portions of this research were conducted on the Lehigh University's Research Computing infrastructure, which is partially supported by a National Science Foundation award 2019035. This work was also supported in part by philanthropic funding and grants from the One Mind-Baszucki Brain Research Fund, the SEAL Future Foundation and the Brain and Behavior Research Foundation to G.A.F. The funders had no role in the design and conduct of the study, and the collection, management, analysis and interpretation of the data, nor were they involved in the decision to submit the manuscript for publication.

Author contributions

X.T. conceptualized and designed the work, wrote the code, analyzed and interpreted the data, and drafted and revised the manuscript. H.X., G.A.F. and K.Z. interpreted the data and refined the design of the work. T.D.S. interpreted the data and revised the manuscript. N.B.C. conceptualized the work and revised the manuscript. Y.Z. conceptualized and designed the work, oversaw the analysis and interpretation of the data, and revised the manuscript.

Competing interests

G.A.F. received monetary compensation for consulting work for SynapseBio AI and owns equity in Alto Neuroscience. None of the other authors declare any competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-023-00195-w.

Correspondence and requests for materials should be addressed to Yu Zhang.

Peer review information *Nature Mental Health* thanks Janine Bijsterbosch, Guillaume Dumas and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Yu Zhang

Last updated by author(s): Dec 6, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No custom code was used for data collection.
Data analysis	EEG data were preprocessed using the EEGLAB-based toolbox (version 2019.1). EEG source localization was performed using Brainstorm (version 3) with the OpenMEEG (version 2.4) plugin. Other data analyses were conducted using a publicly available package (https://github.com/abidlabs/contrastive) in Python v3.8.8 and custom code in MATLAB R2022a. A code availability section is available in the manuscript. The code used in this study is available at https://github.com/Xiaoyu-Tong/Disorder-Specific-Symptom-Linked-rsEEG-FC-Dimension.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data supporting the results in this study are available within this paper and its Supplementary Materials. The ABC-CT dataset is publicly available via the National Institute of Mental Health Data Archive (NDA) at https://nda.nih.gov/edit_collection.html?id=2288. The HBN dataset was obtained from the Child Mind Institute Biobank. The EEG data of the HBN dataset are publicly available through the International Neuroimaging Data-sharing Initiative at http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network. The phenotypical data of the HBN dataset can be accessed under a Data Usage Agreement with the Child Mind Institute Biobank, from Longitudinal Online Research and Imaging System (LORIS) located at http://data.healthybrainnetwork.org/. The manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We have confirmed that the findings from this study is general for both genders. We aimed for discovering brain patterns predictive of treatment outcomes that are general to sex, therefore we did not conduct separate analyses on different sex.
Population characteristics	The ABC-CT dataset comprised 280 children with ASD and 119 TD children whose ages ranged from 6 to 11. Each subject contained 1 to 3 EEG recordings and psychiatric assessments, resulting in 703 ASD samples and 364 TD samples. Due to low EEG data quality, 166 samples were excluded from our analysis (see EEG Preprocessing for details). Another 270 samples were also excluded because of incomplete psychiatric assessments, yielding 392 ASD samples and 239 TD samples that were included in our study. Diagnosis was made using the ADOS. The Full-Scale Intelligence Quotient of subjects spanned from 60 to 150 as measured by Differential Ability Scales. Exclusion criteria included known genetic syndrome, neurological condition putatively causally related to ASD, known metabolic disorder, and mitochondrial dysfunction. Medications were allowed to ensure the sample is representative to subjects with treatment for better generalizability of biomarker measurement, while a stable regimen (at least 8 weeks prior to enrollment) was required for inclusion. The HBN study recruited children and adolescents between 5–21 at four study sites in the New York City area. The recruitment criterion was having adequate verbal communication ability with the help of parents or guardians. Exclusion criteria included severe neurological disorders, severe impairment in cognition (IQ<66), acute encephalopathy, known neurodegenerative disorder, or other abnormalities that may prevent full participation in the protocol. Overall, HBN incorporated a variety of psychiatric disorders, and we focused on ASD in this study. At the time we accessed the dataset, it included 185 TD subjects and 222 ASD patients (regardless of comorbidity) with available behavior measures and neuroimaging data (one for each subject).
Recruitment	See above.
Ethics oversight	ABC-CT: The studies involving human participants were reviewed and approved by the Yale Institutional Review Board. HBN: The study was approved by the Chesapeake Institutional Review Board (https://www.chesapeakeirb.com/).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study was designed to be sufficiently powered to detect associations between brain connectivity features and symptom profiles in ASD patients. A sample size of 34 can provide a sufficient power of 0.8 with significance level of 0.05 to detect a modest correlation of 0.4 between connectivity features and symptoms. The discovery study recruited 280 ASD patients and 119 TD subjects and collected data at 3 time points in 24 weeks. Among them, 392 ASD and 239 TD recordings are with usable EEG data and available behavior assessments. We seek an independent dataset for generalizability analysis cluded the 223 ASD patients with usable EEG data and available behavior assessments.

Data exclusions	Subjects with missing symptom measures, low-quality of rsEEG were excluded.
Replication	The reproducibility of findings successfully confirmed with five rounds of 10-fold cross-validation. A conceptual generalization analysis successfully validates the representation of dimensions on an independent dataset.
Randomization	No randomization was performed since no treatment is administered.
Blinding	The clinical trial does not involve treatment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Autism Biomarker Consortium for Clinical Trials (ABC-CT), NCT02996669
Study protocol	https://doi.org/10.3389/fnint.2020.00016
Data collection	The ABC-CT studies employed four EEG paradigms, including resting-state (viewing abstract videos), human face event-related, biological motion event-related, and visual evoked. In this study, we focused on the rsEEG features to establish neuroimaging biomarkers for ASD symptoms in naturalistic condition. Because the eyes-closed session may not be reliably recorded for the subject age range (6-11), only eyes-open sessions were collected. rsEEG data were recorded using a 128-channel EEG geodesic hydrocel system by Electrical Geodesics Inc. (EGI). The recordings were at a sampling rate of 1000 Hz. The recording reference was set to be Cz (head vertex) and the impedance of each electrode was checked before recording and was kept below 100 kΩ. Data were collected at Yale University, Duke University, Boston Children's Hospital, University of California, Los Angeles (UCLA), and University of Washington with the aforementioned protocols. Written informed consent was provided by the legal guardian of each participant. Monetary compensation was provided for each participant. The data were collected between October 2016 and May 2019.
Outcomes	Primary outcome was defined as understanding the association between brain activity and behavioral assessments in ASD patients. ASD symptom profile was measured by Autism Diagnostic Observation Schedule (ADOS), Autism Impact Measure (AIM), Social Responsiveness Scale (SRS), and Vineland Adaptive Behavior Scale (VABS) at three time points at baseline, 6 week, and 24 week. No treatment was administered.